Stochastik II

Notizen zu einer Vorlesung an der
Johannes-Gutenberg-Universität Mainz, Winter 2024/25

Matthias Birkner

Vorläufige Version, 13. November 2024

Kommentare, Korrekturvorschläge, Hinweise auf (Tipp-)fehler gerne per Email an
`birkner@mathematik.uni-mainz.de` senden

Inhaltsverzeichnis

Diese Notizen beruhen in Teilen auf den Vorlesungsnotizen zur Vorlesung Stochastik II aus dem WS 2014/15, die Matthias Muth in LaTeXgesetzt hatte. Dafür hier nochmals mein herzlicher Dank.

Kapitel 1

Austauschbarkeit

Literaturhinweise Die Themen dieses Kapitels finden sich beispielsweise in [Kl, Kap. 12], [Du, Ch. 4.7], weitergehend in [Ka, Ch. 27 und 28], sehr viel weitergehend z.B. in [K05].

1.1 Grundsätzliches

Sei I eine Indexmenge und $X_i, i \in I$ Zufallsvariablen mit Wertebereich E. E sei ein polnischer Raum (d.h. E ist ein topologischer Raum, der so metrisiert werden kann, dass (E, d) ein vollständiger und separabler metrischer Raum ist, beispielsweise E abzählbar, $E = \mathbb{R}$, $E = \mathbb{R}^d$, $E = \mathcal{C}([0,1])$).

Definition 1.1. $(X_i)_{i\in I}$ heißt *austauschbar*, wenn gilt

$$\mathscr{L}((X_i)_{i\in I}) = \mathscr{L}((X_{\pi(i)})_{i\in I})$$

für jede endliche Permutation $\pi : I \to I$ (d.h. π ist bijektiv und $|\{i \mid \pi(i) \neq i\}| < \infty$).

Bemerkung 1.2. $(X_i)_{i\in I}$ ist genau dann austauschbar, wenn für alle $n \in \mathbb{N}$ und paarweise verschiedene $i_1, \dots i_n \in I$, sowie paarweise verschiedene $j_1, \dots, j_n \in I$ gilt

$$\mathscr{L}((X_{i_1}, \dots, X_{i_n})) = \mathscr{L}((X_{j_1}, \dots, X_{j_n}))$$

Beweis. Ist $(X_i)_{i\in I}$ austauschbar, so gilt $\mathscr{L}((X_{i_1}, \dots, X_{i_n})) = \mathscr{L}((X_{j_1}, \dots, X_{j_n}))$ nach Definition mit $\pi(i_k) = j_k$. Andererseits ist $\mathscr{L}((X_i)_{i\in I})$ festgelegt durch $\{\mathscr{L}((X_j)_{j\in J}) \mid J \subset I \text{ endlich}\}$, denn „endliche Zylindermengen“ $B_{i_1} \times \dots \times B_{i_k} \times \bigtimes_{i\in I\setminus\{i_1,\dots i_k\}} E_i$ erzeugen die Produkt-σ-Algebra auf $\bigtimes_{i\in I} E_i$. $\square$

Insbesondere haben X_i und X_j dieselbe Verteilung für alle $i, j \in I$. Die Umkehrung gilt im Allgemeinen nicht.

Beispiel 1.3. i) Sind $X_i, i \in I$ unabhängig und identisch verteilt, so sind sie auch austauschbar.

ii) Teilfamilien austauschbarer Zufallsvariablen sind austauschbar.

iii) (Ziehen ohne Zurücklegen) Es seien N Kugeln in einer Urne, M schwarze und $N - M$ weiße. Ziehe ohne Zurücklegen. Sei $X_i := \mathbf{1}_{\{i\text{-te Kugel ist schwarz}\}}$. Für $x_1, \ldots, x_N \in \{0,1\}$ mit $x_1 + \ldots + x_N = M$ gilt

$$\mathbb{P}(X_1 = x_1, \ldots, X_N = x_N) = \frac{1}{\binom{N}{M}} = \mathbb{P}(X_1 = x_{\pi(1)}, \ldots, X_N = x_{\pi(N)}),$$

also sind die X_i austauschbar.

iv) (Münzwurf mit zufälliger Erfolgswahrscheinlichkeit) Sei Y eine ZV mit Werten in $[0,1]$. Gegeben $Y = y$ seien $X_1, X_2, \ldots \sim \mathrm{Ber}(y)$ unabhängig und identisch verteilt (zum Beispiel realisierbar als $X_i = \mathbf{1}_{\{U_i \le Y\}}$ mit $U_1, U_2, \ldots \sim \mathrm{Unif}([0,1])$ u.i.v. und unabhängig von Y).

Für $n \in \mathbb{N}$ und $x_1, \ldots, x_n \in \{0,1\}$ mit $x_1 + \ldots + x_n = s$ und jede Permutation π von $\{1, \ldots, n\}$ gilt:

$$\begin{aligned}
\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) &= \mathbb{E}[\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n \mid Y)] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} Y^{x_i}(1-Y)^{1-x_i}\right] = \mathbb{E}[Y^s(1-Y)^{n-s}] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} Y^{x_{\pi^{-1}(i)}}(1-Y)^{1-x_{\pi^{-1}(i)}}\right] \\
&= \mathbb{P}(X_1 = x_{\pi^{-1}(1)}, \ldots, X_n = x_{\pi^{-1}(n)}) \\
&= \mathbb{P}(X_{\pi(1)} = x_1, \ldots, X_{\pi(n)} = x_n)
\end{aligned}$$

Schreib- und Sprechweisen. $\mathcal{S}_n := \big\{\text{Permutationen von } \{1, 2, \ldots, n\}\big\}$. Ein $\pi \in \mathcal{S}_n$ fassen wir auch auf als (endliche) Permutation von $\mathbb{N}$ via $\pi(j) = j$ für $j > n$.
Für $x = (x_1, \ldots, x_n) \in E^n$ definieren wir $x^\pi := (x_{\pi(1)}, \ldots, x_{\pi(n)})$.
Für $x = (x_1, x_2, \ldots) \in E^{\mathbb{N}}$ definieren wir $x^\pi := (x_{\pi(1)}, x_{\pi(2)}, \ldots)$.
Für $f : E^n \to E'$ (mit irgendeinem Wertebereich E') definieren wir $f^\pi((x_1, \ldots, x_n)) := f(x^\pi)$.
Gegebenenfalls setzen wir (manchmal implizit) $f : E^n \to E'$ in naheliegender Weise zu $f : E^{\mathbb{N}} \to E'$ fort, indem wir für $x = (x_1, x_2, \ldots) \in E^{\mathbb{N}}$ definieren $f(x) := f(x_1, x_2, \ldots, x_n)$.

Definition 1.4. $f : E^n \to E'$ heißt (n-) symmetrisch, wenn $f = f^\pi$ für alle $\pi \in \mathcal{S}_n$ gilt. $f\colon E^{\mathbb{N}} \to E'$ heißt n-symmetrisch, wenn $f = f^\pi$ für alle $\pi \in \mathcal{S}_n$. f heißt symmetrisch, falls f n-symmetrisch ist für jedes $n \in \mathbb{N}$.

Beispiel 1.5. i) Ist $E = \mathbb{R}$, $E' = \overline{\mathbb{R}}$, so sind die Funktionen

$$f((x_1, x_2, \ldots)) := \limsup_{n\to\infty} x_n \text{ und } f((x_1, x_2, \ldots)) := \limsup_{n\to\infty} \frac{x_1 + \ldots + x_n}{n} \text{ symmetrisch.}$$

ii) $a_n : \mathbb{R}^\infty \to \mathbb{R},\ a_n(x) = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ ist n-symmetrisch, aber nicht m-symmetrisch für $m > n$.

iii) $s : \mathbb{R}^\infty \to \overline{\mathbb{R}}_+,\ s(x) = \sum_{n=1}^{\infty} |x_n|$ ist symmetrisch.

iv) Für $x \in E^\infty$ ist $\xi(x) = \dfrac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ n-symmetrisch (n-te empirische Verteilung).

Beispiel 1.6 (n-tes symmetrisiertes Mittel). Sei $k \in N$, $\varphi : E^k \to R$ und $A_n(\varphi) : E^{\mathbb{N}} \to \mathbb{R}$ mit

$$A_n(\varphi)(x) = \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n} \varphi(x^\pi).$$

Dann ist $A_n(\varphi)(x) = A_n(\varphi)(x^{\pi'})$ für alle $\pi' \in \mathcal{S}_n$, d.h. A_n ist n-symmetrisch.

Definition 1.7. Sei $X = (X_n)_n$ ein stochastischer Prozess mit Werten in E.

$$\mathscr{E}_n := \sigma\left(F \circ X \mid F\colon E^{\mathbb{N}} \to \mathbb{R} \text{ messbar und } n\text{-symmetrisch}\right)$$

ist die σ-Algebra der unter Permutation der ersten n Koordinaten invarianten Ereignisse.

$$\mathscr{E} := \bigcap_{n\in N} \mathscr{E}_n = \sigma\left(F \circ X \mid F\colon E^{\mathbb{N}} \to \mathbb{R} \text{ messbar und symmetrisch}\right)$$

heißt die *σ-Algebra der austauschbaren Ereignisse für X* (kurz: die austauschbare σ-Algebra).

Bemerkung. $\mathscr{E} = \{X^{-1}(B) \mid B \in \mathscr{B}(E)^{\otimes\mathbb{N}} \text{ mit } B^\pi = B \text{ für alle } \pi \in \mathcal{S}_n, n \in \mathbb{N}\}$, wobei $B^\pi = \{x^\pi \mid x \in B\}$.

Beobachtung 1.8. Es sei $\mathscr{T} = \bigcap_{n\in\mathbb{N}} \sigma(X_n, X_{n+1}, \dots)$ die terminale σ-Algebra für X. Dann gilt $\mathscr{T} \subsetneq \mathscr{E}$.

Beweis. Es gilt $\sigma(X_n, X_{n+1}, \dots) \subset \mathscr{E}_{n-1}$, also $\mathscr{T} \subset \mathscr{E}$.

Betrachte $|E| > 1$ und wähle $B \in \mathscr{B}(E) \smallsetminus \{\varnothing, E\}$. $S := \sum_{n=1}^\infty \mathbf{1}_B(X_n)$ ist $\mathscr{E}$-messbar, aber $\{S = s\} \notin \mathscr{T}$ für $s \in \mathbb{N}_0$. □

Lemma 1.9. *Es sei $X = (X_n)_n$ eine Folge austauschbarer Zufallsvariablen mit Werten in E und $\varphi\colon E^k \to \mathbb{R}$ messbar mit $\mathbb{E}[|\varphi(X)|] < \infty$. Dann gilt für alle $n \geq k$ und $\pi \in \mathcal{S}_n$:*

i) $\mathbb{E}[\varphi(X) \mid \mathscr{E}_n] = \mathbb{E}[\varphi(X^\pi) \mid \mathscr{E}_n]$ *f.s.*

ii) $\mathbb{E}[\varphi(X) \mid \mathscr{E}_n] = \dfrac{1}{n!}\sum_{\pi\in\mathcal{S}_n} \varphi(X^\pi) = A_n(\varphi)(X).$

Beweis. i) Betrachte zunächst $A \in \mathscr{E}_n$ der Form

$$A = \left\{F_1(X) \in B_1, \dots, F_k(X) \in B_k\right\}$$

für n-symmetrische, messbare Funktionen $F_1, \dots, F_k : E^{\mathbb{N}} \to \mathbb{R}$ und $B_1, \dots B_k \in \mathscr{B}(\mathbb{R})$, $k \in \mathbb{N}$.

Sei $\pi \in \mathcal{S}_n$. Dann ist

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_A\varphi(X)] &= \mathbb{E}[\mathbf{1}_{B_1}(F_1(X))\cdots\mathbf{1}_{B_k}(F_k(X))\varphi(X)] \\
&= \mathbb{E}[\mathbf{1}_{B_1}(F_1(X^\pi))\cdots\mathbf{1}_{B_k}(F_k(X^\pi))\varphi(X^\pi)] \\
&= \mathbb{E}[\mathbf{1}_{B_1}(F_1(X))\cdots\mathbf{1}_{B_k}(F_k(X))\varphi(X^\pi)] = \mathbb{E}[\mathbf{1}_A\varphi(X^\pi)].
\end{aligned}$$

Solche As bilden einen $\cap$-stabilen Erzeuger von $\mathscr{E}_n$, daher gilt

$$\mathbb{E}[\varphi(X) \mid \mathscr{E}_n] = \mathbb{E}[\varphi(X^\pi) \mid \mathscr{E}_n] \quad \text{f.s.}$$

ii) Damit gilt auch (f.s.)

$$\begin{aligned}\mathbb{E}[\varphi(X) \mid \mathscr{E}_n] = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \mathbb{E}[\varphi(X^\pi) \mid \mathscr{E}_n] = \mathbb{E}\left[\frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \varphi(X^\pi) \mid \mathscr{E}_n\right] &= \mathbb{E}[A_n(\varphi)(X) \mid \mathscr{E}_n] \\ &= A_n(\varphi)(X),\end{aligned}$$

da $A_n(\varphi)$ als n-symmetrische Funktion $\mathscr{E}_n$-messbar ist. □

1.2 Rückwärtsmartingale

Definition 1.10. Sei $\mathscr{F} = (\mathscr{F}_{-n})_{n \in \mathbb{N}_0}$ eine absteigende Filtration, d.h. $\mathscr{F}_0 \supset \mathscr{F}_{-1} \supset \mathscr{F}_{-2} \supset \dots$ $X = (X_{-n})_{n \in \mathbb{N}_0}$ heißt ein *$\mathscr{F}$-Rückwärtsmartingal (unter $\mathbb{P}$)*, wenn

i) X_{-n} ist $\mathscr{F}_{-n}$-messbar und $\mathbb{E}[|X_{-n}|] < \infty$ für alle $n \in \mathbb{N}_0$.

ii) $\mathbb{E}[X_{-n} \mid \mathscr{F}_{-n-1}] = X_{-n-1}$ f.s. für alle $n \in \mathbb{N}$.

Beispiel 1.11. Sei $X = (X_1, X_2, \dots)$ eine austauschbare Folge von Zufallsvariablen mit Werten in $\mathbb{R}$ und $\mathbb{E}[|X_1|] < \infty$. Sei $\mathscr{F}_{-n} := \mathscr{E}_n$ und $Y_{-n} := \frac{1}{n}\sum_{i=1}^n X_i$, $n \in \mathbb{N}$. Es gilt:

$$\begin{aligned}\mathbb{E}[Y_{-n+1} \mid \mathscr{F}_{-n}] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n-1} X_i \,\middle|\, \mathscr{E}_n\right] = \frac{1}{n-1}\sum_{i=1}^{n-1} \mathbb{E}[X_i \mid \mathscr{E}_n] \\ &= \frac{1}{n-1}\sum_{i=1}^{n-1} \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} X_{\pi(i)} \\ &= \frac{1}{n-1}\sum_{i=1}^{n-1} \frac{1}{n!} \times (n-1)! \big(X_1 + X_2 + \dots + X_n\big) = \frac{1}{n}\sum_{j=1}^{n} X_j = Y_{-n}\end{aligned}$$

(wobei wir in der zweiten Zeile Lemma 1.9 verwenden).

Bemerkung 1.12. Wegen $X_{-n} = \mathbb{E}[X_0 \mid \mathscr{F}_{-n}]$, $n \in \mathbb{N}$ ist ein Rückwärtsmartingal stets gleichgradig integrierbar (vgl. Lemma B.26).

Satz 1.13. *Sei $(X_{-n})_n$ ein Rückwärtsmartingal bezüglich $(\mathscr{F}_{-n})_n$. Dann existiert ein $X_{-\infty}$ mit $X_{-n} \to X_{-\infty}$ f.s. und in $\mathcal{L}^1(\mathbb{P})$. Es gilt $X_{-\infty} = \mathbb{E}[X_0 \mid \mathscr{F}_{-\infty}]$ mit $\mathscr{F}_{-\infty} := \bigcap_n \mathscr{F}_{-n}$.*

Beweis. Sei $-\infty < a < b < \infty$ und $U_{-n}^{(a,b)}$ die Anzahl abgeschlossener Aufkreuzungen von unter a nach über b im Zeitintervall $-n, -n+1, \dots, -1, 0$. Nach Lemma B.20 gilt:

$$\mathbb{E}\big[U_{-n}^{(a,b)}\big] \le \frac{1}{b-a}\mathbb{E}[(X_{-n} - a)^-] \le \frac{1}{b-a}(|a| + \mathbb{E}[|X_{-n}|]) \le \frac{1}{b-a}(|a| + \mathbb{E}[|X_0|]).$$

Also existiert $U^{(a,b)} := \lim_n U_{-n}^{(a,b)}$ mit $\mathbb{E}\big[U^{(a,b)}\big] < \infty$.

Analog zum Beweis von Satz B.21 existiert damit auch $X_{-\infty} := \lim_n X_{-n}$ f.s. Mit Bemerkung 1.12 folgt $X_{-n} \to X_{-\infty}$ in $\mathcal{L}^1(\mathbb{P})$.

$X_{-\infty}$ ist $\mathscr{F}_{-\infty}$-messbar (denn für jedes $k \in \mathbb{N}$ ist $\limsup_{n\to\infty} X_{-n} = \inf_{n\geq k} \sup_{m\geq n} X_{-m}$ nach Definition $\mathscr{F}_{-k}$-messbar).

Sei $A \in \mathscr{F}_{-\infty}$. Dann gilt

$$\mathbb{E}[\mathbf{1}_A X_0] = \mathbb{E}\big[\mathbf{1}_A \mathbb{E}[X_0 \mid \mathscr{F}_{-n}]\big] = \mathbb{E}[\mathbf{1}_A X_{-n}] \xrightarrow[n\to\infty]{} \mathbb{E}[\mathbf{1}_A X_{-\infty}].$$

□

Korollar 1.14. *Sei $X = (X_1, X_2, \ldots)$ eine austauschbare Folge von Zufallsvariablen mit Werten in $\mathbb{R}$ und $\mathbb{E}[|X_1|] < \infty$. Sei $\mathscr{F}_{-n} := \mathscr{E}_n$, $\mathscr{T} = \bigcap_n \sigma(X_n, X_{n+1}, \ldots)$ die terminale σ-Algebra und $\mathscr{E} = \bigcap_n \mathscr{F}_{-n}$ die austauschbare σ-Algebra. Dann gilt*

$$\mathbb{E}[X_1 \mid \mathscr{E}] = \mathbb{E}[X_1 \mid \mathscr{T}] \text{ f.s.} \quad \text{und} \quad \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{} \mathbb{E}[X_1 \mid \mathscr{E}] \text{ f.s. und in } \mathcal{L}^1(\mathbb{P}).$$

Beweis. $Y_{-n} := \frac{1}{n}\sum_{i=1}^n X_i$ ist ein Rückwärtsmartingal (vgl. Bsp. 1.11). Es gilt

$$Y_{-n} \to Y_{-\infty} := \mathbb{E}[X_1 \mid \mathscr{E}] \quad \text{f.s. und in } \mathcal{L}^1(\mathbb{P})$$

nach Satz 1.13 (wobei wir aus Notationsbequemlichkeit den Index um 1 verschoben haben).

$Y_{-\infty}$ ist $\mathscr{T}$-messbar, denn es ist ein Grenzwert. Also gilt

$$\mathbb{E}[X_1 \mid \mathscr{E}] = Y_{-\infty} = \mathbb{E}[Y_{-\infty} \mid \mathscr{T}] = \mathbb{E}\big[\mathbb{E}[X_1 \mid \mathscr{E}] \mid \mathscr{T}\big] = \mathbb{E}[X_1 \mid \mathscr{T}].$$

(für das letzte Gleichheitszeichen verwende Beobachtung 1.8 und die Turmeigenschaft der bedingten Erwartung). □

Korollar 1.15 (Starkes Gesetz der großen Zahlen)**.** *Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte, reelle Zufallsvariablen mit $\mathbb{E}[|X_1|] < \infty$. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{} \mathbb{E}[X_1] \quad \text{f.s. und in } \mathcal{L}^1(\mathbb{P}),$$

denn gemäß Kolmogorovs 0-1-Gesetz (z.B. [St1, Satz 2.9] oder [Kl, Satz 2.37]) gilt $P(A) \in \{0,1\}$ für alle $A \in \mathscr{T}$, demnach $\mathbb{E}[X_1 \mid \mathscr{T}] = \mathbb{E}[X_1]$ f.s.

Satz 1.16. *Sei $(X_n)_n$ austauschbar mit Werten in E und $\varphi : E^k \to \mathbb{R}$ eine messbare Funktion mit $\mathbb{E}[|\varphi(X_1, \ldots, X_k)|] < \infty$. Dann gilt*

$$\lim_{n\to\infty} A_n(\varphi)(X) = \mathbb{E}[\varphi(X) \mid \mathscr{E}] = \mathbb{E}[\varphi(X) \mid \mathscr{T}] \quad \text{f.s. und in } \mathcal{L}^1(\mathbb{P}).$$

Beweis. Nach Lemma 1.9 gilt $A_n(\varphi)(X) = \mathbb{E}[\varphi(X) \mid \mathscr{E}_n]$. Es gilt

$$\mathscr{E}_1 \supset \mathscr{E}_2 \supset \ldots \supset \mathscr{E} = \bigcap_n \mathscr{E}_n,$$

also ist $(A_n(\varphi)(X))_n$ ein Rückwärtsmartingal. Dann gilt nach Satz 1.13:

$$\lim_{n\to\infty} A_n(\varphi)(X) = \mathbb{E}[\varphi(X) \mid \mathscr{E}]$$

Zeige, dass $\lim_n A_n(\varphi)(X)$ $\mathscr{T}$-messbar ist: Sei $\ell \in \mathbb{N}$ und

$$\mathcal{S}_{n,\ell} := \{\pi \in \mathcal{S}_n \mid \pi(1), \dots, \pi(k) \geq \ell\}, \quad A_{n,l}(\varphi) := \frac{1}{n!} \sum_{\pi \in \mathcal{S}_{n,l}} \varphi^\pi$$

Es ist $\mathcal{S}_n \smallsetminus \mathcal{S}_{n,l} = \bigcup_{i=1}^k \{\pi \in \mathcal{S}_n \mid \pi(i) < l\}$ und somit $|\mathcal{S}_n \smallsetminus \mathcal{S}_{n,l}| \leq k(l-1)(n-1)!$. Daher gilt

$$\mathbb{E}[|A_{n,l}(\varphi)(X) - A_n(\varphi)(X)|] \leq \frac{1}{n!}|\mathcal{S}_n \smallsetminus \mathcal{S}_{n,l}| \cdot \mathbb{E}[|\varphi(X)|] \xrightarrow[n\to\infty]{} 0,$$

das heißt $A_{n,l}(\varphi)(X) - A_n(\varphi)(X) \to 0$ in $\mathcal{L}^1(\mathbb{P})$ und damit auch stochastisch.

Wähle eine Teilfolge $n_m \nearrow \infty$ mit

$$A_{n,l}(\varphi)(X) - A_n(\varphi)(X) \to 0 \quad \text{f.s.}$$

Dann ist $\lim_n A_n(\varphi)(X) = \lim_m A_{n_m}(\varphi)(X)$ $\sigma(X_l, X_{l+1}, \dots)$-messbar für jedes $l \in \mathbb{N}$, also $\mathscr{T}$-messbar. Mit der Turmeigenschaft (und $\mathscr{T} \subset \mathscr{E}$ nach Beob. 1.8) folgt

$$\lim_{n\to\infty} A_n(\varphi)(X) = \mathbb{E}\left[\lim_{n\to\infty} A_n(\varphi)(X) \mid \mathscr{T}\right] = \mathbb{E}\big[\mathbb{E}[\varphi(X) \mid \mathscr{E}] \mid \mathcal{T}\big] = \mathbb{E}[\varphi(X) \mid \mathscr{T}]$$

□

Korollar 1.17. *Sei $(X_n)_n$ austauschbar. Dann gibt es für alle $A \in \mathscr{E}$ ein $B \in \mathscr{T}$ mit $\mathbb{P}(A \triangle B) = 0$.*

Beweis. Sei $A \in \mathscr{E} \subset \sigma(X_1, X_2, \dots)$. Wähle $A_k \in \sigma(X_1, \dots, X_k)$ mit

$$P(A \triangle A_k) \xrightarrow[k\to\infty]{} 0$$

Dies ist möglich nach dem Approximationssatz für Maße (vgl. [Kl, Satz 1.65])

Sei $C_k \subset E^k$ messbar, sodass $A_k = \{(X_1, \dots, X_k) \in C_k\}$ und setze $\varphi_k := \mathbf{1}_{C_k \times E^\infty}$. Dann gilt

$$\varphi_k(X) \xrightarrow[k\to\infty]{} \mathbf{1}_A \quad \text{f.s.}$$

(ggf. wähle eine Teilfolge k_m mit $\sum_m \mathbb{P}(A \triangle A_{k_m}) < \infty$). Dominierte Konvergenz für bedingte Erwartungen liefert

$$\mathbf{1}_A = \mathbb{E}[\mathbf{1}_A \mid \mathscr{E}] = \mathbb{E}\left[\lim_{k\to\infty} \varphi_k(X) \mid \mathscr{E}\right] = \lim_{k\to\infty} \mathbb{E}[\varphi_k(X) \mid \mathscr{E}] = \lim_{k\to\infty} \mathbb{E}[\varphi_k(X) \mid \mathscr{T}] =: \psi \quad \text{f.s.}$$

(im letzten Gleichheitszeichen verwende Satz 1.16).

ψ ist $\mathscr{T}$-messbar mit $\psi = \mathbf{1}_A$ f.s. und $B := \{\psi = 1\} \in \mathscr{T}$ leistet das Gewünschte.

(Mit $N := \{\psi \neq \mathbf{1}_A\}$ ist $\mathbb{P}(N) = 0$ und es folgt $P(\{\psi = 1\} \triangle A) \leq \mathbb{P}(N) = 0$. □

Korollar 1.18 (0-1-Gesetz von Hewitt und Savage[1]). *Seien $X_1, X_2, \dots$ unabhängig und identisch verteilt, so gilt $\mathbb{P}(A) \in \{0, 1\}$ für alle $A \in \mathscr{E}$.*

Beweis. Dies folgt aus Korollar 1.17 zusammen mit Kolmogorovs 0-1-Gesetz. □

[1] Edwin Hewitt, 1920–1999 und Jimmie Leonard Savage, 1917–1971

1.3 Struktur unendlicher austauschbarer Familien

Eine Heuristik. Sei $E = \{1, 2, \dots, k\}$ und seien $X_1, X_2, \dots$ austauschbare, E-wertige Zufallsvariablen und sei

$$\xi_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$$

die N-te empirische Verteilung.

Gegeben ξ_N sind $(X_1, X_2, \dots, X_N)$ verteilt wie Züge ohne Zurücklegen aus einer Urne. Für $m_l := |\{1 \le i \le n \mid x_i = l\}|$ mit $m_1 + \dots + m_k = n$ und Notation $(a)_n := a(a-1)\dots(a-n+1)$ gilt

$$\begin{aligned}
\mathbb{P}(X_1 = x_1, \dots, X_n = x_n \mid \xi_N) &= \frac{(N\xi_N(\{1\}))_{m_1} \dots (N\xi_N(\{k\}))_{m_k}}{(N)_n} \\
&\approx \frac{(N\xi_N(\{1\}))^{m_1} \dots (N\xi_N(\{k\}))^{m_k}}{N^n} \\
&= (\xi_N(\{1\}))^{m_1} \dots (\xi_N(\{k\}))_{m_k} \\
&\approx (\xi_\infty(\{1\}))^{m_1} \dots (\xi_\infty(\{k\}))^{m_k},
\end{aligned}$$

für $N \gg 1$, wenn wir für den Moment annehmen, dass sich die empirische Verteilung für $N \to \infty$ stabilisiert mit Grenzwert ξ_∞ (was aus Satz 1.21 folgen wird).

Demnach: Gegeben ξ_∞ sollten $X_1, X_2, \dots$ u.i.v. (mit Verteilung ξ_∞) sein.

Definition 1.19. Es sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathscr{G}, \mathscr{G}_i \subset \mathscr{F}, i \in I$ Teil-σ-Algebren. Die Familie $(\mathscr{G}_i)_{i \in I}$ heißt *unabhängig gegeben* $\mathscr{G}$, wenn für alle endlichen $J \subset I$, $A_j \in \mathscr{G}_j, j \in J$ gilt:

$$\mathbb{P}\left(\bigcap_{j \in J} A_j \mid \mathscr{G}\right) = \prod_{j \in J} \mathbb{P}(A_j \mid \mathscr{G}) \quad \text{f.s.}$$

Analog heißen Zufallsvariablen $(X_i)_{i \in I}$ *unabhängig gegeben* $\mathscr{G}$, falls $(\sigma(X_i))_{i \in I}$ unabhängig gegeben $\mathscr{G}$. Zufallsvariablen $(X_i)_{i \in I}$ heißen *unabhängig und identisch verteilt gegeben* $\mathscr{G}$, wenn die bedingten Verteilungen gegeben $\mathscr{G}$ gleich sind.

Bemerkung 1.20. i) $(\mathscr{G}_i)$ ist stets unabhängig gegeben $\mathscr{G}$, wenn $\mathscr{G}_i \subset \mathscr{G}$ für alle $i \in I$.

ii) Unabhängigkeit gegeben $\{\varnothing, \Omega\}$ ist die „gewöhnliche“ Unabhängigkeit.

iii) Sind $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \mathscr{F}_3$ σ-Algebren und ist $(\mathscr{G}_i)_{i \in I}$ unabhängig gegeben $\mathscr{F}_1$ und unabhängig gegeben $\mathscr{F}_3$, so folgt nicht notwendigerweise die Unabhängigkeit gegeben $\mathscr{F}_2$.

Betrachte zum Beispiel X_1, X_2 unabhängige, reelle Zufallsvariablen, $\mathscr{G}_i = \sigma(X_i)$, $i = 1, 2$ und $\mathscr{F}_1 = \{\varnothing, \Omega\}$, $\mathscr{F}_2 = \sigma(X_1 + X_2)$ und $\mathscr{F}_3 = \sigma(X_1, X_2)$.

Satz 1.21 (Satz von de Finetti[2]). *Sei $(X_n)_n$ eine Folge von Zufallsvariablen mit Werten in E. $(X_n)_n$ ist genau dann austauschbar, wenn es eine σ-Algebra $\mathscr{G}$ gibt, sodass $(X_n)_n$ unabhängig und identisch verteilt gegeben $\mathscr{G}$ ist. In diesem Fall kann man $\mathscr{G} = \mathscr{E}$ oder $\mathscr{G} = \mathscr{T}$ wählen.*

[2] Bruno de Finetti, 1906–1985

Beweis. Sei zunächst $(X_n)_n$ austauschbar und $\mathscr{G} = \mathscr{E}$ oder $\mathscr{G} = \mathscr{T}$.

Seien $f_i\colon E \to \mathbb{R}$ beschränkt und messbar und setze $\varphi_k(X) = \prod_{i=1}^k f_i(X_i)$. Gilt

$$\mathbb{E}[\varphi_k(X) \mid \mathscr{G}] = \mathbb{E}[\varphi_{k-1}(X) \mid \mathscr{G}] \cdot \mathbb{E}[f_k(X_k) \mid \mathscr{G}] \quad \text{für alle } k \in \mathbb{N}, \tag{1.1}$$

so folgt induktiv

$$\mathbb{E}\Big[\prod_{i=1}^k f_i(X_i) \mid \mathscr{G}\Big] = \prod_{i=1}^k \mathbb{E}[f_i(X_i) \mid \mathscr{G}] \quad \text{für alle } k \in \mathbb{N},$$

d.h. $(X_n)_n$ sind bedingt unabhängig gegeben $\mathscr{G}$ (lese $f_i = \mathbf{1}_{B_i}$, um wörtlich an Definition 1.19 anzuschließen).

Zeige also (1.1): Betrachte

$$\begin{aligned}
A_n(\varphi_{k-1})(X) \cdot A_n(f_k)(X) &= \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} f_1(X_{\pi(1)}) \cdots f_{k-1}(X_{\pi(k-1)}) \times \frac{1}{n} \sum_{j=1}^n f_k(X_j) \\
&= \frac{1}{(n)_{k-1}} \sum_{\substack{i_1,\dots,i_{k-1} \in \{1,\dots,n\} \\ \text{p.w. verschieden}}} \prod_{l=1}^{k-1} f_l(X_{i_l}) \times \Bigg(\frac{n-k+1}{n} \cdot \frac{1}{n-k+1} \sum_{\substack{j=1 \\ j \notin \{i_1,\dots,i_{k-1}\}}}^n f_k(X_j) \\
&\qquad + \frac{1}{n} \sum_{j \in \{i_1,\dots,i_{k-1}\}} f_k(X_j) \Bigg) \\
&= \frac{n-k+1}{n} \cdot \underbrace{\frac{1}{(n)_k} \sum_{\substack{i_1,\dots,i_k = 1 \\ \text{p.w. verschieden}}}^n \prod_{l=1}^k f_l(X_{i_l})}_{= A_n(\varphi_k)(X) \xrightarrow[n\to\infty]{} \mathbb{E}[\varphi(X_k) \mid \mathscr{G}]} + R_{n,k}(X)
\end{aligned}$$

mit $|R_{n,k}(X)| \le \frac{k-1}{n} \|f_1\|_\infty \cdots \|f_k\|_\infty \xrightarrow[n\to\infty]{} 0$, also (mit Satz 1.16) folgt (1.1).

Seien nun $(X_n)_n$ unabhängig und identisch verteilt gegeben $\mathscr{G}$ und $\varphi : E^k \to \mathbb{R}$ beschränkt und messbar. Es gilt für $\pi \in \mathcal{S}_n$:

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\big[\mathbb{E}[\varphi(X) \mid \mathscr{G}]\big] = \mathbb{E}\big[\mathbb{E}[\varphi(X^\pi) \mid \mathscr{G}]\big] = \mathbb{E}[\varphi(X^\pi)],$$

das heißt $(X_n)_n$ ist austauschbar. $\square$

Kapitel 2

Ein Intermezzo zur Brownschen Bewegung

Motivation Seien X_i u.i.v. mit $\mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$, wir fassen die Irrfahrt mit solchen Zuwächsen als einen zufälligen Pfad auf und reskalieren die (Zeit- und Raum-)Achsen:

Für $n \in \mathbb{N}$ und $t \geq 0$ sei

$$S_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i$$

(falls $t \in \frac{1}{n}\mathbb{Z}$, sonst linear interpoliert).

Simulationen (die wir beispielsweise im Stochastik-Praktikum betrachtet hatten) zeigen zumindest dem Augenschein nach, dass die Verteilung des zufälligen Pfads $(S_t^{(n)})_{t\geq 0}$ kaum vom Wert von n abzuhängen scheint, sofern nur n groß ist. Der zentrale Grenzwertsatz zeigt, dass für $t_1 < t_2$ gilt

$$S_{t_2}^{(n)} - S_{t_1}^{(n)} \xrightarrow{d} \mathcal{N}(0, t_2 - t_1)$$

(denn – bis auf Rundung, die im Limes keine Rolle spielt – ist $S_{t_2}^{(n)} - S_{t_1}^{(n)} = \sum_{i=\lfloor nt_1 \rfloor}^{\lfloor nt_2 \rfloor} X_i/\sqrt{n}$ eine Summe von $n(t_2 - t_1)$ vielen unabhängigen und zentrierten Summanden mit jeweils Varianz $1/n$). Zudem sind die Zuwächse über disjunkte Zeitintervalle unabhängig: Seien $t_! < t_2 < t_3 < t_4$, so enthalten $S_{t_4}^{(n)} - S_{t_3}^{(n)}$ und $S_{t_2}^{(n)} - S_{t_1}^{(n)}$ Summanden X_i aus disjunkten Indexmengen, sind also nach Konstruktion unabhängig (und analog, wenn man mehr als nur zwei Intervalle betrachtet).

Diese Eigenschaften verwendet man zur Definition der Brownschen Bewegung:

Definition 2.1. Ein stochastischer Prozess $(B_t)_{t\geq 0}$ mit Werten in $\mathbb{R}$ heißt (Standard-)Brownsche Bewegung, falls gilt

$$\begin{aligned}
&B_0 = 0,\\
&\text{für } n \in \mathbb{N},\ 0 = t_0 < t_1 < \cdots < t_n \text{ sind } B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}} \text{ unabhängig}\\
&\text{mit } \ B_{t_i} - B_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})
\end{aligned} \tag{2.1}$$

und $t \mapsto B_t$ ist stetig.

Satz 2.2. *Die Brownsche Bewegung $(B_t)_{t\geq 0}$ im Sinne von Definition 2.1 existiert.*

Beweis. Wir verwenden hier eine „explizite" Konstruktion einer stetigen Version ausführen, dieser Gedanke geht auf Paul Lévy zurück[1] (vgl. z.B. Kapitel 1 in dem Buch von Mörters+Peres [MP] für mehr Details.)

Betrachte zunächst nur $t \in [0,1]$, seien

$$\mathcal{D}_n := \left\{ \frac{k}{2^n} \mid k \in \mathbb{N}_0, k \le 2^n \right\} \quad \text{und} \quad \mathcal{D} := \bigcup_{n=1}^{\infty} \mathcal{D}_n.$$

Für $t \in \mathcal{D}$ seien $Z_t \sim \mathcal{N}(0,1)$ unabhängig. Setze $B(0) = B_0 = 0$ und $B(1) = Z_1$. Sei $B(d')$ konstruiert für $d' \in \mathcal{D}_{n-1}$ und setze für $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$

$$B(d) := \frac{1}{2}\big(B(d-2^{-n}) + B(d+2^{-n})\big) + \frac{Z_d}{\sqrt{2^{n+1}}}.$$

`[Skizze an der Tafel, siehe auch Abbildung 2.1]`

Es gilt (nach Konstruktion)

$$\{B_d \mid d \in \mathcal{D}_n\} \text{ und } \{Z_t \mid t \in \mathcal{D} \setminus \mathcal{D}_n\} \text{ sind unabhängig.} \tag{2.2}$$

Zeige:

$$B(d) - B(d-2^{-n}),\ d \in \mathcal{D}_n \setminus \{0\} \text{ sind u.i.v. mit } B(d) - B(d-2^{-n}) \sim \mathcal{N}(0, 2^{-n}) \tag{2.3}$$

durch Induktion über n. Für $n = 0$ ist die Behauptung klar. Sei also (2.3) für $n-1$ erfüllt. Betrachte $d = \frac{k}{2^n}$ für k ungerade, also $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$. Setze

$$A_{n,k} := \frac{1}{2}\big(B(\underbrace{d+2^{-n}}_{\in \mathcal{D}_{n-1}}) - B(\underbrace{d-2^{-n}}_{\in \mathcal{D}_{n-1}})\big) \sim \mathcal{N}(0, \frac{1}{4}2^{-(n-1)}) = \mathcal{N}(0, 2^{-n-1}),$$

$$B_{n,k} := \frac{Z_d}{\sqrt{2^{n+1}}} \sim \mathcal{N}(0, 2^{-n-1}).$$

Nach (2.2) sind $A_{n,k}$ und $B_{n,k}$ unabhängig, also sind auch $A_{n,k} + B_{n,k} = B(d) - B(d-2^{-n})$ und $A_{n,k} - B_{n,k} = B(d+2^{-n}) - B(d)$ unabhängig und es gilt

$$A_{n,k} + B_{n,k},\ A_{n,k} - B_{n,k} \sim \mathcal{N}(0, 2^{-n}).$$

Demnach ist

$$\left(B\left(\frac{2j}{2^n}\right) - B\left(\frac{2j-1}{2^n}\right),\ B\left(\frac{2j-1}{2^n}\right) - B\left(\frac{2j-2}{2^n}\right)\right)_{j=1,\dots,2^{n-1}} \overset{d}{=} \big(\mathcal{N}(0,2^{-n}) \otimes \mathcal{N}(0,2^{-n})\big)^{\otimes 2^{n-1}},$$

das heißt (2.3) gilt auch für n.

Setze nun

$$F_0(t) := \begin{cases} 0, & t = 0 \\ Z_1, & t = 1 \\ \text{linear interpoliert}, & t \in (0,1) \end{cases}$$

[1] P. Lévy, Processus stochastiques et mouvement brownien, Gauthier-Villars, 1948.

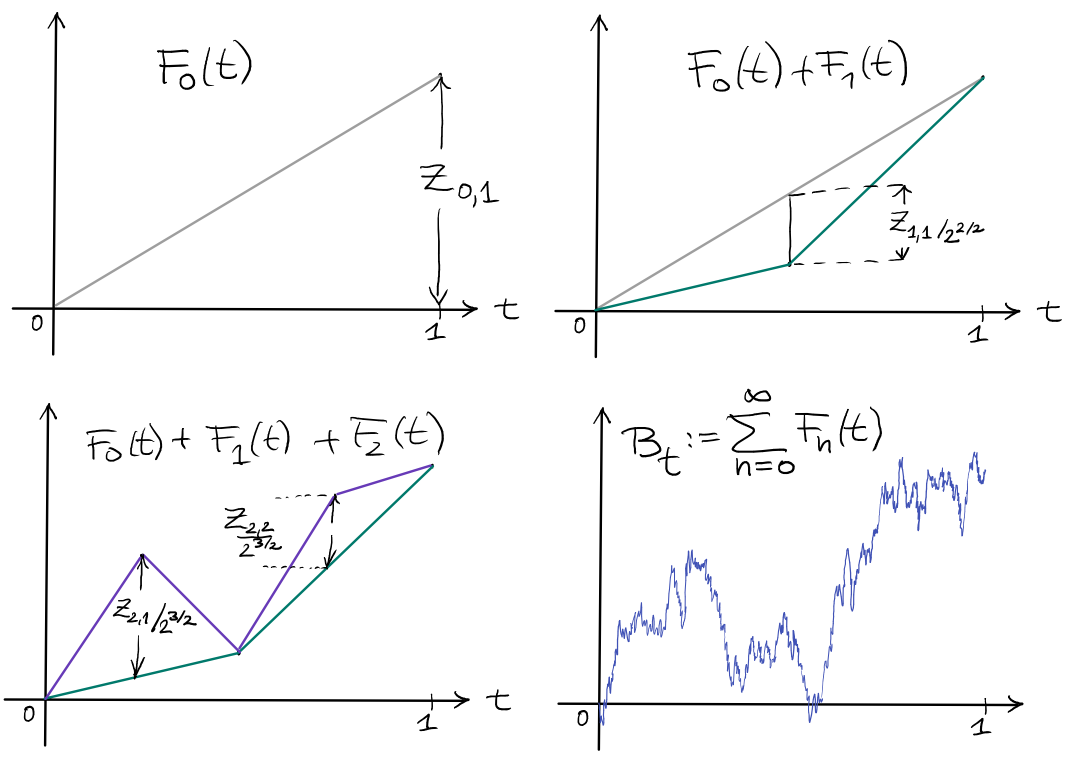


Abbildung 2.1: Skizze der Lévy-Konstruktion der Brownschen Bewegung

und für $n \in \mathbb{N}$

$$F_n(t) := \begin{cases} \frac{Z_t}{\sqrt{2^{n+1}}}, & t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1} \\ 0, & t \in \mathcal{D}_{n-1} \\ \text{linear interpoliert}, & \text{sonst} \end{cases}$$

und zeige für $d \in \mathcal{D}_n$

$$B(d) = \sum_{i=0}^{n} F_i(d) = \sum_{i=0}^{\infty} F_i(d) \tag{2.4}$$

durch Induktion über n. Für $n = 0$ ist die Aussage klar. Sei also (2.4) für $n-1$ erfüllt. Für $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ gilt nach Induktionsvoraussetzung

$$\sum_{i=0}^{n-1} F_i(d) = \sum_{i=0}^{n-1} \frac{F_i(d-2^{-n}) + F_i(d+2^{-n})}{2} = \frac{1}{2}\big(B(d-2^{-n}) + B(d+2^{-n})\big),$$

wobei wir im ersten Schritt die Linearität von F_i auf $[d-2^{-n}, d+2^{-n}]$ ausgenutzt haben.

Nach Konstruktion ist $F_n(d) = \frac{Z_d}{\sqrt{2^{n+1}}}$, demnach ergibt sich zusammen mit der Definition von $B(d)$

$$\sum_{i=0}^{n} F_i(d) = B(d),$$

das heißt (2.4) gilt auch für n.

Für $c > 1$ und $n \in \mathbb{N}$ gilt

$$\mathbb{P}(|Z_d| \geq c\sqrt{n}) = \frac{2}{\sqrt{2\pi}} \int_{c\sqrt{n}}^{\infty} e^{-\frac{x^2}{2}}\, dx \leq \int_{c\sqrt{n}}^{\infty} x e^{-\frac{x^2}{2}}\, dx = \left[-e^{-\frac{x^2}{2}}\right]_{c\sqrt{n}}^{\infty} = e^{-\frac{c^2 n}{2}},$$

das heißt für $c > \sqrt{2\log 2}$ gilt

$$\sum_{n=1}^{\infty} \mathbb{P}(\exists\, d \in \mathcal{D}_n : |Z_d| \geq c\sqrt{n}) \leq \sum_{n=1}^{\infty} (2^n + 1) e^{-\frac{c^2 n}{2}} < \infty.$$

Somit existiert nach Borel-Cantelli ein (zufälliges) N_0, sodass

$$\text{für alle } n \geq N_0 \text{ gilt} \quad \|F_n\|_\infty := \sup_{t\in[0,1]} |F_n(t)| \leq c\sqrt{n} 2^{-\frac{n+1}{2}}, \tag{2.5}$$

das heißt insbesondere $\sum_{n=0}^{\infty} \|F_n\|_\infty < \infty$. Demnach ist

$$B(t) = \sum_{n=0}^{\infty} F_n(t), \quad t \in [0,1] \tag{2.6}$$

als Grenzwert einer gleichmäßig konvergenten Reihe von stetigen Funktionen selbst stetig.

Wegen (2.3) gilt für $t_0 < \ldots < t_n \in \mathcal{D}$

$$\mathscr{L}\big((B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}})\big) = \mathcal{N}(0, t_1 - t_0) \otimes \ldots \otimes \mathcal{N}(0, t_n - t_{n-1}).$$

Der Fall allgemeiner t_i folgt durch Approximation mit $t'_{i_k} \in \mathcal{D}$ (vgl. [MP]).

Für $t \in [0,\infty)$ betrachte nach obiger Konstruktion $(B_0(t))_{t\in[0,1]}$, $(B_1(t))_{t\in[0,1]}, \ldots$ als unabhängige Kopien und setze

$$B(t) := \sum_{i=0}^{\lfloor t\rfloor - 1} B_i(1) + B_{\lfloor t\rfloor}(t - \lfloor t\rfloor).$$

□

Mit Lévys Konstruktion lässt sich relativ leicht einsehen, „wie stetig" die Pfade der Brownschen Bewegung typischerweise sind (zumindest „bis auf Konstante"): Es stellt sich heraus, dass $t \mapsto B_t$ Hölder-stetig der Ordnung γ ist für jedes $\gamma < 1/2$.

Satz 2.3. *1. Es gibt ein $C < \infty$ und ein (zufälliges) $h_0 > 0$, sodass*

$$|B(t+h) - B(t)| \leq C\sqrt{h\log(1/h)} \tag{2.7}$$

für alle $h \in (0, h_0)$ und $t \in [0, 1-h]$ gilt.

2. Für jedes $c < \sqrt{2}$ und $\varepsilon > 0$ gibt es (f.s.) ein $h \in (0,\varepsilon)$ und $t \in [0, t-h]$ mit

$$|B(t+h) - B(t)| \geq c\sqrt{h\log(1/h)}. \tag{2.8}$$

Beweis. 1. Schreibe $B(t) = \sum_{n=0}^{\infty} F_n(t)$ wie oben in (2.6). F_n ist differenzierbar bis auf endlich viele „Knickstellen“, es gilt

$$\|F_n'\|_\infty \leq \frac{\|F_n\|_\infty}{2^{-n}} \leq c_1\sqrt{n}2^{\frac{n-1}{2}}$$

für alle $n \geq N_0$ mit N_0 wie oben in (2.5).

Damit folgt für $\ell > N_0$

$$\begin{aligned}|B(t+h) - B(t)| &\leq \sum_{n=0}^{\infty} |F_n(t+h) - F_n(t)| \leq \sum_{n=0}^{l} h\,\|F_n'\|_\infty + \sum_{n=l+1}^{\infty} 2\,\|F_n\|_\infty \\ &\leq h\underbrace{\sum_{n=0}^{N_0-1}\|F_n'\|_\infty}_{=:S_1} + h\underbrace{\sum_{n=N_0}^{\ell}\frac{c_1}{\sqrt{2}}\sqrt{n}2^{\frac{n}{2}}}_{=:S_2} + \underbrace{\sum_{n=\ell+1}^{\infty} 2\sqrt{n}2^{-\frac{n}{2}}}_{=:S_3}.\end{aligned}$$

Für h genügend klein ist $S_1 \leq \sqrt{\log(1/h)}$. Wähle $\ell > N_0$, sodass $2^{-\ell} \leq h \leq 2^{-\ell+1}$ (dies ist möglich für h genügend klein). Dann gilt

$$S_2 \leq c_2'\sqrt{\ell 2^\ell} \leq c_2''\sqrt{\frac{1}{h}\log\frac{1}{h}}$$

und

$$S_3 \leq c_3'\sqrt{\ell 2^{-\ell}} \leq c_3''\sqrt{h\log\frac{1}{h}}$$

für gewisse Konstanten c_2', c_2'', c_3', c_3''. Damit folgt für $C := 1 + c_2'' + c_3''$

$$|B(t+h) - B(t)| \leq hS_1 + hS_2 + S_3 \leq h\sqrt{\log\frac{1}{h}} + hc_2''\sqrt{\frac{1}{h}\log\frac{1}{h}} + c_3''\sqrt{h\log\frac{1}{h}} \leq C\sqrt{h\log\frac{1}{h}},$$

d.h. (2.7) gilt.

2. Sei

$$A_{k,n} := \left\{B((k+1)e^{-n}) - B(ke^{-n}) > c\sqrt{n}e^{-n/2}\right\},$$

es gilt (mit $Z \sim \mathcal{N}(0,1)$)

$$\begin{aligned}\mathbb{P}(A_{k,n}) &= \mathbb{P}\big(e^{-n/2}Z > c\sqrt{n}e^{-n/2}\big) = P(Z > c\sqrt{n}) \\ &= \frac{1}{\sqrt{2\pi}}\int_{c\sqrt{n}}^{\infty} e^{-x^2/2}\,dx \geq \frac{1}{\sqrt{2\pi}}\frac{c\sqrt{n}}{c^2n+1}e^{-c^2n/2}\end{aligned}$$

Folglich ist (beachte $c < \sqrt{2}$)

$$e^n\mathbb{P}(A_{k,n}) \xrightarrow[n\to\infty]{} \infty,$$

somit

$$\mathbb{P}\Big(\bigcap_{k=0}^{\lfloor e^n-1\rfloor} A_{k,n}^c\Big) = \big(1 - \mathbb{P}(A_{0,n})\big)^{\lfloor e^n-1\rfloor} \leq \exp\big(-e^n\mathbb{P}(A_{0,n})\big) \xrightarrow[n\to\infty]{} 0$$

d.h. $\mathbb{P}\Big(\bigcup_{k=0}^{\lfloor e^n-1\rfloor} A_{k,n}\Big) \to 1$ und (2.8) gilt. □

Beobachtung 2.4. 1. $B = (B_t)_{t\geq 0}$ ist Standard-Brownsche Bewegung g.d.w. B zentrierter Gaußscher Prozess (d.h. $(B_{t_1}, \ldots, B_{t_k})$ ist multivariat normalverteilt für jede Wahl von $t_1 < \cdots < t_k$, $k \in \mathbb{N}$) mit stetigen Pfaden und

$$\mathrm{Cov}[B_s, B_t] = s \wedge t, \quad s, t \geq 0$$

2. (Skalierungsinvarianz) Für $c \neq 0$ ist auch $(\widetilde{B}_t)_{t\geq 0}$ mit $\widetilde{B}_t := \frac{1}{c} B_{c^2 t}$ Brownsche Bewegung.

Beweis. 1. Sei zunächst $(B_t)_{t\geq 0}$ eine (Standard-) Brownsche Bewegung. Für $0 \leq s \leq t$ gilt

$$\mathrm{Cov}[B_s, B_t] = \mathrm{Cov}[B_s, B_s + (B_t - B_s)] = \mathrm{Var}[B_s] + \mathrm{Cov}[B_s, B_t - B_s] = s + 0 = s = s \wedge t.$$

Die Umkehrung folgt aus der Tatsache, dass die endlich-dimensionalen Verteilungen eines (zentrierten) Gaußschen Prozesses durch die Kovarianzen festgelegt sind, vgl. Def. 4.5 (oder z.B. [Kl, Kap. 15.6]).

2. Es gilt $\widetilde{B}_0 = 0$, $\widetilde{B}$ hat stetige Pfade und für die Kovarianzen gilt

$$\mathrm{Cov}[\widetilde{B}_s, \widetilde{B}_t] = \frac{1}{c^2} \mathrm{Cov}[B_{c^2 s}, B_{c^2 t}] = \frac{1}{c^2}(c^2 s \wedge c^2 t) = s \wedge t.$$

□

Bericht 2.5. Detaillierte Auskunft über das Verhalten der Pfade der (eindimensionalen) Brownschen Bewegung geben (präziser als die Schranken aus Satz 2.3 und die Skalierungseigenschaft aus Beob. 2.4) das Gesetz vom iterierten Logarithmus:

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t \log\log t}} = 1 \quad \text{f.s.}, \qquad \limsup_{t\downarrow 0} \frac{B_t}{\sqrt{2t \log\log(1/t)}} = 1 \quad \text{f.s.}$$

und Lévys Stetigkeitsmodul der Brownschen Bewegung:

$$\limsup_{h\searrow 0} \sup_{t\in[0,1]} \frac{B_{t+h} - B_t}{\sqrt{2h \log \frac{1}{h}}} = 1 \quad \text{f.s.} \tag{2.9}$$

(Siehe z.B. [MP, Chapter 1, Chapter 5.1] oder [Kl, Kap. 22].)

Kapitel 3

Schwache Konvergenz und charakteristische Funktionen

Literaturhinweise Die Themen dieses Kapitels finden sich beispielsweise in [Kl, Kap. 13 und 15], [Du, Ch. 3], mit etwas anderem Fokus in [Ka, Ch. 5 und 6], weitergehend in [Bi].

3.1 Vorbemerkungen/Erinnerungen zur mengentheoretischen Topologie

Sei E ein topologischer (meist metrischer) Raum.

Folgende Begriffe werden in diesem Kapitel vorausgesetzt: kompakt, relativkompakt, folgenkompakt, lokalkompakt, totalbeschränkt. Es bezeichne

- $C(E)$ die Menge der stetigen Funktionen von E nach $\mathbb{R}$,
- $C_b(E)$ die Menge der stetigen, beschränkten Funktionen von E nach $\mathbb{R}$,
- $C_c(E)$ die Menge der stetigen Funktionen mit kompaktem Träger von E nach $\mathbb{R}$.

Offenbar: $C_c(E) \subset C_b(E) \subset C(E)$

Eine Teilmenge $A \subset E$ heißt σ-kompakt, wenn es kompakte Mengen $B_n \subset E$ gibt mit $\bigcup_n B_n = A$.

Definition 3.1. Sei μ ein σ-endliches Maß auf $(E, \mathscr{B}(E))$.

i) μ heißt *lokal-endlich* (oder *Borel-Maß*), wenn für alle $x \in E$ eine offene Umgebung U von x existiert mit $\mu(U) < \infty$.

ii) μ heißt *regulär (von innen)*, wenn $\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ kompakt}\}$ für alle $A \in \mathscr{B}(E)$.

iii) μ heißt *regulär (von außen)*, wenn $\mu(A) = \inf\{\mu(U) \mid A \subset U, U \text{ offen}\}$ für alle $A \in \mathscr{B}(E)$.

iv) μ heißt *Radon-Maß*, wenn es lokal endlich und von innen regulär ist.

Es bezeichne

- $\mathcal{M}(E)$ die Menge der Radon-Maße auf E,
- $\mathcal{M}_f(E)$ die Menge der endlichen Radon-Maße auf E,
- $\mathcal{M}_1(E)$ die Menge der Wahrscheinlichkeitsmaße auf E,
- $\mathcal{M}_{\leq 1}(E)$ die Menge der Subwahrscheinlichkeitsmaße auf E.

Beispiel. i) Ist $f\colon \mathbb{R}^d \to [0,\infty]$ messbar mit $f \in \mathcal{L}^1_{\text{loc}}(\mathbb{R}^d)$ und ist λ das Lebesguemaß auf $\mathbb{R}^d$, so ist $\mu = f\lambda$ ein Radon-Maß mit $\mu(B) = \int \mathbf{1}_B(x) f(x) \lambda(\underset{\cdot}{x})$.

ii) Ist $E = \mathbb{R}$, dann ist $\mu(dx) = \mathbf{1}_{\mathbb{R}\setminus\{0\}}(x)\frac{1}{|x|}\lambda(dx) + \delta_0(dx)$ nicht lokal endlich und nicht regulär von außen.

iii) Ist $E = \mathbb{R}$, dann ist $\mu = \sum_{q\in\mathbb{Q}} \delta_q$ zwar σ-endlich, aber nicht regulär.

Beobachtung 3.2. Sei E polnisch und $\mu \in \mathcal{M}_f(E)$. Dann gibt es zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subset E$ mit $\mu(E \setminus K) < \varepsilon$.

Beweis. Wegen der Separabilität von E gibt es für jedes $n \in \mathbb{N}$ eine Folge $(x_i^{(n)})_i \subset E$ mit

$$\bigcup_{i=1}^{\infty} B_{\frac{1}{n}}(x_i^{(n)}) = E.$$

Sei $\varepsilon > 0$, wähle k_n mit

$$\mu\left(\bigcup_{i=1}^{k_n} B_{\frac{1}{n}}(x_i^{(n)})\right) \geq \mu(E) - \frac{\varepsilon}{2^n}.$$

Die Menge $A := \bigcap_{n=1}^{\infty}\bigcup_{i=1}^{k_n} B_{\frac{1}{n}}(x_i^{(n)})$ ist totalbeschränkt, also ist $\overline{A}$ kompakt (man sieht leicht mit einem Diagonalfolgenargument, dass $\overline{A}$ folgenkompakt ist, was für metrische Räume Kompaktheit impliziert) und es gilt

$$\mu(E \setminus \overline{A}) \leq \mu(E \setminus A) \leq \sum_{n=1}^{\infty} \mu\left(E \setminus \bigcup_{i=1}^{k_n} B_{\frac{1}{n}}(x_i^{(n)})\right) \leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

□

Definition 3.3. Sei $\mathcal{F} \subset \mathcal{M}(E)$ und sei $\mathcal{C}$ eine Menge von messbaren Funktionen von E nach $\mathbb{R}$. $\mathcal{C}$ heißt *trennende Familie* (kurz *trennend*) *für* $\mathcal{F}$, falls für alle $\mu, \nu \in \mathscr{F}$ gilt:

$$\forall f \in \mathcal{C} : \int f \, d\mu = \int f \, d\nu \quad \Rightarrow \quad \mu = \nu.$$

Beispiel 3.4. $f : E \to \mathbb{R}$ heißt *Lipschitz-stetig*, wenn eine Konstante $c_f < \infty$ existiert, sodass für alle $x, y \in E$ gilt: $|f(x) - f(y)| \le c_f \cdot d(x, y)$.

$$L_f := \sup_{x \neq y \in E} \frac{|f(x) - f(y)|}{d(x,y)}$$

ist die Lipschitz-Konstante von f. Es bezeichne $\mathrm{Lip}(E)$ die lipschitz-stetigen Funktionen auf E und $\mathrm{Lip}_1(E) = \{f \in \mathrm{Lip}(E) \mid L_f \le 1\}$.

$\mathrm{Lip}_1(E)$ trennend für $\mathcal{M}(E)$, falls E lokalkompakt auch $\mathrm{Lip}_1(E) \cap C_c(E)$.

Beweis. Sei $A \in E$ abgeschlossen und sei $\varepsilon > 0$.

$$\rho_{A,\varepsilon}(x) = 1 - \left(\frac{1}{\varepsilon} d(x, A) \wedge 1\right)$$

erfüllt $\rho_{A,\varepsilon} \equiv 1$ auf A und $\rho_{A,\varepsilon}(x) = 0$, wenn $d(x, A) \ge \varepsilon$. Es ist $\rho_{A,\varepsilon} \in \mathrm{Lip}(E)$ mit $L_{\rho_{A,\varepsilon}} \le \frac{1}{\varepsilon}$.

Seien $\mu_1, \mu_2 \in \mathcal{M}(E)$ mit

$$\int f \, d\mu_1 = \int f \, d\mu_2 \quad \text{für alle } f \in \mathrm{Lip}_1(E) \cap \mathcal{L}^1(\mu_1) \cap \mathcal{L}^1(\mu_2).$$

Wir zeigen

$$\mu_1(K) = \mu_2(K) \quad \text{für alle kompakten } K \subset E, \tag{3.1}$$

dies impliziert $\mu_1 = \mu_2$ wegen der Regularität.

Zu $x \in E$ existiert eine offene Umgebung U_x mit $\mu_1(U_x) < \infty$ und $\mu_2(U_x) < \infty$, denn die μ_i sind lokal endlich. Sei K kompakt,

$$K \subset U := \bigcup_{i=1}^{n} U_{x_i} \quad \text{für geeignete } x_1, \ldots, x_n.$$

Dann ist $\mu_1(U) < \infty$ und $\mu_2(U) < \infty$ und es gilt $\delta := d(U^{\mathrm{c}}, K) > 0$.

Falls $\varepsilon < \delta$, dann gilt

$$\mathbf{1}_K \le \rho_{K,\varepsilon} \le \mathbf{1}_U$$

und

$$\rho_{K,\varepsilon} \to \mathbf{1}_K \text{ punktweise für } \varepsilon \searrow 0,$$

also folgt mit dominierter Konvergenz auch

$$\int \rho_{K,\varepsilon} \, d\mu_i \to \mu_i(K) \text{ punktweise für } \varepsilon \searrow 0.$$

Wegen $\int \rho_{K,\varepsilon} \, d\mu_1 = \frac{1}{\varepsilon} \int \varepsilon \rho_{K,\varepsilon} \, d\mu_1 = \frac{1}{\varepsilon} \int \varepsilon \rho_{K,\varepsilon} \, d\mu_2 = \int \rho_{K,\varepsilon} \, d\mu_2$ folgt $\mu_1(K) = \mu_2(K)$, d.h. (3.1) gilt.

Falls E lokalkompakt ist, kann man die U_x so wählen, dass $\overline{U_x}$ kompakt ist, damit ist auch $\overline{U}$ kompakt und somit gilt dann $\rho_{K,\varepsilon} \in C_c(E)$. □

3.2 Schwache und vage Konvergenz

Definition 3.5. Es sei E ein metrischer Raum.

i) Seien $\mu_1, \mu_2, \ldots, \mu \in \mathcal{M}_f(E)$. Man sagt, μ_n *konvergiert schwach* (engl. *weakly*) *gegen* μ, wenn

$$\text{für alle } f \in C_b(E) \text{ gilt } \int f \, d\mu_n \longrightarrow \int f \, d\mu \text{ für } n \to \infty.$$

Man schreibt auch $\mu_n \to \mu$ schwach oder $\mu_n \xrightarrow{\text{w}} \mu$ oder $\mu = \text{w} - \lim \mu_n$.

ii) Seien $\mu_1, \mu_2, \ldots, \mu \in \mathcal{M}(E)$. Man sagt, μ_n *konvergiert vage* (auch *vag*, engl. *vaguely*) *gegen* μ, wenn

$$\text{für alle } f \in C_c(E) \text{ gilt } \int f \, d\mu_n \longrightarrow \int f \, d\mu \text{ für } n \to \infty.$$

Man schreibt auch $\mu_n \to \mu$ vage oder $\mu_n \xrightarrow{\text{v}} \mu$ oder $\mu = \text{v} - \lim \mu_n$.

Bemerkung 3.6. Wenn E ein polnischer Raum ist, so ist der schwache Limes einer Folge μ_n eindeutig. Falls E zudem lokalkompakt ist, so ist auch der vage Limes eindeutig (dies folgt aus Beispiel 3.4).

Beispiel. Betrachte $E = \mathbb{R}$.

1. Es gilt $\delta_{\frac{1}{n}} \xrightarrow[n\to\infty]{\text{w}} \delta_0$, aber
   $\delta_{\frac{1}{n}}\big((0,\infty)\big) = 1 \not\to 0 = \delta_0\big((0,\infty)\big)$ und $\delta_{\frac{1}{n}}\big((-\infty,0]\big) = 0 \not\to 1 = \delta_0\big((-\infty,0]\big)$.
2. $\delta_n \xrightarrow{\text{v}} 0$-Maß, aber $(\delta_n)_n$ konvergiert nicht schwach.

Beobachtung 3.7. Sei E lokalkompakt und polnisch. Ist $(\mu_n) \subset \mathcal{M}_f(E)$ mit $\mu_n \xrightarrow{\text{v}} \mu$, so gilt $\mu(E) \le \liminf_n \mu_n(E)$.

Beweis. Wähle $f_N \in C_c(E)$ mit $f_N \nearrow 1$. Dann gilt

$$\mu(E) = \int 1 \, d\mu = \lim_{N\to\infty} \int f_N \, d\mu = \lim_{N\to\infty} \lim_{n\to\infty} \int f_N \, d\mu_n \le \lim_{N\to\infty} \lim_{n\to\infty} \mu_n(E) \le \liminf_{n\to\infty} \mu_n(E).$$

$\square$

Satz 3.8 (Portmanteau-Theorem)**.** *Sei E ein metrischer Raum und $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\le 1}(E)$. Dann sind äquivalent:*

i) $\mu = w - \lim \mu_n$.

ii) $\int f \, d\mu_n \to \int f \, d\mu$ *für alle* $f \in C_b(E) \cap \mathrm{Lip}(E)$.

iii) $\int f \, d\mu_n \to \int f \, d\mu$ *für alle beschränkten, messbaren* f *mit* $\mu(U_f) = 0$, *wobei* $U_f \subset E$ *die Unstetigkeitsstellen von* f *bezeichne (d.h.* f *ist* μ*-f.ü. stetig).*

iv) $\liminf_{n\to\infty} \mu_n(E) \ge \mu(E)$ *und* $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ *für alle abgeschlossenen* $F \subset E$.

v) $\limsup_{n\to\infty} \mu_n(E) \le \mu(E)$ *und* $\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$ *für alle offenen* $G \subset E$.

vi) $\mu_n(A) \to \mu(A)$ *für alle* $A \in \mathscr{B}(E)$ *mit* $\mu(\partial A) = 0$.

Wenn E zudem lokalkompakt und polnisch, so sind auch äquivalent:

vii) $\mu = v - \lim \mu_n$ *und* $\mu(E) = \lim \mu_n(E)$.

viii) $\mu = v - \lim \mu_n$ *und* $\mu(E) \geq \limsup \mu_n(E)$.

Beweis. *iv)* $\iff$ *v)* ✓ (F ist abgeschl. $\Leftrightarrow G = E \setminus F$ offen (und $\mu(G) = \mu(E) - \mu(F)$), E ist zugleich offen und abgeschlossen)

iv) & v) $\Longrightarrow$ *vi)* ✓ ($A \setminus \partial A = \mathring{A}$ ist offen, $A \cup \partial A = \overline{A}$ ist abgeschlossen,
$\liminf \mu_n(A) \geq \liminf \mu_n(A \setminus \partial A) = \mu(A \setminus \partial A) = \mu(A)$,
$\limsup \mu_n(A) \leq \limsup \mu_n(A \cup \partial A) = \mu(A \cup \partial A) = \mu(A)$.)

iii) $\Longrightarrow$ *i)* $\Longrightarrow$ *ii)* ✓ (f stetig $\Rightarrow U_f = \varnothing$ und $\mathrm{Lip}(E) \subset C(E)$)

ii) $\Longrightarrow$ *iv)* : $f \equiv 1 \in \mathrm{Lip}(E) \cap C_b(E)$, also $\mu(E) = \int 1 \, d\mu = \lim \mu_n(E)$.

Sei $F \subset E$ abgeschlossen,

$$\rho_{F,\varepsilon}(x) := 1 - \left(\frac{1}{\varepsilon} d(x, F) \wedge 1\right)$$

erfüllt $\rho_{F,\varepsilon} \in C_b(E) \cap \mathrm{Lip}(E)$ und $\mathbf{1}_F \leq \rho_{F,\varepsilon}$ (vgl. Bsp. 3.4). Mit monotoner Konvergenz folgt

$$\limsup_{n\to\infty} \mu_n(F) \leq \inf_{\varepsilon>0} \lim_{n\to\infty} \int \rho_{F,\varepsilon} \, d\mu_n = \inf_{\varepsilon>0} \int \rho_{F,\varepsilon} \, d\mu = \int \mathbf{1}_{\overline{F}} \, d\mu = \int \mathbf{1}_F \, d\mu = \mu(F).$$

vi) $\Longrightarrow$ *iii)* : Sei f beschränkt und messbar mit $\mu(U_f) = 0$. Beobachte zunächst: Stets gilt für $D \in \mathscr{B}(\mathbb{R})$

$$\partial f^{-1}(D) \subset f^{-1}(\partial D) \cup U_f$$

Sei dazu $x \in \partial f^{-1}(D)$ und f stetig in x. Zeige: Dann gilt $x \in f^{-1}(\partial D)$.

Sei $\delta > 0$, da x Stetigkeitspunkt ist, existiert ein $\varepsilon > 0$ mit $f(B_\varepsilon(x)) \subset B_\delta(f(x))$.

Wegen $x \in \partial f^{-1}(D)$ gibt es ein $y \in f^{-1}(D) \cap B_\varepsilon(x)$ und ein $z \notin f^{-1}(D), z \in B_\varepsilon(x)$, d.h. $z \in f^{-1}(\mathbb{R} \setminus D) \cap B_\varepsilon(x)$. Somit ist

$$f(y) \in B_\delta(f(x)) \cap D \text{ und } f(z) \in B_\delta(f(x)) \cap D^{\mathsf{c}}.$$

Da $\delta > 0$ beliebig ist, folgt $x \in f^{-1}(\partial D)$.

Sei nun

$$A = \left\{y \in \mathbb{R} : \mu(f^{-1}(\{y\})) > 0\right\}.$$

(dies sind die Atome des Bildmaßes $\mu \circ f^{-1}$). A ist höchstens abzählbar, also gibt es

$$N \in \mathbb{N} \text{ und } y_0 \leq -\|f\|_\infty < y_1 < y_2 < \ldots < y_{N-1} < \|f\|_\infty < y_N \text{ mit } y_i \notin A \text{ und } |y_{i+1} - y_i| < \varepsilon.$$

Sei $E_i = f^{-1}([y_{i-1}, y_i))$ für $i = 1, \ldots, N$, dann ist $E = \bigcup_{i=1}^N E_i$ und es gilt

$$\mu(\partial E_i) \leq \mu(\{f^{-1}(y_{i-1})\}) + \mu(\{f^{-1}(y_i)\}) + \mu(U_f) = 0.$$

Also folgt

$$\limsup_{n\to\infty} \int f \, d\mu_n \le \limsup_{n\to\infty} \int \sum_{i=1}^N y_i \mathbf{1}_{E_i} \, d\mu_n = \limsup_{n\to\infty} \sum_{i=1}^N \mu_n(E_i) y_i = \sum_{i=1}^N \mu(E_i) y_i \le \varepsilon + \int f \, d\mu.$$

Das analoge Argument für $-f$ und $\varepsilon \searrow 0$ zeigt $iii)$.

i) $\Longrightarrow$ *vii)* ✓ $\quad (C_c(E) \subset C_b(E)$ und $1 \in C_b(E))$

vii) $\Longrightarrow$ *viii)* ✓

viii) $\Longrightarrow$ *vii)* : Es gelte $\mu = \mathrm{v} - \lim \mu_n$ und $\mu(E) \ge \limsup \mu_n(E)$. Nach Beobachtung 3.7 gilt aber auch $\mu(E) \le \liminf_n \mu_n(E)$, d.h. $\mu(E) = \lim \mu_n(E)$.

vii) $\Longrightarrow$ *v)* : Sei $G \subset E$ offen und $\varepsilon > 0$. Wähle (mit Beobachtung 3.2) ein

$$K \subset G \text{ kompakt mit } \mu(G \setminus K) < \varepsilon.$$

Da E lokalkompakt ist, gibt es eine kompakte Menge L mit $K \subset \mathring{L} \subset L \subset G$. Es ist $\delta := d(K, L^\circ) > 0$ und für $\rho_{K,\delta}(x) := 1 - \left(\frac{1}{\delta} d(x,K) \wedge 1\right)$ gilt $\mathbf{1}_K \le \rho_{K,\delta} \le \mathbf{1}_L \le \mathbf{1}_G$ und $\rho_{K,\delta} \in C_c(E)$. Damit folgt

$$\liminf_{n\to\infty} \mu_n(G) \ge \liminf_{n\to\infty} \int \rho_{K,\delta} \, d\mu_n = \int \rho_{K,\delta} \, d\mu \ge \mu(K) \ge \mu(G) - \varepsilon.$$

Mit $\varepsilon \searrow 0$ folgt nun $v)$. □

Definition 3.9. Seien $X, X_1, X_2, \dots$ Zufallsvariablen mit Werten im metrischen Raum E. X_n *konvergiert in Verteilung gegen* X, wenn die Verteilungen $\mathcal{L}(X_n) \in \mathcal{M}_1(E)$ schwach gegen $\mathcal{L}(X)$ konvergieren. In diesem Fall schreibt man $X_n \xrightarrow{\mathrm{D}} X$ oder $X_n \Rightarrow X$ für $n \to \infty$. Gelegentlich schreibt man $X_n \xrightarrow{\mathrm{D}} \mu$ bzw. $X_n \Rightarrow \mu$, wenn $\mu = \mathcal{L}(X)$ und X unspezifiziert bleibt. .

Sind $X, X_1, X_2, \dots$ Zufallsvariablen auf demselben Wahrscheinlichkeitsraum mit Werten in einem separablen metrischen Raum E, so sagt man $X_n \to X$ *stochastisch*, falls $d(X_n, X) \to 0$ stochastisch. In diesem Fall schreibt man $X_n \xrightarrow{\mathbb{P}} X$.

Gilt $X_n \to X$ stochastisch, so gilt auch $X_n \Rightarrow X$, denn für $f \in C_b(E) \cap \mathrm{Lip}(E)$ gilt

$$|\mathbb{E}[f(X_n) - f(X)]| \le \mathbb{E}[L_f \cdot d(X_n, X) \wedge 2 \|f\|_\infty] \to 0$$

(dann verwende Portmanteau-Theorem, Satz 3.8, *ii)*.)

Die Umkehrung gilt i.A. nicht.

Beispiel. Seien $X, X_1, X_2, \dots$ unabhängig und identisch standardnormalverteilt. Dann gilt $X_n \Rightarrow X$, aber nicht $X_n \xrightarrow{\mathbb{P}} X$.

Beobachtung 3.10 (Lemma von Slutsky)**.** Seien $X, X_1, X_2, \dots, Y_1, Y_2, \dots$ Zufallsvariablen mit Werten im separablen metrischen Raum E. Gilt $X_n \Rightarrow X$ und $d(X_n, Y_n) \to 0$ stochastisch, dann gilt auch $Y_n \Rightarrow X$.

Beweis. Sei $f \in C_b(E) \cap \mathrm{Lip}(E)$. Dann gilt $|f(x) - f(y)| \le L_f \cdot d(x,y) \wedge 2\,\|f\|_\infty$. Also gilt

$$\begin{aligned}|\mathbb{E}[f(Y_n) - f(X)]| &\le \mathbb{E}[|f(Y_n) - f(X_n)|] + |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \\ &\le \mathbb{E}[L_f \cdot d(X_n, Y_n) \wedge 2\,\|f\|_\infty] + |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \underset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

□

Lemma 3.11 (Continuous mapping theorem). *Seien (E_1, d_1) und (E_2, d_2) metrische Räume, $\varphi : E_1 \to E_2$ messbar, U_φ die Menge der Unstetigkeitsstellen von φ.*

i) Sind $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\le 1}(E_1)$ mit $\mu_n \overset{w}{\to} \mu$ und $\mu(U_\varphi) = 0$, so gilt $\mu_n \circ \varphi^{-1} \overset{w}{\to} \mu \circ \varphi^{-1}$.

ii) Sind $X, X_1, X_2, \ldots$ E_1-wertige Zufallsvariablen mit $\mathbb{P}(X \in U_\varphi) = 0$ und $X_n \Rightarrow X$, so gilt $\varphi(X_n) \Rightarrow \varphi(X)$.

Beweis. i) Sei $f \in C_b(E_2)$, dann ist $f \circ \varphi : E_1 \to \mathbb{R}$ beschränkt und messbar und es gilt $U_{f\circ\varphi} \subset U_\varphi$, d.h. $\mu(U_{f\circ\varphi}) = 0$. Also gilt nach Satz 3.8 $iii)$

$$\int f\, d(\mu_n \circ \varphi^{-1}) = \int f \circ \varphi\, d\mu_n \underset{n\to\infty}{\longrightarrow} \int f \circ \varphi\, d\mu = \int f\, d(\mu \circ \varphi^{-1}).$$

ii) Setze $\mu_n = \mathcal{L}(X_n)$ und $\mu = \mathcal{L}(X)$. Dann folgt die Aussage mit $i)$.

□

Bemerkung 3.12 (Der Fall $E = \mathbb{R}$). $\mu \in \mathcal{M}_{\le 1}(\mathbb{R})$ ist durch seine Verteilungsfunktion $F_\mu(x) = \mu((-\infty, x])$ festgelegt (siehe z.B. [St1, Satz 1.27] oder [Kl, Satz 1.60]).

Seien $F, F_1, F_2, \ldots$ Verteilungsfunktionen von W'maßen. $(F_n)_n$ *konvergiert schwach* gegen F, geschrieben $F_n \overset{\mathrm{D}}{\to} F$ oder $F_n \Rightarrow F$, :⇔

$$F(x) = \lim_{n\to\infty} F_n(x) \quad \text{für alle Stetigkeitsstellen } x \text{ von } F.$$

Falls $F, F_1, F_2, \ldots$ Verteilungsfunktionen von Sub-W'maßen sind, fordern wir zusätzlich $F(\infty) \ge \limsup_n F_n(\infty)$ (wobei $F(\infty) = \lim_{x\to\infty} F(x)$). Seien $\mu, \mu_1, \mu_2, \cdots \in \mathcal{M}_{\le 1}(\mathbb{R})$ mit zugehörigen Verteilungsfunktionen $F, F_1, F_2, \ldots$, dann gilt

$$\mu_n \overset{w}{\longrightarrow} \mu \quad \Longleftrightarrow \quad F_n \overset{D}{\longrightarrow} F.$$

„⇒“ : x Stetigkeitspunkt von F, so ist $\mu(\{x\}) = \mu(\partial(-\infty, x]) = 0$, also

$$F_n(x) = \mu_n\big((-\infty, x]\big) \underset{n\to\infty}{\longrightarrow} \mu\big((-\infty, x]\big) = F(x)$$

nach Satz 3.8, *vi)*.

„⇐“ : Sei $f \in \mathrm{Lip}(E) \cap C_b(E)$, nach Satz 3.8, *ii)* genügt es zu zeigen, dass

$$\int f\, d\mu_n \underset{n\to\infty}{\longrightarrow} \int f\, d\mu \tag{3.2}$$

Sei $\varepsilon > 0$, wähle $y_0 < y_1 < \cdots < y_N$ Stetigkeitsstellen von F mit $F(y_0) < \varepsilon$, $F(y_N) > F(\infty) - \varepsilon$, $y_i - y_{i-1} < \varepsilon$.

Also (L_f sei die Lipschitz-Konstante von f)

$$\begin{aligned}\int f\,d\mu_n \leq \|f\|_\infty\big(F_n(y_0) + F_n(\infty) - F_n(y_N)\big) \\ + \sum_{i=1}^{N} \big(f(y_i) + L_f\varepsilon\big)\big(F_n(y_i) - F_n(y_{i-1})\big).\end{aligned}$$

Nach Vor. gilt $F_n(y_i) \to F(y_i)$, $F_n(\infty) \to F(\infty)$ für $n \to \infty$, demnach

$$\begin{aligned}\limsup_{n\to\infty} \int f\,d\mu_n &\leq \varepsilon\big(2\|f\|_\infty L_f\big) + \sum_{i=1}^{N} f(y_i)\big(F(y_i) - F(y_{i-1})\big) \\ &\leq 2\varepsilon\big(2\|f\|_\infty L_f\big) + \int f\,d\mu,\end{aligned}$$

mit $\varepsilon \downarrow 0$ folglich

$$\limsup_{n\to\infty} \int f\,d\mu_n \leq \int f\,d\mu.$$

Dasselbe Argument angewendet auf $-f$ zeigt $\liminf_{n\to\infty} \int f\,d\mu_n \geq \int f\,d\mu$, d.h. (3.2) gilt.

3.3 Straffheit

Definition 3.13. Sei E ein metrischer Raum. $\mathcal{K} \subset \mathcal{M}_f(E)$ heißt *straff* (engl. *tight*), falls für jedes $\varepsilon > 0$ eine kompakte Menge $K \subset E$ existiert mit

$$\sup_{\mu\in\mathcal{K}} \mu(E \setminus K) < \varepsilon.$$

Beispiel 3.14. i) Ist E polnisch, so ist $\{\mu\}$ für jedes $\mu \in \mathcal{M}_f(E)$ straff nach Beobachtung 3.2. Genauso ist jede endliche Menge $\{\mu_1, \mu_2, \dots, \mu_n\} \subset \mathcal{M}_f(E)$ straff.

ii) Ist E kompakt, so sind $\mathcal{M}_1(E)$ und $\mathcal{M}_{\leq 1}(E)$ straff.

iii) Sind $X_i, i \in I$ reelle Zufallsvariablen mit $\sup \mathbb{E}[|X_i|] =: c < \infty$, so ist $\{\mathcal{L}(X_i) \mid i \in I\}$ straff, denn mit der Markov-Ungleichung gilt

$$\mathbb{P}\left(|X_i| > \frac{c}{\varepsilon}\right) \leq \frac{\varepsilon}{c}\mathbb{E}[|X_i|] \leq \varepsilon$$

für alle $i \in I$ und für jedes $\varepsilon > 0$.

iv) Auf $E = \mathbb{R}$ sind $\{\delta_n \mid n \in \mathbb{N}\}$, $\{\mathcal{N}(0, n) \mid n \in \mathbb{N}\}$, $\{\mathrm{Unif}([-n, n]) \mid n \in \mathbb{N}\}$ nicht straff.

Satz 3.15 (Satz von Prohorov[1]). *Sei E ein metrischer Raum und $\mathcal{K} \subset \mathcal{M}_{\leq 1}(E)$. Dann gilt:*

i) Ist $\mathcal{K}$ straff, dann ist $\mathcal{K}$ relativ (folgen-)kompakt bezüglich schwacher Konvergenz.

[1] Yuri Vasilyevich Prohorov, 1929–2013

ii) Ist E zudem polnisch, so gilt auch die Umkehrung, d.h. ist $\mathcal{K}$ relativ (folgen-)kompakt bzgl. schwacher Konvergenz, so ist $\mathcal{K}$ straff.

Korollar 3.16. *Ist E kompakt, so sind $\mathcal{M}_{\leq 1}(E)$ und $\mathcal{M}_1(E)$ schwach (folgen-) kompakt.*

Beweis von Satz 3.15. i) Sei $\mathcal{K}$ straff. Wähle kompakte Mengen $K_1 \subset K_2 \subset \ldots \subset E$, sodass für alle $j \in \mathbb{N}$ gilt

$$\sup_{\mu \in \mathcal{K}} \mu\left(E \setminus K_j\right) \leq \frac{1}{j}.$$

$\bigcup_{j=1}^{\infty} K_j$ ist σ-kompakt und es gilt $\mu\left(E \setminus \bigcup_{j=1}^{\infty} K_j\right) = 0$ für alle $\mu \in \mathcal{K}$. Wir nehmen also o.E. an, dass E selbst σ-kompakt (insbesondere separabel) ist, ansonsten schränke die μ ein auf $E' := \bigcup_{j=1}^{\infty} K_j$.

Sei $x_1, x_2, \ldots$ eine Aufzählung einer dichten Teilmenge von E.

$$\mathcal{H} := \left\{ \bigcup_{i=1}^{m} K_{j_i} \cap \overline{B_{\varepsilon_i}(x_{l_i})} \mid m \in \mathbb{N},\ j_i \in \mathbb{N},\ l_i \in \mathbb{N},\ \varepsilon_i > 0,\ \varepsilon_i \in \mathbb{Q} \right\}$$

ist ein abzählbares System von kompakten Teilmengen von E. Sei $(\mu_n) \subset \mathcal{K}$ und sei $\mathcal{H} = \{H_1, H_2, \ldots\}$ eine Aufzählung von $\mathcal{H}$. Wähle eine Teilfolge $n_k \nearrow \infty$, sodass

$$\lim_{k \to \infty} \mu_{n_k}(H) =: \alpha(H)$$

für alle $H \in \mathcal{H}$ existiert (möglich durch Diagonalargument).

Für $G \subset E$ offen sei

$$\mu^*(G) := \sup\left\{\alpha(H) \mid H \in \mathcal{H}, H \subset G\right\}, \quad \mu^*(\varnothing) := 0,$$

für $B \subset E$ beliebig sei

$$\mu^*(B) := \inf\left\{\mu^*(G) \mid G \text{ offen}, G \supset B\right\}$$

(beachte: ist μ^* wohldefiniert, für offenes G fallen die beiden Setzungen zusammen).

Zeige nun: μ^* ist ein äußeres Maß.
Nach Definition und Konstruktion gilt

$$\mu^*(\varnothing) = 0 \quad \text{und} \quad \mu^*(B) \leq \mu^*(B') \text{ für } B \subset B'.$$

Zur σ-Subadditivität: Seien $G_1, G_2, \ldots \subset E$ offene Mengen, $\mathcal{H} \ni H \subset \bigcup_{n=1}^{\infty} G_n$. H ist kompakt, daher existieren $r \in \mathbb{N}$, $j_i \in \mathbb{N}$, $\varepsilon_i \in (0, \infty) \cap \mathbb{Q}$ und paarweise verschiedene $m_i \in \mathbb{N}$ mit

$$H \subset \bigcup_{i=1}^{r} B_{\varepsilon_i}(x_{j_i}) \quad \text{und} \quad \overline{B_{\varepsilon_i}(x_{j_i})} \subset G_{m_i} \text{ für } i = 1, \ldots, r$$

und nach Konstruktion ist $H \subset K_{j_0}$ für ein (genügend großes) $j_0 \in \mathbb{N}$.

Sei

$$H_m := K_{j_0} \cap \bigcup_{i=1}^{r} \left\{ \overline{B_{\varepsilon_i}(x_{j_i})} \mid \overline{B_{\varepsilon_i}(x_{j_i})} \subset G_m \right\}.$$

Es ist $H_m \subset G_m$ und $H \subset H_{m_1} \cup \ldots \cup H_{m_r}$, also

$$\alpha(H) \leq \alpha(H_{m_1}) + \ldots + \alpha(H_{m_r}) \leq \mu^*(G_{m_1}) + \ldots + \mu^*(G_{m_r}) \leq \sum_{n=1}^{\infty} \mu^*(G_n).$$

Nehme das Supremum über alle $H \subset G$, so folgt

$$\mu^*\left(\bigcup_{n=1}^{\infty} G_n\right) \leq \sum_{n=1}^{\infty} \mu^*(G_n).$$

Seien nun $B_1, B_2, \ldots \subset E$ beliebig. Wähle G_n offen mit $B_n \subset G_n$ und

$$\mu^*(B_n) \geq \mu^*(G_n) - \frac{\varepsilon}{2^n} \quad \text{für } \varepsilon > 0.$$

Dann gilt

$$\mu^*\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \mu^*\left(\bigcup_{n=1}^{\infty} G_n\right) \leq \sum_{n=1}^{\infty} \mu^*(G_n) \leq \sum_{n=1}^{\infty} \left(\mu^*(B_n) + \frac{\varepsilon}{2^n}\right) \leq \varepsilon + \sum_{n=1}^{\infty} \mu^*(B_n).$$

Mit $\varepsilon \searrow 0$ folgt die σ-Subadditivität von μ^*.

Zeige weiter, dass

jede offene Menge $G \subset E$ μ^*-messbar ist,

d.h. für alle $B \subset E$ gilt $\mu^*(B) = \mu^*(B \cap G) + \mu^*(B \cap G^c)$.

Seien dazu $G \subset E$ offen, $B \subset E$ beliebig und $\varepsilon > 0$. Wähle

$O \supset B$ offene Obermenge mit $\mu^*(O) \leq \mu^*(B) + \varepsilon$,
$H_1 \in \mathcal{H}$ mit $H_1 \subset O \cap G$ und $\mu^*(O \cap G) \leq \alpha(H_1) + \varepsilon$,
$H_2 \in \mathcal{H}$ mit $H_2 \subset O \cap H_1^c$ und $\mu^*(O \cap H_1^c) \leq \alpha(H_2) + \varepsilon$.

Dann gilt $H_1 \cap H_2 = \varnothing$, $O \cap G^c \subset O \cap H_1^c$ und $H_1 \cup H_2 \subset O$. Mit Monotonie von μ^* folgt nun

$$\begin{aligned}\mu^*(B) + \varepsilon \geq \mu^*(O) \geq \alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2) &\geq \mu^*(O \cap G) + \mu^*(O \cap G^c) - 2\varepsilon \\ &\geq \mu^*(B \cap G) + \mu^*(B \cap G^c) - 2\varepsilon \\ &\geq \mu^*(B) - 2\varepsilon\end{aligned}$$

Mit $\varepsilon \searrow 0$ folgt nun die μ^*-Messbarkeit von G. μ^* ist also ein äußeres Maß und die Erzeugermengen von $\mathcal{B}(E)$, die offenen Mengen, sind μ^*-messbar. Nach dem Satz von Carathéodory (z.B. [St1, Satz 1.22] oder [Kl, Satz 1.41/1.53]) μ^* ein Maß auf $\mathcal{B}(E)$.

Zeige nun, dass $\mu^* = \mathrm{w} - \lim_{k\to\infty} \mu_{n_k}$. Es gilt für alle $j \in \mathbb{N}$

$$\mu^*(E) \geq \alpha(K_j) = \lim_{k\to\infty} \mu_{n_k}(K_j) \geq \limsup_{k\to\infty} \mu_{n_k}(E) - \frac{1}{j},$$

also $\mu^*(E) \geq \limsup \mu_{n_k}(E)$.

Sei $G \subset E$ offen und $H \in \mathcal{H}$ mit $H \subset G$. Dann gilt

$$\alpha(H) = \lim_{k\to\infty} \mu_{n_k}(H) \le \lim_{k\to\infty} \mu_{n_k}(G) \le \liminf_{k\to\infty} \mu_{n_k}(G).$$

Nehme nun das Supremum über $H \subset G, H \in \mathcal{H}$, dann folgt $\mu^*(G) \le \liminf \mu_{n_k}(G)$. Mit Satz 3.8 $v)$ folgt nun die Behauptung.

ii) Sei E polnisch und $x_1, x_2, \dots$ eine Aufzählung einer dichten Teilmenge von E. Setze

$$A_{n,N} := \bigcup_{i=1}^{N} B_{\frac{1}{n}}(x_i),$$

es gilt $A_{n,N} \nearrow_{N\to\infty} E$ für jedes $n \in \mathbb{N}$.

Sei $\mathcal{K} \subset \mathcal{M}_{\le 1}(E)$ schwach relativ (folgen-) kompakt und

$$\delta := \sup_{n\in\mathbb{N}} \inf_{N\in\mathbb{N}} \sup_{\mu\in\mathcal{K}} \mu(A_{n,N}^{\mathsf{c}}).$$

Zeige, dass $\delta = 0$. Sei n so groß, dass es für jedes $N \in \mathbb{N}$ ein $\mu_N \in \mathcal{K}$ gibt mit $\mu_N(A_{n,N}^{\mathsf{c}}) \ge \frac{\delta}{2}$. Sei N_k eine Teilfolge mit $N_k \nearrow \infty$ und $\mu_{N_k} \xrightarrow{\mathrm{w}} \tilde{\mu} \in \mathcal{M}_{\le 1}(E)$. Da $A_{n,N}^{\mathsf{c}}$ abgeschlossen ist, gilt mit Satz 3.8 $iv)$

$$\tilde{\mu}(A_{n,N}^{\mathsf{c}}) \ge \limsup_{k\to\infty} \mu_{N_k}(A_{n,N}^{\mathsf{c}}) \ge \limsup_{k\to\infty} \mu_{N_k}(A_{n,N_k}^{\mathsf{c}}) \ge \frac{\delta}{2}.$$

Mit $N \to \infty$ folgt $0 = \tilde{\mu}(\varnothing) = \lim_{N\to\infty} \tilde{\mu}(A_{n,N}^{\mathsf{c}})$, also auch $\delta = 0$.

Zu $\varepsilon > 0$ wähle $n, N_n \in \mathbb{N}$ mit $\sup_{\mu\in\mathcal{K}} \mu(A_{n,N_n}^{\mathsf{c}}) < \frac{\varepsilon}{2^n}$. $A := \bigcap_{n=1}^{\infty} A_{n,N_n}$ ist totalbeschränkt, d.h. wegen der Vollständigkeit von E ist $\overline{A}$ kompakt. Dann gilt für alle $\mu \in \mathcal{K}$:

$$\mu\big((\overline{A})^{\mathsf{c}}\big) \le \mu(A^{\mathsf{c}}) = \mu\left(\bigcup_{n=1}^{\infty} A_{n,N_n}^{\mathsf{c}}\right) \le \sum_{n=1}^{\infty} \mu(A_{n,N_n}^{\mathsf{c}}) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon,$$

also ist $\mathcal{K}$ straff. □

Beobachtung 3.17. Sei E ein polnischer Raum und seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_{\le 1}(E)$. Dann sind äquivalent:

i) $\mu = \mathrm{w} - \lim \mu_n$.

ii) $\{\mu_n \mid n \in \mathbb{N}\}$ ist straff und für eine trennende Familie $\mathcal{C} \subset C_b(E)$ gilt

$$\int f\, d\mu_n \to \int f\, d\mu \quad \text{für alle } f \in \mathcal{C}.$$

Beweis. $i) \Rightarrow ii)$. Sei $\mu = \mathrm{w} - \lim \mu_n$. Nach Definition gilt $\int f\, d\mu_n \to \int f\, d\mu$ für alle $f \in C_b(E)$, insbesondere also auch für alle $f \in \mathcal{C}$ für jede trennende Familie $\mathcal{C} \subset C_b(E)$. Die Straffheit von $\{\mu_n \mid n \in \mathbb{N}\}$ folgt mit Satz 3.15 $ii)$.

$ii) \Rightarrow i)$. Angenommen $\mu \ne \mathrm{w} - \lim \mu_n$. Dann gibt es ein $g \in C_b(E)$, ein $\varepsilon > 0$ und eine Teilfolge $(n_k)_k, n_k \nearrow \infty$ mit

$$\left|\int g\, d\mu_{n_k} - \int g\, d\mu\right| \ge \varepsilon \quad \text{für alle } k.$$

Nach Satz 3.15 $i)$ gibt es eine Teilfolge $(n_{k_j})_j$ und ein $\nu \in \mathcal{M}_1(E)$ mit $\mu_{n_{k_j}} \xrightarrow[j\to\infty]{\mathrm{w}} \nu$. Dann gilt für alle $f \in \mathcal{C}$

$$\int f\,d\nu = \lim_{j\to\infty} \int f\,d\mu_{n_{k_j}} = \int f\,d\mu.$$

Da $\mathcal{C}$ trennend ist, gilt $\mu = \nu$. Andererseits ist aber auch

$$\left| \int g\,d\nu - \int g\,d\mu \right| \geq \varepsilon,$$

was zu einem Widerspruch führt. Es muss also $\mu = \mathrm{w} - \lim \mu_n$ gelten. □

Bemerkung 3.18. Sei E lokalkompakt und polnisch und seien $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_f(E)$. Dann sind äquivalent:

i) $\mu = \mathrm{w} - \lim \mu_n$.

ii) $\mu = \mathrm{v} - \lim \mu_n$ und $\mu(E) \geq \limsup \mu_n(E)$.

iii) $\mu = \mathrm{v} - \lim \mu_n$ und $\{\mu_n \mid n \in \mathbb{N}\}$ ist straff.

Beweis. i) ⇔ ii) gilt nach Satz 3.8.

i) ⇒ iii) folgt aus Satz 3.15 (eine schwach konvergente Folge ist selbst schwach relativ folgenkompakt).

iii) ⇒ i) : Sei $L \subset E$ kompakt mit $\sup_n \mu_n(E \setminus L) \leq 1$, $h \in C_c(E)$ mit $h \geq \mathbf{1}_L$. Dann gilt

$$\sup_n \mu_n(E) \leq 1 + \sup_n \int h\,d\mu_n < \infty,$$

demnach ist auch $c \coloneqq \mu(E) \vee \sup_n \mu_n(E) < \infty$. Dann sind

$$\mu' = \frac{1}{c}\mu, \mu_n' = \frac{1}{c}\mu_n \in \mathcal{M}_{\leq 1}(E),$$

also $\mu_n' \xrightarrow{\mathrm{v}} \mu'$.

Da E lokalkompakt ist, ist $C_c(E)$ trennend für $\mathcal{M}_{\leq 1}(E)$, also folgt $\mu' = \mathrm{w} - \lim \mu_n'$ mit Beobachtung 3.17 und damit auch $\mu = \mathrm{w} - \lim \mu_n$. □

3.4 Charakteristische Funktionen

Vorbemerkung. Sei E ein messbarer Raum. Eine Funktion

$$f\colon E \to \mathbb{C} \cong \mathbb{R}^2, \quad x \mapsto \mathrm{Re} f(x) + i\,\mathrm{Im} f(x)$$

ist genau dann messbar, wenn $\mathrm{Re} f$ und $\mathrm{Im} f$ messbar sind (beachte $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) = \mathscr{B}(\mathbb{R}^2) \cong \mathscr{B}(\mathbb{C})$). Ist μ ein Maß auf E, dann heißt f μ-integrierbar (schreibe auch $\mathcal{L}^1_{\mathbb{C}}(\mu)$), wenn $\mathrm{Re} f,\ \mathrm{Im} f \in \mathcal{L}^1(\mu)$. Es gilt

$$f \in \mathcal{L}^1_{\mathbb{C}}(\mu) \iff |f| = \sqrt{f\overline{f}} \in \mathcal{L}^1(\mu),$$

denn $|\mathrm{Re} f|, |\mathrm{Im} f| \leq |f| = ((\mathrm{Re} f)^2 + (\mathrm{Im} f)^2)^{\frac{1}{2}} \leq \sqrt{2}\,(|\mathrm{Re} f| + |\mathrm{Im} f|)$. Man setzt

$$\int f\, d\mu := \int \mathrm{Re} f\, d\mu + i \int \mathrm{Im} f\, d\mu.$$

Das Integral ist $\mathbb{C}$-linear (Übung) und es gilt

$$\left| \int f\, d\mu \right| \leq \int |f|\, d\mu.$$

Definition 3.19. Sei $\mu \in \mathcal{M}_f(\mathbb{R}^d)$. Die Funktion

$$\varphi_\mu : \mathbb{R}^d \to \mathbb{C}, \quad t \mapsto \int e^{i\langle t,x\rangle} \mu(\, dx) \tag{3.3}$$

heißt die *charakteristische Funktion* von μ. (Wir notieren das Skalarprodukt als $\langle t, x\rangle = \sum_{i=1}^d t_j x_i$.) Für eine $\mathbb{R}^d$-wertige Zufallsvariable X schreiben wir

$$\varphi_X(t) = \mathbb{E}\left[e^{i\langle t,X\rangle}\right] = \varphi_{\mathcal{L}(X)}(t). \tag{3.4}$$

Bemerkung. In der Analysis ist auch die Bezeichnung *Fourier-Transformierte* üblich, zum Teil mit anderen Vorzeichenkonventionen.

Beispiel 3.20. i) Ist X diskret, so gilt $\varphi_X(t) = \sum_x \mathbb{P}(X = x) e^{i\langle t,x\rangle}$, insbesondere ist

$$\begin{aligned}
\varphi_{\mathrm{Ber}(p)}(t) &= pe^{it} + 1 - p, \\
\varphi_{\mathrm{Bin}(n,p)}(t) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{itk} = (pe^{it} + 1 - p)^n, \\
\varphi_{\mathrm{Poi}(\lambda)}(t) &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk} = e^{\lambda(e^{it}-1)}.
\end{aligned}$$

ii) Es ist $\varphi_{\mathcal{N}(\mu,\sigma^2)}(t) = e^{it\mu - \frac{1}{2}\sigma^2 t}$. Betrachte ohne Einschränkung (vgl. Lemma 3.21, *ii)*) den Fall $\mu = 0, \sigma^2 = 1$. Es gilt

$$\varphi_{\mathcal{N}(\mu,\sigma^2)}(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = e^{-\frac{t^2}{2}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-it)^2}{2}}\, dx}_{=1} = e^{-\frac{t^2}{2}}.$$

iii) $\varphi_{\mathrm{Unif}([-1,1])}(t) = \dfrac{1}{2} \displaystyle\int_{-1}^{1} e^{itx}\, dx = \dfrac{1}{2} \displaystyle\int_{-1}^{1} \cos(tx)\, dx = \dfrac{\sin(t)}{t}.$

Lemma 3.21. *Seien X und Y $\mathbb{R}^d$-wertige Zufallsvariablen, $a \in \mathbb{R}$ und $b \in \mathbb{R}^d$. Dann gilt:*

i) $|\varphi_X(t)| \leq 1 = \varphi_X(0)$ *und* φ_X *ist gleichmäßig stetig.*

ii) $\varphi_{aX+b} = e^{i\langle t,b\rangle} \varphi_X(at)$.

iii) $\varphi_{-X}(t) = \overline{\varphi_X(t)}$. *Insbesondere ist* φ_X *reell, wenn* X *symmetrisch verteilt ist.*

iv) Sind X, Y unabhängig, so ist $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$. Analog ist $\varphi_{\mu*\nu} = \varphi_\mu\varphi_\nu$ für $\mu, \nu \in \mathcal{M}_f(\mathbb{R}^d)$.

v) $0 \le 1 - \mathrm{Re}\varphi_X(2t) \le 4(1 - \mathrm{Re}\varphi_X(t))$.

Beweis. i) Offenbar gilt $|\varphi_X(t)| \le 1 = \varphi_X(0)$.

Zur gleichmäßigen Stetigkeit von φ_X : Sei $\varepsilon > 0$ und K so groß, dass

$$\mathbb{P}\big(X \notin [-K, K]^d\big) < \frac{\varepsilon}{4}.$$

Dann gilt (beachte $|e^{iy} - 1| \le |y|,\ y \in \mathbb{R}$) für $t, t' \in \mathbb{R}^d$ mit $|t - t'| < \delta := \frac{\varepsilon}{4\sqrt{d}K}$

$$\begin{aligned}
|\varphi_X(t) - \varphi_X(t')| &\le \mathbb{E}\left[\left|e^{i\langle t,X\rangle} - e^{i\langle t',X\rangle}\right|\right] \\
&\le 2\mathbb{P}\left(X \notin [-K, K]^d\right) + \mathbb{E}\left[\left|e^{i\langle t-t',X\rangle} - 1\right| \cdot \left|e^{i\langle t',X\rangle}\right| \cdot \mathbf{1}_{[-K,K]^d}(X)\right] \\
&\le 2\mathbb{P}\left(X \notin [-K, K]^d\right) + \mathbb{E}\left[|\langle t - t', X\rangle| \cdot \mathbf{1}_{[-K,K]^d}(X)\right] \\
&\le 2\mathbb{P}\left(X \notin [-K, K]^d\right) + \mathbb{E}\left[\delta\sqrt{d}2K \cdot \mathbf{1}_{[-K,K]^d}(X)\right] \\
&\le \varepsilon.
\end{aligned}$$

ii) Es gilt:

$$\varphi_{aX+b}(t) = \mathbb{E}\left[e^{i\langle t,aX+b\rangle}\right] = \mathbb{E}\left[e^{i\langle t,b\rangle}e^{i\langle at,X\rangle}\right] = e^{i\langle t,b\rangle}\varphi_X(at).$$

iii) Es gilt:

$$\varphi_{-X}(t) = \mathbb{E}\left[e^{i\langle t,-X\rangle}\right] = \mathbb{E}\left[e^{-i\langle t,X\rangle}\right] = \overline{\mathbb{E}\left[e^{i\langle t,X\rangle}\right]} = \overline{\varphi_X(t)}.$$

iv) Da X und Y unabhängig sind, gilt:

$$\varphi_{X+Y} = \mathbb{E}\left[e^{i\langle t,X+Y\rangle}\right] = \mathbb{E}\left[e^{i\langle t,X\rangle}e^{i\langle t,Y\rangle}\right] = \mathbb{E}\left[e^{i\langle t,X\rangle}\right] \cdot \mathbb{E}\left[e^{i\langle t,Y\rangle}\right] = \varphi_X(t)\varphi_Y(t).$$

v) Mit Hilfe der Additionstheoreme des Kosinus gilt:

$$\begin{aligned}
0 \le 1 - \cos(2\langle t, X\rangle) = 2(1 - \cos^2(\langle t, X\rangle)) &= 2(1 + \cos(\langle t, X\rangle))(1 - \cos(\langle t, X\rangle)) \\
&\le 4(1 - \cos(\langle t, X\rangle)).
\end{aligned}$$

Da $\mathrm{Re}\,\varphi_X(t) = \mathbb{E}\left[\cos(\langle t, X\rangle)\right]$, folgt damit die Behauptung.

□

Definition 3.22. Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. $\mathcal{C} \subset C_b(E, \mathbb{K})$ heißt eine *Algebra*, wenn gilt

i) $1 \in \mathcal{C}$.

ii) $f + g,\ fg \in \mathcal{C}$ für alle $f, g \in \mathcal{C}$.

iii) $af \in \mathcal{C}$ für alle $f \in \mathcal{C},\ a \in \mathbb{K}$.

$\mathcal{C}$ heißt *Punkte trennend*, wenn für alle $x \ne y \in E$ ein $f \in \mathcal{C}$ existiert mit $f(x) \ne f(y)$.

Satz 3.23 (Satz von Stone-Weierstraß[2]). *Sei E ein kompakter, topologischer Raum, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $\mathcal{C} \subset C_b(E, \mathbb{K})$ eine Punkte trennende Algebra. Falls $\mathbb{K} = \mathbb{C}$ sei zudem $\mathcal{C}$ abgeschlossen unter komplexer Konjugation. Dann liegt $\mathcal{C}$ bezüglich der Supremumsnorm dicht in $C_b(E, \mathbb{K})$.*

Beweis. Bemerke zunächst: Der Abschluss $\overline{\mathcal{C}}$ von $\mathcal{C}$ bezüglich der Supremumsnormtopologie ist selbst eine Algebra. Betrachte zunächst den Fall $\mathbb{K} = \mathbb{R}$.

Wir zeigen zuerst, dass

$$f, g \in \overline{\mathcal{C}} \quad \Longrightarrow \quad f \wedge g,\ f \vee g \in \overline{\mathcal{C}}.$$

Sei dazu $p_n(t)$ eine Folge von reellen Polynomen mit $\sup_{t \in [0,1]} |p_n(t) - \sqrt{t}| \to 0, n \to \infty$ (wähle zum Beispiel die *Bernstein-Polynome* $p_n(t) = \sum_{k=0}^n \binom{n}{k} t^k (1-t)^{n-k} \sqrt{\frac{k}{n}}$). Sei $0 \neq f \in \overline{\mathcal{C}}$. Dann gilt

$$\overline{\mathcal{C}} \ni \|f\|_\infty p_n\Big(\frac{f(\cdot)^2}{\|f\|_\infty^2}\Big) \xrightarrow[n\to\infty]{\text{gleichmäßig}} |f| \in \overline{\mathcal{C}}.$$

Somit sind auch $f \vee g = \frac{1}{2}(f + g + |f-g|) \in \overline{\mathcal{C}}$ und $f \wedge g = \frac{1}{2}(f + g - |f-g|) \in \overline{\mathcal{C}}$ für $f, g \in \overline{\mathcal{C}}$.

Sei nun $f \in \mathcal{C}_b(E, \mathbb{K})$, $x \in E$ und $\varepsilon > 0$. Zeige, dass es ein $g_{x,\varepsilon} \in \overline{\mathcal{C}}$ gibt mit

$$g_{x,\varepsilon}(x) = f(x) \quad \text{und} \quad g_{x,e} \le f + \varepsilon. \tag{3.5}$$

Da $\mathcal{C}$ Punkte trennend ist, gibt es zu jedem $z \in E,\ z \neq x$

$$\text{ein } H_z \in \mathcal{C} \text{ mit } H_z(z) \neq H_z(x) = 0.$$

Setze

$$h_z(y) := f(x) + \frac{f(z) - f(x)}{H_z(z)} H_z(y), \quad h_x(y) = f(x).$$

Dann ist $h_z \in \mathcal{C}$ und es gilt

$$\text{für alle } z \in E\text{: } h_z(x) = f(x) \text{ und } h_z(z) = f(z).$$

Also existiert eine offene Umgebung U_z von z mit $h_z(y) \le f(y) + \varepsilon$ für alle $y \in U_z$. Da E kompakt ist, lässt sich E mit endlichen vielen solcher Umgebungen überdecken, d.h. $E \subset U_{z_1} \cup \ldots \cup U_{z_n}$ für geeignete $z_1, \ldots, z_n$. Somit erfüllt

$$g_{x,\varepsilon} := \min\{h_{z_1}, \ldots, h_{z_n}\} \in \overline{\mathcal{C}} \tag{3.6}$$

die geforderten Bedingungen in (3.5).

Sei nun $x \in E$ und $\varepsilon > 0$. Wähle $g_{x,\varepsilon}$ gemäß (3.6) und wähle eine offene Umgebung V_x von x mit

$$g_{x,\varepsilon}(y) \ge f(y) - \varepsilon \quad \text{für alle } y \in V_x.$$

Da E kompakt ist, ist $E \subset V_{x_1} \cup \ldots \cup V_{x_m}$ für geeignete $x_1, \ldots, x_m$. Setze

$$g := \max\{g_{x_1,\varepsilon}, \ldots, g_{x_m,\varepsilon}\}.$$

[2] nach Marshall Harvey Stone, 1903–1989 und Karl Theodor Wilhelm Weierstraß, 1815–1897 benannt

Dann gilt $f - \varepsilon \le g \le f + \varepsilon$, also liegt $\mathcal{C}$ dicht in $C_b(E, \mathbb{R})$.

Betrachte nun den Fall $\mathbb{K} = \mathbb{C}$. Für $f \in \mathcal{C}$ ist $\mathrm{Re} f = \frac{1}{2}(f + \overline{f})$, $\mathrm{Im} f = \frac{1}{2i}(f - \overline{f}) \in \mathcal{C}$. $\mathcal{C}' := \{\mathrm{Re} f \mid f \in \mathcal{C}\} \subset C_b(E, \mathbb{R})$ ist eine Punkte trennende Algebra. Es gilt

$$\mathrm{Re}(fg) = \mathrm{Re}(f)\mathrm{Re}(g) - \mathrm{Im}(f)\mathrm{Im}(g) \quad \text{und} \quad \mathrm{Im}(f) = \mathrm{Re}(-if),$$

also ist $\mathcal{C} = \mathcal{C}' + i\mathcal{C}'$ und liegt damit nach dem ersten Fall dicht in $C_b(E, \mathbb{C})$. □

Korollar 3.24. *Sind $\mu, \nu \in \mathcal{M}_f(\mathbb{R}^d)$ mit $\varphi_\mu = \varphi_\nu$, so gilt $\mu = \nu$.*

Beweis. Wir zeigen

$$\int f \, d\mu = \int f \, d\nu \quad \text{für } f \in C_c(\mathbb{R}^d),$$

dies impliziert $\mu = \nu$ nach Beispiel 3.4.

Sei $f \in C_c(\mathbb{R}^d)$, o.E. $f \not\equiv 0$. Sei $0 < \varepsilon < 1$ und K so groß, dass

$$\mathrm{supp}(f) \subset [-K, K]^d, \;\; \mu\left(\mathbb{R}^d \setminus [-K, K]^d\right) \le \frac{\varepsilon}{2\,\|f\|_\infty + 1} \;\; \text{und} \;\; \nu\left(\mathbb{R}^d \setminus [-K, K]^d\right) \le \frac{\varepsilon}{2\,\|f\|_\infty + 1}.$$

Für $m = (m_1, \ldots, m_d) \in \mathbb{Z}^d$ und $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ sei

$$f_{m,K}(x) = \exp\left(i \frac{\pi}{K} \langle m, x \rangle\right).$$

Die von $\{f_{m,K} \mid m \in \mathbb{Z}^d\}$ erzeugte Algebra $\mathcal{C}$ trennt Punkte von $\mathbb{R}^d / 2K\mathbb{Z}^d$. Nach Satz 3.23 gibt es ein $g \in \mathcal{C}$ mit $|f - g| < \varepsilon$ auf $[-K, K]^d$ und es gilt (beachte: $g \in \mathcal{C}$ ist $2K$-periodisch, insbesondere gilt $\|g\|_\infty = \sup_{x \in [-K,K]^d} |g(x)|$)

$$\begin{aligned}\int |f - g| \, d\mu &\le \varepsilon\mu([-K, K]^d) + (2\,\|f\|_\infty + 1)\mu\left(\mathbb{R}^d \setminus [-K, K]^d\right) \le \varepsilon\mu([-K, K]^d) + \varepsilon \\ &\le \varepsilon\mu(\mathbb{R}^d) + \varepsilon.\end{aligned}$$

Die analoge Abschätzung gilt ebenso für ν. Zusammen folgt daher

$$\left|\int f \, d\mu - \int f \, d\nu\right| \le \underbrace{\left|\int g \, d\mu - \int g \, d\nu\right|}_{=0 \text{ da } \varphi_\mu = \varphi_\nu} + \int |f - g| \, d\mu + \int |f - g| \, d\nu \le \varepsilon(\mu(\mathbb{R}^d) + \nu(\mathbb{R}^d)) + 2\varepsilon.$$

Mit $\varepsilon \searrow 0$ folgt damit die Behauptung. □

Korollar 3.25. *$\mu \in \mathcal{M}_f([0, \infty))$ ist durch seine Laplace-Transformierte*

$$L_\mu(\lambda) = \int e^{-\lambda} \mu(dx), \quad \lambda \ge 0$$

eindeutig festgelegt.

Beweis. Betrachte $E = [0, \infty]$. E ist kompakt. Für $\lambda \ge 0$ sei

$$f_\lambda \colon E \to [0, 1], \quad x \mapsto e^{-\lambda x}, x < \infty, \quad \infty \mapsto \begin{cases} 0, & \lambda > 0 \\ 1, & \lambda = 0 \end{cases}$$

$\mathcal{C} = \{\sum_{i=1}^n a_i f_{\lambda_i} \mid n \in \mathbb{N},\ a_i \in \mathbb{R},\ \lambda_i \ge 0\}$ ist eine Punkte trennende Algebra, also folgt die Behauptung mit Satz 3.23. □

Korollar 3.26. *Sei X eine Zufallsvariable mit Werten in $[a, b]$. Dann ist ihre Verteilung durch ihre Momente $m_n := \mathbb{E}[X^n], n = 0, 1, \ldots$ festgelegt, denn die Polynome liegen dicht in $C_b([a, b])$.*

Bemerkung. Die Aussage des Korollars gilt nicht, wenn der Wertebereich von X unbeschränkt ist. Betrachte dazu folgendes Gegenbeispiel: Sei $X \sim \mathcal{N}(0, 1)$ und setze $Y := e^X$. Y ist log-normalverteilt mit Dichte $f_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{y} e^{-\frac{1}{2}(\log y)^2}, \ y > 0$. Es gilt:

$$m_n = \mathbb{E}[Y^n] = \mathbb{E}[e^{nX}] = \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2 + nx - \frac{1}{2}n^2} \, dx}_{=1} \cdot e^{\frac{1}{2}n^2} = e^{\frac{1}{2}n^2}.$$

Sei $b > 0$. Y_b habe die Verteilung $\mathbb{P}(Y_b = be^{bk}) = \frac{1}{C_b} b^{-k} e^{-\frac{1}{2}k^2}, \ k \in \mathbb{Z}$ mit $C_b := \sum_{k \in \mathbb{Z}} b^{-k} e^{-\frac{1}{2}k^2}$. Dann gilt für die Momente von Y_b:

$$e^{-\frac{1}{2}n^2} \mathbb{E}[Y_b^n] = e^{-\frac{1}{2}n^2} \sum_{k \in \mathbb{Z}} \left(be^k\right)^n \frac{b^{-k} e^{-\frac{1}{2}k^2}}{C_b} = \sum_{k \in \mathbb{Z}} \frac{b^{-(k-n)} e^{-\frac{1}{2}(k-n)^2}}{C_b} = \frac{C_b}{C_b} = 1.$$

Es gilt also $\mathbb{E}[Y^n] = \mathbb{E}[Y_b^n] = e^{\frac{1}{2}n^2}$, aber Y und Y_b haben offensichtlich nicht dieselbe Verteilung.

Satz 3.27 (Lévys[3] Stetigkeitssatz)**.** *Seien $P, P_1, P_2, \ldots \in \mathcal{M}_1(\mathbb{R}^d)$ mit charakteristischen Funktionen $\varphi, \varphi_1, \varphi_2, \ldots$.*

i) Ist $P = w - \lim P_n$, so gilt $\varphi_n \to \varphi$ lokal gleichmäßig.

ii) Es gelte $\varphi_n \to f$ punktweise gegen ein in 0 stetiges $f \colon \mathbb{R}^d \to \mathbb{C}$, dann gibt es ein $Q \in \mathcal{M}_1(\mathbb{R}^d)$ mit $f = \varphi_Q$ und $Q = w - \lim P_n$.

Lemma 3.28. *Sei $\mathcal{F} \subset \mathcal{M}_1(\mathbb{R}^d)$ straff. Dann ist $\{\varphi_\mu \mid \mu \in \mathcal{F}\}$ gleichgradig gleichmäßig stetig, das heißt für alle $\varepsilon > 0$ existiert ein $\delta > 0$, sodass für alle $t, t' \in \mathbb{R}^d$ mit $|t - t'| < \delta$ gilt*

$$\sup_{\mu \in \mathcal{F}} |\varphi_\mu(t) - \varphi_\mu(t')| < \varepsilon.$$

Beweis. Es gilt mit Hilfe der Cauchy-Schwarz-Ungleichung:

$$\begin{aligned}
|\varphi_\mu(t) - \varphi_\mu(s)|^2 &\leq \left| \int_{\mathbb{R}^d} \left(e^{i\langle t-s, x\rangle} - 1\right) \left(e^{i\langle s, x\rangle}\right) \mu(dx) \right|^2 \\
&\leq \int_{\mathbb{R}^d} \left|e^{i\langle t-s, x\rangle} - 1\right|^2 \mu(\,dx) \cdot \int_{\mathbb{R}^d} \underbrace{\left|e^{i\langle s, x\rangle}\right|^2}_{=1} \mu(dx) \\
&= \int_{\mathbb{R}^d} \left(e^{i\langle t-s, x\rangle} - 1\right) \left(e^{-i\langle t-s, x\rangle} - 1\right) \mu(dx) \\
&= \int_{\mathbb{R}^d} 2\left(1 - \mathrm{Re}\left(e^{i\langle t-s, x\rangle}\right)\right) \mu(dx) \\
&= 2\big(1 - \mathrm{Re}(\varphi_\mu(t-s))\big).
\end{aligned}$$

[3] nach Paul Lévy, 1886–1971

Wähle K so groß, dass $\sup_{\mu\in\mathcal{F}} \mu\left(\mathbb{R}^d \setminus [-K,K]^d\right) < \frac{\varepsilon^2}{6}$ und wähle δ so klein, dass

$$\text{für alle } |u| < \delta \text{ gilt } \sup_{x\in[-K,K]^d} \left|1 - e^{i\langle u,x\rangle}\right| < \frac{\varepsilon^2}{6}.$$

Dann gilt für alle $\mu \in \mathcal{F}, t, t' \in \mathbb{R}^d$ mit $|t - t'| < \delta$:

$$|\varphi_\mu(t) - \varphi_\mu(t')|^2 \le 2\left(1 - \mathrm{Re}\left(\varphi_\mu(t-t')\right)\right) \le 2\int \left|1 - e^{i\langle t-t',x\rangle}\right| \mu(\,dx) \le 2\left(\frac{\varepsilon^2}{6} + 2\frac{\varepsilon^2}{6}\right) = \varepsilon^2.$$

$\square$

Beweis von Satz 3.27. i) Die punktweise Konvergenz $\varphi_n \to \varphi$ ist klar, lokal gleichmäßige Konvergenz folgt damit aus Lemma 3.28:

Sei $K \subset \mathbb{R}^d$ kompakt, $\varepsilon > 0$, wähle $\delta > 0$ so klein, dass

$$|t - s| < \delta \implies \sup_{n\in\mathbb{N}} |\varphi_n(t) - \varphi_n(s)| \vee |\varphi(t) - \varphi(s)| < \frac{\varepsilon}{3}.$$

Wähle $t_1, \dots, t_m \in K$ mit $K \subset \bigcup_{i=1}^m B_\delta(t_i)$, n_0 so groß, dass

$$\max_{i=1,\dots,m} |\varphi_n(t_i) - \varphi(t_i)| < \frac{\varepsilon}{3} \quad \text{für } n \ge n_0.$$

Für $t \in K, n \ge n_0$ wähle $i \le m$ mit $|t - t_i| < \delta$, so ist

$$|\varphi_n(t) - \varphi(t)| \le |\varphi_n(t) - \varphi_n(t_i)| + |\varphi_n(t_i) - \varphi(t_i)| + |\varphi(t_i) - \varphi(t)| < 3\frac{\varepsilon}{3} = \varepsilon.$$

ii) Zeige

$$\{P_n \mid n \in \mathbb{N}\} \text{ ist straff.} \tag{3.7}$$

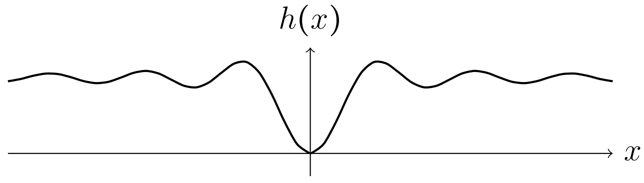
Dazu genügt es zu zeigen, dass $P_n^{(k)} := P_n \circ \pi_k^{-1}$, wobei $\pi_k\colon \mathbb{R}^d \to \mathbb{R}$ die k-te Koordinatenprojektion ist, für jedes $k = 1, \dots, d$ eine straffe Familie sind, denn es gilt

$$P_n\left(\mathbb{R}^d \setminus [-K,K]^d\right) \le \sum_{k=1}^d P_n^{(k)}([-K,K]^c).$$

Für $s \in \mathbb{R}$ ist $\varphi_{P_n^{(k)}}(s) = \varphi_n(s \cdot e_k)$, wobei e_k der k-te Einheitsvektor im $\mathbb{R}^d$ ist. Nach Voraussetzung gilt $\varphi_{P_n^{(k)}}(s) \to f(s \cdot e_k)$ und $s \mapsto f(s \cdot e_k)$ ist stetig in 0 mit $f(0 \cdot e_k) = 1$.

Setze

$$h(x) = \begin{cases} 1 - \dfrac{\sin x}{x}, & x \ne 0 \\ 0, & x = 0 \end{cases}$$



h ist stetig differenzierbar mit $a := \inf\{h(x) \mid |x| \geq 1\} = 1 - \sin(1) > 0$. Für $M > 0$ ist

$$\begin{aligned}
P_n^{(k)}([-M,M]^c) &\leq \frac{1}{a}\int_{[-M,M]^c} h\left(\frac{x}{M}\right) P_n^{(k)}(\,dx) \\
&= \frac{1}{a}\int_{[-M,M]^c}\int_0^1 1-\cos\left(t\frac{x}{M}\right)\,dt\,P_n^{(k)}(\,dx) \\
&= \frac{1}{a}\int_0^1\int_{[-M,M]^c} 1-\cos\left(t\frac{x}{M}\right) P_n^{(k)}(\,dx)\,dt \\
&\leq \frac{1}{a}\int_0^1 1-\mathrm{Re}\left(\varphi_n\left(\frac{t}{M}e_k\right)\right)\,dt.
\end{aligned}$$

(wir verwenden den Satz von Fubini für die zweite Gleichheit). Dann ist

$$\begin{aligned}
\limsup_{n\to\infty} P_n^{(k)}([-M,M]^c) &\leq \frac{1}{a}\limsup_{n\to\infty}\int_0^1 1-\mathrm{Re}\left(\varphi_n\left(\frac{t}{M}e_k\right)\right)\,dt \\
&= \frac{1}{a}\int_0^1 \limsup_{n\to\infty} 1-\mathrm{Re}\left(\varphi_n\left(\frac{t}{M}e_k\right)\right)\,dt \\
&= \frac{1}{a}\int_0^1 1-\mathrm{Re}\left(f\left(\frac{t}{M}e_k\right)\right)\,dt \\
&< \varepsilon
\end{aligned}$$

wenn M so groß gewählt wird, dass $\inf_{|t|\leq 1}\mathrm{Re}\left(f\left(\frac{t}{M}e_k\right)\right) \geq 1 - \frac{\varepsilon}{a}$ gilt, d.h. (3.7) ist erfüllt.

Nach Satz 3.15 gibt es eine Teilfolge $(P_{n_j})_j$ mit $P_{n_j} \xrightarrow[j\to\infty]{w} Q$ und mit $i)$ gilt $f = \varphi_Q$. Somit folgt $Q = Q'$ für jede konvergente Teilfolge $(P_{n_k})_k$ mit $P_{n_k} \xrightarrow{w} Q'$. □

Beobachtung 3.29. *i)* Sei X eine reelle Zufallsvariable mit $\mathbb{E}[|X|^n] < \infty$. Dann ist φ_X n-mal stetig differenzierbar mit

$$\frac{d^n}{dt^n}\varphi_X(t) = \mathbb{E}\left[(iX)^n e^{itX}\right],$$

denn Restgliedabschätzung der Taylorentwicklung von e^{ix} liefert

$$\left|e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!}\right| \leq \frac{|x|^{n+1}}{(n+1)!} \wedge \frac{2|x|^n}{n!}$$

und somit

$$\mathbb{E}\left[\left|e^{i(t+h)X} - e^{itX}\sum_{k=0}^n\frac{(ih)^kX^k}{k!}\right|\right] \leq \mathbb{E}\left[\frac{|h|^{n+1}|X|^{n+1}}{(n+1)!} \wedge \frac{2|h|^n|X|^n}{n!}\right] \quad \text{für } t,h\in\mathbb{R};$$

$h \mapsto \mathbb{E}\left[e^{itX}\sum_{k=0}^n \frac{(ih)^kX^k}{k!}\right] = \sum_{k=0}^n \frac{h^k}{k!}\mathbb{E}\left[(iX)^k e^{itX}\right]$ ist (offenbar) differenzierbar in h.

ii) Falls $\mathbb{E}[|X|^n] < \infty$ für alle $n \in \mathbb{N}$ und

$$\lim_{n\to\infty}\frac{|h|^n\mathbb{E}[|X|^n]}{n!} = 0 \tag{3.8}$$

für ein $h \in \mathbb{R} \setminus \{0\}$ gilt, so ist

$$\varphi_X(t+h) = \sum_{k=0}^{\infty} \frac{(ih)^k}{k!} \mathbb{E}\left[X^k e^{itX}\right]$$

(denn dann gilt $\left|\varphi_X(t+h) - \sum_{k=0}^{n} \frac{(ih)^k}{k!} \mathbb{E}\left[X^k e^{itX}\right]\right| \le \frac{2}{n!} |h|^n \mathbb{E}[|X|^n] \longrightarrow 0$ für $n \to \infty$); insbesondere ist φ_X analytisch.

Korollar 3.30 (Momentenproblem)**.** *X reelle Zufallsvariable mit*

$$a := \limsup_{n\to\infty} \frac{1}{n} \Big(\mathbb{E}[|X|^n]\Big)^{1/n} < \infty.$$

dann ist φ_X analytisch und $\mathcal{L}(X)$ durch die Angabe aller Momente $\mathbb{E}[X^n]$, $n \in \mathbb{N}$ eindeutig festgelegt.

Beweis. Mit Stirling-Approximation ($n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$) ergibt sich für $|h| < 1/(3a)$

$$\limsup_{n\to\infty} \frac{|h|^n \mathbb{E}[|X|^n]}{n!} = \limsup_{n\to\infty} \frac{1}{\sqrt{2\pi n}} \Big(\big(\mathbb{E}[|X|^n]\big)^{1/n} |h| \frac{e}{n}\Big)^n \le \limsup_{n\to\infty} \frac{1}{\sqrt{2\pi n}} \big(e/3\big)^n = 0,$$

d.h. (3.8) ist erfüllt. □

Wir haben in Korollar 3.24 ein „abstraktes" Argument verwendet, um zu zeigen, dass die charakteristische Funktion ein Maß festlegt. In vielen Fällen (speziell in $\mathbb{R}^1$ oder in $\mathbb{Z}^d$ oder falls μ eine Dichte besitzt), erhält man die Eindeutigkeit auch (relativ leicht) durch „explizite" Fourier-Inversion:

Bericht 3.31. Sei $\mu \in \mathcal{M}_1(\mathbb{R})$, so gilt

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\mu(t)\, dt \underset{T\to\infty}{\longrightarrow} \mu((a,b)) + \frac{1}{2}\mu(\{a,b\})$$

für $-\infty < a < b < \infty$ (z.B. [Du, Thm. 3.3.11] oder [St1, Satz 8.4]);
wenn $\mu = f\lambda$ mit $f \in \mathcal{L}^1(\lambda)$, $f \ge 0$ auf $\mathbb{R}^d$ mit $\varphi_\mu \in \mathcal{L}^1(\lambda)$, so gilt

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t,x\rangle} \varphi_\mu(t)\, dt \quad (\lambda\text{-f.ü.})$$

wenn $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ auf $\mathbb{Z}^d$ konzentriert ist, gilt

$$\mu(\{x\}) = \frac{1}{(2\pi)^d} \int_{(-\pi,\pi]^d} e^{-i\langle t,x\rangle} \varphi_\mu(t)\, dt \quad \text{für } x \in \mathbb{Z}^d.$$

Kapitel 4

Zentrale Grenzwertsätze

Erinnerung 4.1 (Zentraler Grenzwertsatz). Seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte, reelle Zufallsvariablen mit $\mu := \mathbb{E}[X_1]$ und $\sigma^2 := \operatorname{Var}[X_1] < \infty$. Sei

$$S_n^* := \frac{1}{\sqrt{n\sigma^2}} \sum_{j=1}^{n} (X_j - \mu).$$

Dann gilt $S_n^* \Rightarrow \mathcal{N}(0,1),\ n \to \infty$, denn es gilt (verwende Taylorentwicklung von φ_{X_1})

$$\begin{aligned}\varphi_{S_n^*}(t) &= \left(\varphi_{\frac{X_1-\mu}{\sqrt{n\sigma^2}}}(t)\right)^n = \left(\varphi_{X_1-\mu}\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right)^n = \left(1 + 0\frac{t}{\sqrt{n\sigma^2}} - \frac{1}{2}\sigma^2\frac{t^2}{n\sigma^2} + o\left(\frac{t^2}{n\sigma^2}\right)\right)^n \\ &\sim \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}t^2} = \varphi_{\mathcal{N}(0,1)}(t) \quad \text{für jedes } t \in \mathbb{R}.\end{aligned}$$

Die aus der Einführung in die Stochastik bekannte Version des Zentralen Grenzwertsatzes

$$\mathbb{P}\left(a \leq S_n^* \leq b\right) \xrightarrow[n\to\infty]{} \int_a^b \frac{1}{2\pi} e^{-\frac{x^2}{2}}\, dx \quad \text{für jedes } -\infty \leq a < b \leq \infty$$

folgt aus $\mathcal{L}(S_n^*) \Rightarrow \mathcal{N}(0,1)$ zusammen mit Satz 3.8 vi).

Satz 4.2 (Zentraler Grenzwertsatz für Dreiecksschemata). *Für $n \in \mathbb{N}$ sei $k_n \in \mathbb{N}$ und $X_{n,1}$, $X_{n,2} \ldots, X_{n,k_n} \in \mathcal{L}^2(\mathbb{P})$. $(X_{n,\ell} \mid \ell = 1, 2, \ldots, k_n,\ n \in \mathbb{N})$ heißt ein Dreiecksschema von Zufallsvariablen. Für $n \in \mathbb{N}$ seien $X_{n,1}, X_{n,2} \ldots, X_{n,k_n}$ unabhängig mit $\mathbb{E}[X_{n,\ell}] = 0$ und $\sum_{\ell=1}^{k_n} \operatorname{Var}[X_{n,\ell}] = 1$ (das Schema ist „zentriert" und „normiert"). Setze*

$$S_n := \sum_{\ell=1}^{k_n} X_{n,\ell}.$$

Es gelte die Lindeberg[1]-Bedingung

$$L_n(\varepsilon) := \sum_{\ell=1}^{k_n} \mathbb{E}\left[X_{n,\ell}^2 \cdot \mathbf{1}_{\left\{X_{n,\ell}^2 > \varepsilon^2\right\}}\right] = \frac{1}{\operatorname{Var}[S_n]} \sum_{l=1}^{k_n} \mathbb{E}\left[X_{n,l}^2 \cdot \mathbf{1}_{\left\{X_{n,l}^2 > \varepsilon^2 \operatorname{Var}[S_n]\right\}}\right] \xrightarrow[n\to\infty]{} 0 \qquad (4.1)$$

für alle $\varepsilon > 0$. Dann gilt $S_n \Rightarrow \mathcal{N}(0,1),\ n \to \infty$.

[1] nach Jarl Waldemar Lindeberg, 1876–1932

Bemerkung 4.3. i) Falls $(X_{n,\ell})$ die *Lyapunov*[2]*-Bedingung*

$$\frac{1}{(\operatorname{Var}[S_n])^{1+\frac{\delta}{2}}}\sum_{\ell=1}^{k_n}\mathbb{E}\left[|X_{n,\ell}|^{2+\delta}\right]\xrightarrow[n\to\infty]{}0 \tag{4.2}$$

für ein $\delta>0$ erfüllt, so gilt auch die Lindeberg-Bedingung (4.1), denn

$$\mathbb{E}\left[X_{n,\ell}^2\cdot\mathbf{1}_{\left\{X_{n,\ell}^2>\varepsilon^2\right\}}\right]\le\varepsilon^{-\delta}\mathbb{E}\left[|X_{n,\ell}|^{2+\delta}\right].$$

ii) Die Lindeberg-Bedingung (4.1) impliziert

$$\lim_{n\to\infty}\max_{1\le\ell\le k_n}\mathbb{P}\left(|X_{n,\ell}|>\varepsilon\right)=0, \tag{4.3}$$

denn

$$\max_{1\le\ell\le k_n}\mathbb{P}\left(|X_{n,\ell}|>\varepsilon\right)\le\sum_{\ell=1}^{k_n}\mathbb{P}\left(|X_{n,\ell}|>\varepsilon\right)\le\frac{1}{\varepsilon^2}\sum_{\ell=1}^{k_n}\mathbb{E}\left[X_{n,\ell}^2\cdot\mathbf{1}_{\left\{X_{n,\ell}^2>\varepsilon^2\right\}}\right]\xrightarrow[n\to\infty]{}0.$$

Man sagt auch: Das Schema ist „asymptotisch vernachlässigbar“. Es gilt auch die Umkehrung: Gilt (4.3) und $S_n\Rightarrow\mathcal{N}(0,1)$, so gilt auch (4.1) (vgl. [Ka, Theorem 5.12]).

Beweis von Satz 4.2. Bemerke zunächst: Sind $z_1,\dots,z_n,z_1',\dots,z_n'\in\{z\in\mathbb{C}\mid|z|\le1\}$, so gilt

$$\left|\prod_{j=1}^n z_j-\prod_{j=1}^n z_j'\right|\le\sum_{j=1}^n|z_j-z_j'|,$$

denn für $n=2$ gilt

$$|z_1z_2-z_1'z_2'|\le|z_1z_2-z_1'z_2|+|z_1'z_2-z_1'z_2'|=|z_2||z_1-z_1'|+|z_1'||z_2-z_2'|\le|z_1-z_1'|+|z_2-z_2'|$$

und der allgemeine Fall folgt induktiv.

Sei $c_{n,\ell}:=\mathbb{E}[X_{n,\ell}^2]=\operatorname{Var}[X_{n,\ell}]$ und $\varphi_{n,\ell}(t):=\mathbb{E}\left[e^{itX_{n,\ell}}\right]$. Es gilt

$$\limsup_{n\to\infty}\max_{1\le\ell\le k_n}c_{n,\ell}=0,$$

denn es gilt wegen (4.1) für alle $\varepsilon>0$

$$\max_{1\le\ell\le k_n}c_{n,\ell}\le\varepsilon^2+\sum_{\ell=1}^{k_n}\mathbb{E}\left[X_{n,\ell}^2\cdot\mathbf{1}_{\left\{X_{n,\ell}^2>\varepsilon^2\right\}}\right]\xrightarrow[n\to\infty]{}\varepsilon^2.$$

Damit folgt

$$\begin{aligned}\left|\mathbb{E}\left[e^{itS_n}\right]-e^{-\frac12t^2}\right|&=\left|\prod_{\ell=1}^{k_n}\varphi_{n,\ell}(t)-\prod_{\ell=1}^{k_n}e^{-\frac12c_{n,\ell}t^2}\right|\le\sum_{\ell=1}^{k_n}\left|\varphi_{n,\ell}(t)-e^{-\frac12c_{n,\ell}t^2}\right|\\&\le\sum_{\ell=1}^{k_n}\left|\varphi_{n,\ell}(t)-1+\frac12c_{n,\ell}t^2\right|+\sum_{\ell=1}^{k_n}\left|e^{-\frac12c_{n,\ell}t^2}-1+\frac12c_{n,\ell}t^2\right|.\end{aligned}$$

[2] nach Alexandr Mihailovich Lyapunov, 1857–1918

Mit Beobachung 3.29 lässt sich das Argument der ersten Summe durch das Restglied der Taylorentwicklung abschätzen:

$$\left|\varphi_{n,\ell}(t) - 1 + \frac{1}{2}c_{n,\ell}t^2\right| \le \mathbb{E}\left[|tX_{n,\ell}|^2 \wedge \frac{|tX_{n,\ell}|^3}{3}\right] \le \varepsilon|t|^3\mathbb{E}[X_{n,\ell}^2] + t^2\mathbb{E}\left[X_{n,\ell}^2 \cdot \mathbf{1}_{\left\{X_{n,\ell}^2 > \varepsilon^2\right\}}\right].$$

Weiterhin gilt

$$\left|e^{-\frac{1}{2}c_{n,\ell}t^2} - 1 + \frac{1}{2}c_{n,\ell}t^2\right| \le \frac{1}{2!}\left(\frac{1}{2}c_{n,\ell}t^2\right)^2 = \frac{1}{8}c_{n,\ell}^2 t^4.$$

Zusammen ergibt sich

$$\begin{aligned}\left|\mathbb{E}\left[e^{itS_n}\right] - e^{-\frac{1}{2}t^2}\right| &\le \sum_{\ell=1}^{k_n}\left|\varphi_{n,\ell}(t) - 1 + \frac{1}{2}c_{n,\ell}t^2\right| + \sum_{\ell=1}^{k_n}\left|e^{-\frac{1}{2}c_{n,\ell}t^2} - 1 + \frac{1}{2}c_{n,\ell}t^2\right| \\ &\le \varepsilon|t|^3\sum_{\ell=1}^{k_n} c_{n,\ell} + t^2 L_n(\varepsilon) + \frac{t^4}{8}\Big(\max_{1\le\ell'\le k_n} c_{n,\ell'}\Big)\sum_{\ell=1}^{k_n} c_{n,\ell} \\ &= \varepsilon|t|^3 + t^2 L_n(\varepsilon) + \frac{t^4}{8}\Big(\max_{1\le\ell'\le k_n} c_{n,\ell'}\Big) \xrightarrow[n\to\infty]{} \varepsilon|t|^3 + 0 + 0 = \varepsilon|t|^3.\end{aligned}$$

Also gilt

$$\limsup_{n\to\infty}\left|\mathbb{E}\left[e^{itS_n}\right] - e^{-\frac{1}{2}t^2}\right| \le \varepsilon.$$

Mit $\varepsilon \searrow 0$ und Levys Stetigkeitssatz (Satz 3.27) folgt die Behauptung. □

4.1 Der mehrdimensionale Fall

Beobachtung 4.4. i) Sei $X = (X_1, \dots, X_d)^{\mathrm{t}}$ eine $\mathbb{R}^d$-wertige Zufallsvariable. Die Kovarianzmatrix $C = (c_{ij})_{i,j=1,\dots,d}$ mit $c_{ij} = \mathrm{Cov}[X_i, X_j]$ ist symmetrisch und positiv semi-definit, denn $c_{ij} = \mathrm{Cov}[X_i, X_j] = \mathrm{Cov}[X_j, X_i] = c_{ji}$ und für $a = (a_1, \dots, a_d)^{\mathrm{t}} \in \mathbb{R}^d$ ist

$$a^{\mathrm{t}}Ca = \sum_{i,j=1}^{d} a_i c_{ij} a_j = \sum_{i,j=1}^{d} a_i a_j \mathrm{Cov}[X_i, X_j] = \mathrm{Cov}\left[\sum_{i=1}^{d} a_i X_i, \sum_{j=1}^{d} a_j X_j\right] = \mathrm{Var}[\langle a, X\rangle] \ge 0.$$

ii) Sind $Z_1, \dots, Z_d$ unabhängig und standardnormalverteilt, so hat $Z = (Z_1, \dots, Z_d)^{\mathrm{t}}$ die Dichte

$$f_Z(z) = (2\pi)^{-\frac{d}{2}}\exp\left(-\frac{1}{2}\left(z_1^2 + \dots + z_d^2\right)\right) = (2\pi)^{-\frac{d}{2}}e^{-\frac{1}{2}\|z\|^2}, \quad z \in \mathbb{R}^d.$$

$\mathcal{L}(Z)$ heißt die d-dimensionale Standardnormalverteilung.

iii) Sei $\mu \in \mathbb{R}^d$ und $A = (a_{ij}) \in \mathbb{R}^{d\times d}$. Dann hat $X := \mu + AZ$ den Erwartungswert(-vektor) $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]) = \mu$ und die Kovarianzmatrix $C := AA^{\mathrm{t}}$, denn

$$\begin{aligned}\mathrm{Cov}[X_k, X_\ell] = \mathrm{Cov}\left[\mu_k + \sum_{i=1}^{d} a_{ki}Z_i,\ \mu_l + \sum_{j=1}^{d} a_{\ell j}Z_j\right] &= \sum_{i,j=1}^{d} a_{ki}a_{\ell j}\mathrm{Cov}[Z_i, Z_j] \\ &= \sum_{i,j=1}^{d} a_{ki}a_{\ell j}\delta_{ij} = \sum_{i=1}^{d} a_{ki}a_{\ell i} = (AA^{\mathrm{t}})_{k\ell}.\end{aligned}$$

Falls A vollen Rang d hat, so hat X die Dichte

$$f_{\mu,C}(x) = \frac{1}{\sqrt{(2\pi)^d \det C}} \exp\left(-\frac{1}{2}\langle x-\mu, C^{-1}(x-\mu)\rangle\right), \quad x \in \mathbb{R}^d,$$

denn für $g(z) := \mu + Az$ gilt $g^{-1}(x) = A^{-1}(x-\mu)$ und $\left(\frac{\partial}{\partial x_i} g_j(z)\right)_{i,j} = \mathrm{D}g(z) = A$. Also folgt mit der Dichtetransformationsformel und mit $\det C = \det(AA^{\mathrm{t}}) = (\det A)^2$:

$$f_{\mu,C}(x) = f_Z\left(g^{-1}(x)\right) \frac{1}{|\det \mathrm{D}g(g^{-1}(x))|}.$$

Falls A nicht vollen Rang hat, so besitzt X keine Dichte bezüglich λ^d.

Was ist jedoch in dem Fall, in dem A (und damit auch C) nicht vollen Rang haben? Betrachte für $u \in \mathbb{R}^d$:

$$\begin{aligned}
\mathbb{E}\left[e^{i\langle u,X\rangle}\right] &= \mathbb{E}\left[e^{i\langle u,\mu\rangle} \cdot e^{i\langle u,AZ\rangle}\right] = e^{i\langle u,\mu\rangle}\mathbb{E}\left[e^{i\sum_{k,\ell=1}^d u_k a_{k\ell} Z_\ell}\right] = e^{i\langle u,\mu\rangle} \prod_{\ell=1}^d \mathbb{E}\left[e^{i\sum_{k=1}^d u_k a_{k\ell} Z_\ell}\right] \\
&= e^{i\langle u,\mu\rangle} \prod_{\ell=1}^d e^{-\frac{1}{2}\left(\sum_{k=1}^d u_k a_{k\ell}\right)^2} = e^{i\langle u,\mu\rangle} e^{-\frac{1}{2}\sum_{\ell=1}^d ((u^{\mathrm{t}}A)_\ell)^2} = e^{i\langle u,\mu\rangle} e^{-\frac{1}{2}\langle u^{\mathrm{t}}A, u^{\mathrm{t}}A\rangle} \\
&= e^{i\langle u,\mu\rangle} e^{-\frac{1}{2}\langle u^{\mathrm{t}}, u^{\mathrm{t}}AA^{\mathrm{t}}\rangle} = e^{i\langle u,\mu\rangle} e^{-\frac{1}{2}\langle u, Cu\rangle}.
\end{aligned}$$

Dies legt folgende Definition nahe:

Definition 4.5. Sei $\mu \in \mathbb{R}^d$, $C \in \mathbb{R}^{d\times d}$ symmetrisch und positiv definit. X heißt *d-dimensional normalverteilt mit Erwartungswert μ und Kovarianzmatrix C*, falls

$$\varphi_X(u) = e^{i\langle u,\mu\rangle} e^{-\frac{1}{2}\langle u,Cu\rangle}.$$

Man schreibt auch $\mathcal{L}(X) =: \mathcal{N}(\mu, C)$.

Bemerkung 4.6. Sei $X \sim \mathcal{N}(\mu, C)$, $A \in \mathbb{R}^{d\times d}$ und $Y := AX$. Dann ist $Y \sim \mathcal{N}(A\mu, ACA^{\mathrm{t}})$, denn

$$\mathbb{E}\left[e^{i\langle u,Y\rangle}\right] = \mathbb{E}\left[e^{i\langle u,AX\rangle}\right] = \mathbb{E}\left[e^{i\langle A^{\mathrm{t}}u,X\rangle}\right] = e^{i\langle A^{\mathrm{t}}u,\mu\rangle} e^{-\frac{1}{2}\langle A^{\mathrm{t}}u, CA^{\mathrm{t}}u\rangle} = e^{i\langle u,A\mu\rangle} e^{-\frac{1}{2}\langle u,ACA^{\mathrm{t}}u\rangle}.$$

Lemma 4.7 (Cramér-Wold device)**.** *Für $n \in \mathbb{N} \cup \{\infty\}$ seien $X_n = (X_{n,1}, \ldots, X_{n,d})^{\mathrm{t}}$ Zufallsvariablen mit Werten in $\mathbb{R}^d$. Dann sind äquivalent:*

i) $X_n \Rightarrow X_\infty$, $n \to \infty$.

ii) $\mathcal{L}(\langle \lambda, X_n\rangle) \xrightarrow[n\to\infty]{w} \mathcal{L}(\langle \lambda, X_\infty\rangle)$ *für alle* $\lambda \in \mathbb{R}^d$.

Beweis. Es gelte zunächst $X_n \Rightarrow X_\infty$, $n \to \infty$. Betrachte $f(x) = e^{i\langle \lambda, x\rangle}$. Es ist $f \in \mathbb{C}_b(\mathbb{R}^d, \mathbb{C})$, also gilt für alle $t \in \mathbb{R}$, $\lambda \in \mathbb{R}^d$

$$\mathbb{E}\left[e^{it\langle \lambda, X_n\rangle}\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[e^{it\langle \lambda, X_\infty\rangle}\right].$$

Somit folgt $ii)$ aus Satz 3.27 (Lévys Stetigkeitssatz).

Es gelte nun $ii)$. Nach Voraussetzung gilt für alle $\lambda \in \mathbb{R}^d$

$$\mathbb{E}\left[e^{i\langle \lambda, X_n\rangle}\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[e^{i\langle \lambda, X_\infty\rangle}\right]$$

und somit folgt $i)$ mit Satz 3.27. $\square$

Satz 4.8 (Zentraler Grenzwertsatz im $\mathbb{R}^d$)**.** *Seien* $(X_n)_{n\in\mathbb{N}}$ *unabhängige, identisch verteilte, d-dimensionale Zufallsvariablen mit* $\mathbb{E}[X_1] = \mu$ *und* $\mathrm{Cov}[X_1] = C$*. Setze* $S_n^* := \frac{X_1+\dots+X_n-n\mu}{\sqrt{n}}$*. Dann gilt* $S_n^* \Rightarrow \mathcal{N}(0, C)$, $n \to \infty$.

Beweis. Sei $\lambda \in \mathbb{R}^d$, $X_n^\lambda := \langle \lambda, X_n \rangle$ und $S_n^\lambda := \langle \lambda, S_n^* \rangle$. Betrachte ohne Einschränkung $\mu = 0$. Mit dem Zentralen Grenzwertsatz in $\mathbb{R}$ gilt $S_n^\lambda \Rightarrow \mathcal{N}(0, \mathrm{Var}[X_1^\lambda]) = \mathcal{N}(0, \langle \lambda, C\lambda \rangle)$, also gilt

$$\mathbb{E}\left[e^{i\langle \lambda, S_n^* \rangle}\right] \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}\langle \lambda, C\lambda \rangle}.$$

Somit folgt die Behauptung mit Lemma 4.7. $\square$

Anhang A

Zur bedingten Erwartung

Dieses Kapitel dient als Referenz und zur Erinnerung an den Begriff der bedingten Erwartung (einer Zufallsvariable gegeben eine σ-Algebra). Es werden wichtige Definitionen und Aussagen und Aussagen zusammengestellt, für die Beweise siehe die Literatur, z.B. [St1, Kap. 6], [Kl, Kap. 8], [Wi, Kap. 9] oder [Ka, Anfang von Kap. 8].

Erinnerung A.1. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ W'raum, $B \in \mathscr{F}$ mit $\mathbb{P}(B) > 0$.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad A \in \mathscr{F}$$

heißt die *bedingte Wahrscheinlichkeit von A, gegeben* B.

$\mathbb{P}(\,\cdot \mid B)$ definiert durch diese Formel ein W'Maß auf $(\Omega, \mathscr{F})$ mit $\mathbb{P}(B \mid B) = 1$, für $Y \in \mathcal{L}^1(\mathbb{P})$ ist

$$\mathbb{E}[Y \mid B] = \int Y(\omega)\,\mathbb{P}(d\omega \mid B) = \frac{\mathbb{E}[\mathbf{1}_B Y]}{\mathbb{P}(B)}.$$

Bemerkung A.2 (Diskreter Fall der bedingten Erwartung)**.** X und Y ZVn (auf einem W'raum $(\Omega, \mathscr{F}, \mathbb{P})$), $Y \in \mathcal{L}^1(\mathbb{P})$, X nehme höchstens abzählbar viele Werte $x_1, x_2, \dots$ an. Setze

$$f(x) := \mathbb{E}[Y \mid \{X = x\}] \quad \text{für } x \in \{x_1, x_2, \dots\}$$

und $\mathbb{E}[Y \mid X] := f(X)$.

Offenbar ist $\mathbb{E}[Y \mid X]$ $\sigma(X)$-messbar (es ist eine Funktion von X) und für $A \in \sigma(X)$ gilt

$$\mathbb{E}[Y \mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[Y \mid X] \mathbf{1}_A\big]$$

(klar für $A = \{X = x_i\}$ und dann auch für $A = \{X \in B\}$ mit $B \subset \{x_1, x_2, \dots\}$; jedes $A \in \sigma(X)$ hat diese Form).

Wir betrachten im folgenden einen W'raum $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition A.3. $X \in \mathcal{L}^1(\mathbb{P})$, $\mathscr{G} \subset \mathscr{F}$ Teil-σ-Algebra. Eine reellwertige ZV Y heißt (eine Version der) bedingte(n) Erwartung von X gegeben $\mathscr{G}$ (geschrieben $\mathbb{E}[X \mid \mathscr{G}]$), wenn gilt

i) Y ist $\mathscr{G}$-messbar,

ii) $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$ für alle $A \in \mathscr{G}$.

Bemerkung. Äquivalent kann ii) durch ii') ersetzt werden:

ii') $\quad \mathbb{E}[Y \cdot H] = \mathbb{E}[X \cdot H] \quad$ für alle reellwertigen, beschr., $\mathscr{G}$-messbaren ZV H.

Falls $\mathscr{G} = \sigma(Z)$ für eine Zufallsvariable Z, so schreibt man oft auch $\mathbb{E}[X \mid Z] := \mathbb{E}[X \mid \sigma(Z)]$.

Satz A.4. *Für $X \in \mathscr{L}^1(\mathbb{P})$ existiert $\mathbb{E}[X \mid \mathscr{G}]$ und ist eindeutig (bis auf $\mathbb{P}$-f.s.-Gleichheit).*

Beweis von Satz A.4 (Eindeutigkeit). Seien $Y, \widetilde{Y}$ bedingte Erwartungen.

$$\mathbb{E}[Y \cdot \mathbf{1}_{\{Y > \widetilde{Y}\}}] = \mathbb{E}[X \cdot 1_{\{Y > \widetilde{Y}\}}] = \mathbb{E}[\widetilde{Y} \cdot \mathbf{1}_{\{Y > \widetilde{Y}\}}],$$

also $\mathbb{E}\big[(Y - \widetilde{Y}) \cdot \mathbf{1}_{\{Y > \widetilde{Y}\}}\big] = 0 \Rightarrow \mathbb{P}(Y > \widetilde{Y}) = 0$; analog gilt $\mathbb{P}(\widetilde{Y} > Y) = 0$, also $Y = \widetilde{Y}$ $\mathbb{P}$-f.s. $\square$

Für die Existenz verwenden wir den folgenden Satz über die Existenz und Eindeutigkeit der $\mathcal{L}^2$-Projektion.

Satz A.5. *$\mathscr{H} \subset \mathcal{L}^2(\mathbb{P})$ abgeschlossener Unterraum. Zu $X \in \mathcal{L}^2(\mathbb{P})$ gibt es (bis auf $\mathbb{P}$-f.s. Gleichheit) genau ein $Y \in \mathscr{H}$ mit*

i) $\|Y - X\|_2 = \inf\big\{\|W - X\|_2 : W \in \mathscr{H}\big\}$

ii) $\mathbb{E}\big[(X - Y)Z\big] = 0$ *für alle* $Z \in \mathscr{H}$

Y ist (die Äquivalenzklasse) der orthogonalen Projektion von X auf $\mathscr{H}$, auch $\mathrm{Proj}_{\mathscr{H}}(X)$ geschrieben.

Beweis von Satz A.4 (Existenz). Sei zunächst $X \in \mathcal{L}^2(\mathbb{P})$,

$$\mathscr{H} := \big\{Y \in \mathcal{L}^2(\mathbb{P}) : Y \text{ ist } \mathscr{G}\text{-messbar}\big\} \subset \mathcal{L}^2(\mathbb{P})$$

ist ein abgeschlossener Unterraum, $Y := \mathrm{Proj}_{\mathscr{H}}(X)$ (nach Satz A.5) leistet das Gewünschte.

Beachte:

$$X \geq 0 \quad \Rightarrow \quad Y := \mathbb{E}[X \mid \mathscr{G}] \geq 0 \quad \text{f.s.},$$

denn dann ist $0 \leq \mathbb{E}[X\mathbf{1}_{\{Y<0\}}] = \mathbb{E}[Y\mathbf{1}_{\{Y<0\}}]$.

Sei nun $X \in \mathcal{L}^1(\mathbb{P})$, $X \geq 0$:

$$Y_n := \mathbb{E}\big[X \wedge n \big| \mathscr{G}\big] \underset{n\to\infty}{\nearrow} Y \ (\geq 0) \quad \text{(f.s.)}$$

($X \wedge n \in \mathcal{L}^2(\mathbb{P})$, nutze dann obige Monotonie-Eigenschaft, um $\mathbb{E}[X \wedge n \mid \mathscr{G}] \leq \mathbb{E}[X \wedge (n+1) \mid \mathscr{G}]$ f.s. zu sehen), für $A \in \mathscr{G}$ gilt

$$\mathbb{E}[X\mathbf{1}_A] = \lim_{n\to\infty} \mathbb{E}\big[(X \wedge n)\mathbf{1}_A\big] = \lim_{n\to\infty} \mathbb{E}[Y_n\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A]$$

(mit monotoner Konvergenz).

Für allgemeines $X \in \mathcal{L}^1(\mathbb{P})$ leistet

$$Y := \mathbb{E}[X^+ \mid \mathscr{G}] - \mathbb{E}[X^- \mid \mathscr{G}]$$

das Gewünschte. $\square$

Satz A.6. *$X, Y \in \mathcal{L}^1(\mathbb{P})$, $\mathscr{G} \subset \mathscr{F}$ Teil-σ-Algebra.*

i) $\mathbb{E}[aX + bY \mid \mathscr{G}] = a\mathbb{E}[X \mid \mathscr{G}] + b\mathbb{E}[Y \mid \mathscr{G}]$ *f.s. für $a, b \in \mathbb{R}$ (Linearität)*

ii) $X \le Y$ *f.s.* $\Rightarrow$ $\mathbb{E}[X \mid \mathscr{G}] \le \mathbb{E}[Y \mid \mathscr{G}]$ *f.s. (Monotonie)*

iii) $\big|\mathbb{E}[X \mid \mathscr{G}]\big| \le \mathbb{E}\big[|X| \,\big|\, \mathscr{G}\big]$ *f.s. (Dreiecksungleichung)*

iv) Es gelte $\mathbb{E}\big[|XY|\big] < \infty$ und Y sei $\mathscr{G}$-messbar, dann ist

$$\mathbb{E}[XY \mid \mathscr{G}] = Y \cdot \mathbb{E}[X \mid \mathscr{G}] \text{ f.s.},$$

insbesondere ist $\mathbb{E}[Y|\mathscr{G}] = Y$ f.s. („Herausziehen von Bekanntem“)

v) Sind $\sigma(X)$ und $\mathscr{G}$ unabhängig, so ist $\mathbb{E}[X \mid \mathscr{G}] = \mathbb{E}[X]$ f.s. (Verhalten bei Unabhängigkeit)

vi) $\mathscr{G}' \subset \mathscr{G}$ Teil-σ-Algebra, so ist

$$\mathbb{E}\big[\mathbb{E}[X \mid \mathscr{G}] \,\big|\, \mathscr{G}'\big] = \mathbb{E}[X \mid \mathscr{G}'] \text{ f.s.},$$

(„Turmeigenschaft“) insbesondere ist

$$\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]] = \mathbb{E}[X]$$

vii) $0 \le X_n \nearrow X$ f.s. für $n \to \infty$, so gilt

$$\mathbb{E}[X_n \mid \mathscr{G}] \underset{n\to\infty}{\nearrow} \mathbb{E}[X \mid \mathscr{G}] \quad \text{f.s. und in } \mathcal{L}^1(\mathbb{P})$$

(monotone Konvergenz)

viii) X_n reelle ZVn mit $|X_n| \le Y$ $\forall\, n$ und $X_n \to X$ f.s., so gilt

$$\mathbb{E}[X_n \mid \mathscr{G}] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[X \mid \mathscr{G}] \quad \text{f.s. und in } \mathcal{L}^1(\mathbb{P})$$

(dominierte Konvergenz)

Satz A.7 (Jensen’sche Ungleichung für die bedingte Erwartung)**.** *$X \in \mathcal{L}^1(\mathbb{P})$, $k : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ konvex, dann gilt*

$$\mathbb{E}\big[k(X) \,\big|\, \mathscr{G}\big] \ge k\big(\mathbb{E}[X \mid \mathscr{G}]\big) \quad \text{f.s.}$$

Bemerkung A.8. X, Y reelle ZVn mit gemeinsamer Dichte $f_{X,Y}$, d.h.

$$\mathbb{P}\big((X, Y) \in A\big) = \int_A f_{X,Y}(x, y)\, \lambda^{\otimes 2}\big(d(x, y)\big) \quad \text{für } A \in \mathcal{B}(\mathbb{R}^2),$$

$Y \in \mathcal{L}^1(\mathbb{P})$.

Sei für $x \in \mathbb{R}$

$$f_X(x) := \int_{\mathbb{R}} f_{X,Y}(x, y)\, \lambda(dy) \quad \text{die Marginaldichte von } X,$$

$$\varphi(x) := \int_{\mathbb{R}} y \frac{f_{X,Y}(x, y)}{f_X(x)}\, \lambda(dy) \mathbf{1}_{f_X(x) > 0}.$$

Dann gilt

$$\mathbb{E}[Y|\sigma(X)] = \varphi(X) \text{ f.s.}$$

denn für $B \in \mathcal{B}(\mathbb{R})$ ist

$$\begin{aligned}\mathbb{E}\big[\mathbf{1}_{\{X\in B\}}\varphi(X)\big] &= \int_{\mathbb{R}} \mathbf{1}_B(x)\varphi(x)f_X(x)\,\lambda(dx)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}_B(x)yf_{X,Y}(x,y)\lambda(dy)\ \lambda(dx)\\ &= \int_{\mathbb{R}^2} \mathbf{1}_B(x)yf_{X,Y}(x,y)\,\lambda^{\otimes 2}\big(d(x,y)\big) = \mathbb{E}\big[\mathbf{1}_{\{X\in B\}}Y\big].\end{aligned}$$

Bericht A.9. (Zu regulären Versionen bedingter Verteilungen) Wenn $Y = 1_B$ für ein Ereignis $B \in \mathscr{A}$, so schreibt man gelegentlich auch $\mathbb{P}(B \mid \mathscr{G}) = \mathbb{E}[Y \mid \mathscr{G}]$. Man muss allerdings etwas vorsichtig sein bei der Interpretation von $\mathbb{P}(\cdot \mid \mathscr{G})$ als ein (zufälliges) Maß, da i.A. überabzählbar viele B in Frage kommen und damit die Kompatibilität der in der Definition der bedingten Erwartung implizit vorkommenden Nullmengen (vgl. Def. A.3) wenigstens a priori unklar bleibt.

In „gutartigen“ Fällen ist eine konsistente Wahl möglich, das Stichwort dazu lautet „reguläre bedingte Verteilung einer Zufallsvariable“. Wir skizzieren hier knapp den reellwertigen Fall:

Sei X reellwertige ZV (auf einem W'raum $(\Omega,\mathscr{F},\mathbb{P})$), $\mathscr{G} \subset \mathscr{F}$ Teil-σ-Algebra. Dann gibt es einen stochastischen Kern $\kappa_{X|\mathscr{G}}$ von $(\Omega,\mathscr{G})$ nach $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ mit

$$\kappa_{X|\mathscr{G}}(\omega,B) = \mathbb{E}\big[\mathbf{1}_{\{X\in B\}}\,\big|\,\mathscr{G}\big](\omega) \text{ f.s. für alle } B \in \mathcal{B}(\mathbb{R}),$$

d.h. (vgl. [St1, Def. 5.4] oder [Kl, Def. 8.24])

für jedes $B \in \mathcal{B}(\mathbb{R})$ ist $\omega \mapsto \kappa_{X|\mathscr{G}}(\omega,B)$ eine Version von $\mathbb{E}\big[\mathbf{1}_{\{X\in B\}}\,\big|\,\mathscr{G}\big]$ und
für jedes ω ist $\kappa_{X|\mathscr{G}}(\omega,\cdot)$ ein W'maß auf $(\mathbb{R},\mathcal{B}(\mathbb{R}))$.

Die Hauptidee besteht darin, das gewünschte Maß $\kappa_{X|\mathscr{G}}(\omega,\cdot)$ auf $\mathbb{R}$ anhand seiner Verteilungsfunktion zu charakterisieren (vgl. [St1, Satz 1.27] oder [Kl, Bsp. 1.44]) die zielführende Beobachtung ist dann, dass eine Verteilungsfunktion (wegen der Monotonie) bereits durch ihre Werte an abzählbar vielen Stellen festgelegt ist. Man betrachtet also $B = (-\infty, r], r \in \mathbb{Q}$ und setzt

$$F_r := \mathbb{E}[\mathbf{1}_{(-\infty,r]}(X) \mid \mathscr{G}].$$

Dann gilt (mit den Eigenschaften der bedingten Erwartung aus Satz A.6) wie gewünscht $\mathbb{P}$-f.s.: $F_r \le F_{r'}$, für $r < r'$, $(r, r' \in \mathbb{Q})$, $\lim\limits_{n\to\infty} F_{r+\frac{1}{n}} = F_r$ für $r \in \mathbb{Q}$, $\lim\limits_{n\to\infty} F_n = 1$, $\lim\limits_{n\to-\infty} F_n = 0$.

Wegen der Abzählbarkeit von $\mathbb{Q}$ gibt es $N \in \mathscr{F}$ mit $\mathbb{P}(N) = 0$, so dass obiges für $\omega \in \Omega \setminus N$ und alle $r, r' \in \mathbb{Q}$ gilt. Dann definiert

$$\widetilde{F}_s := \begin{cases} \inf\{F_r : r \ge s, r \in \mathbb{Q}\}, & \omega \in \Omega \setminus N,\\ \overline{F}_s, & \omega \in N,\end{cases}$$

wobei $\overline{F}$ irgendeine Verteilungsfunktion ist, die (zufällige) Verteilungsfunktion von $k_{X,\mathscr{G}}$. Details finden sich beispielsweise in [Kl, Kap. 8.3, speziell Satz 8.28].

Man kann dieses Argument relativ leicht erweitern auf die Situation, dass der Wertebereich $(E, \mathscr{B})$ von X ein sogenannter Standard-Borel-Raum ist (auch der Name Borel'scher Raum ist üblich), d.h. wenn es ein $A \in \mathcal{B}(\mathbb{R})$ und eine Bijektion $\phi : E \to A$ gibt, so dass ϕ und ϕ^{-1} jeweils messbar sind (dann sind $(E, \mathscr{B})$ und $(A, \mathcal{B}(A))$ isomorph als messbare Räume). Dann ist nämlich $X' := \phi \circ X$ eine reellwertige ZV, und die Argumentation oben greift (vgl. auch [Kl, Satz 8.36]).

Schließlich kann man zeigen, dass jeder separable und vollständige metrische Raum E, versehen mit seiner Borel-σ-Algebra, ein Standard-Borel-Raum ist (siehe z.B. [RW, Bd. 1, Ch. II.82], [Br, Appendix 7]). Solche Wertebereiche heißen *polnische Räume*, sie spielen in der allgemeinen Theorie der Stochastik eine wichtige Rolle (beispielsweise sind $\mathbb{R}^d$ oder $C([0,1])$ mit Supremumsnorm polnisch).

Anhang B

Martingale (in diskreter Zeit)

Martingale[1] sind (u.a.) eine mathematische Formalisierung des Begriffs des fairen Spiels und der Vorstellung, dass man dabei nicht auf systematische Weise gewinnen kann. Wir sammeln hier als Referenz Material aus der Vorlesung Stochastik I, SS 2024, vgl. [St1, Kap. 7]. Siehe auch z.B. [Wi, Teil B], [Kl, Kap. 9–11], [Ka, Kap. 9].

Wir betrachten sozusagen zum „Appetit-Anregen“ folgendes Beispiel:

Beispiel B.1. Betrachte eine faire Münzwurffolge, d.h. seien $W_1, W_2, \dots$ unabhängig und identisch uniform verteilt auf $\{K, Z\}$. Sei

$$R := \min\big\{k \in \mathbb{N} : (W_k, W_{k+1}, W_{k+2}, W_{k+3}) = (Z, K, Z, K)\big\}.$$

Zum Zeitpunkt $R + 3$ können wir sehen, dass das Muster (Z, K, Z, K) zum ersten Mal „fertig“ ist ($R + 3$ ist eine sog. Stoppzeit, siehe Def. B.11 unten).

Um $\mathbb{E}[R]$ zu bestimmen stellen wir uns ein „faires Casino“ vor:

Setze vor dem i-ten Wurf x Euro, erhalte $2x$ Euro oder 0 Euro je nach Ausgang (und jeder mögliche Ausgang hat W'keit $1/2$). Spieler i steigt in Runde i in das Spiel ein und setzt einen Euro auf Z. Falls er gewinnt, setzt er in Runde $i + 1$ zwei Euro auf K. Gewinnt er wieder, setzt er in Runde $i + 2$ vier Euro auf Z. Sollte er wieder gewinnen, setzt er in Runde $i + 3$ acht Euro auf K. Gewinnt er auch dieses Spiel, hört er auf. Sei nun $X_{i,n}$ der Gewinn des i-ten Spielers nach der n-ten Runde und

$$X_n := \sum_{i=1}^{n} X_{i,n}$$

der Gesamtgewinn aller Spieler nach Runde n. Aufgrund der „Fairness“ sollte gelten

$$0 = \mathbb{E}[X_0] = \mathbb{E}[X_n] = \mathbb{E}[X_{R+3}]. \qquad (*)$$

(Wir werden die Theorie hinter $(*)$ entwickeln, siehe Korollar B.19 unten.)

[1] Zur farbigen Geschichte des Begriffs Martingal siehe beispielsweise den Artikel von Roger Mansuy, The origins of the word “martingale”, Electronic Journal for History of Probability and Statistics, Vol. 5 no. 1, (2009), `http://www.jehps.net`.

Zum Zeitpunkt $R+3$ hat Spieler R einen Gewinn von 15 Euro, Spieler $R+2$ hat einen Gewinn von 3 Euro und die anderen $R+1$ Spieler, die bisher mitgespielt haben, haben einen Gewinn von -1 Euro, das heißt

$$X_{R+3} = 15 + 3 - (R+1).$$

$(*)$ liefert

$$0 = \mathbb{E}[X_{R+3}] = \mathbb{E}[R - 17],$$

und damit $\mathbb{E}[R] = 17$.

B.1 Grundlegendes

Im Folgenden sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Definition B.2. Eine Familie $(\mathscr{F}_n)_{n=0,1,\ldots}$ von (Teil-) σ-Algebren mit

$$\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \subset \mathscr{F}$$

heißt *Filtration*. $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n=0,1,\ldots}, \mathbb{P})$ heißt filtrierter Wahrscheinlichkeitsraum.

Bemerkung B.3.
i) Interpretation: $\mathscr{F}_n$ enthält diejenigen Ereignisse, die bis zum Zeitpunkt n entschieden sind.

ii) Ist $X = (X_n)_{n=0,1,\ldots}$ eine Familie von Zufallsvariablen (ein sog. stochastischer Prozess), so ist $\mathscr{F}_n = \sigma(X_1, \ldots X_n)$, $n \in \mathbb{N}_0$ eine Filtration (die von X erzeugte Filtration).

Definition B.4. Es sei $X = (X_n)_n$ ein stochastischer Prozess und $(\mathscr{F}_n)_n$ eine Filtration. X heißt *adaptiert* (an $(\mathscr{F}_n)_n$), wenn X_n $\mathscr{F}_n$-messbar ist für alle $n \in \mathbb{N}_0$.

Definition B.5. Es sei $X = (X_n)_n$ ein (reellwertiger) stochastischer Prozess und $(\mathscr{F}_n)_n$ eine Filtration. X heißt ein *Martingal* (bzgl. $(\mathscr{F}_n)_n$ unter $\mathbb{P}$), wenn gilt:

i) X ist adaptiert (an $(\mathscr{F}_n)_n$).

ii) $X_n \in \mathcal{L}^1(\mathbb{P})$ für $n \in \mathbb{N}_0$.

iii) $\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] = X_n$ f.s. für jedes $n \in \mathbb{N}_0$.

Falls in iii) $\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] \geq X_n$ gilt, so heißt X ein *Submartingal*. Falls $\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] \leq X_n$ gilt, so heißt X ein *Supermartingal*.

Bemerkung B.6. Induktiv folgt für ein Martingal X

$$\mathbb{E}[X_n \mid \mathscr{F}_m] = X_m \text{ f.s. für alle } 0 \leq m \leq n.$$

(bzw. „≥“ für ein Sub- und „≤“ für ein Supermartingal.

Beispiel B.7. i) Seien $Y_1, Y_2, \ldots$ unabhängige, reelle Zufallsvariablen mit $Y_n \in \mathcal{L}^1(\mathbb{P})$ und $\mathbb{E}[Y_n] = 0$ für alle $n \in \mathbb{N}$. Sei $S_0 := 0$ und $S_n := Y_1 + Y_2 + \ldots + Y_n = S_{n-1} + Y_n$ für $n \in \mathbb{N}$. Dann ist $(S_n)_n$ ein Martingal bezgl. $(\mathscr{F}_n)_n$ mit $\mathscr{F}_n = \sigma(S_1, \ldots S_n) = \sigma(Y_1, \ldots Y_n)$, $n \in \mathbb{N}$ (und $\mathscr{F}_0 = \{\varnothing, \Omega\}$), denn $S_n \in \mathcal{L}^1(\mathbb{P})$ als Summe von $\mathcal{L}^1(\mathbb{P})$-Variablen und es gilt

$$\mathbb{E}[S_{n+1} \mid \mathscr{F}_n] = \mathbb{E}[S_n + Y_{n+1} \mid \mathscr{F}_n] = \underbrace{\mathbb{E}[S_n \mid \mathscr{F}_n]}_{=S_n \text{ f.s.}} + \underbrace{\mathbb{E}[Y_{n+1} \mid \mathscr{F}_n]}_{=\mathbb{E}[Y_{n+1}]=0 \text{ f.s.}} = S_n \text{ f.s.}$$

ii) Pólyas Urne: Eine Urne enthalte anfangs $s > 0$ schwarze und $w > 0$ weiße Kugeln. Ziehe jeweils eine Kugel rein zufällig und lege sie zusammen mit einer neuen Kugel der selben Farbe zurück. Sei X_n die Anzahl weißer Kugeln nach n Zügen und $A_n := \frac{X_n}{s+w+n}$ der Anteil weißer Kugeln in der Urne. Dann ist $(A_n)_n$ ein Martingal bzgl. $\mathscr{F}_n = \sigma(A_0, \ldots, A_n)$, denn auf $\{X_n = k\}$ gilt

$$\begin{aligned}\mathbb{E}[A_{n+1} \mid \mathscr{F}_n] &= \frac{k}{s+w+n} \cdot \frac{k+1}{s+w+n+1} + \frac{w+s+n-k}{s+w+n} \cdot \frac{k}{s+w+n+1} \\ &= \frac{k}{s+w+n} \cdot \underbrace{\frac{k+1+w+s+n-k}{w+s+n+1}}_{=1} = A_n.\end{aligned}$$

iii) Seien $Z_1, Z_2, \ldots$ unabhängige, positive Zufallsvariablen mit $Z_n \in \mathcal{L}^1(\mathbb{P})$ und $\mathbb{E}[Z_n] = 1$ für alle $n \in \mathbb{N}$. Sei $M_0 := 1$ und $M_n := Z_1 \cdot Z_2 \cdot \ldots \cdot Z_n = M_{n-1} \cdot Z_n$ für $n \in \mathbb{N}$. Dann ist $(M_n)_n$ ein Martingal bezgl. $(F_n)_n$ mit $\mathscr{F}_n = \sigma(M_1, \ldots M_n)$, denn $M_n \in \mathcal{L}^1(\mathbb{P})$ als Produkt von unabhängigen $\mathcal{L}^1(\mathbb{P})$-Variablen und es gilt

$$\mathbb{E}[M_{n+1} \mid \mathscr{F}_n] = \mathbb{E}[M_n \cdot Z_{n+1} \mid \mathscr{F}_n] = M_n \cdot \underbrace{\mathbb{E}[Z_{n+1} \mid \mathscr{F}_n]}_{=\mathbb{E}[Z_{n+1}]=1 \text{ f.s.}} = M_n \text{ f.s.}$$

Definition B.8. Sei $(\mathscr{F}_n)_n$ eine Filtration. Ein stochastischer Prozess $(C_n)_n$ heißt *previsibel* (bzgl. $(\mathscr{F}_n)_n$), auch vorhersagbar, wenn C_n $\mathscr{F}_{n-1}$-messbar ist für jedes $n \in \mathbb{N}$ (C_0 spielt hier keine Rolle).

Definition B.9. Sei $(X_n)_n$ adaptiert und $(C_n)_n$ previsibel bzgl. $(\mathscr{F}_n)_n$. Setze

$$(C \bullet X)_0 := 0, \quad (C \bullet X)_n := \sum_{m=1}^{n} C_m(X_m - X_{m-1}), \quad n \in \mathbb{N}. \tag{B.1}$$

Der Prozess $C \bullet X = ((C \bullet X)_n)_{n \in \mathbb{N}_0}$ heißt *(diskretes) stochastisches Integral* von C bezüglich X. $C \bullet X$ ist (offenbar) adaptiert.

Spielinterpretation: $C \bullet X$ ist ein akkumulierter Gewinnprozess für einen Spieler, der in der m-ten Runde jeweils C_m-fachen Einsatz setzt.

Lemma B.10. *Es sei $(X_n)_n$ ein Martingal und $(C_n)_n$ ein previsibler Prozess bzgl. $(\mathscr{F}_n)_n$. Es gelte mindestens eine der folgenden drei Bedingungen*

i) $(C_n)_n$ ist lokal beschränkt, d.h. es gibt Konstanten c_n mit $|C_n| \le c_n$ f.s. für alle $n \in \mathbb{N}$.

ii) $(X_n - X_{n-1})_n$ *ist lokal beschränkt und* $C_n \in \mathcal{L}^1(\mathbb{P})$ *für alle* $n \in \mathbb{N}$.

iii) $X_n, C_n \in \mathcal{L}^2(\mathbb{P})$ *für alle* $n \in \mathbb{N}_0$.

Dann ist $C \bullet X$ *ein Martingal. Ist* $C_n \geq 0$ *für alle* $n \in \mathbb{N}_0$ *und* X *ein Sub- bzw. Supermartingal, so auch* $C \bullet X$.

Beweis. i), ii) oder iii) garantieren, dass $C_m(X_m - X_{m-1}) \in \mathcal{L}^1(\mathbb{P})$, denn für iii) gilt mit der Cauchy-Schwarz-Ungleichung

$$\mathbb{E}[|C_m(X_m - X_{m-1})|] \leq (\mathbb{E}[C_m^2])^{\frac{1}{2}}(\mathbb{E}[(X_m - X_{m-1})^2])^{\frac{1}{2}} < \infty.$$

Also gilt

$$\mathbb{E}[(C \bullet X)_{n+1} \mid \mathscr{F}_n] = \underbrace{\mathbb{E}[C_{n+1}(X_{n+1} - X_n) \mid \mathscr{F}_n]}_{=C_{n+1} \cdot \underbrace{\mathbb{E}[X_{n+1} - X_n \mid \mathscr{F}_n]}_{=0} = 0 \text{ f.s.}} + \underbrace{\mathbb{E}[(C \bullet X)_n \mid \mathscr{F}_n]}_{=(C \bullet X)_n \text{ f.s.}} = (C \bullet X)_n \text{ f.s.}$$

□

Definition B.11. Sei $(\mathscr{F}_n)_n$ eine Filtration. Eine Zufallsvariable T mit Werten in $\mathbb{N}_0 \cup \{\infty\}$ heißt eine $((\mathscr{F}_n)_n$-)*Stoppzeit*, wenn $\{T \leq n\} \in \mathscr{F}_n$ für alle $n \in \mathbb{N}_0$ gilt. Für eine Stoppzeit T *ist*

$$\mathscr{F}_T := \{A \in \mathscr{F} \mid A \cap \{T \leq n\} \in \mathscr{F}_n \text{ für alle } n \in \mathbb{N}\}$$

eine σ-Algebra, sie heißt die *(σ-Algebra der) T-Vergangenheit*.

Interpretation: $\mathscr{F}_T$ enthält diejenigen Ereignisse, die sich zu dem (zufälligen) Zeitpunkt T entscheiden lassen.

Bemerkung B.12. T ist genau dann eine Stoppzeit, wenn $\{T = n\} \in \mathscr{F}_n$ für alle $n \in \mathbb{N}_0$, denn $\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^c$.

Beispiel B.13. i) Jede Konstante t_0 ist eine Stoppzeit.

ii) Es sei $(X_n)_n$ ein adaptierter stochastischer Prozess mit Werten in $(E, \mathscr{B})$ und $K \in \mathscr{B}$. Dann ist

$$T := \inf\{n \in \mathbb{N}_0 \mid X_n \in K\}$$

(mit Interpretation $\inf \varnothing = \infty$) eine Stoppzeit, denn $\{T \leq n\} = \bigcup_{m=0}^{n} \{X_m \in K\} \in \mathscr{F}_n$.

Bemerkung. $L := \sup\{n \in \mathbb{N}_0 \mid X_n \in K\}$ ist im Allgemeinen keine Stoppzeit.

Lemma B.14. *Sind* σ, τ *Stoppzeiten, so sind auch* $\sigma \wedge \tau$, $\sigma \vee \tau$ *und* $\sigma + \tau$ *Stoppzeiten.*

Beweis. Sei $n \in \mathbb{N}_0$. Es gilt

$$\{\sigma \vee \tau \leq n\} = \{\sigma \leq n\} \cap \{\tau \leq n\} \in \mathscr{F}_n, \quad \{\sigma \wedge \tau \leq n\} = \{\sigma \leq n\} \cup \{\tau \leq n\} \in \mathscr{F}_n.$$

Also sind $\sigma \wedge \tau$ und $\sigma \vee \tau$ Stoppzeiten. Dann sind auch $\sigma \wedge n$ und $\tau \wedge n$ Stoppzeiten, also gilt insbesondere für $m \leq n$: $\{\sigma \wedge n \leq m\}, \{\tau \wedge n \leq m\} \in \mathscr{F}_m \subset \mathscr{F}_n$. Dann sind

$$\sigma' := \sigma \wedge n + \mathbf{1}_{\{\sigma > n\}}, \quad \tau' := \tau \wedge n + \mathbf{1}_{\{\tau > n\}}$$

$\mathscr{F}_n$-messbar, also ist auch $\sigma' + \tau'$ $\mathscr{F}_n$-messbar. Somit gilt

$$\{\sigma + \tau \leq n\} = \{\sigma' + \tau' \leq n\} \in \mathscr{F}_n,$$

also ist auch $\sigma + \tau$ eine Stoppzeit. □

Bemerkung. $\sigma - \tau$ ist im Allgemeinen keine Stoppzeit.

Lemma B.15. *Sind σ, τ Stoppzeiten mit $\sigma \leq \tau$, dann gilt $\mathscr{F}_\sigma \subset \mathscr{F}_\tau$.*

Beweis. Sei $A \in \mathscr{F}_\sigma$ und $n \in \mathbb{N}_0$. Da $\{\tau \leq n\} \subset \{\sigma \leq n\}$, gilt

$$A \cap \{\tau \leq n\} = \underbrace{(A \cap \{\sigma \leq n\})}_{\in \mathscr{F}_n} \cap \{\tau \leq n\} \in \mathscr{F}_n.$$

□

Beobachtung B.16. Sei $(\mathscr{F}_n)_n$ eine Filtration, T eine Stoppzeit mit $T < \infty$ f.s. und $(X_n)_n$ ein adaptierter stochastischer Prozess mit Werten in $(E, \mathscr{B})$. Dann ist $X_T = X_{T(\omega)}(\omega)$ eine $\mathscr{F}_T$-messbare Zufallsvariable und $X^{(T)} = (X_n^{(T)})_{n \in \mathbb{N}_0} = (X_{T \wedge n})_{n \in \mathbb{N}_0}$ ein adaptierter stochastischer Prozess.

Beweis. Sei $B \in \mathscr{B}$. Dann ist

$$\{X_T \in B\} = \bigcup_{n=0}^{\infty} \underbrace{\{T = n, X_n \in B\}}_{\in \mathscr{F}_n \subset \mathscr{F}} \in \mathscr{F},$$

also ist X_T $\mathscr{F}$-messbar. Ebenso gilt

$$\{X_T \in B\} \cap \{T \leq n\} = \bigcup_{k=0}^{n} \underbrace{\{T = k, X_k \in B\}}_{\in \mathscr{F}_k} \subset \mathscr{F}_n,$$

also ist X_T $\mathscr{F}_T$-messbar.

Weiter ist $X_n^{(T)} = X_{T \wedge n}$ messbar bzgl. $\mathscr{F}_{T \wedge n} \subset \mathscr{F}_n$, d.h. $(X_n^{(T)})_n$ ist adaptiert. □

Bemerkung B.17. $(X_n^{(T)})_n$ ist auch adaptiert an $\mathscr{F}^{(T)} := (\mathscr{F}_{T \wedge n})_n$.

Lemma B.18. *Sei T eine Stoppzeit. Ist $(X_n)_n$ ein (Sub- / Super-) Martingal, so auch $(X_n^{(T)})_n$.*

Beweis. Sei $C_n := \mathbf{1}_{\{T \geq n\}}, n \in \mathbb{N}$. $(C_n)_n$ ist previsibel, denn $\{T \geq n\} = \{T \leq n-1\}^{\mathsf{c}} \in \mathscr{F}_{n-1}$. Schreibe

$$X_{T \wedge n} = X_0 + \sum_{m=1}^{T \wedge n} (X_m - X_{m-1}) = X_0 + \sum_{m=1}^{n} \mathbf{1}_{\{T \geq m\}} (X_m - X_{m-1}) = (C \bullet X)_n + X_0.$$

Damit folgt die Behauptung aus Lemma B.10. □

Korollar B.19. *Sei X ein Supermartingal und T eine Stoppzeit. Es gelte mindestens eine der folgenden Bedingungen*

i) *T ist beschränkt.*

ii) *X ist beschränkt, d.h. $\sup_{n \in \mathbb{N}_0} |X_n| \leq c$ f.s. für ein $c \in \mathbb{R}_+$, und $T < \infty$ f.s.*

iii) $\mathbb{E}[T] < \infty$ *und* $\sup_{n \in \mathbb{N}} |X_n - X_{n-1}| \leq c$ *f.s. für ein* $c \in \mathbb{R}_+$.

Dann gilt $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$. *(Im Falle eines Martingals gilt Gleichheit.)*

Beweis. Angenommen *i)* gilt, dann existiert ein $m \in \mathbb{N}$ mit $T \leq m$. Nach Lemma B.18 ist $(X_{T \wedge n})_n$ ein Supermartingal, also gilt

$$\mathbb{E}[X_0] = \mathbb{E}[X_{T \wedge 0}] \geq \mathbb{E}[X_{T \wedge m}] = \mathbb{E}[X_T].$$

Gilt *ii)*, so ist $\mathbb{E}[X_T] = \lim_{m \to \infty} \mathbb{E}[X_{T \wedge m}] \leq \mathbb{E}[X_0]$ mit dem Satz von der dominierten Konvergenz.

Gilt *iii)*, dann folgt $T < \infty$ f.s. und $X_{T \wedge m} \xrightarrow[m \to \infty]{} X_T$ f.s.
Es gilt

$$\sup_{m \in \mathbb{N}_0} |X_{T \wedge m}| \leq \sup_{m \in \mathbb{N}_0} \Big(|X_0| + (T \wedge m) \cdot c \Big) \leq |X_0| + cT.$$

Da $|X_0| + cT \in \mathcal{L}^1(\mathbb{P})$, ist dies eine integrable Majorante für $X_{T \wedge m}$ und es folgt wiederum mit dem Satz von der dominierten Konvergenz

$$\mathbb{E}[X_T] = \lim_{m \to \infty} \mathbb{E}[X_{T \wedge m}] \overset{i)}{\leq} \mathbb{E}[X_0].$$

□

Nochmal zu Beispiel B.1. An dieser Stelle haben wir genug Theorie entwickelt, um Beispiel B.1 „rigoros" behandeln zu können:

$R + 3$ und X dort erfüllen die Bedingung *iii)* von Korollar B.19: Es gilt

$$\begin{aligned} \mathbb{P}(R > 4k) &\leq \mathbb{P}\Big((W_{4j+1}, W_{4j+2}, W_{4j+3}, W_{4j+4}) \neq (Z, K, Z, K) \text{ für } = j = 0, 1, \ldots, k-1 \Big) \\ &= \Big(1 - (1/2)^4\Big)^k \end{aligned}$$

somit $\mathbb{E}[R] = \sum_{r=1}^{\infty} \mathbb{P}(R \geq r) < \infty$, zudem $|X_n - X_{n-1}| \leq 1 + 2 + 4 + 8 = 15$.

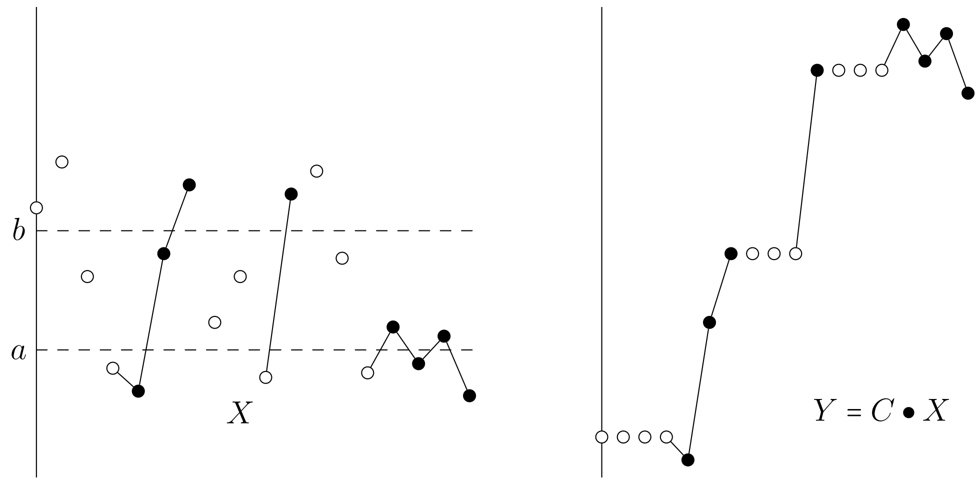


Abbildung B.1: Aufkreuzungen

B.2 Martingalkonvergenzsatz

Sei $(X_n)_n$ ein adaptierter, reellwertiger Prozess und $-\infty < a < b < \infty$.
Setze $C_1 := \mathbf{1}_{\{X_0<a\}}$ und für $n > 1$ rekursiv (siehe auch Abbildung B.1)

$$C_n := \mathbf{1}_{\{C_{n-1}=1,\ X_{n-1}\le b\}} + \mathbf{1}_{\{C_{n-1}=0,\ X_{n-1}<a\}}.$$

$(C_n)_n$ ist previsibel. Sei weiter

$$U_n^{(a,b)} := \sum_{k=1}^{n} \mathbf{1}_{\{C_k=1,\ C_{k+1}=0\}}$$

die Anzahl der abgeschlossenen Aufkreuzungen von unter a nach über b bis zur Zeit n. $U_n^{(a,b)}$ ist $\mathscr{F}_n$-messbar. Setze $Y := C \bullet X$, so gilt

$$Y_n \ge (b-a)U_n^{(a,b)} - (X_n - a)^-.$$

Lemma B.20 (Doobs[2] Aufkreuzungslemma)**.** *Sei X ein Supermartingal. Dann gilt für alle $n \in \mathbb{N}$*

$$\mathbb{E}[U_n^{(a,b)}] \le \frac{1}{b-a}\mathbb{E}[(X_n-a)^-].$$

Beweis. Nach Lemma B.10 ist $Y = C \bullet X$ ein Supermartingal. Also gilt

$$0 = \mathbb{E}[Y_0] \ge \mathbb{E}[Y_n] \ge (b-a)\mathbb{E}\big[U_n^{(a,b)}\big] - \mathbb{E}[(X_n-a)^-].$$

$\square$

Satz B.21 (Doobs (Super-) Martingalkonvergenzsatz)**.** *Ist $(X_n)_n$ ein Supermartingal mit $\sup_n \mathbb{E}[X_n^-] < \infty$, dann gibt es ein $X_\infty \in \mathcal{L}^1(\mathbb{P})$ mit $X_n \to X_\infty$ f.s.*

[2] Joseph L. Doob, 1910-2004

Beweis. Sei $a < b$. Es gilt $U_n^{(a,b)} \nearrow U_\infty^{(a,b)} = \sup_m U_m^{(a,b)}$ nach Konstruktion. Da

$$\mathbb{E}\big[U_\infty^{(a,b)}\big] = \lim_{n\to\infty} \mathbb{E}\big[U_n^{(a,b)}\big] \le \sup_n \frac{1}{b-a}\mathbb{E}[(X_n - a)^-] \le \sup_n \frac{1}{b-a}(\mathbb{E}[X_n^-] + |a|) < \infty$$

ist $U_\infty^{(a,b)} < \infty$ f.s. Für

$$O_{a,b} := \Big\{\liminf_{n\to\infty} X_n < a\Big\} \cap \Big\{\limsup_{n\to\infty} X_n > b\Big\} \subset \Big\{U_\infty^{(a,b)} = \infty\Big\}$$

gilt also $\mathbb{P}(O_{a,b}) = 0$. Damit folgt

$$\mathbb{P}\Big(\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\Big) = \mathbb{P}\left(\bigcup_{\substack{a<b\\ a,b\in\mathbb{Q}}} O_{a,b}\right) = 0.$$

Mit $X_\infty := \limsup_n X_n$ gilt also $X_n \to X_\infty$ f.s. Es bleibt $X_\infty \in \mathcal{L}^1(\mathbb{P})$ zu zeigen. Es gilt:

$$\mathbb{E}[X_\infty^-] = \mathbb{E}[\liminf_{n\to\infty} X_n^-] \overset{\text{Fatou}}{\le} \liminf_{n\to\infty} \mathbb{E}[X_n^-] < \infty$$

nach Voraussetzung und

$$\mathbb{E}[X_\infty^+] = \mathbb{E}[\liminf_{n\to\infty} X_n^+] \le \liminf_{n\to\infty} \mathbb{E}[X_n^+] = \liminf_{n\to\infty}(\mathbb{E}[X_n^-] + \mathbb{E}[X_n]) \le \mathbb{E}[X_0] + \sup_n \mathbb{E}[X_n^-] < \infty.$$

$\square$

Bemerkung B.22. 1. Die analoge Aussage von Satz B.21 gilt für ein Submartingal $(X_n)_n$ mit $\sup_n \mathbb{E}[X_n^+] < \infty$.

2. In der Situation von Satz B.21 liegt im Allgemeinen keine $\mathcal{L}^1$-Konvergenz vor, insbesondere ist $\mathbb{E}[X_\infty] \neq \lim_{n\to\infty} \mathbb{E}[X_n]$ möglich, betrachte zum Beispiel die symmetrische gewöhnliche Irrfahrt startend in 1, gestoppt bei Erreichen der 0.

B.3 Gleichgradig integrierbare Martingale und optionales Stoppen

Erinnerung B.23 (z.B. [St1, Def. 3.29 und Bem. 3.32] oder [Kl, Kap. 6.2]). Eine Familie reeller Zufallsvariablen $(X_n)_n$ heißt gleichgradig integrierbar, falls

$$\lim_{k\to\infty} \sup_n \mathbb{E}\big[|X_n| \cdot \mathbf{1}_{\{|X_n|\ge k\}}\big] = 0.$$

Es gilt:

i) $(X_n)_n$ ist genau dann gleichgradig integrierbar, falls ein $h\colon [0,\infty) \to [0,\infty)$ existiert mit $\frac{h(x)}{x} \xrightarrow[x\to\infty]{} \infty$ und $\sup_n \mathbb{E}[h(|X_n|)] < \infty$.
   Man kann annehmen, dass h monoton wachsend und konvex ist (vgl. [St1, Satz 3.33] oder [Kl, Satz 6.19]).

ii) Sei $X_n \xrightarrow{\mathbb{P}} X_\infty$. Dann gilt $X_n \xrightarrow{\mathcal{L}^1(\mathbb{P})} X_\infty$ genau dann, wenn $(X_n)_n$ gleichgradig integrierbar ist. („Konvergenzsatz von Vitali", vgl. [St1, Satz 3.31] oder[Kl, Kor. 6.26]).

Satz B.24. *Sei $(X_n)_n$ ein gleichgradig integrierbares Supermartingal. Dann existiert $X_\infty \in \mathcal{L}^1(\mathbb{P})$ mit $X_n \to X_\infty$ f.s. und in $\mathcal{L}^1(\mathbb{P})$. Es gilt*

$$\mathbb{E}[X_\infty \mid \mathcal{F}_n] \leq X_n \textit{ f.s. für alle } n \in \mathbb{N}.$$

(Analog gilt $\mathbb{E}[X_\infty \mid \mathcal{F}_n] \geq X_n$, falls $(X_n)_n$ ein gleichgradig integrierbares Submartingal, und $\mathbb{E}[X_\infty \mid \mathcal{F}_n] = X_n$, falls $(X_n)_n$ ein gleichgradig integrierbares Martingal.)

Beweis. Die Existenz von X_∞ mit $X_n \to X_\infty$ f.s. folgt aus Satz B.21. Aufgrund der gleichgradigen Integrierbarkeit gilt $\mathbb{E}[|X_n - X_\infty|] \xrightarrow[n\to\infty]{} 0$. Weiter gilt für $n \geq m$:

$$\begin{aligned}\mathbb{E}\Big[|\mathbb{E}[X_\infty \mid \mathcal{F}_m] - \mathbb{E}[X_n \mid \mathcal{F}_m]|\Big] = \mathbb{E}\Big[|\mathbb{E}[X_\infty - X_n \mid \mathcal{F}_m]|\Big] &\leq \mathbb{E}\Big[\mathbb{E}[|X_\infty - X_n| \mid \mathcal{F}_m]\Big] \\ &= \mathbb{E}[|X_\infty - X_n|] \xrightarrow[n\to\infty]{} 0\end{aligned}$$

und damit ist

$$\mathbb{E}\Big[(\mathbb{E}[X_\infty \mid \mathcal{F}_m] - X_m)^+\Big] \leq \underbrace{\mathbb{E}\Big[(\mathbb{E}[X_\infty \mid \mathcal{F}_m] - \mathbb{E}[X_n \mid \mathcal{F}_m])^+\Big]}_{\to 0} + \mathbb{E}\Big[(\underbrace{\mathbb{E}[X_n \mid \mathcal{F}_m]}_{\leq X_m \text{ für } n>m} - X_m)^+\Big].$$

Also gilt $\mathbb{E}[X_\infty \mid \mathcal{F}_m] \leq X_m$ f.s. □

Satz B.25. *Sei $Y \in \mathcal{L}^1(\mathbb{P})$ und $(\mathcal{F}_n)_n$ eine Filtration. Sei $\mathcal{F}_\infty := \sigma(\mathcal{F}_n, n \in \mathbb{N})$ und $X_n := \mathbb{E}[Y \mid \mathcal{F}_n]$. Dann ist $(X_n)_n$ ein gleichgradig integrierbares Martingal und es gilt*

$$X_n \xrightarrow[n\to\infty]{} X_\infty := \mathbb{E}[Y \mid \mathcal{F}_\infty] \quad \textit{f.s. und in } \mathcal{L}^1(\mathbb{P}).$$

Lemma B.26. *$Y \in \mathcal{L}^1(\mathbb{P})$ und $\mathcal{F}_n \subset \mathcal{F}$ σ-Algebren, dann ist $(\mathbb{E}[Y \mid \mathcal{F}_n])_{n\in\mathbb{N}}$ gleichgradig integrierbar.*

Beweis. Es existiert (vgl. Erinnerung B.23) ein $h\colon [0,\infty) \to [0,\infty)$ konvex und monoton wachsend mit $\frac{h(x)}{x} \xrightarrow[x\to\infty]{} \infty$ und $\mathbb{E}[h(|Y|)] < \infty$.

Es gilt mit der (bedingten) Jensen-Ungleichung (Satz A.7)

$$\sup_n \mathbb{E}\Big[h(|\mathbb{E}[Y \mid \mathcal{F}_n]|)\Big] \leq \sup_n \mathbb{E}\Big[\mathbb{E}[h(|Y|) \mid \mathcal{F}_n]\Big] = \mathbb{E}[h(|Y|)] < \infty.$$

□

Beweis von Satz B.25. $(X_n)_n$ ist ein Martingal und nach Lemma B.26 gleichgradig integrierbar. Nach Satz B.24 gilt

$$X_n \xrightarrow[n\to\infty]{} X_\infty := \limsup_{m\to\infty} X_m \quad \text{f.s. und in } \mathcal{L}^1(\mathbb{P}),$$

X_∞ ist $\mathcal{F}_\infty$-messbar.

Es bleibt $X_\infty = \mathbb{E}[Y \mid \mathcal{F}_\infty]$ zu zeigen. Ohne Einschränkung sei $Y \geq 0$ (ansonsten betrachte Y^+, Y^- separat). Insbesondere ist dann $X_\infty \geq 0$ f.s.

Für $A \in \mathscr{F}_\infty$ sind $\mu_1(A) := \mathbb{E}[X_\infty \mathbf{1}_A]$ und $\mu_2(A) := \mathbb{E}[Y\mathbf{1}_A]$ endliche Maße auf $(\Omega, \mathscr{F}_\infty, \mathbb{P})$. Sei $A \in \mathscr{F}_m \subset \mathscr{F}_\infty$. Dann gilt

$$\mu_1(A) = \mathbb{E}[X_\infty \mathbf{1}_A] = \lim_{n\to\infty} \mathbb{E}[X_n \mathbf{1}_A] \stackrel{n\geq m}{=} \mathbb{E}[Y\mathbf{1}_A] = \mu_2(A).$$

Da $\bigcup_{m\in\mathbb{N}} \mathscr{F}_m$ ein schnittstabiler Erzeuger von $\mathscr{F}_\infty$ ist, folgt $\mu_1 = \mu_2$ (denn σ-endl. Maße sind durch Werte auf $\cap$-stabilem Erzeuger festgelegt, vgl. z.B. [St1, Satz 1.18] oder [Kl, Lemma 1.42]). □

Bemerkung. Für ein gleichgradig integrierbares Martingal $(X_n)_n$ gilt

$$\mathbb{E}[X_\infty \mid \mathscr{F}_n] = X_n \text{ für alle } n \in \mathbb{N}.$$

Solche Martingale heißen *Doob'sche Martingale*.

Satz B.27 (optional-sampling-Theorem)**.** *Es sei $(X_n)_n$ ein gleichgradig integrierbares Martingal, $X_\infty := \lim X_n$ und T eine Stoppzeit. Dann ist $X_T \in \mathcal{L}^1(\mathbb{P})$ und es gilt $\mathbb{E}[X_\infty \mid \mathscr{F}_T] = X_T$ f.s. Insbesondere gilt $\mathbb{E}[X_T] = \mathbb{E}[X_\infty] = \mathbb{E}[X_0]$. Ist S eine Stoppzeit mit $S \leq T$, so gilt $\mathbb{E}[X_T \mid \mathscr{F}_S] = X_S$ f.s.*

Lemma B.28. *Sei $(X_n)_n$ ein Supermartingal und T eine Stoppzeit mit $T \leq m$ f.s. für ein $m \in \mathbb{N}$. Dann ist $X_T \in \mathcal{L}^1(\mathbb{P})$ und es gilt $\mathbb{E}[X_m \mid \mathscr{F}_T] \leq X_T$ f.s. Im Falle eines Martingals gilt Gleichheit.*

Beweis. Nach Lemma B.18 ist $(X_{T\wedge n})_n$ ein Supermartingal. Es gilt $X_T \in \mathcal{L}^1(\mathbb{P})$, denn $X_T = X_{T\wedge m}$ f.s. Für $A \in \mathscr{F}_T$ gilt

$$\mathbb{E}[X_m \mathbf{1}_A] = \sum_{n=0}^m \mathbb{E}\big[X_m \mathbf{1}_{\underbrace{A\cap\{T=n\}}_{\in\mathscr{F}_n}}\big] \leq \sum_{n=0}^m \mathbb{E}\big[X_n \mathbf{1}_{A\cap\{T=n\}}\big] = \mathbb{E}\left[\left(\sum_{n=0}^m X_n \mathbf{1}_{\{T=n\}}\right)\mathbf{1}_A\right] = \mathbb{E}[X_T \mathbf{1}_A].$$

□

Lemma B.29. *Ist $(X_n)_n$ ein gleichgradig integrierbares Martingal, so ist $\{X_T \mid T \text{ Stoppzeit}\}$ gleichgradig integrierbar.*

Beweis. Da $(X_n)_n$ gleichgradig integrierbar ist, existiert (vgl. Erinnerung B.23) ein $h\colon [0,\infty) \to [0,\infty)$ konvex und monoton wachsend mit $\frac{h(x)}{x} \xrightarrow[x\to\infty]{} \infty$ und $\sup_{n\in\mathbb{N}_0} \mathbb{E}[h(|X_n|)] =: M < \infty$. Sei T eine Stoppzeit und $n \in \mathbb{N}$. Es gilt

$$\begin{aligned} \mathbb{E}\big[h(|X_T|)\mathbf{1}_{\{T\leq n\}}\big] &= \mathbb{E}\big[h(|X_{T\wedge n}|)\mathbf{1}_{\{T\leq n\}}\big] \stackrel{B.28}{=} \mathbb{E}\big[h(|\mathbb{E}[X_n \mid \mathscr{F}_{T\wedge n}]|)\mathbf{1}_{\{T\leq n\}}\big] \\ &\leq \mathbb{E}\big[h(\mathbb{E}[|X_n| \mid \mathscr{F}_{T\wedge n}])\mathbf{1}_{\{T\leq n\}}\big] = \mathbb{E}\big[\mathbb{E}[h(|X_n|)\mathbf{1}_{\{T\leq n\}} \mid \mathscr{F}_{T\wedge n}]\big] \\ &\leq M. \end{aligned}$$

Mit $n \to \infty$ folgt $\mathbb{E}\big[h(|X_T|)\mathbf{1}_{\{T<\infty\}}\big] \leq M$, das heißt

$$\sup_{T \text{ Stoppzeit}} \mathbb{E}[h(|X_T|)] \leq 2M < \infty.$$

□

Beweis von Satz B.27. Nach Satz B.25 gilt $\mathbb{E}[X_\infty \mid \mathscr{F}_m] = X_m$ f.s. und nach Lemma B.28 ist

$$\mathbb{E}[X_m \mid \mathscr{F}_{T\wedge m}] = X_{T\wedge m} \text{ f.s.}$$

Also ist

$$\mathbb{E}[X_\infty \mid \mathscr{F}_{T\wedge m}] = \mathbb{E}[\mathbb{E}[X_\infty \mid \mathscr{F}_m] \mid \mathscr{F}_{T\wedge m}] = X_{T\wedge m} \text{ f.s.}$$

Sei nun $A \in \mathscr{F}_T$. Dann ist $A \cap \{T \le m\} \in \mathscr{F}_{T\wedge m}$, denn für $n \in \mathbb{N}_0$ gilt $A \cap \{T \le m\} \cap \{T \le n\} = A \cap \{T \le m \wedge n\} \in \mathscr{F}_{m\wedge n} \subset \mathscr{F}_n$. Somit gilt

$$\begin{aligned}\mathbb{E}\big[X_\infty \mathbf{1}_{A\cap\{T\le m\}}\big] &= \mathbb{E}[\mathbb{E}[X_\infty \mathbf{1}_{A\cap\{T\le m\}} \mid \mathscr{F}_{T\wedge m}]] \\ &= \mathbb{E}\big[X_{T\wedge m}\mathbf{1}_{A\cap\{T\le m\}}\big] = \mathbb{E}\big[X_T \mathbf{1}_{A\cap\{T\le m\}}\big]. \end{aligned} \tag{B.2}$$

Sei ohne Einschränkung $X_\infty \ge 0$ (sonst betrachte X_∞^+, X_∞^- separat). $m \to \infty$ in (B.2) mit monotoner Konvergenz liefert

$$\mathbb{E}\big[X_\infty \mathbf{1}_{A\cap\{T<\infty\}}\big] = \mathbb{E}\big[X_T \mathbf{1}_{A\cap\{T<\infty\}}\big].$$

(Alternativ: Verwende dominierte Konvergenz, das erspart die Zerlegung in Positiv- und Negativteil.)

Nach Definition gilt aber auch

$$\mathbb{E}\big[X_\infty \mathbf{1}_{A\cap\{T=\infty\}}\big] = \mathbb{E}\big[X_T \mathbf{1}_{A\cap\{T=\infty\}}\big],$$

d.h. $\mathbb{E}[X_\infty \mathbf{1}_A] = \mathbb{E}[X_T \mathbf{1}_A]$.

Sei $S \le T$ eine Stoppzeit. Dann ist $\mathscr{F}_S \subset \mathscr{F}_T$, also gilt

$$\mathbb{E}[X_T \mid \mathscr{F}_S] = \mathbb{E}\big[\mathbb{E}[X_\infty \mid \mathscr{F}_T] \mid \mathscr{F}_S\big] = \mathbb{E}[X_\infty \mid \mathscr{F}_S] = X_S \quad \text{f.s.}$$

□

Bemerkung und Definition B.30. Sei $(X_n)_n$ ein adaptierter Prozess mit $X_n \in \mathcal{L}^1(\mathbb{P})$ für $n \in \mathbb{N}_0$. Dann ist $X_n = M_n + A_n$ mit

$$M_0 := X_0, \quad M_n := X_0 + \sum_{k=1}^n \big(X_k - \mathbb{E}[X_k \mid \mathscr{F}_{k-1}]\big),$$

$$A_0 := 0, \quad A_n := \sum_{k=1}^n \big(\mathbb{E}[X_k \mid \mathscr{F}_{k-1}] - X_{k-1}\big).$$

wobei $(M_n)_n$ ein Martingal und $(A_n)_n$ previsibel ist. Die Darstellung $X = M + A$ als Summe eines Martingals M und eines previsiblen Prozesses A mit $A_0 = 0$ heißt *Doob-Zerlegung*, sie ist f.s. eindeutig.

$(X_n)_n$ ist genau dann ein Super- bzw. Submartingal, wenn $(A_n)_n$ nicht-wachsend bzw. nichtfallend ist.

Beweis. Wir zeigen nur die Eindeutigkeit. Angenommen $X = M + A = M' + A'$. Sei

$$\tilde{M}_n := M_n - M_n' = A_n' - A_n.$$

Dann ist $(\tilde{M}_n)_n$ ein previsibles Martingal mit $\tilde{M}_0 = 0$. Also ist $\tilde{M}_n \equiv \tilde{M}_0 \equiv 0$ f.s., denn

$$\tilde{M}_{n-1} = \mathbb{E}[\tilde{M}_n \mid \mathscr{F}_{n-1}] = \tilde{M}_n \text{ f.s.}$$

□

Satz B.31. *Sei $(X_n)_n$ ein gleichgradig integrierbares Supermartingal und seien S, T Stoppzeiten mit $S \leq T$. Dann gilt $\mathbb{E}[X_T \mid \mathscr{F}_S] \leq X_S$ f.s.*

Beweis. Sei $X_n = M_n + A_n$ die Doob-Zerlegung. Dann gilt $A_n \searrow A_\infty \leq 0$. Es ist

$$\mathbb{E}[|A_n|] = \mathbb{E}[-A_n] = \mathbb{E}[M_n - X_n] = \mathbb{E}[M_n - M_0 + X_0 - X_n] \leq \mathbb{E}[|X_n| + \mathbb{E}[|X_0|]] \leq C$$

für alle n mit einer geeigneten Konstante $C < \infty$.

Somit ist $(A_n)_n$ gleichgradig integrierbar und damit auch $(M_n)_n = (X_n - A_n)_n$. Also gilt

$$\mathbb{E}[X_T \mid \mathscr{F}_S] = \underbrace{\mathbb{E}[M_T \mid \mathscr{F}_S]}_{=M_S \text{ f.s.}} + \mathbb{E}[A_T \mid \mathscr{F}_S] \leq M_S + \mathbb{E}[A_S \mid \mathscr{F}_S] = M_S + A_S = X_S.$$

□

B.4 $\mathcal{L}^2$-Martingale

Bemerkung B.32. Sei $(X_n)_n$ ein Martingal und $\varphi\colon \mathbb{R} \to \mathbb{R}$ konvex, sodass $\mathbb{E}[\varphi(X_n)]$ für alle n existiert. Dann ist $(\varphi(X_n))_n$ ein Submartingal, denn

$$\mathbb{E}[\varphi(X_{n+1}) \mid \mathscr{F}_n] \geq \varphi\left(\mathbb{E}[X_{n+1} \mid \mathscr{F}_n]\right) = \varphi(X_n).$$

Die Aussage gilt ebenso, wenn $(X_n)_n$ ein Submartingal und φ konvex und nicht fallend ist.

Bericht B.33. Eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ heißt *harmonisch*, wenn sie die Mittelwerteigenschaft besitzt:

$$f(x) = \frac{1}{2a} \int_{x-a}^{x+a} f(y)\, dy \quad \forall x \in \mathbb{R}, a > 0$$

(man fordert auch, dass f lokal beschränkt und messbar ist, so dass die Integrale stets existieren).

$f\colon \mathbb{R} \to \mathbb{R}$ heißt *subharmonisch*, wenn $f(x) \leq \frac{1}{2a} \int_{x-a}^{x+a} f(y)\, dy$ und *superharmonisch*, wenn $f(x) \geq \frac{1}{2a} \int_{x-a}^{x+a} f(y)\, dy$ für alle $x \in \mathbb{R}, a > 0$.

Man kann zeigen: Auf $\mathbb{R}$ sind genau die konvexen Funktionen subharmonisch, zusammen mit Bemerkung B.32 motiviert dies den Namen Submartingal (und entsprechend auch den Namen Supermartingal).

Beobachtung B.34. Sei $(X_n)_n$ ein quadratintegrables Martingal, d.h. $X_n \in \mathcal{L}^2(\mathbb{P})$ für alle $n \in \mathbb{N}$. Dann gilt $\mathbb{E}[(X_k - X_l)(X_l - X_m)] = 0$ für alle $0 \leq m \leq l \leq k$, denn

$$\begin{aligned}
\mathbb{E}[(X_k - X_l)(X_l - X_m)] &= \mathbb{E}\big[\mathbb{E}[(X_k - X_l)(X_l - X_m) \mid \mathscr{F}_l]\big] \\
&= \mathbb{E}\big[(X_l - X_m)(\mathbb{E}[X_k \mid \mathscr{F}_l] - X_l)\big] \\
&= \mathbb{E}[(X_l - X_m)(X_l - X_l)] = \mathbb{E}[0] = 0.
\end{aligned}$$

Erinnerung. $\|X\|_2 := \sqrt{\mathbb{E}[X^2]}$ ist eine Norm und $\langle X, Y \rangle := \mathbb{E}[XY]$ ist ein Skalarprodukt, $\mathcal{L}^2(\mathbb{P})$ ist ein Hilbertraum (wenn man $\mathbb{P}$-f.s.-Gleichheit „herausfaktorisiert").

Man spricht Beob. B.34 auch aus als: „Martingalinkremente über disjunkte Zeitintervalle sind orthogonal."

Lemma und Definition B.35. *Sei $(X_n)_n$ ein quadratintegrables Martingal.*

$$A_n := \sum_{k=1}^{n} \mathbb{E}[(X_k - X_{k-1})^2 \mid \mathscr{F}_{k-1}]$$

ist der eindeutig bestimmte previsible Prozess mit $A_0 := 0$, sodass $(X_n^2 - A_n)_n$ ein Martingal ist. Man schreibt auch $(\langle X \rangle_n)_n = (A_n)_n$. $\langle X \rangle$ heißt quadratische Variation *von X. (In der Literatur werden auch folgende Namen verwendet: Wachsender Prozess, previsible quadratische Variation, Spitzklammerprozess von X.)*

Beweis. Es gilt:

$$\begin{aligned}
\mathbb{E}[X_{n+1}^2 - A_{n+1} \mid \mathscr{F}_n] &= \mathbb{E}[(X_{n+1} - X_n)^2 + 2(X_{n+1} - X_n)X_n + X_n^2 - A_{n+1} \mid \mathscr{F}_n] \\
&= \underbrace{\mathbb{E}[(X_{n+1} - X_n)^2 \mid \mathscr{F}_n] - A_{n+1}}_{=-A_n} + X_n^2 + 2X_n \underbrace{\mathbb{E}[X_{n+1} - X_n \mid \mathscr{F}_n]}_{=0} \\
&= X_n^2 - A_n.
\end{aligned}$$

Die Eindeutigkeit folgt aus der Eindeutigkeit der Doob-Zerlegung (Bem. und Def. B.30). □

Insbesondere gilt also:

$$\mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \mathbb{E}[\langle X \rangle_n] = \mathbb{E}[X_0^2] + \sum_{k=1}^{n} \mathbb{E}[(X_k - X_{k-1})^2]$$

und

$$\sup_n \mathbb{E}[X_n^2] < \infty \quad \Longleftrightarrow \quad \sup_n \mathbb{E}[\langle X \rangle_n] < \infty.$$

Satz B.36. *Sei $(M_n)_n$ ein $\mathcal{L}^2$-Martingal. Dann gilt:*

i) $\{\langle M \rangle_\infty < \infty\} \overset{f.s.}{\subset} \left\{\lim_{n\to\infty} M_n \textit{ existiert}\right\}$.

ii) Wenn $|M_n - M_{n-1}| \leq c$ für alle n für ein $c < \infty$, so gilt auch

$$\left\{\lim_{n\to\infty} M_n \textit{ existiert}\right\} \overset{f.s.}{\subset} \{\langle M \rangle_\infty < \infty\}.$$

iii) $\{\langle M \rangle_\infty = \infty\} \overset{f.s.}{\subset} \left\{\frac{M_n}{\langle M \rangle_n} \xrightarrow[n\to\infty]{} 0\right\}$.

Bemerkung B.37. iii) impliziert das Starke Gesetz der großen Zahlen für $M_n = Y_1 + \ldots + Y_n$ mit Y_i unabhängig und identisch verteilt mit $\mathbb{E}[Y_1] = 0$ und $\mathrm{Var}[Y_1] < \infty$.

Lemma B.38 (Kroneckers Lemma). *Sei $(x_n)_n \subset \mathbb{R}$ mit $s_k = \sum_{n=1}^{k} x_n \to s_\infty \in \mathbb{R}$. Ist $0 \leq b_n \nearrow \infty$, dann gilt $\frac{1}{b_n} \sum_{k=1}^{n} b_k x_k \xrightarrow[n\to\infty]{} 0$.*

Beweis von Satz B.36. Sei $k \in \mathbb{R}^+$.

$$S_k := \inf\{n \in \mathbb{N}_0 \mid \{M\}_{n+1} > k\}$$

ist eine Stoppzeit. Nach Lemma B.18 und Lemma B.35 ist $(M^2_{n\wedge S_k} - \langle M\rangle_{n\wedge S_k})_n$ ein Martingal, also gilt

$$\sup_n \mathbb{E}[M^2_{n\wedge S_k}] = \mathbb{E}[M_0^2] + \underbrace{\sup_n \mathbb{E}[\langle M\rangle_{n\wedge S_k}]}_{\leq k} \leq \infty,$$

das heißt $(M_{n\wedge S_k})_n$ ist $\mathcal{L}^2$-beschränkt. Also existiert $\lim_n M_{n\wedge S_k}$ f.s. für jedes $k \in \mathbb{R}^+$. Es gilt $\{\langle M\rangle_\infty < \infty\} = \bigcup_{k\in\mathbb{N}} \{S_k = \infty\}$ und $\bigcup_{k\in\mathbb{N}} \{S_k = \infty, \lim_n M_{n\wedge S_k} \text{ existiert}\} \subset \{\lim_n M_n \text{ existiert}\}$, damit gilt $i)$.

Sei $K > 0$. $T_K := \inf\{n \in \mathbb{N}_0 \mid |M_n| > K\}$ ist eine Stoppzeit und es gilt

$$\mathbb{E}[\underbrace{M^2_{n\wedge T_K}}_{\leq (K+c)^2} - \langle M\rangle_{n\wedge T_K}] = \mathbb{E}[M_0^2].$$

Also folgt mit monotoner Konvergenz

$$\mathbb{E}[\langle M\rangle_{T_K}] = \sup_n \mathbb{E}[\langle M\rangle_{n\wedge T_K}] \leq \mathbb{E}[M_0^2] + (K+c)^2 < \infty.$$

Demnach ist $\langle M\rangle_{T_K} < \infty$ f.s. und es gilt

$$\mathbb{P}\left(\underbrace{\left\{\lim_{n\to\infty} M_n \text{ existiert}\right\}}_{\subset \bigcup_{k\in\mathbb{N}}\{T_k=\infty\}} \cap \{\langle M\rangle_\infty = \infty\}\right)$$

$$\leq \mathbb{P}\left(\bigcup_{K\in\mathbb{N}} \{T_K = \infty\} \cap \{\langle M\rangle_\infty = \infty\} \cap \{\langle M\rangle_{T_K} < \infty\}\right) = 0.$$

Also gilt $ii)$.

Weiter sei

$$W_n := \sum_{k=1}^n \frac{M_k - M_{k-1}}{1 + \langle M\rangle_k} = ((1 + \langle M\rangle)^{-1} \bullet M)_n.$$

Nach Lemma B.10 ist $(W_n)_n$ ein Martingal und es gilt

$$\begin{aligned}\mathbb{E}[(W_n - W_{n-1})^2 \mid \mathscr{F}_{n-1}] &= \frac{1}{(1+\langle M\rangle_n)^2}\mathbb{E}[(M_n - M_{n-1})^2 \mid \mathscr{F}_{n-1}] = \frac{\langle M\rangle_n - \langle M\rangle_{n-1}}{(1+\langle M\rangle_n)^2} \\ &\leq \frac{(\langle M\rangle_n + 1) - (\langle M\rangle_{n-1} + 1)}{(1+\langle M\rangle_n)(1+\langle M\rangle_{n-1})} = \frac{1}{1+\langle M\rangle_{n-1}} - \frac{1}{1+\langle M\rangle_n}.\end{aligned}$$

Also gilt

$$\langle W\rangle_\infty = \lim_{N\to\infty} \langle W\rangle_N \leq \limsup_{N\to\infty} \sum_{n=1}^N \left(\frac{1}{1+\langle M\rangle_{n-1}} - \frac{1}{1+\langle M\rangle_n}\right) \leq 1$$

und somit $W_n \to W_\infty$ f.s. nach i). Wähle nun $b_n = 1 + \langle M\rangle_n$ und $x_n = \frac{M_n - M_{n-1}}{1+\langle M\rangle_n}$ in Lemma B.38, dann folgt $\sum_{k=1}^n b_k x_k = \sum_{k=1}^n M_k - M_{k-1} = M_n - M_0$, d.h. $iii)$ gilt. □

B.5 Doob-Ungleichungen

Im Folgenden sei $(X_n)_n$ ein stochastischer Prozess mit Werten in $\mathbb{R}$ und

$$X_n^* := \max_{0\le k\le n} X_k, \quad |X|_n^* := \max_{0\le k\le n} |X_k|.$$

Lemma B.39. *Sei $(X_n)_n$ ein Submartingal und $\lambda \ge 0$. Dann gilt*

$$\lambda \mathbb{P}(X_n^* \ge \lambda) \le \mathbb{E}\big[X_n \mathbf{1}_{\{X_n^* \ge \lambda\}}\big] \le \mathbb{E}\big[|X_n| \mathbf{1}_{\{X_n^* \ge \lambda\}}\big] \; (\le \mathbb{E}[|X_n|]).$$

Beweis. Sei $T := \inf\{k \in \mathbb{N}_0 \mid X_k \ge \lambda\} \wedge n$. T ist eine beschränkte Stoppzeit und es gilt mit Korollar B.19

$$\begin{aligned}\mathbb{E}\big[X_n \mathbf{1}_{\{X_n^* \ge \lambda\}}\big] + \mathbb{E}\big[X_n \mathbf{1}_{\{X_n^* < \lambda\}}\big] = \mathbb{E}[X_n] \overset{B.19}{\ge} \mathbb{E}[X_T] &= \mathbb{E}\big[X_T \mathbf{1}_{\{X_n^* \ge \lambda\}}\big] + \mathbb{E}\big[X_T \mathbf{1}_{\{X_n^* < \lambda\}}\big] \\ &\ge \lambda \mathbb{P}(X_n^* \ge \lambda) + \mathbb{E}\big[X_n \mathbf{1}_{\{X_n^* < \lambda\}}\big].\end{aligned}$$

□

Satz B.40 (Doobs $\mathcal{L}^p$-Ungleichungen)**.** *Sei $(X_n)_n$ ein Martingal oder ein nichtnegatives Submartingal.*

i) Für $p \ge 1$ und $\lambda > 0$ gilt $\lambda^p \mathbb{P}(|X|_n^* \ge \lambda) \le \mathbb{E}[|X_n|^p]$.

ii) Für $p > 1$ gilt $\mathbb{E}[|X_n|^p] \le \mathbb{E}[(|X|_n^*)^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p]$.

Bemerkung. Für ein $\mathcal{L}^2$-Martingal gestattet dies, $\mathbb{E}[(|X|_n^*)^2]$ durch $\mathbb{E}[\langle X\rangle_n]$ zu kontrollieren, denn $\mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \mathbb{E}[\langle X\rangle_n]$.

Beweis von Satz B.40. i) $(|X_n|^p)_n$ ist ein Submartingal, also folgt die Aussage durch Anwenden von Lemma B.39 auf $(|X_n|^p)_n$.

ii) Sei $c > 0$. Es gilt:

$$\begin{aligned}\mathbb{E}[(|X|_n^* \wedge c)^p] &= \mathbb{E}\left[\int_0^{|X|_n^* \wedge c} p\lambda^{p-1}\, d\lambda\right] = \mathbb{E}\left[\int_0^c p\lambda^{p-1} \mathbf{1}_{\{|X|_n^* \ge \lambda\}}\, d\lambda\right] \\ &\overset{\text{Fubini}}{=} \int_0^c p\lambda^{p-1} \mathbb{P}(|X|_n^* \ge \lambda)\, d\lambda \le \int_0^c p\lambda^{p-1} \frac{1}{\lambda} \mathbb{E}[|X_n| \mathbf{1}_{\{|X|_n^* \ge \lambda\}}]\, d\lambda \\ &\overset{\text{Fubini}}{=} p\mathbb{E}\left[|X_n| \int_0^{c\wedge |X|_n^*} \lambda^{p-2}\, d\lambda\right] = \frac{p}{p-1} \mathbb{E}[|X_n| (|X|_n^* \wedge c)^{p-1}] \\ &\overset{\text{Hölder}}{\le} \frac{p}{p-1} E[(|X|_n^* \wedge c)^p]^{\frac{p-1}{p}} \mathbb{E}[|X_n|^p]^{\frac{1}{p}}.\end{aligned}$$

Für $c \to \infty$ folgt mit monotoner Konvergenz

$$\mathbb{E}[(|X|_n^*)^p] \le \frac{p}{p-1} \mathbb{E}[(|X|_n^*)^p]^{\frac{p-1}{p}} \mathbb{E}[|X_n|^p]^{\frac{1}{p}}$$

und durch Umstellen und Potenzieren dieser Ungleichung erhält man

$$\mathbb{E}[(|X|_n^*)^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

□

Korollar B.41 (eine Form der Kolmogorov-Ungleichung). *$X_1, X_2, \dots \in \mathcal{L}^2(\mathbb{P})$ unabhängig, $\mathbb{E}[X_n] = 0$, $S_n := X_1 + \dots + X_n$ ($S_0 := 0$). Es gilt für $x > 0$*

$$\mathbb{P}\Big(\max\{|S_k| : 1 \le k \le n\} \ge x \Big) \le \frac{\mathrm{Var}[S_n]}{x^2}.$$

Beweis. $(S_n)_n$ ist ein Martingal (siehe Bsp. B.7, i)), Satz B.40, *i)* liefert

$$\mathbb{P}\Big(\max\{|S_k| : 1 \le k \le n\} \ge x \Big) \le \frac{\mathbb{E}[S_n^2]}{x^2}.$$

□

Korollar B.42. *$X_1, X_2, \dots$ u.a. mit $\mathbb{E}[X_n] = 0$, $c := \sup_n \mathrm{Var}[X_n] < \infty$, $S_n = X_1 + \dots + X_n$, dann gilt für $\varepsilon > 0$*

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{n}(\log n)^{1/2+\varepsilon}} = 0 \quad \text{f.s.}$$

Beweis. Sei $\ell(n) := \sqrt{n}\,(\log n)^{1/2+\varepsilon}$, $k_n := 2^n$, für $\delta > 0$ sei

$$A_{n,\delta} := \Big\{ \max_{k \le k_n} |S_k| > \delta \ell(k_n) \Big\}.$$

Nach Kor. B.41 ist

$$\sum_{n=1}^{\infty} \mathbb{P}(A_{n,\delta}) \le \sum_{n=1}^{\infty} \frac{1}{\delta^2 \ell(k_n)^2} \mathbb{E}\big[S_{k_n}^2\big] \le \sum_{n=1}^{\infty} \frac{c k_n}{\delta^2 \ell(k_n)^2} = \frac{c}{\delta^2} \sum_{n=1}^{\infty} \frac{1}{\big(\log(2^n)\big)^{1+2\varepsilon}} < \infty,$$

mit Borel-Cantelli also

$$\limsup_{n\to\infty} \frac{1}{\ell(k_n)} \max\big\{|S_k| : k \le k_n\big\} \le \delta \quad \text{f.s.}$$

Mit $\delta \downarrow 0$ folgt $\max\big\{|S_k| : k \le k_n\big\}/\ell(k_n) \to 0$ f.s. für $n \to \infty$ und daher auch

$$\limsup_{m\to\infty} \frac{|S_m|}{\ell(m)} \le \limsup_{m\to\infty} \underbrace{\frac{\ell(k_{\lceil \log_2 m \rceil})}{\ell(m)}}_{\text{ist beschr.}} \frac{\max\big\{|S_k| : k \le k_{\lceil \log_2 m \rceil}\big\}}{\ell(k_{\lceil \log_2 m \rceil})} = 0 \quad \text{f.s.}$$

□

Bericht B.43. Seien $X_1, X_2, \dots$ u.i.v. mit $\mathbb{E}[X_1] = 0$, $\sigma^2 := \mathrm{Var}[X_1] \in (0, \infty)$, $S_n = X_1 + \dots + X_n$. Für sehr großes n ist

$$\frac{S_n}{\sigma\sqrt{n}} \overset{d}{\approx} \mathcal{N}_{0,1}$$

gemäß dem zentralen Grenzwertsatz, d.h. für festes (und großes) n sind die typischen Fluktuationen von S_n von der Ordnung $O(\sqrt{n})$.

Korollar B.42 zeigt, dass die Fluktuationen von S_n „global“ (d.h., wenn wir „irgendwo“ auf dem Pfad nachschauen dürfen), höchstens um einen Faktor $(\log n)^{1/2+\varepsilon}$ größer sind als $\sqrt{n}$. Der Faktor aus Kor. B.42 ist nicht scharf, tatsächlich gilt

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\sigma^2 \log\log n}} = 1 \quad \text{f.s.,}$$

das sogenannte Gesetz vom iterierten Logarithmus (siehe z.B. [Kl, Satz 22.11]).

B.6 Zum Satz von Radon-Nikodým

Sei $(S, \mathscr{A})$ ein messbarer Raum, μ, ν Maße auf $(S, \mathscr{A})$. ν ist *absolut stetig* bezüglich μ, geschrieben $\nu \ll \mu$, wenn für jedes $A \in \mathscr{A}$ gilt

$$\mu(A) = 0 \quad \Longrightarrow \quad \nu(A) = 0.$$

Wenn ν eine Dichte bezüglich μ besitzt ($\nu = h\mu$, also $\nu(A) = \int \mathbf{1}_A h \, d\mu$ mit einem $h : S \to \overline{\mathbb{R}}_+$ messbar, vgl. [St1, Def. 3.16] so gilt offenbar $\nu \ll \mu$.

Satz B.44 (Eine Form des Satzes von Radon-Nikodým[3])**.** *Sei $(S, \mathscr{A})$ separabler messbarer Raum (d.h. $\mathscr{A} = \sigma(A_1, A_2, \dots)$ ist abzählbar erzeugt), μ, ν endliche Maße auf $(S, \mathscr{A})$ mit $\nu \ll \mu$. Dann gibt es ein $h \geq 0$, $h \in \mathcal{L}^1(\mu)$ mit $\nu = h\mu$.*

Beweis. O.E. sei $\mu(S), \nu(S) > 0$, dann sind $P(\cdot) := \mu(\cdot)/\mu(S)$, $Q(\cdot) := \nu(\cdot)/\nu(S)$ W'maße auf $(S, \mathscr{A})$ mit $Q \ll P$ und es genügt zu zeigen, dass $Q = XP$ für ein $X \in \mathcal{L}^1(P)$.

Wir beobachten zunächst:

$$\text{für } \varepsilon > 0 \text{ gibt es } \delta > 0 \text{ mit } \quad P(B) < \delta \;\Rightarrow\; Q(B) < \varepsilon \qquad (*)$$

Wenn $(*)$ nicht gälte, so gäbe es $B_n \in \mathscr{A}$ und ein $\varepsilon_0 > 0$ mit

$$P(B_n) \leq 2^{-n}, \; n \in \mathbb{N}, \qquad \inf_{n \in \mathbb{N}} Q(B_n) \geq \varepsilon_0.$$

Dann gälte für $C := \limsup_{n \to \infty} B_n = \bigcap_n \bigcup_{m \geq n} B_m$ einerseits $P(C) = 0$ mit Borel-Cantelli, andererseits wegen $\inf_{n \in \mathbb{N}} Q\big(\bigcup_{m \geq n} B_m\big) \geq \varepsilon_0$ auch $Q(C) \geq \varepsilon_0$ im Widerspruch zu $Q \ll P$.

Sei

$$\mathscr{F}_n := \sigma(A_1, \dots, A_n) = \{C_{n,1}, \dots, C_{n,m_n}\}$$

(mit $S = \dot\bigcup_{k=1}^{m_n} C_{n,k}$, $C_{n,k} \neq \varnothing$), für $\omega \in S$ setze

$$X_n(\omega) = \begin{cases} \frac{Q(C_{n,k})}{P(C_{n,k})}, & \omega \in C_{n,k} \text{ und } P(C_{n,k}) > 0, \\ 0, & \omega \in C_{n,k} \text{ und } P(C_{n,k}) = 0. \end{cases}$$

$X_n \geq 0$ ist $\mathscr{F}_n$-messbar, $X_n \in \mathcal{L}^1(P)$ mit $\mathbb{E}_P[X_n] = 1$ und für $B \in \mathscr{F}_n$ (d.h. $B = \dot\bigcup_{j \in J} C_{n,j}$ für ein $J \subset \{1, \dots, m_n\}$) ist

$$\mathbb{E}_P[X_n \mathbf{1}_B] = \sum_{j \in J} \mathbb{E}_P[X_n \mathbf{1}_{C_{n,j}}] = \sum_{j \in J} Q(C_{n,j}) = Q(B),$$

d.h. X_n ist die Dichte von $Q|_{\mathscr{F}_n}$ bzgl. $P|_{\mathscr{F}_n}$.

Offenbar ist $(\mathscr{F}_n)_n$ eine Filtration, für $B' \in \mathscr{F}_n$ $(\subset \mathscr{F}_{n+1})$ ist $\mathbb{E}_P[X_{n+1} \mathbf{1}_{B'}] = Q(B') = \mathbb{E}_P[X_n \mathbf{1}_{B'}]$ nach obigem, d.h.

$$\mathbb{E}_P[X_{n+1} \mid \mathscr{F}_n] = X_n \quad \text{(f.s.)}$$

[3]Johann Radon, 1887–1956; Otton Nikodým, 1887–1974

$(X_n)_n$ ist also ein nicht-negatives (P-)Martingal.

Zeige: $\{X_n, n \in \mathbb{N}\}$ ist gleichgradig (P-)integrierbar.
Zu $\varepsilon > 0$ wähle $\delta > 0$ gemäß $(*)$, setze $M := 1/\delta$, dann gilt für jedes $n \in \mathbb{N}$

$$P(X_n > M) \le \frac{1}{M}\mathbb{E}_P[X_n] = \delta,$$

mit $(*)$ also

$$\mathbb{E}_P\big[X_n \mathbf{1}_{\{X_n \ge M\}}\big] = Q(X_n > M) \le \varepsilon.$$

Mit Satz B.24 gibt es $X_\infty \in \mathcal{L}^1(P)$ mit $\mathbb{E}_P[X_\infty] = 1$ und $X_n = \mathbb{E}_p[X_\infty \mid \mathscr{F}_n] \to X_\infty$ f.s. und in $\mathcal{L}^1(P)$.

X_∞ leistet das Gewünschte: Das W'maß $\widetilde{Q}(A) := \mathbb{E}_P[X_\infty \mathbf{1}_A]$, $A \in \mathscr{A}$ erfüllt

$$\widetilde{Q}(B) = \mathbb{E}_P[X_\infty \mathbf{1}_B] = \mathbb{E}_P[X_n \mathbf{1}_B] = Q(B) \quad \text{falls } B \in \mathscr{F}_n \text{ für ein } n \in \mathbb{N},$$

d.h. $Q = \widetilde{Q}$ auf $\bigcup_n \mathscr{F}_n$ (und dies ist ein $\cap$-stabiler Erzeuger von $\mathscr{A}$). Mit dem Eindeutigkeitssatz für Maße (z.B. [St1, Satz 1.18] oder [Kl, Lemma 1.42])) folgt $Q = \widetilde{Q}$. □

Bericht B.45. 1. Man nennt die Dichte h von $\nu = h\mu$ bezüglich μ auch die Radon-Nikodým-Ableitung (von ν bezüglich μ) und notiert dies auch als $\frac{d\nu}{d\mu} = h$.

2. Satz B.44 gilt genauso, wenn μ und ν σ-endlich sind (zerlege $S = \dot{\bigcup}_k S_k$, so dass $\mu(S_k), \nu(S_k) < \infty$, dann gibt es jeweils $\frac{d\mathbf{1}_{S_k}\nu}{d\mathbf{1}_{S_k}\mu}$).

Auch auf die Forderung, dass $\mathscr{A}$ einen abzählbaren Erzeuger besitzt, kann man verzichten, siehe z.B. [Wi, Ch. 14.13].

Literaturverzeichnis

[Kl] A. Klenke, Wahrscheinlichkeitstheorie, 4. Aufl., Springer, 2020.

[Ka] O. Kallenberg, Foundations of modern probability, 3rd ed., Springer, 2021.

[Du] R. Durrett, Probability : theory and examples, 5th ed., Cambridge Univ. Press, 2019.

[Wi] D. Williams, Probability with martingales, Cambridge Univ. Press, 1995.

[Fe] W. Feller, An Introduction to Probability Theory, Band 1 und Band 2, Wiley 1968 und 1971.

[Bi] P. Billingsley, Convergence of probability measures, 2. ed., Wiley, 1999.

[Br] L. Breiman: Probability, Wiley, 1968.

[MP] P. Mörters, Y. Peres, Brownian motion, Cambridge University Press, 2010.

[RW] L.C.G. Rogers, D. Williams, Diffusions, Markov processes and martingales, Band I und II, Wiley, 1994.

[K05] O. Kallenberg, Probabilistic symmetries and invariance principles, Springer, 2005.

[St1] M. Birkner, Notizen zur Vorlesung Stochastik I, JGU Mainz, SS 2024. https://www.staff.uni-mainz.de/birkner/StochI_24/Stochastik_I_SS24.pdf