Zusammenfassung der Vorlesung

Einführung in die Wahrscheinlichkeitstheorie

Margherita Disertori
Universität Bonn
Wintersemester 2025-2026

Dies ist eine gekürzte Zusammenfassung und *kein* vollständiges Skript der Vorlesung. Deshalb kann diese Zusammenfassung ein Lehrbuch *nicht* ersetzen. Folgende Bücher werden empfohlen:

[G] Hans-Otto Georgii, Einführung in die Wahrscheinlichkeitstheorie und Statistik, de Gruyter Verlag

[K] Achim Klenke, Wahrscheinlichkeitstheorie, Springer 2006

[F] William Feller, An introduction to probability and its applications Vol. 1., John Wiley, 1978

Insbesondere enthalten diese Notizen nur ausgewählte Beweise. Die anderen sind entweder leichte Übungsaufgaben oder in der angegebenen Literatur zu finden. Diese Zusammenfassung basiert auf den oben genannten Büchern und einer Vorlesungszusammenfassung von Prof. A. Bovier, Prof. A. Eberle und Prof. P. Ferrari und ist nur für Hörer der Vorlesung V2F1 Einführung in die Wahrscheinlichkeitstheorie an der Universität Bonn, Wintersemester 2025-2026, bestimmt.

Empfehlung: Die Vorlesung Analysis III befasst sich mit Maßtheorie und ergänzt diese Vorlesung gut.

Hinweise auf Tippfehler und Korrekturen bitte an `mdiserto@uni-bonn.de`

Inhaltsverzeichnis

Notation. Wir werden folgende Notationen verwenden:

$$\mathbb{N} = \{1, 2, \dots\}, \quad \mathbb{N}_0 = \{0, 1, 2, \dots\}, \quad \mathbb{R}_+ = \mathbb{R}_{>0} = (0, \infty), \quad \mathbb{R}_{\geq 0} = [0, \infty).$$

Wir werden die Notation $\subset$ für eine Teilmenge und $\subsetneq$ für eine strikte Teilmenge verwenden.

1 Wahrscheinlichkeiten modellieren

1.1 Einleitung

Um ein Zufallsexperiment zu modellieren, brauchen wir drei Zutaten: Die Menge Ω der möglichen Ergebnisse, das Mengensystem der möglichen Ereignisse Er und eine Funktion $\mathbb{P}\colon\ Er\ \to [0,1]$, die jedem Ereignis $\tilde{E} \in Er$ die Wahrscheinlichkeit $\mathbb{P}(\tilde{E})$ zuordnet.
Das Mengensystem der möglichen Ereignisse muss u.a. folgende Eigenschaften haben:

- $\tilde{\Omega}$:="ein beliebiges $\omega \in \Omega$ tritt ein" und $\tilde{\varnothing}$:="kein $\omega \in \Omega$ tritt ein" müssen in Er liegen.
- Falls $\tilde{E}_1, \tilde{E}_1 \in Er$ gilt, dann gilt $\tilde{E}_1 \wedge \tilde{E}_1 \in Er$, wobei $\tilde{E}_1 \wedge \tilde{E}_1$ genau dann eintritt, wenn $\tilde{E}_1$ und $\tilde{E}_2$ eingetreten sind.
- Falls $\tilde{E}_1, \tilde{E}_1 \in Er$ gilt, dann gilt $\tilde{E}_1 \vee \tilde{E}_1 \in Er$, wobei $\tilde{E}_1 \vee \tilde{E}_1$ genau dann eintritt, wenn $\tilde{E}_1$ oder $\tilde{E}_2$ eingetreten ist.
- Falls $\tilde{E} \in Er$ gilt, dann gilt $\neg\tilde{E} \in Er$, wobei $\neg\tilde{E}$ genau dann eintritt, wenn $\tilde{E}$ nicht eingetreten ist.

Man kann Ereignisse durch Teilmengen von Ω wie folgt darstellen: Für $\tilde{E} \in Er$ definieren wir

$$\tilde{E}(E) := \{\omega \in \Omega \,|\ \text{falls } \omega \text{ eintritt, dann ist } \tilde{E} \text{ eingetreten }\}.$$

Das Mengensystem der möglichen Ereignissen Er wird damit eine Familie von Teilmengen $\mathcal{F} \subset \mathcal{P}(\Omega)$. Insbesondere gilt

$$E(\tilde{E}_1 \wedge \tilde{E}_1) = E_1 \cup E_2, \quad E(\tilde{E}_1 \vee \tilde{E}_1) = E_1 \cap E_2, \quad E(\neg\tilde{E}) = E^c = \Omega \setminus E.$$

Die obigen Eigenschaften sind durch das Konzept der $\sigma-$Algebra kodiert (siehe Definition 1.1).

Das Wahrscheinlichkeitmaß muss also eine Funktion $\mathbb{P}\colon \mathcal{F} \to [0,1]$ sein, die jedem $E \in \mathcal{F}$ die Wahrscheinlichkeit $\mathbb{P}(E)$ zuordnet. Diese Funktion muss u.a. folgende Eigenschaften haben:

- $\mathbb{P}(\Omega) = 1$ (das Ereignis "irgendein $\omega \in \Omega$ tritt" ist sicher),
- $\mathbb{P}(\varnothing) = 0$ (das Ereignis "kein $\omega \in \Omega$ tritt ein" ist unmöglich),
- Falls E_1 und E_2 unvereinbar (disjunkt) (d.h. $E_1 \cap E_2 = \varnothing$) sind, gilt $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$.

Beispiel. Wie werfen zwei fairen Münzen. Der Spieler 1 (bzw. 2) gewinnt, falls zwei Mal Kopf (bzw. Zahl) geworfen wird. Die Ergebnismenge Ω hat daher vier Elemente,

$$\Omega = \{(K,K),(K,Z),(Z,K),(Z,Z)\}.$$

Mögliche Ereignisse sind z.B.

$$A := \text{“Spieler 1 gewinnt”}, \quad B := \text{“Spieler 2 gewinnt”}, \quad C := \text{“niemand gewinnt”}.$$

- Falls das Ergebnis des Zufallsexperiments $\omega = (K,K)$ ist, dann sagen wir, dass A eingetreten ist (B, C sind nicht eingetreten)
- Falls das Ergebnis des Zufallsexperiments $\omega = (Z,Z)$ ist, dann sagen wir, dass B eingetreten ist (A, C sind nicht eingetreten)
- Falls das Ergebnis des Zufallsexperiments $\omega \in \{(K,Z),(Z,K)\}$ erfüllt, dann sagen wir, dass C eingetreten ist (A, B sind nicht eingetreten).

Da die Münzen fair sind, können wir annehmen, dass alle Ergebnisse gleichwahrscheinlich sind, d.h.

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{4} \qquad \forall \omega \in \Omega.$$

Daher rechnen wir

$$\mathbb{P}(A) = \mathbb{P}(\{(K,K)\}) = \frac{1}{4} = \mathbb{P}(\{(Z,Z)\}) = \mathbb{P}(B).$$

Da $\{(K,Z)\}$ und $\{(Z,K)\}$ unvereinbar sind, gilt auch

$$\mathbb{P}(C) = \mathbb{P}(\{(K,Z),(Z,K)\}) = \mathbb{P}(\{(K,Z)\}) + \mathbb{P}(\{(Z,K)\}) = \frac{1}{2}.$$

1.2 Endliche Ergebnismenge: wichtige Beispiele

Beispiel I. Wir verteilen $K \geq 1$ Kugeln auf $N \geq 1$ Fächer.
Für $i = 1, \ldots, K$, setzen wir $\omega_i = j$, falls Kugel i im Fach j liegt. Eine Konfiguration (Ergebnis) ist also ein $K-$Tupel $\omega = (\omega_1, \ldots, \omega_K) \in \{1, \ldots, N\}^K$ und die Ergebnismenge ist

$$\Omega_I := \Omega_I(N,K) := \{\omega = (\omega_1, \ldots, \omega_K) \in \{1, \ldots, N\}^K\}.$$

Es gilt $|\Omega_I(N,K)| = N^K$.
Im Beispiel von zwei Münzen oben kann man die zwei Münzen durch zwei Kugeln und die Werte K, Z durch zwei Fächer darstellen. Dann entspricht z.B. das Ergebnis (K,K) dem Ergebnis $(\omega_1, \omega_2) = (1,1)$.

Beispiel II. Wir verteilen $K \geq 1$ Kugeln auf $N \geq 1$ Fächer unter der Bedingung, dass jedes Fach *maximal eine* Kugel enthalten darf.
Die Ergebnismenge ist

$$\Omega_{II} := \Omega_{II}(N,K) := \{\omega \in \Omega_I \mid \omega_i \neq \omega_j \, \forall i \neq j\}.$$

Falls $K > N$, gilt $\Omega_{II}(N,K) = \varnothing$ und $|\Omega_{II}(N,K)| = 0$.
Falls $K \leq N$ ist $\Omega_{II}(N,K)$ nicht leer und

$$|\Omega_{II}(N,K)| = N(N-1)\cdots(N-K+1) = \frac{N!}{(N-K)!}.$$

Im Spezialfall $N = K$ haben wir genau eine Kugel pro Fach und

$$|\Omega_{II}(N, N)| = N!,$$

was der Anzahl von Permutationen von N Elementen entspricht.

Beispiel III. Wir verteilen $K \geq 1$ Kugeln auf $N \geq 1$ Fächer.
Jedes Fach darf *maximal eine* Kugel enthalten. Die Kugeln sind *ununterscheidbar.*
In diesem Fall sind z.B. die Konfigurationen $(1, 2, 3, 4)$ und $(2, 1, 3, 4)$ ununterscheidbar.
Um den Ergebnisraum Ω zu beschreiben, definieren wir die Äquivalenzrelation $\sim$ auf Ω, sodass $\omega \sim \omega'$, falls eine Permutation $\sigma \in S_K$ existiert, sodass $\omega'_i = \omega_{\sigma(i)}$ für alle $i \in \{1, \dots, K\}$. Damit definieren wir

$$\Omega_{III} \coloneqq \Omega_{III}(N, K) \coloneqq \{[\omega] \mid \omega \in \Omega_{II}\}.$$

Jede Äquivalenzklasse enthält $K!$ Elemente und es folgt

$$|\Omega_{III}(N, K)| = \frac{|\Omega_{II}(N, K)|}{K!} = \frac{N!}{(N - K)!K!} = \binom{N}{K}.$$

Beispiel IV. Wir verteilen $K \geq 1$ Kugeln auf $N \geq 1$ Fächer.
Die Kugel sind *unterscheidbar* und die Anzahl der *Kugeln pro Fach* ist gegeben.
Für jede Konfiguration $\omega \in \Omega_I$ definieren wir $N_l(\omega)$ als die Anzahl der Kugeln, welche im Fach $l = 1, \dots, N$ liegen. Es gilt

$$0 \leq N_l(\omega) \leq K \text{ und } \sum_{l=1}^{N} N_l(\omega) = K.$$

Seien $n_1, \dots, n_N \in \{0, 1, \dots, N\}$,, sodass $\sum_{l=1}^{N} n_l = K$. Wir definieren

$$\Omega_{IV} \coloneqq \Omega_{IV}(N, K, n_1, \dots, n_N) \coloneqq \{\omega \in \Omega_I \mid N_l(\omega) = n_l \ \forall l = 1, \dots, N\}.$$

Es gilt

$$|\Omega_{IV}| = \binom{K}{n_1}\binom{K - n_1}{n_2} \cdots \binom{K - (n_1 + \cdots + n_{N-1})}{n_N} = \frac{K!}{\prod_{l=1}^{N} n_l!} =: \binom{K}{n_1, n_2, \cdots, n_N},$$

wobei $(K - (n_1 + \cdots + n_N))! = (K - K)! = 0! = 1$ angewendet wurde.

Beispiel V. Wir verteilen $K \geq 1$ Kugeln auf $N \geq 1$ Fächer.
Die Kugel sind *unterscheidbar* und *die Kugeln im Fach sind auch geordnet.*
Um ein Ergebnis zu beschreiben, ist es nicht genug zu sagen, in welchem Fach die Kugel i liegt.
Wir brauchen auch eine Ordnung, welche die Anordnung der $N_l(\omega)$ Kugeln im Fach l beschreibt.
Eine Konfiguration ist eindeutig durch das Paar (σ, s) beschrieben, wobei:

- $\sigma = (\sigma(1), \dots \sigma(K)) \in S_K$ ist eine Permutation der Kugeln, also eine Ordnung,
- $s = (s_0, \dots, s_N)$ ist eine Familie von Schranken, sodass

$$0 = s_0 \leq s_1 \leq s_2 \leq \cdots \leq s_{N-1} \leq s_N = N.$$

Das Fach F_l, $l = 1, \dots N$ enthält die geordneten Kugeln $(\sigma(s_{l-1}+1), \sigma(s_{l-1}+2) \dots, \sigma(s_l))$. Falls $s_{l-1} = s_l$, so ist das Fach F_l leer. Die Ergebnismenge ist

$$\Omega_V := \Omega_V(N, K) := \{(\sigma, s) \mid \sigma \in S_K, 0 = s_0 \leq s_1 \leq s_2 \leq \cdots \leq s_{N-1} \leq s_N = K\}.$$

Es gilt

$$|\Omega_V| = \frac{(K+N-1)!}{(N-1)!},$$

wobei $(K+N-1)!$ der Anzahl der Permutationen von $(K+N-1)$ Symbolen entspricht und wir teilen durch $(N-1)!$, da die $N-1$ Schranken $s_1, \dots, s_{N-1}$ durch die Bedingung $0 \leq s_1 \leq s_2 \leq \cdots \leq s_{N-1} \leq N$ geordnet sind.
Wir merken, dass

$$\begin{aligned} |\Omega_V| = \frac{(K+N-1)!}{(N-1)!} &= (N+K-1)(N+K-2) \cdots (N+K-K) \\ &= N^K \left(1 + \frac{K-1}{N}\right)\left(1 + \frac{K-2}{N}\right) \cdots 1 > N^K = |\Omega_I|. \end{aligned}$$

[1: 14.10.2025]

[2: 17.10.2025]

1.3 Messbare Mengen und Wahrscheinlichkeitsmaße

1.3.1 $\sigma-$Algebra

Definition 1.1 (σ-Algebra)**.** *Sei Ω eine nichtleere Menge.*
Eine Menge $\mathcal{F} \subset \mathcal{P}(\Omega)$ heißt eine σ-Algebra *über Ω, falls folgendes gilt:*

(i) $\mathcal{F}$ ist nichtleer;

(ii) Für alle $A \in \mathcal{F}$ gilt $A^c := X \setminus A \in \mathcal{F}$;

(iii) Für jede Folge $A\colon \mathbb{N} \to \mathcal{F}$ gilt

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}\,. \tag{1.1}$$

Bemerkung 1. Die Elemente in $\mathcal{F}$ werden Ereignisse oder "messbare Mengen" genannt.

Bemerkung 2. Die Potenzmenge $\mathcal{P}(\Omega)$ ist eine $\sigma-$Algebra. Für unendliche (überabzählbare) Mengen Ω ist diese aber oftmals zu groß, um vernünftige Wahrscheinlichkeitverteilungen zu definieren.

Lemma 1.2 (Wichtige Eigenschaften)**.** *Sei Ω eine nichtleere Menge.*

(i) Sei $\mathcal{F}$ eine σ-Algebra über Ω. Folgendes gilt:

  (a) Falls $A, B \in \mathcal{F}$, so folgt: $A \cup B$ und $A \setminus B \in \mathcal{F}$;

  (b) $\Omega, \varnothing \in \mathcal{F}$;

  (c) Falls $A_j \in \mathcal{F}$ $\forall j \in \mathbb{N}$ gilt, so folgt: $\bigcap_{j \in \mathbb{N}} A_j \in \mathcal{F}$.

(ii) Sei I eine beliebige Indexmenge und für jedes $\alpha \in I$ sei $\mathcal{F}_\alpha \subset \mathcal{P}(\Omega)$ eine σ-Algebra. Folgendes gilt:

   (a) Der Durchschnitt $\mathcal{F} \coloneqq \bigcap_{\alpha\in I} \mathcal{F}_\alpha$ ist eine σ-Algebra.

   (b) Die Vereinigung $\bigcup_{\alpha\in I} \mathcal{F}_\alpha$ ist im Allgemeinen keine σ-Algebra.

Beweis. Übung. □

Definition 1.3. *Sei Ω eine nichtleere Menge und $\mathcal{E} \subset \mathcal{P}(\Omega)$ ein Mengensystem.*

(i) Die von $\mathcal{E}$ erzeugte $\sigma-$Algebra über Ω ist die kleinste $\sigma-$Algebra, die $\mathcal{E}$ enthält und definiert durch

$$\sigma(\mathcal{E}) \coloneqq \bigcap_{\mathcal{F} \supset \mathcal{E},\ \mathcal{F}\ \sigma\text{-Algebra}} \mathcal{F}.$$

$\sigma(\mathcal{E})$ ist eine $\sigma-$Algebra nach Lemma 1.2 (ii)(a).

(ii) Sei $\mathcal{F}$ eine $\sigma-$Algebra über Ω. Das Mengensystem $\mathcal{E}$ ist ein Erzeuger (Generator) von $\mathcal{F}$, falls $\mathcal{F} = \sigma(\mathcal{E})$.

Beispiel Sei $\Omega = \{1, \ldots, N\}$, $N > 2$ und $\mathcal{F} = \mathcal{P}(\Omega)$.

(i) Das Mengensystem $\mathcal{E}_1 \coloneqq \{\{1\}, \ldots, \{N\}\}$ erzeugt $\mathcal{F}$, d.h. $\sigma(\mathcal{E}_1) = \mathcal{F}$.

(ii) Sei $\mathcal{E}_2 \coloneqq \{A, B\}$ wobei

$$A \coloneqq \{j \in \{1, \ldots, N\} |\ j \text{ gerade}\}, \qquad B \coloneqq \{j \in \{1, \ldots, N\} |\ j \text{ ungerade}\}.$$

Dann gilt $\sigma(\mathcal{E}_2) = \{A, B, \Omega, \varnothing\} \subsetneq \mathcal{F}$.

1.3.2 Maß

Definition 1.4 (Maß)**.** *Sei $\Omega \neq \varnothing$ eine Menge und $\mathcal{F}$ eine σ-Algebra über Ω.*
Eine Funktion $\mu\colon \mathcal{F} \to [0, \infty]$ heißt ein Maß *auf $\mathcal{F}$, falls*

(i) $\mu(\varnothing) = 0$ und

(ii) μ ist σ-additiv, das heißt, für alle Folge $A\colon \mathbb{N} \to \mathcal{F}$ mit $A_j \cap A_k = \varnothing\ \forall j \neq k$ gilt

$$\mu\left(\bigcup_{j\in\mathbb{N}} A_j\right) = \sum_{j\in\mathbb{N}} \mu(A_j)\,. \tag{1.2}$$

- *Das Tripel $(X, \mathcal{F}, \mu)$ heißt* Maßraum.
- *Das Maß μ heißt* endlich, *falls $\mu(\Omega) < \infty$.*
- *Das Maß μ heißt $\sigma-$*endlich, *falls eine Folge $A\colon \mathbb{N} \to \mathcal{F}$ existiert, sodass $\Omega = \cup_{n\in\mathbb{N}} A_n$, $\mu(A_n) < \infty$ und $A_n \subset A_{n+1}\ \forall n \geq 1$.*
- *Das Maß μ heißt* Wahrscheinlichkeitsmaß *(oder* Wahrscheinlichkeitsverteilung*), falls gilt $\mu(\Omega) = 1$. In diesem Fall heißt das Tripel $(X, \mathcal{F}, \mu)$* Wahrscheinlichkeitsraum.
- *Eine Menge $N \subset \Omega$ heißt μ-*Nullmenge *(oder auch einfach* Nullmenge*), falls $N \in \mathcal{F}$ und $\mu(N) = 0$.*

Wir werden meist die Notation $\mathbb{P}$ statt μ für ein Wahrscheinlichkeitsmaß anwenden.

Bemerkungen

1. Aus der $\sigma-$Additivität folgt unmittelbar, dass μ auch <u>additiv</u> ist. Das heißt, für alle $A_0, \ldots, A_k \in \mathcal{F}$ paarweise disjunkte Teilmenge gilt

$$\mu\left(\bigcup_{j=0}^{k} A_j\right) = \sum_{j=0}^{k} \mu(A_j)\,.$$

2. Die Funktion μ ist monoton: $A, B \in \mathcal{F}$, mit $A \subset B \Rightarrow \mu(A) \leq \mu(B)$. (Übungsblatt)

3. Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß, d.h. $\mathbb{P}(\Omega) = 1$.

- $\forall A \in \mathcal{F}$ gilt $A \subset \Omega$ und deshalb (da $\mathbb{P}$ monoton ist) gilt $0 \leq \mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$
- $\forall A\colon \mathbb{N} \to \mathcal{F}$ mit $A_j \cap A_k = \varnothing\ \forall j \neq k$ gilt

$$0 \leq \sum_{j\in\mathbb{N}} \mathbb{P}(A_j) = \mathbb{P}\left(\bigcup_{j\in\mathbb{N}} A_j\right) \leq 1. \tag{1.3}$$

Beispiele

1. Sei $\Omega = \mathbb{N}$ und $\mathcal{F} = \mathcal{P}(\Omega)$. Wir definieren $\forall A \subset \mathbb{N}\ \mu(A) := |A| = \#A$. Dann ist μ ein Maß (das Zählmaß). μ ist $\sigma-$endlich, aber nicht endlich.

2. Sei Ω eine endliche Menge $|\Omega| = \#\Omega < \infty$, und $\mathcal{F} = \mathcal{P}(\Omega)$. Wir definieren $\forall A \subset \Omega\ \mu(A) := \frac{|A|}{|\Omega|}$. Insbesondere gilt $\mu(\{\omega\}) = \frac{1}{|\Omega|}\ \forall \omega \in \Omega$. Dann ist μ ein Wahrscheinlichkeitsmaß (die Gleichverteilung).

Satz 1.5 (Maß als Grenzwert)**.**
Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum, und sei $A\colon \mathbb{N} \to \mathcal{F}$ eine Folge messbarer Mengen. Folgende Aussagen gelten.

(i) Falls die Folge monoton wachsend ist ($A_j \subset A_{j+1}\ \forall j \in \mathbb{N}$), dann folgt

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{j\to\infty} \mu(A_j)\,. \tag{1.4}$$

(ii) Falls die Folge monoton fallend ist, ($A_j \supset A_{j+1}\ \forall j \in \mathbb{N}$), <u>und</u> $\mu(A_1) < \infty$, dann folgt

$$\mu\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{j\to\infty} \mu(A_j)\,. \tag{1.5}$$

Beweis. Übungsblatt □

Bemerkung. Die Annahme $\mu(A_1) < \infty$ gilt automatisch, falls μ endlich ist.

Lemma 1.6 (Subadditivität). *Sei* $(\Omega, \mathcal{F}, \mu)$ *ein Maßraum.*
Für alle Folgen $A\colon \mathbb{N} \to \mathcal{F}$ *gilt*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \le \sum_{n\in\mathbb{N}} \mu(A_n)\,.$$

Beweis. Da $\mathcal{F}$ eine $\sigma-$Algebra ist, gilt $\bigcup_{n\in\mathbb{N}} A_n \in, \mathcal{F}$ also ist $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right)$ wohldefiniert. Wir konstruieren eine neue Folge $B\colon \mathbb{N} \to \mathcal{F}$ wie folgt:

$$B_1 \coloneqq A_1, \qquad B_n \coloneqq A_n \setminus \left(\bigcup_{j=1}^{n-1} A_j\right) \qquad \forall n \ge 2.$$

Folgendes gilt.

- $B_n \in \mathcal{F}$ $\forall n$ folgt aus Lemma 1.2 (a) und Def. 1.1.
- $B_j \cap B_k = \varnothing$ $\forall j \neq k$ und $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$ (Übung).
- Wir rechnen

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) \overset{\mu\ \sigma-\text{Add.}}{=} \sum_{n\in\mathbb{N}} \mu(B_n) \overset{\mu\ \text{monot.}}{\le} \sum_{n\in\mathbb{N}} \mu(A_n).$$

Damit ist der Beweis beendet. □

[2: 17.10.2025]

[3: 21.10.2025]

1.3.3 Wahrscheinlichkeitsräume: wichtige Beispiele

1. Dirac Maß. Sei Ω beliebig und $a \in \Omega$ ein festes Element. Das *Dirac-Maß* in a ist die durch

$$\delta_a[A] \coloneqq I_A(a) = \begin{cases} 1 & \text{falls } a \in A, \\ 0 & \text{sonst,} \end{cases} \qquad \forall A \subset \Omega$$

definierte Wahrscheinlichkeitsverteilung $\mathbb{P} = \delta_a$ auf der σ-Algebra $\mathcal{F} = \mathcal{P}(\Omega)$. Dirac-Maße sind ‚deterministische Verteilungen" – es gilt

$$\delta_a[\{a\}] = 1, \qquad \delta_a[\{a\}^c] = 0.$$

2. Konvexkombinationen von Dirac-Maßen. Ist C eine abzählbare Teilmenge von Ω, und $p\colon C \to [0,1]$ eine Gewichtsfunktion mit $\sum_{\omega\in C} p(\omega) = 1$, dann ist durch

$$\mathbb{P}[A] \coloneqq \sum_{a\in A\cap C} p(a) = \sum_{a\in C} p(a)\,\delta_a[A] \quad \forall A \subset \Omega$$

ein eindeutiges Wahrscheinlichkeitsmaß $\mathbb{P}$ auf der Potenzmenge $\mathcal{F} = \mathcal{P}(\Omega)$ gegeben. Die Verteilung $\mathbb{P}$ ist ‚rein atomar", d.h. die Masse sitzt auf abzählbaren vielen ‚Atomen" (den Elementen von C). Jede diskrete Wahrscheinlichkeitsverteilung ist von dieser Form.

Beispiel. Wir werfen n faire Münzen. Wir stellen das Ergebnis 'Kopf' mit 0 und 'Zahl' mit 1 dar. Die Ergebnismenge ist also

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) \in \{0,1\}^n\}.$$

Insbesondere gilt $|\Omega| = 2^n$. Sei $\mathcal{F} = \mathcal{P}(\Omega)$. Die Gleichverteilung $\mathbb{P}(A) := \frac{|A|}{|\Omega|}$ ist eine Konvexkombination von Dirac-Maßen:

$$\mathbb{P} = \sum_{\omega \in \Omega} \frac{1}{|\Omega|} \delta_{\{\omega\}}.$$

In diesem Fall gibt es $|\Omega| = 2^n$ Atome ($C = \Omega$) und die Gewichtsfunktion ist durch $p(\{\omega\}) := \frac{1}{|\Omega|}$ definiert.

3. Unendliches Produktmodell (z.B. Münzwurffolge). Im einfachsten Fall möchten wir eine Folge unabhängiger, fairer Münzwürfe (0-1-Experimente) modellieren. Die Ergebnismenge ist also

$$\Omega \;=\; \{\omega = (\omega_1, \omega_2, \ldots) \in \{0,1\}^{\mathbb{N}}\}.$$

Ω ist überabzählbar!
Wir suchen eine σ-Algebra $\mathcal{F}$ und ein Wahrscheinlichkeitsmaß $\mathbb{P}$, sodass Folgendes gilt:
$\forall n \in \mathbb{N}$ und $x_1, \ldots, x_n \in \{0,1\}$

- ist die Menge (das Ergebnis für die ersten n Münzwurfen ist gegeben)

  $$A(x_1, \ldots, x_n) := \{\omega \in \Omega \mid (\omega_1, \ldots, \omega_n) = (x_1, \ldots, x_n)\}$$

  messbar und es gilt

- $\mathbb{P}(A(x_1, \ldots, x_n)) = \frac{1}{2^n}$ (Gleichverteilung für die ersten n Münzen).

Für $n \geq 1$ und $A \in \mathcal{P}(\{0,1\}^n)$ sei

$$C_{n,A} := \{\omega \in \Omega | \; (\omega_1, \ldots, \omega_n) \in A\}. \tag{1.6}$$

$C_{n,A}$ heißt Zylindermenge. Sei

$$\mathcal{C} := \{C_{n,A} \mid n \geq 1, A \in \mathcal{P}(\{0,1\}^n)\} \tag{1.7}$$

die Menge der Zylindermengen. Wir setzen $\mathcal{F} := \sigma(\mathcal{C})$. $\mathcal{F}$ heißt Produkt-$\sigma-$Algebra und es gilt $\mathcal{F} \subsetneq \mathcal{P}(\Omega)$.
Wir werden im Abschnitt 1.3.5 sehen, dass genau ein Wahrscheinlichkeitsmaß $\mathbb{P}\colon \sigma(\mathcal{C}) \to [0,1]$ existiert, sodass $\mathbb{P}(C_{n,A}) = \frac{|A|}{2^n}$ $\forall n \geq 1, A \in \mathcal{P}(\{0,1\}^n)$ gilt.

Definition 1.7 (**Produkt-$\sigma-$Algebra**). *Sei I eine beliebige Indexmenge und für jedes $i \in I$ sei Ω_i nichtleer und $\mathcal{F}_i \subset \mathcal{P}(\Omega_i)$ eine σ-Algebra über Ω_i. Wir betrachten die Ergebnismenge*

$$\Omega = \prod_{i \in I} \Omega_i := \{\omega = (\omega_i)_{i \in I} | \; \omega_i \in \Omega_i\}.$$

Sei $\mathcal{C}$ die Kollektion aller Zylindermengen von der Form

$$\{\omega = (\omega_i)_{i\in I} \in \Omega \mid \omega_{i_1} \in A_{i_1}, \omega_{i_2} \in A_{i_2}, \dots, \omega_{i_n} \in A_{i_n}\},$$

mit $n \in \mathbb{N}, i_1, \dots, i_n \in I$, und $A_{i_1} \in \mathcal{F}_{i_1}, \dots, A_{i_n} \in \mathcal{F}_{i_n}$,
Die Produkt-$\sigma-$Algebra $\bigotimes_{i\in I} \mathcal{F}_i$ *ist die von $\mathcal{C}$ erzeugte σ-Algebra auf Ω*

$$\bigotimes_{i\in I} \mathcal{F}_i := \sigma(\mathcal{C}).$$

4. Kontinuierliche Gleichverteilung. Sei $\Omega = (a, b)$ oder $\Omega = [a, b]$, mit $-\infty < a < b < \infty$. Wir suchen eine σ-Algebra $\mathcal{F}$ und ein Wahrscheinlichkeitsmaß $\mathbb{P}$, sodass $\forall\ a \leq c < d \leq b$ Folgendes gilt:

- Die Intervalle (c, d) und $[c, d]$ liegen in $\mathcal{F}$ und
- $P[(c, d)] = P[[c, d]] = \dfrac{d-c}{b-a}$ (Gleichverteilung).

Für $\mathcal{F}$ müssen wir die Borel σ-Algebra $\mathcal{B}((a, b))$ einführen. Wir werden später sehen, dass genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$: $\mathcal{B}((a, b)) \to [0, 1]$ existiert, sodass $P[(c, d)] = P[[c, d]] = \dfrac{d-c}{b-a}$ gilt.

1.3.4 Borel $\sigma-$Algebra

Definition 1.8. *Sei X ein topologischer Raum und*

$$\mathcal{T} = \{A \subset X \mid A \text{ offen}\}$$

seine Topologie. Die Borel σ-Algebra $\mathcal{B}(X)$ *ist die kleinste σ-Algebra auf X, die $\mathcal{T}$ enthält,*

$$\mathcal{B}(X) := \sigma(\mathcal{T}) = \bigcap_{\mathcal{F} \supset \mathcal{T},\ \mathcal{F}\ \sigma\text{-Algebra}} \mathcal{F}.$$

Die Mengen in $\mathcal{B}(X)$ werden Borel-Mengen *genannt.*

Bemerkung 1. Beachte, dass wir für $X = \mathbb{R}$ stets die euklidische Topologie nutzen.

Bemerkung 2. Die Borel σ-Algebra $\mathcal{B}(\mathbb{R})$ enthält alle für uns relevanten Teilmengen. Insbesondere enthält $\mathcal{B}(\mathbb{R})$

- alle offene und abgeschlossene Mengen (zB alle kompakte Intervalle $[a, b]$),
- alle einzelne Punkte $\{x\}$ $x \in \mathbb{R}$,
- alle Intervalle der Form $[a, b)$ oder $(a, b]$.

Bemerkung 3. $\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{P}(\mathbb{R})$ (siehe Analysis 3).

Erzeugende Mengensysteme für $\mathcal{B}(\mathbb{R})$. Wir betrachten folgende Mengensysteme

$$\begin{aligned} \mathcal{F}_1 &:= \{(-\infty, a) \mid a \in \mathbb{R}\}, \quad & \mathcal{F}_2 &:= \{(-\infty, a] \mid a \in \mathbb{R}\}, \\ \mathcal{F}_3 &:= \{(a, \infty) \mid a \in \mathbb{R}\}, \quad & \mathcal{F}_4 &:= \{[a, \infty) \mid a \in \mathbb{R}\}. \end{aligned}$$

Es gilt $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_i)\ \forall i = 1, \dots, 4$ (Übung).

1.3.5 Eindeutigkeitssatz, Carathéodory und Verteilungsfunktion

Sei Ω eine nichtleere Menge und $\mathcal{F}$ eine $\sigma-$Algebra über Ω. In der Regel sind die Wahrscheinlichkeiten $\mathbb{P}(A)$ zunächst nur für Ereignisse A aus einer Teilmenge $\mathcal{E} \subsetneq \mathcal{F}$ gegeben, z.B. für Intervalle bei Wahrscheinlichkeitsverteilungen auf $\mathbb{R}$. Es stellt sich die Frage, ob hierdurch bereits die Wahrscheinlichkeiten aller Ereignisse in $\mathcal{F}$ eindeutig festgelegt sind, und ob sich $\mathbb{P}$ zu einer Wahrscheinlichkeitsverteilung auf $\mathcal{F}$ fortsetzen lässt. Wir werden sehen, dass folgende Eigenschaften genügen: $\sigma(\mathcal{E}) = \mathcal{F}$ und $\mathcal{E}$ ist durchschnittsstabil.

Definition 1.9 (Durchschnittsstabiles Mengensystem, Mengenalgebra)**.**
Sei Ω eine nichtleere Menge, $\mathcal{A} \subset \mathcal{P}(\Omega)$ ein Mengensystem.

(i) $\mathcal{A}$ heißt durchschnittsstabil *(*$\cap$-stabil*), falls*

$$A, B \in \mathcal{A} \quad \Rightarrow \quad A \cap B \in \mathcal{A}.$$

(ii) $\mathcal{A}$ heißt Mengenalgebra *(*Algebra*), falls*

(a) $\mathcal{A}$ nichtleer ist,

(b) $A \in \mathcal{A} \quad \Rightarrow \quad A^c \in \mathcal{A}$,

(c) $A, B \in \mathcal{A} \quad \Rightarrow \quad A \cup B \in \mathcal{A}$.

Bemerkung Folgendes gilt.

- $\mathcal{A}$ Algebra $\Rightarrow \Omega, \varnothing \in \mathcal{A}$.
- $\mathcal{A}$ $\sigma-$Algebra $\Rightarrow$ $\mathcal{A}$ Algebra.
  Die andere Richtung gilt nicht.
- $\mathcal{F}$ $\sigma-$Algebra $\Rightarrow$ $\mathcal{F}$ durchschnittsstabil.
  Die andere Richtung gilt nicht. Die Menge $\mathcal{U}(\mathbb{R}) := \{I \subset \mathbb{R} \mid I \text{ offen}\}$ ist durchschnittsstabil, aber keine Algebra.

Satz 1.10 (Eindeutigkeitssatz)**.**
Sei Ω eine nichtleere Menge, $\mathcal{F}$ eine $\sigma-$Algebra über Ω und $\mathbb{P}_1, \mathbb{P}_2 \colon \mathcal{F} \to [0,1]$ zwei Wahrscheinlichkeitsmaße. Sei $\mathcal{E} \subset \mathcal{P}(\Omega)$ ein Mengensystem, sodass

- *$\mathcal{F} = \sigma(\mathcal{E})$ und*
- *$\mathcal{E}$ ist durchschnittsstabil.*

Falls $\mathbb{P}_1(E) = \mathbb{P}_2(E)\ \forall E \in \mathcal{E}$ gilt, dann gilt auch $\mathbb{P}_1 = \mathbb{P}_2$.

Beweis. siehe Abschnitt 1.3.6. □

Satz 1.11 (Fortsetzungssatz von Carathéodory)**.** *Sei Ω eine nichtleere Menge, $\mathcal{E} \subset \mathcal{P}(\Omega)$ eine Algebra über Ω und $\tilde{\mu}\colon \mathcal{E} \to [0,\infty]$ eine Funktion, sodass Folgendes gilt:*

- $\tilde{\mu}(\varnothing) = 0$ *und*
- $\tilde{\mu}$ *ist σ-additiv, d.h.* $\forall A\colon \mathbb{N} \to \mathcal{E}$*, sodass* $A_n \cap A_m = \varnothing$ $\forall n \neq m$ *und* $\cup_{n\in\mathbb{N}} A_n \in \mathcal{E}$ *gilt*

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

In diesem Fall heißt $\tilde{\mu}$ ein Prämaß.

Dann besitzt $\tilde{\mu}$ eine Fortsetzung zu einem Maß auf $\sigma(\mathcal{E})$, d.h. $\exists \mu\colon \sigma(\mathcal{E}) \to [0,\infty]$ Maß, sodass $\mu(E) = \tilde{\mu}(E)$ $\forall E \in \mathcal{E}$.

Falls $\tilde{\mu}$ $\sigma-$endlich ist (d.h. $\exists A\colon \mathbb{N} \to \mathcal{E}$, sodass $\tilde{\mu}(A_n) < \infty$ $\forall n$ gilt und $\cup_{n\in\mathbb{N}} A_n = \Omega$), dann ist das Maß μ $\sigma-$endlich und eindeutig.

Beweis. Der Beweis dieser Aussage findet sich in fast jedem Lehrbuch der Maßtheorie, z.B. Satz 5.3 (Existenz) und Satz 5.4 (Eindeutigkeit) in *Mass und Integrationstheorie* von Heinz Bauer.

[*Dieser Teil wurde nicht in der Vorlesung besprochen.*
Beweis der Eindeutigkeit: Sei $\mathcal{E} \subset \mathcal{P}(\Omega)$ eine Algebra über Ω, $\tilde{\mu}\colon \mathcal{E} \to [0,\infty]$ ein $\sigma-$endliches Prämaß, $\mu_1, \mu_2\colon \sigma(\mathcal{E}) \to [0,\infty]$ zwei Maße, die $\tilde{\mu}$ fortsetzen $\mu_{1|\mathcal{E}} = \mu_{2|\mathcal{E}} = \tilde{\mu}$.
Da $\tilde{\mu}$ $\sigma-$endlich ist, existiert eine Folge $E\colon \mathbb{N} \to \mathcal{E}$ sodass $E_n \subset E_{n+1}$, $\cup_{n\in\mathbb{N}} E_n = \Omega$ und $\mu_1(E) = \mu_2(E) = \tilde{\mu}(E_n) < \infty$ $\forall n$ gilt. Wir betrachten das Mengensystem

$$\mathcal{D} := \{A \in \sigma(\mathcal{E}) \mid \mu_1(A \cap E_m) = \mu_2(A \cap E_m)\ \forall m \in \mathbb{N}\} \subset \sigma(\mathcal{E}).$$

$\mathcal{D}$ ist ein Dynkin-System (siehe Def. 1.16 unten) und $\mathcal{E} \subset \mathcal{D}$. $\mathcal{E}$ ist eine Algebra, also auch $\cap-$stabil. Daraus folgt (Lemma 1.19 unten), dass $\sigma(\mathcal{E}) = \mathcal{D}(\mathcal{E}) \subset \mathcal{D}$, und deshalb $\sigma(\mathcal{E}) = \mathcal{D}$. Die Folge $n \mapsto A_n := (A \cap E_n)$ ist monoton steigend und $\cup_{n\in\mathbb{N}} A_n = A$. Aus dem Satz 1.5 folgt dann, dass für alle $A \in \sigma(\mathcal{E}) = \mathcal{D}$,

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A \cap E_n) = \lim_{n\to\infty} \mu_2(A \cap E_n) = \mu_2(A).$$]

□

Wahrscheinlichkeitsmaße konstruieren. Sei Ω eine nichtleere Menge. Man kann ein Wahrscheinlichkeitsmaß über $(\Omega, \mathcal{F})$ konstruieren wie folgt:

1. Wir suchen ein $\cap-$stabil Mengensystem $\mathcal{E} \subset \mathcal{P}(\Omega)$, sodass $\mathcal{F} = \sigma(\mathcal{E})$ und konstruieren eine σ-additive Abbildung $\tilde{\mu}\colon \mathcal{E} \to [0,1]$.
2. Sei $\mathcal{A}(\mathcal{E}) := \bigcap_{\mathcal{A} \supset \mathcal{E},\, \mathcal{A} \text{ Algebra}} \mathcal{A}$ die von $\mathcal{E}$ erzeugte Algebra. Insbesondere gilt $\Omega, \varnothing \in \mathcal{A}(\mathcal{E})$. Wir konstruieren eine σ-additive Fortsetzung $\tilde{\mu}_1\colon \mathcal{A}(\mathcal{E}) \to [0,1]$ $\tilde{\mu}_{1|\mathcal{E}} = \tilde{\mu}$, sodass $\tilde{\mu}_1(\varnothing) = 0$ und $\tilde{\mu}_1(\Omega) = 1$.

- Aus dem Fortsetzungssatz 1.11 ($\mathcal{A}(\mathcal{E})$ ist eine Algebra, $\tilde{\mu}_1(\varnothing) = 0$, $\tilde{\mu}_1(\Omega) = 1$ und $\tilde{\mu}_1$ ist σ-additiv) folgt, dass genau ein Wahrscheinlichkeitsmaß $\mu\colon \mathcal{F} \to [0,1]$ existiert, sodass $\mu(A) = \tilde{\mu}_1(A)$ $\forall A \in \mathcal{A}(\mathcal{E})$.

- Aus dem Eindeutigkeitssatz 1.10 (μ ist ein Wahrscheinlichkeitsmaß, $\mathcal{E}$ $\cap$-stabil) folgt, dass μ das einzige Wahrscheinlichkeitsmaß mit $\mu_{|\mathcal{E}} = \tilde{\mu}$ ist.

Beispiel. Im Fall der Münzwurffolge oben ist das Mengensystem (siehe (1.6) und (1.7))

$$\mathcal{E} \coloneqq \mathcal{C}$$

eine Algebra (Übung). Wir definieren $\tilde{\mu}\colon \mathcal{E} \to [0,1]$ durch

$$\tilde{\mu}(C_{n,A}) \coloneqq \frac{|A|}{2^n} \quad \forall n \geq 1, A \in \mathcal{P}(\{0,1\}^n).$$

Diese Funktion ist wohldefiniert, σ-additiv und $\tilde{\mu}(A(x_1,\dots,x_n)) = \frac{1}{2^n}$ (Übung). Nach Satz 1.11 folgt, dass genau ein Wahrscheinlichkeitsmaß $\mathbb{P}\colon \sigma(\mathcal{C}) \to [0,1]$, existiert, sodass $\mathbb{P}(C_{n,A}) := \frac{|A|}{2^n}$ für $C_{n,A} \in \mathcal{C}$ gilt.

Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Wir zeigen jetzt, dass ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ durch seine Verteilungsfunktion eindeutig charakterisiert ist.

Satz 1.12.

(i) Sei $F\colon \mathbb{R} \to \mathbb{R}$ monoton wachsend und rechtsstetig.
Dann $\exists!$ $\sigma-$endliches Maß $\mu\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$, sodass

$$\mu((s,t]) = F(t) - F(s) \qquad \forall s < t \in \mathbb{R}. \tag{1.8}$$

Falls $F(\infty) - F(-\infty) = 1$, wobei $F(\pm\infty) \coloneqq \lim_{t\to\pm\infty} F(t)$, dann ist μ ein Wahrscheinlichkeitsmaß.

(ii) Sei $\mu\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ ein $\sigma-$endliches Maß. Dann existiert eine monoton wachsende und rechtsstetige Funktion $F\colon \mathbb{R} \to [-\infty,\infty]$, sodass (1.8) *gilt. F ist bis auf eine additive Konstante festgelegt.*

Beweis. siehe Abschnitt 1.3.7. □

Erinnerung: $F\colon \mathbb{R} \to \mathbb{R}$ ist monoton wachsend, falls $t \leq t' \Rightarrow F(t) \leq F(t')$ gilt.
$F\colon \mathbb{R} \to \mathbb{R}$ ist rechtsstetig, falls $\lim_{\varepsilon\downarrow 0} F(t+\varepsilon) = F(t)$ $\forall t \in \mathbb{R}$ gilt.

Korollar 1.13 (Lebesgue-Maß)**.** *Es existiert genau ein $\sigma-$endliches Maß $\mathcal{L}\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$, , sodass*

$$\mathcal{L}((s,t]) = t - s \qquad \forall s < t \in \mathbb{R}. \tag{1.9}$$

Insbesondere gilt

$$\mathcal{L}((s,t)) = \mathcal{L}([s,t]) = \mathcal{L}([s,t)) = t - s \quad \forall s < t \in \mathbb{R} \quad \textit{und} \quad \mathcal{L}(\{x\}) = 0 \quad \forall x \in \mathbb{R}.$$

Dieses Maß heißt das Lebesgue-Maß.

Beweis. Wir wählen $F(t) = t$. Dann ist F monoton wachsend und stetig. Aus dem Satz 1.12 folgt, dass ein eindeutiges $\sigma-$endliches Maß $\mathcal{L}\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ existiert, sodass $\mathcal{L}((s,t]) = t - s$ $\forall s < t \in \mathbb{R}$.

Aus $[s,t] = \bigcap_n \left(s - \frac{1}{n}, t\right]$ und Satz 1.5 (die Folge $n \mapsto A_n \coloneqq (s - \frac{1}{n}, t]$ ist monoton fallend und $\mathcal{L}(A_1) < \infty$) folgt

$$\mathcal{L}([s,t]) = \mathcal{L}\left(\bigcap_n \left(s - \frac{1}{n}, t\right]\right) = \lim_{n\to\infty} \mathcal{L}\left(\left(s - \frac{1}{n}, t\right]\right) = \lim_{n\to\infty}\left(t - s + \frac{1}{n}\right) = t - s.$$

Die andere Fälle werden analog bewiesen (Übung). Z.B. folgt $\mathcal{L}(\{x\}) = 0$ aus $\{x\} = \bigcap \left(x - \frac{1}{n}, x\right]$ und Satz 1.5.

□

Bemerkung. Man kann das Lebesgue-Maß auf einer größeren $\sigma-$Algebra (die *Lebesgue* $\sigma-$Algebra $\mathcal{M}$) definieren durch

$$\mathcal{L}(A) \coloneqq \inf\left\{\sum_{n\in\mathbb{N}} vol(I_j) | A \subset \bigcup_{j\in\mathbb{N}} I_j,\ I_j \text{ Intervall } \forall j\right\} \quad \forall A \in \mathcal{M},$$

wobei I_j ein (möglicherwise leeres) Intervall $I_j = [a_j, b_j], (a_j, b_j), [a_j, b_j)$, oder $(a_j, b_j]$ mit $a_j \le b_j$ ist. Zudem gilt $vol(I_j) \coloneqq b_j - a_j$.

Es gilt $\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{M} \subsetneq \mathcal{P}(\mathbb{R})$ (siehe Analysis 3).

Definition 1.14 (Verteilungsfunktion)**.** *Sei* $\mathbb{P}$ *ein Wahrscheinlichkeitsmaß auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*.* *Die Funktion* $F\colon \mathbb{R} \to [0,1]$ *definiert durch*

$$F(t) \coloneqq \mathbb{P}((-\infty, t])$$

heißt Verteilungsfunktion von $\mathbb{P}$.

Lemma 1.15 (Eigenschaften der Verteilungsfunktion)**.** *Sei* $\mathbb{P}$ *ein Wahrscheinlichkeitsmaß auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. Die Verteilungsfunktion* $F\colon \mathbb{R} \to [0,1]$ *von* $\mathbb{P}$ *hat folgende Eigenschaften:*

(i) F *ist* monoton wachsend*,*

(ii) $\lim_{c\to-\infty} F(c) = 0$ *und* $\lim_{c\to\infty} F(c) = 1$*,*

(iii) F *ist* rechtsstetig*, d.h.* $F(c) = \lim_{y\searrow c} F(y)$ *für alle* $c \in \mathbb{R}$*,*

(iv) $F(c) = \lim_{y\nearrow c} F(y) + \mathbb{P}[\{c\}]$.
Insbesondere ist F *genau dann stetig in* c*, wenn* $\mathbb{P}[\{c\}] = 0$ *gilt.*

Beweis. Übung. □

Bemerkung. Aus dem Satz 1.12 folgt, dass die Verteilungsfunktion eindeutig das Wahrscheinlichkeitsmaß charakterisiert.

1.3.6 Beweis vom Eindeutigkeitssatz

Wir beweisen jetzt Satz 1.10. Sei Ω eine nichtleere Menge, $\mathcal{F}$ eine $\sigma-$Algebra über Ω und $\mathbb{P}_1, \mathbb{P}_2 \colon \mathcal{F} \to [0,1]$ zwei Wahrscheinlichkeitsmaßen. Sei $\mathcal{E} \subset \mathcal{P}(\Omega)$ ein Mengensystem, sodass

- $\mathcal{F} = \sigma(\mathcal{E})$ und
- $\mathcal{E}$ ist durchschnittsstabil.

Wir nehmen an, dass $\mathbb{P}_1(E) = \mathbb{P}_2(E)\ \forall E \in \mathcal{E}$, und definieren

$$\mathcal{D}_0 \coloneqq \{A \in \mathcal{F} \mid \mathbb{P}_1(A) = \mathbb{P}_2(A)\} \subset \mathcal{F}.$$

Unser Ziel ist zu beweisen, dass $\mathcal{F} \subset \mathcal{D}_0$ und deshalb $\mathcal{D}_0 = \mathcal{F}$ gilt. Es gilt $\mathcal{E} \subset \mathcal{D}_0$.
Dazu stellen wir fest, dass $\mathcal{D}_0$ folgende Eigenschaften hat:

- $\Omega \in \mathcal{D}_0$,
- $A \in \mathcal{D}_0 \ \Rightarrow\ A^c \in \mathcal{D}_0$,
- Für jede Folge $A \colon \mathbb{N} \to \mathcal{D}_0$ mit $A_n \cap A_m = \varnothing\ \forall n \neq m$ gilt $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}_0$.

Also ist $\mathcal{D}_0$ ein Dynkin-System (siehe Def. 1.16 unten). Aus Lemma 1.17 unten folgt, dass

$$\mathcal{D}(\mathcal{E}) \coloneqq \bigcap_{\mathcal{D} \supset \mathcal{E},\ \mathcal{D} \text{ Dynkin-System}} \mathcal{D}$$

ein Dynkin-System ist (das kleinste Dynkin-System, das $\mathcal{E}$ enthält). Da $\mathcal{E} \subset \mathcal{D}_0$ gilt und $\mathcal{D}_0$ ein Dynkin-System ist, folgern wir, dass $\mathcal{D}(\mathcal{E}) \subset \mathcal{D}_0$ gilt.

Behauptung: $\mathcal{E}$ ist $\cap-$stabil $\Rightarrow \sigma(\mathcal{E}) = \mathcal{D}(\mathcal{E})$. (Beweis in Lemma 1.19 unten)

Aus der Behauptung, zusammen mit der Annahme $\sigma(\mathcal{E}) = \mathcal{F}$, folgt

$$\mathcal{F} = \sigma(\mathcal{E}) = \mathcal{D}(\mathcal{E}) \subset \mathcal{D}_0 \subset \mathcal{F}.$$

Deshalb gilt $\mathcal{F} = \mathcal{D}_0$ und damit ist der Beweis beendet. □

Definition 1.16 (Dynkin-System)**.** *Ein Mengensystem $\mathcal{D} \subset \mathcal{P}(\Omega)$ heißt* Dynkin-System*, falls Folgendes gilt:*

(i) $\Omega \in \mathcal{D}$,

(ii) $A \in \mathcal{D} \ \Rightarrow\ A^c \in \mathcal{D}$,

(iii) Ist $A \colon \mathbb{N} \to \mathcal{D}$ eine Folge mit $A_n \cap A_m = \varnothing\ \forall n \neq m$, dann gilt auch $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}$.

Lemma 1.17. *Sei Ω eine nichtleere Menge, I eine beliebige Indexmenge und für jedes $\alpha \in I$ sei $\mathcal{D}_\alpha \subset \mathcal{P}(\Omega)$ ein Dynkin-System.*

Der Durchschnitt $\mathcal{D} \coloneqq \bigcap_{\alpha \in I} \mathcal{S}_\alpha$ ist ein Dynkin-System.

Beweis. $\mathcal{D}$ ist ein Dynkin-System, falls es (i) – (iii) aus Def. 1.16 erfüllt.

(i) Da $\mathcal{D}_\alpha$ ein Dynkin-System $\forall \alpha \in I$ ist, gilt $\Omega \in \mathcal{D}_\alpha\ \forall \alpha \in I$ und deshalb gilt auch $\Omega \in \mathcal{D}$.

(ii) und (iii) werden analog bewiesen.

□

Bemerkung 1. $\Omega \in \mathcal{D} \overset{(ii)}{\Rightarrow} \varnothing \in \mathcal{D}$.
Daraus folgt, dass $\forall A_1, \dots, A_n \in \mathcal{D}$, mit $A_i \cap A_j = \varnothing\ \forall i \neq j$ auch $A_1 \cup A_2 \cup \dots \cup A_n \in \mathcal{D}$ gilt (wir nehmen die Folge $k \mapsto B_k$, welche durch $B_k \coloneqq A_k$ für $k = 1, \dots, n$ und $B_k \coloneqq \varnothing\ \forall k > n$ definiert ist).

Bemerkung 2. Anders als in der Definition von $\sigma-$Algebra können wir nicht (i) durch '$\mathcal{D}$ ist nicht leer' ersetzen, denn ohne $\varnothing \in \mathcal{D}$ haben wir keine Additivität.

Bemerkung 3. $A, B \in \mathcal{D}$ mit $A \subset B \Rightarrow A \setminus B \in \mathcal{D}$.
Beweis. Es gilt $B \setminus A = B \cap A^c = (B^c \cup A)^c$. Es gilt $B \in \mathcal{D} \Rightarrow B^c \in \mathcal{D}$. Zudem folgt aus $A \subset B \Rightarrow A \cap B^c = \varnothing$ und Bemerkung 1, dass $B^c \cup A \in \mathcal{D}$.
Die Aussage folgt jetzt aus (ii). □

Bemerkung 4. Eine $\sigma-$Algebra ist auch ein Dynkin-System. Ein Dynkin-System $\mathcal{D}$ ist eine $\sigma-$Algebra nur dann, wenn $\mathcal{D}$ auch $\cap-$stabil ist (siehe Lemma 1.18 unten).

[4: 24.10.2025]

[5: 28.10.2025]

Lemma 1.18. *Sei Ω eine nichtleere Menge, $\mathcal{D} \subset \mathcal{P}(\Omega)$ ein Dynkin-System.*

Falls $\mathcal{D}$ $\cap$-stabil ist, dann ist $\mathcal{D}$ eine $\sigma-$Algebra.

Beweis. Übungsblatt

□

Lemma 1.19. *Sei Ω eine nichtleere Menge, $\mathcal{E} \subset \mathcal{P}(\Omega)$ ein $\cap-$stabiles Mengensystem und sei*

$$\mathcal{D}(\mathcal{E}) \coloneqq \bigcap_{\mathcal{D} \supset \mathcal{E},\, \mathcal{D} \text{ Dynkin-System}} \mathcal{D} \tag{1.10}$$

das kleinste Dynkin-System, das $\mathcal{E}$ enthält ($\mathcal{D}(\mathcal{E})$ ist ein Dynkin-System nach Lemma 1.18). Dann gilt

$$\mathcal{D}(\mathcal{E}) = \sigma(\mathcal{E}).$$

Beweis. Wir zeigen, dass $\mathcal{D}(\mathcal{E}) \subset \sigma(\mathcal{E})$ und $\sigma(\mathcal{E}) \subset \mathcal{D}(\mathcal{E})$ gilt.

$\mathcal{D}(\mathcal{E}) \subset \sigma(\mathcal{E})$: $\sigma(\mathcal{E})$ ist eine $\sigma-$Algebra, also ist $\sigma(\mathcal{E})$ ein Dynkin-System, welches $\mathcal{E}$ enthält. Die Aussage folgt aus (1.10).

$\sigma(\mathcal{E}) \subset \mathcal{D}(\mathcal{E})$: *Behauptung.* $\mathcal{E}$ $\cap-$stabil $\Rightarrow \mathcal{D}(\mathcal{E})$ $\sigma-$Algebra.

Aus der Behauptung folgt $\sigma(\mathcal{E}) \subset \mathcal{D}(\mathcal{E})$, da $\sigma(\mathcal{E})$ die kleinste $\sigma-$Algebra ist, welche $\mathcal{E}$ enthält.

Beweis der Behauptung. Nach Lemma 1.18, genügt es zu zeigen, dass $\mathcal{D}(\mathcal{E})$ $\cap-$stabil ist. Für $A \in \mathcal{D}(\mathcal{E})$ definieren wir

$$\mathcal{D}_A \coloneqq \{B \in \mathcal{D}(\mathcal{E}) \mid A \cap B \in \mathcal{D}(\mathcal{E})\} \subset \mathcal{D}(\mathcal{E}).$$

Es gilt

$$\mathcal{D}(\mathcal{E}) \cap -\text{stabil} \quad \Leftrightarrow \quad \mathcal{D}_A = \mathcal{D}(\mathcal{E}) \quad \forall A \in \mathcal{D}(\mathcal{E}) \quad \Leftrightarrow \quad \mathcal{D}(\mathcal{E}) \subset \mathcal{D}_A \quad \forall A \in \mathcal{D}(\mathcal{E}),$$

wobei $\mathcal{D}_A \subset \mathcal{D}(\mathcal{E})$ verwendet wurde. Wir zeigen jetzt, dass $\mathcal{D}(\mathcal{E}) \subset \mathcal{D}_A \ \forall A \in \mathcal{D}(\mathcal{E})$ gilt.

Schritt 1. $\mathcal{D}_A$ ist ein Dynkin-System (Übung).

Schritt 2. Wir zeigen, dass $\mathcal{D}(\mathcal{E}) \subset \mathcal{D}_A \ \forall A \in \mathcal{E}$ gilt. Sei $A \in \mathcal{E}$ fest. Da $\mathcal{E}$ $\cap-$stabil ist, folgt $A \cap B \in \mathcal{E} \subset \mathcal{D}(\mathcal{E}) \ \forall B \in \mathcal{E}$. Dies impliziert $\mathcal{E} \subset \mathcal{D}_A$. Also ist $\mathcal{D}_A$ ein Dynkin-System, das $\mathcal{E}$ enthält. Daraus folgt, dass $\mathcal{D}(\mathcal{E}) \subset \mathcal{D}_A$ gilt.

Schritt 3. Wir zeigen, dass $\mathcal{D}(\mathcal{E}) \subset \mathcal{D}_A \ \forall A \in \mathcal{D}(\mathcal{E})$. Sei $A \in \mathcal{D}(\mathcal{E})$ fest. Aus Schritt 2 folgt, dass $A \in \mathcal{D}_E \ \forall E \in \mathcal{E}$ gilt. Daraus folgt $A \cap E \in \mathcal{D}(\mathcal{E}) \ \forall E \in \mathcal{E}$ und daher $\mathcal{E} \subset \mathcal{D}_A$.

Also ist $\mathcal{D}_A$ ein Dynkin-System, das $\mathcal{E}$ enthält. Daraus folgt, dass $\mathcal{D}(\mathcal{E}) \subset \mathcal{D}_A$ gilt.

Damit ist der Beweis der Behauptung beendet.

Die beiden Inklusionen implizieren $\mathcal{D}(\mathcal{E}) = \sigma(\mathcal{E})$. □

1.3.7 Beweis vom Eindeutigkeitssatz auf $\mathbb{R}$

Beweis von Satz 1.12.

(ii) Sei $\mu\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ ein $\sigma-$endliches Maß. Unser Ziel ist zu zeigen, dass eine monoton wachsende und rechtsstetige Funktion $F\colon \mathbb{R} \to [-\infty,\infty]$ existiert, sodass

$$\mu((s,t]) = F(t) - F(s) \quad \forall s < t \in \mathbb{R} \tag{1.11}$$

gilt. Wir setzen

$$F(t) \coloneqq \begin{cases} \mu((0,t]) & \text{falls } t > 0, \\ 0 & \text{falls } t = 0, \\ -\mu((t,0]) & \text{falls } t < 0. \end{cases}$$

Diese Funktion ist monton wachsend und rechtsstetig (Übung). Falls $\mu((a,b]) < \infty \ \forall a < b$, gilt auch $F\colon \mathbb{R} \to \mathbb{R}$. Das ist z.B. der Fall für das Lebesgue Maß. Sei jetzt $\tilde{F}\colon \mathbb{R} \to \mathbb{R}$ eine andere monton wachsende und rechtsstetige Funktion, sodass $\mu((s,t]) = \tilde{F}(t) - \tilde{F}(s)$ gilt. Dann gilt

$$F(t) - \tilde{F}(t) = F(s) - \tilde{F}(s) \qquad \forall s < t.$$

Daraus folgt, dass eine Konstante $C \in \mathbb{R}$ existiert, sodass $\tilde{F} = F + C$.

(i) Sei $F\colon \mathbb{R} \to \mathbb{R}$ monoton wachsend und rechtsstetig. Unser Ziel ist zu zeigen, dass genau ein $\sigma-$endliches Maß $\mu\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ existiert, sodass (1.11) gilt. Wir betrachten das Mengensystem

$$\mathcal{E} \coloneqq \{(s,t] \mid -\infty \le s \le t < \infty\} \cup \{(s,\infty) \mid -\infty \le s < \infty\}.$$

Sei $\mathcal{A}(\mathcal{E})$ die von $\mathcal{E}$ erzeugte Algebra

$$\mathcal{A}(\mathcal{E}) = \bigcap_{\mathcal{A} \supset \mathcal{E},\ \mathcal{A} \text{ Algebra}} \mathcal{A}.$$

Das Mengensystem $\mathcal{A}(\mathcal{E})$ ist eine Algebra und $\forall A \in \mathcal{A}(\mathcal{E})\ \exists I_1, \ldots, I_n \in \mathcal{E}$, $n \geq 1$ disjunkt, sodass $A = \cup_{j=1}^n I_j$ (Übung). Wir definieren $\tilde{\mu} \colon \mathcal{E} \to [0, \infty]$ durch

$$\begin{aligned}
\tilde{\mu}((s,t]) &\coloneqq F(t) - F(s) \quad \forall s,t \in \mathbb{R} \\
\tilde{\mu}((-\infty,t]) &\coloneqq F(t) - F(-\infty) \quad \forall t \in \mathbb{R} \\
\tilde{\mu}((s,\infty)) &\coloneqq F(\infty) - F(s) \quad \forall s \in \mathbb{R} \\
\tilde{\mu}((-\infty,\infty)) &\coloneqq F(\infty) - F(-\infty) \\
\tilde{\mu}(\varnothing) &\coloneqq 0.
\end{aligned}$$

Mit $A = \cup_{j=1}^n I_j$ wird $\tilde{\mu}$ auf $\mathcal{A}(\mathcal{E})$ fortgesetzt. Die Funktion $\tilde{\mu} \colon \mathcal{A}(\mathcal{E}) \to [0, \infty]$ ist wohldefiniert, erfüllt $\tilde{\mu}(\varnothing) = 0$ und ist additiv, d.h. für jede Folge $A_1, \ldots, A_n \in \mathcal{A}(\mathcal{E})$ mit $A_i \cap A_j = \varnothing\ \forall i \neq j$ gilt $\tilde{\mu}\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n \tilde{\mu}(A_j)$.

Behauptung. $\tilde{\mu}$ ist $\sigma-$additiv und $\sigma-$endlich.

Aus der Behauptung, zusammen mit $\sigma(\mathcal{A}(\mathcal{E})) = \sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$ und Satz 1.11, folgt, dass genau ein $\sigma-$endliches Maß $\mu \colon \mathcal{B}(\mathbb{R}) \to [0, \infty]$ existiert, sodass (1.11) gilt. Falls $F(\infty) - F(-\infty) = 1$, so ist μ auch ein Wahrscheinlichkeitsmaß.

Beweisstrategie für die Behauptung.

Schritt 1. Wir zeigen, dass $\tilde{\mu}$ $\sigma-$*nullstetig* (stetig bei der leeren Menge) ist, d.h. für jede Folge $A \colon \mathbb{N} \to \mathcal{A}(\mathcal{E})$ mit $\tilde{\mu}(A_1) < \infty$ und $A_n \downarrow \varnothing$ (d.h. $A_{n+1} \subset A_n$ und $\bigcap_{n \in \mathbb{N}} A_n = \varnothing$) gilt $\lim_{n \to \infty} \tilde{\mu}(A_n) = 0$.

Schritt 2. Wir zeigen, dass entweder $\tilde{\mu}(\mathbb{R}) < \infty$ gilt oder für jedes $A \in \mathcal{A}(\mathcal{E})$ existiert eine Folge $B \colon \mathbb{N} \to \mathcal{A}(\mathcal{E})$, sodass

(i) $B_n \uparrow A$ (d.h. $B_n \subset B_{n+1}$ und $\bigcup_{n \in \mathbb{N}} B_n = A$),

(ii) $\tilde{\mu}(B_n) < \infty\ \forall n$,

(iii) $\lim_{n \to \infty} \tilde{\mu}(B_n) = \tilde{\mu}(A)$.

Insbesondere ist $\tilde{\mu}$ $\sigma-$endlich.

• Schritt 1 und 2 implizieren, dass $\tilde{\mu}$ $\sigma-$*stetig von unten* ist, d.h. für jede Folge $A \colon \mathbb{N} \to \mathcal{A}(\mathcal{E})$ mit $A_n \uparrow A \in \mathcal{A}(\mathcal{E})$ gilt $\lim_{n \to \infty} \tilde{\mu}(A_n) = \tilde{\mu}(A)$.

[*Dieser Teil wurde nicht in der Vorlesung besprochen.*

Um diese Implikation zu beweisen, sei $A \colon \mathbb{N} \to \mathcal{A}(\mathcal{E})$ mit $A_n \uparrow A \in \mathcal{A}(\mathcal{E})$ gegeben. Dann gilt $B_n \coloneqq (A \setminus A_n) \downarrow \varnothing$ (Übung).

Falls $\tilde{\mu}(A) < \infty$, gilt $\tilde{\mu}(B_1) < \infty$ und aus der Additivität von $\tilde{\mu}$ auf $\mathcal{A}(\mathcal{E})$ sowie der $\sigma-$Nullstetigkeit folgt

$$\tilde{\mu}(A) - \tilde{\mu}(A_n) = \tilde{\mu}(B_n) \overset{n \to \infty}{\to} 0.$$

Falls $\tilde{\mu}(A) = \infty$, gilt auch $\tilde{\mu}(\mathbb{R}) = \infty$ und aus Schritt 2 folgt, dass eine Folge $E \colon \mathbb{N} \to \mathcal{A}(\mathcal{E})$ existiert, , sodass $E_m \uparrow A$, $\tilde{\mu}(E_m) < \infty\ \forall m$, und $\lim_{m \to \infty} \tilde{\mu}(E_m) = \tilde{\mu}(A)$. Für jedes m gilt $\tilde{\mu}(A \cap E_m) = \tilde{\mu}(E_m) < \infty$ und $(A_n \cap E_m) \uparrow (A \cap E_m)$. Die $\sigma-$Nullstetigkeit impliziert erneut

$$\lim_{n \to \infty} \tilde{\mu}(A_n \cap E_m) = \tilde{\mu}(A \cap E_m) = \tilde{\mu}(E_m).$$

Wir rechnen jetzt

$$\tilde{\mu}(A) \geq \lim_{n\to\infty} \tilde{\mu}(A_n) \geq \lim_{n\to\infty} \tilde{\mu}(A_n \cap E_m) = \tilde{\mu}(E_m) \qquad \forall m \in \mathbb{N}.$$

Die Aussage folgt aus $\lim_{m\to\infty} \tilde{\mu}(E_m) = \tilde{\mu}(A)$.]

• $\sigma-$Additivität folgt jetzt aus der $\sigma-$Stetigkeit von unten (siehe Übungsblatt 2).

Beweisidee für Schritt 1. Sei $A\colon \mathbb{N} \to \mathcal{A}(\mathcal{E})$ mit $\tilde{\mu}(A_1) < \infty$ und $A_n \downarrow \varnothing$.
Die Funktion $\tilde{\mu}$ ist monoton (Übung), also ist die Folge $n \mapsto \tilde{\mu}(A_n)$ monoton fallend und daher existiert der Grenzwert. Wir argumentieren durch Widerspruch, wir nehmen also an, dass $\lim_{n\to\infty} \tilde{\mu}(A_n) = a > 0$. Daraus folgt, dass eine Folge $n \mapsto K_n$ existiert, sodass

- $K_n \subset \mathbb{R}$ kompakt ist, $K_{n+1} \subset K_n$ und $K_n \subset A_n$ $\forall n \in \mathbb{N}$ und
- $K_1 \cap \cdots \cap K_n \neq \varnothing$ $\forall n$.

Aus $K_1 \cap \cdots \cap K_n \neq \varnothing$ $\forall n$ folgt, dass (Resultat aus der Topologie) $\cap_{n\in\mathbb{N}} K_n \neq \varnothing$ und da $K_n \subset A_n \forall n$, gilt auch $\cap_{n\in\mathbb{N}} A_n \neq \varnothing$, was der Annahme $A_n \downarrow \varnothing$ widerspricht.

Beweisidee für Schritt 2. Für $A \in \mathcal{A}(\mathcal{E})$ gilt auch $B_n \coloneqq A \cap [-n, n] \in \mathcal{A}(\mathcal{E})$. Die Eigenschaften (i), (ii) und (iii) folgen aus der Definition von $\tilde{\mu}$. □

[5: 28.10.2025]

[6: 31.10.2025]

1.4 Zufallsvariablen und Integration

1.4.1 Zufallsvariablen

Informell ist eine Zufallsvariable eine Funktion $X\colon \Omega \to \mathbb{R}$. Um die Wahrscheinlichkeit für das Ereignis

$$(X(\omega) \in [a, b]) = X^{-1}([a, b]) \coloneqq \{\omega \in \Omega \mid X(\omega) \in [a, b]\}$$

zu definieren, muss $X^{-1}([a, b])$ eine messbare Menge sein. Das motiviert folgende Definition.

Definition 1.20.

(i) Sei $\Omega \neq \varnothing$ eine Menge und $\mathcal{F}$ eine σ-Algebra über Ω. Das Paar $(\Omega, \mathcal{F})$ heißt Messraum.

(ii) Sei $(\Omega, \mathcal{F})$ ein Messraum. Eine Funktion $f\colon \Omega \to \mathbb{R}$ heißt messbar, *falls*

$$B \in \mathcal{B}(\mathbb{R}) \;\Rightarrow\; f^{-1}(B) \in \mathcal{F}, \quad \textit{wobei } f^{-1}(B) \coloneqq \{\omega \in \Omega \mid f(\omega) \in B\} \in \mathcal{F}.$$

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine messbare Funktion $X\colon \Omega \to \mathbb{R}$ heißt reellwertige Zufallsvariable *(oder nur* Zufallsvariable*).*

(iii) Seien $(\Omega, \mathcal{F})$ und $(\tilde{\Omega}, \tilde{\mathcal{F}})$ Messräume. Eine Funktion $f\colon \Omega \to \tilde{\Omega}$ heißt messbar, *falls*

$$B \in \tilde{\mathcal{F}} \;\Rightarrow\; f^{-1}(B) \in \mathcal{F}.$$

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine messbare Funktion $X\colon \Omega \to \tilde{\Omega}$ heißt $\tilde{\Omega}-$wertige Zufallsvariable.

Notation. Man schreibt manchmal $f\colon (\Omega, \mathcal{F}) \to (\tilde{\Omega}, \tilde{\mathcal{F}})$, um die $\sigma-$Algebren zu betonen.

Beispiel. Sei $(\Omega, \mathcal{F})$ ein Messraum.
Für $A \in \mathcal{F}$ ist die Indikatorfunktion $\mathbf{1}_A \colon \Omega \to \mathbb{R}$ definiert durch $\mathbf{1}_A(\omega) \coloneqq 1$, falls $\omega \in A$ und $\mathbf{1}_A(\omega) \coloneqq 0$, falls $\omega \in A^c$. Diese Funktion ist messbar, da für $B \in \mathcal{B}(\mathbb{R})$ gilt

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

Analog ist für alle $x_1, \ldots, x_n \in \mathbb{R}$, $A_1, \ldots, A_n \in \mathcal{F}$ die Funktion $f \colon \Omega \to \mathbb{R}$ definiert durch $f \coloneqq \sum_{j=1}^n x_j \mathbf{1}_{A_j}$ messbar (Übung).

Messbare Funktionen erkennen

Lemma 1.21. *Sei $(\Omega, \mathcal{F})$ ein Messraum, $\tilde{\Omega}$ eine nichleere Menge und $f \colon \Omega \to \tilde{\Omega}$ eine Funktion. Wir betrachten das Mengensystem*

$$\mathcal{U}_f \coloneqq \{A \subset \tilde{\Omega} | \ f^{-1}(A) \in \mathcal{F}\} \subset \mathcal{P}(\tilde{\Omega}).$$

Dann ist $\mathcal{U}_f$ eine $\sigma-$Algebra über $\tilde{\Omega}$.

Beweis. Übung □

Korollar 1.22. *Seien $(\Omega, \mathcal{F})$ $(\tilde{\Omega}, \tilde{\mathcal{F}})$ zwei Messräume, und $f \colon \Omega \to \tilde{\Omega}$ eine Funktion. Sei $\tilde{\mathcal{E}} \subset \mathcal{P}(\tilde{\Omega})$ eine erzeugendes Mengensystem, also $\sigma(\tilde{\mathcal{E}}) = \tilde{\mathcal{F}}$.*

Falls $f^{-1}(A) \in \mathcal{F} \ \forall A \in \tilde{\mathcal{E}}$ gilt, so ist f messbar.

Beweis. Das Mengensystem $\mathcal{U}_f \coloneqq \{A \subset \tilde{\Omega} | \ f^{-1}(A) \in \mathcal{F}\} \subset \mathcal{P}(\tilde{\Omega})$ ist eine $\sigma-$Algebra (Lemma 1.21).
Es gilt $\tilde{\mathcal{E}} \subset \mathcal{U}_f$, denn $\forall E \in \tilde{\mathcal{E}}$ gilt $f^{-1}(E) \in \mathcal{F}$.
$\mathcal{U}_f$ ist also eine $\sigma-$Algebra über $\tilde{\Omega}$, die $\tilde{\mathcal{E}}$ enthält. Daraus folgt, dass

$$\tilde{\mathcal{F}} = \sigma(\tilde{\mathcal{E}}) \subset \mathcal{U}_f$$

gilt und damit ist der Beweis beendet. □

Korollar 1.23. *Sei $f \colon (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Funktion.*

Falls f stetig ist, dann ist f auch messbar.

Beweis. f stetig $\Rightarrow \forall A \subset \mathbb{R}$ offen ist $f^{-1}(A) \subset \mathbb{R}$ auch offen. Da die offene Teilmengen von $\mathbb{R}$ die Borel $\sigma-$Algebra $\mathcal{B}(\mathbb{R})$ erzeugen, folgt die Aussage aus Korollar 1.22. □

Bemerkung. Es gilt: $f \colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist messbar
$\Leftrightarrow f^{-1}((-\infty, a)) \in \mathcal{F} \ \forall a \in \mathbb{R}$
$\Leftrightarrow f^{-1}((-\infty, a]) \in \mathcal{F} \ \forall a \in \mathbb{R}$
$\Leftrightarrow f^{-1}((a, \infty)) \in \mathcal{F} \ \forall a \in \mathbb{R}$
$\Leftrightarrow f^{-1}([a, \infty)) \in \mathcal{F} \ \forall a \in \mathbb{R}$.

Lemma 1.24. *Sei* $(\Omega, \mathcal{F})$ *ein Messraum und* $f\colon \Omega \to \mathbb{R}$ *eine messbare Funktion. Wir definieren*

$$\begin{aligned} f_+(\omega) &:= \max\{f(\omega), 0\} = f(\omega)\mathbf{1}_{[0,\infty)}(f(\omega)) = f(\omega)\mathbf{1}_{f(\omega)\geq 0} \\ f_-(\omega) &:= -\min\{f(\omega), 0\} = -f(\omega)\mathbf{1}_{(-\infty,0)}(f(\omega)) = -f(\omega)\mathbf{1}_{f(\omega)<0}. \end{aligned}$$

Die Funktionen $f_\pm\colon \Omega \to \mathbb{R}$ *sind messbar und*

$$f(\omega) = f_+(\omega) - f_-(\omega), \qquad |f(\omega)| = f_+(\omega) + f_-(\omega) \qquad \forall \omega \in \Omega.$$

Außerdem gilt $f_\pm(\omega) > 0 \Rightarrow f_\mp(\omega) = 0$.

Beweis. Übung □

Definition 1.25. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $X\colon \Omega \to \mathbb{R}$ *eine Zufallsvariable. Die Funktion*

$$\begin{aligned} F_X\colon \quad \mathbb{R} &\to \quad [0,1] \\ t &\mapsto \quad F_X(t) := \mathbb{P}(X^{-1}(-\infty, t]) = \mathbb{P}(X(\omega) \leq t) \end{aligned}$$

heißt Verteilungsfunktion *von* X.

Die Funktion F_X ist wohldefiniert, da X messbar ist und deshalb $X^{-1}((-\infty, t]) \in \mathcal{F}$ $\forall t \in \mathbb{R}$ gilt.

Lemma 1.26. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $X\colon \Omega \to \mathbb{R}$ *eine Zufallsvariable. Die Verteilungsfunktion* $F_X\colon \mathbb{R} \to [0,1]$ *von* X *hat folgende Eigenschaften:*

(i) F_X *ist* monoton wachsend*,*

(ii) $\lim_{t\to-\infty} F_X(t) = 0$ *und* $\lim_{t\to\infty} F_X(t) = 1$,

(iii) F_X *ist* rechtsstetig*, d.h.* $F(t) = \lim_{y\searrow t} F_X(y)$ *für alle* $t \in \mathbb{R}$,

(iv) $F_X(t) = \lim_{y\nearrow t} F_X(y) + \mathbb{P}(X^{-1}(\{t\}))$.
Insbesondere ist F_X *genau dann stetig in* t*, wenn* $\mathbb{P}^{-1}(\{t\}) = 0$ *gilt.*

Beweis. Übung □

1.4.2 Integration

Definition 1.27. *Sei* $(\Omega, \mathcal{F})$ *ein Messraum.*
Eine Funktion $f\colon \Omega \to \mathbb{R}$ *heißt* einfach *(oder* elementare Zufallsvariable*), wenn* f *messbar und* $f(\Omega)$ *endlich ist.*

Wir definieren $\mathcal{S}(\Omega) := \{f\colon \Omega \to \mathbb{R} |\ f \text{ einfach}\}$ *und* $\mathcal{S}_+(\Omega) := \{f\colon \Omega \to [0,\infty) |\ f \text{ einfach}\}$.

Lemma 1.28. *Sei* $(\Omega, \mathcal{F})$ *ein Messraum und* $f\colon \Omega \to \mathbb{R}$ *eine Funktion.*

(i) $f \in \mathcal{S}(\Omega) \Leftrightarrow \exists c_1, \ldots, c_n \in \mathbb{R}$ *und* $A_1, \ldots, A_n \in \mathcal{F}$ *mit* $c_i \neq c_j$ *und* $A_i \cap A_j = \varnothing$ $\forall i \neq j$,, *sodass*

$$f = \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}.$$

(ii) Für $f, g \in \mathcal{S}(\Omega)$ gilt

(a) $af + bg \in \mathcal{S}(\Omega)$ $\forall a, b \in \mathbb{R}$

(b) $fg \in \mathcal{S}(\Omega)$

(c) $\max\{f, g\} \in \mathcal{S}(\Omega)$ *und* $\min\{f, g\} \in \mathcal{S}(\Omega)$.

Beweis. Übung. □

Definition 1.29. *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum.*

(i) Für $\varphi = \sum_{j=1}^n c_j \mathbf{1}_{A_j} \in \mathcal{S}_+(\Omega)$ definieren wir

$$\int_\Omega \varphi \; d\mu := \sum_{y \in \varphi(\Omega)} y \; \mu(\varphi^{-1}(\{y\})) = \sum_{j=1}^n c_j \; \mu(A_j) \in [0, \infty].$$

(ii) Für $f \colon \Omega \to [0, \infty)$ messbar definieren wir

$$\int_\Omega f \; d\mu := \sup_{\varphi \in \mathcal{S}_+(\Omega), \varphi \leq f} \int_\Omega \varphi \; d\mu \in [0, \infty].$$

*Die Funktion f heißt $\mu-$*integrierbar *(oder* integrierbar*), falls $\int_\Omega f \; d\mu < \infty$.*

*(iii) Eine messbare Funktion $f \colon \Omega \to \mathbb{R}$ heißt $\mu-$*integrierbar *(oder* integrierbar*), falls f_+ und f_- integrierbar sind. In diesem Fall definieren wir*

$$\int_\Omega f \; d\mu := \int_\Omega f_+ \; d\mu \; - \int_\Omega f_- \; d\mu.$$

(iv) Wir definieren $L^1(\Omega, \mathcal{F}, \mu) = L^1(\Omega, \mu) := \{f \colon \Omega \to \mathbb{R} |\; f \text{ integrierbar}\}$.

Notation. Folgende Notationen sind äquivalent:

$$\int_\Omega f \; d\mu = \int_\Omega f(\omega) \; d\mu(\omega) = \int_\Omega f(\omega) \; \mu(d\omega).$$

Falls $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L})$ bekommen wir das Lebesgue Integral $\int_\mathbb{R} f \; d\mathcal{L}$. Wir werden oft folgende Notationen anwenden

$$\int_\mathbb{R} f \; d\mathcal{L} = \int_\mathbb{R} f(x) \; d\mathcal{L}(x) = \int_\mathbb{R} f(x) \; dx.$$

Bemerkung 1. Für $\varphi = \sum_{j=1}^n c_j \mathbf{1}_{A_j} \in \mathcal{S}(\Omega)$ mit $\mu(A_j) < \infty$ $\forall j = 1, \dots, n$ gilt

$$\int_\Omega \varphi \; d\mu := \sum_{y \in \varphi(\Omega)} y \; \mu(\varphi^{-1}(\{y\})) = \sum_{j=1}^n c_j \; \mu(A_j).$$

Bemerkung 2. Die Funktion $f \colon \Omega \to \mathbb{R}$ ist integrierbar $\Leftrightarrow$ $|f|$ ist integrierbar und es gilt

$$\left| \int_\Omega f \; d\mu \right| = \left| \int_\Omega f_+ \; d\mu \; - \int_\Omega f_- \; d\mu \right| \leq \int_\Omega f_+ \; d\mu + \int_\Omega f_- \; d\mu = \int_\Omega |f| \; d\mu.$$

[6: 31.10.2025]

[7: 04.11.2025]

Bemerkung 3. Das Lebesgue Integral verallgemeinert das Riemann Integral insofern, dass deutlich Lebesgue-integrierbare Funktionen weniger regulär sein müssen als Riemann-integrierbare Funktionen. Andererseits gilt, dass jede Riemann integrierbare Funktion auch Lebesgue integrierbar ist, und dass in diesem Fall beide Integrale übereinstimmen.

Bemerkung 4. In der Definition 1.29 kann man $f\colon (\Omega,\mathcal{F}) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ mit $f\colon (\Omega,\mathcal{F}) \to (\overline{\mathbb{R}},\mathcal{B}(\overline{\mathbb{R}}))$ ersetzen, wobei

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty,\infty\} = [-\infty,\infty].$$

Die Topologie auf $\overline{\mathbb{R}}$ ist erzeugt durch die offenen Mengen $(a,b) \subset \overline{\mathbb{R}}$, $a < b \in \mathbb{R}$ sowie die Mengen $(a,\infty]$ und $[-\infty,a)$, $a \in \mathbb{R}$.

Satz 1.30 (Beppo Levi, Monotone Konvergenz)**.** *Sei $(\Omega,\mathcal{F},\mu)$ ein Maßraum. Sei $f\colon \Omega \to [0,\infty]$ eine messbare Funktion und $n \mapsto f_n\colon \Omega \to [0,\infty]$ eine Funktionenfolge, sodass*

- *f_n messbar ist und $f_n \le f_{n+1} \le f$ $\forall n \in \mathbb{N}$,*
- *$f_n \overset{n\to\infty}{\to} f$ punktweise, d.h. $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ $\forall \omega \in \Omega$.*

Dann gilt

$$\int_\Omega f\,d\mu = \int_\Omega \Big[\lim_{n\to\infty} f_n\Big]\,d\mu = \lim_{n\to\infty}\int_\Omega f_n\,d\mu.$$

Beweis. Analysis 3. □

Anwendungsbeispiel: Integrale rechnen. Sei $f\colon \Omega \to [0,\infty]$ eine messbare Funktion. Aus Def. 1.29 folgt, dass $\int_\Omega f\ d\mu = \sup_{\varphi\in\mathcal{S}_+(\Omega),\varphi\le f}\int_\Omega \varphi\ d\mu \in [0,\infty]$. Anstatt ein solches Supremum über alle einfachen Funktionen zu betrachten, reicht es, geeignete Folgen simpler Funktionen zu konstruieren, welche gegen das Supremum konvergieren.

1. Es gibt wenigstens eine Folge $\varphi\colon \mathbb{N} \to \mathcal{S}_+(\Omega)$, sodass $\varphi_n \le \varphi_{n+1} \le f$ $\forall n$ und $\varphi_n \overset{n\to\infty}{\to} f$ punktweise.

   Z.B. die folgende Folge funktioniert:

$$\varphi_n := \sum_{k=0}^{n2^n-1}\frac{k}{2^n}\mathbf{1}_{I_k} \qquad \text{mit} \qquad I_k := f^{-1}\left(\left[\frac{k}{2^n},\frac{k+1}{2^n}\right)\right). \tag{1.12}$$

   Dabei sind I_k messbar, da f messbar ist.

2. Aus der monotone Konvergenz folgt, dass

$$\int_\Omega f\ d\mu = \lim_{n\to\infty}\sum_{k=0}^{n2^n-1}\frac{k}{2^n}\mu(I_k).$$

Folgende Resultate sind weitere Anwendungen der monotone Konvergenz.

Lemma 1.31. *Sei $(\Omega,\mathcal{F},\mu)$ ein Maßraum.*

(i) *$f\colon \Omega \to \mathbb{R}$ und $g\colon \Omega \to \mathbb{R}$ messbar $\Rightarrow af+bg$ messbar $\forall a,b \in \mathbb{R}$.*

(ii) *Für alle $f,g \in L^1(\Omega,\mu)$ und $a,b \in \mathbb{R}$ gilt $af+bg \in L^1(\Omega,\mu)$ und*

$$\int_\Omega (af+bg)\,d\mu \ = \ a\int_\Omega f\,d\mu + b\int_\Omega g\,d\mu.$$

Beweis. Analysis 3. □

Lemma 1.32. *Sei* $(\Omega, \mathcal{F}, \mu)$ *ein Maßraum.*

(i) Für alle $f, g \in L^1(\Omega, \mu)$ *mit* $0 \leq f \leq g$ *punktweise (d.h.* $0 \leq f(\omega) \leq g(\omega)$ $\forall \omega \in \Omega$*) gilt*

$$0 \leq \int_\Omega f \, d\mu \leq \int_\Omega g \, d\mu.$$

(ii) Sei $f \in L^1(\Omega, \mu)$ *und* $n \mapsto f_n \in L^1(\Omega, \mu)$*, sodass* $f, f_n \geq 0$ *und* $f(\omega) = \sum_{n \in \mathbb{N}} f_n(\omega)$ $\forall \omega \in \Omega$*. Dann gilt*

$$\sum_{n \in \mathbb{N}} \int_\Omega f_n \, d\mu = \int_\Omega \left[\sum_{n \in \mathbb{N}} f_n \right] d\mu = \int_\Omega f \, d\mu.$$

Beweis.
(i) Aus der Linearität (Lemma 1.31) folgt

$$\int_\Omega g \, d\mu - \int_\Omega f \, d\mu = \int_\Omega (g - f) \, d\mu.$$

Da $g - f \geq 0$ gilt, folgt die Aussage aus der Definition 1.29(ii).

(ii) Sei $F_N := \sum_{n=1}^N f_n$. Aus Lemma 1.31 folgt, dass F_N messbar und

$$\int_\Omega F_N \, d\mu = \sum_{n=1}^N \int_\Omega f_n \, d\mu.$$

Die Folge $N \mapsto F_N$ ist monoton wachsend $0 \leq F_N \leq F_{N+1}$ und konvergiert punktweise gegen f. Aus der monotonen Konvergenz folgt, dass

$$\sum_{n \in \mathbb{N}} \int_\Omega f_n \, d\mu = \lim_{N \to \infty} \sum_{n=1}^N \int_\Omega f_n \, d\mu = \lim_{N \to \infty} \int_\Omega F_N \, d\mu = \int_\Omega \left[\lim_{N \to \infty} F_N \right] d\mu = \int_\Omega f \, d\mu.$$

□

Satz 1.33. *(Lemma von Fatou) Sei* $(\Omega, \mathcal{F}, \mu)$ *ein Maßraum.*
Sei $f \colon \Omega \to [0, \infty]$ *eine messbare Funktion und* $n \mapsto f_n \colon \Omega \to [0, \infty]$ *eine Funktionenfolge, sodass*

- f_n *ist messbar* $\forall n \in \mathbb{N}$ *und*
- $f = \liminf_{n \to \infty} f_n$ *punktweise, d.h.* $\liminf_{n \to \infty} f_n(\omega) = f(\omega)$ $\forall \omega \in \Omega$.

Dann gilt

$$\int_\Omega f \, d\mu = \int_\Omega \left[\liminf_{n \to \infty} f_n \right] d\mu \leq \liminf_{n \to \infty} \int_\Omega f_n \, d\mu.$$

Beweis. Es gilt $\liminf_{n \to \infty} f_n = \lim_{n \to \infty} \inf_{k \geq n} f_k$. Die Funktion $F_n := \inf_{k \geq n} f_k$ ist messbar (Übung) und nicht-negativ, also $F_n \geq 0$. Dazu ist die Folge $n \mapsto F_n$ monoton wachsend und $\lim_{n \to \infty} F_n = f$ punktweise. Monotone Konvergenz ergibt

$$\int_\Omega \left[\liminf_{n \to \infty} f_n \right] d\mu = \int_\Omega \left[\lim_{n \to \infty} F_n \right] d\mu = \lim_{n \to \infty} \int_\Omega F_n \, d\mu.$$

Es gilt $F_n(\omega) \leq f_k \ \forall k \geq n$ und aus Lemma 1.32(i) folgt

$$\int_\Omega F_n \, d\mu \leq \int_\Omega f_k \, d\mu \quad \forall k,$$

und deshalb

$$\int_\Omega F_n \, d\mu \leq \inf_{k \geq n} \int_\Omega f_k \, d\mu.$$

Daher gilt

$$\int_\Omega \left[\liminf_{n\to\infty} f_n\right] d\mu = \lim_{n\to\infty} \int_\Omega F_n \, d\mu \leq \lim_{n\to\infty} \inf_{k\geq n} \int_\Omega f_k \, d\mu = \liminf_{n\to\infty} \int_\Omega f_n \, d\mu.$$

□

Definition 1.34 (fast überall)**.** *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $\omega \mapsto A(\omega)$ eine Aussage. Die Aussage A gilt* $\mu-$fast überall*, falls eine messbare Menge $N \in \mathcal{F}$ existiert, sodass*

- $\mu(N) = 0$ *(N ist eine Nullmenge),*
- $A(\omega)$ *gilt* $\forall \omega \in \Omega \setminus N$.

Lemma 1.35. *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum.*

(i) Sei $f\colon \Omega \to \overline{\mathbb{R}}$ eine messbare Funktion.
Falls $f = 0$ $\mu-$fast überall, dann ist f integrierbar und $\int_\Omega f \, d\mu = 0$.

(ii) Seien $f, g\colon \Omega \to \overline{\mathbb{R}}$ zwei integrierbare Funktionen.
Falls $f = g$ $\mu-$fast überall, dann gilt $\int_\Omega f \, d\mu = \int_\Omega g \, d\mu$.

Beweis. (i) Sei $f = 0$ $\mu-$fast überall, also $\exists N \in \mathcal{F}$, sodass $\mu(N) = 0$ und $f(\omega) = 0 \ \forall \omega \notin N$.

- Sei f einfach. In diesem Fall $f = \sum_{j=1}^n x_j \mathbf{1}_{A_j}$, wobei $x_1, \ldots x_n \in \mathbb{R}$, $A_1, \ldots, A_n \in \mathcal{F}$, $x_j \neq x_k \neq 0$ und $A_j \cap A_k = \varnothing \ \forall j \neq k$.
  Aus $f(\omega) = 0 \ \forall \omega \notin N$ folgt, dass $A_j \subset N \ \forall j$ und deshalb $0 \leq \mu(A_j) \leq \mu(N) = 0$ gilt. Wir rechnen
  $$\int_\Omega f \, d\mu = \sum_{j=1}^n x_j \mu(A_j) = \sum_{j=1}^n x_j \cdot 0 = 0.$$

- [*Dieser Teil wurde nicht in der Vorlesung besprochen.*
  Sei $f\colon \Omega \to [0, \infty]$ eine nichtnegative messbare Funktion. Es gibt eine Folge einfacher Funktionen $n \mapsto \varphi_n \in \mathcal{S}_+(\Omega)$, sodass $0 \leq \varphi_n \leq \varphi_{n+1} \leq f$ und $\lim_{n\to\infty} \varphi_n = f$. Aus $0 \leq \varphi_n \leq f$ folgt, dass $\varphi_n(\omega) = 0 \ \forall \omega \notin N$ und deshalb
  $$\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega \varphi_n \, d\mu = \lim_{n\to\infty} 0 = 0.$$

- Sei $f\colon \Omega \to \overline{\mathbb{R}}$ eine messbare Funktion. Dann gilt $f = f_+ - f_-$ mit $f_\pm$ nichtnegative messbare Funktionen. Aus $f_\pm(\omega) > 0 \Rightarrow f_\mp(\omega) = 0$ folgt, dass $f_\pm(\omega) = 0 \ \forall \omega \notin N$ und deshalb
  $$\int_\Omega f \, d\mu = \int_\Omega f_+ \, d\mu - \int_\Omega f_- \, d\mu = 0 - 0 = 0.$$

(ii) Folgt aus $\int_\Omega f\,d\mu - \int_\Omega g\,d\mu = \int_\Omega (f-g)\,d\mu$ und (i). $\Big]$ □

Satz 1.36 (dominierte Konvergenz)**.** *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum.*
Sei $f\colon \Omega \to \overline{\mathbb{R}}$ eine messbare Funktion und $n \mapsto f_n\colon \Omega \to \overline{\mathbb{R}}$ eine Funktionenfolge, sodass

- *f_n messbar ist,*
- *$f_n \overset{n\to\infty}{\to} f$ $\mu-$fast überall, d.h. $\exists N \in \mathcal{F}$, sodass $\mu(N) = 0$ und $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ $\forall \omega \in \Omega \setminus N$.*

Sei $g\colon \Omega \to [0,\infty]$ eine messbare Funktion, sodass

- *$\int_\Omega g\,d\mu < \infty$,*
- *$|f_n| \le g$ $\mu-$fast überall.*

Dann sind f_n und f integrierbar und

$$\int_\Omega f\,d\mu = \lim_{n\to\infty} \int_\Omega f_n\,d\mu.$$

Beweis. Analysis 3. □

Beispiel Sei $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L})$. Wir betrachten die Folge $f_n \coloneqq \mathbf{1}_{[n,n+1]}$. Die Funktion f_n ist messbar und $\int_\mathbb{R} f_n\,d\mathcal{L} = \mathcal{L}([n,n+1]) = 1\ \forall n$. Die Folge konvergiert punktweise gegen 0 $\lim_{n\to\infty} f_n = 0$, weswegen

$$\int_\mathbb{R} \left[\lim_{n\to\infty} f_n\right] d\mathcal{L} = \int_\mathbb{R} 0\,d\mathcal{L} = 0 \ne 1 = \lim_{n\to\infty} \int_\mathbb{R} f_n\,d\mathcal{L}.$$

Dies widerspricht *nicht* Satz 1.36, da die kleinste Majorante für die Folge f_n die Funktion $1_{[0,\infty)}$ ist, welche allerdings nicht integrierbar ist:

$$\int_\mathbb{R} 1_{[0,\infty)}\,d\mathcal{L} = \lim_{n\to\infty} \int 1_{[0,n)}\,d\mathcal{L} = \lim_{n\to\infty} n = \infty.$$

Satz 1.37 (Satz von Fubini (Version 1))**.** *Seien $(\Omega_1, \mathcal{F}_1, \mu_1)$ $(\Omega_2, \mathcal{F}_2, \mu_2)$ zwei Maßräume, mit $\sigma-$endlichen Maßen.*
Sei $f\colon \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ eine $\mathcal{F}_1 \otimes \mathcal{F}_2 - \mathcal{B}(\overline{\mathbb{R}})-$messbare Funktion. (siehe auch Definition 1.7).
Für $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$ betrachten wir die Funktionen

$$\begin{aligned} f_{\omega_1}\colon\ & \Omega_2 \to \overline{\mathbb{R}} \\ & \omega_2 \mapsto f_{\omega_1}(\omega_2) \coloneqq f(\omega_1,\omega_2), \end{aligned} \qquad \begin{aligned} f_{\omega_2}\colon\ & \Omega_1 \to \overline{\mathbb{R}} \\ & \omega_1 \mapsto f_{\omega_2}(\omega_1) \coloneqq f(\omega_1,\omega_2). \end{aligned}$$

(i) Die Funktion f_{ω_1} ist $\mathcal{F}_2 - \mathcal{B}(\overline{\mathbb{R}})$ messbar $\forall \omega_1 \in \Omega_1$.
Analog ist die Funktion f_{ω_2} $\mathcal{F}_1 - \mathcal{B}(\overline{\mathbb{R}})$ messbar $\forall \omega_2 \in \Omega_2$.

(ii) Sei $f \ge 0$.
Dann ist die Funktion

$$\begin{aligned} F_{f,1}\colon\ & \Omega_1 \to \overline{\mathbb{R}} \\ & \omega_1 \mapsto F_{f,1}(\omega_1) \coloneqq \textstyle\int_{\Omega_2} f_{\omega_1} d\mu_2 = \int_{\Omega_2} f(\omega_1,\omega_2) d\mu_2(\omega_2) \end{aligned}$$

wohldefiniert und $\mathcal{F}_1 - \mathcal{B}(\overline{\mathbb{R}})-$messbar. Die analog definierte Funktion $F_{f,2}\colon \Omega_2 \to \overline{\mathbb{R}}$ ist auch wohldefiniert und $\mathcal{F}_2 - \mathcal{B}(\overline{\mathbb{R}})-$messbar.

Weiter gilt

$$\int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \;=\; \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \qquad (1.13)$$

Insbesondere ist $F_{f,1}$ integrierbar genau dann, wenn $F_{f,2}$ integrierbar ist und in diesem Fall stimmen die zwei Integrale überein.

(iii) Sei $f\colon \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$, sodass $\int_{\Omega_1} \left(\int_{\Omega_2} |f(\omega_1, \omega_2)|\, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) < \infty$.

Dann $\exists N_1 \in \mathcal{F}_1, N_2 \in \mathcal{F}_2$ Nullmengen, also $\mu_i(N_i) = 0, i = 1, 2$, sodass

$$\int_{\Omega_2} |f(\omega_1, \omega_2)| d\mu_2(\omega_2) < \infty \qquad \forall \omega_1 \in \Omega_1 \setminus N_1$$
$$\int_{\Omega_1} |f(\omega_1, \omega_2)| d\mu_1(\omega_1) < \infty \qquad \forall \omega_2 \in \Omega_2 \setminus N_2,$$

gilt und die Funktionen

$$\begin{aligned} F_{f,1}\colon \quad & \Omega_1 \to \quad \mathbb{R} \\ & \omega_1 \mapsto \quad F_{f,1}(\omega_1) := \begin{cases} \int_{\Omega_2} f_{\omega_1} d\mu_2 = \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) & \text{falls } \omega_1 \in \Omega_1 \setminus N_1, \\ 0 & \text{falls } \omega_1 \in N_1, \end{cases} \\ F_{f,2}\colon \quad & \Omega_2 \to \quad \mathbb{R} \\ & \omega_2 \mapsto \quad F_{f,2}(\omega_2) := \begin{cases} \int_{\Omega_1} f_{\omega_2} d\mu_1 = \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) & \text{falls } \omega_2 \in \Omega_2 \setminus N_2, \\ 0 & \text{falls } \omega_2 \in N_2, \end{cases} \end{aligned}$$

sind wohldefiniert und integrierbar. In diesem Fall gilt auch (1.13).

Beweis. Wir werden den Beweis im Fall von Wahrscheinlichkeitsmaßen im Abschnitt 2.4 zeigen (siehe Satz 2.11). Der Beweis im allgemeinen Fall ist fast identisch. □

Folgende Resultate über das Lebesgue Integral werden sehr nützlich sein.

Lemma 1.38 (partielle Integration). *Sei $f \in C^1(\mathbb{R}; \mathbb{R})$. Falls $f, f' \in L^1(\mathbb{R}, \mathcal{L})$, dann gilt*

$$\int_{\mathbb{R}} f'(x)\; d\mathcal{L}(x) = 0.$$

Beweis. Analysis 3. □

Satz 1.39 (Transformationssatz in $\mathbb{R}$). *Sei $U, V \subset \mathbb{R}$ offene Mengen und $\varphi\colon U \to V$ ein C^1-Diffeomorphismus (d.h. $\varphi \in C^1(U; V)$, φ ist invertierbar mit $\varphi^{-1} \in C^1(U; V)$).*
Sei $f\colon V \to \overline{\mathbb{R}}$ eine $\mathcal{B}(V) - \mathcal{B}(\overline{\mathbb{R}})-$messbare Funktion.

Die Funktion $f\colon V \to \overline{\mathbb{R}}$ ist genau dann integrierbar, wenn die Funktion $(f \circ \varphi) \cdot \varphi'\colon U \to \overline{\mathbb{R}}$ integrierbar ist.
In diesem Fall gilt

$$\int_V f(x)\, d\mathcal{L}(x) = \int_U f \circ \varphi(y)\, \varphi'(y)\, d\mathcal{L}(y).$$

Beweis. Analysis 3. □

[7: 04.11.2025]

[8: 07.11.2025]

1.4.3 Abbildungen von Maßen

Definition 1.40. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $(\tilde{\Omega}, \tilde{\mathcal{F}})$ *ein Messraum und* $f\colon \Omega \to \tilde{\Omega}$ *eine* $\mathcal{F} - \tilde{\mathcal{F}}-$*messbare Funktion. Die Funktion*

$$\begin{aligned} \mathbb{P}_f\colon \quad \tilde{\mathcal{F}} &\to [0,1] \\ A &\mapsto \mathbb{P}_f(A) := \mathbb{P} \circ f^{-1}(A) = \mathbb{P}(f^{-1}(A)) \end{aligned}$$

heißt das von f auf $(\tilde{\Omega}, \tilde{\mathcal{F}})$ induzierte Maß *oder das* Bildmaß von $\mathbb{P}$ unter f. *Falls* $(\tilde{\Omega}, \tilde{\mathcal{F}}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, so heißt* $\mathbb{P}_f$ *die* Verteilung des Zufallsvariable f.

Lemma 1.41. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $(\tilde{\Omega}, \tilde{\mathcal{F}})$ *ein Messraum.*

(i) Für jede $\mathcal{F} - \tilde{\mathcal{F}}-$*messbare Funktion* $f\colon \Omega \to \tilde{\Omega}$ *ist die Funktion* $\mathbb{P}_f$ *wohldefiniert und ein Wahrscheinlichkeitsmaß auf* $(\tilde{\Omega}, \tilde{\mathcal{F}})$.

(ii) Sei $(\tilde{\Omega}, \tilde{\mathcal{F}}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *und* $f \in L^1(\Omega, \mathbb{P})$.

  (a) Es gilt $\int_\Omega f(\omega)\, d\mathbb{P}(\omega) = \int_\mathbb{R} x\, d\mathbb{P}_f(x)$.

  (b) Sei $g\colon \mathbb{R} \to \mathbb{R}$ *eine* $\mathcal{B}(\mathbb{R}) - \mathcal{B}(\mathbb{R})$*-messbare Funktion mit* $g \circ f \in L^1(\Omega, \mathbb{P})$. *Dann gilt*

$$\int_\Omega g \circ f(\omega)\, d\mathbb{P}(\omega) = \int_\mathbb{R} g(x)\, d\mathbb{P}_f(x).$$

Notation: Für $g\colon \mathbb{R} \to \mathbb{R}$ messbar mit $g \circ X \in L^1(\Omega, \mathbb{P})$ schreiben wir

$$\mathbb{E}[g(X)] := \int_\Omega g \circ X\, d\mathbb{P} = \int_\mathbb{R} g(x)\, d\mathbb{P}_f(x). \tag{1.14}$$

Beweis. (i) Übung.

(ii) (a) folgt aus (b) mit $g(x) = x$.

(ii) (b) Für $g = \mathbf{1}_B$ mit $B \in \mathcal{B}(\mathbb{R})$ ist $g \circ f(\omega) = \mathbf{1}_{f^{-1}(B)}(\omega)$ die Indikatorfunktion der messbaren Menge $f^{-1}(B)$. Wir rechnen

$$\begin{aligned} \int_\Omega g \circ f(\omega)\, d\mathbb{P}(\omega) &= \int_\Omega \mathbf{1}_{f^{-1}(B)}(\omega)\, d\mathbb{P}(\omega) = \mathbb{P}(f^{-1}(B)) = \mathbb{P}_f(B) = \int_\mathbb{R} \mathbf{1}_B(x)\, d\mathbb{P}_f(x) \\ &= \int_\mathbb{R} g(x)\, d\mathbb{P}_f(x). \end{aligned}$$

Die andere Fälle werden durch maßtheoretische Induktion bewiesen (Übung):

- Für einfache Funktionen wendet man direkte Rechnung an;
- Nichtnegative Funktionen können immer als monoton steigend punktweise Grenzwert nichtnegativer einfachen Funktionen (siehe (1.12)) dargestellt und damit durch monotone Konvergenz behandelt werden, siehe Satz 1.30;
- Schließlich wird der Fall einer allgemeinen Funktion f durch die Darstellung $f = f_+ - f_-$ und den Resultaten für nichtnegative Funktionen behandelt.

□

Wichtige Beispiele von diskrete Verteilungen Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X\colon \Omega \to \mathbb{R}$ eine Zufallsvariable.

Bernoulli Verteilung: $Ber(p)$ mit $p \in [0,1]$.
Die Zufallsvariable heißt *Bernoulli verteilt*, geschrieben $X \sim Ber(p)$, falls

$$\mathbb{P}_X = p\delta_1 + (1-p)\delta_0. \tag{1.15}$$

Die mögliche Werte (mit Wahrscheinlichkeit 1) sind also $X \in \{0,1\}$ und wir rechnen $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}_X(1) + \mathbb{P}_X(0) = p + (1-p) = 1$.
Die Verteilungsfunktion $F_X(t) = \mathbb{P}(X(\omega) \leq t) = \int_{(-\infty,t]} 1 d\mathbb{P}_X$ ist gegeben durch

$$F_X(t) = \begin{cases} 0 & \text{falls } t < 0 \\ (1-p) & \text{falls } 0 \leq t < 1 \\ 1 & \text{falls } 1 \leq t. \end{cases}$$

Z.B. gilt für $X \coloneqq \mathbf{1}_A$ mit $A \in \mathcal{F}$, $X \sim Ber(\mathbb{P}(A))$.

Binomial Verteilung: $Bin(n,p) = B_{n,p}$ mit $n \in \mathbb{N}$ und $p \in [0,1]$.
Die Zufallsvariable heißt *binomial verteilt*, geschrieben $X \sim Bin(n,p)$, falls

$$\mathbb{P}_X = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k. \tag{1.16}$$

Die mögliche Werte sind also $X \in \{0, 1, \ldots, n\}$ und wir rechnen

$$\mathbb{P}_X(\mathbb{R}) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Die Verteilungsfunktion $F_X(t) = \mathbb{P}(X(\omega) \leq t) = \int_{(-\infty,t]} 1 d\mathbb{P}_X$ ist gegeben durch

$$F_X(t) = \begin{cases} 0 & \text{falls } t < 0 \\ \sum_{k=0}^{q} \binom{n}{k} p^k (1-p)^{n-k} & \text{falls } q \leq t < q+1,\ q = 0, \ldots, n-1 \\ 1 & \text{falls } n \leq t. \end{cases}$$

Betrachten wir z.B. n unabhängige (wird später definiert) Münzwürfe, modelliert durch $\omega \in \Omega \coloneqq \{\text{Kopf, Zahl}\}^n$. Hierbei sei $p \in [0,1]$ die Wahrscheinlichkeit, dass bei einem Wurf Kopf eintritt. $X(\omega)$ ist definiert als die Anzahl der aufgetretenen Köpfe in einer Wurffolge ω. Dann gilt $X \sim Bin(n,p)$.

Poisson Verteilung: $Poi(\lambda)$ mit $\lambda > 0$.
Die Zufallsvariable heißt *Poisson-verteilt*, geschrieben $X \sim Poi(\lambda)$, falls

$$\mathbb{P}_X = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta_k. \tag{1.17}$$

Es gilt also $X \in \mathbb{N}$ mt Wahrscheinlichkeit 1. Wir rechnen

$$\mathbb{P}_X(\mathbb{R}) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Die Verteilungsfunktion $F_X(t) = \mathbb{P}(X(\omega) \leq t) = \int_{(-\infty,t]} 1 d\mathbb{P}_X$ ist gegeben durch

$$F_X(t) = \begin{cases} 0 & \text{falls } t < 0 \\ e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} & \text{falls } n \leq t < n+1, \; n \in \mathbb{N}. \end{cases}$$

Es gilt $0 < F_X(t) < 1 \; \forall t > 0$ und $\lim_{\to\infty} F_X(t) = 1$.
Diese Verteilung kommt z.B. in Nuklearzerfall vor.

Geometrische Verteilung: $Geo(q)$ mit $q \in [0,1)$.
Die Zufallsvariable heißt *geometrisch verteilt*, geschrieben $X \sim Geo(q)$, falls

$$\mathbb{P}_X = \sum_{k=0}^{\infty} (1-q) q^k \delta_k. \tag{1.18}$$

Es gilt also auch hier $X \in \mathbb{N}$ mit Wahrscheinlichkeit 1. Wir rechnen

$$\mathbb{P}_X(\mathbb{R}) = (1-q) \sum_{k=0}^{\infty} q^k = (1-q) \frac{1}{(1-q)} = 1.$$

Die Verteilungsfunktion $F_X(t) = \mathbb{P}(X(\omega) \leq t) = \int_{(-\infty,t]} 1 d\mathbb{P}_X$ ist gegeben durch

$$F_X(t) = \begin{cases} 0 & \text{falls } t < 0 \\ 1 - q^{n+1} & \text{falls } n \leq t < n+1, \; n \in \mathbb{N}. \end{cases}$$

Es gilt $0 < F_X(t) < 1 \; \forall t > 0$ und $\lim_{\to\infty} F_X(t) = 1$.

Betrachten wir z.B. abzählbar viele unabhängige (wird später definiert) Münzwürfe, modelliert durch $\omega \in \Omega := \{\text{Kopf, Zahl}\}^{\mathbb{N}} = \{(\omega_1, \omega_2, \ldots) | \; \omega_i \in \{\text{Kopf, Zahl}\} \; \forall i \in \mathbb{N}\}$. Hierbei sei $p \in [0,1]$ die Wahrscheinlichkeit, dass bei einem Wurf Kopf eintritt. $X(\omega)$ ist definiert durch

$$X(\omega) := \sup\{j \in \mathbb{N} | \; \omega_i = \text{Kopf} \; \forall 1 \leq i \leq j\}.$$

mit der Konvention $X(\omega) = 0$, falls $\omega_1 = \text{Zahl}$. Dann gilt $X \sim Bin(n,p)$ und insbesondere $\mathbb{P}(X(\omega) = \infty) = 0$.

Stetige Verteilungen

Definition 1.42. *Sei $\mathbb{P} \colon \mathcal{B}(\mathbb{R}) \to [0,1]$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*
$\mathbb{P}$ heißt absolut stetig *bezüglich des Lebesgue Maßes $\mathcal{L}$, falls eine integrierbare Funktion $\rho \colon \mathbb{R} \to [0, \infty)$ existiert, sodass*

$$\mathbb{P}(B) = \int_B \rho \, d\mathcal{L} = \int_{\mathbb{R}} \mathbf{1}_B \, \rho \, d\mathcal{L}$$

gilt $\forall B \in \mathcal{B}(\mathbb{R})$. Insbesondere gilt $\int_{\mathbb{R}} \rho \, d\mathcal{L} = 1$.
Die Funktion ρ heißt Dichte *von $\mathbb{P}$.*

Notation. Folgende Notationen werden oft verwendet:

$$\mathbb{P} = \rho \mathcal{L}, \quad d\mathbb{P}(x) = \rho(x) d\mathcal{L}(x), \quad \mathbb{P}(dx) = \rho(x) \mathcal{L}(dx).$$

Bemerkungen.

- $\forall B \in \mathcal{B}(\mathbb{R})$ gilt $\mathcal{L}(B) = 0 \Rightarrow \mathbb{P}(B) = 0:$
$$\mathbb{P}(B) = \int_{\mathbb{R}} \mathbf{1}_B \rho \, d\mathcal{L} = 0$$
da $\mathbf{1}_B(x)\rho(x) = 0$ $\mathcal{L}-$fast überall gilt.
- ρ ist nicht eindeutig: $\forall \tilde{\rho} \colon \Omega \to [0, \infty)$ messbar, sodass $\rho = \tilde{\rho}$ $\mathcal{L}-$fast überall, gilt
$$\int_B \rho \, d\mathcal{L} = \int_B \tilde{\rho} \, d\mathcal{L} \qquad \forall B \in \mathcal{B}(\mathbb{R}).$$
- Da $\mathcal{L}(\{t\}) = 0$ $\forall t \in \mathbb{R}$ gilt
$$\int_{[a,b]} \rho \, d\mathcal{L} = \int_{[a,b)} \rho \, d\mathcal{L} = \int_{(a,b]} \rho \, d\mathcal{L} = \int_{(a,b)} \rho \, d\mathcal{L}.$$
Wir werden also oft $\int_a^b \rho \, d\mathcal{L}$ statt $\int_{[a,b]} \rho \, d\mathcal{L}$ schreiben. Insbesondere gilt
$$F(t) = \mathbb{P}(X \leq t) = \int_{(-\infty,t]} \rho \, d\mathcal{L} = \int_{(-\infty,t)} \rho \, d\mathcal{L} = \int_{-\infty}^{t} \rho \, d\mathcal{L}.$$

Wichtige Beispiele von stetige Verteilungen Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable.

Gleichverteilung: $\mathcal{U}([a,b])$ mit $a < b \in \mathbb{R}$.
Die Zufallsvariable heißt *gleichverteilt*, geschrieben $X \sim \mathcal{U}([a,b])$, falls $\mathbb{P}_X = \rho\mathcal{L}$ mit Dichte
$$\rho(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x). \tag{1.19}$$
Wir rechnen
$$\mathbb{P}_X(\mathbb{R}) = \int_{[a,b]} \frac{1}{b-a} \, d\mathcal{L} = \frac{\mathcal{L}([a,b])}{b-a} = 1.$$
Die Verteilungsfunktion $F_X(t) = \mathbb{P}(X(\omega) \leq t) = \int_{(-\infty,t]} 1 d\mathbb{P}_X$ ist gegeben durch
$$F_X(t) = \begin{cases} 0 & \text{falls } t < a \\ \frac{t-a}{b-a} & \text{falls } a \leq t < b \\ 1 & \text{falls } 1 \leq t. \end{cases}$$
Diese Funktion ist stetig.

Exponentialverteilung: $Exp(\lambda)$ mit $\lambda > 0$.
Die Zufallsvariable heißt *exponential verteilt*, geschrieben $X \sim Exp(\lambda)$, falls $\mathbb{P}_X = \rho\mathcal{L}$ mit Dichte
$$\rho(x) = \mathbf{1}_{[0,\infty]}(x)\lambda e^{-\lambda x}. \tag{1.20}$$
Wir rechnen
$$\mathbb{P}_X(\mathbb{R}) = \lambda \int_{[0,\infty)} e^{-\lambda x} d\mathcal{L}(x) = \lim_{N\to\infty} \lambda \int_{[0,N]} e^{-\lambda x} d\mathcal{L}(x) = \lim_{N\to\infty} (1 - e^{-\lambda N}) = 1.$$

Die Verteilungsfunktion ist

$$F_X(t) = \begin{cases} 0 & \text{falls } t < 0 \\ 1 - e^{-\lambda t} & \text{falls } 0 \leq t. \end{cases}$$

Diese Funktion ist stetig und erfüllt $0 < F_X(t) < 1 \ \forall t > 0$.

[8: 07.11.2025]

[9: 11.11.2025]

Gaußverteilung (Normalverteilung): $\mathcal{N}(m, \sigma^2)$ mit $m \in \mathbb{R}$, $\sigma^2 > 0$.
Die Zufallsvariable heißt *Gaußverteilt* (oder *normal verteilt*), geschrieben $X \sim \mathcal{N}(m, \sigma^2)$, falls $\mathbb{P}_X = \rho\mathcal{L}$ mit Dichte

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \tag{1.21}$$

Wir rechnen mithilfe der Transformation (siehe Lemma 1.39) $\tilde{x} = \frac{x-m}{\sigma}$,

$$\mathbb{P}_X(\mathbb{R}) = \tfrac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{-\frac{(x-m)^2}{2\sigma^2}} d\mathcal{L}(x) = \tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} d\mathcal{L}(x) = \tfrac{1}{\sqrt{2\pi}} \lim_{N\to\infty} \int_{[-N,N]} e^{-\frac{x^2}{2}} d\mathcal{L}(x) = \tfrac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1.$$

Die Verteilungsfunktion ist

$$F_X(t) = \tfrac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-\frac{(x-m)^2}{2\sigma^2}} d\mathcal{L}(x).$$

Diese Funktion ist glatt und erfüllt $0 < F_X(t) < 1 \ \forall t \in \mathbb{R}$.

Cauchy Verteilung: $Cauchy(a)$ mit $a > 0$.
Die Zufallsvariable heißt *Cauchy verteilt*, geschrieben $X \sim Cauchy(a)$, falls $\mathbb{P}_X = \rho\mathcal{L}$ mit Dichte

$$\rho(x) = \frac{a}{\pi} \frac{1}{x^2 + a^2}. \tag{1.22}$$

Wir rechnen mithilfe der Transformation (siehe Lemma 1.39) $\tilde{x} = \frac{x}{a}$,

$$\begin{aligned} \mathbb{P}_X(\mathbb{R}) &= \tfrac{a}{\pi} \int_{\mathbb{R}} \frac{1}{x^2 + a^2} d\mathcal{L}(x) = \tfrac{1}{\pi} \int_{\mathbb{R}} \frac{1}{x^2 + 1} d\mathcal{L}(x) \\ &= \tfrac{1}{\pi} \lim_{N\to\infty} \int_{[-N,N]} \frac{1}{x^2 + 1} d\mathcal{L}(x) = \tfrac{1}{\pi} \lim_{N\to\infty} [\arctan N - \arctan(-N)] = 1. \end{aligned}$$

Die Verteilungsfunktion ist

$$\begin{aligned} F_X(t) &= \tfrac{a}{\pi} \int_{-\infty}^{t} \frac{1}{x^2 + a^2} d\mathcal{L}(x) = \tfrac{1}{\pi} \int_{-\infty}^{\frac{t}{a}} \frac{1}{x^2 + 1} d\mathcal{L}(x) \\ &= \tfrac{1}{\pi} \lim_{N\to\infty} \int_{-N}^{\frac{t}{a}} \frac{1}{x^2 + 1} d\mathcal{L}(x) = \tfrac{1}{\pi} \left[\arctan \frac{t}{a} + \frac{\pi}{2} \right]. \end{aligned}$$

Diese Funktion ist glatt und erfüllt $0 < F_X(t) < 1 \ \forall t \in \mathbb{R}$.

Lemma 1.43. *(Absolute Stetigkeit der Verteilungsfunktion) Sei* $\mathbb{P}\colon \mathcal{B}(\mathbb{R}) \to [0, 1]$ *ein Wahrscheinlichkeitsmaß auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

(i) Falls $\mathbb{P} = \rho\mathcal{L}$ *mit* $\rho\colon \mathbb{R} \to [0,\infty)$ *integrierbar, dann ist die Verteilungsfunktion* $F(t) = \mathbb{P}(X \leq t)$ *absolut stetig.* F *ist differenzierbar* $\mathcal{L}-$*fast überall und es gilt* $F'(t) = \rho(t)$ $\mathcal{L}-$*fast überall. 0*

(ii) Falls die Verteilungsfunktion $F(t) = \mathbb{P}(X \leq t)$ *absolut stetig ist, dann existiert eine integrierbare Funktion* $\rho\colon \mathbb{R} \to [0,\infty)$, *, sodass* $\mathbb{P} = \rho\mathcal{L}$ *und* (1.23) *gilt.*

Erinnerung: $f\colon [a,b] \to \mathbb{R}$ ist absolut stetig, falls Folgendes gilt: Es existiert eine integrierbare Funktion $g\colon [a,b] \to \mathbb{R}$, sodass für alle $t \in [a,b]$

$$f(t) = f(a) + \int_a^t g\, d\mathcal{L}. \tag{1.23}$$

Insbesondere ist f stetig und für $\mathcal{L}-$fast alle $t \in \mathbb{R}$ ist f in t differenzierbar und $f'(t) = g(t)$ (diese Resultat wird in Funktional Analysis bewiesen).

Beweis. [*Dieser Teil wurde nicht in der Vorlesung besprochen.*
(i) Falls $\mathbb{P} = \rho\mathcal{L}$, dann gilt, für alle $a < b \in \mathbb{R}$,

$$F(a) - F(b) = \int_{(a,b]} \rho\, d\mathcal{L} = \int_{(a,b)} \rho\, d\mathcal{L} = \int_a^b \rho\, d\mathcal{L}.$$

Dann ist die Funktion differenzierbar $\mathcal{L}-$fast überall mit $F'(t) = \rho(t)$ (siehe Erinnerung).

(ii) Folgt aus der Definition von absolut Stetigkeit und Verteilungsfunktion.] □

Definition 1.44. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $X\colon \Omega \to \mathbb{R}$ *eine Zufallsvariable.*

(i) Für $p > 1$ *heißt* X $p-$fach integrierbar, *geschrieben* $X \in L^p(\Omega, \mathbb{P})$, *falls* $|X|^p \in L^1(\Omega, \mathbb{P})$.

(ii) Sei $X \in L^1(\Omega, \mathbb{P})$. *Der Erwartungswert von* X *ist definiert durch (siehe auch* (1.14)*)*

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega).$$

(iii) Sei $X \in L^1(\Omega, \mathbb{P}) \cap L^2(\Omega, \mathbb{P})$ *(also* X *ist integrierbar und 2-fach integrierbar). Die Varianz von* X *ist definiert durch*

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

Lemma 1.45. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $X\colon \Omega \to \mathbb{R}$ *eine Zufallsvariable. Für* $X \in L^p(\Omega, \mathbb{P})$ *gilt*

$$\mathbb{E}[|X|^p] = \int_\Omega |X|^p(\omega)\, d\mathbb{P}(\omega) = \int_\mathbb{R} |x|^p\, d\mathbb{P}_X(x).$$

Insbesondere gilt

$$\operatorname{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_\mathbb{R} (x - \mathbb{E}[X])^2\, d\mathbb{P}_X(x).$$

Beweis. Folgt direkt aus Lemma 1.41.

□

Beispiele
Für $X \sim Ber(p)$ rechnen wir

$$\mathbb{E}[X] = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p.$$

Für $X \sim \mathcal{U}([a, b])$ rechnen wir

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x \, d\mathcal{L}(x) = \frac{1}{b-a} \frac{1}{2} \left[b^2 - a^2\right] = \frac{a+b}{2}.$$

Für $X \sim Cauchy(1)$ gilt $X \notin L^1(\Omega, \mathbb{P})$, da

$$\begin{aligned}
\int_\Omega |X(\omega)| \, d\mathbb{P}(\omega) &= \frac{1}{\pi} \int_{\mathbb{R}} \frac{|x|}{1+x^2} d\mathcal{L}(x) \geq \frac{1}{\pi} \int_0^\infty \frac{|x|}{1+x^2} d\mathcal{L}(x) \\
&= \frac{1}{\pi} \int_0^\infty \frac{x}{1+x^2} d\mathcal{L}(x) = \lim_{N\to\infty} \frac{1}{\pi} \int_0^N \frac{x}{1+x^2} d\mathcal{L}(x) \\
&= \lim_{N\to\infty} \frac{1}{2\pi} \ln[1+x^2]\Big|_0^N = \lim_{N\to\infty} \frac{1}{2\pi} \ln(1+N^2) = \infty.
\end{aligned}$$

Die Cauchy Verteilung hat also keinen Erwartungswert.

2 Bedingte Wahrscheinlichkeiten, Unabhängigkeit, Produktmaße

2.1 Bedingte Wahrscheinlichkeit

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Für alle $A, B \in \mathcal{F}$ gilt $\mathbb{P}(A \cap B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\}$. Was uns nun interessiert, ist wie Information über das Ereignis B unsere Annahmen über das Ereignis A beeinflussen. Dazu definieren wir die bedingte Wahrscheinlichkeit.

Definition 2.1. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, und $A, B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$. Dann heißt*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2.1}$$

die bedingte Wahrscheinlichkeit von A gegeben B.

Satz 2.2. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, und $B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$.*

(i) Das Mengensystem

$$\mathcal{F} \cap B := \{A \cap B \mid A \in \mathcal{F}\} \tag{2.2}$$

ist eine $\sigma-$Algebra über B.

(ii) Die durch (2.1) definierte Funktion $\mathbb{P}(\cdot|B) \colon \mathcal{F} \to [0, 1]$ ist ein Wahrscheinlichkeitsmaß auf $(B, \mathcal{F} \cap B)$.

(iii) Sei $B_n \in \mathcal{F}$ $n \in \mathbb{N}$, eine paarweise disjunkte Folge von Mengen, so dass $\cup_{n\in\mathbb{N}} B_n = \Omega$ und $\mathbb{P}(B_n) > 0$ $\forall n \in \mathbb{N}$ gilt. Dann gilt $\forall A \in \mathcal{F}$

$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}} \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

Beweis. (*i*) Übung.

(*ii*) Es gilt $\mathbb{P}(\varnothing|B) = \frac{\mathbb{P}(\varnothing)}{\mathbb{P}(B)} = 0$ und $\mathbb{P}(B|B) = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$. Um $\sigma-$Additivität zu beweisen, sei $\tilde{A}_n \in \mathcal{F} \cap B$, $n \in \mathbb{N}$, eine paarweise disjunkte Folge von Mengen. Aus der Definition von $\mathcal{F} \cap B$ folgt, dass $\tilde{A}_n = A_n \cap B \subset B$ mit $A_n \in \mathcal{F}$ gilt und die $\sigma-$Additivität von $\mathbb{P}$ impliziert

$$\mathbb{P}(\cup_{n\in\mathbb{N}}\tilde{A}_n|B) = \frac{\mathbb{P}(B \cap \cup_{n\in\mathbb{N}}\tilde{A}_n)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_{n\in\mathbb{N}}\tilde{A}_n)}{\mathbb{P}(B)} = \sum_{n\in\mathbb{N}} \frac{\mathbb{P}(\tilde{A}_n)}{\mathbb{P}(B)} = \sum_{n\in\mathbb{N}} \frac{\mathbb{P}(\tilde{A}_n \cap B)}{\mathbb{P}(B)} = \sum_{n\in\mathbb{N}} \mathbb{P}(\tilde{A}_n|B).$$

(*iii*) Übung.

□

Korollar 2.3 (Formel von Bayes)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum, und* $A, B \in \mathcal{F}$ *mit* $\mathbb{P}(A) > 0$ *und* $\mathbb{P}(B) > 0$. *Dann gilt*

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

Beweis. Folgt direkt aus der Formel (2.1). □

2.2 Unabhängige Ereignisse

Definition 2.4. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum.*

(i) $A, B \in \mathcal{F}$ *heißen* unabhängig *bezüglich* $\mathbb{P}$*, falls*

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

(ii) Sei $\mathcal{A} := \{A_1, \dots, A_n\} \subset \mathcal{F}$ *ein endliches Mengensystem.*

  (a) Die Mengen in $\mathcal{A}$ *heißen* paarweise unabhängig *bezüglich* $\mathbb{P}$*, falls* A_i *und* A_j $\forall i \neq j$ *unabhängig sind.*

  (b) Die Mengen in $\mathcal{A}$ *heißen* unabhängig *bezüglich* $\mathbb{P}$*, falls für alle* $k \in \{1, \dots, n\}$ *und* $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ *gilt*

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_j\right) = \prod_{j=1}^{k} \mathbb{P}(A_j).$$

(iii) Seien $\mathcal{F}_1, \dots, \mathcal{F}_n$ *Unter-*$\sigma-$*Algebren von* $\mathcal{F}$ *(also* $\mathcal{F}_i \subset \mathcal{F}$ *und* $\mathcal{F}_i$ *ist eine* σ*-Algebra). Dann heißen* $\mathcal{F}_1, \dots, \mathcal{F}_n$ unabhängig *bezüglich* $\mathbb{P}$*, falls* $\forall A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$ *die Mengen* $A_1, \dots, A_n$ *unabhängig sind.*

Bemerkung 1. Falls $\mathbb{P}(A) > 0$ und $\mathbb{P}(B) > 0$, dann gilt

$$A \text{ und } B \text{ unabhängig} \quad \Leftrightarrow \quad \mathbb{P}(A|B) = \mathbb{P}(A) \quad \Leftrightarrow \quad \mathbb{P}(B|A) = \mathbb{P}(B).$$

Bemerkung 2. A und B sind unabhängig $\Leftrightarrow$ A^c und B^c sind unabhängig.

Beweis:

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) = 1 - [\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)] \\ &= 1 - [\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)] = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = \mathbb{P}(A^c)\mathbb{P}(B^c).\end{aligned}$$

Bemerkung 3. Es gilt $\mathcal{A}$ unabhängig $\Rightarrow \mathcal{A}$ paarweise unabhängig.
Die andere Richtung gilt allgemein nicht.

Beispiel. Ein Test auf eine Krankheit liefert mit Wahrscheinlichkeit $y \in (0,1)$ ein korrektes Ergebnis. Die Wahrscheinlichkeit, krank zu sein, ist $x \in (0,1)$. Nehmen wir an, dass $y = 0.99$. Was ist die Wahrscheinlichkeit krank zu sein, gegeben, dass der Test positiv ist?
Wir schreiben K für das Ergebnis, krank zu sein, TP für das Ergebnis 'Test positiv' und TR für das Ergebnis 'Test richtig'. Es gilt also $\mathbb{P}(K) = x$ und $\mathbb{P}(TR) = y = 0.99$.
Wir nehmen an, dass die Ergebnisse K und TR unabhängig sind.
Unser Ziel ist es, $\mathbb{P}(K|TP)$ zu finden.
Wir merken, dass $TP \cap K = TR \cap K$ und $TP \cap K^c = TR^c \cap K^c$ gilt (Übung). Daher gilt

$$\begin{aligned} \mathbb{P}(TP) &= \mathbb{P}(TP \cap K) + \mathbb{P}(TP \cap K^c) = \mathbb{P}(TR \cap K) + \mathbb{P}(TR^c \cap K^c) \\ &= \mathbb{P}(TR)\mathbb{P}(K) + \mathbb{P}(TR^c)\mathbb{P}(K^c) = yx + (1-y)(1-x). \end{aligned}$$

Daraus folgt

$$\mathbb{P}(K|TP) = \frac{\mathbb{P}(K \cap TP)}{\mathbb{P}(TP)} = \frac{\mathbb{P}(K \cap TR)}{\mathbb{P}(TP)} = \frac{xy}{xy + (1-x)(1-y)} = \frac{x \cdot 0.99}{x \cdot 0.99 + (1-x) \cdot 0.01}.$$

Falls die Wahrscheinlichkeit, krank zu sein, sehr klein ist (z.B. $x = 10^{-4}$), gilt auch

$$\mathbb{P}(K|TP) \sim 100 \cdot x = 10^{-2} \ll 1.$$

Der Test hat eigentlich keine neue Information gebracht, bzw. fast alle positiv getesteten erweisen sich im Nachhinein als gesund.

2.3 Unabhängige Zufallsvariablen

Definition 2.5. *Sei $(\Omega, \mathcal{F})$ ein Messraum und $f \colon \Omega \to \mathbb{R}$ $\mathcal{F}-\mathcal{B}(\mathbb{R})-$messbar. Die von f erzeugte $\sigma-$Algebra ist definert durch*

$$\sigma(f) := \bigcap_{\substack{\tilde{\mathcal{F}} \subset \mathcal{F},\ \text{sodass} \\ \tilde{\mathcal{F}}\ \sigma-\text{Algebra, und } f \text{ ist } \tilde{\mathcal{F}}-\mathcal{B}(\mathbb{R})-\text{messbar}}} \tilde{\mathcal{F}}.$$

Lemma 2.6. *Sei $(\Omega, \mathcal{F})$ ein Messraum und $f \colon \Omega \to \mathbb{R}$ $\mathcal{F} - \mathcal{B}(\mathbb{R})-$messbar.*

(i) Das Mengensystem $f^{-1}(\mathcal{B}(\mathbb{R})) := \{f^{-1}(B) |\ B \in \mathcal{B}(\mathbb{R})\}$ ist eine $\sigma-$Algebra über Ω.

(ii) Es gilt $\sigma(f) = f^{-1}(\mathcal{B}(\mathbb{R}))$.

Beweis. Übungsblatt. □

Definition 2.7. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.*

(i) Seien $X_1, X_2 \colon \Omega \to \mathbb{R}$ Zufallsvariablen.

(a) X_1 und X_2 heißen unabhängig*, falls $\sigma(X_1)$ und $\sigma(X_2)$ unabhängig sind.*

(b) Nehmen wir an, dass $X_1, X_2 \in L^1(\Omega, \mathbb{P})$ und $X_1 \cdot X_2 \in L^1(\Omega, \mathbb{P})$ gilt.
Dann heißen X_1 und X_2 unkorreliert*, falls $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$ gilt.*

(ii) Seien $X_1, X_2, \dots X_n \colon \Omega \to \mathbb{R}$ Zufallsvariablen.

(a) $X_1, \dots, X_n$ heißen unabhängig*, falls $\sigma(X_1), \dots, \sigma(X_n)$ unabhängig sind.*

(b) Nehmen wir an $X_j \in L^1(\Omega, \mathbb{P})$ $\forall j = 1, \dots, n$ und $X_i \cdot X_j \in L^1(\Omega, \mathbb{P})$ $\forall i \neq j$ gilt. Dann heißen $X_1, \dots, X_n$ unkorreliert*, falls $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j]$ $\forall i \neq j$ gilt.*

[9: 11.11.2025]

[10: 14.11.2025]

Lemma 2.8. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X_1, X_2, \dots, X_n \colon \Omega \to \mathbb{R}$ Zufallsvariablen. Folgende Aussagen sind äquivalent.*

(i) $X_1, X_2, \dots, X_n$ sind unabhängig.

(ii) $\forall B_1, B_2, \dots B_n \in \mathcal{B}(\mathbb{R})$ gilt

$$\mathbb{P}\Big(\{X_1 \in B_1\} \cap \{X_2 \in B_2\} \cap \dots \cap \{X_n \in B_n\}\Big) = \prod_{j=1}^n \mathbb{P}(X_j \in B_j).$$

(iii) $\forall g_1, \dots, g_n \colon \mathbb{R} \to \mathbb{R}$ $\mathcal{B}(\mathbb{R}) - \mathcal{B}(\mathbb{R})-$messbare Fuktionen gilt

$$\int_\Omega \prod_{j=1}^n g_j(X_j(\omega))\ d\mathbb{P}(\omega) = \prod_{j=1}^n \int_\Omega g_j(X_j(\omega))\ d\mathbb{P}(\omega),$$

solange alle Integrale wohldefiniert sind.

Beweis. Wir betrachten hier nur den Fall $n = 2$. Der Fall $n > 2$ wird analog bewiesen.

$(i) \Rightarrow (ii)$ aus der Definition von Unabhängigkeit.

$(ii) \Rightarrow (i)$ Aus Lemma 2.6 folgt, dass $\sigma(X_i) = X_i^{-1}(\mathcal{B}(\mathbb{R}))$. Für alle $A_1, A_2 \in \sigma(X_i)$ gilt also $A_i = X_i^{-1}(B_i)$ mit $B_i \in \mathcal{B}(\mathbb{R})$. Wir rechnen

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) &= \mathbb{P}\Big(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)\Big) = \mathbb{P}\Big(\{X_1 \in B_1\} \cap \{X_2 \in B_2\}\Big) \\ &\overset{(ii)}{=} \mathbb{P}(X_1 \in B_1)\ \mathbb{P}(X_2 \in B_2) = \mathbb{P}(A_1)\ \mathbb{P}(A_2),\end{aligned}$$

also sind X_1 und X_2 unabhängig.

$(iii) \Rightarrow (ii)$ Für $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ definieren wir $g_i \coloneqq \mathbf{1}_{B_i}$ $i = 1, 2$. Dann gilt

$$g_i \circ X_i(\omega) = \mathbf{1}_{B_i}(X_i(\omega)) = \mathbf{1}_{X_i^{-1}(B_i)}(\omega) \tag{2.3}$$

$$g_1(X_1(\omega))g_2(X_2(\omega)) = \mathbf{1}_{X_1^{-1}(B_1) \cap X_2^{-1}(B_2)}(\omega), \tag{2.4}$$

und daher

$$\begin{aligned}\mathbb{P}\Big(\{X_1 \in B_1\} \cap \{X_2 \in B_2\}\Big) &= \int_\Omega \mathbf{1}_{X_1^{-1}(B_1) \cap X_2^{-1}(B_2)}(\omega)\, d\mathbb{P}(\omega) = \int_\Omega g_1(X_1(\omega))g_2(X_2(\omega))\, d\mathbb{P}(\omega) \\ &\overset{(iii)}{=} \int_\Omega g_1(X_1(\omega))\, d\mathbb{P}(\omega) \int_\Omega g_2(X_2(\omega))\, d\mathbb{P}(\omega) \\ &= \int_\Omega \mathbf{1}_{X_1^{-1}(B_1)}(\omega)\, d\mathbb{P}(\omega) \int_\Omega \mathbf{1}_{X_2^{-1}(B_2)}(\omega)\, d\mathbb{P}(\omega) \\ &= \mathbb{P}(X_1 \in B_1)\ \mathbb{P}(X_2 \in B_2).\end{aligned}$$

$(ii) \Rightarrow (iii)$ Falls $g_i \coloneqq \mathbf{1}_{B_i}$ $i = 1,2$, mit $B_i \in \mathcal{B}(\mathbb{R})$, rechnen wir

$$\int_\Omega g_1(X_1(\omega))g_2(X_2(\omega))\,d\mathbb{P}(\omega) = \int_\Omega \mathbf{1}_{X_1^{-1}(B_1)\cap X_2^{-1}(B_2)}(\omega)\,d\mathbb{P}(\omega) = \mathbb{P}\Big(\{X_1 \in B_1\} \cap \{X_2 \in B_2\}\Big)$$
$$\overset{(ii)}{=} \mathbb{P}(X_1 \in B_1)\ \mathbb{P}(X_2 \in B_2) = \int_\Omega g_1(X_1(\omega))\,d\mathbb{P}(\omega) \int_\Omega g_2(X_2(\omega))\,d\mathbb{P}(\omega).$$

Die andere Fälle (g_i einfach, $g_i \geq 0$, g_i allgemein) werden durch maßtheoretische Induktion bewiesen (Übung).

□

Bemerkung 1. Seien X_1, X_2 unabhängige Zufallsvariablen und $f\colon \mathbb{R} \to \mathbb{R}$ eine messbare Funktion. Dann sind $f(X_1), f(X_2)$ unabhängige Zufallsvariablen. (Übung)

Bemerkung 2. X_1, X_2 unabhängig $\Rightarrow$ X_1, X_2 unkorreliert.
Die andere Richtung gilt nicht, außer im Falle von Gaußverteilungen.

Beispiel. Seien $X, Y \sim Ber(\frac{1}{2})$ unabhängige Zufallsvariablen. Die Funktionen $Z_\pm \coloneqq X \pm Y$ definieren ebenso Zufallsvariablen.
Die Zufallsvariablen Z_+, Z_- sind unkorreliert, da

$$\mathbb{E}[Z_+Z_-] = \mathbb{E}[(X^2 - Y^2)] = \mathbb{E}[X^2] - \mathbb{E}[Y^2] = 0.$$

Sie sind aber nicht unabhängig:

$$\begin{aligned}
\mathbb{P}(Z_+ = 2) &= \mathbb{P}\Big(\{X = 1\} \cap \{Y = 1\}\Big) = \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = \frac{1}{4} \\
\mathbb{P}(Z_- = 0) &= \mathbb{P}\Big(\{X = 1\} \cap \{Y = 1\}\Big) + \mathbb{P}\Big(\{X = 0\} \cap \{Y = 0\}\Big) \\
&= \mathbb{P}(X = 1)\mathbb{P}(Y = 1) + \mathbb{P}(X = 0)\mathbb{P}(Y = 0) = \frac{1}{2}
\end{aligned}$$
$$\mathbb{P}\Big(\{Z_+ = 2\} \cap \{Z_- = 0\}\Big) = \mathbb{P}\Big(\{X = 1\} \cap \{Y = 1\}\Big) = \frac{1}{4} \neq \frac{1}{4} \cdot \frac{1}{2} = \mathbb{P}(Z_+ = 2)\mathbb{P}(Z_- = 0).$$

2.4 Unabhängige Zufallsvariablen konstruieren: Produkträume.

Seien $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ zwei Messräume. Die Produkt-$\sigma-$Algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ ist definiert durch (siehe Def. 1.7)

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{C}),$$

wobei

$$\mathcal{C} \coloneqq \{C = A \times B \mid A \in \mathcal{F}_1, B \in \mathcal{F}_2\}. \tag{2.5}$$

Bemerkung 1. Es gilt $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$, weil $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{Q}_2)$ wobei
$\mathcal{Q}_2 = \{Q = (a_1, b_1) \times (a_2, b_2) \mid a_i < b_i \in \overline{\mathbb{R}}\}$ das Mengensystem aller offenen Quader ist.

Es gilt aber $\mathcal{M}(\mathbb{R}) \otimes \mathcal{M}(\mathbb{R}) \neq \mathcal{M}(\mathbb{R}^2)$ (wobei $\mathcal{M}(\mathbb{R}^2)$ die Lebesgue $\sigma-$Algebra über $\mathbb{R}^2$ ist): die Menge $A = \{0\} \times M$, mit M die Vitali Menge, erfüllt $A \in \mathcal{M}(\mathbb{R}^2) \setminus \mathcal{M}(\mathbb{R}) \otimes \mathcal{M}(\mathbb{R})$ weil M nicht Lebesgue-messbar ist.

Bemerkung 2. Seien $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ zwei Wahrscheinlichkeiträume. Die Mengensysteme $\mathcal{F}_1 \times \Omega_2$ und $\Omega_1 \times \mathcal{F}_2$ sind Unter-$\sigma-$Algebren von $\mathcal{F}_1 \otimes \mathcal{F}_2$. Wir suchen ein Wahrscheinlichkeitmaß auf $\mathcal{F}_1 \otimes \mathcal{F}_2$,, sodass die $\sigma-$Algebren $\mathcal{F}_1 \times \Omega_2$ und $\Omega_1 \times \mathcal{F}_2$ unabhängig sind.
In diesem Fall sind alle messbare Funktionen $X_1, X_2 \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$, sodass $X_1(\omega_1, \omega_2) = X_1(\omega_1)$ und $X_2(\omega_1, \omega_2) = X_2(\omega_2)$, unabhängige Zufallsvariablen.

Für $C \subset \Omega_1 \times \Omega_2$ definieren wir die Mengen

$$\begin{aligned} C_{\omega_1} &:= \{\omega_2 \in \Omega_2 \mid (\omega_1, \omega_2) \in C\} \subset \Omega_2, \qquad && \forall \omega_1 \in \Omega_1, \\ C_{\omega_2} &:= \{\omega_2 \in \Omega_2 \mid (\omega_1, \omega_2) \in C\} \subset \Omega_1, \qquad && \forall \omega_2 \in \Omega_2. \end{aligned} \tag{2.6}$$

C_{ω_i} hießt der ω_i-Schnitt von C.
Für $f \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ definieren wir die Funktionen

$$\begin{aligned} f_{\omega_1} \colon \quad & \Omega_2 \to \mathbb{R} & \\ & \omega_2 \mapsto f_{\omega_1}(\omega_2) := f(\omega_1, \omega_2) \qquad & \forall \omega_1 \in \Omega_1, \\ f_{\omega_2} \colon \quad & \Omega_1 \to \mathbb{R} & \\ & \omega_1 \mapsto f_{\omega_2}(\omega_1) := f(\omega_1, \omega_2) \qquad & \forall \omega_2 \in \Omega_2. \end{aligned} \tag{2.7}$$

Lemma 2.9. *Seien $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ Messräume.*

(i) Sei $C \in \mathcal{F}_1 \otimes \mathcal{F}_2$.
Dann gilt $C_{\omega_1} \in \mathcal{F}_2$ und $C_{\omega_2} \in \mathcal{F}_1$ für alle $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$.

(ii) Sei $f \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ $\mathcal{F}_1 \otimes \mathcal{F}_2 - \mathcal{B}(\mathbb{R})-$messbar.
Dann ist die Funktion f_{ω_1} $\mathcal{F}_2 - \mathcal{B}(\mathbb{R})-$messbar für alle $\omega_1 \in \Omega_1$.
Analog ist die Funktion f_{ω_2} $\mathcal{F}_1 - \mathcal{B}(\mathbb{R})-$messbar für alle $\omega_2 \in \Omega_2$.

Beweis.
(i) Sei $\omega_1 \in \Omega_1$ gegeben. Wir zeigen, dass $C_{\omega_1} \in \mathcal{F}_2$ $\forall C \in \mathcal{F}_1 \otimes \mathcal{F}_2$ gilt.
Dafür definieren wir das Mengensystem

$$\mathcal{A}_{\omega_1} := \{C \in \mathcal{F}_1 \otimes \mathcal{F}_2 \mid C_{\omega_1} \in \mathcal{F}_2\} \subset \mathcal{F}_1 \otimes \mathcal{F}_2.$$

Das Mengensystem $\mathcal{A}_{\omega_1}$ ist eine $\sigma-$Algebra (Übung) und enthält $\mathcal{C}$ (siehe (2.5)), also $\mathcal{C} \subset \mathcal{A}_{\omega_1}$, da für $C = A \times B$ mit $A \in \mathcal{F}_1, B \in \mathcal{F}_2$ gilt

$$C_{\omega_1} = \begin{cases} \varnothing \text{ falls } & \omega_1 \in \Omega_1 \setminus A \\ B \text{ falls } & \omega_1 \in A. \end{cases}$$

Daraus folgt

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{C}) \subset \mathcal{A}_{\omega_1} \subset \mathcal{F}_1 \otimes \mathcal{F}_2.$$

Also gilt $C_{\omega_1} \in \mathcal{F}_2$ $\forall C \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

Die Aussage für C_{ω_2} wird analog bewiesen.

(ii) Sei $f \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ $\mathcal{F}_1 \otimes \mathcal{F}_2 - \mathcal{B}(\mathbb{R})-$messbar und $\omega_1 \in \Omega_1$ gegeben. Wir zeigen, dass f_{ω_1} $\mathcal{F}_2 - \mathcal{B}(\mathbb{R})-$messbar ist.
Sei $B \in \mathcal{B}(\mathbb{R})$. Aus der Messbarkeit von f folgt, dass $f^{-1}(B) \in \mathcal{F}_1 \otimes \mathcal{F}_2$ gilt, und daher, mit (i) gilt auch $\left(f^{-1}(B)\right)_{\omega_1} \in \mathcal{F}_2$. Aus

$$(\omega_1, \omega_2) \in f^{-1}(B) \Leftrightarrow f(\omega_1, \omega_2) \in B \Leftrightarrow f_{\omega_1}(\omega_2) \in B \Leftrightarrow \omega_2 \in (f_{\omega_1})^{-1}(B),$$

folgt

$$(f_{\omega_1})^{-1}(B) = \{\omega_2 \in \Omega_2 \mid \omega_2 \in (f_{\omega_1})^{-1}(B)\} = \{\omega_2 \in \Omega_2 \mid (\omega_1,\omega_2) \in f^{-1}(B)\} = \big(f^{-1}(B)\big)_{\omega_1} \in \mathcal{F}_2.$$

Also ist f_{ω_1} messbar.

Die Aussage für f_{ω_2} wird analog bewiesen.

□

Bemerkung. Die Aussage gilt nicht für $\mathcal{M}(\mathbb{R}^2)$, da die Menge $A = \{0\} \times M$ (wobei M erneut eine Vitali-Menge ist) erfüllt

$$A_x = \begin{cases} \varnothing \in \mathcal{M}(\mathbb{R}) & \text{falls} \quad x \neq 0 \\ M \notin \mathcal{M}(\mathbb{R}) & \text{falls} \quad x = 0. \end{cases}$$

Das widerspricht allerdings nicht Lemma 2.9, da $\mathcal{M}(\mathbb{R}^2)$ nicht eine Produkt-$\sigma-$Algebra ist.

[10: 14.11.2025]

[11: 18.11.2025]

Satz 2.10. *Seien $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ zwei Wahrscheinlichkeiträume.*

(i) $\exists!$ *Wahrscheinlichkeitsmaß* $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2 \colon \mathcal{F}_1 \otimes \mathcal{F}_2 \to [0,1]$, *sodass*

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\, \mathbb{P}_2(A_2) \qquad \forall A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2 \tag{2.8}$$

gilt.

(ii) $\forall C \in \mathcal{F}_1 \otimes \mathcal{F}_2$ *gilt*

$$\mathbb{P}(C) = \int_{\Omega_1} \mathbb{P}_2(C_{\omega_1}) d\mathbb{P}_1(\omega_1) = \int_{\Omega_2} \mathbb{P}_1(C_{\omega_2}) d\mathbb{P}_2(\omega_2), \tag{2.9}$$

wobei C_{ω_i} der ω_i-Schnitt von C ist (siehe (2.6)*).*

Bemerkung. Dasselbe Resultat (mit fast demselben Beweis) gilt, wenn $\mathbb{P}_1, \mathbb{P}_2$ durch zwei σ-endliche Maße μ_1, μ_2 ersetzt werden. Z.B. gilt das Resultat für $\mathcal{L} \otimes \mathcal{L}$.

Beweis.
Eindeutigkeit. Seien $\mathbb{P}, \tilde{\mathbb{P}}$ zwei Wahrscheinlichkeitsmaße, die (2.8) erfüllen.
Dann gilt $\mathbb{P}_{|\mathcal{C}} = \tilde{\mathbb{P}}_{|\mathcal{C}}$, wobei

$$\mathcal{C} \coloneqq \{C = A \times B \mid A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}. \tag{2.10}$$

Das Mengensystem $\mathcal{C}$ ist $\cap-$stabil und erzeugt $\mathcal{F}_1 \otimes \mathcal{F}_2$. Aus dem Eindeutigkeitssatz Satz 1.10 folgt, dass $\mathbb{P} = \tilde{\mathbb{P}}$.

Existenz. Für $C \in \mathcal{F}_1 \otimes \mathcal{F}_2$ definiert man

$$\mathbb{P}(C) \coloneqq \int_{\Omega_1} \mathbb{P}_2(C_{\omega_1}) d\mathbb{P}_1(\omega_1).$$

- Wir zeigen, dass $\mathbb{P}$ wohldefiniert ist.

  Für $C \in \mathcal{F}_1 \otimes \mathcal{F}_2$ sei
  $$f_C\colon \quad \begin{array}{rl} \Omega_1 \to & [0,1] \\ \omega_1 \mapsto & f_C(\omega_1) \coloneqq \mathbb{P}_2(C_{\omega_1}). \end{array}$$
  Diese Funktion ist wohldefiniert, da $C_{\omega_1} \in \mathcal{F}_2 \ \forall \omega_1 \in \Omega_1$ (Lemma 2.9).

  f_C ist auch $\mathcal{F}_1 - \mathcal{B}(\mathbb{R})-$messbar: Sei
  $$\mathcal{E} \coloneqq \{C \in \mathcal{F}_1 \otimes \mathcal{F}_2 \mid f_C \text{ ist messbar}\}.$$
  Das Mengensystem $\mathcal{E}$ ist ein Dynkin-System (Übung). Weiter gilt für $C = A \times B \in \mathcal{C}$
  $$(A \times B)_{\omega_1} = \begin{cases} \varnothing & \text{falls } \omega_1 \in \Omega_1 \setminus A \\ B & \text{falls } \omega_1 \in A \end{cases}$$
  und daher ist die Funktion $f_{A\times B} = \mathbb{P}_2(B)\mathbf{1}_A$ messbar. Daraus folgt, dass $\mathcal{C} \subset \mathcal{E}$ gilt. Das Mengensystem $\mathcal{C}$ ist $\cap-$stabil und erzeugt $\mathcal{F}_1 \otimes \mathcal{F}_2$. Aus Lemma 1.19 folgt, dass
  $$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{C}) = \mathcal{D}(\mathcal{C}) \subset \mathcal{E}.$$
  Die Funktion f_C ist also messbar und nichtnegativ $\forall C \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Daher ist das Integral $\int_{\Omega_1} f_C \, d\mathbb{P}_1 \in [0,\infty]$ wohldefiniert. Schließlich folgt aus $f_C \leq 1$, dass $0 \leq \int_{\Omega_1} f_C d\mathbb{P}_1 \leq 1$ gilt.

- Wir zeigen, dass $\mathbb{P}$ ein Wahrscheinlichkeitsmaß ist.

  Es gilt
  $$\mathbb{P}(\varnothing) = \int_{\Omega_1} \mathbb{P}_2(\varnothing_{\omega_1}) d\mathbb{P}_1(\omega_1) = \int_{\Omega_1} \mathbb{P}_2(\varnothing) d\mathbb{P}_1(\omega_1) = 0$$
  und
  $$\mathbb{P}(\Omega_1 \times \Omega_2) = \int_{\Omega_1} \mathbb{P}_2(\Omega_2) d\mathbb{P}_1(\omega_1) = \int_{\Omega_1} 1 d\mathbb{P}_1(\omega_1) = 1.$$
  $\mathbb{P}$ ist zudem $\sigma-$additiv (Übung).

- Wir zeigen, dass (2.8) gilt.

  Für $C = A \times B$ mit $A \in \mathcal{F}_1, B \in \mathcal{F}_2$ rechnen wir, mit $\mathbb{P}_2((A \times B)_{\omega_1}) = \mathbb{P}_2(B)\mathbf{1}_A(\omega_1)$ und folglich gilt
  $$\mathbb{P}(C) = \mathbb{P}(A \times B) = \int_{\Omega_1} \mathbb{P}_2((A \times B)_{\omega_1}) d\mathbb{P}_1(\omega_1) = \mathbb{P}_2(B) \int_{\Omega_1} \mathbf{1}_A d\mathbb{P}_1 = \mathbb{P}_2(B)\mathbb{P}_1(A).$$

- Wir zeigen, dass zusätzlich $\mathbb{P}(C) = \int_{\Omega_2} \mathbb{P}_1(C_{\omega_2}) d\mathbb{P}_2(\omega_2)$ gilt.

  Sei $\tilde{\mathbb{P}}$ durch $\tilde{\mathbb{P}}(C) \coloneqq \int_{\Omega_2} \mathbb{P}_1(C_{\omega_2}) d\mathbb{P}_2(\omega_2)$. Dann ist $\tilde{\mathbb{P}}$ auch ein Wahrscheinlichkeitsmaß, welches (2.8) erfüllt (die obigen Schritte wiederholen). Aus der Eindeutigkeit folgt, dass $\mathbb{P} = \tilde{\mathbb{P}}$ gilt.

$\square$

Satz 2.11 (Satz von Fubini). *Seien* $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ *zwei Wahrscheinlichkeiträume. Sei* $f\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ *eine* $\mathcal{F}_1 \otimes \mathcal{F}_2 - \mathcal{B}(\mathbb{R})-$*messbare Funktion.*
Für $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$ *betrachten wir die Funktionen*

$$f_{\omega_1}\colon \begin{array}{rl} \Omega_2 \to & \mathbb{R} \\ \omega_2 \mapsto & f_{\omega_1}(\omega_2) := f(\omega_1, \omega_2), \end{array} \qquad f_{\omega_2}\colon \begin{array}{rl} \Omega_1 \to & \mathbb{R} \\ \omega_1 \mapsto & f_{\omega_2}(\omega_1) := f(\omega_1, \omega_2). \end{array}$$

(i) Die Funktion f_{ω_1} *ist* $\mathcal{F}_2 - \mathcal{B}(\mathbb{R})$ *messbar* $\forall \omega_1 \in \Omega_1$.
Analog ist die Funktion f_{ω_2} $\mathcal{F}_1 - \mathcal{B}(\mathbb{R})$ *messbar* $\forall \omega_2 \in \Omega_2$.

(ii) Sei $f \geq 0$. *Dann ist die Funktion*

$$F_{f,1}\colon \begin{array}{rl} \Omega_1 \to & \overline{\mathbb{R}} \\ \omega_1 \mapsto & F_{f,1}(\omega_1) := \int_{\Omega_2} f_{\omega_1} d\mathbb{P}_2 = \int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) = \int_{\Omega_2} f(\omega_1, \cdot) d\mathbb{P}_2 \end{array}$$

wohldefiniert und $\mathcal{F}_1 - \mathcal{B}(\overline{\mathbb{R}})-$*messbar. Die analog definierte Funktion* $F_{f,2}\colon \Omega_2 \to \overline{\mathbb{R}}$ *ist auch wohldefiniert und* $\mathcal{F}_2 - \mathcal{B}(\overline{\mathbb{R}})-$*messbar.*
Weiter gilt

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f \; d(\mathbb{P}_1 \otimes \mathbb{P}_2) &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \, d\mathbb{P}_2(\omega_2) \right) d\mathbb{P}_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \, d\mathbb{P}_1(\omega_1) \right) d\mathbb{P}_2(\omega_2). \end{aligned} \tag{2.11}$$

Insbesondere ist $F_{f,1}$ *integrierbar genau dann, wenn* $F_{f,2}$ *integrierbar ist und in diesem Fall stimmen die zwei Integrale überein.*

(iii) Sei $f \geq 0$ *und* $f \in L^1(\Omega_1 \times \Omega_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$, *d.h.,* $\int_{\Omega_1 \times \Omega_2} f \; d(\mathbb{P}_1 \otimes \mathbb{P}_2) < \infty$.
Dann $\exists N_1 \in \mathcal{F}_1, N_2 \in \mathcal{F}_2$ *Nullmengen, also* $\mathbb{P}_i(N_i) = 0, i = 1, 2$, *sodass* $F_{f,i}(\omega_i) \in \mathbb{R}$ $\forall \omega_i \in \Omega_i \setminus N_i$, $i = 1, 2$.
Wir definieren

$$\tilde{F}_{f,i}\colon \begin{array}{rl} \Omega_i \to & \mathbb{R} \\ \omega_i \mapsto & \begin{cases} F_{f,i}(\omega_i) & \text{falls } \omega_i \in \Omega_i \setminus N_i, \\ 0 & \text{falls } \omega_i \in N_i. \end{cases} \end{array}$$

Dann gilt $F_{f,i} = \tilde{F}_{f,i}$ $\mathbb{P}_i-$*fast überall und* $0 \leq \int_{\Omega_i} F_{f,i} \, d\mathbb{P}_i \; = \; \int_{\Omega_i} \tilde{F}_{f,i} \, d\mathbb{P}_i < \infty$.

(iv) Sei $f \in L^1(\Omega_1 \times \Omega_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$, *d.h.,* $\int_{\Omega_1 \times \Omega_2} |f| \; d(\mathbb{P}_1 \otimes \mathbb{P}_2) < \infty$.
Dann $\exists N_1 \in \mathcal{F}_1, N_2 \in \mathcal{F}_2$ *Nullmengen, also* $\mathbb{P}_i(N_i) = 0, i = 1, 2$, *sodass*

$$\int_{\Omega_2} |f(\omega_1, \cdot)| d\mathbb{P}_2 < \infty \quad \forall \omega_1 \in \Omega_1 \setminus N_1, \qquad \int_{\Omega_2} |f(\cdot, \omega_2)| d\mathbb{P}_1 < \infty \quad \forall \omega_1 \in \Omega_2 \setminus N_2$$

gilt und die Funktionen

$$F_{f,1}\colon \begin{array}{rl} \Omega_1 \to & \mathbb{R} \\ \omega_1 \mapsto & F_{f,1}(\omega_1) := \begin{cases} \int_{\Omega_2} f(\omega_1, \cdot) d\mathbb{P}_2 & \text{falls } \omega_1 \in \Omega_1 \setminus N_1, \\ 0 & \text{falls } \omega_1 \in N_1, \end{cases} \end{array}$$

$$F_{f,2}\colon \begin{array}{rl} \Omega_2 \to & \mathbb{R} \\ \omega_2 \mapsto & F_{f,2}(\omega_2) := \begin{cases} \int_{\Omega_1} f(\cdot, \omega_2) d\mathbb{P}_1 & \text{falls } \omega_2 \in \Omega_2 \setminus N_2, \\ 0 & \text{falls } \omega_2 \in N_2, \end{cases} \end{array}$$

sind wohldefiniert und integrierbar. In diesem Fall gilt auch (2.11).

Bemerkung. Dasselbe Resultat (mit fast demselben Beweis) gilt, wenn $\mathbb{P}_1, \mathbb{P}_2$ durch zwei σ-endlichen Maße μ_1, μ_2 ersetzt werden. Z.B. gilt das Resultat für $\mathcal{L} \otimes \mathcal{L}$.

Beweis.
(i) gilt nach Lemma 2.9.

(ii) Die Funktion f_{ω_1} ist nichtnegativ und $\mathcal{F}_2 - \mathcal{B}(\mathbb{R})-$messbar, also ist die Funktion $F_{f,1}$ wohldefiniert. Wir zeigen, dass $F_{f,1}$ $\mathcal{F}_1 - \mathcal{B}(\mathbb{R})-$messbar ist:
Sei $f = \mathbf{1}_C$ für $C \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Es gilt $f(\omega_1, \omega_2) = \mathbf{1}_C(\omega_1, \omega_2) = \mathbf{1}_{C_{\omega_1}}(\omega_2)$ und daher

$$F_{f,1}(\omega_1) = \int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) = \int_{\Omega_2} \mathbf{1}_{C_{\omega_1}}(\omega_2) d\mathbb{P}_2(\omega_2) = \mathbb{P}_2(C_{\omega_1}).$$

Die Funktion $\omega_1 \mapsto \mathbb{P}_2(C_{\omega_1})$ ist messbar (siehe Beweis von Satz 2.10). Die anderen Fälle (f einfach, nichtnegativ, allgemein) werden durch maßtheoretische Induktion bewiesen (Übung).
Wir zeigen jetzt, dass (2.11) gilt.
Sei $f = \mathbf{1}_C$ für $C \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Wir rechnen, mit Satz 2.10),

$$\int_{\Omega_1 \times \Omega_2} f \, d(\mathbb{P}_1 \otimes \mathbb{P}_2) = \mathbb{P}_1 \otimes \mathbb{P}_2(C) = \int_{\Omega_1} \mathbb{P}_2(C_{\omega_1}) d\mathbb{P}_1(\omega_1) = \int_{\Omega_2} \mathbb{P}_1(C_{\omega_2}) d\mathbb{P}_2(\omega_2),$$

Die Aussage folgt aus $\mathbb{P}_2(C_{\omega_1}) = F_{f,1}(\omega_1)$, $\mathbb{P}_1(C_{\omega_2}) = F_{f,2}(\omega_2)$. Die andere Fälle ($f$ einfach, nichtnegativ, allgemein) werden durch maßtheoretische Induktion bewiesen (Übung).

(iii) Sei $f \geq 0$ und $f \in L^1(\Omega_1 \times \Omega_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$.
Aus $\int_{\Omega_1 \times \Omega_2} f \, d(\mathbb{P}_1 \otimes \mathbb{P}_2) < \infty$ und (2.11) folgt $0 \leq \int_{\Omega_1} F_{f,1} d\mathbb{P}_1 < \infty$.
Sei $A \in \mathcal{F}_1$, sodass $\mathbb{P}_1(A) > 0$ und $F_{1,f}(\omega_1) = \infty \ \forall \omega_1 \in A$ gilt. Wir rechnen

$$\infty > \int_{\Omega_1} F_{1,f}(\omega_1) d\mathbb{P}_1(\omega_1) \geq \int_A F_{1,f}(\omega_1) d\mathbb{P}_1(\omega_1) = \infty,$$

was einen Widerspruch ergibt.

(iv) Falls $f \in L^1(\Omega_1 \times \Omega_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$, dann gilt $f = f_+ - f_-$ mit $f_\pm \in L^1(\Omega_1 \times \Omega_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ und $f_\pm \geq 0$. Die Aussage folgt aus (ii) und (iii) (Übung).

□

Unser nächste Ziel ist es, eine äquivalente Definition von Unabhängigkeit durch Produktmaße zu definieren.
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X_1, X_2 \colon \Omega \to \mathbb{R}$ zwei Zufallsvariablen. Wir definieren

$$\begin{aligned} X\colon \quad \Omega &\to \quad \mathbb{R}^2 \\ \omega &\mapsto \quad X(\omega) := (X_1(\omega), X_2(\omega)). \end{aligned} \tag{2.12}$$

Die Funktion X ist $\mathcal{F} - \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})-$messbar: $\forall C = A \times B$ mit $A, B \in \mathcal{B}(\mathbb{R})$ gilt

$$X^{-1}(A \times B) = X_1^{-1}(A) \cap X_2^{-1}(B) \in \mathcal{F},$$

da die Funktionen X_1, X_2 messbar sind.
Die Messbarkeit folgt jetzt aus $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \sigma(C = A \times B \mid A, B \in \mathcal{B}(\mathbb{R}))$ und Korollar 1.22.
Die Funktion X ist also eine $\mathbb{R}^2-$wertige Zufallsvariable. Ihr Bildmaß ist $\mathbb{P}_X = \mathbb{P} \circ X^{-1} \colon \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \to [0, 1]$.

Definition 2.12 (gemeinsame Verteilung).
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $X_1, X_2 \colon \Omega \to \mathbb{R}$ *zwei Zufallsvariablen und* $X \colon \Omega \to \mathbb{R}^2$ *die in* (2.12) *definierte Zufallsvariable.*

Die Verteilung $\mathbb{P}_{X_1,X_2} \coloneqq \mathbb{P}_X = \mathbb{P} \circ X^{-1}$ *heißt* gemeinsame Verteilung *der Zufallsvariablen* X_1 *und* X_2.

Satz 2.13.
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $X_1, X_2 \colon \Omega \to \mathbb{R}$ *zwei Zufallsvariablen. Folgende Aussagen sind äquivalent.*

(i) X_1 *und* X_2 *sind unabhängig.*

(ii) $\mathbb{P}_{X_1,X_2}$ *ist ein Produktmaß, d.h.* $\exists\ \mu_1, \mu_2 \colon \mathcal{B}(\mathbb{R}) \to [0,1]$ *Wahrscheinlichkeitsmaße, sodass* $\mathbb{P}_{X_1,X_2} = \mu_1 \otimes \mu_2$.

(iii) $\mathbb{P}_{X_1,X_2} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$.

Beweis.
$(i) \Rightarrow (iii)$ Für $C = A \times B$ mit $A, B \in \mathcal{B}(\mathbb{R})$ rechnen wir

$$\mathbb{P}_{X_1,X_2}(A \times B) = \mathbb{P}(\{X_1 \in A\} \cap \{X_2 \in B\}) \overset{(i)}{=} \mathbb{P}(X_1 \in A)\mathbb{P}(X_2 \in B) = \mathbb{P}_{X_1}(A)\mathbb{P}_{X_2}(B).$$

Die Aussage folgt aus Satz 2.10. $(iii) \Rightarrow (i)$ Für $A, B \in \mathcal{B}(\mathbb{R})$ rechnen wir

$$\mathbb{P}(\{X_1 \in A\} \cap \{X_2 \in B\}) = \mathbb{P}_{X_1,X_2}(A \times B) \overset{(iii)}{=} \mathbb{P}_{X_1}(A)\mathbb{P}_{X_2}(B) = \mathbb{P}(X_1 \in A)\mathbb{P}(X_2 \in B).$$

$(iii) \Rightarrow (ii)$ automatisch.

$(ii) \Rightarrow (iii)$ Für $A, B \in \mathcal{B}(\mathbb{R})$ rechnen wir

$$\mathbb{P}_{X_1,X_2}(A \times B) \overset{(ii)}{=} \mu_1(A)\mu_2(B).$$

Die Aussage folgt aus

$$\begin{aligned}
\mu_1(A) &= \mu_1(A)\mu_2(\mathbb{R}) = \mathbb{P}_{X_1,X_2}(A \times \mathbb{R}) = \mathbb{P}(\{X_1 \in A\} \cap \{X_2 \in \mathbb{R}\}) = \mathbb{P}(X_1 \in A) = \mathbb{P}_{X_1}(A),\\
\mu_2(B) &= \mu_1(\mathbb{R})\mu_2(B) = \mathbb{P}_{X_1,X_2}(\mathbb{R} \times B) = \mathbb{P}(\{X_1 \in \mathbb{R}\} \cap \{X_2 \in B\}) = \mathbb{P}(X_2 \in B) = \mathbb{P}_{X_2}(B).
\end{aligned}$$

□

[11: 18.11.2025]

[12: 21.11.2025]

2.5 Summe von Zufallsvariablen: Faltung

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X_1, X_2 \colon \Omega \to \mathbb{R}$ zwei Zufallsvariablen mit Verteilungen $\mathbb{P}_{X_1}, \mathbb{P}_{X_2}$. Wir suchen die Verteilung von $X_1 + X_2$. Für jede messbare Funktion $f \colon \mathbb{R}^2 \to \mathbb{R}$, sodass $f \circ (X_1, X_2) \in L^1(\Omega, \mathbb{P})$, gilt

$$\mathbb{E}[f(X_1, X_2)] = \int_{\mathbb{R}^2} f(x, y)\, d\mathbb{P}_{X_1,X_2}(x, y),$$

wobei $\mathbb{P}_{X_1,X_2}$ die gemeinsame Verteilung von X_1, X_2 ist. Insbesondere gilt

$$\mathbb{P}_{X_1+X_2}(A) = \int_{\mathbb{R}^2} \mathbf{1}_A(x + y)\, d\mathbb{P}_{X_1,X_2}(x, y).$$

Lemma 2.14. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X_1, X_2 \colon \Omega \to \mathbb{R}$ zwei unabhängige Zufallsvariablen.*

(i) $\forall A \in \mathcal{B}(\mathbb{R})$ *gilt*

$$\mathbb{P}_{X_1+X_2}(A) = \int_{\mathbb{R}} \mathbb{P}_{X_1}(A-y)\, d\mathbb{P}_{X_2}(y) = \int_{\mathbb{R}} \mathbb{P}_{X_2}(A-x)\, d\mathbb{P}_{X_1}(x),$$

wobei $A - y = \{z - y \,|\, z \in A\} \subset \mathbb{R}$.

(ii) $\forall t \in \mathbb{R}$ *gilt*

$$F_{X_1+X_2}(t) = \int_{\mathbb{R}} F_{X_1}(t-y) d\mathbb{P}_{X_2}(y) = \int_{\mathbb{R}} F_{X_2}(t-x) d\mathbb{P}_{X_1}(x).$$

(iii) Falls $\mathbb{P}_{X_1} = \rho_{X_1}\mathcal{L}$ *und* $\mathbb{P}_{X_2} = \rho_{X_2}\mathcal{L}$, *dann gilt* $\mathbb{P}_{X_1+X_2} = \rho_{X_1+X_2}\mathcal{L}$ *mit Dichte*

$$\rho_{X_1+X_2}(s) = \int_{\mathbb{R}} \rho_{X_1}(s-y)\rho_{X_2}(y) d\mathcal{L}(y) = \int_{\mathbb{R}} \rho_{X_2}(s-x)\rho_{X_1}(x) d\mathcal{L}(x).$$

Beweis. Da X_1, X_2 unabhängig sind, gilt $\mathbb{P}_{X_1,X_2} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$. Für jede messbare Funktion $f \colon \mathbb{R}^2 \to \mathbb{R}$, sodass $f \circ (X_1, X_2) \in L^1(\Omega, \mathbb{P})$, gilt also

$$\int_{\mathbb{R}^2} f(x,y)\, d\mathbb{P}_{X_1,X_2}(x,y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\, d\mathbb{P}_{X_1}(x) \right) d\mathbb{P}_{X_2}(y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\, d\mathbb{P}_{X_2}(y) \right) d\mathbb{P}_{X_1}(x).$$

(i) Wir rechnen, mit $\mathbf{1}_A(x+y) = \mathbf{1}_{A-y}(x)$,

$$\begin{aligned}
\mathbb{P}_{X_1+X_2}(A) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(x+y)\, d\mathbb{P}_{X_1}(x) \right) d\mathbb{P}_{X_2}(y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{A-y}(x)\, d\mathbb{P}_{X_1}(x) \right) d\mathbb{P}_{X_2}(y) = \int_{\mathbb{R}} \mathbb{P}_{X_1}(A-y)\, d\mathbb{P}_{X_2}(y).
\end{aligned}$$

Die andere Gleichung wird analog bewiesen.

(ii) Für $t \in \mathbb{R}$ rechnen wir, mit $\mathbf{1}_{(-\infty,t]}(x+y) = \mathbf{1}_{(-\infty,t-y]}(x)$,

$$F_{X_1+X_2}(t) = \mathbb{P}_{X_1+X_2}((-\infty,t]) = \int_{\mathbb{R}} \mathbb{P}_{X_1}((-\infty,t-y])\, d\mathbb{P}_{X_2}(y) = \int_{\mathbb{R}} F_{X_1}(t-y)\, d\mathbb{P}_{X_2}(y).$$

Die andere Gleichung wird analog bewiesen.

(iii) Wir rechnen

$$\begin{aligned}
\mathbb{P}_{X_1+X_2}(A) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(x+y)\, d\mathbb{P}_{X_1}(x) \right) d\mathbb{P}_{X_2}(y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(x+y)\, \rho_{X_1}(x) d\mathcal{L}(x) \right) \rho_{X_2}(y) d\mathcal{L}(y).
\end{aligned}$$

Wir ersetzen x mit $x = \tilde{x} - y$. Aus dem Transformationssatz 1.39 und Fubini folgt

$$\begin{aligned}
\mathbb{P}_{X_1+X_2}(A) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(\tilde{x})\, \rho_{X_1}(\tilde{x}-y) d\mathcal{L}(\tilde{x}) \right) \rho_{X_2}(y) d\mathcal{L}(y) \\
&= \int_{\mathbb{R}} \mathbf{1}_A(\tilde{x}) \left(\int_{\mathbb{R}} \rho_{X_1}(\tilde{x}-y) \rho_{X_2}(y) d\mathcal{L}(y) \right) d\mathcal{L}(\tilde{x}),
\end{aligned}$$

und daher gilt $\mathbb{P}_{X_1+X_2} = \rho_{X_1+X_2}\mathcal{L}$ mit Dichte $\rho_{X_1,X_2}(s) = \int_{\mathbb{R}} \rho_{X_1}(s-y)\rho_{X_2}(y) d\mathcal{L}(y)$. Die andere Gleichung wir analog bewiesen. $\square$

Definition 2.15 (Faltung)**.**

(i) Seien $\mathbb{P}_1, \mathbb{P}_2$ zwei Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfuktionen $F_{\mathbb{P}_1}, F_{\mathbb{P}_2}$. Die Faltung von $\mathbb{P}_1$ mit $\mathbb{P}_2$ ist das Wahrscheinlichkeitsmaß $\mathbb{P}_1 * \mathbb{P}_2$ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfuktion

$$F_{\mathbb{P}_1 * \mathbb{P}_2}(t) := \int_{\mathbb{R}} F_{\mathbb{P}_1}(t-y)\, d\mathbb{P}_{X_2}(y).$$

(ii) Seien ρ_1, ρ_2 zwei Wahrscheinlichkeitsdichten. Die Faltung (convolution) der Funktionen ρ_1 und ρ_2 ist die Funktion

$$\rho_1 * \rho_2(t) := \int_{\mathbb{R}} \rho_{X_1}(s-y)\rho_{X_2}(y) d\mathcal{L}(y).$$

Bemerkung. Die Faltung ist kommutativ: $\mathbb{P}_1 * \mathbb{P}_2 = \mathbb{P}_2 * \mathbb{P}_1$ und $\rho_1 * \rho_2(t) = \rho_2 * \rho_1(t)$.

Man kann sich die Frage stellen, ob es Typen von Verteilungen gibt, die unter der Faltungsoperation invariant bleiben. Solche Verteilungen nennt man *stabil.* Wir werden diese Frage hier nicht im Allgemeinen untersuchen, sondern nur ein wichtiges Beispiel betrachten.

Satz 2.16 (Stabilität der Gaußverteilung)**.** *Seien $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ und $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ zwei unabhängige Gaußverteilte Zufallsvariablen. Dann gilt $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Beweis. Es gilt $X \sim \mathcal{N}(m, \sigma^2) \Leftrightarrow X - m \sim \mathcal{N}(0, \sigma^2)$ (Übung). Wir können also nur den Fall $m_1 = m_2 = 0$ betrachten. Aus Lemma 2.14 folgt, dass $\mathbb{P}_{X_1+X_2} = \rho_{X_1+X_2}\mathcal{L}$ mit Dichte

$$\rho_{X_1+X_2}(t) = \int_{\mathbb{R}} \rho_{X_1}(s-y)\rho_{X_2}(y) d\mathcal{L}(y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \int_{\mathbb{R}} e^{-\frac{(t-y)^2}{2\sigma_1^2}} e^{-\frac{y^2}{2\sigma_2^2}} d\mathcal{L}(y).$$

Sei $\Sigma := \sigma_1^2 + \sigma_2^2$. Wir rechnen,

$$-\frac{(t-y)^2}{\sigma_1^2} - \frac{y^2}{\sigma_2^2} = -\frac{t^2}{\Sigma} - \frac{\Sigma}{\sigma_1^2\sigma_2^2}\left(y - \frac{t\sigma_2^2}{\Sigma}\right)^2.$$

Daher gilt, mit der Formel

$$\int_{\mathbb{R}} e^{-\frac{a}{2}x^2} d\mathcal{L}(x) = \frac{\sqrt{2\pi}}{\sqrt{a}} \qquad \forall a > 0,$$

und dem Transformationssatz 1.39,

$$\begin{aligned}\int_{\mathbb{R}} e^{-\frac{(t-y)^2}{2\sigma_1^2}} e^{-\frac{y^2}{2\sigma_2^2}} d\mathcal{L}(y) &= e^{-\frac{t^2}{2\Sigma}} \int_{\mathbb{R}} e^{-\frac{1}{2}\frac{\Sigma}{\sigma_1^2\sigma_2^2}\left(y - \frac{t\sigma_2^2}{\Sigma}\right)^2} d\mathcal{L}(y) = e^{-\frac{t^2}{2\Sigma}} \int_{\mathbb{R}} e^{-\frac{1}{2}\frac{\Sigma}{\sigma_1^2\sigma_2^2}y^2} d\mathcal{L}(y) \\ &= \frac{\sqrt{2\pi\sigma_1^2\sigma_2^2}}{\sqrt{\Sigma}} e^{-\frac{t^2}{2\Sigma}}.\end{aligned}$$

Daher gilt

$$\rho_{X_1+X_2}(t) = \frac{1}{\sqrt{2\pi\Sigma}} e^{-\frac{t^2}{2\Sigma}}.$$

□

Andere Beispiele. Folgende Verteilungen sind stabil:

$$\begin{aligned} X_1 \sim Poi(\lambda_1), X_2 \sim Poi(\lambda_2) \quad &\text{unabhängig} \;\Rightarrow \quad & X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2), \\ X_1 \sim Bin(n_1, p), X_2 \sim Bin(n_2, p) \quad &\text{unabhängig} \;\Rightarrow \quad & X_1 + X_2 \sim Bin(n_1 + n_2, p), \\ X_1 \sim Cauchy(a_1), X_2 \sim Cauchy(a_2) \quad &\text{unabhängig} \;\Rightarrow \quad & X_1 + X_2 \sim Cauchy(a_1 + a_2). \end{aligned}$$

Die Exponentialverteilung ist nicht stabil: Seien $X_1 \sim Exp(\lambda_1), X_2 \sim Exp(\lambda_2)$ unabhängig. Die Zufallsvariable $X_1 + X_2$ ist nicht exponentialverteilt.

2.6 Unendliche Produkte

Erinnerung: Sei $(\Omega_i, \mathcal{F}_i)_{i\in\mathbb{N}}$ eine Folge von Messräumen, und sei

$$\Omega =: \times_{i\in\mathbb{N}}\Omega_i = \prod_{i\in\mathbb{N}} \Omega_i = \big\{\omega = (\omega_i)_{i\in\mathbb{N}} \mid \omega_i \in \Omega_i \ \forall i \in \mathbb{N}\big\}.$$

Die Produkt-$\sigma-$Algebra auf Ω ist $\otimes_{n\in\mathbb{N}}\mathcal{F}_n \coloneqq \sigma(\mathcal{C})$, wobei

$$\begin{aligned} \mathcal{C} &\coloneqq \left\{ C = \prod_{i\in I} A_i \prod_{i\notin I} \Omega_i \mid \text{für ein } I \subset \mathbb{N} \text{ endlich, } A_i \in \mathcal{F}_i \ \forall i \in I \right\} \\ &= \bigcup_{I\subset\mathbb{N} \text{ endlich}} \ \bigcup_{A_i\in\mathcal{F}_i \, \forall i\in I} \{\omega = (\omega_i)_{i\in\mathbb{N}} \mid \omega_i \in A_i \, \forall i \in I\} \end{aligned}$$

die Menge der Zylindermengen ist.

Definition 2.17. *Sei $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)_{i\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsräumen. Das Produktmaß $\otimes_{i\in\mathbb{N}}\mathbb{P}_i$ ist das (eindeutige) Wahrscheinlichkeitsmaß auf $\otimes_{n\in\mathbb{N}}\mathcal{F}_n$, sodass*

$$\otimes_{i\in\mathbb{N}}\mathbb{P}_i \left(\prod_{i\in I} A_i \prod_{i\notin I} \Omega_i \right) = \prod_{i\in I} \mathbb{P}_i(A_i) \tag{2.13}$$

für alle Zylindermengen.

Definition 2.18. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.*

(i) Eine Funktion $X : \Omega \to \mathbb{R}^{\mathbb{N}}$, die $\mathcal{F} - \otimes_{n\in\mathbb{N}}\mathcal{B}(\mathbb{R})-$messbar ist, heißt Zufallsfolge *oder* stochastischer Prozess *(mit diskreter Zeit).*

(ii) Eine Zufallsfolge X heißt eine Folge unabhängiger Zufallsvariablen*, wenn die Verteilung $\mathbb{P}_X$ von X ein Produktmaß ist, also von der Form* (2.13) *ist.*

Bemerkung $X = (X_n)_{n\in\mathbb{N}}$ ist eine Folge unabhängiger Zufallsvariablen $\Leftrightarrow$ $\forall n \in \mathbb{N}$ und $h_1, \ldots, h_n : \mathbb{R} \to \mathbb{R}$ messbare Funktionen gilt

$$\mathbb{E}\left[\prod_{j=1}^{n} h_i(X_i)\right] = \prod_{j=1}^{n} \mathbb{E}\left[h_i(X_i)\right],$$

solange die Integrale wohldefiniert sind.

Definition 2.19 (Irrfahrt). *Sei $X = (X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit*

$$\mathbb{P}_{X_n} = p\delta_1 + (1-p)\delta_{-1} \qquad \forall n \in \mathbb{N}.$$

Dann heißt die Folge $n \mapsto S_n := \sum_{j=1}^n X_j$ die einfache Irrfahrt *auf $\mathbb{Z}$ (simple random walk). Falls $p = \frac{1}{2}$ heißt S_n* symmetrische Irrfahrt.

[12: 21.11.2025]

[13: 25.11.2025]

Die Irrfahrt erfüllt, unter anderem,

$$\begin{aligned}\mathbb{P}(S_{n+1} = k|\ S_n = q) &= \mathbb{P}(S_n + X_{n+1} = k|\ S_n = q) = \mathbb{P}(X_{n+1} = k - q|\ S_n = q)\\ &= \mathbb{P}(X_{n+1} = k - q) = \mathbf{1}_{k-q=1}p + \mathbf{1}_{k-q=-1}(1-p) \qquad \forall n \in \mathbb{N}.\end{aligned}$$

Hier gilt $\mathbb{P}(X_{n+1} = k-q|\ S_n = q) = \mathbb{P}(X_{n+1} = k-q)$, weil $\{X_1, \ldots, X_n\}$ und X_{n+1} unabhängig sind. Man rechnet auch

$$\begin{aligned}\mathbb{E}[S_n] &= \mathbb{E}\left[\sum_{j=1}^n X_j\right] = \sum_{j=1}^n \mathbb{E}[X_j] = n\mathbb{E}[X_1]\\ \mathrm{var}(S_n) &= \mathbb{E}[S_n^2] - \mathbb{E}[S_n]^2 = \sum_{j,k=1}^n \mathbb{E}[X_jX_k] - \sum_{j,k=1}^n \mathbb{E}[X_j]\mathbb{E}[X_k]\\ &= \sum_{j=1}^n \left(\mathbb{E}[X_j^2] - \mathbb{E}[X_j]^2\right) + \sum_{j\neq k} (\mathbb{E}[X_jX_k] - \mathbb{E}[X_j]\mathbb{E}[X_k])\\ &= n\ \mathrm{var}(X_1),\end{aligned}$$

wobei $\mathbb{E}[X_jX_k] = \mathbb{E}[X_j]\mathbb{E}[X_k]$ gilt, da X_j und X_k $\forall j \neq k$ unabhängig sind. Es gilt also

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \frac{\mathbb{E}[S_n]}{n} = \mathbb{E}[X_1] \quad \text{und} \quad \mathrm{var}\left(\frac{S_n}{n}\right) = \frac{\mathrm{var}(S_n)}{n^2} = \frac{\mathrm{var}(X_1)}{n} \overset{n\to\infty}{\to} 0.$$

Wir werden später sehen, dass $\frac{S_n}{n} \overset{n\to\infty}{\to} \mathbb{E}[X_1]$ fast sicher gilt (Gesetz der großen Zahlen).
Sei jetzt $p = \frac{1}{2}$,, sodass $\mathbb{E}[X_1] = 0$ und $\mathrm{var}(X_1) = 1$ gilt. Wir rechnen

$$\mathbb{E}\left[\frac{S_n}{\sqrt{n}}\right] = \frac{\mathbb{E}[S_n]}{\sqrt{n}} = 0 \quad \text{und} \quad \mathrm{var}\left(\frac{S_n}{\sqrt{n}}\right) = \frac{\mathrm{var}(S_n)}{n} = 1.$$

Wir werden später sehen, dass $\frac{S_n}{\sqrt{n}} \overset{n\to\infty}{\to} \mathcal{N}(0,1)$ in Verteilung gilt (zentraler Grenzwertsatz).
Diese Konvergenzbegriffe werden im nächsten Kapitel eingeführt.

3 Konvergenzbegriffe

3.1 Konvergenz von Verteilungsfunktionen

Erinnerung:

- Eine Funktion $F\colon \mathbb{R} \to [0,1]$ ist eine Verteilungsfunktion, falls F monoton wachsend und rechtsstetig ist und zusätzlich $\lim_{t\to-\infty} F(t) = 0$ sowie $\lim_{t\to\infty} F(t) = 1$ erfüllt.

- F ist eine Verteilungsfunktion $\Leftrightarrow \exists! \mathbb{P}\colon \mathcal{B}(\mathbb{R}) \to [0,1]$ Wahrscheinlichkeitsmaß, sodass $F(t) = \mathbb{P}((-\infty, t])$.

Definition 3.1.

(i) Sei $n \mapsto F_n$ eine Folge von Verteilungsfunktionen, und F eine Verteilungsfunktion.

Die Folge F_n konvergiert schwach *gegen F, falls Folgendes gilt:*

$$\lim_{n\to\infty} F_n(t) = F(t) \qquad \forall t \in \mathbb{R}, \text{ wo die Funktion } F \text{ stetig ist.}$$

(ii) Sei Ω ein topologischer Raum, $n \mapsto \mathbb{P}_n\colon \mathcal{B}(\Omega) \to [0,1]$ eine Folge von Wahrscheinlichkeitsmaßen und $\mathbb{P}\colon \mathcal{B}(\Omega) \to [0,1]$ ein Wahrscheinlichkeitsmaß.

Die Folge $\mathbb{P}_n$ konvergiert schwach *gegen $\mathbb{P}$, falls für alle beschränkten, stetigen Funktionen $g \in C_b(\mathbb{R})$ Folgendes gilt:*

$$\lim_{n\to\infty} \int_\Omega g(\omega) d\mathbb{P}_n = \int_\Omega g(\omega) d\mathbb{P}.$$

Bemerkung. Die Anforderung, dass F in t stetig ist in (i) wirkt zunächst unnatürlich, ist aber wichtig. Zum Beispiel konvergiert die Folge von Verteilungsfunktionen

$$F_n(t) := \frac{1 + \tanh(nt)}{2} = \frac{1}{1 + e^{-2nt}}$$

gegen eine nicht-rechtstetige Funktion

$$\lim_{n\to\infty} F_n(t) = \begin{cases} 1 & \text{falls } t > 0 \\ \frac{1}{2} & \text{falls } t = 0 \\ 0 & \text{falls } t < 0. \end{cases}$$

Wir wollen jedoch dennoch die rechtsstetige Variante $\mathbf{1}_{[0,\infty)}$ als schwachen Limes akzeptieren wollen, d.h. F_n konvergiert schwach gegen $F = \mathbf{1}_{[0,\infty)}$.

Satz 3.2. *Sei $n \mapsto \mathbb{P}_n\colon \mathcal{B}(\mathbb{R}) \to [0,1]$ eine Folge von Wahrscheinlichkeitsmaßen und $n \mapsto F_n$ die Folge der zugehörigen Verteilungsfunktionen. Sei $\mathbb{P}\colon \mathcal{B}(\mathbb{R}) \to [0,1]$ ein Wahrscheinlichkeitsmaß und F seine Verteilungsfunktion.*

Folgende Aussagen sind äquivalent.

(i) $\mathbb{P}_n$ konvergiert schwach gegen $\mathbb{P}$.

(ii) F_n konvergiert schwach gegen F.

Beweis.
$(i) \Rightarrow (ii)$ Es gilt

$$F_n(t) = \mathbb{P}_n((-\infty, t]) = \int_\mathbb{R} \mathbf{1}_{(-\infty,t]} d\mathbb{P}_n.$$

Da $\mathbf{1}_{(-\infty,t]}$ beschränkt, aber nicht stetig ist, können wir nicht (i) direkt anwenden.

Sei $\varepsilon > 0$. Es gibt eine Funktion $g_\varepsilon \in C_b(\mathbb{R})$, sodass $\mathbf{1}_{(-\infty,t]} \leq g_\varepsilon \leq \mathbf{1}_{(-\infty,t+\varepsilon]}$ gilt. Zum Beispiel funktioniert die Funktion

$$g_\varepsilon(x) := \begin{cases} 1 & \text{falls } x < t \\ 1 - \frac{(x-t)}{\varepsilon} & \text{falls } t \leq x < t+\varepsilon \\ 0 & \text{falls } t+\varepsilon \leq x. \end{cases}$$

Wir rechnen

$$F_n(t) = \int_{\mathbb{R}} \mathbf{1}_{(-\infty,t]} d\mathbb{P}_n \leq \int_{\mathbb{R}} g_\varepsilon d\mathbb{P}_n \overset{(i)}{\to}_{n\to\infty} \int_{\mathbb{R}} g_\varepsilon d\mathbb{P} \leq \int_{\mathbb{R}} \mathbf{1}_{(-\infty,t+\varepsilon]} d\mathbb{P} = F(t+\varepsilon).$$

Daher gilt

$$\limsup_{n\to\infty} F_n(t) \leq F(t+\varepsilon).$$

Analog gibt es eine Funktion $\tilde{g}_\varepsilon \in C_b(\mathbb{R})$, sodass $\mathbf{1}_{(-\infty,t-\varepsilon]} \leq \tilde{g}_\varepsilon \leq \mathbf{1}_{(-\infty,t]}$ gilt, und daher

$$F(t-\varepsilon) \leq \liminf_{n\to\infty} F_n(t) \leq \limsup_{n\to\infty} F_n(t) \leq F(t+\varepsilon) \qquad \forall t \in \mathbb{R}.$$

Sei jetzt $t \in \mathbb{R}$ einer Stetigkeitspunkt für F. Die Aussage folgt dann aus $\lim_{\varepsilon\to 0} F(t \pm \varepsilon) = F(t)$.

(ii) $\Rightarrow$ *(i)* Wir nehmen an, dass F_n schwach gegen F konvergiert. Sei $g \in C_b(\mathbb{R})$ mit $\sup_{x\in\mathbb{R}} |g(x)| = M < \infty$ fest. Wir müssen zeigen, dass

$$\lim_{n\to\infty} \left| \int_{\mathbb{R}} g \, d\mathbb{P}_n - \int_{\mathbb{R}} g \, d\mathbb{P} \right| = 0$$

gilt.

- Wir ersetzen $\int_{\mathbb{R}}$ durch $\int_{(a,b]}$ wie folgt.

  Aus $\lim_{t\to-\infty} F(t) = 0, \lim_{t\to\infty} F(t) = 1$ folgt, dass $\forall \varepsilon > 0$ zwei Punkte $a = a_\varepsilon, b = b_\varepsilon \in \mathbb{R}$ existieren, sodass $a < b$, $F(a) < \varepsilon$ und $F(b) > 1 - \varepsilon$ gilt. Da F maximal abzählbare Unstetigkeitspunkte besitzt, können wir auch a, b wählen, sodass F stetig in a, b ist.

  Sei $\sup_{x\in\mathbb{R}} |g(x)| \leq M$. Wir rechnen

$$\left| \int_{(b,\infty)} g(\omega) d\mathbb{P} \right| \leq \int_{(b,\infty)} |g| \, d\mathbb{P} \leq M\mathbb{P}((b,\infty)) = M(1 - F(b)) \leq M\varepsilon$$
$$\left| \int_{(-\infty,a]} g(\omega) d\mathbb{P} \right| \leq \int_{(-\infty,a]} |g| \, d\mathbb{P} \leq M\mathbb{P}((-\infty,a]) = MF(a) \leq M\varepsilon.$$

  Aus der Stetigkeit von F in a, b folgt, dass $\lim_{n\to\infty} F_n(a) = F(a)$ und $\lim_{n\to\infty} F_n(b) = F(b)$ gilt. Es gibt also $n = n_\varepsilon \in \mathbb{N}$, sodass $F_n(a) < 2\varepsilon$ und $F_n(b) > 1 - 2\varepsilon$ $\forall n \geq n_\varepsilon$ gilt. Daraus folgt, dass

$$\left| \int_{(b,\infty)} g(\omega) d\mathbb{P}_n \right| \leq M2\varepsilon \text{ und}$$
$$\left| \int_{(-\infty,a]} g(\omega) d\mathbb{P}_n \right| \leq M2\varepsilon$$

gilt. Insbesondere ist

$$\left|\int_{\mathbb{R}} g\,d\mathbb{P}_n - \int_{\mathbb{R}} g\,d\mathbb{P}\right| \le \left|\int_{(a,b]} g\,d\mathbb{P}_n - \int_{(a,b]} g\,d\mathbb{P}\right| + 6M\varepsilon.$$

- Wir zeigen jetzt, dass $\left|\int_{(a,b]} g\,d\mathbb{P}_n - \int_{(a,b]} g\,d\mathbb{P}\right|$ klein ist.

  Die Funktion g ist stetig und daher auf dem kompaktem Intervall $[a,b]$ gleichmäßig stetig. Für jedes $\delta > 0$ können wir daher ein $N = N(\delta)$ und Stetigkeitsstellen von F, $a = a_1 < a_2 < \cdots < a_N = b$, finden, sodass $\sup_{x\in(a_{k-1},a_k]} |g(x) - g(a_k)| \le \delta$. Definiere

  $$h_\delta \coloneqq \sum_{j=2}^{N} g(a_k)\mathbf{1}_{(a_{k-1},a_k]}.$$

  Es gilt also $|g(x) - h_\delta(x)| \le \delta\ \forall x \in [a,b]$. Wir rechnen

  $$\left|\int_{(a,b]} g\,d\mathbb{P}_n - \int_{(a,b]} h_\delta\,d\mathbb{P}_n\right| = \left|\int_{(a,b]} (g - h_\delta)\,d\mathbb{P}_n\right| \le \int_{(a,b]} |g - h_\delta|\,d\mathbb{P}_n \le \delta\,\mathbb{P}_n((a,b]) \le \delta.$$

  Analog gilt $\left|\int_{(a,b]} g\,d\mathbb{P} - \int_{(a,b]} h_\delta\,d\mathbb{P}\right| \le \delta$ und daher

  $$\left|\int_{\mathbb{R}} g\,d\mathbb{P}_n - \int_{\mathbb{R}} g\,d\mathbb{P}\right| \le \left|\int_{(a,b]} h_\delta\,d\mathbb{P}_n - \int_{(a,b]} h_\delta\,d\mathbb{P}\right| + 2\delta + 6M\varepsilon.$$

  Wir rechnen

  $$\begin{aligned}&\int_{(a,b]} h_\delta\,d\mathbb{P}_n = \sum_{j=2}^{N} g(a_k)\mathbb{P}_n((a_{k-1},a_k])\\ &= \sum_{j=2}^{N} g(a_k)(F_n(a_k) - F_n(a_{k-1})) \overset{n\to\infty}{\to} \sum_{j=2}^{N} g(a_k)(F(a_k) - F(a_{k-1})) = \int_{(a,b]} h_\delta\,d\mathbb{P}.\end{aligned}$$

  Also gilt

  $$\limsup_{n\to\infty} \left|\int_{\mathbb{R}} g\,d\mathbb{P}_n - \int_{\mathbb{R}} g\,d\mathbb{P}\right| \le 2\delta + 6M\varepsilon$$

  für alle $\varepsilon, \delta > 0$. Daraus folgt die gewünschte Konvergenz.

□

3.2 Konvergenz von Zufallsvariablen

3.2.1 Konvergenz in Verteilung und in Wahrscheinlichkeit

Definition 3.3 (Konvergenz in Verteilung). *Sei $n \mapsto X_n$ eine Folge von (reellen) Zufallsvariablen, wobei X_n auf einem Wahrscheinlichkeitsraum $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ definiert ist. Sei X eine Zufallsvariable definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$.*
X_n konvergiert in Verteilung *(distribution) gegen X, geschrieben $X_n \overset{D}{\to} X$, falls F_{X_n} schwach gegen F_X konvergiert.*

Bemerkung. Die schwache Konvergenz einer Folge von Zufallsvariablen erfordert nicht, dass diese auf demselben Wahrscheinlichkeitsraum definiert sind.

Definition 3.4 (Konvergenz in Wahrscheinlichkeit)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $n \mapsto X_n : \Omega \to \mathbb{R}$ eine Folge von Zufallsvariablen und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.*
X_n konvergiert in Wahrscheinlichkeit *(probability) gegen X, geschrieben $X_n \overset{P}{\to} X$, falls*

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

[13: 25.11.2025]

[14: 28.11.2025]

Lemma 3.5. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $n \mapsto X_n : \Omega \to \mathbb{R}$ eine Folge von Zufallsvariablen und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.*
Es gilt

$$X_n \overset{P}{\to} X \qquad \Rightarrow \qquad X_n \overset{D}{\to} X.$$

Die andere Richtung stimmt nicht.

Beweis.
Wir nehmen an, dass $X_n \overset{P}{\to} X$ gilt. Wir rechnen

$$\begin{aligned} |F_{X_n}(t) - F_X(t)| &= |\mathbb{P}(X_n \le t) - \mathbb{P}(X \le t)| \\ &= \Big|\mathbb{P}\Big(\{X_n \le t\} \cap \{X > t\}\Big) - \mathbb{P}\Big(\{X \le t\} \cap \{X_n > t\}\Big)\Big| \\ &\le \mathbb{P}\Big(\{X_n \le t\} \cap \{X > t\}\Big) + \mathbb{P}\Big(\{X \le t\} \cap \{X_n > t\}\Big) \\ &= \mathbb{P}(A_n) + \mathbb{P}(B_n), \end{aligned}$$

wobei $A_n := \{X_n \le t\} \cap \{X > t\}$, und $B_n := \{X \le t\} \cap \{X_n > t\}$. Es gilt

$$\begin{aligned} \mathbb{P}(A_n) &= \mathbb{P}\Big(A_n \cap \{|X - X_n| \le \varepsilon\}\Big) + \mathbb{P}\Big(A_n \cap \{|X - X_n| > \varepsilon\}\Big) \\ &\le \mathbb{P}\Big(A_n \cap \{|X - X_n| \le \varepsilon\}\Big) + \mathbb{P}(|X - X_n| > \varepsilon). \end{aligned}$$

Analog gilt $\mathbb{P}(B_n) \le \mathbb{P}\Big(B_n \cap \{|X - X_n| \le \varepsilon\}\Big) + \mathbb{P}(|X - X_n| > \varepsilon)$. Daraus folgt

$$|F_{X_n}(t) - F_X(t)| \le \mathbb{P}\Big(A_n \cap \{|X - X_n| \le \varepsilon\}\Big) + \mathbb{P}\Big(B_n \cap \{|X - X_n| \le \varepsilon\}\Big) + 2\mathbb{P}(|X - X_n| > \varepsilon).$$

Wir merken, dass

$$\begin{aligned} A_n \cap \{|X - X_n| \le \varepsilon\} &= \{X_n \le t\} \cap \{X > t\} \cap \{|X - X_n| \le \varepsilon\} \subset \{t < X \le t + \varepsilon\} \\ B_n \cap \{|X - X_n| \le \varepsilon\} &= \{X \le t\} \cap \{X_n > t\} \cap \{|X - X_n| \le \varepsilon\} \subset \{t - \varepsilon < X \le t\} \end{aligned}$$

gilt und daher

$$\mathbb{P}\Big(A_n \cap \{|X - X_n| \le \varepsilon\}\Big) + \mathbb{P}\Big(B_n \cap \{|X - X_n| \le \varepsilon\}\Big) \le \mathbb{P}(t - \varepsilon < X \le t + \varepsilon) + 2\mathbb{P}(|X - X_n| > \varepsilon).$$

Aus der Konvergenz in Wahrscheinlichkeit folgt, dass $\limsup_{n\to\infty} \mathbb{P}(|X - X_n| > \varepsilon) = 0$ und daher $\forall t \in \mathbb{R}$ und $\forall \varepsilon > 0$

$$\limsup_{n\to\infty} |F_{X_n}(t) - F_X(t)| \le \mathbb{P}(t-\varepsilon < X \le t+\varepsilon) = F_X(t+\varepsilon) - F_X(t-\varepsilon)$$

gilt. Sei jetzt t ein Stetigkeitspunkt von F_X. Aus dem Grenzwert $\varepsilon \to 0$ folgt

$$\lim_{n\to\infty} |F_{X_n}(t) - F_X(t)| = \limsup_{n\to\infty} |F_{X_n}(t) - F_X(t)| = 0.$$

Damit ist der Beweis von $X_n \overset{D}{\to} X$ beendet.

Um zu zeigen, dass die andere Richtung nicht gilt, genügt es, ein Gegenbeispiel zu finden (Übungsblatt).

□

3.2.2 Konvergenz in L^p

Erinnerung: Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $p \ge 1$, $f\colon \Omega \to \mathbb{R}$ eine messbare Funktion. Die Funktion ist $p-$fach integrierbar, geschrieben $f \in L^p(\Omega, \mathbb{P})$, falls $|f|^p \in L^1(\Omega, \mathbb{P})$ gilt. Wir definieren

$$\|f\|_{L^p} := \left(\int_\Omega |f|^p \, d\mathbb{P}\right)^{\frac{1}{p}}.$$

Dann gilt $f \in L^p(\Omega, \mathbb{P}) \Leftrightarrow \|f\|_{L^p} < \infty$. Die Funktion $\|\cdot\|_{L^p}$ definiert eine Seminorm auf $L^p(\Omega, \mathbb{P})$.

[*Dieser Teil wurde nicht in der Vorlesung besprochen.*
Mit der Äquivalenzrelation $f \sim g$ für $f,\, g \in L^p(\Omega, \mathbb{P})$, falls $f(\omega) = g(\omega)$ für $\mathbb{P}$-fast alle $\omega \in \Omega$, wird

$$|||[f]|||_{L^p} := \|f\|_{L^p}, \quad [f] \in L^p(\Omega, \mathbb{P})/\sim$$

zu einer Norm auf $L^p(\Omega, \mathbb{P})/\sim$, dem Raum der Äquivalenzklassen bezüglich $\sim$. In der Literatur wird typischerweise mit diesem Raum gearbeitet, die Notation $L_p(\Omega, \mathbb{P})$ für die Menge der $p-$fach integrierbaren Funktionen und $L^p(\Omega, \mathbb{P})$ für den Raum der Äquivalenzklassen angewendet. Für unsere Zwecke ist dieser Unterschied aber nicht von Belang und wir werden der Einfachheit halber weiter ohne diese Äquivalenzrelation arbeiten.]

Definition 3.6 (Konvergenz in L^p)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $p \ge 1$, $n \mapsto X_n \in L^p(\Omega, \mathbb{P})$ eine Folge von Zufallsvariablen und $X \in L^p(\Omega, \mathbb{P})$ eine Zufallsvariable. X_n konvergiert* in L^p *gegen X, geschrieben $X_n \overset{L^p}{\to} X$, falls*

$$\lim_{n\to\infty} \|X_n - X\|_{L^p} = 0.$$

Lemma 3.7. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $p \ge 1$, $n \mapsto X_n \in L^p(\Omega, \mathbb{P})$ eine Folge von Zufallsvariablen und $X \in L^p(\Omega, \mathbb{P})$ eine Zufallsvariable.*

Es gilt

$$X_n \overset{L^p}{\to} X \qquad \Rightarrow \qquad X_n \overset{P}{\to} X.$$

Die andere Richtung stimmt nicht.

Beweis.
Sei $\varepsilon > 0$. Wir rechnen, mit der Abschätzung $\mathbf{1}_{|x|>\varepsilon} \leq \left(\frac{|x|}{\varepsilon}\right)^p \ \forall x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(|X - X_n| > \varepsilon) &= \int_\Omega \mathbf{1}_{|X-X_n|>\varepsilon}(\omega)\, d\mathbb{P}(\omega) \leq \int_\Omega \frac{|X - X_n|^p}{\varepsilon^p}\, d\mathbb{P}(\omega) \\ &= \frac{1}{\varepsilon^p} \|X - X_n\|_{L^p}^p.\end{aligned}$$

Aus $X_n \xrightarrow{L^p} X$ folgt, dass $\lim_{n\to\infty} \|X - X_n\|_{L^p} = 0$ und daher $\lim_{n\to\infty} \mathbb{P}(|X - X_n| > \varepsilon) = 0$ $\forall \varepsilon > 0$. Damit ist die Konvergenz $X_n \xrightarrow{P} X$ bewiesen.

Um zu zeigen, dass die andere Richtung nicht gilt, genügt es, ein Gegenbeispiel zu finden (Übungsblatt).

□

3.2.3 Fast sichere Konvergenz

Definition 3.8 (fast sichere Konvergenz)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $n \mapsto X_n \colon \Omega \to \mathbb{R}$ *eine Folge von Zufallsvariablen und* $X \colon \Omega \to \mathbb{R}$ *eine Zufallsvariable.*

X_n *konvergiert* fast sicher *(oder punktweise* $\mathbb{P}$*-fast überall) gegen* X*, geschrieben* $X_n \xrightarrow{f.s.} X$*, falls*

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

Bemerkung 1. Die Menge $\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ ist messbar: Es gilt

$$\begin{aligned}\left\{\lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} &= \left\{\liminf_{n\to\infty} (X_n - X)(\omega) = 0\right\} \cap \left\{\limsup_{n\to\infty} (X_n - X)(\omega) = 0\right\} \\ &= (\liminf_{n\to\infty} (X_n - X))^{-1}(\{0\}) \cap \ (\limsup_{n\to\infty} (X_n - X))^{-1}(\{0\}).\end{aligned}$$

Die Funktion $X_n - X$ ist messbar, und daher ist auch $\liminf_{n\to\infty}(X_n - X)$ und $\limsup_{n\to\infty}(X_n - X)$ messbar (siehe Übungsblatt). Da $\{0\}$ eine Borel Menge ist, gilt

$$(\liminf_{n\to\infty} (X_n - X))^{-1}(\{0\}) \in \mathcal{F}, \qquad (\limsup_{n\to\infty} (X_n - X))^{-1}(\{0\}) \in \mathcal{F}.$$

Also gilt $\{\lim_{n\to\infty} X_n(\omega) = X(\omega)\} \in \mathcal{F}$.

Bemerkung 2. $X_n \xrightarrow{f.s.} X$ genau dann, wenn eine Menge $N \in \mathcal{F}$ existiert, sodass $\mathbb{P}(N) = 0$ und $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ $\forall \omega \in \Omega \setminus N$ gilt.

Lemma 3.9. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $n \mapsto X_n \colon \Omega \to \mathbb{R}$ *eine Folge von Zufallsvariablen und* $X \colon \Omega \to \mathbb{R}$ *eine Zufallsvariable. Folgendes gilt.*

(i) Falls $X_n \xrightarrow{f.s.} X$*, dann gilt auch* $X_n \xrightarrow{P} X$*.*

(ii) Falls $X_n \xrightarrow{f.s.} X$ *und* $\sup_n |X_n|^p \in L^1(\Omega, \mathbb{P})$ *gilt, dann gilt auch* $X_n, X \in L^p$ *und* $X_n \xrightarrow{L^p} X$*.*

Beweis.
(i) Sei $\varepsilon > 0$. Wir rechnen

$$\mathbb{P}(|X - X_n| > \varepsilon) = \int_\Omega \mathbf{1}_{|X-X_n|>\varepsilon}(\omega)\, d\mathbb{P}(\omega) = \int_\Omega g_n(\omega)\, d\mathbb{P}(\omega).$$

Die Funktion g_n erfüllt

$$|g_n| \leq 1 \in L^1(\Omega, \mathbb{P}) \qquad \forall n \in \mathbb{N}.$$

Aus $X_n \overset{f.s.}{\to} X$ folgt, dass eine Menge $N \in \mathcal{F}$ existiert, sodass $\mathbb{P}(N) = 0$ und $\lim_{n\to\infty} X_n(\omega) = X(\omega)\ \forall \omega \in \Omega \setminus N$ gilt. Daraus folgt, dass für jedes $\omega \in \Omega \setminus N$ ein $n_{\omega,\varepsilon}$ existiert, sodass $|X_n(\omega) - X(\omega)| < \varepsilon\ \forall n \geq n_{\omega,\varepsilon}$ gilt und daher

$$\lim_{n\to\infty} g_n(\omega) = 0 \qquad \forall \omega \in \Omega \setminus N.$$

Aus der dominierter Konvergenz folgt

$$\lim_{n\to\infty} \int_\Omega g_n(\omega)\, d\mathbb{P}(\omega) = \int_\Omega \lim_{n\to\infty} g_n(\omega)\, d\mathbb{P}(\omega) = 0.$$

Da $\varepsilon > 0$ beliebig ist, gilt $X_n \overset{P}{\to} X$.

(ii) Übungsblatt.

□

Definition 3.10. *Sei $(\Omega, \mathcal{F})$ ein Messraum, $n \mapsto A_n \in \mathcal{F}$ eine Folge messbaren Mengen. Wir definieren*

$$\limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \left(\bigcup_{k\geq n} A_k \right),$$
$$\liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \left(\bigcap_{k\geq n} A_k \right).$$

Bemerkung. Die Mengen $\limsup_{n\to\infty} A_n, \liminf_{n\to\infty} A_n$ sind messbar, da sie aus abzählbaren Durchschnitten (Vereinigungen) messbaren Mengen kontruiert werden. Folgendes gilt:

(i) $\omega \in \limsup_{n\to\infty} A_n \Leftrightarrow \forall n\ \exists k = k_{n,\omega} \geq n$, sodass $\omega \in A_k \Leftrightarrow \omega$ liegt in unendlichen vielen Mengen A_k.

(ii) $\omega \in \liminf_{n\to\infty} A_n \Leftrightarrow \exists n_\omega \in \mathbb{N}$, sodass $\omega \in A_k\ \forall k \geq n_\omega \Leftrightarrow \omega$ liegt in allen A_k ab einem bestimmten n, welches aber von ω abhängt.

(iii) $(\limsup_{n\to\infty} A_n)^c = \liminf_{n\to\infty} A_n^c$.

Beispiel. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängigen Zufallsvariablen, mit $X_n \sim Ber(p)\ \forall n \in \mathbb{N}$. Sei $A_n := \{X_n = 1\}$. Dann gilt

- $\limsup_{n\to\infty} A_n =$ '1 tritt unendlich oft ein',
- $\liminf_{n\to\infty} A_n =$ 'ab einem zufälligen n gilt $X_k = 1\, \forall k \geq n$'.

Lemma 3.11. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $n \mapsto X_n \colon \Omega \to \mathbb{R}$ *eine Folge von Zufallsvariablen und* $X \colon \Omega \to \mathbb{R}$ *eine Zufallsvariable. Dann gilt*

$$X_n \overset{f.s.}{\to} X \qquad \Leftrightarrow \qquad \mathbb{P}\Big(\limsup_{n\to\infty} \Big\{ |X_n - X| > \frac{1}{k} \Big\} \Big) = 0 \quad \forall k \in \mathbb{N}.$$

Beweis. Sei $\omega \in \Omega$ gegeben. Es gilt $\lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0 \Leftrightarrow \forall k \in \mathbb{N}\ \exists n_{k,\omega} \in \mathbb{N}$, sodass $|X_n(\omega) - X(\omega)| \leq \frac{1}{k}\ \forall n \geq n_{k,\omega} \Leftrightarrow \omega \in \liminf_{n\to\infty} \Big\{ |X_n - X| \leq \frac{1}{k} \Big\}\ \forall k \in \mathbb{N}$.
Daraus folgt, dass

$$\Big\{ \omega \in \Omega |\ \lim_{n\to\infty} X_n(\omega) = X(\omega) \Big\} = \bigcap_{k\in\mathbb{N}} \liminf_{n\to\infty} \Big\{ |X_n - X| \leq \frac{1}{k} \Big\}$$

gilt. Wir rechnen, mit $(\limsup_{n\to\infty} A_n)^c = \liminf_{n\to\infty} A_n^c$,

$$\mathbb{P}\Big(\lim_{n\to\infty} X_n = X \Big) = \mathbb{P}\Big(\bigcap_{k\in\mathbb{N}} \liminf_{n\to\infty} \Big\{ |X_n - X| \leq \frac{1}{k} \Big\} \Big) = 1 - \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} \limsup_{n\to\infty} \Big\{ |X_n - X| > \frac{1}{k} \Big\} \Big).$$

Also gilt $\mathbb{P}\Big(\lim_{n\to\infty} X_n = X \Big) = 1$ genau dann, wenn $\mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} \limsup_{n\to\infty} \Big\{ |X_n - X| > \frac{1}{k} \Big\} \Big) = 0$.
Aus

$$\sup_{n\in\mathbb{N}} \mathbb{P}(A_n) \leq \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n \Big) \leq \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$$

folgt

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n \Big) = 0 \quad \Leftrightarrow \quad \mathbb{P}(A_n) = 0\ \forall n \in \mathbb{N}.$$

Daher gilt $\mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} \limsup_{n\to\infty} \Big\{ |X_n - X| > \frac{1}{k} \Big\} \Big) = 0$ genau dann, wenn

$$\mathbb{P}\Big(\limsup_{n\to\infty} \Big\{ |X_n - X| > \frac{1}{k} \Big\} \Big) = 0 \qquad \forall k \in \mathbb{N}$$

gilt. Damit ist der Beweis beendet.

□

3.2.4 Beispiel: Der Satz von de Moivre-Laplace.

Folgendes Resultat ist unsere erste Version des zentralen Grenzwertsatzes, wie er im 17. Jahrhundert zuerst von de Moivre bewiesen wurde.

Satz 3.12 (Der Satz von de Moivre-Laplace)**.** *Seien* $n \mapsto X_n$ *eine Folge von unabhängigen Bernoulli-verteilten Zufallsvariablen mit Parameter* $0 < p < 1$*. Sei* $Z_n \coloneqq \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - p)$*.*

Dann gilt $Z_n \overset{D}{\to} Z \sim \mathcal{N}(0, p(1-p))$*.*

Bemerkung. Es gilt $\mathbb{E}[X_n] = p$ und $\mathrm{var}[X_n] = p(1-p)\ \forall n$, also gilt $Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - \mathbb{E}[X_j])$ und $Z \sim \mathcal{N}(0, \mathrm{var}[X_1])$.

Beweis. Da F_Z glatt ist, muss man zeigen, dass $\lim_{n\to\infty} F_{Z_n}(t) = F_Z(t)\ \forall t \in \mathbb{R}$ gilt, d.h.

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le t) = \mathbb{P}(Z \le t) = \int_{-\infty}^{t} e^{-\frac{y^2}{2p(1-p)}} \frac{1}{\sqrt{2\pi p(1-p)}} d\mathcal{L}(y).$$

Wir zeigen hier nur den einfacheren Grenzwert

$$\lim_{n\to\infty} \mathbb{P}(a \le Z_n \le b) = \mathbb{P}(a \le Z \le b) = \int_{a}^{b} e^{-\frac{y^2}{2p(1-p)}} \frac{1}{\sqrt{2\pi p(1-p)}} d\mathcal{L}(y), \qquad \forall a < b \in \mathbb{R}.$$

Das Beweisverfahren für den Fall $a = -\infty$ ist ähnlich.
Sei jetzt $S_n := \sum_{j=1}^{n} X_j$. Es gilt $0 \le S_n \le n$, $S_n \sim Bin(n,p)$ und $Z_n = \frac{S_n - np}{\sqrt{n}}$.
Wir rechnen, mit $p_n(k) := \binom{n}{k} p^k (1-p)^{n-k}$,

$$\mathbb{P}(a \le Z_n \le b) = \mathbb{P}(np + \sqrt{n}a \le S_n \le np + \sqrt{n}b) = \sum_{\substack{k\in\mathbb{N} \\ np+\sqrt{n}a \le k \le np+\sqrt{n}b}} p_n(k) \ = \sum_{\substack{x\in\frac{\mathbb{N}}{n} \\ p+\frac{a}{\sqrt{n}} \le x \le p+\frac{b}{\sqrt{n}}}} p_n(xn).$$

Aus Stirling's Formel (Stirlings Formel besagt, dass $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right)$ für $n \to \infty$) folgt

$$p_n(nx) = \frac{1}{\sqrt{2\pi p(1-p)}} \frac{1}{\sqrt{n}} e^{-nI_p(x)} \left(1 + O\left(\frac{1}{n}\right)\right),$$

wobei, für $0 < x < 1$,

$$I_p(x) := x \ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right).$$

Diese Funktion ist glatt und erfüllt $I_p(p) = 0 = I_p'(p)$, $I_p''(p) = \frac{1}{p(1-p)}$. Die Taylorentwicklung um $x = p$ ergibt also

$$I_p(x) = \frac{1}{2p(1-p)}(x-p)^2 + O\left(\frac{1}{n^{\frac{3}{2}}}\right) \qquad \forall |x-p| \le \frac{C}{\sqrt{n}},$$

wobei $C > 0$ ist eine Konstante. Daraus folgt, für $n \gg 1$,

$$\begin{aligned}
\mathbb{P}(a \le Z_n \le b) &= \sum_{\substack{x\in\frac{\mathbb{N}}{n} \\ p+\frac{a}{\sqrt{n}} \le x \le p+\frac{b}{\sqrt{n}}}} \frac{1}{\sqrt{2\pi p(1-p)}} \frac{1}{\sqrt{n}} e^{-\frac{n(x-p)^2}{2p(1-p)}} e^{O(n^{-\frac{1}{2}})} \left(1 + O\left(\frac{1}{n}\right)\right) \\
&= \frac{1}{\sqrt{n}} \sum_{\substack{y\in\frac{\mathbb{N}}{\sqrt{n}} - p\sqrt{n} \\ a \le y \le b}} \frac{1}{\sqrt{2\pi p(1-p)}} e^{-\frac{y^2}{2p(1-p)}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \overset{n\to\infty}{\to} \int_a^b e^{-\frac{y^2}{2p(1-p)}} \frac{1}{\sqrt{2\pi p(1-p)}} d\mathcal{L}(y),
\end{aligned}$$

wo im letzten Schritt die Konvergenz der Riemann Summe $\frac{1}{\sqrt{n}} \sum_{y\in\frac{\mathbb{N}}{\sqrt{n}}} f(y) \overset{n\to\infty}{\to} \int f(y)\, d\mathcal{L}(y)$ für Riemann-integrierbare Funktionen f angewendet wurde. □

[14: 28.11.2025]

[15: 02.12.2025]

Lemma 3.13 (Borel-Cantelli)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum.*

(i) Sei $n \mapsto A_n \in \mathcal{F}$ eine Folge von Ereignissen.
Falls $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$, dann gilt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.

(ii) Sei $n \mapsto A_n \in \mathcal{F}$ eine Folge von unabhängigen Ereignisse.
Falls $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$, dann gilt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Beweis.
(i) Die Folge $n \mapsto B_n := \bigcup_{k\geq n} A_k$ ist monoton fallend und daher gilt mit dem Satz 1.5

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcap_{n\in\mathbb{N}}\left(\bigcup_{k\geq n} A_k\right)\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geq n} A_k\right).$$

Wir rechnen

$$\mathbb{P}\left(\bigcup_{k\geq n} A_k\right) \leq \sum_{k\geq n} \mathbb{P}(A_k).$$

Aus $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$ folgt, dass $\lim_{n\to\infty} \sum_{k\geq n} \mathbb{P}(A_k) = 0$ und daher

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geq n} A_k\right) = 0.$$

Damit ist der Beweis von (i) beendet.

(ii) Es gilt

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1 \quad \Leftrightarrow \quad 0 = \mathbb{P}\left(\left(\limsup_{n\to\infty} A_n\right)^c\right) = \mathbb{P}\left(\liminf_{n\to\infty} A_n^c\right).$$

Die Folge $n \mapsto B_n := \bigcap_{k\geq n} A_k^c$ ist monoton wachsend und daher gilt mit dem Satz 1.5

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n^c\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}}\left(\bigcap_{k\geq n} A_k^c\right)\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right).$$

Aus $\bigcap_{k\geq n} A_k^c = \bigcap_{N\geq n}\left(\bigcap_{n\leq k\leq N} A_k^c\right)$ und dem Satz 1.5 folgt

$$\mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = \lim_{N\to\infty} \mathbb{P}\left(\bigcap_{n\leq k\leq N} A_k^c\right).$$

Da die Ereignisse unabhängig sind, gilt, mit der Abschätzung $(1-x) \leq e^{-x}\ \forall x \in \mathbb{R}$,

$$\mathbb{P}\left(\bigcap_{n\leq k\leq N} A_k^c\right) = \prod_{k=n}^{N} \mathbb{P}(A_k^c) = \prod_{k=n}^{N}(1-\mathbb{P}(A_k)) \leq \prod_{k=n}^{N} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{N}\mathbb{P}(A_k)},$$

und daher, mit $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$,

$$\limsup_{N\to\infty} \mathbb{P}\left(\bigcap_{n\leq k\leq N} A_k^c\right) \leq \limsup_{N\to\infty} e^{-\sum_{k=n}^{N}\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{\infty}\mathbb{P}(A_k)} = e^{-\infty} = 0.$$

Daraus folgt

$$\mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = 0,$$

und damit ist der Beweis von (ii) beendet. □

Folgendes Lemma sagt, dass Konvergenz in Wahrscheinlichkeit fast sichere Konvergenz impliziert, wenn die Wahrscheinlichkeit $\mathbb{P}(|X - X_n| > \varepsilon)$ schnell genug gegen 0 konvergiert.

Lemma 3.14. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $n \mapsto X_n \colon \Omega \to \mathbb{R}$ eine Folge von Zufallsvariablen und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable.*

(i) Falls $\sum_{n\in\mathbb{N}} \mathbb{P}(|X - X_n| > \varepsilon) < \infty \ \forall \varepsilon > 0$ dann gilt $X_n \overset{f.s.}{\to} X$.

(ii) Wir nehmen an, die Zufallsvariablen $(X_n - X)_{n\in\mathbb{N}}$ sind unabhängig.
Falls $X_n \overset{f.s.}{\to} X$, dann gilt $\sum_{n\in\mathbb{N}} \mathbb{P}(|X - X_n| > \varepsilon) < \infty$.

Beweis.
(i) Aus $\sum_{n\in\mathbb{N}} \mathbb{P}(|X - X_n| > \varepsilon) < \infty \ \forall \varepsilon > 0$ folgt, mit dem Lemma von Borel-Cantelli, Lemma 3.13 (i),

$$\mathbb{P}\left(\limsup_{n\to\infty}\{|X - X_n| > \varepsilon\}\right) = 0 \qquad \forall \varepsilon > 0.$$

Die Aussage folgt jetzt aus dem Lemma 3.11.

(ii) Durch Widerspruch nehmen wir an, dass es ein $\varepsilon_0 > 0$ gibt, sodass

$$\sum_{n\in\mathbb{N}} \mathbb{P}(|X - X_n| > \varepsilon_0) = \infty.$$

Aus dem Lemma von Borel-Cantelli (ii) folgt, dass $\mathbb{P}(\limsup_{n\to\infty}\{|X - X_n| > \varepsilon_0\}) = 1$. Also gilt $\lim_{n\to\infty}(X_n - X) \neq 0$ $\mathbb{P}-$fast überall, was $X_n \overset{f.s.}{\to} X$ widerspricht.

□

Lemma 3.15. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $n \mapsto X_n \colon \Omega \to \mathbb{R}$ eine Folge von Zufallsvariablen und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable.*

Falls $X_n \overset{D}{\to} X$ und X $\mathbb{P}-$fast überall konstant ist, dann gilt $X_n \overset{P}{\to} X$.

Beweis. Da X $\mathbb{P}-$fast überall konstant ist, existiert eine Konstante $c \in \mathbb{R}$, sodass

$$\mathbb{P}(X = c) = 1.$$

Insbesondere gilt $\mathbb{P}(A \cap \{X \neq c\}) \leq \mathbb{P}(X \neq c) = 0$ für alle $A \in \mathcal{F}$. Sei nun $\varepsilon > 0$ gegeben. Wir rechnen

$$\begin{aligned}
\mathbb{P}(|X - X_n| > \varepsilon) &= \mathbb{P}\Big(\{|X_n - X| > \varepsilon\} \cap \{X = c\}\Big) + \mathbb{P}\Big(\{|X_n - X| > \varepsilon\} \cap \{X \neq c\}\Big) \\
&= \mathbb{P}\Big(\{|X_n - X| > \varepsilon\} \cap \{X = c\}\Big) = \mathbb{P}\Big(\{|X_n - c| > \varepsilon\} \cap \{X = c\}\Big) \\
&= \mathbb{P}(|X_n - c| > \varepsilon) - \mathbb{P}\Big(\{|X_n - c| > \varepsilon\} \cap \{X \neq c\}\Big) = \mathbb{P}(|X_n - c| > \varepsilon) \\
&= \mathbb{P}(X_n > c + \varepsilon) + \mathbb{P}(X_n < c - \varepsilon) \leq \mathbb{P}(X_n > c + \varepsilon) + \mathbb{P}(X_n \leq c - \varepsilon) \\
&= (1 - F_{X_n}(c + \varepsilon)) + F_{X_n}(c - \varepsilon).
\end{aligned}$$

$X = c$ fast überall, also gilt

$$F_X(t) = \begin{cases} 0 & t < c \\ 1 & t \geq c, \end{cases}$$

und daher ist die Verteilungsfunktion F_X stetig in $t \neq c$. Aus der Konvergenz in Verteilung dann folgt

$$\lim_{n\to\infty} F_{X_n}(c+\varepsilon) = F_X(c+\varepsilon) = 1, \qquad \lim_{n\to\infty} F_{X_n}(c-\varepsilon) = F_X(c-\varepsilon) = 0,$$

und daher

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

□

Lemma 3.16. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $n \mapsto X_n \colon \Omega \to \mathbb{R}$ eine Folge von Zufallsvariablen und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable.*

Falls $X_n \overset{P}{\to} X$, dann existiert eine Teilfolge $j \mapsto n_j$, sodass $X_{n_j} \overset{f.s.}{\to} X$.

Beweis. Übungsblatt □

$$\begin{tikzcd}
X_n \overset{f.s.}{\to} X \arrow[r, shift left] \arrow[dr, "\sup_n |X_n|^p \in L^1"'] & X_n \overset{P}{\to} X \arrow[l, shift left, "\text{Teilfolge}"] \arrow[r, shift left] & X_n \overset{D}{\to} X \arrow[l, shift left, "X = c \text{ f.ü.}"] \\
& X_n \overset{L^p}{\to} X \arrow[u] &
\end{tikzcd}$$

4 Momente, Ungleichungen

Definition 4.1. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable.*

(i) Sei $p \in \mathbb{N}$ und $X \in L^p(\Omega, \mathbb{P})$, mit $p \geq 1$. Das $p-$te Moment *von X ist definiert durch*

$$M_p := \mathbb{E}[X^p].$$

(ii) Die Momenterzeugende Funktion *von X ist definiert durch*

$$\begin{aligned} M \colon \quad \mathbb{R} &\to \quad [0, \infty] \\ t &\mapsto \quad M(t) := \mathbb{E}[e^{tX}]. \end{aligned}$$

Bemerkungen.

- Falls $X \in L^p(\Omega, \mathbb{P})$, dann gilt $\mathbb{E}[|X|^p]$ und daher ist M_p wohldefiniert.
- Die Funktion $t \mapsto M(t)$ ist messbar und nichtnegativ. Das integral $\mathbb{E}[e^{tX}] = \int_\Omega e^{tX} d\mathbb{P}$ ist also immer wohldefiniert, darf aber den Wert $+\infty$ nehmen.
- Es gilt $M(0) = \mathbb{E}[1] = 1$.
- $M(t)$ heißt auch *Laplace Transformierte* des Maßes $\mathbb{P}$.

Lemma 4.2. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X\colon \Omega \to \mathbb{R}$ eine Zufallsvariable. Wir nehmen an, es gibt $h > 0$, sodass $M(h) + M(-h) < \infty$ gilt. Dann gilt Folgendes.*

(i) $M(t) < \infty \ \forall t \in [-h, h]$,

(ii) $M \in C^\infty((-h, h))$,

(iii) $X \in L^n(\Omega, \mathbb{P}) \ \forall n \in \mathbb{N}$ *und* $M_n = \frac{d^n}{dt^n}|_{t=0} M(t)$.

Beweis. Es gilt, für $|t| \leq h$,

$$e^{|tX|} \leq e^{h|X|} \leq e^{hX} + e^{-hX}.$$

Daraus folgt

$$\int_\Omega e^{|tX|} d\mathbb{P} \leq \int_\Omega e^{hX} d\mathbb{P} + \int_\Omega e^{-hX} d\mathbb{P} = M(h) + M(-h) < \infty,$$

und daher ist $M(t) < \infty$ für alle $|t| < h$. Wir rechnen, mit Lemma 1.32,

$$\int_\Omega e^{|tX|} d\mathbb{P} = \int_\Omega \left(\sum_{n \in \mathbb{N}} \frac{|tX|^n}{n!} \right) d\mathbb{P} = \sum_{n \in \mathbb{N}} \frac{|t|^n}{n!} \int_\Omega |X|^n \ d\mathbb{P}.$$

Aus $M(t) < \infty$ folgt, dass $\int_\Omega |X|^n \ d\mathbb{P} < \infty$ und daher $X \in L^n(\Omega, \mathbb{P}) \ \forall n \in \mathbb{N}$ gilt. Wir rechnen

$$M(t) = \int_\Omega e^{tX} d\mathbb{P} = \int_\Omega \left(\lim_{N \to \infty} f_N \right) d\mathbb{P}$$

wobei $f_N := \sum_{n=0}^N \frac{(tX)^n}{n!}$. Aus $\lim_{N \to \infty} f_N = e^{tX}$ und $|f_N| \leq e^{|tX|} \in L^1(\Omega, \mathbb{P})$ folgt, durch dominierte Konvergenz,

$$M(t) = \lim_{N \to \infty} \int_\Omega f_N \ d\mathbb{P} = \sum_{n \in \mathbb{N}} \frac{t^n}{n!} \mathbb{E}[X^n] \qquad \forall |t| \leq h.$$

$M(t)$ ist also eine Potenzreihe und daher gilt $M \in C^\infty((-h, h))$ und $M_n = \frac{d^n}{dt^n}|_{t=0} M(t)$.

□

[15: 02.12.2025]

[16: 05.12.2025]

Beispiele.

- Falls $X \sim \mathcal{N}(m, \sigma^2)$, dann gilt $M(t) = e^{\frac{\sigma^2 t^2}{2} + mt}$.
- Falls $X \sim Bin(n, p)$, dann gilt $M(t) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{tk} = \left((e^t p) + (1-p) \right)^n$.
- Falls $X \sim Cauchy(a)$, dann gilt $M(t) = +\infty \ \forall t \neq 0$:

$$M(t) = \frac{a}{\pi} \int_\mathbb{R} \frac{e^{tx}}{a^2 + x^2} d\mathcal{L}(x) \geq \frac{a}{\pi} \int_0^\infty \frac{e^{|t|x}}{a^2 + x^2} d\mathcal{L}(x) \geq C_a \int_0^\infty d\mathcal{L}(x) = \infty,$$

  für eine Konstante $C_a > 0$.

Satz 4.3. *Sei* $(\Omega, \mathcal{F}, \mu)$ *ein Massraum.*

(i) (Cauchy-Schwarz) $f, g \in L^2(\Omega, \mu) \Rightarrow f \cdot g \in L^1(\Omega, \mu)$ *und*

$$\|fg\|_{L^1} \leq \|f\|_{L^2}\|g\|_{L^2},$$

(ii) (Hölder) Seien $1 < p, q < \infty$ *mit* $\frac{1}{p} + \frac{1}{q} = 1$, *oder sei* $p = 1, q = \infty$.
Dann gilt $f \in L^p(\Omega, \mu), g \in L^q(\Omega, \mu) \Rightarrow f \cdot g \in L^1(\Omega, \mu)$ *und*

$$\|fg\|_{L^1} \leq \|f\|_{L^p}\|g\|_{L^q},$$

(iii) (Minkowski) $f, g \in L^p(\Omega, \mu) \Rightarrow f + g \in L^p(\Omega, \mu)$ *und*

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}.$$

(iv) (Jensen) Sei $\mu = \mathbb{P}$ *ein Wahrscheinlichkeitsmaß,* $\Phi\colon \mathbb{R} \to \mathbb{R}$ *eine konvexe Funktion (also* $\Phi(tx + (1-t)y) \leq t\Phi(x) + (1-t)\Phi(y)$ *für alle* $x,\ y \in \mathbb{R}$ *und* $t \in [0,1]$*).*
Falls $f \in L^1(\Omega, \mathbb{P})$, *dann ist* $(\Phi \circ f)_+$ *messbar,* $(\Phi \circ f)_- \in L^1(\Omega, \mathbb{P})$ *und*

$$\Phi\left(\int_\Omega f \, d\mathbb{P}\right) \leq \int_\Omega \Phi \circ f \, d\mathbb{P}.$$

Beweis. Analysis 3. □

Korollar 4.4. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum. Falls* $f \in L^p(\Omega, \mathbb{P})$, *dann gilt* $f \in L^q(\Omega, \mathbb{P})\ \forall 1 \leq q \leq p$.

Beweis. Übungsblatt. □

Satz 4.5 (allgemeine Markov-Ungleichung)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum.*
Sei $h\colon \mathbb{R} \to [0, \infty)$ *monoton wachsend und messbar,* $X\colon \Omega \to \mathbb{R}$ *eine Zufallsvariable.*
Dann gilt $\forall \delta \in \mathbb{R}$, *sodass* $h(\delta) > 0$,

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

Beweis. Es gilt $\mathbb{P}(X \geq \delta) = \int_\Omega \mathbf{1}_{X \geq \delta}(\omega) d\mathbb{P}(\omega) = \int_\mathbb{R} \mathbf{1}_{[\delta,\infty)}(x) d\mathbb{P}_X(x)$.
Sei $f(x) := \frac{h(x)}{h(\delta)}$. Diese Funktion ist monoton wachsend und erfüllt

$$\mathbf{1}_{[\delta,\infty)}(x) \leq f(x),$$

daher

$$\mathbb{P}(X \geq \delta) \leq \int_\mathbb{R} f(x) d\mathbb{P}_X(x) = \frac{\mathbb{E}[h(X)]}{h(\delta)}.$$

□

Wichtige Speziallfälle.

(i) *Markov Ungleichung*: Es gilt $\forall \delta > 0$

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

Beweis. Setze $Y := |X|$ und $h(x) := x\mathbf{1}_{[0,\infty)}(x)$. Dann gilt

$$\mathbb{P}(|X| \geq \delta) = \mathbb{P}(Y \geq \delta) \leq \frac{\mathbb{E}[h(Y)]}{h(\delta)} = \frac{\mathbb{E}[Y\mathbf{1}_{[0,\infty)}(Y)]}{\delta} = \frac{\mathbb{E}[|X|]}{\delta}.$$

□

(ii) *Polynomiale Markov (oder Tschebishev) Ungleichung*: Es gilt $\forall \delta > 0, n \in \mathbb{N}$

$$\mathbb{P}(|X| \geq \delta) \leq \frac{\mathbb{E}[|X|^n]}{\delta^n}.$$

Beweis. Setze $Y := |X|$ und $h(x) := x^n\mathbf{1}_{[0,\infty)}(x)$. Dann gilt

$$\mathbb{P}(|X| \geq \delta) = \mathbb{P}(Y \geq \delta) \leq \frac{\mathbb{E}[h(Y)]}{h(\delta)} = \frac{\mathbb{E}[Y\mathbf{1}_{[0,\infty)}(Y)]}{\delta} = \frac{\mathbb{E}[|X|^n]}{\delta^n}.$$

□

(iii) *Tschebishev Ungleichung*: Es gilt $\forall \delta > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta) \leq \frac{\mathrm{var}(X)}{\delta^2}.$$

Beweis. Setze $Y := |X - \mathbb{E}[X]|$ und $h(x) := x^2\mathbf{1}_{[0,\infty)}(x)$. Dann gilt

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta) = \mathbb{P}(Y \geq \delta) \leq \frac{\mathbb{E}[h(Y)]}{h(\delta)} = \frac{\mathbb{E}[Y^2\mathbf{1}_{[0,\infty)}(Y)]}{\delta^2} = \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\delta^2}.$$

□

(iv) *Exponentielle Tschebishev (oder Markov, oder Chernov) Ungleichung*: Es gilt $\forall \delta \in \mathbb{R}$

$$\mathbb{P}(X \geq \delta) \leq e^{-\mu\delta}\mathbb{E}[e^{\mu X}] \quad \forall \mu > 0.$$

Insebesondere gilt

$$\mathbb{P}(X \geq \delta) \leq \inf_{\mu > 0} e^{-\mu\delta}\mathbb{E}[e^{\mu X}].$$

Die Ungleichung folgt aus dem Satz 4.5 mit $h(x) := e^{\mu x}$.

Beispiel 1. Sei $X \sim \mathcal{N}(0, \sigma^2)$ und $\delta > 0$. Aus dem exponentiellen Tschebishev Ungleichung folgt, für $\mu > 0$,

$$\mathbb{P}(X \geq \delta) \leq e^{-\mu\delta}\mathbb{E}[e^{\mu X}] = e^{-\mu\delta + \frac{1}{2}\sigma^2\mu^2}$$

und daher

$$\mathbb{P}(X \geq \delta) \leq \inf_{\mu > 0} e^{-\mu\delta + \frac{1}{2}\sigma^2\mu^2} = e^{-\frac{1}{2}\frac{\delta^2}{\sigma^2}}.$$

Beispiel 2. Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen, $\delta \in \mathbb{R}$. Aus der exponentiellen Tschebishev Ungleichung folgt, für $\mu > 0$,

$$\mathbb{P}\left(\frac{\sum_{j=1}^n X_j}{n} \geq \delta\right) = \mathbb{P}\left(\sum_{j=1}^n X_j \geq n\delta\right) \leq e^{-\mu n\delta}\mathbb{E}[e^{\mu\sum_{j=1}^n X_j}] = \prod_{j=1}^n e^{-\mu\delta}\mathbb{E}[e^{\mu X_j}].$$

Falls die Zufallsvariablen auch identisch verteilt sind, gilt

$$\mathbb{P}\left(\frac{\sum_{j=1}^n X_j}{n} \geq \delta\right) \leq \left(\inf_{\mu>0} e^{-\mu\delta}\mathbb{E}[e^{\mu X_1}]\right)^n.$$

5 Gesetze der großen Zahlen

Unser Ziel ist es, das empirische Mittel $\frac{1}{n}\sum_{j=1}^n X_j$ im Grenzwert $n \to \infty$ zu studieren. Dafür brauchen wir den Begriff von Covarianz.

Definition 5.1 (Covarianz)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X, Y\colon \Omega \to \mathbb{R}$ zwei 2-fach integrierbar Zufallsvariablen.*
Die Covarianz von X, Y ist definiert durch

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

Bemerkung 1. Aus $X, Y \in L^2(\Omega, \mathbb{P})$ folgt, dass $X, Y \in L^1(\Omega, \mathbb{P})$ und $XY \in L^1(\Omega, \mathbb{P})$ gelten und daher ist die Covarianz wohldefiniert.

Bemerkung 2. Es gilt $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$ und $\operatorname{Cov}(X, X) = \operatorname{var}(X)$. Im Gegensatz zu der Varianz darf die Covarianz auch negative Werte annehmen.

Bemerkung 3 Falls X, Y unkorreliert sind gilt $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ und daher gilt $\operatorname{Cov}(X, Y) = 0$.

Lemma 5.2. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X_1, \ldots, X_n \in L^2(\Omega, \mathbb{P})$ Zufallsvariablen. Wir setzen $S_n := \sum_{j=1}^n X_j$.*

(i) $S_n \in L^2(\Omega, \mathbb{P})$,

(ii) $\mathbb{E}[\sum_{j=1}^n X_j] = \sum_{j=1}^n \mathbb{E}[X_j]$,

(iii) $\operatorname{var}(S_n) = \sum_{j=1}^n \operatorname{var}(X_j) + \sum_{j\neq k} \operatorname{Cov}(X_j, X_k) = \sum_{j=1}^n \operatorname{var}(X_j) + 2\sum_{j<k} \operatorname{Cov}(X_j, X_k)$.

Beweis. (i) folgt aus Minkowski Ungleichung. (ii) folgt aus der Linearität des Erwartungswert. (iii) Wir rechnen

$$\begin{aligned}
\operatorname{var}(S_n) &= \mathbb{E}\left[S_n^2\right] - \mathbb{E}\left[S_n\right]^2 = \mathbb{E}\left[(S_n - \mathbb{E}\left[S_n\right])^2\right] \\
&= \mathbb{E}\left[\left(\sum_{j=1}^n (X_j - \mathbb{E}[X_j])\right)^2\right] = \sum_{j,k=1}^n \mathbb{E}\left[(X_j - \mathbb{E}[X_j])(X_k - \mathbb{E}[X_k])\right] \\
&= \sum_{j=1}^n \mathbb{E}\left[(X_j - \mathbb{E}[X_j])^2\right] + \sum_{j\neq k} \mathbb{E}\left[(X_j - \mathbb{E}[X_j])(X_k - \mathbb{E}[X_k])\right] \\
&= \sum_{j=1}^n \operatorname{var}(X_j) + \sum_{j\neq k} \operatorname{Cov}(X_j, X_k) = \sum_{j=1}^n \operatorname{var}(X_j) + 2\sum_{j<k} \operatorname{Cov}(X_j, X_k).
\end{aligned}$$

$\square$

Bemerkung. Falls die Zufallsvariablen unkorreliert sind, gilt

$$\operatorname{var}(S_n) = \sum_{j=1}^{n} \operatorname{var}(X_j).$$

5.1 Schwache Version

Satz 5.3 (Schwaches Gesetz der großen Zahlen, L^2 Version)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, und $(X_n)_{n\in\mathbb{N}}$ eine Folge Zufallsvariablen mit $X_n \in L^2(\Omega, \mathbb{P})$, $\forall n \in \mathbb{N}$.*
Wir nehmen an, es gibt eine Folge $(c_n)_{n\geq 0} \in [0, \infty)^{\mathbb{N}}$, sodass

- $C \coloneqq \sum_{n\in\mathbb{N}_0} c_n < \infty$,
- $\operatorname{Cov}(X_j, X_k) \leq c_{|j-k|}$ $\forall j, k \in \mathbb{N}$.

Sei $S_n \coloneqq \sum_{j=1}^{n} X_j$. Folgendes gilt.

(i) $\operatorname{var}\left(\frac{S_n}{n}\right) \leq \frac{2C}{n}$,

(ii) Falls $m \in \mathbb{R}$ und $\mathbb{E}[X_j] = m$ $\forall j \in \mathbb{N}$, dann gilt $\frac{S_n}{n} \xrightarrow{L^2} m$ und deshalb auch $\frac{S_n}{n} \xrightarrow{P} m$.

Beweis.
(i) Wir rechnen, mit $\sum_{k=j}^{n} c_{k-j} \leq \sum_{k=1}^{\infty} c_k$,

$$\begin{aligned}\operatorname{var}\left(\frac{S_n}{n}\right) &= \frac{1}{n^2}\operatorname{var}(S_n) = \frac{1}{n^2}\left[\sum_{j=1}^{n} \operatorname{var}(X_j) + 2\sum_{j<k} \operatorname{Cov}(X_j, X_k)\right] \\ &\leq \frac{1}{n^2}\left[c_0 \sum_{j=1}^{n} 1 + 2\sum_{j=1}^{n}\sum_{k=j}^{n} c_{k-j}\right] \leq \frac{1}{n^2} n 2C = \frac{2C}{n}. \end{aligned} \tag{5.1}$$

(ii) Wir rechnen

$$\left\|\frac{S_n}{n} - m\right\|_{L^2} = \left(\int_\Omega \left(\frac{S_n}{n} - m\right)^2 d\mathbb{P}\right)^{\frac{1}{2}} = \operatorname{var}\left(\frac{S_n}{n}\right)^{\frac{1}{2}} \leq \frac{\sqrt{2C}}{\sqrt{n}} \overset{n\to\infty}{\to} 0.$$

$\square$

[16: 05.12.2025]

[17: 09.12.2025]

Erinnerung: Letztes Mal haben wir die schwache L^2-Version vom Gesetz der großen Zahlen besprochen, Satz 5.3: Falls $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen ist, sodass $X_n \in L^2(\Omega, \mathbb{P})$ $\forall n \in \mathbb{N}$ und es eine Folge $(c_n)_{n\geq 0} \in [0, \infty)^{\mathbb{N}}$ gibt, sodass

- $C \coloneqq \sum_{n\in\mathbb{N}_0} c_n < \infty$,
- $\operatorname{Cov}(X_j, X_k) \leq c_{|j-k|}$ $\forall j, k \in \mathbb{N}$,

dann gilt Folgendes für die empirischen Mittel $S_n \coloneqq \sum_{j=1}^n X_j$:

(i) $\operatorname{var}\left(\frac{S_n}{n}\right) \leq \frac{2C}{n}$,

(ii) Falls $m \in \mathbb{R}$ und $\mathbb{E}[X_j] = m \ \forall j \in \mathbb{N}$, dann gilt $\frac{S_n}{n} \xrightarrow{L^2} m$ und deshalb auch $\frac{S_n}{n} \xrightarrow{P} m$.

Anwendungsbeispiele: Konzentration von Produktmaßen in hoher Dimension. Seien $X_j : \Omega \to \mathbb{R}$, $j \in \{1, \ldots, n\}$ unabhängige, identisch verteilte Zufallsvariablen und $X = (X_1, \ldots, X_n)$. Die gemeinsame Verteilung $\mathbb{P}_X = \mathbb{P}_{X_1,\ldots,X_n}$ ist gegeben durch

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n} = \mathbb{P}_{X_1}^n.$$

Wir nehmen an, dass $X_1 \in L^4(\Omega, \mathbb{P})$ mit

$$\mathbb{E}[X_1^2] = 1, \quad 1 < M_4 = \mathbb{E}[X_1^4] < \infty.$$

Sei $\|X\|_2^2 \coloneqq \sum_{j=1}^n X_j^2$. Dann gilt

$$\forall a > 0 : \quad \mathbb{P}\left(|\|X\|_2^2 - n| > a\sqrt{n}\right) \leq \frac{M_4 - 1}{a^2}.$$

Z.B. gilt für $a = 10\sqrt{M_4 - 1}$ auch $\mathbb{P}\left(|\|X\|_2^2 - n| > 10\sqrt{M_4 - 1}\sqrt{n}\right) \leq \frac{1}{100}$.

Beweis. Sei $Y_i \coloneqq X_i^2$. $Y_1, \ldots, Y_n$ sind unabhängig und identisch verteilt mit

$$\mathbb{E}[Y_i] = \mathbb{E}[X_i^2] = 1, \quad \mathbb{E}[Y_i^2] = \mathbb{E}[X_i^4] = M_4.$$

Somit gilt $\operatorname{var}(Y_i) = M_4 - 1$ und

$$S_n \coloneqq \sum_{j=1}^n Y_j = \sum_{j=1}^n X_j^2 = \|X\|_2^2$$

mit $\mathbb{E}[S_n] = n\mathbb{E}[Y_1] = n$. Nach der Tschebishev-Ungleichung gilt

$$\begin{aligned}
\mathbb{P}\left(|\|X\|_2^2 - n| > a\sqrt{n}\right) &= \mathbb{P}\left(\left|\frac{\|X\|_2^2}{n} - 1\right| > \frac{a}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| > \frac{a}{\sqrt{n}}\right) \\
&\leq \frac{\operatorname{var}\left(\frac{S_n}{n}\right)}{\frac{a^2}{n}}.
\end{aligned}$$

Es gilt aufgrund der Unabhängigkeit der Y_i und (5.1)

$$\operatorname{var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{j=1}^n \operatorname{var}(Y_j) = \frac{M_4 - 1}{n}.$$

Es folgt daher

$$\mathbb{P}\left(|\|X\|_2^2 - n| > a\sqrt{n}\right) \leq \frac{M_4 - 1}{a^2}.$$

□

Wir wollen als nächstes die L^p-Konvergenz im Gesetz der großen Zahlen verbessern zu fast sicherer Konvergenz. Es gilt nach Borel-Cantelli

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) < \infty \quad \forall \varepsilon > 0 \implies \frac{S_n}{n} \overset{f.s.}{\to} m.$$

Das schwache Gesetz der großen Zahlen impliziert

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) \le \frac{\operatorname{var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} \le \frac{C}{n\varepsilon^2}.$$

Da $\sum_n \frac{1}{n} = \infty$, ist dies nicht genug, um die fast sichere Konvergenz zu folgern und wir brauchen eine stärkere Abschätzung.

Satz 5.4 (Starkes Gesetz der großen Zahlen, L^4 Version)**.** *Seien* $(X_n)_{n\in\mathbb{N}}$ unabhängige, identisch verteilte *Zufallsvariablen mit* $X_1 \in L^4(\Omega, \mathbb{P})$*. Dann gilt*

$$S_n = \frac{\sum_{i=1}^n X_i}{n} \overset{f.s.}{\to} \mathbb{E}[X_1].$$

Beweis. Wir können annehmen, dass $\mathbb{E}[X_1] = 0$. Ansonsten beweisen wir das Theorem zuerst für $\tilde{X}_i \coloneqq X_i - \mathbb{E}[X_i]$. Es gilt nach der polynomialen Markov-Ungleichung

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \le \mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right]\frac{1}{\varepsilon^4} = \frac{1}{n^4\varepsilon^4}\mathbb{E}[S_n^4].$$

Außerdem gilt

$$\mathbb{E}[S_n^4] = \sum_{(i_1,i_2,i_3,i_4)\in\{1,\dots,n\}^4} \mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}].$$

Falls $i_1 \notin \{i_2, i_3, i_4\}$, so folgt aus der Unabhängigkeit $\mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}] = 0$. Es bleiben also nur zwei Fälle übrig. Entweder gilt $i_1 = i_2 = i_3 = i_4$ oder $i_1 = i_2 \ne i_3 = i_4$ sowie Permutationen dieser Fälle. Insgesamt folgt

$$\mathbb{E}[S_n^4] = \sum_{i=1}^n \mathbb{E}[X_i^4] + 3\sum_{i_1\ne i_2} \mathbb{E}[X_{i_1}^2]\mathbb{E}[X_{i_2}^2] = nM_4 + 3n(n-1)M_2^2.$$

Somit

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \le \frac{1}{n^4\varepsilon^4}(nM_4 + 3n(n-1)M_2^2) \le \frac{\tilde{C}}{\varepsilon^4}\frac{1}{n^2}$$

und

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \le \frac{\tilde{C}}{\varepsilon^4}\sum_{n\in\mathbb{N}}\frac{1}{n^2} < \infty \quad \forall \varepsilon > 0.$$

Mit Borel-Cantelli folgt nun $\frac{S_n}{n} \overset{f.s.}{\to} 0$. □

Satz 5.5 (Starkes Gesetz der großen Zahlen, L^2 Version)**.** *Seien* $(X_n)_{n\in\mathbb{N}}$ *Zufallsvariablen mit*

- $X_n \in L^2(\Omega, \mathbb{P})$ $\forall n \in \mathbb{N}$ *und* $\sup_n \operatorname{var}(X_n) =: C_0 < \infty$,
- $(X_n)_{n\in\mathbb{N}}$ *sind unkorreliert, also* $\operatorname{Cov}(X_i, X_j) = 0$ *für* $i \ne j$.

Sei $S_n := \sum_{j=1}^n X_j$. Dann gilt

$$\frac{S_n}{n} - \frac{\mathbb{E}[S_n]}{n} \overset{f.s.}{\to} 0 \quad \text{für } n \to \infty.$$

Insbesondere gilt: Falls $\mathbb{E}[X_j] = m\ \forall j$, dann folgt $\frac{S_n}{n} \overset{f.s.}{\to} m$.

Beweis. Wir nehmen erneut an, dass $\mathbb{E}[X_i] = 0\ \forall i$, und somit $\mathbb{E}[S_n] = 0\ \forall n$.

Schritt 1. Wir zeigen fast sichere Konvergenz der Teilfolge $\frac{S_{k^2}}{k^2}$:
Da die X_i unkorelliert sind, folgt

$$\operatorname{var}(S_n) = \sum_{j=1}^n \operatorname{var}(X_j) \leq n \sup_{j\in\mathbb{N}} \operatorname{var}(X_j) = C_0 n.$$

Sei $n_k = k^2$. Nach Tschebishev gilt

$$\mathbb{P}\left(\left|\frac{S_{n_k}}{n_k}\right| > \varepsilon\right) \leq \frac{\operatorname{var}\left(\frac{S_{n_k}}{n_k}\right)}{\varepsilon^2} \leq \frac{C_0}{\varepsilon^2}\frac{n_k}{n_k^2} = \frac{C_0}{\varepsilon^2}\frac{1}{k^2}.$$

Daher gilt

$$\sum_{k\in\mathbb{N}} \mathbb{P}\left(\left|\frac{S_{n_k}}{n_k}\right| > \varepsilon\right) \leq \frac{C_0}{\varepsilon^2}\sum_{k\in\mathbb{N}}\frac{1}{k^2} < \infty.$$

Somit existiert $N_1 \in \mathcal{F}$, sodass $\mathbb{P}(N_1) = 0$ und $\frac{S_{k^2}(\omega)}{k^2} \overset{k\to\infty}{\to} 0\ \forall \omega \notin N_1$.

Schritt 2. Wir betrachten Fluktuationen der Folge S_n um S_{k^2}:
Sei $D_k := \max_{k^2 \leq l < (k+1)^2} |S_l - S_{k^2}|$. Es gilt

$$\left\{\max_{k^2\leq l<(k+1)^2} |S_l - S_{k^2}| > \varepsilon k^2\right\} = \bigcup_{l=k^2}^{(k+1)^2-1} \{|S_l - S_{k^2}| > \varepsilon k^2\}$$

und somit

$$\mathbb{P}\left(\left|\frac{D_k}{k^2}\right| > \varepsilon\right) = \mathbb{P}(|D_k| > \varepsilon k^2) \leq \sum_{l=k^2}^{(k+1)^2-1} \mathbb{P}(|S_l - S_{k^2}| > \varepsilon k^2) \leq \sum_{l=k^2}^{(k+1)^2-1} \frac{\operatorname{var}(S_l - S_{k^2})}{\varepsilon^2 k^4}.$$

Da $S_l - S_{k^2} = \sum_{j=k^2+1}^l X_j$, gilt $\operatorname{var}(S_l - S_{k^2}) = \sum_{j=k^2+1}^l \operatorname{var}(X_j) \leq C_0(l-k^2) \leq 2kC_0$.
Es folgt

$$\mathbb{P}\left(\left|\frac{D_k}{k^2}\right| > \varepsilon\right) \leq \sum_{l=k^2}^{(k+1)^2-1} \frac{2kC_0}{\varepsilon^2 k^4} = 2k\frac{2kC_0}{\varepsilon^2 k^4} = \frac{4C_0}{\varepsilon^2 k^2}.$$

Es folgt also mit Borel-Cantelli die fast sichere Konvergenz der Mittel $\frac{D_k}{k^2}$ gegen 0, nämlich

$$\sum_{k\in\mathbb{N}} \mathbb{P}\left(\left|\frac{D_k}{k^2}\right| > \varepsilon\right) < \infty \quad \forall \varepsilon > 0 \implies \exists N_2 \in \mathcal{F}, \quad \mathbb{P}(N_2) = 0, \quad \frac{D_k}{k^2} \overset{k\to\infty}{\to} 0 \quad \forall \omega \notin N_2.$$

Schritt 3. Für alle $n \in \mathbb{N}$ existiert genau ein $k_n \in \mathbb{N}$, sodass $k_n^2 \leq n < (k_n + 1)^2$. Zudem ist

$$S_n(\omega) = S_{k_n^2}(\omega) + [S_n(\omega) - S_{k_n^2}(\omega)].$$

Es folgt für $\omega \notin N_1 \cup N_2$

$$\begin{aligned}\frac{|S_n(\omega)|}{n} &\leq \frac{1}{n}|S_{k_n^2}(\omega)| + \frac{1}{n}|S_n(\omega) - S_{k_n^2}(\omega)| \leq \frac{1}{n}|S_{k_n^2}(\omega)| + \frac{1}{n}|D_{k_n}(\omega)| \\ &\leq \frac{1}{k_n^2}|S_{k_n^2}(\omega)| + \frac{1}{k_n^2}|D_{k_n}(\omega)| \overset{n\to\infty}{\to} 0.\end{aligned}$$

Daher folgt $\frac{S_n}{n} \to 0$ fast sicher.

□

[17: 09.12.2025]

[18: 12.12.2025]

5.2 Starke Version

Satz 5.6 (Starkes Gesetz der großen Zahlen, L^1-Version)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum. Sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge von Zufallsvariablen, sodass*

- $X_n \in L^1(\Omega, \mathbb{P})$ $\forall n \in \mathbb{N}$,
- X_n *sind* paarweise unabhängig *und* identisch verteilt *mit* $\mathbb{E}[X_i] = \mathbb{E}[X_1] = m$.

Dann gilt $\frac{1}{n}\sum_{j=1}^n X_j \overset{f.s.}{\to} m$.

Beweis. Wir beweisen den Satz zuerst im Falle, dass $X_n \geq 0$ $\forall n \in \mathbb{N}$.

Schritt 1. Wir schneiden die X_j bei Höhe j ab, um später L^2-Abschätzungen nutzen zu können: Wir definieren

$$Y_n(\omega) := X_n(\omega)\mathbf{1}_{X_n(\omega)\leq n} = \begin{cases} X_n & \text{falls } X_n \leq n, \\ 0 & \text{sonst.} \end{cases}$$

Es gilt $0 \leq Y_n \leq n$, also $Y_n \in L^\infty(\Omega, \mathbb{P})$ $\forall n \in \mathbb{N}$. Da $Y_n = f_n(X_n)$ für $f_n(x) = x\mathbf{1}_{x\leq n}$, sind $(Y_n)_{n\in\mathbb{N}}$ paarweise unabhängige Zufallsvariablen (siehe Bemerkung 1 nach Lemma 2.8). Sie sind allerdings nicht identisch verteilt. Es gilt

$$\begin{aligned}\mathbb{E}[X_n] &= \mathbb{E}[X_1] = m \quad \forall n \\ \mathbb{E}[Y_n] &= \mathbb{E}[X_n\mathbf{1}_{X_n\leq n}] = \int_{\mathbb{R}} x\mathbf{1}_{(-\infty,n]}(x)\ d\mathbb{P}_{X_n} \\ &\overset{X_n \text{ i.v.}}{=} \int_{\mathbb{R}} x\mathbf{1}_{(-\infty,n]}\ d\mathbb{P}_{X_1} = \mathbb{E}[X_1\mathbf{1}_{X_1\leq n}] \leq \mathbb{E}[X_1] = m.\end{aligned}$$

Wir zeigen

$$\mathbb{P}\left(\limsup_{n\to\infty}\{X_n \neq Y_n\}\right) = 0. \tag{5.2}$$

Beweis von (5.2)*:* Es gilt $X_j \neq Y_j$ genau dann, wenn $X_j > j$ und daher

$$\sum_{j=1}^{\infty} \mathbb{P}(X_j \neq Y_j) = \sum_{j=1}^{\infty} \mathbb{P}(X_j > j)$$

Weiterhin ist die Funktion $x \mapsto \mathbb{P}(X_1 > x)$ monoton fallend. Wir rechnen

$$\begin{aligned}
\sum_{j=1}^{\infty} \mathbb{P}(X_1 > j) &= \sum_{j=1}^{\infty} \int_{j-1}^{j} \mathbb{P}(X_1 > j)\ d\mathcal{L}(x) \leq \sum_{j=1}^{\infty} \int_{j-1}^{j} \mathbb{P}(X_1 > x)\ d\mathcal{L}(x) \\
&= \int_0^{\infty} \mathbb{P}(X_1 > x)\ d\mathcal{L}(x) = \int_{[0,\infty)} \int_\Omega \underbrace{\mathbf{1}_{(x,\infty)}(X_1(\omega))}_{=\mathbf{1}_{[0,X_1(\omega))}(x)}\ d\mathbb{P}(\omega)\ d\mathcal{L}(x) \\
&= \int_\Omega \left(\int_{[0,\infty)} \mathbf{1}_{[0,X_1(\omega))}(x)\ d\mathcal{L}(x) \right)\ d\mathbb{P}(\omega) \\
&= \int_\Omega X_1(\omega)\ d\mathbb{P}(\omega) = \mathbb{E}[X_1] < \infty.
\end{aligned}$$

Dabei haben wir im vorletzten Schritt Fubini verwendet. Nach Borel-Cantelli gilt also (5.2). □

Es gibt also $N_1 \in \mathcal{F}$ mit $\mathbb{P}(N_1) = 0$, sodass $\forall \omega \notin N_1\, \exists n_\omega \in \mathbb{N}$, sodass $X_n(\omega) = Y_n(\omega)\ \forall n \geq n_\omega$. Wir folgern jetzt

$$\frac{1}{n} \sum_{j=1}^{n} X_j \overset{f.s.}{\to} m \iff \frac{1}{n} \sum_{j=1}^{n} Y_j \overset{f.s.}{\to} m. \tag{5.3}$$

Beweis von (5.3)*:* Es gilt für $\omega \notin N_1$ und $n > n_\omega$

$$\frac{1}{n} \sum_{j=1}^{n} X_i(\omega) = \frac{1}{n} \sum_{i=1}^{n_\omega - 1} X_i(\omega) + \frac{1}{n} \sum_{i=n_\omega}^{n} \underbrace{X_i(\omega)}_{=Y_i(\omega)} = \frac{1}{n} \sum_{i=1}^{n} Y_i(\omega) + \frac{1}{n} \sum_{i=1}^{n_\omega - 1} [X_i(\omega) - Y_i(\omega)].$$

Es gilt für $\omega \notin N_1$ fest

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n_\omega - 1} [X_i(\omega) - Y_i(\omega)] = 0.$$

Somit folgt

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = m \iff \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Y_i(\omega) = m.$$

□

Schritt 2. Sei $S_n = \sum_{j=1}^n Y_j$. Nach Schritt 1 reicht es also zu zeigen, dass $\frac{S_n}{n} \overset{f.s.}{\to} m$. Wir zeigen zunächst

$$\lim_{n\to\infty} \frac{\mathbb{E}[S_n]}{n} = m. \tag{5.4}$$

Beweis von (5.4)*:* Sei $a_j = \mathbb{E}[Y_j]$. Dann gilt

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[Y_j] = \frac{1}{n}\sum_{j=1}^{n} a_j.$$

Es gilt also $0 \le a_j \le m \ \forall j$. Mit monotoner Konvergenz folgt auch

$$a_j = \mathbb{E}[X_1 \mathbf{1}_{X_1 \le j}] = \int_{[0,\infty)} x \mathbf{1}_{(-\infty,j]} \, d\mathbb{P}_{X_1}(x) \to \int_{\mathbb{R}} x \, d\mathbb{P}_{X_1} = \mathbb{E}[X_1] = m.$$

(5.4) folgt nun aus Analysis 1. $\square$

Als nächstes beweisen wir die fast sichere Konvergenz von bestimmten Teilfolgen: Sei $\alpha > 1$, $k_n = k_{n,\alpha} = \lfloor \alpha^n \rfloor$. Wir zeigen

$$\frac{S_{k_n} - \mathbb{E}[S_{k_n}]}{k_n} \overset{f.s.}{\to} 0. \tag{5.5}$$

Beweis von (5.5)*:* Nach Borel-Cantelli genügt es zu zeigen, dass

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{k_n} - \mathbb{E}[S_{k_n}]}{k_n}\right| > \varepsilon\right) < \infty \quad \forall \varepsilon > 0.$$

Es gilt nach Tschebishev und der paarweisen Unabhängigkeit der Y_j

$$\mathbb{P}\left(\left|\frac{S_{k_n} - \mathbb{E}[S_{k_n}]}{k_n}\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\operatorname{var}\left(\frac{S_{k_n}}{k_n}\right) = \frac{1}{\varepsilon^2 k_n^2}\operatorname{var}(S_{k_n}) = \frac{1}{\varepsilon^2 k_n^2}\sum_{j=1}^{k_n} \operatorname{var}(Y_j).$$

Es gilt

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{k_n} - \mathbb{E}[S_{k_n}]}{k_n}\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty} \frac{1}{k_n^2}\sum_{j=1}^{k_n} \operatorname{var}(Y_j) = \frac{1}{\varepsilon^2}\sum_{j=1}^{\infty} \operatorname{var}(Y_j) \sum_{\substack{n\in\mathbb{N} \\ k_n \ge j}}^{\infty} \frac{1}{k_n^2}.$$

Sei n_j so, dass $k_n \ge j \iff n \ge n_j$. Dann ist

$$\sum_{\substack{n\in\mathbb{N} \\ k_n \ge j}}^{\infty} \frac{1}{k_n^2} = \sum_{n \ge n_j}^{\infty} \frac{1}{\lfloor \alpha^n \rfloor^2} \approx \sum_{n \ge n_j}^{\infty} \frac{1}{\alpha^{2n}} = \frac{\alpha^{-2n_j}}{1 - \frac{1}{\alpha^2}} \le C_\alpha \frac{1}{\alpha^{2n_j}} \le C_\alpha \frac{1}{j^2},$$

wobei C_α von α abhängt. Es folgt

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{k_n} - \mathbb{E}[S_{k_n}]}{k_n}\right| > \varepsilon\right) \le \frac{C_\alpha}{\varepsilon^2}\sum_{j=1}^{\infty} \operatorname{var}(Y_j)\frac{1}{j^2}.$$

Zudem ist

$$\begin{aligned}
0 \le \operatorname{var}(Y_j) &= \mathbb{E}[Y_j^2] - \mathbb{E}[Y_j]^2 \le \mathbb{E}[Y_j]^2 = \mathbb{E}[X_j^2 \mathbf{1}_{X_j \le j}] = \mathbb{E}[X_1^2 \mathbf{1}_{X_1 \le j}] \\
&= \sum_{l=1}^{j} \mathbb{E}[X_1^2 \mathbf{1}_{l-1 < X_1 \le l}] \le \sum_{l=1}^{j} l^2 \mathbb{P}(X_1 \in (l-1, l]).
\end{aligned}$$

Es gilt also

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{S_{k_n}-\mathbb{E}[S_{k_n}]}{k_n}\right|>\varepsilon\right) &\leq \frac{C_\alpha}{\varepsilon^2}\sum_{j=1}^{\infty}\mathrm{var}(Y_j)\frac{1}{j^2} \leq \frac{C_\alpha}{\varepsilon^2}\sum_{j=1}^{\infty}\sum_{l=1}^{j}\frac{1}{j^2}l^2\mathbb{P}(X_1\in(l-1,l]) \\ &= \frac{C_\alpha}{\varepsilon^2}\sum_{l=1}^{\infty}l^2\mathbb{P}(X_1\in(l-1,l])\underbrace{\sum_{j=l}^{\infty}\frac{1}{j^2}}_{\leq\frac{C}{l}} \\ &\leq \frac{C_\alpha}{\varepsilon^2}\sum_{l=1}^{\infty}l\,\mathbb{P}(X_1\in(l-1,l]) = \frac{C_\alpha}{\varepsilon^2}\sum_{l=1}^{\infty}\mathbb{E}[l\ \mathbf{1}_{l-1<X_1\leq l}] \\ &\leq \frac{C_\alpha}{\varepsilon^2}\sum_{l=1}^{\infty}\mathbb{E}[(X_1+1)\mathbf{1}_{l-1<X_1\leq l}] = \frac{C_\alpha}{\varepsilon^2}\mathbb{E}[X_1+1]<\infty.\end{aligned}$$

Für alle $\alpha>1$ existiert also $N_\alpha\in\mathcal{F}$, sodass $\lim_{n\to\infty}\frac{S_{k_{n,\alpha}}(\omega)-\mathbb{E}[S_{k_{n,\alpha}}]}{k_{n,\alpha}}=0\ \forall\omega\notin N_\alpha$, was (5.5) ist. □

Aus (5.4) und (5.5) folgt, dass für alle $\alpha>1$ ein $\tilde{N}_\alpha\in\mathcal{F}$ existiert, sodass

$$\lim_{n\to\infty}\frac{S_{k_{n,\alpha}}(\omega)}{k_{n,\alpha}}=\lim_{n\to\infty}\frac{S_{k_{n,\alpha}}(\omega)-\mathbb{E}[S_{k_{n,\alpha}}]}{k_{n,\alpha}}+\lim_{n\to\infty}\frac{\mathbb{E}[S_{k_{n,\alpha}}]}{k_{n,\alpha}}=m \qquad \forall\omega\notin\tilde{N}_\alpha. \tag{5.6}$$

Wir können nun $\frac{S_n}{n}\overset{f.s.}{\to}m$ beweisen. Sei $\alpha_q=1+\frac{1}{q}$ für $q\in\mathbb{N}$ und $N=\bigcup_{q\in\mathbb{N}}\tilde{N}_{\alpha_q}$. Dann ist $\mathbb{P}(N)=0$ und $\forall\omega\notin N,\ \forall q\in\mathbb{N}$ gilt

$$\lim_{n\to\infty}\frac{S_{k_{n,\alpha_q}}(\omega)}{k_{n,\alpha_q}}=m. \tag{5.7}$$

Sei $\alpha=\alpha_q$ fest. Für $l\in\mathbb{N}$ gibt es genau ein n, sodass $k_n\leq l<k_{n+1}$. Da $Y_j\geq 0$, gilt

$$\sum_{j=1}^{k_n}Y_j\leq\sum_{j=1}^{l}Y_j\leq\sum_{j=1}^{k_{n+1}}Y_j.$$

Da $k_n\leq l\implies\frac{1}{l}\leq\frac{1}{k_n}$ und $l<k_{n+1}\implies\frac{1}{l}>\frac{1}{k_{n+1}}$, folgt

$$\frac{k_n}{k_{n+1}}\frac{S_{k_n}}{k_n}=\frac{S_{k_n}}{k_{n+1}}\leq\frac{S_{k_n}}{l}\leq\frac{S_l}{l}\leq\frac{S_{k_{n+1}}}{l}\leq\frac{S_{k_{n+1}}}{k_n}\leq\frac{S_{k_{n+1}}}{k_{n+1}}\frac{k_{n+1}}{k_n}.$$

Es gilt mit (5.7)

$$\lim_{n\to\infty}\frac{k_n}{k_{n+1}}=\frac{1}{\alpha}\quad\text{und}\quad\lim_{n\to\infty}\frac{S_{k_n}}{k_n}=m\quad\forall\alpha_q=1+\frac{1}{q},\omega\notin N.$$

Wir folgern

$$\frac{m}{\alpha_q}\leq\liminf_{l\to\infty}\frac{S_l}{l}\leq\limsup_{l\to\infty}\frac{S_l}{l}\leq m\alpha_q\quad\forall q\in\mathbb{N}.$$

Da $\lim_{q\to\infty}\alpha_q=1$, folgt

$$\lim_{l\to\infty}\frac{S_l}{l}=m\quad\forall\omega\notin N.$$

Das beweist Satz 5.6 im Falle, dass $X_n \geq 0$ für alle $n \in \mathbb{N}$. Es verbleibt, den Fall $X_n \in \mathbb{R}$ zu betrachten. Wir zerlegen $X_n = X_{n+} - X_{n-}$, $X_{n\pm} \geq 0$. Da X_n paarweise unabhängig und identisch verteilt sind, folgt, dass X_{n+} paarweise unabhängig und identisch verteilt sind. Aus dem ersten Fall folgt, dass $\exists N_+, N_- \in \mathcal{F}$ mit $\mathbb{P}(N_\pm) = 0$, sodass

$$\frac{1}{n}\sum_{j=1}^{n} X_{j\pm}(\omega) \to m_\pm = \mathbb{E}[X_{1\pm}] \quad \forall \omega \notin N_\pm.$$

Es folgt $\forall \omega \notin (N_+ \cup N_-)$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\frac{\sum_{i=1}^{n} X_{i+}}{n} - \frac{1}{n}\frac{\sum_{i=1}^{n} X_{i+}}{n} \to m_+ - m_- = \mathbb{E}[X_1] = m.$$

□

Bemerkung: Die *paarweise Unabhängigkeit* der X_n wird benötigt, um zu zeigen, dass die Y_n immer noch unabhängig sind, was benötigt wird, um $\operatorname{var}(\sum_{i=1}^n Y_i) = \sum_{i=1}^n \operatorname{var}(Y_i)$ zu folgern. Sind die X_n nur paarweise unkorreliert, so folgt nicht, dass die Y_n ebenfalls unkorreliert sind, auch im Fall wenn $\mathbb{E}[X_i X_j]$ wohldefiniert ist.

Satz 5.7 (Starkes Gesetz der großen Zahlen im Falle ohne Integrierbarkeit)**.** *Seien $(X_n)_{n\in\mathbb{N}}$ Zufallsvariablen, sodass $X_n \geq 0$ $\forall n \in \mathbb{N}$ und (X_n) sind unabhängig und identisch verteilt. Dann gilt*

$$\frac{1}{n}\sum_{j=1}^{n} X_j \overset{f.s.}{\to} \mathbb{E}[X_1].$$

Beweis. Es gilt $0 \leq \mathbb{E}[X_1]$, da $X_1 \geq 0$. Falls $\mathbb{E}[X_1] < \infty$, können wir Satz 5.6 anwenden.
Sei $\mathbb{E}[X_1] = \infty$. Für $k \in \mathbb{N}$ sei $Z_j := \min(X_j, k)$. Es gilt $0 \leq Z_j \leq k$, und daher $Z_j \in L^\infty$. Die Zufallsvariablen Z_j sind auch paarweise unabhängig und identisch verteilt. Nach Satz 5.6 gilt also, für fast alle $\omega \in \Omega$

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j \geq \liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \underbrace{\min(X_j, k)}_{=Z_j} = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} Z_j = \mathbb{E}[Z_1].$$

Es existieren also $N_k \in \mathcal{F}$ mit $\mathbb{P}(N_k) = 0$, sodass

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j(\omega) \geq \mathbb{E}[Z_1] \quad \forall \omega \notin N_k.$$

Es gilt

$$\mathbb{E}[Z_1] = \mathbb{E}[X_1 \mathbf{1}_{X_1 \leq k}] + k\mathbb{P}(X_1 > k) \geq \mathbb{E}[X_1 \mathbf{1}_{X_1 \leq k}] \overset{k\to\infty}{\to} \infty$$

nach monotoner Konvergenz. Es folgt also $\forall \omega \notin \bigcup_{k\in\mathbb{N}} N_k$, dass

$$\frac{1}{n}\sum_{j=1}^{\infty} X_j(\omega) \to \infty = \mathbb{E}[X_1].$$

□

[18: 12.12.2025]
[19: 16.12.2025]

6 Charakteristische Funktion ϕ_X

6.1 Definition und Eigenschaften

Definition 6.1 (Charakteristische Funktion)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $X\colon \Omega \to \mathbb{R}$ *eine Zufallsvariable.*
Die Funktion

$$\begin{aligned} \phi_X\colon \quad \mathbb{R} &\to \quad \mathbb{C} \\ t &\mapsto \quad \phi_X(t) := \mathbb{E}\left[e^{itX}\right] := \mathbb{E}\left[\cos(tX)\right] + i\mathbb{E}\left[\sin(tX)\right] \end{aligned}$$

heißt charakteristische Funktion von X.

Bemerkung. Aus $|\cos(tX)| \leq 1$ und $|\sin(tX)| \leq 1$ folgt, dass das komplexwertige Integral $\mathbb{E}\left[e^{itX}\right]$ wohldefiniert ist. Außerdem gilt, mit der Cauchy-Schwarz Ungleichung,

$$|\phi_X| = \sqrt{\mathbb{E}\left[\cos(tX)\right]^2 + \mathbb{E}\left[\sin(tX)\right]^2} \leq \sqrt{\mathbb{E}\left[\cos(tX)^2 + \sin(tX)^2\right]} = \sqrt{\mathbb{E}\left[1\right]} = 1.$$

Falls $\mathbb{P}_X = \rho_X d\mathcal{L}$, dann ist ϕ_X die Fourier Transformierte der Funktion ρ_X bis auf eine multiplikative Konstante (siehe unten).

Definition 6.2 (Fourier Transformation)**.** *Sei* $f \in L^1(\mathbb{R}, \mathcal{L})$*. Die Fourier Transformierte von* f *ist für* $k \in \mathbb{R}$ *definiert durch*

$$\hat{f}(k) = \mathcal{F}(f)(k) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ikx} f(x)\, d\mathcal{L}(x).$$

Wichtige Eigenschaften (siehe Analysis 3):

(i) Die Funktion $k \mapsto \hat{f}(k)$ ist stetig und beschränkt: $|\hat{f}(k)| \leq \frac{1}{\sqrt{2\pi}} \|f\|_{L^1}$;

(ii) Falls $f, \hat{f} \in L^1(\mathbb{R}, \mathcal{L})$, dann gilt

$$f(x) = \mathcal{F}^{-1}(\hat{f})(x) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ikx} \hat{f}(k)\, d\mathcal{L}(k) = \mathcal{F}(\hat{f})(-x);$$

(iii) Falls $x \mapsto |x\, f(x)| \in L^1(\mathbb{R}, \mathcal{L})$, dann gilt $\hat{f} \in C^1(\mathbb{R}; \mathbb{C})$ und

$$\hat{f}'(k) = \mathcal{F}(f)'(k) = \mathcal{F}(ix\, f(x))(k);$$

(iv) Falls $f \in L^1(\mathbb{R}, \mathcal{L}) \cap C^1(\mathbb{R}; \mathbb{C})$ und $f' \in L^1(\mathbb{R}, \mathcal{L})$, dann gilt

$$\mathcal{F}(f')(k) = -ik\mathcal{F}(f)(k).$$

Lemma 6.3 (Fourier Transformierte der Gauß Funktion)**.** *Für* $\sigma > 0$ *sei* $p_\sigma\colon \mathbb{R} \to [0, \infty)$ *definiert durch* $p_\sigma(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$*. Folgendes gilt:*

(i) $x \mapsto |x|^n p_\sigma(x) \in L^q(\mathbb{R}, \mathcal{L})\ \forall q \geq 1, n \in \mathbb{N}_0$.

(ii) $\hat{p}_\sigma(k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sigma^2 x^2} = \frac{1}{\sigma}\, p_{\frac{1}{\sigma}}(k)\ \forall k \in \mathbb{R}$.

(iii) Falls $X \sim \mathcal{N}(0, \sigma^2)$ dann gilt $\phi_X(t) = e^{-\frac{1}{2}\sigma^2 t^2}$.

Beweis. Da p_σ die Dichte für eine Zufallsvariable $X \sim \mathcal{N}(0, \sigma^2)$ ist, gilt

$$\int_{\mathbb{R}} p_\sigma \, d\mathcal{L} = 1 \qquad \forall \sigma > 0.$$

(i) Wir rechnen für $y > 0$, $e^y = \sum_{j \geq 0} \frac{y^j}{j!} \geq \frac{y^{j_0}}{j_0!}$ $\forall j_0 \geq 0$ und daher

$$\begin{aligned} x^{2n} p_\sigma(x) &= \left[\frac{1}{n!}\left(\frac{x}{2\sigma}\right)^{2n} p_\sigma(x)\right] n! \, (2\sigma)^{2n} \leq \left[e^{\frac{x^2}{4\sigma^2}} p_\sigma(x)\right] n! \, (2\sigma)^{2n} \\ &= \left[\sqrt{2} \, p_{\sigma\sqrt{2}}(x)\right] n! \, (2\sigma)^{2n}. \end{aligned}$$

Mit der Cauchy-Schwarz Ungleichung erhalten wir

$$\begin{aligned} &\left(\int_{\mathbb{R}} |x^n p_\sigma(x)| \, d\mathcal{L}(x)\right)^2 \leq \int_{\mathbb{R}} x^{2n} p_\sigma(x) \, d\mathcal{L}(x) \int_{\mathbb{R}} p_\sigma(x) \, d\mathcal{L}(x) \\ &= \int_{\mathbb{R}} x^{2n} p_\sigma(x) \, d\mathcal{L}(x) \leq \sqrt{2} \, n! \, (2\sigma)^{2n} \int_{\mathbb{R}} p_{\sigma\sqrt{2}} d\mathcal{L} = \sqrt{2} \, n! \, (2\sigma)^{2n} < \infty. \end{aligned}$$

Damit gilt $x^n p_\sigma(x) \in L^1$. Die Aussage $x^n p_\sigma(x) \in L^q$ wird analog bewiesen.

(ii) Wir zeigen, dass $\hat{p}_\sigma$ eine ODG erfüllt.
Aus (i) folgt, dass $|x| e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} \in L^1$ gilt und daher $\hat{p}_\sigma$ is stetig differenzierbar mit

$$\hat{p}_\sigma'(k) = \mathcal{F}(ix \, p_\sigma(x))(k).$$

Wir rechnen $x \, p_\sigma(x) = -\sigma^2 p_\sigma'(x)$. Aus $p_\sigma' \in L^1$ folgt, dass

$$\mathcal{F}(p_\sigma')(k) = -ik \, \hat{p}_\sigma(k).$$

Wir erhalten also

$$\hat{p}_\sigma'(k) = \mathcal{F}(ix \, p_\sigma(x))(k) = -i\sigma^2 \, \mathcal{F}(p_\sigma')(k) = -\sigma^2 k \, \hat{p}_\sigma(k).$$

Die Funktion $\hat{p}_\sigma$ muss also folgende Form haben

$$\hat{p}_\sigma(k) = C \, e^{-\frac{1}{2}\sigma^2 k^2}.$$

Um die Konstante C zu bestimmen, rechnen wir

$$C = \hat{p}_\sigma(0) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} p_\sigma(x) \, d\mathcal{L}(x) = \frac{1}{\sqrt{2\pi}}.$$

(iii) Es gilt

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x) = \int_{\mathbb{R}} e^{itx} p_\sigma(x) \, d\mathcal{L}(x) = \sqrt{2\pi} \, \hat{p}_\sigma(t).$$

□

Wir können p_σ anwenden, um andere Verteilungen zu studieren.

Satz 6.4. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X, Y_\sigma \colon \Omega \to \mathbb{R}$ zwei unabhängige Zufallsvariablen, mit $Y_\sigma \sim \mathcal{N}(0, \sigma^2)$, $\sigma > 0$. Sei $\mathbb{P}_\sigma := \mathbb{P}_{X+Y_\sigma} = \mathbb{P}_{Y_\sigma} * \mathbb{P}_X$. Dann gilt Folgendes:*

(i) $\mathbb{P}_\sigma = \rho_\sigma \mathcal{L}$ *mit*

$$\rho_\sigma(z) \coloneqq \int_{\mathbb{R}} p_\sigma(z-x)\, d\mathbb{P}_X(x).$$

(ii) Die Dichte ρ_σ *ist von* ϕ_X *bestimmt wie folgt*

$$\rho_\sigma(z) = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-izk} p_{\frac{1}{\sigma}}(k)\, \phi_X(k)\ d\mathcal{L}(k) = \frac{1}{\sigma}\mathcal{F}\left(p_{\frac{1}{\sigma}}\, \phi_X\right)(-z).$$

(iii) Wenn $\sigma \to 0$, *gilt* $\mathbb{P}_\sigma \overset{\text{schwach}}{\to} \mathbb{P}_X$.

Beweis. Später. □

Korollar 6.5. *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $X\colon \Omega \to \mathbb{R}$ *eine Zufallsvariable. Die charakteristiche Funktion* ϕ_X *bestimmt die Verteilung von* X *eindeutig.*

Beweis. Sei $t \in \mathbb{R}$. Dann für alle $\varepsilon > 0$ gibt es eine Funktion $g_\varepsilon \in C_b(\mathbb{R})$, sodass $g_\varepsilon(x) = 1$ für $x \le t$, $g_\varepsilon(x) = 0$ für $x > t + \varepsilon$ und $|g_\varepsilon(x)| \le 1\ \forall x \in \mathbb{R}$. Dann, via dominierte Konvergenz, gilt

$$F_X(t) = \mathbb{P}_X((-\infty, t]) = \lim_{\varepsilon \to 0+} \int_{\mathbb{R}} g_\varepsilon\ d\mathbb{P}_X.$$

Um die Verteilung von X zu bestimmen, genügt es also, $\mathbb{E}[g(X)]$ für alle $g \in C_b(\mathbb{R})$ zu kennen. Aus dem Satz 6.4 folgt, für alle $g \in C_b(\mathbb{R})$,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g\ d\mathbb{P}_X = \lim_{\sigma\to 0} \int_{\mathbb{R}} g\ \rho_\sigma\, d\mathcal{L},$$

wobei $\rho_\sigma(z) = \frac{1}{\sigma}\mathcal{F}\left(p_{\frac{1}{\sigma}}\, \phi_X\right)(-z)$ und deshalb eindeutig durch ϕ_X bestimmt wird. □

Bemerkung. Die momenterzeugende Funktion $M_X(t) \coloneqq \mathbb{E}[e^{tX}]$ bestimmt, im Allgemeinen, die Verteilung von X nicht!

Beweis von Satz 6.4.
(i) Für jede $A \in \mathcal{B}(\mathbb{R})$ gilt

$$\begin{aligned}
\mathbb{P}_{X+Y_\sigma}(A) &= \int_{\mathbb{R}^2} \mathbf{1}_A(x+y) d\mathbb{P}_{Y_\sigma}(y) d\mathbb{P}_X(x) = \int_{\mathbb{R}^2} \mathbf{1}_A(x+y) p_\sigma(y)\ d\mathcal{L}(y)\, d\mathbb{P}_X(x) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(x+y) p_\sigma(y)\ d\mathcal{L}(y) \right) d\mathbb{P}_X(x) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(z) p_\sigma(z-x)\ d\mathcal{L}(z) \right) d\mathbb{P}_X(x) \\
&= \int_{\mathbb{R}} \mathbf{1}_A(z) \left(\int_{\mathbb{R}} p_\sigma(z-x)\ d\mathbb{P}_X(x) \right) d\mathcal{L}(z) = \int_{\mathbb{R}} \mathbf{1}_A(z) \rho_\sigma(z)\, d\mathcal{L}(z).
\end{aligned}$$

Hier konnten wir die Integrationsordnung vertauchen, da alle Funktionen nicht-negativ sind.

(*ii*) Wir rechnen, mit Lemma 6.3,

$$\begin{aligned}
\rho_\sigma(z) &= \int_{\mathbb{R}} p_\sigma(z-x)\ d\mathbb{P}_X(x) = \int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ik(z-x)} \hat{p}_\sigma(k)\ d\mathcal{L}(k) \right) d\mathbb{P}_X(x) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-ik(z-x)} p_{\frac{1}{\sigma}}(k)\ d\mathcal{L}(k) \right) d\mathbb{P}_X(x) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{-ikz} p_{\frac{1}{\sigma}}(k) \left(\int_{\mathbb{R}} e^{ikx}\ d\mathbb{P}_X(x) \right) d\mathcal{L}(k) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{-ikz} p_{\frac{1}{\sigma}}(k) \phi_X(k)\, d\mathcal{L}(k).
\end{aligned}$$

Hier konnten wir die Integrationsordnung vertauchen, da

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left| e^{-ik(z-x)} p_{\frac{1}{\sigma}}(k) \right|\ d\mathcal{L}(k) \right) d\mathbb{P}_X(x) = \int_{\mathbb{R}} p_{\frac{1}{\sigma}}(k) d\mathcal{L}(k)\ \int_{\mathbb{R}} d\mathbb{P}_X(x) = 1 < \infty.$$

(*iii*) Sei $g \in C_b(\mathbb{R})$. Unser Ziel ist zu zeigen, dass $\lim_{\sigma\to 0} \int_{\mathbb{R}} g\, d\mathbb{P}_\sigma = \int_{\mathbb{R}} g\, d\mathbb{P}_X$.
Wir schreiben

$$\begin{aligned}
\int_{\mathbb{R}} g\, d\mathbb{P}_\sigma &= \int_{\mathbb{R}} g(z)\, \rho_\sigma(z)\, d\mathcal{L}(z) = \int_{\mathbb{R}} g(z) \left(\int_{\mathbb{R}} p_\sigma(z-x) d\mathbb{P}_X(x) \right) d\mathcal{L}(z) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(z) p_\sigma(z-x)\, d\mathcal{L}(z) \right) d\mathbb{P}_X(x) = \int_{\mathbb{R}} g_\sigma(x)\, d\mathbb{P}_X(x)
\end{aligned}$$

wobei, mit $p_\sigma(x-z) = p_\sigma(z-x)$,

$$g_\sigma(x) \coloneqq p_\sigma * \rho_\sigma(x) = \int_{\mathbb{R}} g(z) p_\sigma(z-x)\, d\mathcal{L}(z) = \int_{\mathbb{R}} p_\sigma(x-z) g(z)\, d\mathcal{L}(z).$$

Oben konnten wir die Integrationsordnung vertauchen, da

$$\begin{aligned}
&\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |g(z)| p_\sigma(z-x)\, d\mathcal{L}(z) \right) d\mathbb{P}_X(x) \le \|g\|_{L^\infty} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} p_\sigma(z-x)\, d\mathcal{L}(z) \right) d\mathbb{P}_X(x) \\
&= \|g\|_{L^\infty} \int_{\mathbb{R}} p_\sigma(z)\, d\mathcal{L}(z)\ \int_{\mathbb{R}} d\mathbb{P}_X(x) = \|g\|_{L^\infty} < \infty.
\end{aligned}$$

Behauptung. Es gilt

(i) $|g_\sigma(x)| \le \|g\|_{L^\infty} \in L^1(\mathbb{R}, \mathbb{P}_X)\ \forall x \in \mathbb{R}$,

(ii) $\lim_{\sigma\to 0} g_\sigma(x) = g(x)\ \forall x \in \mathbb{R}$.

Aus der Behauptung folgt, mit dominierter Konvergenz,

$$\lim_{\sigma\to 0} \int_{\mathbb{R}} g_\sigma\ d\mathbb{P}_X = \int_{\mathbb{R}} g\ d\mathbb{P}_X$$

und damit ist der Beweis des Satzes beendet.

[19: 16.12.2025]

[20: 19.12.2025]

Beweis der Behauptung.

(i) Wir rechnen

$$|g_\sigma(x)| \leq \int_{\mathbb{R}} p_\sigma(x-z)|g(z)|\, d\mathcal{L}(z) \leq \|g\|_{L^\infty} \int_{\mathbb{R}} p_\sigma(x-z)\; d\mathcal{L}(z) = \|g\|_{L^\infty}.$$

(ii) Da $\int_{\mathbb{R}} p_\sigma(z) d\mathcal{L}(z) = 1$ gilt, können wir schreiben

$$\begin{aligned} g_\sigma(x) - g(x) &= \int_{\mathbb{R}} p_\sigma(x-z)[g(z)-g(x)]\, d\mathcal{L}(z) = \int_{\mathbb{R}} p_\sigma(z)[g(x+z)-g(x)]\, d\mathcal{L}(z) \\ &= \int_{|z|\leq M} p_\sigma(z)[g(x+z)-g(x)]\, d\mathcal{L}(z) + \int_{|z|>M} p_\sigma(z)[g(x+z)-g(x)]\, d\mathcal{L}(z). \end{aligned}$$

Es gilt

$$\begin{aligned} &\left| \int_{|z|\leq M} p_\sigma(z)[g(x+z)-g(x)]\, d\mathcal{L}(z) \right| \leq \int_{|z|\leq M} p_\sigma(z)\; |g(x+z)-g(x)|\; d\mathcal{L}(z) \\ &\leq \max_{|z|\leq M} |g(x+z)-g(x)| \int_{|z|\geq M} p_\sigma(z) d\mathcal{L}(z) \\ &\leq \max_{|z|\leq M} |g(x+z)-g(x)| \int_{\mathbb{R}} p_\sigma(z) d\mathcal{L}(z) = \max_{|z|\geq M} |g(x+z)-g(x)|. \end{aligned}$$

Weiter gilt

$$\begin{aligned} &\left| \int_{|z|>M} p_\sigma(z)[g(x+z)-g(x)]\, d\mathcal{L}(z) \right| \leq \int_{|z|>M} p_\sigma(z)\; |g(x+z)-g(x)|\; d\mathcal{L}(z) \\ &\quad \leq 2\|g\|_{L^\infty} \int_{|z|>M} p_\sigma(z)\; d\mathcal{L}(z). \end{aligned}$$

Wir rechnen, für $|z| > M$,

$$p_\sigma(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{z^2}{\sigma^2}} \leq e^{-\frac{1}{4}\frac{M^2}{\sigma^2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{4}\frac{z^2}{\sigma^2}} = e^{-\frac{1}{4}\frac{M^2}{\sigma^2}} \sqrt{2}\; p_{\sigma\sqrt{2}}(z),$$

und daher

$$\int_{|z|>M} p_\sigma(z)\, d\mathcal{L}(z) \leq e^{-\frac{1}{4}\frac{M^2}{\sigma^2}} \sqrt{2} \int_{|z|>M} p_{\sigma\sqrt{2}}(z)\, d\mathcal{L}(z) \leq e^{-\frac{1}{4}\frac{M^2}{\sigma^2}} \sqrt{2} \int_{\mathbb{R}} p_{\sigma\sqrt{2}}(z)\, d\mathcal{L}(z) = e^{-\frac{1}{4}\frac{M^2}{\sigma^2}} \sqrt{2}.$$

Also gilt, für alle $M, \sigma > 0$

$$|g_\sigma(x) - g(x)| \leq \max_{|z|\leq M} |g(x+z)-g(x)| \;+\; 2^{\frac{3}{2}} \|g\|_{L^\infty} e^{-\frac{1}{4}\frac{M^2}{\sigma^2}}.$$

Wir setzen $M = M_\sigma \coloneqq \sqrt{\sigma}$. Dann gilt

$$\lim_{\sigma\to 0} M_\sigma = 0, \qquad \lim_{\sigma\to 0} \frac{M_\sigma}{\sigma} = \infty.$$

Die Aussage folgt jetzt aus der Stetigkeit von g. □

Lemma 6.6 (Eigenschaften von ϕ_X)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum.*

(i) Für alle Zufallsvariablen $X\colon \Omega \to \mathbb{R}$ mit $a \neq 0$ und $b \in \mathbb{R}$ gilt

$$\phi_{aX+b}(t) = e^{ibt}\phi_X(at).$$

(ii) Seien $X_1, \ldots, X_n\colon \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen und $S_n := \sum_{j=1}^n X_j$, so gilt

$$\phi_{S_n}(t) = \prod_{j=1}^n \phi_{X_j}(t).$$

Falls die Zufallsvariablen auch identisch verteilt sind, gilt

$$\phi_{S_n}(t) = \phi_{X_1}(t)^n.$$

Beweis.
(i) Wir rechnen $\phi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = e^{itb}\mathbb{E}[e^{i(ta)X}] = e^{itb}\phi_X(at)$.
(ii) Wir rechnen $\mathbb{E}[e^{itS_n}] = \mathbb{E}[\prod_{j=1}^n e^{itX_j}] = \prod_{j=1}^n \mathbb{E}[e^{itX_j}]$, wo der letzter Schritt wegen Unabhängigkeit gilt. Falls die Zufallsvariablen auch identisch verteilt sind, gilt $\mathbb{E}[e^{itX_j}] = \mathbb{E}[e^{itX_1}]$ $\forall j = 2, \ldots, n$. □

Beispiele:

- Falls $X \sim Bin(n,p)$, dann gilt $\phi_X(t) = (1-p+pe^{it})^n$:
$$\phi_X(t) = \sum_{k=1}^n \binom{n}{k} p^k(1-p)^{n-k}e^{ikt} = \sum_{k=1}^n \binom{n}{k}(1-p)^{n-k}(pe^{it})^k = (1-p+pe^{it})^n.$$
- Falls $X \sim \mathcal{N}(\mu,\sigma^2)$, dann gilt $\phi_X(t) = e^{i\mu t}e^{-\frac{1}{2}\sigma^2t^2}$:
$$\phi_X(t) = \phi_{(X-\mu)+\mu}(t) = e^{i\mu t}\phi_{X-\mu}(t).$$
Da $X-\mu \sim \mathcal{N}(0,\sigma^2)$, die Aussage folgt aus Lemma 6.3.
- Falls $X \sim Cauchy(a)$, dann gilt $\phi_X(t) = e^{-a|t|}$ (Übungsblatt).

6.2 ϕ_X erzeugt die Momente

Lemma 6.7. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X\colon \Omega \to \mathbb{R}$ eine Zufallsvariable. Die charakteristische Funktion ϕ_X ist gleichmäßig stetig auf $\mathbb{R}$.*

Beweis. Sei $\varepsilon > 0$. Wir suchen $\delta = \delta_\varepsilon > 0$, sodass

$$|t-s| \leq \delta \quad \Rightarrow \quad |\phi_X(t) - \phi_X(s)| \leq \varepsilon.$$

Wir nehmen an, dass $|t-s| \leq \delta$, wobei $\delta > 0$ später bestimmt wird, und rechnen

$$\begin{aligned}|\phi_X(t) - \phi_X(s)| &= \left|\int_{\mathbb{R}} (e^{itx} - e^{isx})\, d\mathbb{P}_X(x)\right| \leq \int_{\mathbb{R}} |(e^{itx} - e^{isx})|\, d\mathbb{P}_X(x)\\ &= \int_{\mathbb{R}} |(1-e^{i(s-t)x})|\, d\mathbb{P}_X(x) = \int_{\mathbb{R}} \sqrt{2(1-\cos(s-t)x)}\, d\mathbb{P}_X(x)\\ &= \int_{|x|\leq M} \sqrt{2(1-\cos(s-t)x)}\, d\mathbb{P}_X(x) + \int_{|x|>M} \sqrt{2(1-\cos(s-t)x)}\, d\mathbb{P}_X(x).\end{aligned}$$

Aus $1-\cos\mu \leq \mu^2$ folgt

$$\int_{|x|\leq M} \sqrt{2(1-\cos(s-t)x)}\, d\mathbb{P}_X(x) \leq \max_{|x|\leq M} \sqrt{2(1-\cos(s-t)x)} \leq \sqrt{2}|(t-s)|M \leq \sqrt{2}\ \delta M.$$

Weiter gilt, mit $1-\cos\mu \leq 2$,

$$\int_{|x|>M} \sqrt{2(1-\cos(s-t)x)}\, d\mathbb{P}_X(x) \leq 2\ \mathbb{P}(|X|>M).$$

Aus $\lim_{M\to\infty} \mathbb{P}(|X|>M) = 0$ folgt, dass ein $M = M_\varepsilon > 0$ existiert, sodass

$$2\ \mathbb{P}(|X|>M_\varepsilon) \leq \frac{\varepsilon}{2}.$$

Damit gilt

$$|\phi_X(t)-\phi_X(s)| \leq \sqrt{2}\ \delta M_\varepsilon + \frac{\varepsilon}{2}.$$

Sei jetzt $\delta_\varepsilon := \frac{\epsilon}{2^{\frac{3}{2}} M_\varepsilon}$, dann gilt $|\phi_X(t)-\phi_X(s)| \leq \varepsilon$.

□

Satz 6.8. *Sei $(\Omega,\mathcal{F},\mathbb{P})$ ein Wahrscheinlichkeitsraum und $X\colon \Omega\to\mathbb{R}$ eine Zufallsvariable, mit $X \in L^m(\Omega,\mathbb{P})$, $m\in\mathbb{N}$.*

Dann gilt $\phi_X \in C^m(\mathbb{R};\mathbb{C})$ und

$$\phi_X^{(k)}(t) = i^k \mathbb{E}[X^k e^{itX}] \quad \forall k = 1,\dots,n,\ \forall t\in\mathbb{R}.$$

Insbesondere gilt $\mathbb{E}[X^k] = (-i)^k \phi_X^{(k)}(0)$.

Beweis. Wir argumentieren duch Induktion in n.

Im Fall $n=1$ rechnen wir

$$\begin{aligned}\phi_X(t) = \mathbb{E}[e^{itX}] &= \mathbb{E}\left[1+\int_0^t (iX)e^{isX} d\mathcal{L}(s)\right]\\ &= 1+\mathbb{E}\left[\int_0^t (iX)e^{isX} d\mathcal{L}(s)\right].\end{aligned}$$

Aus $\mathbb{E}[|X|] = \int_{\mathbb{R}} |x|\ d\mathbb{P}_X(x) < \infty$ folgt

$$\int_{\mathbb{R}} \left(\int_0^t |(ix)e^{isx}| d\mathcal{L}(s)\right) d\mathbb{P}_X(x) = \int_{\mathbb{R}} |x|\ d\mathbb{P}_X(x)\ \int_0^t d\mathcal{L}(s) = \int_{\mathbb{R}} |x|\ d\mathbb{P}_X(x) < \infty,$$

und daher, mit Fubini,

$$\begin{aligned}\mathbb{E}\left[\int_0^t (iX)e^{isX} d\mathcal{L}(s)\right] &= \int_{\mathbb{R}} \left(\int_0^t (ix)e^{isx} d\mathcal{L}(s)\right) d\mathbb{P}_X(x)\\ &= \int_0^t \left(\int_{\mathbb{R}} (ix)e^{isx} \mathbb{P}_X(x)\right) d\mathcal{L}(s) = \int_0^t \mathbb{E}[(iX)e^{isX}]\ d\mathcal{L}(s).\end{aligned}$$

Schließlich erreichen wir

$$\phi_X(t) = 1+\mathbb{E}\left[\int_0^t (iX)e^{isX} d\mathcal{L}(s)\right] = \phi_X(0) + \int_0^t \mathbb{E}[(iX)e^{isX}]\ d\mathcal{L}(s).$$

Da $\mathbb{E}[|X|] < \infty$ gilt, ist die Funktion $s \mapsto \mathbb{E}[(iX)e^{isX}]$ gleichmäßig stetig auf $\mathbb{R}$ (siehe Beweis von Lemma 6.7). Daraus folgt, dass ϕ_X stetig differenzierbar ist mit $\phi'_X(t) = \mathbb{E}[(iX)e^{itX}]$.

Wir nehmen jetzt an, die Aussage gilt für n und betrachten den Fall $\mathbb{E}[|X|^{n+1}] < \infty$.
Aus $\mathbb{E}[|X|^n] < \infty$ folgt, mit der Induktionhypothese, $\phi_X \in C^n$ mit $\phi_X^{(n)}(t) = \mathbb{E}[(iX)^n e^{itX}]$. Wir rechnen

$$\begin{aligned}\phi_X^{(n)}(t) = \mathbb{E}[(iX)^n e^{itX}] &= \mathbb{E}\left[(iX)^n + \int_0^t (iX)^{n+1} e^{isX} d\mathcal{L}(s)\right] \\ &= \phi_X^{(n)}(0) + \mathbb{E}\left[\int_0^t (iX)^{n+1} e^{isX} d\mathcal{L}(s)\right].\end{aligned}$$

Da $\mathbb{E}[|X|^{n+1}] < \infty$, können wir wieder Fubini anwenden

$$\phi_X^{(n)}(t) = \phi_X^{(n)}(0) + \int_0^t \mathbb{E}\left[(iX)^{n+1} e^{isX}\right] \, d\mathcal{L}(s).$$

Da $\mathbb{E}[|X|^{n+1}] < \infty$ gilt, ist die Funktion $s \mapsto \mathbb{E}[(iX)^{n+1}e^{isX}]$ gleichmäßig stetig auf $\mathbb{R}$ (siehe Beweis von Lemma 6.7). Daraus folgt, dass $\phi_X^{(n)}$ stetig differenzierbar ist mit $(\phi_X^{(n)})'(t) = \mathbb{E}[(iX)^{n+1}e^{itX}]$. □

6.3 Charakteristische Funktion und Konvergenz

Satz 6.9. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen, X eine Zufallsvariable. Dann gilt*

$$X_n \overset{D}{\to} X \qquad \Longleftrightarrow \qquad \phi_{X_n} \overset{\text{punktweise}}{\to} \phi_X.$$

Beweis.
$\Rightarrow$ Wir nehmen an, dass $X_n \overset{D}{\to} X$. Dann gilt

$$\int_{\mathbb{R}} g d\mathbb{P}_{X_n} \to \int_{\mathbb{R}} g d\mathbb{P}_X \qquad \forall g \in C_b(\mathbb{R}).$$

Die Aussage folgt, da die Funktion $x \mapsto e^{itx}$ (aufgeteilt in reelle und imaginäre Komponente) stetig und beschränkt ist.

[20: 19.12.2025]

[21: 23.12.2025]

$\Leftarrow$ Wir nehmen an, dass ϕ_{X_n} punktweise gegen ϕ_X konvergiert.
Unser Ziel ist zu zeigen, dass $\int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} \to \int_{\mathbb{R}} g \, d\mathbb{P}_X$ $\forall g \in C_b(\mathbb{R})$ gilt.

Schritt 1. Sei $g \in C_c(\mathbb{R})$ stetig mit kompaktem Träger gegeben.
Wir zeigen, dass $\lim_{n\to\infty} \int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} = \int_{\mathbb{R}} g \, d\mathbb{P}_X$ gilt.

Es gilt, für alle $\sigma > 0$,

$$\begin{aligned}\left|\int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} - \int_{\mathbb{R}} g \, d\mathbb{P}_X\right| \leq &\int_{\mathbb{R}} |g - p_\sigma * g| \, d\mathbb{P}_{X_n} \\ &+ \left|\int_{\mathbb{R}} p_\sigma * g \, d\mathbb{P}_{X_n} - \int_{\mathbb{R}} p_\sigma * g \, d\mathbb{P}_X\right| \\ &+ \int_{\mathbb{R}} |g - p_\sigma * g| \, d\mathbb{P}_X,\end{aligned}$$

wobei

$$p_\sigma * g(x) = \int_{\mathbb{R}} p_\sigma(x-z) g(z) \; d\mathcal{L}(z).$$

Wir rechnen

$$\int_{\mathbb{R}} |g - p_\sigma * g| \; d\mathbb{P}_{X_n} \leq \sup_{x\in\mathbb{R}} |g(x) - p_\sigma * g(x)| \; \int_{\mathbb{R}} 1 \, d\mathbb{P}_{X_n} = \sup_{x\in\mathbb{R}} |g(x) - p_\sigma * g(x)|$$
$$\int_{\mathbb{R}} |g - p_\sigma * g| \; d\mathbb{P}_{X} \leq \sup_{x\in\mathbb{R}} |g(x) - p_\sigma * g(x)| \; \int_{\mathbb{R}} 1 \, d\mathbb{P}_{X} = \sup_{x\in\mathbb{R}} |g(x) - p_\sigma * g(x)|,$$

und daher

$$\begin{aligned} &\limsup_{n\to\infty} \left| \int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} \; - \; \int_{\mathbb{R}} g \, d\mathbb{P}_X \right| \\ \leq \; & 2 \sup_{x\in\mathbb{R}} |g(x) - p_\sigma * g(x)| + \limsup_{n\to\infty} \left| \int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_X \right|. \end{aligned}$$

Behauptung. Es gilt

$$\lim_{n\to\infty} \left| \int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_X \right| = 0. \tag{6.1}$$

Falls (6.1) gilt, erhalten wir

$$\limsup_{n\to\infty} \left| \int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} \; - \; \int_{\mathbb{R}} g \, d\mathbb{P}_X \right| \leq \; 2 \sup_{x\in\mathbb{R}} |g(x) - p_\sigma * g(x)|.$$

Im Beweis von Satz 6.4 haben wir die folgende Abschätzung für alle $x \in \mathbb{R}$ und $\sigma > 0$ bewiesen:

$$|g(x) - p_\sigma * g(x)| \leq \; \max_{|z|\leq\sqrt{\sigma}} |g(x+z) - g(x)| \; + \; 2^{\frac{3}{2}} \|g\|_{L^\infty} e^{-\frac{1}{4\sigma}}.$$

Da $g \in C_c(\mathbb{R})$ gilt, ist g gleichmäßig stetig auf $\mathbb{R}$ und daher gilt

$$\lim_{\sigma\to 0} \left[\sup_{x\in\mathbb{R}, |z|\sqrt{\sigma}} |g(x+z) - g(x)| \; + \; 2^{\frac{3}{2}} \|g\|_{L^\infty} e^{-\frac{1}{4\sigma}} \right] = 0,$$

und damit ist der Beweis von Schritt 1 beendet.

Beweis der Behauptung.
Im Beweis von Satz 6.4 haben wir folgende Identität bewiesen

$$\int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_X = \int_{\mathbb{R}} g(z) \rho_\sigma(z) \, d\mathcal{L}(z) = \int_{\mathbb{R}} g(z) \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ikz} p_{\frac{1}{\sigma}}(k) \phi_X(k) \, d\mathcal{L}(k) \right) d\mathcal{L}(z).$$

Analog gilt

$$\int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_X = \int_{\mathbb{R}} g(z) \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ikz} p_{\frac{1}{\sigma}}(k) \phi_{X_n}(k) \, d\mathcal{L}(k) \right) d\mathcal{L}(z),$$

und daher

$$\int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} p_\sigma * g \; d\mathbb{P}_X = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} g(z) \left(\int_{\mathbb{R}} e^{-ikz} p_{\frac{1}{\sigma}}(k) \left[\phi_{X_n}(k) - \phi_X(k) \right] \; d\mathcal{L}(k) \right) d\mathcal{L}(z).$$

Aus $g \in C_c(\mathbb{R})$ folgt, dass $g \in L^1(\mathbb{R}, \mathcal{L})$ gilt. Mit $|\phi_{X_n}(k) - \phi_X(k)| \leq 2$ erhalten wir

$$\int_{\mathbb{R}^2} |g(z)|\, p_{\frac{1}{\sigma}}(k)\, |\phi_{X_n}(k) - \phi_X(k)|\, d\mathcal{L}(k) d\mathcal{L}(z) \leq 2 \int_{\mathbb{R}} |g(z)| d\mathcal{L}(z) \int_{\mathbb{R}} p_{\frac{1}{\sigma}}(k) d\mathcal{L}(k) = 2 \int_{\mathbb{R}} |g(z)| d\mathcal{L}(z) < \infty.$$

Daher, mithilfe von Fubini, können wir die Integrationsordnung im Integral oben tauschen:

$$\begin{aligned}\int_{\mathbb{R}} p_\sigma * g \, d\mathbb{P}_{X_n} - \int_{\mathbb{R}} p_\sigma * g \, d\mathbb{P}_X &= \frac{1}{\sigma} \int_{\mathbb{R}} p_{\frac{1}{\sigma}}(k) \left[\phi_{X_n}(k) - \phi_X(k)\right] \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ikz} g(z) d\mathcal{L}(z) \right) d\mathcal{L}(k) \\ &= \frac{1}{\sigma} \int_{\mathbb{R}} p_{\frac{1}{\sigma}}(k) \hat{g}(-k) \; \left[\phi_{X_n}(k) - \phi_X(k)\right] \; d\mathcal{L}(k),\end{aligned}$$

wobei $\hat{g}$ wohldefiniert ist, weil $g \in L^1(\mathbb{R}, \mathcal{L})$. Es gilt

- $|p_{\frac{1}{\sigma}}(k)\hat{g}(-k) \; [\phi_{X_n}(k) - \phi_X(k)]\,| \leq 2\frac{1}{\sqrt{2\pi}} \|g\|_{L^1} \; p_{\frac{1}{\sigma}}(k) \in L^1(\mathbb{R}, \mathcal{L})$, und
- $\lim_{n\to\infty} p_{\frac{1}{\sigma}}(k)\hat{g}(-k) \; [\phi_{X_n}(k) - \phi_X(k)] = 0$ da $\lim_{n\to\infty} \phi_{X_n}(k) = \phi_X(k)$.

Die Aussage jetzt folgt aus dominierter Konvergenz.

Schritt 2. Sei $g \in C_b(\mathbb{R})$ gegeben. Wir zeigen, dass $\lim_{n\to\infty} \int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} = \int_{\mathbb{R}} g \, d\mathbb{P}_X$ gilt.

Wir approximieren g durch eine stetige Funktion mit kompaktem Träger wie folgt:
Sei $\chi \colon \mathbb{R} \to [0, 1]$ definiert durch

$$\chi(x) = \begin{cases} 1 & \text{falls } |x| \leq 1 \\ 0 & \text{falls } |x| \geq 2 \\ 2 - x & \text{falls } x \in (1, 2) \\ 2 + x & \text{falls } x \in (-2, -1), \end{cases}$$

und für $k \in \mathbb{N}$ sei $g_k(x) := g(x)\chi_k(x)$, wobei $\chi_k(x) := \chi\left(\frac{x}{k}\right)$. Dann gilt $g_k \in C_c(\mathbb{R})$ mit $g_k(x) = g(x) \; \forall |x| \leq k$ und $g_k(x) = 0 \; \forall |x| \geq 2k$.
Es gilt, für alle $k \in \mathbb{N}$,

$$\left| \int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} - \int_{\mathbb{R}} g \, d\mathbb{P}_X \right| \leq \int_{\mathbb{R}} |g - g_k| \; d\mathbb{P}_{X_n} + \left| \int_{\mathbb{R}} g_k \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} g_k \; d\mathbb{P}_X \right| + \int_{\mathbb{R}} |g - g_k| \; d\mathbb{P}_X$$

Da $g_k \in C_c(\mathbb{R})$ folgt aus Schritt 1 $\lim_{n\to\infty} \left| \int_{\mathbb{R}} g_k \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} g_k \; d\mathbb{P}_X \right| = 0$ und daher

$$\limsup_{n\to\infty} \left| \int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} - \int_{\mathbb{R}} g \, d\mathbb{P}_X \right| \leq \limsup_{n\to\infty} \int_{\mathbb{R}} |g - g_k| \; d\mathbb{P}_{X_n} + \int_{\mathbb{R}} |g - g_k| \; d\mathbb{P}_X.$$

Es gilt

$$|g(x) - g_k(x)| = |g(x)|(1 - \chi_k(x)) \leq \|g\|_{L^\infty} \; (1 - \chi_k(x)),$$

und daher, mit $(1 - \chi_k) = 0$ für $|x| \leq k$

$$\int_{\mathbb{R}} |g - g_k| \; d\mathbb{P}_X \leq \|g\|_{L^\infty} \int_{\mathbb{R}} (1 - \chi_k) \; d\mathbb{P}_X \leq \|g\|_{L^\infty} \int_{|x|>k} (1 - \chi_k) \; d\mathbb{P}_X \leq \|g\|_{L^\infty} \mathbb{P}(|X| > k).$$

Weiter rechnen wir

$$\begin{aligned}
&\int_{\mathbb{R}} |g - g_k| \; d\mathbb{P}_{X_n} \leq \|g\|_{L^\infty} \int_{\mathbb{R}} (1 - \chi_k) \; d\mathbb{P}_{X_n} = \|g\|_{L^\infty} \left[1 - \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_{X_n} \right] \\
&\leq \|g\|_{L^\infty} \left[\left| \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_X \right| + 1 - \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_X \right] \\
&= \|g\|_{L^\infty} \left[\left| \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_X \right| + \int_{\mathbb{R}} (1 - \chi_k) \; d\mathbb{P}_X \right] \\
&\leq \|g\|_{L^\infty} \left[\left| \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_X \right| + \mathbb{P}(|X| > k) \right].
\end{aligned}$$

Da $\chi_k \in C_c(\mathbb{R})$, folgt aus dem Schritt 1 $\lim_{n\to\infty} \left| \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_{X_n} - \int_{\mathbb{R}} \chi_k \; d\mathbb{P}_X \right| = 0$ und daher

$$\limsup_{n\to\infty} \left| \int_{\mathbb{R}} g \, d\mathbb{P}_{X_n} \; - \; \int_{\mathbb{R}} g \, d\mathbb{P}_X \right| \leq 2\|g\|_{L^\infty} \mathbb{P}(|X| > k) \qquad \forall k \in \mathbb{N}.$$

Die Aussage folgt aus $\lim_{k\to\infty} \mathbb{P}(|X| > k) = 0$. □

Bemerkung. Um Konvergenz in Verteilung zu beweisen, genügt es, punktweise Konvergenz fast überall der charakteristischen Funktionen zu haben (da wir dominierte Konvergenz anwenden). Aus der Konvergenz in Verteilung folgt aber punktweise Konvergenz der charakteristischen Funktionen.

7 Der zentrale Grenzwertsatz

7.1 Der zentrale Grenzwertsatz im Fall L^2

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten und integrierbaren (d.h. $X_1 \in L^1(\Omega, \mathbb{P})$) Zufallsvariablen.
Aus dem Gesetz der großen Zahlen (Satz 5.6) folgt

$$\frac{S_n - \mathbb{E}[S_n]}{n} \overset{f.s.}{\to} 0 \quad \text{für } n \to \infty.$$

Frage 1: wie schnell ist die Konvergenz gegen 0? Gibt es ein $\alpha > 0$, sodass

$$n^\alpha \frac{S_n - \mathbb{E}[S_n]}{n} \overset{D}{\to} Z \quad \text{für } n \to \infty,$$

wobei Z eine nicht-triviale Zufallsvariable ist? Z ist trivial falls $Z \in \{0, \pm\infty\}$ f.ü.

Frage 2: Falls α existiert, was ist die Verteilung von Z?

Beispiel. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängige identisch verteilte Zufallsvariablen mit $X_1 \sim Ber(p)$ $0 < p < 1$.
Es gilt $\mathbb{E}[X_1] = p$, $\text{var}(X_1) = p(1-p)$ und $S_n = \sum_{j=1}^n X_j \sim Bin(n, p)$.

Aus dem Gesetz der großen Zahlen folgt $\frac{S_n - np}{n} \overset{f.s.}{\to} 0$ für $n \to \infty$.

Aus dem Satz von de Moivre-Laplace folgt $\frac{S_n - np}{\sqrt{n}} \overset{D}{\to} Z \sim \mathcal{N}(0, p(1-p))$.

In diesem Fall gilt also $\alpha = \frac{1}{2}$ und Z ist normalverteilt. Der zentrale Grenzwertsatz sagt, dass dieses Resultat immer gilt, wenn die Zufallsvariablen zweifach integrierbar sind.

Satz 7.1 (Der zentrale Grenzwertsatz (central limit theorem))**.**
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $(X_n)_{n\in\mathbb{N}}$ *eine Folge von unabhängigen, identisch verteilte und zweifach integrierbaren (d.h.* $X_1 \in L^2(\Omega, \mathbb{P})$*) Zufallsvariablen.*
Sei $\mu := \mathbb{E}[X_1]$, $\sigma^2 := \mathrm{var}(X_1)$ *und* $Z_n := \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mu)$.

Dann gilt $Z_n \xrightarrow{D} Z \sim \mathcal{N}(0, \sigma^2)$ *für* $n \to \infty$.

Beweis. Wegen Satz 6.9 genügt es zu zeigen, dass $\lim_{n\to\infty} \phi_{Z_n}(t) = \phi_Z(t)$ gilt $\forall t \in \mathbb{R}$.
Da $Z \sim \mathcal{N}(0, \sigma^2)$, gilt (siehe Lemma 6.3) $\phi_Z(t) = e^{-\frac{1}{2}\sigma^2 t^2}$. Da die Zufallsvariables $(X_j - \mu)$ unabhängig identisch verteilt sind, folgt aus Lemma 6.6

$$\phi_{Z_n}(t) = \phi_{X_1-\mu}\left(\frac{t}{\sqrt{n}}\right)^n.$$

Da $(X_1 - \mu) \in L^2(\Omega, \mathbb{P})$, folgt aus dem Satz 6.8, dass $\phi := \phi_{X_1-\mu} \in C^2(\mathbb{R}, \mathbb{C})$ mit

$$\phi(0) = 1, \qquad \phi'(0) = i\mathbb{E}[(X_1 - \mu)] = 0, \qquad \phi''(0) = i^2\mathbb{E}[(X_1 - \mu)^2] = -\sigma^2.$$

Sei $t \in \mathbb{R}$ gegeben. Für $n \gg 1$ gilt $\frac{|t|}{\sqrt{n}} \ll 1$ und daher gilt $|\phi\left(\frac{t}{\sqrt{n}}\right) - 1| \ll 1$.
Daraus folgt, dass $\ln\left[\phi\left(\frac{t}{\sqrt{n}}\right)\right] \in \mathbb{C}$ wohldefiniert ist. Wir können also schreiben

$$\phi_{Z_n}(t) = \phi\left(\frac{t}{\sqrt{n}}\right)^n = e^{nf\left(\frac{t}{\sqrt{n}}\right)} \qquad \text{wobei } f(x) := \ln[\phi(x)].$$

Aus $\phi \in C^2$ folgt $f \in C^2$ mit

$$f(0) = \ln[\phi(0)] = 0, \qquad f'(0) = \frac{\phi'(0)}{\phi(0)} = 0, \qquad f''(0) = \frac{\phi''(0)}{\phi(0)} - \left(\frac{\phi'(0)}{\phi(0)}\right)^2 = -\sigma^2.$$

Außerdem gilt, für x klein (sodass die Funktion f wohldefiniert ist),

$$f(x) = f(0) + f'(0)x + x^2 \int_0^1 (1-s) f''(xs) d\mathcal{L}(s) = x^2 \int_0^1 (1-s) f''(xs) d\mathcal{L}(s).$$

Wir erhalten

$$nf\left(\frac{t}{\sqrt{n}}\right) = t^2 \int_0^1 (1-s) f''\left(\frac{ts}{\sqrt{n}}\right) d\mathcal{L}(s).$$

Aus dominierter Konvergenz und der Stetigkeit von f'' folgt

$$\lim_{n\to\infty} nf\left(\frac{t}{\sqrt{n}}\right) = t^2 \int_0^1 (1-s) f''(0) d\mathcal{L}(s) = -\sigma^2 t^2 \int_0^1 (1-s) d\mathcal{L}(s) = -\frac{1}{2}\sigma^2 t^2.$$

□

[21: 23.12.2025]

[22: 09.01.2026]

Bemerkung 1:
Im Falle, dass $\sigma = 0$, ist $X_n = \mu$ f.s. und gilt $\frac{S_n - \mu n}{\sqrt{n}} = 0$ f.s. . Das stimmt überein mit $\phi_{\frac{S_n - \mu n}{\sqrt{n}}}(t) \to e^0 = 1$ $\forall t \in \mathbb{R}$. Es gilt somit $\mathbb{P}_Z = \delta_0$ und wir schreiben $\mathcal{N}(0, 0) = \delta_0$.

Bemerkung 2:
Falls $X_1 \in L^2$ und $\sigma \neq 0$, so gilt für $\gamma > 0$

$$\frac{S_n - \mu n}{n^\gamma} \overset{D}{\to} \begin{cases} 0 & \text{falls } \gamma > \frac{1}{2} \\ \mathcal{N}(0, \sigma^2) & \text{falls } \gamma = \frac{1}{2} \\ \text{nicht definiert} & \text{falls } \gamma < \frac{1}{2}. \end{cases}$$

Beweis. Es gilt

$$\phi_{\frac{S_n - \mu n}{n^\gamma}}(t) = \phi_{X_1 - \mu}\left(\frac{t}{n^\gamma}\right)^n = e^{n \log\left(1 - \frac{\sigma^2}{2}\frac{t^2}{n^{2\gamma}} + O\left(\frac{t^3}{n^{3\gamma}}\right)\right)} = e^{n^{1-2\gamma} n^{2\gamma} \log\left(1 - \frac{\sigma^2}{2}\frac{t^2}{n^{2\gamma}} + O\left(\frac{t^3}{n^{3\gamma}}\right)\right)}.$$

Es gilt

$$n^{2\gamma} \log\left(1 - \frac{\sigma^2}{2}\frac{t^2}{n^{2\gamma}} + O\left(\frac{t^3}{n^{3\gamma}}\right)\right) \overset{n \to \infty}{\to} -\frac{\sigma^2}{2} t^2$$

und somit folgt

$$\phi_{\frac{S_n - \mu n}{n^\gamma}}(t) \overset{n \to \infty}{\to} \begin{cases} e^0 = 1 & \text{falls } \gamma > \frac{1}{2} \\ e^{-\frac{\sigma^2}{2} t^2} & \text{falls } \gamma = \frac{1}{2} \\ \mathbf{1}_{\{0\}}(t) & \text{falls } \gamma < \frac{1}{2} \end{cases}.$$

Es galt $1 = \phi_Z(t)$ für $Z = 0$ und $e^{-\frac{\sigma^2}{2} t^2} = \phi_Z(t)$ für $Z \sim \mathcal{N}(0, \sigma^2)$. Da $\mathbf{1}_{\{0\}}(t)$ nicht stetig ist, ist dies somit nicht die charakteristische Funktion einer Zufallsvariablen. □

Bemerkung 3:
Die Voraussetzung $X_1 \in L^2$ ist notwendig. Seien zum Beispiel $(X_n)_{n \in \mathbb{N}}$ unabhängige, identisch verteilte Zufallsvariablen mit $X_1 \sim Cauchy(a)$, wobei $a > 0$, so gilt $X_1 \notin L^1 \supset L^2$ und

$$\frac{S_n}{n^\gamma} \overset{D}{\to} \begin{cases} Z \sim Cauchy(a) & \text{falls } \gamma = 1 \\ 0 & \text{falls } \gamma > 1 \\ \text{nicht definiert} & \text{falls } \gamma < 1. \end{cases}$$

Beweis. Siehe Bemerkung 2 und Übung. □

Bemerkung 4:
Falls $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabhängige Zufallsvariablen sind, so gilt

$$\frac{S_n - \mu n}{\sqrt{n}} \sim \mathcal{N}(0, \sigma^2).$$

Beweis.

$$\phi_{\frac{S_n - \mu n}{\sqrt{n}}}(t) = \phi_{X_1 - \mu}\left(\frac{t}{\sqrt{n}}\right)^n = e^{-\frac{\sigma^2}{2}\frac{t^2}{n} \cdot n} = e^{-\frac{\sigma^2}{2} t^2} = \phi_Z(t)$$

für $Z \sim \mathcal{N}(0, \sigma^2)$. □

Allgemein: Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \in L^1$, $\mu = \mathbb{E}[X_1]$, $S_n = \sum_{i=1}^n X_i$ und $\gamma > 0$, dann gilt

$$\frac{S_n - \mu n}{n^\gamma} \xrightarrow{D} Z, \tag{7.1}$$

falls

i) $\phi_{\frac{S_n-\mu n}{n^\gamma}} \to \phi$ punktweise, wobei $\phi : \mathbb{R} \to \mathbb{C}$, und

ii) $\phi = \phi_Z$, d.h. ϕ ist die charakteristische Funktion einer Zufallsvariable.

Dementsprechend ist es wichtig, zu verstehen, welche Funktionen als charakteristische Funktion einer Zufallsvariable darstellbar sind. Ein Kriterium bietet der Satz von Bochner.

Satz 7.2 (Bochner)**.** *Sei $\phi : \mathbb{R} \to \mathbb{C}$. Folgende Aussagen sind äquivalent:*

i) ϕ ist die charakteristische Funktion einer Zufallsvariable.

ii) ϕ erfüllt

a) $\phi(0) = 1$,

b) ϕ ist gleichmäßig stetig,

c) ϕ ist nicht-negativ definit, d.h. $\forall n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{R}$ und $\forall z_1, \ldots, z_n \in \mathbb{C}$ gilt

$$\sum_{i,j=1}^n \overline{z_i}\phi(t_i - t_j)z_j \geq 0.$$

Beweis. Wir beweisen nur $i) \implies ii)$. Für den vollständigen Beweis, siehe Feller Vol II, Kapitel XIX 2. Sei $\phi = \phi_X$. Wir haben bereits gesehen, dass $a)$ und $b)$ gelten, es bleibt also $c)$ zu zeigen. Es gilt

$$\sum_{i,j=1}^n \overline{z_i}\phi(t_i - t_j)z_j = \mathbb{E}\left[\sum_{i,j=1}^n \overline{z_i}e^{i(t_i-t_j)X}z_j\right] = E\left[\left|\sum_{i=1}^n e^{it_iX}z_i\right|^2\right] \geq 0.$$

$\square$

7.2 Mögliche Grenzverteilungen

Wir sind interessiert and den Verteilungen, die als Grenzwert in (7.1) auftreten können. Eine solche Verteilung muss stabil sein, d.h. es muss gelten für unabhängige, gleichverteilte $Z_1, \ldots, Z_n$ mit $\mathbb{P}_{Z_i} = \mathbb{P}_Z$, dass

$$\mathbb{P}_{\frac{Z_1+\ldots+Z_n}{n^\gamma}} = \mathbb{P}_Z.$$

Beispiele: Für $\gamma = \frac{1}{2}$ erfüllt $\mathbb{P}_Z = \mathcal{N}(0, \sigma^2)$ dies. Für $\gamma = 1$ erfüllt $\mathbb{P}_Z = Cauchy(a)$ dies.

Lemma 7.3. *Sei $\phi : \mathbb{R} \to \mathbb{C}$, $\phi(t) = e^{-c|t|^\alpha}$, wobei $c > 0$ und $\alpha \in (0, 2]$.*

i) ϕ ist die charakteristische Funktion einer Zufallsvariable mit Verteilung $\nu_{\alpha,c}$.

ii) Für $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit Verteilung $\nu_{\alpha,c}$ gilt

$$\frac{X_1 + \ldots + X_n}{n^{1/\alpha}} \sim \nu_{\alpha,c}.$$

Beweis. i) $\Big[$ *Dieser Teil wurde nicht in der Vorlesung besprochen.*

Das Ziel ist zu zeigen, dass die Funktion $e^{-|t|^\alpha}$ nicht-negativ definit ist für $\alpha \in (0,2]$. Der Fall $e^{-c|t|^\alpha}$ folgt durch Skalierung von t.

Wir sagen, dass $f : \mathbb{R} \to \mathbb{C}$ *nicht-negativ definit* ist, falls $f(t) = \overline{f(-t)}$ und $\forall n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{R}$, die Matrix $M = (M_{jk})$ definert durch $M_{jk} := f(t_j - t_k) - f(t_j) - \overline{f(t_k)}$ nicht-negativ definit (als quadratische Form) ist, d.h. $\forall c_1, \ldots, c_n \in \mathbb{C}$, gilt

$$\sum_{j,k=1}^{n} \Big(f(t_j - t_k) - f(t_j) - \overline{f(t_k)} \Big) c_j \overline{c_k} \geq 0. \tag{7.2}$$

Diese Bedingung ist wohldefiniert, weil M hermitesch ist (das folgt aus $f(t) = \overline{f(-t)}$) und daher ist die Summe eine reelle Zahl. Wir zeigen zuerst:

Ist f nicht-negativ definit, so ist die Funktion e^f nicht-negativ definit im Sinne von Bochner.

Beweis: Das Schur-Produkt zweier Matrizen $A = (a_{jk})$ und $B = (b_{jk})$ ist die $n \times n$-Matrix $C := (a_{jk} b_{jk})$. Falls A und B nicht-negativ definite hermitesche Matrizen sind, so ist auch C nicht-negativ definit hermitisch. Um dies zu sehen, schreiben wir mithilfe des Spektralsatzes $B = PP^*$, wobei $P = (p_{jk}) \in \mathbb{C}^{n \times n}$ und $P^* = (\overline{p_{kj}})$ ist die adjungierte Matrix von P. Dann gilt für $c_1, \ldots, c_n \in \mathbb{C}$

$$\sum_{j,k=1}^{n} a_{jk} b_{jk} c_j \overline{c_k} = \sum_{l=1}^{n} \sum_{j,k=1}^{n} a_{jk} (p_{jl} c_j) \overline{(p_{kl} c_k)} \geq 0.$$

Falls $A = (a_{jk})$ also nicht-negativ definit hermitesch ist, so ist auch $\Big(\sum_{l=0}^{n} \frac{(a_{jk})^l}{l!}\Big)$ nicht-negativ hermitesch und daher auch $(e^{a_{jk}})$. Wir wenden dies auf die nicht-negativ definite hermitesche Matrix $M = (M_{jk})$ mit $M_{jk} = f(t_j - t_k) - f(t_j) - \overline{f(t_k)}$ an, es gilt also, dass die Matrix $(e^{M_{jk}})_{jk}$ ebenfalls nicht-negativ hermitesch ist. Es folgt

$$\begin{aligned} \sum_{j,k=1}^{n} e^{f(t_j - t_k)} c_j \overline{c_k} &= \sum_{j,k=1}^{n} e^{f(t_j - t_k) - f(t_j) - \overline{f(t_k)}} e^{f(t_j)} e^{\overline{f(t_k)}} c_j \overline{c_k} \\ &= \sum_{j,k=1}^{n} e^{f(t_j - t_k) - f(t_j) - \overline{f(t_k)}} \gamma_j \overline{\gamma_k} \geq 0, \end{aligned}$$

wobei $\gamma_j = c_j e^{f(t_j)} \in \mathbb{C}$. □

Es bleibt zu zeigen, dass $f(t) := -|t|^\alpha$ nicht-negativ definit für $\alpha \in (0,2]$ im Sinne von (7.2) ist. Dazu schreiben wir für $\alpha \in (0,2)$

$$\int_0^\infty (1 - \cos(ts)) s^{-1-\alpha} ds = |t|^\alpha \underbrace{\int_0^\infty (1 - \cos(u)) u^{-1-\alpha} du}_{=: C_\alpha > 0} \tag{7.3}$$

mithilfe von $u = |t|s$. Hierbei ist $\alpha \in (0,2)$ notwendig, damit das Integral wohldefiniert ist. Es gilt also

$$f(t) = -|t|^\alpha = \frac{1}{C_\alpha} \int_0^\infty (\cos(ts) - 1) s^{-1-\alpha} ds.$$

Wir zeigen nun, dass für fixes s die Funktion $f_s(t) := \cos(ts) - 1$ nicht-negativ definit im Sinne von (7.2) ist. Da f_s reellwertig ist, reicht es nach Aufspaltung von Real- und Imaginärteil, (7.2) für reellwertige c_j zu prüfen.

$$
\begin{aligned}
0 \le \left|\sum_{i=1}^{n}(e^{ist_j}-1)c_j\right|^2 &= \sum_{j,k=1}^{n} \operatorname{Re}(e^{is(t_j-t_k)} - e^{-ist_k} - e^{ist_j} + 1)c_j c_k \\
&= \sum_{j,k=1}^{n} (\cos(s(t_j-t_k)) - \cos(st_k) - \cos(st_j) + 1)c_j c_k \\
&= \sum_{j,k=1}^{n} (f_s(t_j-t_k) - f_s(t_j) - f_s(t_k))c_j c_k.
\end{aligned}
$$

Somit ist f_s nicht-negativ definit. Mit (7.3) folgt nun, dass auch $-|t|^\alpha = \frac{1}{C_\alpha}\int_0^\infty f_s(t)s^{-1-\alpha}ds$ nicht-negativ definit ist (Übung).

Für $\alpha = 2$ lässt sich direkt prüfen, dass $-|t|^2$ nicht-negativ definit ist. Somit erfüllt $e^{-|t|^\alpha}$ die Voraussetzungen im Satz von Bochner für $\alpha \in (0, 2]$. $\rceil$

ii)

$$
\phi_{\frac{X_1+\ldots+X_n}{n^{1/\alpha}}}(t) = \phi_{X_1}\left(\frac{t}{n^{1/\alpha}}\right)^n = e^{-c\frac{|t|^\alpha}{n}n} = e^{-c|t|^\alpha}.
$$

□

Sowohl die Normalverteilung $\mathcal{N}(0, 2c)$ ($\alpha = 2$) als auch die Cauchy-Verteilung $Cauchy(c)$ ($\alpha = 1$) sind hierbei Beispiele.

Definition 7.4. *Sei $\phi(t) := e^{itm}e^{-c|t|^\alpha}$ für $m \in \mathbb{R}$, $\alpha \in (0, 2]$ und $c > 0$. Die Wahrscheinlichkeitsverteilung mit charakteristischer Funktion ϕ heißt symmetrische α-stabile Verteilung (mit Mittelwert m). Für $m = 0$ wird die Verteilung mit $\nu_{\alpha,c}$ notiert.*

Bemerkung. ϕ ist eine charakteristische Funktion (Übung). Für $\alpha > 1$ ist $\phi \in C^1(\mathbb{R})$. In diesem Falle ist $m = \mathbb{E}[X]$ wohldefiniert.

7.3 Der zentrale Grenzwertsatz für heavy tails

Als nächstes formulieren wir eine Version des zentralen Grenzwertsatzes für $X_1 \notin L^2$ (heavy tail). Dazu benötigen wir folgende Asymptotik:

Lemma 7.5. *Sei X eine Zufallsvariable mit $\mathbb{P}_X = \rho_X \mathcal{L}$. Wir nehmen an, dass $r > 0$ existiert sowie $\alpha \in (0, 2)$, sodass*

$$
\rho(x) = \frac{1}{|x|^{1+\alpha}} \quad \forall |x| > r.
$$

i) Falls $\alpha \in (1, 2)$, so gilt

$$
\phi_X(t) = 1 + imt - c|t|^\alpha + O(t^2), \quad t \to 0,
$$

wobei $m = \mathbb{E}[X]$ und $c = \int_\mathbb{R} \frac{1-\cos u}{|u|^{1+\alpha}}du$.

ii) Falls $\alpha \in (0, 1)$, so gilt

$$
\phi_X(t) = 1 - c|t|^\alpha + O(t), \quad t \to 0,
$$

wobei $c = \int_\mathbb{R} \frac{1-\cos u}{|u|^{1+\alpha}}du$.

iii) Falls $\alpha = 1$, so gilt

$$\phi_X(t) = 1 + imt - c|t| + O(t^2), \quad t \to 0,$$

wobei $m = \int_{-r}^{r} x\rho(x)dx$ und $c = \int_{\mathbb{R}} \frac{1-\cos u}{|u|^2} du$.

Beweis. Für $\alpha > 0$ ist

$$\int_{\mathbb{R}} \rho dx = \int_{-r}^{r} \rho dx + \underbrace{\int_{|x|>r} \frac{1}{|x|^{1+\alpha}} dx}_{<\infty},$$

weswegen $\mathbb{P}_X$ wohldefiniert ist.

(i) Sei $\alpha \in (1,2)$. Es gilt $\int_{\mathbb{R}} |x|\rho(x)dx < \infty$, wonach $m = \mathbb{E}[X] < \infty$ wohldefiniert ist. Es gilt

$$\begin{aligned}\phi_X(t) - 1 - imt &= \int_{\mathbb{R}} (e^{itx} - 1 - itx)\rho(x)dx \\ &= \int_{\mathbb{R}} (e^{itx} - 1 - itx)\frac{1}{|x|^{1+\alpha}} dx + \underbrace{\int_{\mathbb{R}} (e^{itx} - 1 - itx)\left[\rho(x) - \frac{1}{|x|^{1+\alpha}}\right] dx}_{=:R}.\end{aligned}$$

Dabei ist das erste Integral wohldefiniert, da $e^{itx} - 1 - itx = O((tx)^2)$ für $x \to 0$ und $e^{itx} - 1 - itx = O(tx)$ für $x \to \infty$. Daher gilt mit $u = |t|x$

$$\begin{aligned} &= |t|^\alpha \int_{\mathbb{R}} \frac{e^{iu} - 1 - iu}{|u|^{1+\alpha}} du + R \\ &= |t|^\alpha \frac{1}{2} \int_{\mathbb{R}} \frac{e^{iu} - 1 - iu + e^{-iu} - 1 + iu}{|u|^{1+\alpha}} du + R \\ &= |t|^\alpha \underbrace{\int_{\mathbb{R}} \frac{\cos(u) - 1}{|u|^{1+\alpha}} du}_{=-c} + R.\end{aligned}$$

Desweiteren gilt für R

$$\begin{aligned}|R| &= \left| \int_{|x|\le r} (e^{itx} - 1 - itx)\left[\rho(x) - \frac{1}{|x|^{1+\alpha}}\right] dx \right| \\ &\le Ct^2 \left[r^2 \int_{\mathbb{R}} \rho dx + \int_{-r}^{r} x^2 \frac{1}{|x|^{1+\alpha}} dx \right] \le Ct^2.\end{aligned}$$

(ii) und (iii) Übung. □

Satz 7.6 (zentrale Grenzwertsatz für heavy tails)**.** *Sei $(X_n)_n$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \sim \rho\mathcal{L}$, wobei $\rho(x) = \frac{1}{|x|^{1+\alpha}} \forall |x| > r$. Sei $S_n = \sum_{i=1}^n X_i$.*

i) Falls $\alpha \in (1,2)$, so gilt

$$\frac{S_n - n\mathbb{E}[X_1]}{n^{1/\alpha}} \xrightarrow{D} Z \sim \nu_{\alpha,c},$$

wobei $\phi_Z(t) = e^{-c|t|^\alpha}$ und $c = \int_{\mathbb{R}} \frac{1-\cos u}{|u|^{1+\alpha}} du$.

ii) Falls $\alpha \in (0,1)$, so gilt

$$\frac{S_n}{n^{1/\alpha}} \xrightarrow{D} Z \sim \nu_{\alpha,c},$$

wobei $\phi_Z(t) = e^{-c|t|^\alpha}$.

iii) Falls $\alpha = 1$, so gilt

$$\frac{S_n - nm}{n} \xrightarrow{D} Z \sim \nu_{1,c} = Cauchy(c),$$

wobei $m = \int_{-r}^{r} x\rho(x)dx$ und $c = \int_{\mathbb{R}} \frac{1-\cos u}{|u|^2} du$.

Beweis.
(i) Es gilt mit $\mu = \mathbb{E}[X_1]$

$$\begin{aligned}
\phi_{\frac{S_n - n\mu}{n^{1/\alpha}}}(t) &= \phi_{X_1-\mu}\left(\frac{t}{n^{1/\alpha}}\right)^n \\
&= \left(1 - c\frac{|t|^\alpha}{n} + O\left(\frac{t^2}{n^{2/\alpha}}\right)\right)^n \\
&= e^{n\ln\left(1-c\frac{|t|^\alpha}{n}+O\left(\frac{t^2}{n^{2/\alpha}}\right)\right)} \\
&\to e^{-c|t|^\alpha}.
\end{aligned}$$

(ii) und (iii) Übung.

□

[22: 09.01.2026]

[23: 13.01.2026]

8 Markov Prozesse

8.1 Definition: zeitlich homogene Markov-Ketten

Erinnerung: die einfache Irrfahrt auf $\mathbb{Z}$ ist die Folge $n \mapsto S_n$ wobei $S_n = \sum_{j=1}^n X_j$ und $(X_j)_{j\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit $\mathbb{P}_{X_1} = p\delta_1 + (1-p)\delta_{-1}$ (siehe Def. 2.19). S_n beschreibt die Position des Fahrers an der Zeit n.
Allgemein nehmen die mögliche Zeiten Werte in einer Indexmenge I (abzählbar oder überabzählbar) und die Position des Fahrers liegt in einer Menge S, die *Zustandsraum* genannt wird.
In dieser Vorlesung werden wir nur diskrete Markov Prozesse mit Indexmenge $I = \mathbb{N}_0$ betrachten.

Definition 8.1. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und S ein topologischer Raum. Eine Funktion $X\colon \Omega \to S^{\mathbb{N}_0}$, die $\mathcal{F} - \otimes_{n\in\mathbb{N}_0}\mathcal{B}(S) -$ messbar ist, heißt* stochastischer Prozess *mit Zustandsraum S und Indexmenge $\mathbb{N}_0$.*
Wir schreiben auch $X = (X_n)_{n\in\mathbb{N}_0}$.

Bemerkung. Ein stochastischer Prozess mit Zustandsraum S und Indexmenge $\mathbb{N}_0$ ist auch eine Folge von Zufallsvariablen $X = (X_n)_{n\in\mathbb{N}_0}$, $X_n : \Omega \to S$.

Definition 8.2. *Sei $X : \Omega \to S^{\mathbb{N}_0}$ ein stochastischer Prozess mit* endlichem *Zustandsraum S.*

(i) X heißt Markov-Kette*, falls $\forall n \in \mathbb{N}_0, x_0, \ldots, x_n \in S$, sodass $\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0) > 0$ gilt*

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n). \tag{8.1}$$

Diese Gleichung heißt Markov Eigenschaft.

(ii) Sei X eine Markov-Kette.

- *Die* Anfangsverteilung *ist die Funktion $\pi_0 : S \to [0,1]$ definiert durch* $\pi_0(x) := \mathbb{P}(X_0 = x)$.
- *Für $n \geq 0$ und $x \in S$, sodass $\mathbb{P}(X_n = x) > 0$, ist die* Übergangswahrscheinlichkeit *an der Zeit n die Funktion $p_n(x,\cdot) : S \to [0,1]$ definiert durch* $p_n(x,y) := \mathbb{P}(X_{n+1} = y \mid X_n = x)$.

Bemerkungen.

- π_0 ist ein Vektor $\pi_0 \in \mathbb{R}^S$, sodass $\pi_0(x) \in [0,1]$ $\forall x \in S$ und $\sum_{x\in S} \pi_0(x) = 1$.
- Falls $\mathbb{P}(X_n = x) > 0$, dann ist $p_n(x,y) \in [0,1]$ $\forall y \in S$ und $\sum_{y\in S} p_n(x,y) = 1$.
- Falls $\mathbb{P}(X_n = x) = 0$, dann ist die bedingte Wahrscheinlichkeit $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ nicht definiert.

  In diesem Fall können wie eine Funktion $p_n(x,\cdot) : S \to [0,1]$ frei wählen, sodass $\sum_{y\in S} p_n(x,y) = 1$ gilt.

  Die so konstruierte Matrix $p_n \in [0,1]^{S\times S}$ heißt *Übergangsmatrix* an der Zeit n.
- Eine Matrix $M \in \mathbb{R}^{N\times N}$ heißt *stochastisch*, falls $M_{ij} \in [0,1]$ $\forall i,j = 1,\ldots,N$ und $\sum_{j=1}^N M_{ij} = 1$ $\forall i = 1,\ldots,N$ gilt. Die Übergangsmatrix ist also immer eine stochastische Matrix.
- Sei $M \in [0,1]^{N\times N}$ eine stochastische Matrix. Dann gilt (Übung)

$$M 1_N = 1_N, \qquad \text{wobei } 1_N^t := (1,\ldots,1). \tag{8.2}$$

- Die Anfangsverteilung einer Markov-Kette ist eindeutig bestimmt. Die Übergangsmatrix hingegen ist im Allgemeinen nicht eindeutig bestimmt.

Satz 8.3. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und S eine endliche Menge.*

(i) Sei $X : \Omega \to S^{\mathbb{N}_0}$ eine Markov-Kette mit Anfangsverteilung π_0 und Übergangsmatrizen $(p_n)_{n\in\mathbb{N}_0}$. Dann ist die Verteilung von X eindeutig durch π_0 und $(p_n)_{n\in\mathbb{N}_0}$ bestimmt.

Insbesondere gilt $\forall n \geq 0$ und $x_0, \cdots, x_n \in S$

$$\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0) = \pi_0(x_0) p_0(x_0, x_1) \cdots p_{n-1}(x_{n-1}, x_n). \tag{8.3}$$

(ii) Sei $\pi_0\colon S \to [0,1]$, sodass $\sum_{x\in S}\pi_0(x) = 1$, und sei $(p_n)_{n\in\mathbb{N}_0}$ eine Folge von stochastischen Matrizen $p_n \in [0,1]^{S\times S}$. Dann existiert eine Markov-Kette $X\colon \Omega \to S^{\mathbb{N}_0}$ mit Anfangsverteilung π_0 und Übergangsmatrizen $(p_n)_{n\in\mathbb{N}_0}$.

Beweis.
(i) Die Zylindermengen sind ein durchschnitt-stabiles Mengensystem, das $\otimes_{n\in\mathbb{N}_0}\mathcal{B}(S)$ erzeugt. Nach Satz 1.10 genügt es also zu zeigen, dass $\forall n \geq 0$ und $x_0 \cdots, x_n \in S$ die Wahrscheinlichkeit $\mathbb{P}\left(X_n = x_n, \ldots, X_0 = x_0\right)$ eindeutig durch π_0 und die Übergangsmatrizen $(p_k)_{k=0}^{n-1}$ bestimmt wird.
Wir werden (8.3) durch Induktion in n beweisen.

Für $n = 0$ gilt $\mathbb{P}(X_0 = x_0) = \pi_0(x_0)$.

Für $n = 1$ und $\pi_0(x_0) = \mathbb{P}(X_0 = x_0) > 0$ gilt

$$\mathbb{P}(X_1 = x_1, X_0 = x_0) = \mathbb{P}(X_1 = x_1 | X_0 = x_0)\mathbb{P}(X_0 = x_0) = p_0(x_0, x_1)\pi_0(x_0).$$

Für $n = 1$ und $\pi_0(x_0) = \mathbb{P}(X_0 = x_0) = 0$ gilt

$$\mathbb{P}(X_1 = x_1, X_0 = x_0) = 0 = p_0(x_0, x_1)0 = p_0(x_0, x_1)\pi_0(x_0).$$

Wir nehmen jetzt $\mathbb{P}(\{X_j = x_j\}_{j=1}^n) = \pi_0(x_0)p_0(x_0, x_1)\cdots p_{n-1}(x_{n-1}, x_n)\ \forall x_0, \ldots, x_n \in S$ an.
Falls $\mathbb{P}\left(\{X_j = x_j\}_{j=1}^n\right) > 0$, dann gilt $\forall x_{n+1} \in S$,

$$\begin{aligned}
\mathbb{P}\left(\{X_j = x_j\}_{j=1}^{n+1}\right) &= \mathbb{P}\left(X_{n+1} = x_{n+1} \mid \{X_j = x_j\}_{j=1}^n\right)\mathbb{P}\left(\{X_j = x_j\}_{j=1}^n\right) \\
&\overset{(8.1)}{=} \mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right)\mathbb{P}\left(\{X_j = x_j\}_{j=1}^n\right) \\
&= p_n(x_n, x_{n+1})\mathbb{P}\left(\{X_j = x_j\}_{j=1}^n\right) \overset{IH}{=} \pi_0(x_0)p_0(x_0, x_1)\cdots p_{n-1}(x_{n-1}, x_n)p_n(x_n, x_{n+1}).
\end{aligned}$$

(ii) Wir definieren $\mathbb{P}\left(X_n = x_n, \ldots, X_0 = x_0\right)$ via (8.3). Dann gilt $\mathbb{P} = \mathbb{P}_X$ für eine Markov-Kette mit Anfangsverteilung π_0 und Übergangsmatrizen $(p_n)_{n\geq 0}$. Dies wird aber später in der Vorlesung *Stochastische Prozesse* bewiesen. □

Bemerkung. Die Verteilung an der Zeit n ist definiert durch $\pi_n(x) \coloneqq \mathbb{P}(X_n = x)$. Aus (8.3) folgt

$$\begin{aligned}
\pi_n(x) = \mathbb{P}(X_n = x) &= \sum_{x_0,\ldots,x_n\in S}\mathbb{P}\left(X_n = x, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0\right) \\
&= \sum_{x_0,\ldots,x_{n-1}\in S}\pi_0(x_0)p_0(x_0, x_1)\cdots p_{n-1}(x_{n-1}, x) = (\pi_0^t p_0 p_1 \cdots p_{n-1})(x).
\end{aligned} \tag{8.4}$$

Definition 8.4. *Eine Markov-Kette heißt* zeitlich homogen *(oder hat* stationäre Übergagnswahrscheinlichkeiten*), falls die Übergangsmatrizen nicht von der Zeit abhängen.*

Beispiel. Sei $\pi_0 \in [0,1]^N$ mit $\sum_{j=1}^N \pi_0(j) = 1$, $X = (X_n)_{n\in\mathbb{N}_0}$ eine Familie unabhängigen identisch verteilten Zufallsvariablen mit Werte in $\{1, \ldots, N\}$ und $X_0 \sim \pi_0$.
Dann ist X eine zeitlich homogene Markov-Kette mit Übergangsmatrix $p \in [0,1]^{N\times N}$ definiert durch

$$p = \begin{pmatrix} \pi_0(1) & \pi_0(2) & \cdots & \pi_0(N) \\ \pi_0(1) & \pi_0(2) & \cdots & \pi_0(N) \\ \vdots & & & \vdots \\ \pi_0(1) & \pi_0(2) & \cdots & \pi_0(N) \end{pmatrix}.$$

Lemma 8.5. *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S, Anfangsverteilung π_0 und Übergangsmatrix p. Dann ist die Verteilung an der Zeit $n \geq 1$ π_n bestimmt durch*

$$\pi_n^t = \pi_0^t p^n. \tag{8.5}$$

Beweis. Folgt aus (8.4) mit $p_j = p \; \forall j \geq 0$. □

Bemerkung. Sei p eine stochastische Matrix. Dann ist die Matrix p^n auch eine stochastische Matrix, für alle $n \in \mathbb{N}$.

Beweis durch Induktion in n :

Für $n = 1$ ist die Aussage wahr. Wir nehmen an, dass p^n auch eine stochastische Matrix ist. Wir rechnen

$$p^{n+1}(x,y) = \sum_{z\in S} p(x,z)p^n(z,y).$$

Aus $p(x,z) \geq 0$ und $p^n(z,y) \geq 0$ folgt, dass $p^{n+1}(x,y) \geq 0$. Weiter gilt

$$\sum_{y\in S} p^{n+1}(x,y) = \sum_{y,z\in S} p(x,z)p^n(z,y) = \sum_{z\in S} \left[\sum_{y\in S} p^n(z,y)\right] \; p(x,z) = \sum_{z\in S} p(x,z) = 1.$$

Daher ist p^{n+1} eine stochastische Matrix. □

8.2 Invariante Verteilungen und Ergodensatz

Definition 8.6. *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S, Übergangsmatrix p. Eine Anfangsverteilung $\pi_0 \colon S \to [0,1]$ heißt* invariante Verteilung *für X, falls $\pi_n = \pi_0$ $\forall n \geq 1$.*

Lemma 8.7. *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergagnsmatrix p. Eine Anfangsverteilung $\pi_0 \colon S \to [0,1]$ ist eine invariante Verteilung für X genau dann wenn*

$$\pi_0^t p = \pi_0^t, \tag{8.6}$$

d.h. der Vektor π_0 ist ein links-Eigenvektor von p (äquivalent dazu ein rechts-Eigenvektor von p^t) zum Eigenwert 1.

Beweis. Aus (8.5) gilt $\pi_n^t = \pi_0^t p^n \; \forall n \geq 1$.
Die Verteilung π_0 ist also invariant $\Leftrightarrow \pi_0^t p^n = \pi_0^t \; \forall n \geq 1 \Leftrightarrow \pi_0^t p = \pi_0^t$. □

Beispiel. Sei $\pi_0 \in [0,1]^N$ mit $\sum_{j=1}^N \pi_0(j) = 1$, $X = (X_n)_{n\in\mathbb{N}_0}$ eine Familie von unabhängigen, identisch verteilten Zufallsvariablen mit Werten in $\{1,\dots,N\}$ und $X_0 \sim \pi_0$. Dann ist π_0 eine invariante Verteilung.

Satz 8.8. *[Existenz und Eindeutigkeit] Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix p.*

(i) Die Zahl 1 ist ein Eigenwert der Matrix p^t und es gibt wenigstens einen Vektor $\mu \in [0,1]^S$, sodass $p^t\mu = \mu$ und $\sum_{x\in S} \mu(x) = 1$. Weiter gilt für alle Eigenwerte $\lambda \in \mathbb{C} \setminus \{1\}$ von p^t, dass $|\lambda| \leq 1$.

Insbesondere besizt X wenigstens eine invariante Verteilung.

(ii) Falls $\exists k \in \mathbb{N}$,, sodass $p^k(x,y) > 0\ \forall x, y \in S$, dann ist 1 ein einfacher Eigenwert für p^t und es gibt einen Vektor $\mu \in (0,1]^S$, sodass $p^t\mu = \mu$ und $\sum_{x\in S}\mu(x) = 1$. Weiter gilt für alle Eigenwerte $\lambda \in \mathbb{C}\setminus\{1\}$ von p^t, dass $|\lambda| < 1$.

Insbesondere besizt X genau eine invariante Verteilung μ. Dazu gilt $\mu(x) > 0\ \forall x \in S$.

Um diese Resultat zu beweisen, brauchen wir die folgende zwei Versionen des Satzes von Perron-Frobenius.

Satz 8.9. *[Perron-Frobenius 1] Sei $A \in \mathbb{R}^{N\times N}$, $N > 1$, eine Matrix mit $A_{ij} > 0\ \forall i, j = 1, \ldots N$. Dann existiert ein $\lambda_0 > 0$, sodass*

(i) λ_0 ist ein einfacher Eigenwert von A,

(ii) $\exists v_0 \in \mathbb{R}^N$, sodass $Av_0 = \lambda_0 v_0$ und $v_0(j) > 0\ \forall j = 1, \ldots, N$,

(iii) $\forall \lambda \in \mathbb{C}\setminus\{\lambda_0\}$ Eigenwert von A gilt $|\lambda| < \lambda_0$.

Beweis. Siehe z.B. die Vorlesungsnotizen von Prof. Bovier. □

[23: 13.01.2026]

[24: 16.01.2026]

Satz 8.10. *[Perron-Frobenius 2] Sei $A \in \mathbb{R}^{N\times N}$, $N > 1$, eine Matrix mit $A_{ij} \geq 0\ \forall i, j = 1, \ldots N$. Sei*

$$\lambda_0 \coloneqq \sup\left\{\lambda \geq 0 \middle| \ \exists v \in [0,1]^N \text{ s.d. } \sum_{j=1}^N v_j = 1 \text{ und } (Av)_j \geq \lambda v_j\, \forall j = 1, \ldots N\right\}.$$

Dann gilt $\lambda_0 \geq 0$ und

(i) λ_0 ist ein Eigenwert von A,

(ii) $\exists v_0 \in \mathbb{R}^N\setminus\{0\}$, sodass $Av_0 = \lambda_0 v_0$ und $v_0(j) \geq 0\ \forall j = 1, \ldots, N$,

(iii) $\forall \lambda \in \mathbb{C}\setminus\{\lambda_0\}$ Eigenwert von A gilt $|\lambda| \leq \lambda_0$.

Beweis. Siehe z.B. die Vorlesungsnotizen von Prof. Bovier. □

Beweis von Satz 8.8. Im Fall $|S| = 1$ gelten die Aussage automatisch (Übung). Wir nehmen an $|S| > 1$. π_0 ist eine invariante Verteilung genau dann, wenn $p^t\pi_0 = \pi_0$ (siehe Lemma 8.7).

(*i*) Da $p^t(x,y) \geq 0$ folgt, mit Satz 8.10, dass $\lambda_0 \geq 0$ und $v \in [0,1]^S$ existiert, sodass $p^t v = \lambda_0 v$ und $\sum_{x\in S} v(x) = 1$. Weiter gilt: Für $\lambda \in \mathbb{C}\setminus\{\lambda_0\}$ Eigenwert von p gilt $|\lambda| \leq \lambda_0$.
Es genügt also zu zeigen, dass $\lambda_0 = 1$.
Aus $(p^t v)(x) = \lambda_0 v(x)\ \forall x \in S$ und $\sum_{x\in S} v(x) = 1$, folgt

$$\sum_{x\in S}(p^t v)(x) = \lambda_0 \sum_{x\in S} v(x) = \lambda_0.$$

Wir rechnen, mit $\sum_{x\in S} v(x) = 1$ und $\sum_{x\in S} p(y,x) = 1$ (p ist eine stochastiche Matrix),

$$\lambda_0 = \sum_{x\in S}(p^t v)(x) = \sum_{x,y\in S} p(y,x)v(y) = \sum_{y\in S}\left[\sum_{x\in S} p(y,x)\right] v(y) = \sum_{y\in S} v(y) = 1.$$

(ii) Aus $(p^k)^t(y,x) = p^k(x,y) > 0\ \forall x,y \in S$ folgt, mit dem Satz 8.9, dass $\lambda_0 > 0$ und $v_0 \in (0,1]^S$ existieren, sodass $(p^k)^t v_0 = \lambda_0 v_0$ und $\sum_{x\in S} v_0(x) = 1$. Weiter ist λ_0 ein einfacher Eigenwert von $(p^k)^t$ und es gilt: Für $\lambda \in \mathbb{C} \setminus \{\lambda_0\}$ Eigenwert von $(p^k)^t$ gilt $|\lambda| < \lambda_0$.
Da p^k eine stochastische Matrix ist (siehe Bemerkung nach Lemma 8.5), gilt

$$\lambda_0 = \lambda_0 \sum_{x\in S} v_0(x) = \sum_{x\in S}((p^k)^t v_0)(x) = \sum_{y\in S}\left[\sum_{x\in S}(p^k)(y,x)\right] v_0(y) = \sum_{y\in S} v_0(y) = 1.$$

Die Zahl 1 ist also ein *einfacher* Eigenwert für $(p^k)^t$, d.h.:

$$(p^k)^t v' = v' \quad \Rightarrow \quad v' = \alpha v_0 \quad \text{für ein } \alpha \in \mathbb{C} \setminus \{0\}.$$

Sei $v \in \mathbb{R}^S \setminus \{0\}$. Falls $p^t v = v$, dann gilt $(p^t)^k v = (p^k)^t v = v$, und daher $v = \alpha v_0$ für ein $\alpha \in \mathbb{C}$.
Die Zahl 1 ist also ein einfacher Eigenwert für p^t.
Aus (i) existiert $\mu = \alpha v_0$, sodass $p^t \mu = \mu$ und $\sum_{x\in S} \mu(x) = 1$. Wir rechnen

$$1 = \sum_{x\in S} \mu(x) = \alpha \sum_{x\in S} v_0(x) = \alpha,$$

daher gilt $\mu = v_0$. Insbesondere ist die invariante Verteilung μ eindeutig und erfüllt $\mu(x) > 0$ $\forall x \in S$.
Sei jetzt $\lambda \in \mathbb{C} \setminus \{1\}$ Eigenwert von p^t. Dann ist λ^k Eigenwert von $(p^t)^k$ und daher gilt $|\lambda| < 1$.
Damit ist der Beweis beendet. $\square$

Satz 8.11 (Ergodensatz)**.** *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix p, sodass ein $k \in \mathbb{N}$ existiert mit $p^k(x,y) > 0\ \forall x,y \in S$. Sei μ die eindeutige invariante Verteilung. Dann gilt*

$$\lim_{n\to\infty} \pi_n = \mu$$

für jede Anfangsverteilung π_0.

Bemerkung. Markovketten, die eine einzige invariante Verteilung besitzen, gegen welche die Verteilung π_n für jede Anfangsverteilung π_0 konvergiert, nennt man auch *ergodisch.*

Beweis. Im Fall $|S| = 1$ gelten die Aussage automatisch (Übung). Wir nehmen an $|S| > 1$ also $S = \{x_1, \dots, x_N\}$, mit $N > 1$. Um die Notation zu vereinfachen, schreiben wir hier $S = \{1, \dots, N\}$.
Sei jetzt $P_\mu \in [0,1]^{N\times N}$ die Matrix definiert durch

$$P_\mu \coloneqq \begin{pmatrix} \mu(1) & \mu(2) & \cdots & \mu(N) \\ \mu(1) & \mu(2) & \cdots & \mu(N) \\ \vdots & & & \vdots \\ \mu(1) & \mu(2) & \cdots & \mu(N) \end{pmatrix}.$$

Behauptung: Es gilt:

(a) P_μ ist eine stochastische Matrix. Außerdem gilt $P_\mu^2 = P_\mu$ und $\forall v \in \mathbb{C}^N$ mit $\sum_{j=1}^N v_j = 1$,

$$v^t P_\mu = \mu^t.$$

(b) $\lim_{n\to\infty} \|(p^n)^t - P_\mu^t\|_{op} = 0$, wobei

$$\|M\|_{op} \coloneqq \sup_{v\in\mathbb{C}^N,\ |v|=1} |Mv| \qquad \forall M \in \mathbb{C}^{N\times N}.$$

Konsequenz: Aus (b) folgt $\lim_{n\to\infty} p^n(i,j) = P_\mu(i,j)\ \forall i,j = 1,\ldots,N$ und daher

$$\lim_{n\to\infty} \pi_n(i) = \lim_{n\to\infty} (\pi_0^t p^n)(i) = \sum_{j=1}^N \pi_0(j) \lim_{n\to\infty} p^n(j,i) = \sum_{j=1}^N \pi_0(j) P_\mu(j,i) = (\pi_0^t P_\mu)(i) = \mu(i).$$

Beweis der Behauptung:
(a) P_μ ist eine stochastiche Matrix, da $P_\mu(i,j) = \mu(j) \geq 0$ und $\sum_{j=1}^N P_\mu(i,j) = \sum_{j=1}^N \mu(j) = 1$. Weiter gilt

$$P_\mu^2(i,j) = \sum_k P_\mu(i,k) P_\mu(k,j) = \sum_k \mu(k)\mu(j) = \mu(j) = P_\mu(i,j)$$

und $(v^t P_\mu)_j = \sum_{i=1}^N v_i \mu(j) = \mu(j)\ \forall v$ mit $\sum_j v_j = 1$.

(b) Sei $A \coloneqq p^t$. Da 1 ein einfacher Eigenwert für A ist und $|\lambda| < 1$ für alle andere Eigenwerte, existiert eine invertierbare Matrix $U \in \mathbb{C}^{N\times N}$, sodass $A = U A_J U^{-1}$ wobei

$$A_J = \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$$

und Q eine blockdiagonale Matrix ist (die Jordan-Normalform). Jeder Block B in Q ist entweder diagonal $B = \lambda Id_d$ oder der Form

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 \\ 0 & 0 & \lambda & 1 & \cdots & 0 \\ \vdots & & & & & 0 \\ \vdots & & & & & 1 \\ 0 & 0 & 0 & 0 & \cdots & \lambda \end{pmatrix},$$

wobei $\lambda \neq 1$ ein Eigenwert von A ist und daher $|\lambda| < 1$. Wir schreiben

$$A = U \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} U^{-1} + U \begin{pmatrix} 0 & 0 \\ 0 & Q \end{pmatrix} U^{-1} \coloneqq A_0 + A_1$$

Es gilt $A_0 = P_\mu^t$ und $A_0 A_1 = 0 = A_1 A_0$ (Übung). Daraus folgt, mit $(P_\mu^t)^2 = P_\mu^t$,

$$(p^t)^n = A^n = (A_0 + A_1)^n = A_0^n + A_1^n = P_\mu^t + A_1^n.$$

Unser Ziel ist also, zu beweisen, dass $\lim_{n\to\infty} \|A_1^n\|_{op} = 0$.

[24: 16.01.2026]

[25: 20.01.2026]

$\lambda \in \mathbb{C}$ ist ein Eigenwert von $A_1 \Leftrightarrow \det(A_1 - \lambda \mathrm{Id}_N) = 0 \Leftrightarrow \lambda \cdot \det(Q - \lambda \mathrm{Id}_{N-1}) = 0 \Leftrightarrow \lambda = 0$ oder λ ist ein Eigenwert von p^t mit $\lambda \neq 1$. Daher gilt

$$r \coloneqq \sup\{|\lambda| \in \mathbb{C} \mid \lambda \text{ Eigenwert von } A_1\} < 1.$$

Insbesondere $\exists 0 < \varepsilon < 1,$, sodass $r + \epsilon < 1$. Es gilt (Beweis unten)

$$\limsup_{n\to\infty} \|A_1^n\|_{op}^{\frac{1}{n}} = r. \tag{8.7}$$

Daher $\exists n_0 \in \mathbb{N}$, sodass $\forall n \geq n_0$ gilt $\|A_1^n\|_{op}^{\frac{1}{n}} \leq r + \varepsilon < 1$ und daher auch

$$\|A_1^n\|_{op} \leq (r+\varepsilon)^n \to_{n\to\infty} 0.$$

Daraus folgt, dass $\lim_{n\to\infty} \|A_1^n\|_{op} = 0$ und damit ist die Behauptung bewiesen.

Um (8.7) zu beweisen, zeigen wir zwei Ungleichungen.
Es gilt

$$\|A_1^n\|_{op}^{\frac{1}{n}} = \left(\sup_{v\in\mathbb{C}^N,\ |v|=1} |A_1^n v|\right)^{\frac{1}{n}} \geq |\lambda| \qquad \forall \lambda \text{ Eigenwert von } \ A_1$$

und daher $\|A_1^n\|_{op}^{\frac{1}{n}} \geq r\ \forall n \in \mathbb{N}$. Daraus folgt die untere Schranke $\limsup_{n\to\infty} \|A_1^n\|_{op}^{\frac{1}{n}} \geq r$. Wir zeigen jetzt die obere Schranke $\limsup_{n\to\infty} \|A_1^n\|_{op}^{\frac{1}{n}} \leq r$. Wir zerlegen jeden Jordanblock in Q wie folgt:

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 \\ 0 & 0 & \lambda & 1 & \cdots & 0 \\ \vdots & & & & & 0 \\ \vdots & & & & & 1 \\ 0 & 0 & 0 & 0 & \cdots & \lambda \end{pmatrix} = \lambda \mathrm{Id} + \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & & 0 \\ \vdots & & & & & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Daraus folgt

$$\begin{pmatrix} 0 & 0 \\ 0 & Q \end{pmatrix} = (D+N)$$

wobei D diagonal und N nilpotent ist (d.h. $\exists k \in \mathbb{N}$ mit $N^{k+1} = 0$). Weiter gilt $DN = ND$ (Übung). Wir rechnen, mit $N^{k+1} = 0$,

$$A_1^n = U(D+N)^n U^{-1} = U\left[\sum_{j=1}^{n} \binom{n}{j} D^{n-j} N^j\right] U^{-1} = U\left[\sum_{j=1}^{k} \binom{n}{j} D^{n-j} N^j\right] U^{-1}, \qquad \forall n \geq k.$$

Die Norm $\|\cdot\|_{op}$ erfüllt $\|M_1 + M_2\|_{op} \leq \|M_1\|_{op} + \|M_2\|_{op}$ und $\|M_1 M_2\|_{op} \leq \|M_1\|_{op}\|M_2\|_{op}$ $\forall M_1, M_2 \in \mathbb{C}^{N\times N}$. Wir rechnen also, zusammen mit

$$\binom{n}{j} \leq \frac{n^j}{j!} \leq \frac{n^k}{j!} \forall j \leq k, \quad \|D\|_{op} = r, \quad \text{und} \quad \sum_{j=1}^{k} \frac{x^j}{j!} \leq \sum_{j=1}^{\infty} \frac{x^j}{j!} = e^x\ \forall x \geq 0,$$

$$\begin{aligned} \|A_1^n\|_{op} &\leq \|U\|_{op} \left\|\sum_{j=1}^{k} \binom{n}{j} D^{n-j} N^j\right\|_{op} \|U^{-1}\|_{op} \leq \|U\|_{op}\|U^{-1}\|_{op} \sum_{j=1}^{k} \binom{n}{j} \|D\|_{op}^{n-j} \|N\|_{op}^j \\ &\leq \|U\|_{op}\|U^{-1}\|_{op}\ n^k r^n \sum_{j=1}^{k} \frac{1}{j!} \left(\frac{\|N\|_{op}}{\|D\|_{op}}\right)^j \leq \|U\|_{op}\|U^{-1}\|_{op}\ n^k r^n e^{\frac{\|N\|_{op}}{\|D\|_{op}}}. \end{aligned}$$

Die gesuchte Ungleichung folgt aus

$$\lim_{n\to\infty}\left(\|U\|_{op}\|U^{-1}\|_{op}n^k r^n e^{\frac{\|N\|_{op}}{\|D\|_{op}}}\right)^{\frac{1}{n}} = r.$$

□

8.3 Irreduzible Markov-Ketten

Wir fragen uns, wann $p^k(x,y) > 0\ \forall x,y \in S$ gilt. Dafür brauchen wir eine Darstellung durch gerichtete Graphen.

Definition 8.12. *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix p. Sei $G_X = (S,E)$ der gerichtete Graph mit Knotenmenge S und Kantenmenge $E \subset S \times S$, wobei $e = (x,y) \in E$ genau dann, wenn $p(x,y) > 0$.*

(i) Sei $x,y \in S$. Ein Pfad auf G von x nach y ist eine Familie von Kanten $\gamma = (e_1,\dots,e_n)$ wobei $e_j = (x_j,y_j)$, $x_1 = x$, $y_n = y$ und $y_j = x_{j+1}\ \forall j = 1,\dots,n-1$. Wir nennen Γ_{xy} die Menge aller Pfade auf G von x nach y. Wir schreiben $|\gamma| = n$ für die Länge von γ.

(ii) Wir sagen, dass zwei unterschiedliche Knoten $x \neq y \in S$ 'kommunizieren', $x \sim y$, falls $\Gamma_{xy} \neq \varnothing$ und $\Gamma_{yx} \neq \varnothing$ (d.h., es gibt einen Pfad auf G von x nach y und einen Pfad auf G von y nach x).

Wir sagen, dass x immer mit sich selbst kommuniziert, $x \sim x$.

Lemma 8.13. *$\sim$ definiert eine Äquivalenzrelazion auf S.*

Beweis. Übung. □

Definition 8.14. *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix p.*

(i) Die Äquivalenzklassen von $(S,\sim)$ heißen kommunizierende Klassen *(oder* Klassen*).*

(ii) Die Markov-Kette X heißt irreduzibel*, falls S nur eine Klasse besitzt.*

Beispiel 1 Sei $S = \{1,2,3\}$ und

$$p = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Dann ist G_X nicht zusammenhängend und es gibt zwei Klassen $\{1,2\},\{3\}$. X ist also nicht irreduzibel.

Beispiel 2 Sei $S = \{1,2,3\}$ und

$$p = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Dann ist G_X zusammenhängend aber es gibt zwei Klassen $\{1,2\},\{3\}$. X ist also nicht irreduzibel.

Beispiel 2 Sei $S = \{1,2,3\}$ und
$$p = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$
Dann ist G_X zusammenhängend und es gibt nur eine Klasse $\{1,2,3\}$. X ist also irreduzibel.

Lemma 8.15. *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix p.*

(i) Sei $x, y \in S$ und $k \in \mathbb{N}$. Es gilt
$$p^k(x,y) > 0 \;\Leftrightarrow\; \exists \gamma \in \Gamma_{xy} \text{ mit } |\gamma| = k.$$

(ii) Sei $x \neq y \in S$. Dann gilt
$$x \sim y \;\Leftrightarrow\; \exists k = k_{xy}, k' = k'_{yx} \in \mathbb{N}, \text{ sodass } p^k(x,y) > 0 \text{ und } p^{k'}(y,x) > 0.$$

(iii) X ist irreduzibel $\Leftrightarrow \forall x, y \in S \; \exists k = k_{xy} \in \mathbb{N}$, sodass $p^k(x,y) > 0$.

Beweis. Im Fall $|S| = 1$ gelten die Aussagen trivial. Wir nehmen jetzt an $|S| > 1$.

(i) Wir rechnen
$$\begin{aligned} p^k(x,y) &= \sum_{x_1,\ldots,x_{k-1} \in S} p(x_0,x_1)p(x_1,x_2)\cdots p(x_{n-1},y) \\ &= \sum_{\gamma \in \Gamma_{xy}, |\gamma| = k} \prod_{e \in \gamma} p(x_e, y_e) > 0 \\ &\Leftrightarrow \exists \gamma \in \Gamma_{xy} \text{ mit } |\gamma| = k. \end{aligned}$$

(ii) Sei $x \neq y$. Es gilt $x \sim y \Leftrightarrow \exists \gamma \in \Gamma_{xy}, \gamma' \in \Gamma_{yx} \Leftrightarrow p^{|\gamma|}(x,y) > 0$ und $p^{|\gamma'|}(y,x) > 0$.

$(iii) \Rightarrow$ Sei X irreduzibel. Dann gilt $x \sim y \; \forall x, y \in S$.
Aus (ii) folgt, dass für $\forall x \neq y \; \exists k, k' \in N$, sodass $p^k(x,y) > 0$ und $p^{k'}(y,x) > 0$. Weiter gilt
$$p^{k+k'}(x,x) = (p^k p^{k'})(x,x) = \sum_{z \in S} p^k(x,z)p^{k'}(z,y) \geq p^k(x,y)p^{k'}(y,x) > 0$$
und
$$p^{k+k'}(y,y) = (p^{k'} p^k)(y,y) = \sum_{z \in S} p^{k'}(y,z)p^k(z,x) \geq p^{k'}(y,x)p^k(x,y) > 0.$$

$(iii) \Leftarrow$ Wir nehmen an $\forall x, y \in S \; \exists k = k_{xy} \in \mathbb{N}$, sodass $p^k(x,y) > 0$. Daher gilt, mit (ii), $\forall x \neq y \in S \; \Gamma_{xy} \neq \varnothing$ und $\Gamma_{yx} \neq \varnothing$ d.h. $x \sim y$. Also ist X irreduzibel.

□

Oben ist die Potenz $k = k_{xy}$ von x, y abhängig. Wir suchen aber ein Kriterium, um die Existenz einer Potenz k unabhängig von x, y zu garantieren. Dafür brauchen wir folgende Definition.

Definition 8.16. *Sei X eine zeitlich homogene und irreduzibel Markov-Kette mit Zustandsraum S und Übergangsmatrix p.*
Für $x \in S$ sei
$$d(x) \coloneqq ggT\{n \in \mathbb{N} |\; p^n(x,x) > 0\}.$$

(i) Der Zustand x heißt aperiodisch, *falls $d(x) = 1$.*

(ii) Die Markov-Kette X heißt aperiodisch, *falls $d(x) = 1 \; \forall x \in S$.*

Bemerkung X ist irreduzibel und daher ist die Menge $\{n \in \mathbb{N} | \ p^n(x,x) > 0\}$ nicht leer.

Lemma 8.17. *Sei X eine zeitlich homogene und irreduzible Markov-Kette mit Zustandsraum S und Übergangsmatrix p.*

(i) Sei $x \neq y \in S$. Falls $x \sim y$, dann gilt $d(x) = d(y)$.

(ii) $\forall x \in S \ \exists m = m_x \in \mathbb{N}$, sodass $p^{nd(x)}(x,x) > 0 \ \forall n \geq m_x$.

Beweis. Im Fall $|S| = 1$ gelten die Aussagen trivial. Wir nehmen jetzt an $|S| > 1$.

(i) Sei $x \neq y \in S$ mit $x \sim y$. Aus Lemma 8.15 folgt, dass $k, k', l \in \mathbb{N}$ existieren sodass $p^k(x,y) > 0$ und $p^{k'}(y,x) > 0$ und $p^l(x,x) > 0$. Wir rechnen

$$\begin{aligned}
p^{k+k'+l}(y,y) &= \sum_{zz' \in S} p^{k'}(y,z) p^l(z,z') p^k(z',y) \geq p^{k'}(y,x) p^l(x,x) p^k(x,y) > 0, \\
p^{2l}(x,x) &= \sum_{z \in S} p^l(x,z) p^l(z,x) \geq p^l(x,x) p^l(x,x) > 0, \\
p^{k+k'+2l}(y,y) &= \sum_{zz' \in S} p^{k'}(y,z) p^{2l}(z,z') p^k(z',y) \geq p^{k'}(y,x) p^{2l}(x,x) p^k(x,y) > 0.
\end{aligned}$$

Dann teilt $d(y)$ $k + k' + l$ und $k + k' + 2l$. Daher teilt $d(y)$ auch l $\forall l \in \mathbb{N}$, sodass $p^l(x,x) > 0$. Aus der Definition von $d(x)$ folgt $d(y) \leq d(x)$. Da wir das Argument auch umdrehen können, folgt genauso gut, dass $d(x) \leq d(y)$ und mithin die Behauptung.

(ii) siehe zB. das Skript von Prof. Bovier.

□

Beispiel Sei $S = \{1, 2, \ldots, 6\}$ und

$$p = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Dann ist X irreduzibel. Weiter gilt $p^l(1,1) > 0$ genau dann, wenn $l = n_1 4 + n_2 6 \ n_1, n_2 \in \mathbb{N}$ und daher $d(1) = 2$ und $m_1 = 2$, aber $p^2(1,1) = 0$.

[25: 20.01.2026]

[26: 23.01.2026]

Bemerkung Falls $p(x,x) > 0$, dann gilt $d(x) = 1$ und $p^n(x,x) > 0 \ \forall n \in \mathbb{N}$.
Falls X irreduzibel ist und $p(x,x) > 0 \ \forall x \in S$ gilt, dann ist X auch aperiodisch.

Lemma 8.18. *Sei X eine zeitlich homogene, irreduzible und aperiodische Markov-Kette mit Zustandsraum S und Übergangsmatrix p.*
Dann $\exists k \in \mathbb{N}$, sodass $p^k(x,y) > 0 \ \forall x, y \in S$.

Beweis. Da X aperiodisch ist, gilt $\forall x \in S\ \exists m_x \in \mathbb{N}$,, sodass $p^n(x,x) > 0\ \forall n \geq m_x$ (siehe Lemma 8.17(ii)). Wir definieren

$$m := \max_{x \in S} m_x.$$

Aus $|S| < \infty$ folgt, dass $m \in \mathbb{N}$ wohldefiniert ist.
Da X irreduzibel ist, existiert $\forall x \neq y \in S\ k_{xy} \in \mathbb{N}$, sodass $p^{k_{xy}}(x,y) > 0$. Wir definieren

$$\tilde{m} := \max_{x \neq y \in S} k_{xy}.$$

Aus $|S| < \infty$ folgt, dass $\tilde{m} \in \mathbb{N}$ wohldefiniert ist. Sei jetzt

$$k := m + \tilde{m}.$$

Wir zeigen, dass $p^k(x,y) > 0\ \forall x, y \in S$.
Falls $x \neq y$ gilt

$$p^k(x,y) = (p^{k_{xy}} p^{k-k_{xy}})(x,y) \geq p^{k_{xy}}(x,y) p^{k-k_{xy}}(y,y).$$

Aus $k - k_{xy} \geq m \geq m_y$ folgt $p^{k-k_{xy}}(y,y) > 0$ und daher $p^k(x,y) > 0$. Falls $x = y$, gilt $k \geq m \geq m_x$ und daher

$$p^k(x,x) > 0.$$

□

Bemerkung X irreduzibel und aperiodisch $\Rightarrow \exists k \in \mathbb{N}$, sodass $p^k(x,y) > 0\ \forall x, y \in S \Rightarrow \exists!$ invariante Verteilung μ und X ist ergodisch $\pi_n \to \mu$.
Falls X nicht aperiodisch ist, dann konvergiert π_n im Allgemeinen nicht gegen μ.

Beispiel Sei $S = \{1,2\}$ und

$$p = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Dann ist X irreduzibel. Weiter gilt

$$p^{2m} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad p^{2m+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \forall m \in \mathbb{N}.$$

Also gilt $d(1) = d(2) = 2$ und X ist nicht aperiodisch. Die Matrix besitzt genau eine invariante Verteilung (Übung)

$$\mu = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

und es gilt $\mu(x) = \frac{1}{2} > 0\ \forall x \in S$. Die Folge p^n hat aber kein Grenzwert, also ist X nicht ergodisch.

Irreduzibilität garantiert Eindeutigkeit der invarianten Verteilung auch ohne Aperiodizität.

Satz 8.19. *Sei X eine zeitlich homogene und irreduzible Markov-Kette mit Zustandsraum S und Übergangsmatrix p.*
Dann besizt X genau eine invariante Verteilung μ. Dazu gilt $\mu(x) > 0\ \forall x \in S$.

Beweis. Für $0 < \varepsilon < 1$ betrachten wir die Markov-Kette X_ε mit Zustandsraum S und Übergangsmatrix p_ε definiert durch

$$p_\varepsilon := \varepsilon \mathrm{Id} + (1-\varepsilon) p.$$

Es gilt

- $p_\varepsilon(x,x) = \varepsilon + (1-\varepsilon)p(x,x) > 0 \ \forall x \in S$,
- $p_\varepsilon(x,y) = (1-\varepsilon)p(x,y) \geq 0 \ \forall x \neq y \in S$,
- $\sum_y p_\varepsilon(x,y) = \varepsilon + (1-\varepsilon)\sum_y p(x,y) = \varepsilon + (1-\varepsilon) = 1$.

Daher ist p_ε eine stochastische Matrix mit $p_\varepsilon(x,x) > 0 \ \forall x \in S$.
Weiter gilt $p(x,y) > 0 \Rightarrow p_\varepsilon(x,y) > 0$ und daher ist die Markov-Kette X_ε irreduzibel.
Daraus folgt, mit $p_\varepsilon(x,x) > 0$, dass X_ε eine aperiodische irreduzibel Markov-Kette ist und daher besitzt X_ε genau eine invariante Verteilung μ_ε. Dazu gilt $\mu_\varepsilon(x) > 0 \ \forall x \in S$. Sei jetzt μ eine invariante Verteilung für X. Wir rechnen

$$\begin{aligned}\mu^t p = \mu^t \ \Leftrightarrow\ (1-\varepsilon)\mu^t p = (1-\varepsilon)\mu^t \ \Leftrightarrow\ \varepsilon\mu^t + (1-\varepsilon)\mu^t p = \varepsilon\mu^t + (1-\varepsilon)\mu^t = \mu^t \\ \Leftrightarrow\ \mu^t p_\varepsilon = \mu^t \ \Leftrightarrow\ \mu = \mu_\varepsilon.\end{aligned}$$

Daraus folgt $\mu = \mu_\varepsilon \ \forall < \varepsilon < 1$. Insbesondere ist μ_ε unabhängig von ε. Die invariante Verteilung μ ist also eindeutig und erfüllt $\mu(x) > 0 \ \forall x \in S$. □

8.4 Wesentliche und unwesentliche Klassen

Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix p. Wir haben bis jetzt folgende Resultaten bewiesen:

(A) $\exists$ wenigstens eine invariante Verteilung μ.

(B) Falls X irreduzibel ist, dann ist die invariante Verteilung eindeutig und $\mu(x) > 0 \ \forall x \in S$.

(C) Falls X irreduzibel und aperiodisch ist, dann ist X auch ergodisch $\lim_{n\to\infty} \pi_n = \mu$.

Falls X nicht irreduzibel ist, gibt es mehrere invariante Verteilungen. Um diese zu klassifizieren brauchen wir folgende Definition.

Definition 8.20. *Sei X eine zeitlich homogene nicht irreduzible Markov-Kette mit Zustandsraum S und Übergangsmatrix p. Seien $\{C_j\}_{j=1}^N$ die Äquivalenzklassen von S, d.h., $S = \bigcup_{j=1}^N C_j$. Da X nicht irreduzibel ist, gilt $N \geq 2$.*

(i) Die Klasse C_j heißt wesentlich, *falls $\Gamma_{xy} = \varnothing \ \forall x \in C_j, y \notin C_j$ (d.h., es gibt keinen Pfad von C_j nach C_j^c.)*

(ii) Die Klasse C_j heißt unwesentlich, *falls sie nicht wesentlich ist.*

Bemerkung Im Fall endlichen Zustandsraums können wir wesentliche Klassen auch als rekurrent, unwesentliche als transient bezeichnen. Im Fall von unendlichem Zustandsraum (den Fall haben wir aber nicht betrachtet) sind diese Begriffe aber zu unterscheiden.

Beispiel Sei $S = \{1,2,3\}$

$$p = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Es gibt zwei Klassen $C_1 = \{1,2\}, C_2 = \{3\}$. Die Klasse C_1 ist wesentlich, die Klasse C_2 ist unwesentlich.

Satz 8.21. *Sei X eine zeitlich homogene, nicht irreduzible Markov-Kette mit Zustandsraum S und Übergangsmatrix p. Seien $\{C_j\}_{j=1}^N$ die wesentlichen Äquivalenzklassen von S und $\{D_j\}_{j=1}^{N'}$ die unwesentlichen Äquivalenzklassen von S, d.h., $S = (\bigcup_{j=1}^N C_j) \bigcup (\bigcup_{j=1}^{N'} D_j)$.*

(i) Sei $C \coloneqq \bigcup_{j=1}^N C_j$ und $D \coloneqq \bigcup_{j=1}^{N'} D_j$. Dann $\forall x \in D\ \exists y \in C$ sodass $\Gamma_{xy} \neq \varnothing$.

(ii) Es gibt für jedes $j = 1, \dots, N$ genau eine invariante Verteilungen μ_j, sodass $\mu_j(x) > 0 \Leftrightarrow x \in C_j$ (d.h., μ_j hat Träger auf C_j).

(iii) Falls μ eine invariante Verteilung ist, dann gibt es eine Konvexkombination $\alpha_1, \dots, \alpha_N \in [0,1]$, sodass $\sum_{j=1}^N \alpha_j = 1$ und

$$\mu = \sum_{j=1}^N \alpha_j \mu_j.$$

[26: 23.01.2026]

[27: 27.01.2026]

Beweis.
(i) [*Dieser Teil wurde nicht in der Vorlesung besprochen.*
Sei $x \in D_1$. Die Klasse D_1 ist unwesentlich, also $\exists y_1 \notin D_1$, sodass $\Gamma_{xy_1} \neq \varnothing$. Falls $y_1 \in C$, sind wir fertig. Falls $y_1 \notin C$, gilt $y_1 \in D_{l_1}$ mit $l_1 \neq 1$. Die Klasse D_{l_1} ist unwesentlich, also $\exists y_2 \notin D_{l_1}$, sodass $\Gamma_{y_1y_2} \neq \varnothing$. Der Knoten y_2 darf nicht in D_1 liegen, sonst wäre $D_1 \cup D_{l_1}$ eine Äquivalenzklasse, daher gilt $y_2 \in C \cup (\bigcup_{j \neq 1, l_1} D_j)$. Falls $y_2 \in C$, sind wir fertig. Sonst gilt $y_2 \in D_{l_2}$ mit $l_2 \neq 1, l_1$. Wir wiederholen dies endlich oft (da nur endlich viele Äquivalenzklassen existieren), bis wir einen Punkt in C finden.]

(ii) Sei $p_j \in [0,1]^{C_j \times C_j}$ definiert durch $p_j(x,y) = p(x,y)\ \forall x, y \in C_j$. Wir zeigen, dass p_j eine stochastische Matrix ist: Da C_j eine wesentliche Klasse ist, gilt $p(x,y) = 0\ \forall x \in C_j, y \notin C_j$ und daher, für alle $x \in C_j$

$$1 = \sum_{y \in S} p(x,y) = \sum_{y \in C_j} p(x,y) = \sum_{y \in C_j} p_j(x,y).$$

Sei jetzt $X^{(j)}$ eine zeitlich homogene Markov-Kette mit Zustandsraum C_j und Übergangsmatrix p_j. Diese Markov-Kette ist irreduzibel. Aus dem Satz 8.19 folgt, dass $\exists!\ \tilde{\mu}_j \in [0,1]^{C_j}$, sodass $\sum_{x \in C_j} \tilde{\mu}_j(x) = 1$, $\tilde{\mu}_j(x) > 0\ \forall x \in C_j$ und $\tilde{\mu}_j^t p = \tilde{\mu}_j^t$. Wir definieren

$$\mu_j(x) \coloneqq \begin{cases} \tilde{\mu}_j(x) & x \in C_j \\ 0 & x \notin C_j \end{cases}$$

Dann gilt $\mu_j(x) \geq 0$ und $\sum_{x \in S} \mu_j(x) = \sum_{x \in C_j} \tilde{\mu}_j(x) = 1$. Wir zeigen, dass $\mu_j^t p = \mu_j^t$.
Falls $x \notin C_j$, rechnen wir

$$(\mu_j^t p)(x) = \sum_{y \in S} \mu_j(y) p(y,x) = \sum_{y \in C_j} \tilde{\mu}_j(y) p(y,x) = 0 = \mu_j(x),$$

da $p(y,x) = 0\ \forall y \in C_j, x \notin C_j$. Falls $x \in C_j$, rechnen wir

$$(\mu_j^t p)(x) = \sum_{y \in S} \mu_j(y) p(y,x) = \sum_{y \in C_j} \tilde{\mu}_j(y) p_j(y,x) = \tilde{\mu}_j(x) = \mu_j(x).$$

μ_j ist also eine invariante Verteilung für X mit Träger auf C_j.

Wir zeigen, dass μ_j die einzige invariante Verteilung für X mit Träger auf C_j ist. Sei μ eine invariante Verteilung für X mit Träger auf C_j, d.h. $\mu(x) > 0 \Leftrightarrow x \in C_j$.
Wir rechnen, $\sum_{x\in S} \mu(x) = \sum_{x\in C_j} \mu(x) = 1$ und, für $x \in C_j$,

$$\mu(x) = \sum_{y\in S} \mu_j(y)p(y,x) = \sum_{y\in C_j} \mu_j(y)p(y,x) = \sum_{y\in C_j} \mu_j(y)p_j(y,x).$$

$\mu_{|C_j}$ ist also eine invariante Verteilung für $X^{(j)}$ und daher (die invariante Verteilung für $X^{(j)}$ ist eindeutig) $\mu_{|C_j} = \tilde{\mu}_j$. Es folgt $\mu = \mu_j$.

(iii) Sei $\mu \in [0,1]^S$ eine invariante Verteilung für X.

Wir zeigen dass $\mu(x) = 0\ \forall x \in D := \bigcup_{l=1}^{N'} D_l$. Durch Widerspruch, sei $x_0 \in D_l$ mit $\mu(x_0) > 0$. Da D_j unwesentlich ist, folgt aus (i), dass $y_0 \in C$ und $k = k_{x_0} \in \mathbb{N}$ existieren, sodass $p^k(x_0, y_0) > 0$. Da μ eine invariante Verteilung gilt auch $\mu^t p^k = \mu^t$. Wir rechnen $\forall y \in D$

$$\mu(y) = \sum_{x\in S} \mu(x)p^k(x,y) = \sum_{x\in D} \mu(x)p^k(x,y).$$

Die zweite Gleichung gilt, weil $p^k(x,y) = 0\ \forall x \in C, y \in D$. Daher

$$0 < \sum_{y\in D} \mu(y) = \sum_{y\in D} \mu(x)[\sum_{y\in D} p^k(x,y)].$$

Wir rechnen, für $x = x_0$,

$$\sum_{y\in D} p^k(x_0,y) = 1 - \sum_{y\in C} p^k(x_0,y) \leq 1 - p^k(x_0,y_0) < 1.$$

Für alle anderen $x \in D$ gilt $\sum_{y\in D} p^k(x,y) \leq 1$. Daher

$$0 < \sum_{y\in D} \mu(y) = \sum_{y\in D} \mu(x)[\sum_{y\in D} p^k(x,y)] < \sum_{y\in D} \mu(x).$$

Das ergibt den gesuchten Widerspruch.

Sei jetzt $y \in C_j$. Wir rechnen, mit $\mu(x) = 0\ \forall x \in D$,

$$\mu(y) = \sum_{x\in S} \mu(x)p(x,y) = \sum_{x\in C} \mu(x)p(x,y) = \sum_{x\in C_j} \mu(x)p_j(x,y) + \sum_{j'\neq j}\sum_{x\in C_{j'}} \mu(x)p(x,y) = \sum_{x\in C_j} \mu(x)p_j(x,y).$$

Die letzte Gleichung gilt, weil $p(x,y) = 0\ \forall x \in C_{j'}, y \notin C_{j'}$. $\mu_{|C_j}$ ist also ein links-Eigenvektor für p_j und daher $\mu_{|C_j} = \alpha_j \tilde{\mu}_j$, mit $\alpha_j := \sum_{x\in C_j} \mu(x)$. Daher gilt

$$\mu(x) = \sum_{j=1}^{N} \alpha_j \mu_j(x).$$

Weiter gilt

$$1 = \sum_{x\in S} \mu(x) = \sum_{j=1}^{N} \alpha_j.$$

□

8.5 Markov Eigenschaft

Sei X eine Markov-Kette mit Zustandsraum S, Anfangsverteilung π_0 und Übergangsmatrizen $(p_n)_{n\in\mathbb{N}_0}$. Falls $\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0) > 0$, gilt die Markov Eigenschaft (siehe (8.1))

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

Lemma 8.22. *Sei X eine Markov-Kette mit Zustandsraum S, Anfangsverteilung π_0 und Übergangsmatrizen $(p_n)_{n\in\mathbb{N}_0}$. Wir nehmen $\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0) > 0$ an. Folgendes gilt:*

(i) $\forall k \geq 1, x_{n+1}, \ldots, x_{n+k} \in S$ *gilt*

$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1}, \ldots, X_{n+k} = x_{n+k} \mid X_n = x_n, \ldots, X_0 = x_0)& \\ = \mathbb{P}(X_{n+1} = x_{n+1} \ldots, X_{n+k} = x_{n+k} \mid X_n = x_n)& \\ = p_n(x_n, x_{n+1}) p_{n+1}(x_{n+1}, x_{n+2}) \cdots p_{n+k-1}(x_{n+k-1}, x_{n+k}).&\end{aligned}$$

Insbesondere gilt $\forall k \geq 1, x_n, \ldots, x_{n+k} \in S$

$$\mathbb{P}(X_n = x_n, \ldots, X_{n+k} = x_{n+k}) = \pi_n(x_n) \prod_{j=1}^{k} p_{n+j-1}(x_{n+j-1}, x_{n+j}).$$

(ii) $\forall k \geq 1, f\colon S^{k+1} \to \mathbb{R}$ *messbar gilt*

$$\mathbb{E}[f(X_n, \ldots, X_{n+k}) \mid X_n = x_n, \ldots, X_0 = x_0] = \mathbb{E}[f(X_n, \ldots, X_{n+k}) \mid X_n = x_n]$$

wobei, für $F\colon \Omega \to \mathbb{R}$ messbar und integrierbar und $A \in \mathcal{F}$, definieren wir

$$\mathbb{E}[F|A] := \frac{\mathbb{E}[F\mathbf{1}_A]}{\mathbb{E}[\mathbf{1}_A]}.$$

Beweis.
(i) Übung

(ii) Wir rechnen

$$\begin{aligned}&\mathbb{E}[f(X_n, \ldots, X_{n+k}) \mid X_n = x_n, \ldots, X_0 = x_0] \\ &\quad= \sum_{x_{n+1},\ldots,x_{n+k}\in S} f(x_n, x_{n+1}, \ldots, x_{n+k}) \mathbb{P}(X_{n+1} = x_{n+1} \ldots, X_{n+k} = x_{n+k} \mid X_n = x_n, \ldots, X_0 = x_0) \\ &\quad= \sum_{x_{n+1},\ldots,x_{n+k}\in S} f(x_n, x_{n+1}, \ldots, x_{n+k}) \mathbb{P}(X_{n+1} = x_{n+1} \ldots, X_{n+k} = x_{n+k} \mid X_n = x_n) \\ &\quad= \mathbb{E}[f(X_n, \ldots, X_{n+k}) \mid X_n = x_n].\end{aligned}$$

□

Lemma 8.23. *Sei X eine Markov-Kette mit Zustandsraum S, Anfangsverteilung π_0 und Übergangsmatrizen $(p_n)_{n\in\mathbb{N}_0}$. Folgendes gilt:*

(i) Sei $k \geq 1$, und $f\colon S^{k+1} \to \mathbb{R}$ messbar. Dann gilt $\forall m \geq 1$

$$\mathbb{E}[f(X_m, \ldots, X_{m+k})] = \mathbb{E}^m\left[f(X'_0, \ldots, X'_k)\right], \tag{8.8}$$

wobei $\mathbb{E}^m$ ist der Erwartungswert bezüglich einer Markov-Kette mit Zustandsraum S, Anfangsverteilung $\pi'_0 := \pi_m$ und Übergangsmatrizen $p'_n := p_{n+m}, n \geq 0$.

Für alle $x \in S$, sodass $\mathbb{P}(X_m = x) > 0$, gilt

$$\mathbb{E}[f(X_m, \ldots, X_{m+k}) \mid X_m = x] = \mathbb{E}^m\left[f(X'_0, \ldots, X'_k) | X'_0 = x\right]. \tag{8.9}$$

(ii) $\forall f\colon S^{\mathbb{N}_0} \to [0,\infty)$ *messbar, gilt*

$$\mathbb{E}\left[f(X_m, X_{m+1}, \dots)\right] = \mathbb{E}^m\left[f(X_0', X_1', \dots)\right]. \tag{8.10}$$

Für alle $x \in S$*, sodass* $\mathbb{P}(X_m = x) > 0$*, gilt*

$$\mathbb{E}\left[f(X_m, X_{m+1}, \dots) \mid X_m = x\right] = \mathbb{E}^m\left[f(X_0', X_1', \dots) | X_0' = x\right]. \tag{8.11}$$

Beweis.
(i) Um (8.8) zu beweisen, rechnen wir

$$\begin{aligned}
\mathbb{E}\left[f(X_m, \dots, X_{m+k})\right] &= \sum_{x_m,\dots,x_{m+k}\in S} f(x_m, \dots, x_{m+k})\mathbb{P}\left(X_m = x_m, \dots, X_{m+k} = x_{m+k}\right)\\
&= \sum_{x_m,\dots,x_{m+k}\in S} f(x_m, \dots, x_{m+k})\pi_m(x_m)\prod_{j=1}^{k} p_{n+j-1}(x_{n+j-1}, x_{n+j})\\
&= \sum_{x_m,\dots,x_{m+k}\in S} f(x_m, \dots, x_{m+k})\pi_0'(x_m)\prod_{j=1}^{k} p_{j-1}'(x_{n+j-1}, x_{n+j})\\
&= \mathbb{E}^m\left[f(X_0', \dots, X_k')\right].
\end{aligned}$$

Der Beweis für (8.9) ist analog.

(ii) Aus (i), folgt dass (8.10) gilt $\forall f = \mathbf{1}_A$ mit $A \in \mathcal{C}$ Zylindermenge.
Wir zeigen jetzt, dass (8.10) gilt für alle $f = \mathbf{1}_A$ mit $A \in \mathcal{B}(S)^{\mathbb{N}_0}$. Wir betrachten das Mengensystem

$$\mathcal{D} \coloneqq \{A \in \mathcal{B}(S)^{\mathbb{N}_0} \mid \text{(8.10) gilt für } f = \mathbf{1}_A\}.$$

Insbesondere gilt $\mathcal{C} \subset \mathcal{D}$. $\mathcal{D}$ ist ein Dynkin-System (Übung). Es folgt, da $\mathcal{C}$ $\cap-$stabil ist und $\sigma(\mathcal{C}) = \mathcal{B}(S)^{\mathbb{N}_0}$, dass $\mathcal{D} = \mathcal{B}(S)^{\mathbb{N}_0}$.
Die Aussage für messbare nicht-negative Funktionen folgt durch maßtheoretische Induktion. (8.11) wird analog bewiesen.

□

Wir sind jetzt bereit die Markov-Eigenschaft allgemein anzugeben.

Satz 8.24 (Markov-Eigenschaft (Version 1))**.** *Sei X eine Markov-Kette mit Zustandsraum S, Anfangsverteilung π_0 und Übergangsmatrizen $(p_n)_{n\in\mathbb{N}_0}$. Folgendes gilt:*

(i) *Für $m \geq 0$ und $k \geq 1$ seien $f\colon S^{m+1} \to \mathbb{R}$ und $g\colon S^{k+1} \to \mathbb{R}$ messbare Funktionen.*
Dann gilt

$$\mathbb{E}\left[f(X_0, \dots, X_m)g(X_m, \dots, X_{m+k})\right] = \mathbb{E}\left[f(X_0, \dots, X_m)\ \mathbb{E}^m\left[g(X_0', \dots, X_k') \mid X_0' = X_m\right]\right]$$

(ii) *Für $m \geq 0$ seien $f\colon S^{m+1} \to [0,\infty)$ und $g\colon S^{\mathbb{N}_0} \to [0,\infty)$ messbare Funktionen.*
Dann gilt

$$\mathbb{E}\left[f(X_0, \dots, X_m)g(X_m, X_{m+1}, \dots)\right] = \mathbb{E}\left[f(X_0, \dots, X_m)\ \mathbb{E}^m\left[g(X_0', X_2', \dots) \mid X_0' = X_m\right]\right]$$

Beweis.
(i) Wir rechnen mit $\mathbb{E}^m\left[g(X'_0,\dots,X_{k'})|X'_0 = x_m\right] = h(x_m)$

$$
\begin{aligned}
&\mathbb{E}\left[f(X_0,\dots,X_m)g(X_m,X_{m+1},\dots)\right]\\
&= \sum_{x_0,\dots,x_m\in S} f(x_m,\dots,x_{m+k})\ \mathbb{E}\left[g(X_m,\dots,X_{m+k})|X_0 = x_0,\dots,X_m = x_m\right]\ \mathbb{P}\left(X_0 = x_0,\dots,X_m = x_m\right)\\
&= \sum_{x_0,\dots,x_m\in S} f(x_m,\dots,x_{m+k})\ \mathbb{E}\left[g(X_m,\dots,X_{m+k})|X_m = x_m\right]\ \mathbb{P}\left(X_0 = x_0,\dots,X_m = x_m\right)\\
&= \sum_{x_0,\dots,x_m\in S} f(x_m,\dots,x_{m+k})\ \mathbb{E}^m\left[g(X'_0,\dots,X_{k'})|X'_0 = x_m\right]\ \mathbb{P}\left(X_0 = x_0,\dots,X_m = x_m\right)\\
&= \sum_{x_0,\dots,x_m\in S} f(x_m,\dots,x_{m+k})h(x_m)\ \mathbb{P}\left(X_0 = x_0,\dots,X_m = x_m\right)\\
&= \mathbb{E}\left[f(X_n,\dots,X_{n+k})h(X_m)\right] = \mathbb{E}\left[f(X_0,\dots,X_m)\ \mathbb{E}^m\left[g(X'_0,\dots,X'_k)\mid X'_0 = X_m\right]\right].
\end{aligned}
$$

(ii) Übung.

□

8.6 Starke Markov Eigenschaft

Wir formulieren jetzt die Markov Eigenschaft für zufällige Zeit m.
In diesem Abschnitt betrachten wir nur *zeitlich homogene* Markov-Ketten mit Angangsverteilung $\pi_0 = \delta_x$ für ein $x \in S$. Die zugehörige Wahrscheinlichkeit/Erwartungswert werden mit $\mathbb{P}_x, \mathbb{E}_x$ notiert. Insbesondere gilt

$$\mathbb{E}_x[f(X)|\, X_0 = x] = \mathbb{E}_x[f(X)].$$

Erinnerung: Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

- Sei $f\colon \Omega \to \mathbb{R}$ eine Funktion. Die aus f erzeugte $\sigma-$Algebra ist $\sigma(f) = f^{-1}(\mathcal{B}(\mathbb{R}))$.
  f ist $\mathcal{F} - \mathcal{B}(\mathbb{R})-$messbar $\Leftrightarrow \sigma(f) \subset \mathcal{F}$.
- Sei $X\colon \Omega \to S^{\mathbb{N}_0}$ messbar. Dann ist die Funktion $(X_0,\dots,X_n)\colon \Omega \to S^{n+1}$ auch messbar.

Die Information, die über einen stochastischen Prozess $(X_n)_{n\in\mathbb{N}_0}$ bis zur Zeit n vorliegt, wird beschrieben durch die $\sigma-$Algebra

$$\mathcal{F}_n \coloneqq \sigma(X_0,\dots,X_n).$$

Damit können wir die Markov-Eigenschaft ein bisschen abstrakter formulieren.

Satz 8.25 (Markov-Eigenschaft (Version 2))**.** *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S, Anfangsverteilung $\pi_0 = \delta_x$, $x \in S$ und Übergangsmatrix p. Für $m \in \mathbb{N}_0$ definieren wir den Shiftoperator $\theta_m\colon S^{\mathbb{N}_0} \to S^{\mathbb{N}_0}$ durch*

$$\theta_m(X_0, X_1, \dots) \coloneqq (X_m, X_{m+1}, \dots).$$

Seien $f\colon S^{\mathbb{N}_0} \to [0,\infty)$ und $g\colon S^{\mathbb{N}_0} \to [0,\infty)$ messbare Funktionen.
Falls $\omega \mapsto f(X(\omega))$ $\mathcal{F}_n - \mathcal{B}(\mathbb{R})-$messbar ist, gilt

$$\mathbb{E}_x\left[f(X)g(\theta_m(X))\right] = \mathbb{E}_x\left[f(X)\ \mathbb{E}_{X_m}\left[g(X')\right]\right]$$

Beweis.
Aus $\omega \mapsto f(X(\omega))$ $\mathcal{F}_n - \mathcal{B}(\mathbb{R})-$messbar folgt, dass $f(X) = F(X_0, \dots, X_n)$. Wir rechnen

$$\begin{aligned}
\mathbb{E}_x\left[f(X)g(\theta_m(X))\right] &= \mathbb{E}_x\left[F(X_0, \dots, X_m)g(X_m, X_{m+1}, \dots)\right] \\
&= \mathbb{E}_x\left[F(X_0, \dots, X_m)\mathbb{E}^m\left[g(X') \middle| X_0' = X_m\right]\right] \\
&= \mathbb{E}_x\left[F(X_0, \dots, X_m)\mathbb{E}\left[g(X') \middle| X_0' = X_m\right]\right] \\
&= \mathbb{E}_x\left[F(X_0, \dots, X_m)\mathbb{E}_{X_m}\left[g(X')\right]\right],
\end{aligned}$$

wobei $\mathbb{E}^m = \mathbb{E}$, weil die Markov-Kette zeitliche homogen ist. □

[27: 27.01.2026]

[28: 30.01.2026]

Definition 8.26 (Stoppzeit)**.** *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix p. Eine messbare Funktion*

$$T\colon \Omega \to \mathbb{N}_0 \cup \{\infty\},$$

bzw. die Familie $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$, heißt Stoppzeit*, falls*

$$\{T = n\} = \{\omega \in \Omega |\ T(\omega) = n\} \in \mathcal{F}_n \qquad \forall n \in \mathbb{N}_0.$$

Beispiel Sei $D \subset S$.

1. Die erste Eintrittszeit (oder Rückkehrzeit) in D ist definiert durch

$$T_D(\omega) \coloneqq \min\{n \geq 1 |\ X_n(\omega) \in D\}$$

mit der Konvention $\min \varnothing \coloneqq +\infty$.
Für $n = 0$ gilt $\{T_D = 0\} = \varnothing \in \mathcal{F}_0$.
Für $n \geq 1$ gilt $\{T_D = n\} = \{X_n \in D\} \cap \left(\bigcap_{j=1}^{n-1}\{X_j \notin D\}\right) \in \mathcal{F}_n$.
Für $n = \infty$ gilt $\{T_D = \infty\} = \bigcap_{j=1}^{\infty}\{X_j \notin D\} \in \mathcal{F}$.
Die Funktion ist also messbar und eine Stoppzeit.

2. Die letzte Austrittstzeit aus D ist definiert durch

$$L_D(\omega) \coloneqq \sup\{n \geq 0 |\ X_n(\omega) \in D\}$$

mit der Konvention $\sup \varnothing \coloneqq 0$.
Für $n = 0$ gilt $\{L_D = 0\} = \left(\bigcap_{j=1}^{\infty}\{X_j \notin D\}\right) \in \mathcal{F} \setminus \mathcal{F}_0$.
Für $n \geq 1$ gilt $\{L_D = n\} = \{X_n \in D\} \cap \left(\bigcap_{j=n+1}^{\infty}\{X_j \notin D\}\right) \in \mathcal{F} \setminus \mathcal{F}_n$.
Für $n = \infty$ gilt $\{L_D = \infty\} = \bigcap_{n=0}^{\infty}\bigcup_{m\geq n}\{X_m \in D\} = \limsup_{n\to\infty}\{X_n \in D\} \in \mathcal{F}$.
Die Funktion ist also messbar aber keine Stoppzeit.

Lemma 8.27. *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S, und Übergangsmatrix p, und sei $T\colon \Omega \to \mathbb{N}_0 \cup \{\infty\}$ eine Stoppzeit. Wir betrachten das Mengensystem*

$$\mathcal{F}_T \coloneqq \{A \in \mathcal{F} | \forall n \in \mathbb{N}_0 \text{ gilt } A \cap \{T = n\} \in \mathcal{F}_n\} \subset \mathcal{F}. \tag{8.12}$$

Folgendes gilt:

(i) $\mathcal{F}_T$ *ist eine* $\sigma-$*Algebra.*

(ii) Eine Funktion $f\colon \Omega \to \mathbb{R}$ *ist* $\mathcal{F}_T - \mathcal{B}(\mathbb{R})-$*messbar* $\Leftrightarrow \forall n \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R})$ *gilt* $\{f(\omega) \in B\} \cap \{T(\omega) = n\} \in \mathcal{F}_n$

$\mathcal{F}_T$ ist also das Analog von $\mathcal{F}_n$ für eine zufällige Zeit n.

Beweis.
(i) Übung

(ii) f ist $\mathcal{F}_T - \mathcal{B}(\mathbb{R})-$messbar $\Leftrightarrow \forall B \in \mathcal{B}(\mathbb{R})\ f^{-1}(B) \in \mathcal{F}_T$
$\Leftrightarrow \forall B \in \mathcal{B}(\mathbb{R}), n \in \mathbb{N}_0\ f^{-1}(B) \cap \{T = n\} \in \mathcal{F}_n\}$
$\Leftrightarrow \forall B \in \mathcal{B}(\mathbb{R}), n \in \mathbb{N}_0\ \{f(\omega) \in B\} \cap \{T(\omega) = n\} \in \mathcal{F}_n\}$

□

Satz 8.28 (starke Markov Eigenschaft)**.** *Sei X eine zeitlich homogene Markov-Kette mit Zustandsraum S, Anfangsverteilung $\pi_0 = \delta_x$ und Übergangsmatrix p.*
Sei $T\colon \Omega \to \mathbb{N}_0 \cup \{\infty\}$ eine Stoppzeit.

Für alle Funktionen $f, g\colon S^{\mathbb{N}_0} \to [0, \infty)$ messbar, sodass $\omega \mapsto f(X(\omega))$ $\mathcal{F}_T - \mathcal{B}(\mathbb{R})-$messbar ist, gilt

$$\mathbb{E}_x [\mathbf{1}_{T<\infty} f(X) g(\theta_T(X))] = \mathbb{E}_x \left[\mathbf{1}_{T<\infty} f(X)\ \mathbb{E}_{X_T} \left[g(X')\right]\right].$$

Beweis. Wir rechnen

$$\mathbb{E}_x [\mathbf{1}_{T<\infty} f(X) g(\theta_T(X))] = \sum_{m \in \mathbb{N}_0} \mathbb{E}_x [\mathbf{1}_{T=m} f(X) g(\theta_m(X))].$$

Die Funktion $\omega \mapsto \mathbf{1}_{T(\omega)=m} f(X(\omega))$ ist $\mathcal{F}_m - \mathcal{B}(R)-$messbar, und daher $\mathbf{1}_{T=m} f(X) = h(X_0, \dots, X_m)$. Daraus folgt, mit der Markov Eingeschaft,

$$\begin{aligned} \mathbb{E}_x [\mathbf{1}_{T<\infty} f(X) g(\theta_T(X))] &= \sum_{m \in \mathbb{N}_0} \mathbb{E}_x [\mathbf{1}_{T=m} f(X) g(\theta_m(X))] \\ &= \sum_{m \in \mathbb{N}_0} \mathbb{E}_x [h(X) g(\theta_m(X))] = \sum_{m \in \mathbb{N}_0} \mathbb{E}_x \left[h(X) \mathbb{E}_{X_m} \left[g(X')\right]\right] \\ &= \sum_{m \in \mathbb{N}_0} \mathbb{E}_x \left[\mathbf{1}_{T=m} f(X) \mathbb{E}_{X_m} \left[g(X')\right]\right] = \mathbb{E}_x \left[\mathbf{1}_{T<\infty} f(X)\ \mathbb{E}_{X_T} \left[g(X')\right]\right]. \end{aligned}$$

□

Eine Anwendung der starken Markoveigenschaft liefert eine neue Interpretation der invarianten Verteilung.

Satz 8.29. *Sei X eine zeitlich homogene und irreduzibel Markov-Kette mit Zustandsraum S und Übergangsmatrix p. Sei $\mu \in (0,1)^S$ die eindeutige invariante Verteilung für X.*

(i) Es gilt $\forall y \in S \ 1 \leq \mathbb{E}_y[\tau_y] < \infty$, *wobei* $\tau_y \colon \Omega \to \mathbb{N}_0 \cup \{\infty\}$ *die erste Rückkehrzeit nach* y *ist:*

$$\tau_y(\omega) = T_{\{y\}}(\omega) \coloneqq \min\{n \geq 1 | \, X_n = y\}, \qquad \textit{mit} \quad \min \varnothing \coloneqq +\infty..$$

(ii) $\forall x, y \in S$ *gilt*

$$\mu(x) = \frac{\mathbb{E}_y\left[\sum_{n=1}^{\tau_y} \mathbf{1}_{X_n=x}\right]}{\mathbb{E}_y[\tau_y]} \tag{8.13}$$

Bemerkung Im Fall $x \neq y$ die Summe $\sum_{n=1}^{\tau_y} \mathbf{1}_{X_n=x}$ ist der Anzahl von Besuchen in x vor den ersten Rückkehr nach y.

Beweis. später. □

Korollar 8.30. *Sei* X *eine zeitlich homogene und irreduzibel Markov-Kette mit Zustandsraum* S *und Übergangsmatrix* p. *Sei* $\mu \in (0,1)^S$ *die eindeutige invariante Verteilung für* X.
Dann gilt

$$\mu(x) = \frac{1}{\mathbb{E}_x[\tau_x]} \qquad \forall x \in S.$$

Beweis. Wir setzen $y = x$ in (8.13)

$$\mu(x) = \frac{\mathbb{E}_x\left[\sum_{n=1}^{\tau_x} \mathbf{1}_{X_n=x}\right]}{\mathbb{E}_x[\tau_x]}.$$

Aus der Definition von τ_x folgt, dass $\sum_{n=1}^{\tau_x} \mathbf{1}_{X_n=x} = \mathbf{1}_{X_{\tau_x}=x} = 1$ und daher gilt $\mathbb{E}_x\left[\sum_{n=1}^{\tau_x} \mathbf{1}_{X_n=x}\right] = \mathbb{E}_x[1] = 1$. □

Beweis von Satz 8.29.
(i) [*Dieser Teil wurde nicht in der Vorlesung besprochen.*

• Wir zeigen, dass ein $k \geq 1$ und $1 > c > 0$ existieren, sodass $\mathbb{P}_y(\tau_y > nk) \leq (1-c)^n \ \forall n \geq 1$.

Nehmen wir an, die MK ist aperiodisch. Dann $\exists k \geq 1$, sodass $p^k(x,y) > 0 \ \forall x, y$. Insbesondere gilt

$$c \coloneqq \min_{z \in S} p^k(z,y) > 0.$$

Wir rechnen

$$\begin{aligned}
&\mathbb{P}_y(\tau_y > nk) = \mathbb{P}_y(X_j \neq y, j = 1, \ldots, nk) \leq \mathbb{P}_y(X_{lk} \neq y, l = 1, \ldots, n) \\
&= \sum_{z_1, \ldots, z_n \in S \setminus \{y\}} p^k(y, z_1) \cdots p^k(z_{n-1}, z_n) = \sum_{z_1, \ldots, z_{n-1} \in S \setminus \{y\}} p^k(y, z_1) \cdots p^k(z_{n-2}, z_{n-1}) \left(1 - p^k(z_{n-1}, y)\right) \\
&\leq (1-c) \sum_{z_1, \ldots, z_{n-1} \in S \setminus \{y\}} p^k(y, z_1) \cdots p^k(z_{n-2}, z_{n-1}) \leq (1-c)^n.
\end{aligned}$$

Falls die MK nicht aperiodisch ist, gibt es trotzdem $\forall z \in S$ ein $k_z \geq 1$ sodass $p^{k_z}(z,y) > 0$. Dann gilt

$$c \coloneqq \min_{z \in S} p^{k_z}(z,y) > 0$$

Sei $k \coloneqq \max_{z \in S} k_z$. Wir rechnen, für alle $z \in S$,

$$\mathbb{P}_z(\tau_y > k) = \mathbb{P}_z(X_j \neq y, j = 1, \ldots, k) \leq \mathbb{P}_z(X_{k_z} \neq y) = (1 - p^{k_z}(z,y)) \leq (1-c).$$

Daraus folgt

$$\begin{aligned}
\mathbb{P}_y(\tau_y > nk) &= \sum_{z\in S\setminus\{y\}} \mathbb{P}_y\Big(\tau_y > (n-1)k,\ X_{(n-1)k} = z\Big)\mathbb{P}_z\Big(X'_j \neq y, j = 1,\dots,k\Big)\\
&\le (1-c) \sum_{z\in S\setminus\{y\}} \mathbb{P}_y\Big(\tau_y > (n-1)k,\ X_{(n-1)k} = z\Big)\\
&= (1-c)\mathbb{P}_y(\tau_y > (n-1)k) \ \le\ (1-c)^n.
\end{aligned}$$

• Wir zeigen, dass $\mathbb{P}_y(\tau_y = \infty) = 0$.

Die Aussage folgt aus $\mathbb{P}_y(\tau_y = \infty) = \lim_{n\to\infty} \mathbb{P}_y(\tau_y > nk) \le \lim_{n\to\infty}(1-c)^n = 0$.

• Aus $\mathbb{P}_y(\tau_y = \infty) = 0$ folgt, dass

$$\begin{aligned}
\mathbb{E}_y[\tau_y] &= \sum_{j\ge1} j\ \mathbb{P}_y(\tau_y = j) = \sum_{n\ge0}\left(\sum_{j=nk+1}^{(n+1)k} j\ \mathbb{P}_y(\tau_y = j)\right)\\
&\le \sum_{n\ge0}(n+1)k\ \mathbb{P}_y\Big(nk < \tau_y \le (n+1)k\Big)\\
&\le \sum_{n\ge0}(n+1)k\ \mathbb{P}_y(nk < \tau_y) \le k\sum_{n\ge0}(n+1)(1-c)^n < \infty.
\end{aligned}$$

]

(ii) Wir definieren, für $x, y \in S$,

$$\nu_y(x) := \mathbb{E}_x\left[\sum_{n=1}^{\tau_y} \mathbf{1}_{X_n=x}\right], \qquad \mu_y(x) := \frac{\nu_y(x)}{\mathbb{E}_y[\tau_y]}.$$

Unser Ziel ist, zu zeigen, dass

- $0 \le \nu_y(x) \le \mathbb{E}_y[\tau_y]$ und $\sum_{x\in S}\nu_y(x) = \mathbb{E}_y[\tau_y]$. Somit ist μ_y eine Verteilung.
- $\nu_y^t p = \nu_y^t$ und deshalb ist μ_y eine invariante Verteilung.

Die Aussage dann folgt aus der Eindeutigkeit der invarianten Verteilung.

Aus $\mathbb{P}_y(\tau_y = \infty) = 0$ folgt, dass $\tau_y < \infty$ fast sicher. Die erste Aussage folgt aus $\sum_{n=1}^{\tau_y}\mathbf{1}_{X_n=x} \le \tau_y$. Weiter gilt, mit $\sum_{x\in S}\mathbf{1}_{X_n=x} = 1$,

$$\sum_{x\in S}\nu_y(x) = \mathbb{E}_y\left[\sum_{n=1}^{\tau_y}\sum_{x\in S}\mathbf{1}_{X_n=x}\right] = \mathbb{E}_y\left[\tau_y\right].$$

Wir zeigen jetzt, dass $\nu_y(x) = \sum_{z\in S}\nu_y(z)p(z,y)$ gilt.
Wir rechnen, mit $1 = \sum_{z\in S}\mathbf{1}_{X_{n-1}=z}$,

$$\begin{aligned}
\nu_y(x) = \mathbb{E}_y\left[\sum_{n=1}^{\tau_y}\mathbf{1}_{X_n=x}\right] &= \sum_{n=1}^{\infty}\mathbb{E}_y\left[\mathbf{1}_{X_n=x}\mathbf{1}_{n\le\tau_y}\right]\\
&= \sum_{n=1}^{\infty}\sum_{z\in S}\mathbb{E}_y\left[\mathbf{1}_{X_{n-1}=z}\mathbf{1}_{n\le\tau_y}\mathbf{1}_{X_n=x}\right].
\end{aligned}$$

Die Funktion $f := \mathbf{1}_{X_{n-1}=z}\mathbf{1}_{n\le\tau_y}$ ist $\mathcal{F}_{n-1} - \mathcal{B}(\mathbb{R})-$messbar. Sei $g(X) := \mathbf{1}_{X_1=x}$.

Aus der Markov Eigenschaft folgt

$$\begin{aligned}\mathbb{E}_y\left[\mathbf{1}_{X_{n-1}=z}\mathbf{1}_{n\leq\tau_y}\mathbf{1}_{X_n=x}\right] &= \mathbb{E}_y\left[f(X)g(\theta_{n-1}(X))\right] = \mathbb{E}_y\left[f(X)\,\mathbb{E}_{X_{n-1}}\left[g(X')\right]\right] \\ &= \mathbb{E}_y\left[\mathbf{1}_{X_{n-1}=z}\mathbf{1}_{n\leq\tau_y}\mathbb{E}_z\left[\mathbf{1}_{X_1=x}\right]\right] = \mathbb{E}_y\left[\mathbf{1}_{X_{n-1}=z}\mathbf{1}_{n\leq\tau_y}\right]\ p(z,x).\end{aligned}$$

Daraus folgt

$$\nu_y(x) = \sum_{n=1}^{\infty}\sum_{z\in S} p(z,x)\ \mathbb{E}_y\left[\mathbf{1}_{X_{n-1}=z}\mathbf{1}_{n\leq\tau_y}\right] = \sum_{z\in S} p(z,x)\ \mathbb{E}_y\left[\sum_{n=1}^{\tau_y}\mathbf{1}_{X_{n-1}=z}\right].$$

Die Aussage folgt jetzt aus

$$\begin{aligned}\mathbb{E}_y\left[\sum_{n=1}^{\tau_y}\mathbf{1}_{X_{n-1}=z}\right] &= \mathbb{E}_y\left[\sum_{n=0}^{\tau_y-1}\mathbf{1}_{X_n=z}\right] = \nu_y(z) + \mathbb{E}_y\left[\mathbf{1}_{X_0=z}\right] - \mathbb{E}_y\left[\mathbf{1}_{X_{\tau_y}=z}\right] \\ &= \nu_y(z) + \delta_{zy} - \delta_{zy} = \nu_y(z).\end{aligned}$$

□

[28: 30.01.2026]

[29: 03.02.2026]

9 Ein paar Anwendungen in der Statistik

9.1 Einleitung

In der Wahrscheinlichkeitstheorie erforscht man für einen gegebenen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ die Eigenschaften von n Zufallsvariablen $X_1, \ldots, X_n$. In der Statistik hingegen ist eine Folge von Beobachtungen (also Ausgänge von Zufallsexperimenten) gegeben, $z_1, \ldots, z_n$, die man als Realisierungen von n Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, interpretieren möchte. Das Ziel ist es, Informationen über die unbekannte gemeinsame Verteilung der entsprechenden n Zufallsvariablen herauszufinden.

Beispiel 1. Der Wirkungsgrad eines neuen Medikament ist der Parameter $\theta \in [0,1]$. Um θ zu bestimmen, testen wir das Medikament auf n Patienten. Das Ergebnis sind n Zahlen $x_1, \ldots, x_n \in \{0,1\}$ wobei $x_i = 1$, falls der Patient geheilt wird.
Wir machen folgende Modellannahme: $x_1, \ldots, x_n$ sind Realisierungen von n unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \sim Ber(\theta)$.

- Frage 1: Können wir θ aus den Werten $x_1, \ldots, x_n$ schätzen? Dafür brauchen wir eine Funktion $T_n(x_1, \ldots, x_n)$, die θ "schätzt".
- Frage 2: Wie groß muss n sein, sodass $|T_n(x_1, \ldots, x_n) - \theta|$ klein genug ist? Wir müssen einen Kompromiss zwischen den Kosten (Anzahl n von Messungen) und der Sicherheit (wie sicher sind wir, eine gute Abschätzung für θ zu haben?) finden.

Definition 9.1 (Schätzer)**.** *Sei S ein topologischer Raum, $(\Omega, \mathcal{F}, \mathbb{P})$ ein (unbekannter) Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von $S-$wertigen Zufallsvariablen, dessen Verteilungen durch den $\mathbb{R}^k$-wertigen Parameter $\theta \in \mathbb{R}^k$ parametrisiert sind.*
Ein Schätzer *für θ ist eine Familie $(T_n)_{n\in\mathbb{N}}$ von messbaren Funktionen $T_n \colon S^n \to \mathbb{R}^k$.*

Mit dieser Definition ist jede messbare Funktion ein Schätzer. Im Beispiel 1 oben gilt $S = \{0, 1\}$ und folgende zwei Funktionen sind beide Schätzer für den Wirkungsgrad $\theta \in [0, 1]$:

$$T_n^{(1)}(x_1, \dots, x_n) \coloneqq \frac{1}{n} \sum_{j=1}^{n} x_j$$
$$T_n^{(2)}(x_1, \dots, x_n) \coloneqq \frac{1}{2}.$$

Die erste Funktion sieht vernünftig aus, da $\mathbb{E}[X_1] = \theta$. Die zweite Funktion ist kein guter Schätzer, da sie unabhängig von $(x_1, \dots, x_n)$ ist. Die minimale Bedingung, um einen guten Schätzer zu haben, ist Konsistenz.

Definition 9.2 (Konsistenz)**.** *Sei S ein topologischer Raum, $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von $S-$wertigen Zufallsvariablen, und $T_n \colon S^n \to \mathbb{R}^k$, $n \in \mathbb{N}$ ein Schätzer für den Parameter $\theta \in \mathbb{R}^k$.*
Der Schätzer T_n heißt konsistent*, falls die $\mathbb{R}^k-$wertigen Zufallsvariablen $\omega \to T_n(X_1(\omega), \dots, X_n(\omega))$ folgende Bedingungen erfüllen:*

$$\lim_{n\to\infty} T_n(X_1(\omega), \dots, X_n(\omega)) = \theta \qquad \mathbb{P} - f.s.$$

Im Beispiel 1 oben gilt aufgrund des Gesetzes der großen Zahlen

$$T_n^{(1)}(X_1, \dots, X_n) = \frac{1}{n} \sum_{j=1}^{n} X_j \overset{n\to\infty}{\to} \mathbb{E}[X_1] = \theta \qquad f.s.$$

und daher ist $T_n^{(1)}$ ein konsistenter Schätzer. $T_n^{(2)}$ ist kein konsistenter Schätzer.

Folgende zusätzliche Bedingung ist oft erwünscht.

Definition 9.3 (Erwartungtreue)**.** *Sei S ein topologischer Raum, $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von $S-$wertigen Zufallsvariablen, und $T_n \colon S^n \to \mathbb{R}^k$, $n \in \mathbb{N}$ ein Schätzer für den Parameter $\theta \in \mathbb{R}^k$.*
Der Schätzer T_n heißt erwartungstreu*, falls die $\mathbb{R}^k-$wertigen Zufallsvariablen $\omega \mapsto T_n(X_1(\omega), \dots, X_n(\omega))$*

$$\mathbb{E}\left[T_n(X_1(\omega), \dots, X_n(\omega))\right] = \theta \qquad \forall n \in \mathbb{N}$$

erfüllen.

Im Beispiel 1 oben ist der Schätzer $T_n^{(1)}$ erwartungstreu, da

$$\mathbb{E}\left[\frac{1}{n} \sum_{j=1}^{n} X_j\right] = \mathbb{E}[X_1] = \theta.$$

9.2 Schätzer für Produktmaßen

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \sim \nu$, wobei das Wahrscheinlichkeitsmaß ν auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ unbekannt ist.
Sei $A \in \mathcal{B}(\mathbb{R})$ gegeben. Die Wahrscheinlichkeit $\nu(A)$ kann durch die Frequenz der Beobachtungen in A geschätzt werden. Wir definieren

$$\nu_n(x_1, \dots, x_n) \coloneqq \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_A(x_j) = \int_{\mathbb{R}} \mathbf{1}_A \, d\nu_n$$

wobei

$$\nu_n(x_1, \ldots, x_n) \coloneqq \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j}.$$

Insbesondere ist

$$\nu_n(X_1(\omega), \ldots, X_n(\omega)) \coloneqq \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}$$

ein zufälliges Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
$\nu_n(X_1(\omega), \ldots, X_n(\omega))$ heißt *empirische Verteilung* von ν.

Lemma 9.4. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \sim \nu$.*
Für jedes $A \in \mathcal{B}(\mathbb{R})$ ist die Funktionenfamilie $(\nu_n)_{n\in\mathbb{N}}$ definiert durch

$$(x_1, \ldots, x_n) \mapsto \nu_n(x_1, \ldots, x_n) \coloneqq \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_A(x_j)$$

ein konsistenter und erwartungstreuer Schätzer für den Parameter $\nu(A)$.

Beweis. Wir schreiben

$$\nu_n(\omega) \coloneqq \nu_n(X_1(\omega), \ldots, X_n(\omega))(A). \tag{9.1}$$

• ν_n ist konsistent:
Die Funktion $\omega \mapsto Z_i(\omega) \coloneqq \mathbf{1}_A(X_j(\omega))$ ist eine Zufallsvariable mit Werten in $\{0, 1\}$ und $\mathbb{E}[Z_j] = \mathbb{E}[\mathbf{1}_A(X_j)] = \nu(A)$. Weiter sind die Zufallsvariablen $(Z_1, \ldots, Z_n)$ unabhängig und identisch verteilt. Aus dem Gesetz der großen Zahlen folgt

$$\nu_n(X_1, \ldots, X_n)(A) = \frac{1}{n} \sum_{j=1}^{n} Z_j \overset{n\to\infty}{\to} \mathbb{E}[Z_1] = \nu(A) \qquad f.s.$$

• ν_n ist erwartungstreu:

$$\mathbb{E}[\nu_n(X_1, \ldots, X_n)(A)] = \mathbb{E}\left[\frac{1}{n} \sum_{j=1}^{n} Z_j\right] = \nu(A) \qquad \forall n \in \mathbb{N}.$$

□

Bemerkung Wir haben keine Annahme über die Integrierbarkeit von X_1 gebraucht.

Das Maß ν wird durch die Verteilungsfunktion $t \mapsto F(t) \coloneqq \nu(-\infty, t])$ bestimmt. Wegen Lemma 9.4, definiert, für jede $t \in \mathbb{R}$, die Zufallsvariable $F_n^w(t) \coloneqq \nu_n(\omega)((-\infty, t])$ einen Schätzer für $F(t)$.

Satz 9.5. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Verteilung $X_1 \sim \nu$ und Verteilungsfunktion $F(t) = \nu((-\infty, t])$.*
Für $A \in \mathcal{B}(\mathbb{R})$ sei $\omega \mapsto \nu_n(\omega)(A)$ definiert wie in (9.1) und $F_n^w(t) \coloneqq \nu_n(\omega)((-\infty, t])$.

(i) $\exists \overline{\Omega} \subset \Omega$, *sodass* $\mathbb{P}(\overline{\Omega}) = 1$ *und* $\forall \omega \in \overline{\Omega}$ *ist die Folge* F_n^ω *schwach konvergent gegen* F *(F_n ist ein konsistenter Schätzer für F).*

(ii) $\mathbb{E}[F_n^\omega(t)] = F(t)$ $\forall n \in \mathbb{N}$ $\forall t \in \mathbb{R}$ *(F_n ist erwartungstreu).*

Beweis.
(i) Aus Lemma 9.4 folgt, dass $\forall t \in \mathbb{R}\ \exists \Omega_t \subset \Omega$, sodass $\mathbb{P}(\Omega_t) = 1$ und

$$F_n^\omega(t) \overset{n\to\infty}{\to} F(t) \qquad \forall \omega \in \Omega_t.$$

Wir definieren

$$\overline{\Omega} := \bigcap_{q\in\mathbb{Q}} \Omega_q.$$

Dann gilt $\overline{\Omega} \in \mathcal{F}$, $\mathbb{P}(\overline{\Omega}) = 1$ und, $\forall \omega \in \overline{\Omega}$ gegeben,

$$F_n^\omega(t) \overset{n\to\infty}{\to} F(t) \qquad \forall t \in \mathbb{Q}.$$

Die Funktion $t \mapsto F_n^\omega(t)$ ist monoton wachsend und deshalb gilt $F_n^\omega(t) \overset{n\to\infty}{\to} F(t)$ für alle $t \in \mathbb{R}$, welche Stetigkeitspunkte von F sind. Damit ist (i) bewiesen.

(ii) folgt aus

$$\mathbb{E}[F_n^\omega(t)] = \mathbb{E}[\nu_n((-\infty,t])] = \nu((-\infty,t]) = F(t).$$

□

Bemerkung Es ist leider nicht praktikabel, $F(t) = \nu((-\infty,t])$ für alle $t \in \mathbb{Q}$ zu schätzen, außer wir machen einige zusätzliche Annahmen über ν, und daher über F.
Das nächste Resultat gibt eine Schätzung über den relativen Fehler $\frac{|\nu_n(\omega)(A)-\nu(A)|}{\nu(A)}$ für eine gegebene Borelmenge A.

Lemma 9.6. *Sei* $(\Omega,\mathcal{F},\mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $(X_n)_{n\in\mathbb{N}}$ *eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit* $X_1 \sim \nu$*. Für* $A \in \mathcal{B}(\mathbb{R})$ *sei* $\omega \mapsto \nu_n(\omega)(A)$ *definiert wie in* (9.1)*.*
Falls $\nu(A) > 0$*, dann gilt*

$$\mathbb{P}\left(\frac{|\nu_n(\omega)(A)-\nu(A)|}{\nu(A)} \geq \varepsilon\right) \leq \frac{1}{n\nu(A)\varepsilon^2}. \tag{9.2}$$

Bemerkung $n\nu(A)$ ist der erwartete Anzahl von Beobachtungen in A. Falls $\nu(A) \ll 1$, dann brauchen wir viele Messungen, um $n\nu(A)$ groß zu machen.

Beweis. Aus $\mathbb{E}[\nu_n(A)] = \nu(A)$ und Tschebishev Ungleichung folgt

$$\mathbb{P}\left(\frac{|\nu_n(\omega)(A)-\nu(A)|}{\nu(A)} \geq \varepsilon\right) = \mathbb{P}\Big(|\nu_n(\omega)(A) - \mathbb{E}[\nu_n(A)]| \geq \nu(A)\varepsilon\Big) \leq \frac{\operatorname{var}(\nu_n(A))}{\nu(A)^2\varepsilon^2}.$$

Wir rechnen

$$\operatorname{var}(\nu_n(A)) = \frac{1}{n^2}\sum_{j=1}^n \operatorname{var}(Z_j) = \frac{n}{n^2}\nu(A)(1-\nu(A)) \leq \frac{\nu(A)}{n}.$$

Daher

$$\mathbb{P}\left(\frac{|\nu_n(\omega)(A)-\nu(A)|}{\nu(A)} \geq \varepsilon\right) \leq \frac{\operatorname{var}(\nu_n(A))}{\nu(A)^2\varepsilon^2} \leq \frac{1}{n\nu(A)\varepsilon^2}.$$

□

Bemerkung Sei $0 < \delta < 1$ gegeben. Wir möchten die Schranke (9.2) anwenden, um $\overline{n}$ zu identifizieren, sodass $\mathbb{P}\Big(|\nu_n(\omega)(A) - \nu(A)| \geq \nu(A)\varepsilon\Big) \leq \delta$ für $n \geq \overline{n}$.
Die Schranke $\frac{1}{n\nu(A)\varepsilon^2}$ hängt aber vom gesuchten Parameter $\nu(A)$ ab. Falls $\nu(A) > 0$ gilt, setzen wir $R \coloneqq \frac{1}{\nu(A)\varepsilon^2}$. Die Funktion

$$(x_1, \dots, x_n) \mapsto \frac{1}{\nu_n(x_1, \dots, x_n)(A)\varepsilon^2}$$

definiert einen Schätzer für R. Aus $\nu_n(\omega)(A) \overset{n\to\infty}{\to} \nu(A) > 0$ fast sicher folgt

$$R_n(\omega) \coloneqq \frac{1}{\nu_n(\omega)(A)\varepsilon^2} \overset{n\to\infty}{\to} \frac{1}{\nu(A)\varepsilon^2} \quad f.s.$$

und daher ist R_n konsistent.
Ein mögliches Verfahren wäre Folgendes: Wir berechnen R_{n_1} für n_1 Messungen. Falls $\frac{1}{n_1}R_{n_1}(x_1, \dots, x_{n_1}) \leq \delta$, dann setzen wir $\overline{n} \coloneqq n_1$. Falls $\frac{1}{n_1}R_{n_1}(x_1, \dots, x_{n_1}) > \delta$, machen wir noch mehr Messungen, bis wir ein $n_1 + n_2$ erreichen, sodass $\frac{1}{n_1+n_2}R_{n_1+n_2}(x_1, \dots, x_{n_1+n_2}) \leq \delta$ und setzen $\overline{n} \coloneqq n_1 + n_2$.

9.3 Schätzer für Erwartungswert und Varianz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \sim \nu$, wobei das Wahrscheinlichkeitsmaß ν auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ unbekannt ist.
Wir nehmen jetzt$X_1 \in L^2$ an und daher sind $\mu \coloneqq \mathbb{E}[X_1]$ und $\sigma^2 \coloneqq \mathrm{var}(X_1)$ wohldefiniert.
Unser Ziel ist es, μ und σ^2 zu bestimmen.

Lemma 9.7. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \in L^2$, $\mathbb{E}[X_1] = \mu$ und $\mathrm{var}(X_1) = \sigma^2$. Für $n \in \mathbb{N}$, betrachten wir die Funktionen*

$$\mu_n(x_1, \dots, x_n) \coloneqq \frac{1}{n}\sum_{j=1}^{n} x_j, \tag{9.3}$$

$$V_n(x_1, \dots, x_n) \coloneqq \frac{1}{n}\sum_{j=1}^{n} \Big(x_j - \mu_n(x_1, \dots, x_n)\Big)^2. \tag{9.4}$$

(i) μ_n (bzw V_n) ist ein konsistenter Schätzer für μ (bzw σ^2).

(ii) μ_n ist erwartungstreu, V_n ist nicht erwartungstreu.

Beweis.
(i) Aus dem Gesetz der großen Zahlen (die X_j sind unabhängig, identisch verteilt und integrierbar) folgt

$$\mu_n(\omega) \coloneqq \frac{1}{n}\sum_{j=1}^{n} X_j(\omega) \overset{n\to\infty}{\to} \mathbb{E}[X_1] = \mu \qquad f.s.$$

und daher ist μ_n konsistent. Wir zeigen jetzt dass V_n auch konsistent ist. Wir rechnen

$$V_n(\omega) \coloneqq \frac{1}{n}\sum_{j=1}^{n} \Big(X_j(\omega) - \mu_n(\omega)\Big)^2 = \frac{1}{n}\Big(\sum_{j=1}^{n} X_j(\omega)^2\Big) - \mu_n(\omega)^2.$$

Die Zufallsvariablen $Z_j := X_j^2$ sind unabhängig, identisch verteilt und integrierbar (da $X_j^2 \in L^1$). Aus dem Gesetz der großen Zahlen folgt dann

$$\frac{1}{n}\sum_{j=1}^{n} X_j(\omega)^2 \to_{n\to\infty} \mathbb{E}\left[X_1^2\right] = \sigma^2 + \mu^2 \quad f.s.$$

Die Aussage folgt aus $\mu_n^2 \overset{n\to\infty}{\to} \mu^2 f.s.$

(ii) Die erste Aussage folgt aus $\mathbb{E}[\mu_n] = \mathbb{E}[X_1] = \mu$. Wir rechnen jetzt

$$\mathbb{E}\left[\mu_n^2\right] = \frac{1}{n^2}\sum_{j,k=1}^{n} \mathbb{E}[X_j X_k] = \frac{1}{n}\mathbb{E}\left[X_1^2\right] + \frac{n^2-n}{n^2}\mathbb{E}[X_1]^2 .$$

Daraus folgt

$$\mathbb{E}[V_n] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[X_j^2\right] - \mathbb{E}\left[\mu_n^2\right] = \mathbb{E}\left[X_1^2\right] - \frac{1}{n}\mathbb{E}\left[X_1^2\right] - \left(1-\frac{1}{n}\right)\mathbb{E}[X_1]^2 = \left(1-\frac{1}{n}\right)\sigma^2 \neq \sigma^2.$$

□

Lemma 9.8. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $X_1 \in L^2$, $\mathbb{E}[X_1] = \mu$ und $\operatorname{var}(X_1) = \sigma^2$. Seien μ_n, V_n die Funktionen definiert in* (9.3) *und* (9.4)*.*
Dann ist die Funktion $V_n^* := \frac{n}{n-1}V_n$ ein konsistenter und erwartungstreuer Schätzer von σ^2.

Beweis. Aus $\lim_{n\to\infty}\frac{n}{n-1} = 1$ und $\lim_{n\to\infty} V_n = \sigma^2 f.s$ folgt dass V_n^* ist konsistent. Weiter gilt $\mathbb{E}[V_n^*] = \frac{n}{n-1}\mathbb{E}[V_n] = \sigma^2$.

□

Bemerkung Mithilfe der Tschebishev Ungleichung (siehe Beweis von Lemma 9.6) kann man zeigen

$$\mathbb{P}\left(\frac{|\mu_n(\omega)-\mu|}{\mu} \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2\mu^2}.$$

Sei $0 < \delta < 1$ gegeben. Wir möchten diese Schranke anwenden um $\overline{n}$ zu identifizieren, sodass $\mathbb{P}(|\mu_n(\omega)-\mu| \geq \mu\varepsilon) \leq \delta$ für $n = \overline{n}$. Der Fehler hängt aber von den unbekannten Parametern μ und σ^2 ab. Wir setzen $R := \frac{\sigma^2}{\varepsilon^2\mu^2}$. Die Funktion

$$(x_1,\ldots,x_n) \mapsto R_n := \frac{V_n}{\varepsilon^2\mu_n^2}.$$

definiert einen Schätzer für R. Aus $\mu_n \to \mu$ und $V_n \to \sigma^2$ fast sicher folgt dass $R_n \overset{n\to\infty}{\to} R$ fast sicher und daher ist R_n konsistent.

[29: 03.02.2026]

[30: 06.02.2026]

9.4 Das Maximum-Likelihood Prinzip

Es ist nicht immer klar, wie man einen guten Schätzer konstruiert.
Die Idee vom *maximum-likelihood Prinzip (MLP)* besteht darin, die Parameter so zu schätzen, dass den beobachteten Werten $x_1,\ldots,x_n$ die größte Wahrscheinlichkeit zukommt.

Beispiel Wir haben n Messungen $z_1, \dots, z_n \in \{0,1\}$. Wir wollen diese modellieren als Realisierung von unabhängigen, identisch verteilten Bernoulli Zufallsvariablen, $X_1, \dots, X_n$ mit Parameter $0 < p < 1$. Aus den Beobachtungen wollen wir nun den Wert von p schätzen. Da $\mathbb{E}[X_1] = p$, ist ein möglicher Schätzer

$$\theta_n(z_1, \dots, z_n) \coloneqq \frac{1}{n} \sum_{j=1}^{n} z_j.$$

Dieser Schätzer ist konsistent (wegen des Gesetzes der großen Zahlen) und erwartungstreu. Wir können denselben Schätzer via MLP konstruieren wie folgt.

Das MLP sagt, man schätze $p_n \coloneqq p_n(z_1, \dots, z_n)$ so, dass die Wahrscheinlichkeit der Beobachtungen maximal wird.
Für $z_i = 0, 1$ gilt $\mathbb{P}(X_i = z_i) = p^{z_i}(1-p)^{z_i}$ und daher

$$\rho_n(p) \coloneqq \mathbb{P}(X_1 = x_1, \dots, X_n = z_n) = \prod_{j=1}^{n} p^{z_i}(1-p)^{z_i} = e^{(\ln p) \sum_{j=1}^{n} z_j + (\ln(1-p)) \sum_{j=1}^{n} (1-z_j)}.$$

Wir suchen $p_n^* = p_n^*(z_1, \dots, z_n)$ sodass

$$\rho_n(p_n^*) = \max_{0<p<1} \rho_n(p).$$

Lemma 9.9. *Die Funktion ρ_n ist maximal in*

$$p_n^* = \frac{1}{n} \sum_{j=1}^{n} z_j.$$

Beweis. ρ_n is maximal $\Leftrightarrow$ $f(p) \coloneqq (\ln p) \sum_{j=1}^{n} z_j + (\ln(1-p)) \sum_{j=1}^{n} (1-z_j)$ maximal wird. Die Aussage folgt aus

$$f'(p) = \frac{1}{p} \sum_{j=1}^{n} z_j - \frac{1}{1-p} \sum_{j=1}^{n} (1-z_j) = \frac{n}{p(1-p)} \left(\left(\frac{1}{n} \sum_{j=1}^{n} z_j \right) - p \right).$$

□

9.5 Schätzer für Funktionen

Sei $f \colon [0,1] \to \mathbb{R}$ eine unbekannte Funktion. Für n Zeiten $0 \le t_1 < t_2 < \cdots < t_n \le 1$ messen wir $f(t_j)$. Wir haben also n Beobachtungen $z_1, \dots, z_n \in \mathbb{R}$. Falls die Messung exakt ist, gilt $z_j = f(t_j)$. Durch Fehler ist aber diese Gleichung verfälscht und wir sollen annehmen, dass die Differenz $X_j = z_j - f(t_j)$ eine Zufallsvariable ist.
Unser Ziel ist es, f zu bestimmen. Ohne weitere Vorabinformation ist dieses Problem praktisch unlösbar, da es unendlich viele Parameter involviert. Wir müssen also vereinfachende Annahmen machen. Wir werden jetzt Folgendes annehmen:

- Die Funktion ist linear $f(t) = a + bt$, mit unbekannte Parameter $a, b \in \mathbb{R}$,
- Die Variablen $X_1, \dots, X_n$ sind unabhängig und identisch verteilt mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ für unbekannte μ und σ^2 (das folgt aus dem zentralen Grenzwertsatz, wenn wir der Überzeugung sind, dass die Fehler X_j sich als Summen vieler kleiner "Elementarfehler", die unseren Messapparat beeinflussen, ergeben). Wir hoffen auch, dass unser Messapparat kein Bias hat und wir können daher $\mu = 0$ annehmen.

Wir versuchen also, drei Parameter durch MLP zu schätzen: a, b, σ^2. Wegen der Stetigkeit der Gaußverteilung ist die Wahrscheinlichkeit jeder Beobachtung null. Es liegt aber nahe, als "likelihood Funktion" statt der Wahrscheinlichkeit der Beobachtung die Wahrscheinlichkeitsdichte zu wählen, also

$$\rho_n(a, b, \sigma^2) := \prod_{j=1}^n h_{\sigma^2}(x_j) = \prod_{j=1}^n h_{\sigma^2}(z_j - f(t_j)) = \prod_{j=1}^n h_{\sigma^2}(z_j - a - bt_j)$$

wobei $x_j = z_j - f(t_j)$ die Beobachtung j, und

$$h_{\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

die Dichte der Gaußverteilung $\mathcal{N}(0, \sigma^2)$ ist. Wir suchen $a^* = a_n^*, b^* = b_n^*, \sigma^{*2} = \sigma^{*2}_n$, sodass

$$\rho_n(a^*, b^*, \sigma^{*2}) = \max_{a,b\in\mathbb{R},\sigma^2>0} \rho_n(a, b, \sigma^2).$$

Lemma 9.10. *Die Funktion ρ_n ist maximal in*

$$\begin{aligned}
b_n^* &:= \frac{\left(\frac{\sum_{j=1}^n t_j z_j}{n}\right) - \left(\frac{\sum_{j=1}^n t_j}{n}\right)\left(\frac{\sum_{j=1}^n z_j}{n}\right)}{\frac{\sum_{j=1}^n t_j^2}{n} - \left(\frac{\sum_{j=1}^n t_j}{n}\right)^2} \\
a_n^* &:= \frac{1}{n}\sum_{j=1}^n (z_j - b^* t_j) = \left(\frac{\sum_{j=1}^n z_j}{n}\right) - b^*\left(\frac{\sum_{j=1}^n t_j}{n}\right) \\
\sigma^{*2} &:= \frac{\sum_{j=1}^n (z_j - a^* - b^* t_j)^2}{n}.
\end{aligned}$$

Beweis. ρ_n is maximal $\Leftrightarrow$

$$F := -\sum_{j=1}^n \frac{(z_j - a - bt_j)^2}{2\sigma^2} - \frac{n}{2}\ln(2\pi\sigma^2)$$

maximal wird. Wir rechnen

$$\begin{aligned}
\partial_a F &= \frac{1}{\sigma^2}\sum_{j=1}^n (z_j - a - bt_j) = \frac{n}{\sigma^2}\left[\frac{\sum_{j=1}^n z_j}{n} - a - b\frac{\sum_{j=1}^n t_j}{n}\right] \\
\partial_b F &= \frac{1}{\sigma^2}\sum_{j=1}^n (z_j - a - bt_j)t_j = \frac{n}{\sigma^2}\left[\frac{\sum_{j=1}^n z_j t_j}{n} - a\frac{\sum_{j=1}^n t_j}{n} - b\frac{\sum_{j=1}^n t_j^2}{n}\right] \\
\partial_{\sigma^2} F &= \frac{1}{\sigma^4}\sum_{j=1}^n \frac{(z_j - a - bt_j)^2}{2} - \frac{n}{2\sigma^2} = \frac{n}{2\sigma^4}\left[\frac{\sum_{j=1}^n (z_j - a - bt_j)^2}{n} - \sigma^2\right].
\end{aligned}$$

Es gilt $\partial_b F = \partial_a F = \partial_{\sigma^2} F = 0 \Leftrightarrow a = a^*, b = b^*, \sigma^2 = \sigma^{*2}$ (Übung).
Weiter gilt $\partial^2_{a,b,\sigma^2} F(a^*, b^*, \sigma^{*2}) \le 0$ (Übung), also ist die Funtion maximal in a^*, b^*, σ^{*2}.

□

Durch MLP wir haben also folgende Schätzer für a, b gefunden:

$$b_n^*(z_1,\dots,z_n) := \frac{\left(\frac{\sum_{j=1}^n t_j z_j}{n}\right) - \left(\frac{\sum_{j=1}^n t_j}{n}\right)\left(\frac{\sum_{j=1}^n z_j}{n}\right)}{\frac{\sum_{j=1}^n t_j^2}{n} - \left(\frac{\sum_{j=1}^n t_j}{n}\right)^2}$$

$$a_n^*(z_1,\dots,z_n) := \frac{1}{n}\sum_{j=1}^n (z_j - b^* t_j) = \left(\frac{\sum_{j=1}^n z_j}{n}\right) - b^*\left(\frac{\sum_{j=1}^n t_j}{n}\right).$$

Lemma 9.11. *a_n^* (bzw b_n^*) ist ein erwartungstreuer Schätzer für a (bzw b).*

Beweis. Die Funktion $Z_j := f(t_j) + X_j = a + bt_j + X_j$ ist eine Zufallsvariable mit $\mathbb{E}[Z_j] = a + bt_j = f(t_j)$. Daher gilt

$$\mathbb{E}[b_n^*(Z_1,\dots,Z_n)] = \frac{1}{\frac{\sum_{j=1}^n t_j^2}{n} - \left(\frac{\sum_{j=1}^n t_j}{n}\right)^2}\left[\frac{\sum_{j=1}^n t_j \mathbb{E}[Z_j]}{n} - \left(\frac{\sum_{j=1}^n t_j}{n}\right)\left(\frac{\sum_{j=1}^n \mathbb{E}[Z_j]}{n}\right)\right] = b.$$

Analog rechnen wir

$$\mathbb{E}[a_n^*(Z_1,\dots,Z_n)] = \frac{1}{n}\sum_{j=1}^n (\mathbb{E}[Z_j] - \mathbb{E}[b^*]t_j) = a.$$

□

Lemma 9.12. *Wir nehmen an $t_j = \frac{j}{n}$, $j = 1,\dots,n$.*
Dann ist a_n^* (bzw b_n^*) ist ein konsistenter Schätzer für a (bzw b).

Bemerung Die Annahme $t_j = \frac{j}{n}$ heißt, die Zeiten, wo wir f messen, müssen approximiert gleichverteilt auf $[0, 1]$ sein.

Beweis.
• Wir zeigen, dass b_n^* ist konsistent.
Es gilt

$$b_n^*(Z_1,\dots,Z_n) = b_n + R_n(X_1,\dots,X_n),$$

wobei

$$R_n = R_n(X_1,\dots,X_n) := \frac{\left(\frac{\sum_{j=1}^n t_j X_j}{n}\right) - \left(\frac{\sum_{j=1}^n t_j}{n}\right)\left(\frac{\sum_{j=1}^n X_j}{n}\right)}{\frac{\sum_{j=1}^n t_j^2}{n} - \left(\frac{\sum_{j=1}^n t_j}{n}\right)^2}.$$

Unser Ziel ist, zu zeigen, dass $\lim_{n\to\infty} R_n = 0$ fast sicher.
Aus $t_j = \frac{j}{n}$ folgt

$$\lim_{n\to\infty}\frac{\sum_{j=1}^n t_j}{n} = \frac{1}{2}, \quad \lim_{n\to\infty}\frac{\sum_{j=1}^n t_j^2}{n} = \frac{1}{3}.$$

Daher gilt

$$\lim_{n\to\infty}\frac{\sum_{j=1}^n t_j^2}{n} - \left(\frac{\sum_{j=1}^n t_j}{n}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} > 0.$$

Das Gesetz der großen Zahlen ergibt

$$\lim_{n\to\infty}\frac{\sum_{j=1}^n X_j}{n} = \mathbb{E}[X_1] = 0 \qquad f.s.$$

Wir zeigen jetzt

$$\lim_{n\to\infty} \frac{\sum_{j=1}^n t_j X_j}{n} = 0 \qquad f.s.,$$

was den Beweis beendet.

Behauptung: Für $n \in \mathbb{N}, 0 < \alpha$ sei $A_n := \left\{ \left| \sum_{j=1}^n t_j X_j \right| \leq n^{\frac{1}{2}+\alpha} \right\}$. Es gilt

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n \right) = 1.$$

Konsequenz: Wir wählen $\alpha < \frac{1}{2}$. Aus der Behauptung $\exists N_\alpha \in \mathcal{F}$ Nullmenge, sodass:
$\forall \omega \in \Omega \setminus N_\alpha \; \exists n_0 = n_0(\omega) \in \mathbb{N}$ mit

$$\left| \sum_{j=1}^n t_j X_j(\omega) \right| \leq n^{\frac{1}{2}+\alpha} \quad \forall n \geq n_0.$$

Insbesondere gilt

$$\limsup_{n\to\infty} \frac{\left| \sum_{j=1}^n t_j X_j(\omega) \right|}{n} \leq \lim_{n\to\infty} \frac{n^\alpha}{n^{\frac{1}{2}}} = 0.$$

Beweis der Behauptung Wegen Borel-Cantelli genügt es, zu zeigen, dass $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n^c) < \infty$ gilt.
Wir kürzen $Z := \sum_{j=1}^n t_j X_j$, $\delta := n^{\frac{1}{2}+\alpha}$. Wir suchen eine Abschätzung von $\mathbb{P}(|Z| > \delta)$.
Aus der exponentiellen Tschebishev Ungleichung folgt für alle $t > 0$

$$\mathbb{P}(|Z| > \delta) \leq e^{-\delta t} \left(\mathbb{E}[e^{tZ}] + \mathbb{E}[e^{-tZ}] \right).$$

Aus $X_j \sim \mathcal{N}(0, \sigma^2)$ und $t_j \geq 0$ folgt dass

$$Z \sim \mathcal{N}(0, \Sigma^2), \qquad \text{mit} \quad \Sigma^2 = \sigma^2 \sum_{j=1}^n t_j^2,$$

und daher $\mathbb{E}[e^{\pm tZ}] = e^{\frac{\Sigma^2 t^2}{2}}$. Nach Minimierung in $t > 0$ erhalten wir

$$\mathbb{P}(|Z| > \delta) \leq \; 2 \; e^{-\frac{\delta^2}{2\Sigma^2}} = 2 \; e^{-\frac{n^{2\alpha}}{2\sigma^2 c_n}}, \quad \text{mit} \quad c_n := \frac{\sum_{j=1}^n t_j^2}{n}.$$

Da $\lim_{n\to\infty} c_n = \frac{1}{3}$, $\exists n_1 \in \mathbb{N}$, sodass

$$\mathbb{P}(A_n^c) \leq 2 \; e^{-\frac{n^{2\alpha}}{2\sigma^2}} \quad \forall n \geq n_1,$$

daher

$$\sum_{n\geq n_1} \mathbb{P}(A_n^c) \leq 2 \sum_{n\geq n_1} e^{-\frac{n^{2\alpha}}{2\sigma^2}} < \infty.$$

Es gilt also $\sum_n \mathbb{P}(A_n^c) < \infty$.

• Wir zeigen, dass a_n^* konsistent ist.
Die Aussage folgt aus

$$a_n^*(Z_1, \ldots, Z_n) = a + (b - b_n^*) \frac{\sum_{j=1}^n t_j}{n} + \frac{\sum_{j=1}^n t_j X_j}{n},$$

mit $\lim_{n\to\infty} b_n^* = b$ und $\lim_{n\to\infty} \frac{\sum_{j=1}^n t_j X_j}{n} = 0$ fast sicher. □