Stochastik II

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

Matthias Birkner

Version vom 27. Februar 2025

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 LaTeX gesetzt hatte. Dafür hier nochmals mein herzlicher Dank. Ich danke Anton Vogt, der durch seine kritische Lektüre im WS 2024/25 an verschiedenen Stellen zur Korrektur von Tippfehlern beigetragen hat.

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 $i_1, \ldots i_n \in I$, sowie paarweise verschiedene $j_1, \ldots, j_n \in I$ gilt

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

Beweis. Ist $(X_i)_{i \in I}$ austauschbar, so gilt $\mathscr{L}((X_{i_1}, \ldots, X_{i_n})) = \mathscr{L}((X_{j_1}, \ldots, X_{j_n}))$ 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“ $B_{i_1} \times \ldots \times B_{i_k} \times \bigtimes_{i \in I \setminus \{i_1, \ldots i_k\}} E_i$ erzeugen die Produkt-σ-Algebra auf $\bigtimes_{i \in I} E_i$. □

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 $X_i = \mathbf{1}_{\{U_i \leq Y\}}$ 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. $\mathcal{S}_n := \{\text{Permutationen von } \{1, 2, \ldots, n\}\}$. 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. $f : E^{\mathbb{N}} \to E'$ 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) $a_n : \mathbb{R}^\infty \to \mathbb{R},\ a_n(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ ist n-symmetrisch, aber nicht m-symmetrisch für $m > n$.

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

iv) Für $x \in E^\infty$ ist $\xi(x) = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$ 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.

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

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

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

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 $B \in \mathscr{B}(E) \setminus \{\varnothing, E\}$. $S := \sum_{n=1}^\infty \mathbf{1}_B(X_n)$ 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 $n \ge k$ 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) $\mathbb{E}[\varphi(X) \mid \mathscr{E}_n] = \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n} \varphi(X^\pi) = A_n(\varphi)(X)$.

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

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

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:

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

(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:

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

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

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

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

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

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

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 $\lim_n A_n(\varphi)(X) = \lim_m A_{n_m}(\varphi)(X)$ $\sigma(X_l, X_{l+1}, \ldots)$-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

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

$\square$

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 $A \in \mathscr{E} \subset \sigma(X_1, X_2, \ldots)$. Wähle $A_k \in \sigma(X_1, \ldots, X_k)$ 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 $A_k = \{(X_1, \ldots, X_k) \in C_k\}$ 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 $P(\{\psi = 1\} \triangle A) \leq \mathbb{P}(N) = 0$. $\square$

Korollar 1.18 (0-1-Gesetz von Hewitt und Savage[1]). *Seien $X_1, X_2, \ldots$ 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. $\square$

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

1.3 Struktur unendlicher austauschbarer Familien

Eine Heuristik. Sei $E = \{1, 2, \ldots, k\}$ und seien $X_1, X_2, \ldots$ 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 $(X_1, X_2, \ldots, X_N)$ verteilt wie Züge ohne Zurücklegen aus einer Urne. Für $m_l := |\{1 \le i \le n \mid x_i = l\}|$ mit $m_1 + \ldots + m_k = n$ und Notation $(a)_n := a(a-1)\ldots(a-n+1)$ gilt

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

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 $X_1, X_2, \ldots$ 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

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

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. □

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 $t_1 < t_2 < t_3 < t_4$, 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.

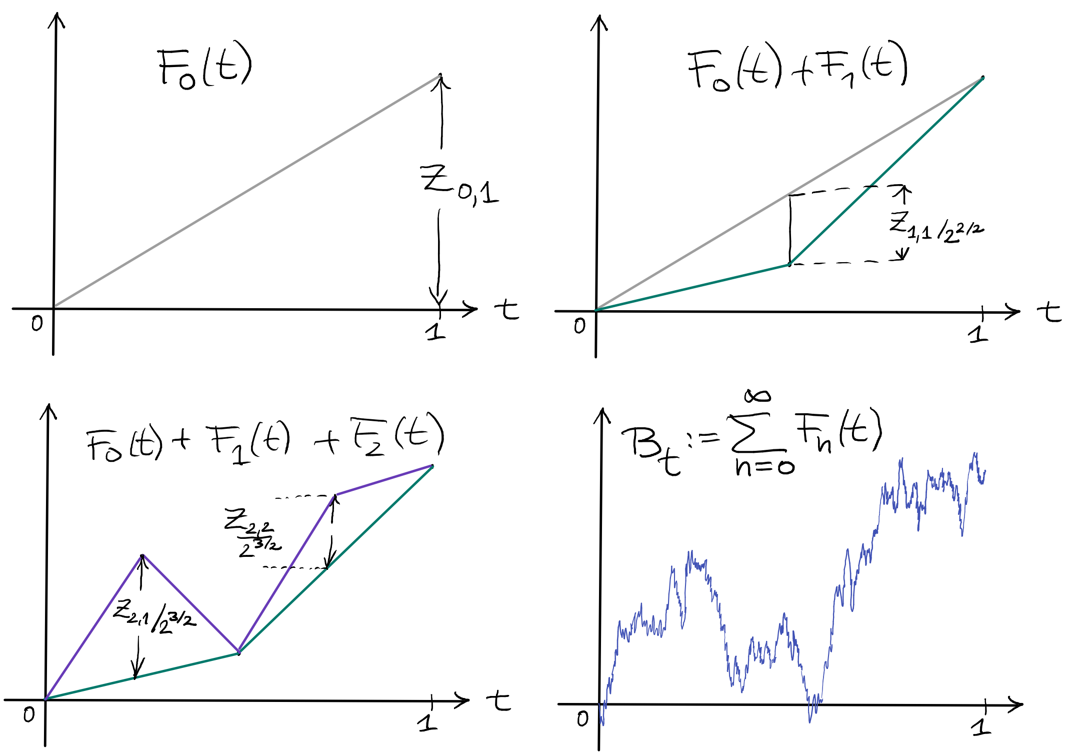


Abbildung 2.1: Skizze der Lévy-Konstruktion der Brownschen Bewegung

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 $\sum\limits_{n=0}^{\infty} \|F_n\|_\infty < \infty$. 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 $\ell > N_0$

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

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

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

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 $A \in \mathscr{B}(E)$.

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

   μ heißt *regulär*, wenn es von innen und von außen regulär ist.

iii) 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 $f: \mathbb{R}^d \to [0, \infty]$ 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 $\mu, \nu \in \mathcal{F}$ gilt:

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

Beispiel 3.4. $f\colon E \to \mathbb{R}$ 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 $\mu_1, \mu_2, \dots, \mu \in \mathcal{M}_f(E)$. 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 $\mu_1, \mu_2, \dots, \mu \in \mathcal{M}(E)$. 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

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

□

Satz 3.8 (Portmanteau-Theorem)**.** *Sei E ein metrischer Raum und $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_{\le 1}(E)$. 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,
$\liminf \mu_n(A) \geq \liminf \mu_n(A \setminus \partial A) \geq \mu(A \setminus \partial A) = \mu(A)$,
$\limsup \mu_n(A) \leq \limsup \mu_n(A \cup \partial A) \leq \mu(A \cup \partial A) = \mu(A)$.)

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

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

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 $X, X_1, X_2, \ldots$ 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 $X, X_1, X_2, \ldots$ 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 $X, X_1, X_2, \ldots$ 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 $X, X_1, X_2, \ldots, Y_1, Y_2, \ldots$ 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 $|f(x) - f(y)| \leq L_f \cdot d(x,y) \wedge 2\|f\|_\infty$. Also gilt

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

□

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

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 $iii)$

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

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

□

Bemerkung 3.12 (Der Fall $E = \mathbb{R}$)**.** $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$ 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 $F_n \xrightarrow{\mathrm{D}} F$ oder $F_n \Rightarrow F$, $:\Leftrightarrow$

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

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

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

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

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

nach Satz 3.8, *vi)*.

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

$$\int f \, d\mu_n \xrightarrow[n\to\infty]{} \int f \, d\mu \tag{3.2}$$

Sei $\varepsilon > 0$, wähle $y_0 < y_1 < \dots < y_N$ 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)

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

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

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

mit $\varepsilon \downarrow 0$ folglich

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

Dasselbe Argument angewendet auf $-f$ zeigt $\liminf\limits_{n\to\infty} \int f\, d\mu_n \ge \int f\, d\mu$, 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 $\mathcal{M}_{\le 1}(E)$ 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

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

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 $\mathcal{K} \subset \mathcal{M}_{\le 1}(E)$. 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 $\mathcal{M}_{\le 1}(E)$ 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

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

$\bigcup_{j=1}^{\infty} K_j$ 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

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

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

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

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

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

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 $x_1, x_2, \ldots$ 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

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

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

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

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

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

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

also ist $\mathcal{K}$ straff. □

Beobachtung 3.17. Sei E ein polnischer Raum und seien $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\le 1}(E)$. 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 $\mu_{n_{k_j}} \xrightarrow[j\to\infty]{\text{w}} \nu$. 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 $\mu = \text{w} - \lim \mu_n$ 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) $\mu = \text{w} - \lim \mu_n$.

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

iii) $\mu = \text{v} - \lim \mu_n$ 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 $c := \mu(E) \vee \sup_n \mu_n(E) < \infty$. Dann sind

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

also $\mu'_n \xrightarrow{\text{v}} \mu'$.

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

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

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

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 $\langle t, x\rangle = \sum_{j=1}^d t_j x_j$.) 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 $\varphi_{\mathcal{N}(\mu,\sigma^2)}(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$. 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) $|\varphi_X(t)| \le 1 = \varphi_X(0)$ *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) $0 \leq 1 - \mathrm{Re}\varphi_X(2t) \leq 4(1 - \mathrm{Re}\varphi_X(t))$.

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

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

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

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:

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

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 $x \neq y \in E$ ein $f \in \mathcal{C}$ existiert mit $f(x) \neq f(y)$.

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

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

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 $h_z(y) \leq f(y) + \varepsilon$ 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

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

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 $\mathbb{P}(Y_b = be^k) = \frac{1}{C_b} b^{-k} e^{-\frac{1}{2}k^2},\ k \in \mathbb{Z}$ 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 $|t - t'| < \delta$:

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

□

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

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

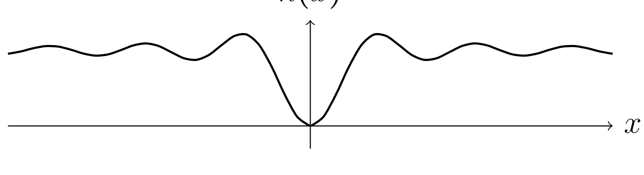
Dazu genügt es zu zeigen, dass $P_n^{(k)} := P_n \circ \pi_k^{-1}$, wobei $\pi_k \colon \mathbb{R}^d \to \mathbb{R}$ die k-te Koordinatenprojektion ist, für jedes $k = 1, \dots, d$ eine straffe Familie sind, denn es gilt

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

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

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



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

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

$h \mapsto \mathbb{E}\Big[e^{itX}\sum_{k=0}^n \frac{(ih)^k X^k}{k!}\Big] = \sum_{k=0}^n \frac{h^k}{k!}\mathbb{E}\left[(iX)^k e^{itX}\right]$ 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

Literaturhinweise Die Themen dieses Kapitels finden sich beispielsweise in [Kl, Kap. 15], [Du, Ch. 3.4], mit breiterem Fokus in [Ka, Ch. 6] und in [Fe, Vol. 2, Ch. XV.5–XV.7].

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

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

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* $\sum_{\ell=1}^{k_n} \mathrm{Var}[X_{n,\ell}] = 1$ *(das Schema ist „zentriert“ und „normiert“). Setze*

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

Es gelte die Lindeberg[1]-Bedingung

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

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*

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

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

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

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

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

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 $z_1, \dots, z_n, z_1', \dots, z_n' \in \{z \in \mathbb{C} \mid |z| \le 1\}$, 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

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

und der allgemeine Fall folgt induktiv.

Sei $c_{n,\ell} := \mathbb{E}[X_{n,\ell}^2] = \mathrm{Var}[X_{n,\ell}]$ 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 $\varepsilon > 0$

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

Damit folgt

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

[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 $X = (X_1, \dots, X_d)^{\mathsf{t}}$ 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 $a = (a_1, \dots, a_d)^{\mathsf{t}} \in \mathbb{R}^d$ ist

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

ii) Sind $Z_1, \dots, Z_d$ unabhängig und standardnormalverteilt, so hat $Z = (Z_1, \dots, Z_d)^{\mathsf{t}}$ 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 $C := AA^{\mathsf{t}}$, denn

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

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. □

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* $\operatorname{Cov}[X_1] = C$. *Setze* $S_n^* := \frac{X_1+\cdots+X_n-n\mu}{\sqrt{n}}$. *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 $S_n^\lambda \Rightarrow \mathcal{N}(0, \operatorname{Var}[X_1^\lambda]) = \mathcal{N}(0, \langle\lambda, C\lambda\rangle)$, 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. $\square$

Kapitel 5

Unendlich teilbare Verteilungen

Literaturhinweise Die Themen dieses Kapitels finden sich beispielsweise in [Kl, Kap. 16], [Du, Ch. 3.8–3.9], [Ka, Ch. 7], [Fe, Vol. 2, Ch. XVII].

Definition 5.1. Eine reelle Zufallsvariable X heißt *unendlich teilbar* (auch *unbegrenzt teilbar*), wenn es zu jedem $n \in \mathbb{N}$ unabhängige und identisch verteilte Zufallsvariablen $X_{n,1}, \dots, X_{n,n}$ gibt mit

$$X \overset{\mathrm{d}}{=} X_{n,1} + \cdots + X_{n,n}.$$

Analog heißt $\mu \in \mathcal{M}_1(\mathbb{R})$ *unendlich teilbar*, wenn es zu jedem $n \in \mathbb{N}$ ein μ_n gibt mit $\mu = \mu_n * \cdots * \mu_n = \mu_n^{*n}$.

Eine charakteristische Funktion φ eines Wahrscheinlichkeitsmaßes μ auf $\mathbb{R}$ heißt *unendlich teilbar*, wenn es zu jedem $n \in \mathbb{N}$ eine charakteristische Funktion φ_n eines Wahrscheinlichkeitsmaßes gibt mit $\varphi = \varphi_n^n$.

Beispiel. i) δ_x ist unendlich teilbar, denn $\delta_x = \left(\delta_{\frac{x}{n}}\right)^{*n}$.

ii) $\mathcal{N}(\mu, \sigma^2)$ ist unendlich teilbar, denn $\mathcal{N}(\mu, \sigma^2) = \mathcal{N}\left(\frac{\mu}{n}, \frac{\sigma^2}{n}\right)^{*n}$.

iii) Die Gammaverteilung $\mathrm{Gamma}(r, \lambda)$ mit Formparameter r und Skalenparameter λ mit Dichte $f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x)$ ist unendlich teilbar, denn $\mathrm{Gamma}(r, \lambda) = \mathrm{Gamma}\left(\frac{r}{n}, \lambda\right)^{*n}$.

iv) Die Cauchyverteilung $\mathrm{Cauchy}(a)$ mit Parameter a und Dichte $f(x) = \frac{1}{a\pi} \frac{1}{1+\left(\frac{x}{a}\right)^2}$ ist unendlich teilbar, denn $\mathrm{Cauchy}(a) = \mathrm{Cauchy}\left(\frac{a}{n}\right)^{*n}$.

v) $\mathrm{Poi}(\lambda)$ ist unendlich teilbar, denn $\mathrm{Poi}(\lambda) = \mathrm{Poi}\left(\frac{\lambda}{n}\right)^{*n}$.

Beispiel und Definition 5.2. Zu $\nu \in \mathcal{M}_f(\mathbb{R}) \setminus \{0\}$ heißt

$$\mathrm{CPoi}(\nu) = e^{-\nu(\mathbb{R})} \sum_{n=0}^{\infty} \frac{1}{n!} \nu^{*n}$$

mit $\nu^{*0} := \delta_0$ die *zusammengesetzte Poisson-Verteilung* (engl. *compound Poisson distribution*) mit Intensitätsmaß ν.

Sind $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit $X_i \sim \tilde{\nu} := \frac{\nu}{\nu(\mathbb{R})}$ für $i \in \mathbb{N}$ und ist $N \sim \mathrm{Poi}(\nu(\mathbb{R}))$ unabhängig von $X_1, X_2, \ldots$, so hat

$$S := \sum_{j=1}^{N} X_j$$

die Verteilung $\mathrm{CPoi}(\nu)$ und die charakteristische Funktion ist

$$\varphi_{\mathrm{CPoi}(\nu)}(t) = \exp\left(\int_{\mathbb{R}} \left(e^{itx} - 1\right) \nu(dx) \right).$$

Es gilt $\mathrm{CPoi}(\nu + \nu') = \mathrm{CPoi}(\nu) * \mathrm{CPoi}(\nu')$, insbesondere ist $\mathrm{CPoi}(\nu)$ unendlich teilbar.

Beweis. Sei $A \in \mathscr{B}(\mathbb{R})$. Dann gilt:

$$\begin{aligned}
\mathbb{P}(S \in A) &= \sum_{k=0}^{\infty} \mathbb{P}(N = k, S \in A) = \sum_{k=0}^{\infty} \mathbb{P}(N = k, X_1 + \cdots + X_k \in A) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N = k) \cdot \mathbb{P}(X_1 + \cdots + X_k \in A) = \sum_{k=0}^{\infty} e^{-\nu(\mathbb{R})} \frac{\nu(\mathbb{R})^k}{k!} \cdot \tilde{\nu}^{*k}(A) \\
&= \sum_{k=0}^{\infty} e^{-\nu(\mathbb{R})} \frac{\nu(\mathbb{R})^k}{k!} \cdot \frac{\nu^{*k}(A)}{\nu(\mathbb{R})^k} = \sum_{k=0}^{\infty} e^{-\nu(\mathbb{R})} \frac{1}{k!} \cdot \nu^{*k}(A) = \mathrm{CPoi}(\nu)(A).
\end{aligned}$$

Für die charakteristische Funktion betrachte

$$\begin{aligned}
\varphi_{\mathrm{CPoi}(\nu)}(t) &= \mathbb{E}\left[e^{itS}\right] = \sum_{k=0}^{\infty} \mathbb{E}\left[e^{it(X_1 + \cdots + X_k)} \mathbf{1}_{\{N=k\}}\right] = \sum_{k=0}^{\infty} \left(\mathbb{E}\left[e^{itX_1}\right]\right)^k \cdot \mathbb{P}(N = k) \\
&= \sum_{k=0}^{\infty} e^{-\nu(\mathbb{R})} \frac{\nu(\mathbb{R})^k}{k!} \left(\int_{\mathbb{R}} e^{itx} \, \tilde{\nu}(dx) \right)^k = \exp\left(\int_{\mathbb{R}} e^{itx} \, \nu(dx) - \nu(\mathbb{R}) \right) \\
&= \exp\left(\int_{\mathbb{R}} \left(e^{itx} - 1\right) \nu(dx) \right).
\end{aligned}$$

□

Satz 5.3. *Ein $\mu \in \mathcal{M}_1(\mathbb{R})$ ist genau dann unendlich teilbar, wenn es eine Folge $(\nu_n)_n \subset \mathcal{M}_f(\mathbb{R}) \setminus \{0\}$ gibt mit $\mathrm{CPoi}(\nu_n) \xrightarrow[n\to\infty]{w} \mu$.*

Lemma 5.4. *Sei $(\varphi_n)_n$ eine Folge von charakteristischen Funktionen von Wahrscheinlichkeitsmaßen. Dann sind äquivalent:*

i) $\varphi(t) = \lim_n \varphi_n^n(t)$ *existiert für alle* $t \in \mathbb{R}$ *und* φ *ist stetig in* 0.

ii) $\psi(t) = \lim_n n\left(\varphi_n(t) - 1\right)$ *existiert für alle* $t \in \mathbb{R}$ *und* ψ *ist stetig in* 0.

Dann gilt $\varphi = e^{\psi}$ *(inbesondere* $\varphi(t) \neq 0$ *für alle* t*) und* φ *ist die charakteristische Funktion eines Wahrscheinlichkeitsmaßes.*

Beweis. Für alle $z \in \mathbb{C}$ mit $|z-1| < \frac{1}{2}$ gilt mit Taylorentwicklung

$$|\log(z) - (z-1)| \leq \frac{|z-1|^2}{2}.$$

Insbesondere gilt für eine Folge $(z_n) \subset \mathbb{C}$:

$$\limsup_{n\to\infty} n|z_n - 1| < \infty \quad \Leftrightarrow \quad \limsup_{n\to\infty} n \log(z_n) < \infty.$$

Sofern einer der Limiten existiert, gilt also

$$\lim_{n\to\infty} n(z_n - 1) = \lim_{n\to\infty} n \log(z_n). \tag{5.1}$$

$ii) \Rightarrow i)$: Wähle $z_n = \varphi_n(t)$, dann folgt mit (5.1), dass $\lim_n \log\left(\varphi_n^n(t)\right) = \log\left(\lim_n \varphi_n^n(t)\right)$ existiert für alle $t \in \mathbb{R}$.

$i) \Rightarrow ii)$: Wir nehmen zunächst an, dass $\varphi(t) \neq 0$ für alle $t \in \mathbb{R}$ gilt. Dann können wir (den komplexen) $\log(\cdot)$ auf die Voraussetzung $\varphi(t) = \lim_n \varphi_n^n(t)$ anwenden (wir verwenden für gegebenes t einen Zweig, der in einer Umgebung von $\varphi(t) \neq 0$ analytisch ist) und erhalten mit $z_n = \varphi_n(t)$ aus (5.1) die Behauptung $ii)$.

Offensichtlich gilt die Beziehung $\varphi(t) = e^{\psi(t)}$.

Um $\varphi(t) \neq 0$ sicherzustellen, zeigen wir, dass ein $\gamma > 0$ existiert mit

$$|\varphi(t)| \geq \frac{1}{2} e^{-\gamma t^2} \quad \text{für jedes } t \in \mathbb{R}. \tag{5.2}$$

Nach Satz 3.27 (Lévys Stetigkeitssatz) ist φ die charakteristische Funktion eines Wahrscheinlichkeitsmaßes. Ebenso sind auch $|\varphi|^2 = \varphi\overline{\varphi}$ und $|\varphi_n|^2$ charakteristische Funktionen von Wahrscheinlichkeitsmaßen, daher gilt $|\varphi_n(t)|^{2n} \to |\varphi(t)|^2$ lokal gleichmäßig nach Satz 3.27. Es existiert ein $\varepsilon > 0$ mit $\inf_{|t|\leq\varepsilon} |\varphi(t)| > \frac{1}{2}$, denn $\varphi(0) = 1$ und φ ist stetig. Es gilt also

$$\limsup_{n\to\infty} \sup_{|t|\leq\varepsilon} n\left(1 - |\varphi_n(t)|^2\right) < \infty.$$

Nach Lemma 3.21 $v)$ gilt für jede charakteristische Funktion $\tilde{\varphi}$ eines Wahrscheinlichkeitsmaßes

$$0 \leq 1 - \operatorname{Re}\tilde{\varphi}(2t) \leq 4(1 - \operatorname{Re}\tilde{\varphi}(t)).$$

Damit gilt also auch $\sup_n n\left(1 - |\varphi_n(2t)|^2\right) < \infty$ und

$$|\varphi(2t)|^2 \geq \liminf_{n\to\infty} \exp\left(4n(|\varphi_n(t)|^2 - 1)\right) = \left(|\varphi(t)|^2\right)^4.$$

Also ist $|\varphi(t)| > \frac{1}{2}$ für $|t| \leq \varepsilon$, $|\varphi(t)| > \left(\frac{1}{2}\right)^4$ für $|t| \leq 2\varepsilon, \ldots, |\varphi(t)| > \left(\frac{1}{2}\right)^{4^k}$ für $|t| \leq 2^k\varepsilon$, d.h.

$$|\varphi(t)| \geq \frac{1}{2}\mathbf{1}(|t| \leq \varepsilon) + \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{4^k} \mathbf{1}(\varepsilon 2^{k-1} < |t| \leq \varepsilon 2^k)$$

und dies impliziert (5.2). □

Korollar 5.5. *i) Unter den Voraussetzungen von Lemma 5.4 ist* $\varphi^r = e^{r\psi}$ *für jedes* $r > 0$ *die charakteristische Funktion eines Wahrscheinlichkeitsmaßes. Insbesondere ist* $\varphi = \left(\varphi^{\frac{1}{n}}\right)^n$ *unendlich teilbar.*

ii) Sei $\varphi : \mathbb{R} \to \mathbb{C}$ *stetig in* 0*. Dann ist* φ *genau dann die charakteristische Funktion eines Wahrscheinlichkeitsmaßes und unendlich teilbar, wenn es eine Folge* $(\varphi_n)_n$ *von charakteristischen Funktionen von Wahrscheinlichkeitsmaßen gibt mit* $\varphi_n^n(t) \to \varphi(t)$ *für alle* $t \in \mathbb{R}$*.*

iii) Sei $(\mu_n)_n \subset \mathcal{M}_1(\mathbb{R})$*,* μ_n *unendlich teilbar und* $\mu_n \xrightarrow[n\to\infty]{w} \mu$*. Dann ist auch* μ *unendlich teilbar.*

Beweis. i) Sei φ_n wie in Lemma 5.4 und μ_n mit $\varphi_{\mu_n} = \varphi_n$. Es ist $e^{rn(\varphi_n - 1)}$ die charakteristische Funktion von $\mathrm{CPoi}(rn\mu_n)$ (nach Bsp. und Def. 5.2), also ist

$$\varphi^r = \left(\lim_{n\to\infty} e^{n(\varphi_n - 1)}\right)^r = e^{r\psi}$$

eine CFW (gemäß Lévys Stetigkeitssatz, Satz 3.27).

ii) Sei φ die charakteristische Funktion eines Wahrscheinlichkeitsmaßes und unendlich teilbar. $\varphi_n = \varphi^{\frac{1}{n}}$ leistet das Gewünschte, denn

$$\varphi_n^n(t) = \varphi(t) \to \varphi(t) \quad \text{für alle } t \in \mathbb{R}.$$

Die Rückrichtung folgt aus Lemma 5.4 zusammen mit Teil *i)*.

iii) Sei μ_n unendlich teilbar, d.h. es gibt eine Folge $(\nu_n)_n$ mit $\mu_n = \nu_n^{*n}$. Sei φ_n die zugehörige charakteristische Funktion von ν_n und φ die charakteristische Funktion von μ. Es gilt $\nu_n^{*n} = \mu_n \to \mu$ und somit auch $\varphi_n^n \to \varphi$. Also ist μ unendlich teilbar nach ii).

□

Beweis von Satz 5.3. Sei zunächst $(\nu_n) \subset \mathcal{M}_f(\mathbb{R})$ mit $\mathrm{CPoi}(\nu_n) \xrightarrow{\mathrm{w}} \mu$. Jedes $\mathrm{CPoi}(\nu_n)$ ist unendlich teilbar, also ist auch μ unendlich teilbar nach Korollar 5.5 iii).

Sei nun $\mu \in \mathcal{M}_1(\mathbb{R})$ unendlich teilbar, $\mu = \mu_n^{*n}$ für $n \in \mathbb{N}$, $\varphi_n = \varphi_{\mu_n}$ und $\varphi = \varphi_\mu$. Es ist $e^{n(\varphi_n - 1)} = \varphi_{\mathrm{CPoi}(n\mu_n)}$ und

$$e^{n(\varphi_n(t) - 1)} \xrightarrow[n\to\infty]{} \varphi(t) \quad \text{für alle } t \in \mathbb{R}$$

(gemäß Lemma 5.4), also gilt $\mathrm{CPoi}(n\mu_n) \xrightarrow{\mathrm{w}} \mu$. □

Bemerkung 5.6. Sei $\nu \in \mathcal{M}_f(\mathbb{R})$ mit $\int_{\mathbb{R}} x^2 \,\nu(dx) < \infty$ und $X \sim \mathrm{CPoi}(\nu)$. Dann ist

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \,\nu(dx) \qquad \text{und} \qquad \mathrm{Var}[X] = \int_{\mathbb{R}} x^2 \,\nu(dx).$$

Beweis. Es gilt für die Ableitungen von $\varphi_X(t) = \exp\left(\int_{\mathbb{R}} e^{itx} - 1\,\nu(dx)\right)$:

$$\varphi'_X(t) = \int_{\mathbb{R}} ixe^{itx}\,\nu(dx) \cdot \varphi_X(t),$$

$$\varphi''_X(t) = \int_{\mathbb{R}} -x^2 e^{itx}\,\nu(dx) \cdot \varphi_X(t) + \left(\int_{\mathbb{R}} ixe^{itx}\,\nu(dx)\right)^2 \cdot \varphi_X(t).$$

Also gilt

$$\mathbb{E}[X] = -i\varphi'_X(0) = \int_{\mathbb{R}} x\,\nu(dx),$$

$$\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = -\varphi''_X(0) - \left(\int_{\mathbb{R}} x\,\nu(dx)\right)^2 = \int_{\mathbb{R}} x^2\,\nu(dx).$$

□

Beobachtung 5.7. Sei $b \in \mathbb{R}$, $\sigma^2 \geq 0$ und $\nu_k \in \mathcal{M}_f(\mathbb{R})$ mit

$$\operatorname{supp}(\nu_k) \subset \left(-\frac{1}{k}, -\frac{1}{k+1}\right] \cup \left[\frac{1}{k+1}, \frac{1}{k}\right) =: I_k, \quad k = 0, 1, 2, \ldots$$

wobei $\frac{1}{0} := \infty$. Sei weiter

$$\alpha_k := \int x\,\nu_k(dx), \quad k = 1, 2, \ldots$$

und $Z \sim \mathcal{N}(0,1)$, $X_k \sim \operatorname{CPoi}(\nu_k)$, $k = 0, 1, 2, \ldots$ seien unabhängig. Sei

$$X := b + \sigma Z + X_0 + \sum_{k=1}^{\infty} (X_k - \alpha_k).$$

Sofern $M_n := \sum_{k=1}^{n} (X_k - \alpha_k)$ f.s. konvergiert, ist X unendlich teilbar (nach Korollar 5.5, *iii)*, denn alle „Bauteile" sind unendlich teilbar). Ist

$$\sum_{k=1}^{\infty} \operatorname{Var}[X_k] = \sum_{k=1}^{\infty} \int x^2\,\nu_k(dx) < \infty$$

dann existiert $\lim_{n\to\infty} M_n$, denn $(M_n)_n$ ist ein Martingal mit

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[M_n^2\right] = \sum_{k=1}^{\infty} \operatorname{Var}[X_k] = \sum_{k=1}^{\infty} \int x^2\,\nu_k(dx) < \infty.$$

Demnach: Wenn $\nu = \sum_{k=0}^{\infty} \nu_k$ die Bedingung

$$\int \left(x^2 \wedge 1\right)\,\nu(dx) < \infty$$

erfüllt, so ist X eine unendlich teilbare Zufallsvariable und ihre charakteristische Funktion ist

$$\begin{aligned}
\log\left(\varphi_X(t)\right) &= \log\left(\mathbb{E}\left[e^{itb}\right]\right) + \log\left(\mathbb{E}\left[e^{it\sigma Z}\right]\right) + \log\left(\mathbb{E}\left[e^{itX_0}\right]\right) + \sum_{k=1}^{\infty} \log\left(\mathbb{E}\left[e^{it(X_k-\alpha_k)}\right]\right) \\
&= itb - \frac{1}{2}\sigma^2 t^2 + \int \left(e^{itx} - 1\right)\,\nu_0(dx) + \sum_{k=1}^{\infty} \int \left(e^{itx} - 1 - itx\right)\,\nu_k(dx) \\
&= itb - \frac{1}{2}\sigma^2 t^2 + \int \left(e^{itx} - 1 - \mathbf{1}_{\{|x|<1\}}(x) \cdot itx\right)\,\nu(dx).
\end{aligned}$$

Definition 5.8. Ein σ-endliches Maß ν auf $\mathbb{R}$ mit $\nu(\{0\}) = 0$ und $\int (x^2 \wedge 1)\, \nu(dx) < \infty$ heißt ein *kanonisches Maß*. (b, σ^2, ν) mit $b \in \mathbb{R}$, $\sigma^2 \in [0, \infty)$ und einem kanonischen Maß ν heißt ein *kanonisches Tripel*.

Satz 5.9 (Lévy-Khinchin-Formel[1]). *Sei $\mu \in \mathcal{M}_1(\mathbb{R})$ und $\psi(t) = \log \int e^{itx}\, \mu(dx)$. μ ist genau dann unendlich teilbar, wenn es ein kanonisches Tripel (b, σ^2, ν) gibt mit*

$$\psi(t) = itb - \frac{1}{2}\sigma^2 t^2 + \int \left(e^{itx} - 1 - \mathbf{1}_{\{|x|<1\}}(x) \cdot itx\right) \nu(dx). \tag{5.3}$$

Das kanonische Tripel ist dabei eindeutig festgelegt. ν heißt das Lévy-Maß, *σ^2 der* Gauß'sche Koeffizient, *b die* Zentrierungskonstante.

Beispiel. Für $m \in \mathbb{R}$, $\sigma^2 > 0$ hat $\mathcal{N}(m, \sigma^2)$ charakteristische Funktion $\varphi_{\mathcal{N}(m,\sigma^2)}(t) = \exp(itm - \frac{1}{2}\sigma^2 t)$, d.h. das kanonische Tripel ist $(m, \sigma^2, 0)$.

Bemerkung. Anstelle der Abschneidefunktion $x \cdot \mathbf{1}_{\{|x|<1\}}$ in (5.3) kann prinzipiell jede Funktion $\tilde{f}$ mit $\tilde{f}(x) \sim x$ für $x \to 0$ und $\int |\tilde{f}(x) - x \cdot \mathbf{1}_{\{|x|<1\}}|\, \nu(dx) < \infty$ gewählt werden. In der Literatur üblich sind auch $\tilde{f}(x) = \sin(x)$ (vgl. Taylorentwicklung von $\sin(x)$ bis zur Ordnung 1 ist x) oder $\tilde{f}(x) = \frac{x}{1+x^2}$. In der Lévy-Khinchin-Formel ändert sich dann b zu $\tilde{b} = b + \int \tilde{f}(x) - x \cdot \mathbf{1}_{\{|x|<1\}}\, \nu(dx)$.

Beweis von Satz 5.9. „$\Leftarrow$": Sei ein solches kanonisches Tripel (b, σ^2, ν) gegeben, sodass (5.3) gilt. Dann folgt aus Beobachtung 5.7, dass μ unendlich teilbar ist.

Zeige: ψ legt das kanonische Tripel fest. Sei

$$g_t(x) := e^{itx} - 1 - \mathbf{1}_{\{|x|<1\}} itx. \tag{5.4}$$

Dann ist $\frac{g_t(x)}{t^2(x^2 \wedge 1)}$ gleichmäßig in x und $t \geq 1$ beschränkt und es gilt

$$\frac{g_t(x)}{t^2(x^2 \wedge 1)} \xrightarrow[t \to \infty]{} 0.$$

Damit folgt mit dominierter Konvergenz:

$$\lim_{t \to \infty} \frac{\psi(t)}{t^2} = 0 - \frac{1}{2}\sigma^2 + \lim_{t \to \infty} \int \frac{g_t(x)}{t^2(x^2 \wedge 1)} (x^2 \wedge 1)\, \nu(dx) = -\frac{1}{2}\sigma^2.$$

Also ist σ^2 festgelegt. Sei nun ohne Einschränkung $\sigma^2 = 0$ (sonst gehe über zu $\psi(t) + \frac{1}{2}\sigma^2 t$).

Es ist

$$\begin{aligned} \widetilde{\psi}(t) &:= \psi(t) - \frac{1}{2}\int_{t-1}^{t+1} \psi(s)\, ds \\ &= \int_{\mathbb{R}} e^{itx} \left(1 - \frac{1}{2}\int_{-1}^{1} e^{isx}\, ds\right) \nu(dx) = \int_{\mathbb{R}} e^{itx} h(x)\, \nu(dx) \end{aligned} \tag{5.5}$$

[1] nach Paul Lévy, 1886–1971 und Alexandr Jakovlevich Khinchin, 1894–1959

(in der zweiten Zeile verwende, dass sich die Terme itb und $\mathbf{1}_{\{|x|<1\}}(x) \cdot itx$ jeweils herausheben) mit

$$h(x) = \begin{cases} 1 - \dfrac{\sin x}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases} \tag{5.6}$$

Demnach ist $\widetilde{\psi}$ die charakteristische Funktion von $\widetilde{\nu} := h\nu \in \mathcal{M}_f(\mathbb{R})$ (beachte: $\widetilde{\nu}(dx) = \widetilde{h}(x)(1 \wedge x^2)\nu(dx)$ mit

$$\widetilde{h}(x) := \begin{cases} \frac{h(x)}{1 \wedge x^2}, & x \neq 0 \\ \frac{1}{6}, & x = 0 \end{cases}$$

und $\widetilde{h}$ ist stetig, beschränkt und strikt positiv auf $\mathbb{R}$).

Also ist $\widetilde{\nu}$ und somit auch ν durch ψ festgelegt und damit schließlich auch b.

„$\Rightarrow$": Sei nun μ unendlich teilbar und $\psi = \log \varphi_\mu$. Dann ist

$$\mathrm{Re}(\psi) \leq 0$$

(denn $|\varphi_\mu(t)| \leq 1$) und

$$t \mapsto \mathrm{Im}\,(\psi(t)) \quad \text{ist ungerade.}$$

Daher ist $\widetilde{\psi}(0)$ reell und $\widetilde{\psi}(0) = \psi(0) - \frac{1}{2}\int_{-1}^{1} \mathrm{Re}(\psi(u))\,du \geq 0$ (für $\widetilde{\psi}$ aus (5.5)).

Falls $\widetilde{\psi}(0) = 0$, so ist

$$\mu = \delta_b \quad \text{für ein } b \in \mathbb{R},$$

denn dann ist $\mathrm{Re}(\psi(t)) = 0$ für alle $t \in [-1, 1]$, also $|\varphi_\mu(t)| = 1$ für alle $t \in [-1, 1]$ und falls μ nicht trivial wäre, so wäre

$$1 = \left| \int e^{itx}\,\mu(dx) \right| < \int \left|e^{itx}\right|\,\mu(dx) = 1$$

für (zumindest einige) $t \in [-1, 1]$ (eine nicht-triviale Konvexkombination von komplexen Zahlen vom Betrag 1 liegt strikt im Inneren des Einheitskreises), ein Widerspruch. δ_b hat kanonisches Tripel $(b, 0, 0)$.

Sei $\widetilde{\psi}(0) > 0$. Wähle gemäß Satz 5.3 eine Folge $(\nu_n)_n \subset \mathcal{M}_f(\mathbb{R} \setminus \{0\})$ mit $\mathrm{CPoi}(\nu_n) \xrightarrow{\mathrm{w}} \mu$. Setze

$$\begin{aligned} b_n &:= \int x \cdot \mathbf{1}_{(-1,1)}(x)\,\nu_n(dx), \\ \psi_n(t) &:= \log \varphi_{\mathrm{CPoi}(\nu_n)}(t) = \int \left(e^{itx} - 1\right)\,\nu_n(dx) = \int g_t\,d\nu_n + itb_n \end{aligned}$$

(mit g_t aus (5.4)),

$$\widetilde{\psi}_n(t) := \psi_n(t) - \frac{1}{2}\int_{t-1}^{t+1} \psi_n(s)\,ds = \int e^{itx} h(x)\,\nu_n(dx).$$

Dann gilt nach Lévys Stetigkeitssatz (Satz 3.27) $\psi_n \to \psi$ lokal gleichmäßig und ψ ist stetig, insbesondere gilt

$$\widetilde{\psi}_n(t) \underset{n\to\infty}{\longrightarrow} \widetilde{\psi}(t) \quad \text{für alle } t \in \mathbb{R}$$

und $\widetilde{\psi}_n(0) > 0$ für n genügend groß.

Somit ist (mit h aus (5.6))

$$\widetilde{\nu}_n(dx) = \frac{1}{\widetilde{\psi}_n(0)} h(x)\nu_n(dx) \in \mathcal{M}_1(\mathbb{R})$$

(für n genügend groß) und

$$\frac{\widetilde{\psi}(t)}{\widetilde{\psi}(0)} = \lim_{n\to\infty} \int e^{itx}\,\widetilde{\nu}_n(dx)$$

ist die charakteristische Funktion eines Wahrscheinlichkeitsmaßes. Es gibt also ein

$$\widetilde{\nu} \in \mathcal{M}_1(\mathbb{R}) \text{ mit } \widetilde{\nu}_n \xrightarrow{\text{w}} \widetilde{\nu} \quad \text{und} \quad \widetilde{\psi}(t) = \widetilde{\psi}(0) \int e^{itx}\,\widetilde{\nu}(dx).$$

Sei $\sigma^2 := 6\widetilde{\psi}(0)\widetilde{\nu}(\{0\})$, $\nu(dx) := \frac{\widetilde{\psi}(0)}{h(x)} \cdot \mathbf{1}_{\{x\neq 0\}}\widetilde{\nu}(dx)$ (dies ist ein kanonisches Maß) und

$$f_t : \mathbb{R} \to \mathbb{R},\ x \mapsto \begin{cases} \dfrac{g_t(x)}{h(x)}, & x \neq 0 \\ -3t^2, & x = 0 \end{cases}$$

(mit g_t aus (5.4), h aus (5.6)). f_t ist stetig und beschränkt, daher gilt

$$\int f_t\, d\widetilde{\nu}_n \xrightarrow[n\to\infty]{} \int f_t\, d\widetilde{\nu} = \frac{1}{\widetilde{\psi}(0)}\Big(-\frac{1}{2}\sigma^2 t^2 + \int g_t\, d\nu\Big)$$

und damit

$$\psi(t) = \lim_{n\to\infty} \psi_n(t) = \lim_{n\to\infty} \left(\widetilde{\psi}_n(0) \int f_t\, d\widetilde{\nu}_n + itb_n\right) = -\frac{1}{2}\sigma^2 t^2 + \int g_t\, d\nu + \lim_{n\to\infty} itb_n.$$

Also existiert $b := \lim_{n\to\infty} b_n$ und es gilt

$$\psi(t) = itb - \frac{1}{2}\sigma^2 t^2 + \int g_t\, d\nu.$$

$\square$

Beobachtung 5.10. Sei μ ein unendlich teilbares Wahrscheinlichkeitsmaß auf $\mathbb{R}$ mit zugehörigem kanonischem Maß ν. Dann ist $\nu := \text{v} - \lim_{n\to\infty} n\mu^{\frac{1}{n}}\,|_{\mathbb{R}\setminus\{0\}}$.

Beweisskizze. μ habe kanonische Tripel (b, σ^2, ν), dann hat $\mu^{*1/n}$ kanonisches Tripel $(b/n, \sigma^2/n, \nu/n)$.

Stelle wie in Beob 5.7 $X^{(n)} \sim \mu^{*1/n}$ dar als

$$X^{(n)} = \frac{b}{n} + \frac{\sigma}{\sqrt{n}} Z + X_0^{(n)} + \sum_{k=1}^{\infty} (X_k^{(n)} - \alpha_k^{(n)})$$

mit unabhängigen $Z \sim \mathcal{N}(0,1)$, $X_k^{(n)} \sim \text{CPoi}(\frac{1}{n}\nu|_{I_K})$ ($I_k = (-\frac{1}{k}, -\frac{1}{k+1}] \cup [\frac{1}{k+1}, \frac{1}{k})$, $\alpha_k = \int_{I_k} x\, \frac{1}{n}\nu(dx)$).

Sei $A = [a, b] \subset \mathbb{R}$ ein kompaktes Intervall mit $d(0, A) > 0$ und $\nu(\partial A) = 0$. Man kann zeigen, dass man für $\mathbb{P}(X^{(n)} \in A)$ bei geeignet gewähltem m die Beiträge von $\frac{b}{n}$, $\frac{\sigma}{\sqrt{n}} Z$, $\sum_{k=m+1}^{\infty} (X_k^{(n)} - \alpha_k)$ und von $\sum_{k=1}^{m} \alpha_k^{(n)}$ mit Limes $n \to \infty$ (nahezu) vernachlässigen kann: Zu $\varepsilon_1, \varepsilon_2 > 0$ kann man m und n_0 so groß wählen, dass

$$\mathbb{P}\Big(\Big|\frac{b}{n} + \frac{\sigma}{\sqrt{n}} Z + \sum_{k=m+1}^{\infty} (X_k^{(n)} - \alpha_k^{(n)}) + \sum_{k=1}^{m} \alpha_k^{(n)}\Big| > \varepsilon_1\Big) \le \frac{\varepsilon_2}{n} \quad \text{für } n \ge n_0.$$

Ersetzen wir also $X^{(n)}$ durch

$$Y := \sum_{k=0}^{\infty} X_k^{(n)} \sim \mathrm{CPoi}(\nu|_S)$$

mit $S := (-\infty, \frac{1}{m}] \cup [\frac{1}{m}, \infty)$ (und wir wählen m so groß, dass $A \subset S$).

Somit ist für n gen. groß

$$\mathbb{P}\big(Y \in (a + \varepsilon_1, b - \varepsilon_1)\big) - \frac{\varepsilon_2}{n} \le \mathbb{P}(X^{(n)} \in A) \le \mathbb{P}\big(Y \in (a - \varepsilon_1, b + \varepsilon_1)\big) + \frac{\varepsilon_2}{n}.$$

Mit der Darstellung (vg. Bsp. und Def. 5.2)

$$Y \overset{\mathrm{d}}{=} \sum_{j=1}^{N} \widetilde{Y}_j \quad \text{mit } N \sim \mathrm{Poi}(\tfrac{1}{n}\nu(S)),\ \widetilde{Y}_1, \widetilde{Y}_2, \dots \sim \frac{\nu(\cdot \cap S)}{\nu(s)} \ \text{ u.a.}$$

ist

$$\begin{aligned} n\mathbb{P}(Y \in A) &= n\mathbb{P}(N = 1, Y_1 \in A) + n\mathbb{P}\big(N \ge 2, \textstyle\sum_{j=1}^{N} \widetilde{Y}_j \in A\big) \\ &= ne^{-\nu(S)/n} \frac{1}{n} \nu(S) \frac{\nu(A)}{\nu(S)} + O\big(n\mathbb{P}(N \ge 2)\big) \\ &= e^{-\nu(S)/n} \nu(A) + O\Big(\frac{1}{n}\Big) \xrightarrow[n\to\infty]{} \nu(A). \end{aligned}$$

□

Beispiel 5.11. *i)* $\mathrm{Gamma}(r, \lambda)$ mit Formparameter r und Skalenparameter λ hat Dichte $f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x)$ und charakteristische Funktion $\varphi(t) = \frac{1}{(1-it/\lambda)^r}$, erfüllt $\mathrm{Gamma}(r, \lambda) = \mathrm{Gamma}\left(\frac{r}{n}, \lambda\right)^{*n}$.

Somit ist für $A \subset \mathbb{R} \setminus \{0\}$ kompakt

$$\lim_{n\to\infty} n\mu^{\frac{1}{n}}|_{\mathbb{R}\setminus\{0\}}(A) = \lim_{n\to\infty} n \int_A \frac{\lambda^{r/n}}{\Gamma(r/n)} x^{r/n - 1} e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x)\, dx = \int_A \frac{r}{x} e^{-\lambda x}\, dx$$

(verwende $\Gamma(x) \sim 1/x$ für $x \to 0$, was z.B. aus $\Gamma(x+1) = x\Gamma(x)$ folgt), d.h. das kanonische Maß ist $\nu(dx) = \frac{r}{x} e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x)\, dx$.

Weiter ist $\sigma^2 = -2 \lim_{t\to\infty} \frac{1}{t^2} \log \varphi_{\mathrm{Gamma}(r,\lambda)}(t) = 0$ und $b = 0$ (da $\mathrm{supp}(\mathrm{Gamma}(r, \lambda)) = [0, \infty)$).

ii) Sei $a > 0$. $\mathrm{Cauchy}(a)$ hat Dichte $f(x) = \frac{1}{a\pi} \frac{1}{1+(\frac{x}{a})^2}$ und charakteristische Funktion $\varphi_\nu(t) = e^{-a|t|}$. Sei $A \subset \mathbb{R} \setminus (-\varepsilon, \varepsilon)$ und betrachte $a = 1$:

$$n\mathrm{Cauchy}\Big(\frac{1}{n}\Big)(A) = \frac{1}{\pi} \int_A \frac{n^2}{1 + n^2 x^2}\, dx \xrightarrow[n\to\infty]{} \frac{1}{\pi} \int_A \frac{1}{x^2}\, dx.$$

Folglich ist $\nu = \mathbf{1}_{\{x \neq 0\}} \frac{1}{x^2}\, dx$, $\sigma^2 = -2 \lim_{t\to\infty} \frac{1}{t^2} \log \varphi_{\mathrm{Cauchy}(1)}(t) = 0$ und $b = 0$ aus Symmetrie.

5.1 Ein Bericht über stabile Verteilungen

Beispiel 5.12. Sei $\alpha \in (0,2)$ und $\nu_\alpha(dx) := \frac{1}{\theta_\alpha}\mathbf{1}_{\{x\neq 0\}}|x|^{-\alpha-1}\,dx$ mit

$$\theta_\alpha := \int_{\mathbb{R}\setminus\{0\}} (1-\cos(x))|x|^{-\alpha-1}\,dx = \begin{cases} -2\Gamma(-\alpha)\cos(\frac{\alpha\pi}{2}), & \alpha \neq 1 \\ \pi, & \alpha = 1 \end{cases}$$

(beachte $0 \le 1-\cos(x) \le x^2 \wedge 2$, d.h. das Integral ist endlich, vgl. z.B. [Fe, Kap. XVII, S. 568-569] für den expliziten Wert). ν_α ist ein kanonisches Maß (denn $\int_0^\infty (x^2 \wedge 1)x^{-1-\alpha}\,dx < \infty$ für $\alpha \in (0,2)$).

Sei μ_α die unendlich teilbare Verteilung mit kanonischem Tripel $(0,0,\nu_\alpha)$. μ_α heißt *(standard-) symmetrisch stabile Verteilung von Index* α. Gemäß Lévy-Khinchin-Formel ist

$$\psi_{\mu_\alpha}(t) = \log\varphi_{\mu_\alpha}(t) = \int_{\mathbb{R}\setminus\{0\}} \left(e^{itx} - 1 - itx\mathbf{1}_{\{|x|<1\}}\right)\frac{1}{\theta_\alpha|x|^{\alpha+1}}\,dx.$$

Der Imaginärteil des Integrals ist aus Symmetriegründen $= 0$, daher folgt

$$\begin{aligned}\psi_{\mu_\alpha}(t) &= \int_{\mathbb{R}\setminus\{0\}} \left(e^{itx} - 1 - itx\mathbf{1}_{\{|x|<1\}}\right)\frac{1}{\theta_\alpha|x|^{\alpha+1}}\,dx \\ &= -\frac{1}{\theta_\alpha}\int_{\mathbb{R}\setminus\{0\}} (1-\cos(tx))\,|x|^{-\alpha-1}\,dx \\ &= -\frac{1}{\theta_\alpha}\int_{\mathbb{R}\setminus\{0\}} (1-\cos(y))\,|y|^{-\alpha-1}|t|^{\alpha+1}\frac{1}{|t|}\,dy = -|t|^\alpha\end{aligned}$$

(wir substituieren $tx = y$ für das dritte Gleichheitszeichen).

Insbesondere gilt also: Sind $X_1,\dots,X_n \sim \mu_\alpha$ unabhängig, so ist

$$\frac{X_1+\cdots+X_n}{n^{\frac{1}{\alpha}}} \overset{\mathrm{d}}{=} X_1, \tag{5.7}$$

denn $\left(\varphi_{\mu_\alpha}\left(\frac{t}{n^{1/\alpha}}\right)\right)^n = \left(e^{-\frac{|t|^\alpha}{n}}\right)^n = \varphi_{\mu_\alpha}(t)$.

Bemerkung. Die analoge Verteilungsidentität (5.7) für $\alpha = 2$ ist für $X_i \sim \mathcal{N}(0,\sigma^2)$ erfüllt, in diesem Sinne sind die (zentrierten) Normalverteilungen die symmetrisch stabilen Verteilungen von Index 2.

Definition 5.13. Sei $\mu \in \mathcal{M}_1(\mathbb{R})$ und seien $X_1,\dots,X_n \sim \mu$ unabhängig. μ heißt *(strikt) stabil mit Index* $\alpha \in (0,2)$, wenn

$$\frac{X_1+\cdots+X_n}{n^{\frac{1}{\alpha}}} \overset{\mathrm{d}}{=} X_1.$$

μ heißt *stabil mit Index* α (auch *im weiteren Sinne stabil*), falls es $b_n \in \mathbb{R}$ gibt, sodass

$$\frac{X_1+\cdots+X_n-b_n}{n^{\frac{1}{\alpha}}} \overset{\mathrm{d}}{=} X_1.$$

Beobachtung 5.14. Sei X unendlich teilbar mit kanonischem Tripel (b,σ^2,ν) und sei $a > 0$. Dann gilt:

$$\begin{aligned}\log\varphi_{aX}(t) = \log\varphi_X(at) &= ibat - \frac{1}{2}\sigma^2a^2t^2 + \int e^{itax} - 1 - \mathbf{1}_{\{|x|<1\}}itax\,\nu(dx) \\ &= ibat - \frac{1}{2}\sigma^2a^2t^2 + \int e^{itax} - 1 - \mathbf{1}_{\{|ax|<1\}}itax\,\nu(dx) + it\int ax\left(\mathbf{1}_{\{|ax|<1\}} - \mathbf{1}_{\{|x|<1\}}\right)\nu(dx).\end{aligned}$$

Also hat aX das kanonische Tripel

$$\left(ab + a \int x \left(\mathbf{1}_{\{|ax|<1\}} - \mathbf{1}_{\{|x|<1\}}\right) \nu(dx),\ \sigma^2 a^2,\ \nu \circ f_a^{-1}\right)$$

mit $f_a(x) = ax$.

Satz 5.15. *X ist genau dann stabil mit Index $\alpha \in (0,2)$, wenn es $b \in \mathbb{R}$ und $c_+, c_- \geq 0$ gibt, sodass X das kanonische Tripel $(b, 0, \nu)$ besitzt mit*

$$\nu(dx) = c_- \mathbf{1}_{\{x<0\}} |x|^{-\alpha-1}\,dx + c_+ \mathbf{1}_{\{x>0\}} |x|^{-\alpha-1}\,dx. \tag{5.8}$$

Bericht 5.16. Stabile Verteilungen von Index $\alpha = 2$ sind die Normalverteilungen, es gibt keine stabile Verteilungen von Index $\alpha > 2$.

Beweisksizze für Satz 5.15. X habe kanonisches Tripel $(b, 0, \nu)$ mit ν aus (5.8), $X_1, \ldots, X_n$ seien u.a. Kopien von X. Dann hat

$$X_1 + \cdots + X_n$$

kanonisches Tripel $(nb, 0, n\nu)$ und mit $\widetilde{b}_n := nb - bn^{1/\alpha} - n^{1/\alpha} \int_{n^{-1/\alpha} \leq |x| < 1} x\, \nu(dx)$ hat

$$n^{1/\alpha} X + \widetilde{b}_n$$

kanonisches Tripel $(nb, 0, \nu \circ f_{n^{1/\alpha}}^{-1})$ (vgl. Beob. 5.14).

Es ist $\nu \circ f_{n^{1/\alpha}}^{-1} = n\nu$: Betrachte z.B. $a > 0$, so ist

$$\begin{aligned}
(\nu \circ f_{n^{1/\alpha}}^{-1})\big([a,\infty)\big) &= \nu\big([n^{-1/\alpha}a, \infty)\big) = c_+ \int_{n^{-1/\alpha}a}^{\infty} x^{-\alpha-1}\,dx \\
&= nc_+ \int_a^\infty y^{-\alpha-1}\,dy = n\nu([a,\infty))
\end{aligned}$$

mit Substitution $y = n^{1/\alpha}x$. Somit gilt

$$X_1 + \cdots + X_n \overset{\mathrm{d}}{=} n^{1/\alpha} X + \widetilde{b}_n.$$

(Wir sehen auch: Für strikte Stabilität muss $\widetilde{b}_n = 0$ sein, dies führt für $\alpha \neq 1$ zur Bedingung $b = (c_+ - c_-)/(\alpha - 1)$.)

Sei andererseits X stabil mit Index $\alpha \in (0,2)$ (also insbesondere X unendlich teilbar), dann muss analog zu oben das zugehörige Lévy-Maß ν die Bedingung

$$\nu \circ f_{n^{1/\alpha}}^{-1} = n\nu \quad \text{für jedes } n \in \mathbb{N}$$

erfüllen. Setze

$$h_+(x) := \nu([x,\infty)), \quad h_-(x) := \nu((-\infty, -x]), \qquad x > 0,$$

diese erfüllen

$$nh_\pm(x) = h_\pm(n^{-1/\alpha}x), \ x > 0, n \in \mathbb{N}.$$

Mit der Wahl $x = (n/j)^{1/\alpha}$ mit $n, j \in \mathbb{N}$ (und dann $n = j$ für die zweite Gleichung) ergibt sich

$$nh_\pm\big((n/j)^{1/\alpha}\big) = h_\pm(j^{-1/\alpha}) = jh_\pm(1),$$

d.h.

$$h_\pm(x) = x^{-\alpha}h_\pm(1) \quad \text{für } x = (n/j)^{1/\alpha} \text{ mit } n, j \in \mathbb{N}.$$

Da $h_\pm$ monoton fallend ist, gilt $h_\pm(x) = x^{-\alpha}h_\pm(1)$ für beliebiges $x > 0$. Somit muss ν die Form (5.8) haben. □

Bericht 5.17. Die stabilen X aus Satz 5.15 haben

$$\log \varphi_X(t) = \begin{cases} ict - d|t|^\alpha \left(1 + i\theta \operatorname{sgn}(t) \tan\left(\alpha \frac{\pi}{2}\right)\right), & \alpha \neq 1 \\ ict - d|t| \left(1 + i\theta \operatorname{sgn}(t) \frac{2}{\pi} \log(|t|)\right), & \alpha = 1 \end{cases}$$

mit $c \in \mathbb{R}$, $d > 0$, $\theta = \frac{c_+ - c_-}{c_+ + c_-} \in [-1, 1]$ (vgl. [Br, Theorem 9.32]).

Bericht 5.18. Ist X stabil mit Index α, so ist

$$\mathbb{E}\left[|X|^\beta\right] \begin{cases} < \infty, & 0 \le \beta < \alpha \\ = \infty, & \beta \ge \alpha \end{cases}$$

Bericht 5.19 (Stabile Analoga zum Zentralen Grenzwertsatz)**.** Seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte, reelle Zufallsvariablen mit Verteilungsfunktion F.

i) Gibt es Konstanten a_n, b_n, für die

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \xrightarrow[n\to\infty]{\mathrm{d}} Y \tag{5.9}$$

gilt, so ist Y stabil.

ii) Genau dann existieren $(a_n)_n$ und $(b_n)_n$ mit (5.9), sodass Y stabil ist mit Index $\alpha \in (0, 2)$, wenn es $c_+, c_- \ge 0$ gibt mit $c_+ + c_- > 0$ und

a) $\lim\limits_{x\to\infty} \dfrac{F(-x)}{1 - F(x)} = \dfrac{c_-}{c_+}$.

b) Falls $c_+ > 0$, so gilt für alle $\xi > 0$: $\lim\limits_{x\to\infty} \dfrac{\mathbb{P}(X > \xi x)}{\mathbb{P}(X > x)} = \dfrac{1}{\xi^\alpha}$, und falls $c_- > 0$, so gilt für alle $\xi > 0$: $\lim\limits_{x\to\infty} \dfrac{\mathbb{P}(X < -\xi x)}{\mathbb{P}(X < -x)} = \dfrac{1}{\xi^\alpha}$.

Bemerkung. Ein „Paradebeispiel" für die Bedingung ii) aus Bericht 5.19 ist eine Zufallsvariable X mit Dichte

$$f(x) = c_1 \mathbf{1}_{\{x \ge 1\}} x^{-\alpha-1} + c_2 \mathbf{1}_{\{x \le -1\}} (-x)^{-\alpha-1}.$$

Kapitel 6

Markovprozesse

6.1 Grundlegendes: Stochastische Kerne, projektive Familien

Definition 6.1. Seien $(S_1, \mathscr{A}_1)$ und $(S_2, \mathscr{A}_2)$ messbare Räume. $\kappa\colon S_1 \times \mathscr{A}_2 \to [0,1]$ heißt *stochastischer Kern* (auch *Markov-Kern*) (*von* (S_1, A_1) *nach* $(S_2, \mathscr{A}_2)$), falls gilt

i) Für alle $A_2 \in \mathscr{A}_2$ gilt: $S_1 \ni x \mapsto \kappa(x, A_2)$ ist $(\mathscr{A}_1 - \mathscr{B}(\mathbb{R}))$-messbar.

ii) Für alle $x \in S_1$ gilt: $\kappa(x, \cdot\,) \in \mathcal{M}_1(S_2)$.

κ heißt *substochastisch*, wenn in $ii)$ gefordert wird, dass $\kappa(x, \cdot\,) \in \mathcal{M}_{\leq 1}(S_2)$.

Beispiel 6.2. i) Sei $S_1 = S_2 = S$ höchstens abzählbar, $\mathscr{A}_1 = \mathscr{A}_2 = 2^S$ und $(p_{xy})_{x,y \in S}$ eine stochastische Matrix. Dann ist $\kappa(x, A) := \sum_{y \in A} p_{xy}$ ein stochastischer Kern (von S nach S).

ii) Sei $S_1 = S_2 = \mathbb{R}$, $\mathscr{A}_1 = \mathscr{A}_2 = \mathscr{B}(\mathbb{R})$ und $\nu \in \mathcal{M}_1(\mathbb{R})$. Dann ist $\kappa(x, A) := (\delta_x * \nu)(A) = \nu(A - x)$ ein stochastischer Kern.

(Interpretation: $\kappa(x, \cdot)$ beschreibt einen zufälligen Sprung gemäß ν von x aus.)

Erinnerung (Produkt σ-Algebra). Seien $(S_i, \mathscr{A}_i)$, $i \in I$ messbare Räume und sei $S = \bigtimes_{i \in I} S_i$. Dann ist $\mathscr{A} := \bigotimes_{i \in I} \mathscr{A}_i$ die kleinste σ-Algebra, bezüglich der alle kanonischen Projektionen $\pi_i\colon S \to S_i$ messbar sind.

Bemerkung und Definition 6.3. i) Seien $(S_i, \mathscr{A}_i)$, $i = 0, 1, 2$ messbare Räume, κ_1 ein stochastischer Kern von S_0 nach S_1 und κ_2 ein stochastischer Kern von $S_0 \times S_1$ nach S_2. Dann ist

$$\begin{aligned} \kappa_1 \otimes \kappa_2 : S_0 \times (\mathscr{A}_1 \otimes \mathscr{A}_2) &\to [0,1] \\ (x_0, A) &\mapsto \int_{S_1} \int_{S_2} \mathbf{1}_A(x_1, x_2)\, \kappa_2((x_0, x_1),\, dx_2)\, \kappa_1(x_0,\, dx_1) \end{aligned}$$

ein stochastischer Kern von S_0 nach $S_1 \times S_2$ („Produkt von κ_1 und κ_2“).

ii) Seien $(S_i, \mathscr{A}_i)$, $i = 0, 1, 2$ messbare Räume, κ_1 ein stochastischer Kern von S_0 nach S_1 und κ_2 ein stochastischer Kern von S_1 nach S_2. Dann ist

$$\begin{aligned} \kappa_1 \circ \kappa_2 : S_0 \times \mathscr{A}_2 &\to [0,1] \\ (x, A) &\mapsto \int_{S_1} \kappa_2(y, A)\, \kappa_1(x, dy) \end{aligned}$$

ein stochastischer Kern von S_0 nach S_2 („Verkettung von κ_1 und κ_2“).

Definition 6.4. Ein messbarer Raum $(S, \mathscr{A})$ heißt *Borel-Raum* (oder auch *Standard-Borel-Raum*), wenn es ein $B \in \mathscr{B}(\mathbb{R})$ und eine Bijektion $\varphi : S \to B$ gibt, sodass φ und φ^{-1} messbar sind.

Satz 6.5. *Jeder polnische Raum (ausgestattet mit seiner Borel-σ-Algebra) ist ein Borel-Raum.*

Beweisskizze. Betrachte zunächst $[0,1]^\infty$ mit Metrik

$$d\left((x_i), (y_i)\right) := \sum_{i=1}^{\infty} 2^{-i} |x_i - y_i|.$$

Für $x = (x_1, x_2, \dots) \in [0,1]^\infty$ sei

$$x_i = \sum_{j=1}^{\infty} 2^{-j} x_{i,j} \quad \text{mit } x_{i,j} \in \{0, 1\}$$

die (eindeutige, nicht-abbrechende) Binärdarstellung. Sei $(a(n), b(n))_{n \in \mathbb{N}}$ eine Aufzählung von $\mathbb{N}^2$ und setze

$$\psi(x) := \sum_{n=1}^{\infty} 2^{-n} x_{a(n), b(n)}.$$

ψ ist bijektiv und bi-messbar (beachte: $y \mapsto k$-te Ziffer der Binärentwicklung von y ist messbar, sie ist stückweise konstant). Also ist $[0,1]^\infty$ ein Borel-Raum.

Zeige: E ein polnischer Raum (mit Metrik d),

$$\text{so gibt es } S \subset [0,1]^\infty \text{ messbar und eine bi-messbare Bijektion } \varphi : E \to S. \tag{6.1}$$

Sei dazu $x_1, x_2, \dots$ eine dichte Folge in E und definiere

$$\varphi(x) := \left(d(x, x_1) \wedge 1,\ d(x, x_2) \wedge 1, \dots\right) \in [0,1]^\infty.$$

Es gilt

$$y_n \xrightarrow[n\to\infty]{} y \text{ in } E \quad \iff \quad d(y_n, x_m) \wedge 1 \xrightarrow[n\to\infty]{} d(y, x_m) \wedge 1 \text{ für alle } m \in \mathbb{N}.$$

Sei dazu zunächst $y_n \to y$. Für alle $m \in \mathbb{N}$ ist $d(\,\cdot\,, x_m)$ stetig, also gilt $d(y_n, x_m) \wedge 1 \to d(y, x_m) \wedge 1$ für alle $m \in \mathbb{N}$.

Sei umgekehrt $d(y_n, x_m) \wedge 1 \to d(y, x_m) \wedge 1$ für alle $m \in \mathbb{N}$. Es gilt $d(y_n, y) \wedge 1 \leq (d(y_n, x_m) \wedge 1) + (d(y, x_m) \wedge 1)$, also ist $\limsup_{n\to\infty} (d(y_n, y) \wedge 1) \leq 2\,(d(y, x_m) \wedge 1)$. Wähle nun eine Folge $x_{m_k} \to y$ und erhalte $\lim_{n\to\infty} d(y_n, y) = 0$.

Demnach ist also φ stetig und injektiv und $\varphi^{-1}: \varphi(E) \to E$ ist stetig.

Zeige:

$$S := \varphi(E) \subset [0,1]^\infty \quad \text{ist messbar.} \tag{6.2}$$

Sei dazu

$$U_n := \left\{(x_i) \in \overline{S} \mid \exists\, V \subset [0,1]^\infty \text{ offen, } x \in V \text{ und } \operatorname{diam}(\varphi^{-1}(V \cap S)) < \frac{1}{n}\right\}.$$

Es ist $S \subset U_n$, denn φ^{-1} ist stetig auf $S = \varphi(E)$, und U_n ist relativ offen in $\overline{S}$.

Zeige

$$\bigcap_n U_n \subset S.$$

Sei $x \in \bigcap_n U_n \subset \overline{S}$. Zu $n \in \mathbb{N}$ wähle

$$V_n \subset [0,1]^\infty \text{ offen mit } x_n \in V_n \subset V_{n-1} \text{ und } \operatorname{diam}(\varphi^{-1}(V \cap S)) < \frac{1}{n}$$
$$\text{sowie } x_n' \in V_n \cap S \text{ mit } x_n' \to x.$$

$y_n' := \varphi^{-1}(x_n')$ ist Cauchyfolge in E. Da E vollständig ist, existiert ein $y \in E$ mit $y_n' \to y$. Wegen Stetigkeit von φ gilt

$$\varphi(y) = \lim_{n\to\infty} \varphi(\varphi^{-1}(x_n')) = \lim_{n\to\infty} x_n' = x,$$

also ist $x \in S = \varphi(E)$.

Damit ist

$$S = \bigcap_n U_n \quad \text{messbar,}$$

denn U_n ist relativ offen in $\overline{S}$ und damit messbar, d.h. (6.2) gilt. Insgesamt folgt (6.1).

Mehr Details finden sich z.B. in [RW, Ch. II.82] oder [Br, Appendix 7]. □

Satz 6.6 (Existenz einer regulären Version der bedingten Wahrscheinlichkeiten für Borel-Wertebereiche, vergl. auch Bericht A.9)**.** *Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathscr{G} \subset \mathscr{F}$ eine Teil-σ-Algebra und X eine Zufallsvariable mit Werten im Borel-Raum $(S, \mathscr{A})$. Dann gibt es einen stochastischen Kern κ von $(\Omega, \mathscr{G})$ nach $(S, \mathscr{A})$ mit*

$$\kappa(\omega, B) = \mathbb{E}\left[\mathbf{1}_B(X) \mid \mathscr{G}\right](\omega) \quad \mathbb{P}-\text{f.s.}$$

für jedes $B \in \mathscr{A}$.

Beweis. Sei ohne Einschränkung $S \subset \mathbb{R}$ Borel-messbar, sonst wähle $\varphi: S \to S' \subset \mathbb{R}$ bijektiv und bi-messbar und betrachte $X' := \varphi \circ X$.

Für $r \in \mathbb{Q}$ sei $F_r := \mathbb{E}\left[\mathbf{1}_{(-\infty, r]}(X) \mid \mathscr{G}\right]$. Für $r, r' \in \mathbb{Q}$ mit $r \le r'$ gilt

$$F_r \le F_{r'}, \qquad \lim_{n\to\infty} F_n = 1, \qquad \lim_{n\to\infty} F_{-n} = 0, \qquad \lim_{n\to\infty} F_{r+\frac{1}{n}} = F_r \tag{6.3}$$

$\mathbb{P}$-fast sicher, d.h. es gibt ein $N \in \mathscr{F}$ mit $\mathbb{P}(N) = 0$, sodass die Ereignisse aus (6.3) auf $\Omega \setminus N$ gelten.

Setze

$$\widetilde{F}_s(\omega) := \begin{cases} \inf\{F_r(\omega) \mid r \geq s,\ r \in \mathbb{Q}\}, & \omega \in \Omega \setminus N \\ \mathbf{1}_{\{s\geq 0\}}, & \omega \in N \end{cases}$$

$\widetilde{F}_\cdot$ ist für jedes $\omega \in \Omega$ die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes auf $\mathbb{R}$. Sei $\kappa(\omega, \cdot)$ das zu ω gehörige Wahrscheinlichkeitsmaß. Für $r \in \mathbb{Q}$ ist

$$\kappa(\omega, (-\infty, r]) = F_r \quad \mathscr{G}\text{-messbar nach Konstruktion.}$$

Demnach ist

$$\omega \mapsto \kappa(\omega, B) \quad \mathscr{G}\text{-messbar für alle } B \in \mathscr{B}(\mathbb{R}),$$

denn die Menge aller B mit dieser Eigenschaft ist ein Dynkin-System, das den schnittstabilen Erzeuger $\{(-\infty, r] \mid r \in \mathbb{Q}\}$ von $\mathscr{B}(\mathbb{R})$ umfasst (vgl. [Kl, Satz 1.19] oder [St1, Satz 1.14]).

Zeige nun, dass

$$\kappa(\cdot, B) \text{ für jedes } B \in \mathscr{B}(\mathbb{R}) \text{ eine Version von } \mathbb{E}[\mathbf{1}_B(X) \mid \mathscr{G}] \text{ ist.}$$

Sei dazu $A \in \mathscr{G}$,

$$\nu_1(B) := \mathbb{E}[\mathbf{1}_A \kappa(\cdot, B)] \quad \text{und} \quad \nu_2(B) := \mathbb{E}[\mathbf{1}_A \mathbf{1}_B(X)].$$

ν_1 und ν_2 sind endliche Maße auf $\mathbb{R}$ und es gilt $\nu_1((-\infty, r]) = \nu_2((-\infty, r])$ für alle $r \in \mathbb{Q}$ nach Konstruktion. Nach dem Eindeutigkeitssatz für Maße (vgl. [Kl, Lemma 1.42] oder [St1, Satz 1.18]) folgt damit $\nu_1(B) = \nu_2(B)$ für alle $B \in \mathscr{B}(\mathbb{R})$, also gilt

$$\mathbb{E}[\mathbf{1}_A \kappa(\cdot, B)] = \mathbb{E}[\mathbf{1}_A \mathbf{1}_B(X)]$$

für jedes $A \in \mathscr{G}$. □

Korollar 6.7. *In der Situation von Satz 6.6 sei $\mathscr{G} = \sigma(Y)$ für eine Zufallsvariable Y mit Werten im messbaren Raum $(S', \mathscr{A}')$. Dann gibt es einen stochastischen Kern κ' von S' nach S mit*

$$\kappa'(Y, B) = \mathbb{E}[\mathbf{1}_B(X) \mid Y] \quad \textit{fast sicher für alle } B \in \mathscr{A}.$$

Erinnerung 6.8 (Faktorisierungslemma). Sei Y eine Zufallsvariable mit Werten im messbaren Raum $(S', \mathscr{A}')$ und sei $f : \Omega \to \mathbb{R}$ $\sigma(Y)$-messbar. Dann gibt es eine messbare Funktion $g : S' \to \mathbb{R}$ mit $f = g \circ Y$ (siehe z.B. [St1, Lemma 1.48] oder [Kl, Korollar 1.97]).

Beweis. Sei zunächst $f = \mathbf{1}_A$ für ein $A \in \sigma(Y) = \{Y^{-1}(B) \mid B \in \mathscr{A}'\}$, also gibt es ein $B \in \mathscr{A}'$ mit $A = Y^{-1}(B)$. Dann ist $g := \mathbf{1}_B$ messbar und es gilt $f = g \circ Y$.

Sei nun $f \geq 0$. Schreibe $f = \sum_{n=1}^\infty \alpha_n \mathbf{1}_{A_n}$ für geeignete $\alpha_n \geq 0$, $A_n \in \sigma(Y)$, beispielsweise

$$f = \sum_{n=1}^\infty n \mathbf{1}_{\{f \in [n, n+1)\}} + \sum_{k=1}^\infty 2^{-k} \mathbf{1}_{\{k\text{-te Ziffer in der Binärentwicklung von } f - \lfloor f \rfloor \text{ ist } 1\}}.$$

Da die Behauptung für Indikatorfunktionen gilt, folgt sie damit auch für nichtnegative f.

Sei nun f beliebig. Schreibe $f = f^+ - f^-$ mit $f^+, f^- \geq 0$. Da die Behauptung für nichtnegative Funktionen gilt, folgt sie damit auch für f. □

Beweis von Korollar 6.7. Die F_r, $r \in \mathbb{Q}$ aus dem Beweis von Satz 6.6 sind $\sigma(Y)$-messbar, also gibt es nach Erinnerung 6.8 eine messbare Funktion $F'_r: S' \to \mathbb{R}$ mit $F_r = F'_r \circ Y$.

Führe die Konstruktion aus dem Beweis von Satz 6.6 für die F'_r durch, erhalte $\kappa'(x', \cdot)$ für $x' \in S'$ als das Wahrscheinlichkeitsmaß mit Verteilungsfunktion $\widetilde{F}'(x')$. □

Lemma 6.9. *Sei $(S, \mathscr{A})$ ein Borel-Raum, $(S', \mathscr{A}')$ ein messbarer Raum, X eine S-wertige und Y eine S'-wertige Zufallsvariable auf demselben Wahrscheinlichkeitsraum. Dann gibt es eine messbare Funktion $f: S' \times [0,1] \to S$, sodass gilt:*

Ist $\widetilde{Y}$ eine Zufallsvariable mit $\widetilde{Y} \overset{d}{=} Y$ und $\widetilde{U} \sim \mathrm{unif}([0,1])$ unabhängig von $\widetilde{Y}$, so ist

$$\left(f(\widetilde{Y}, \widetilde{U}), \widetilde{Y}\right) \overset{d}{=} (X, Y).$$

Beweis. Sei ohne Einschränkung $S \subset \mathbb{R}$ (messbar), ansonsten gibt es (wie im Beweis von Satz 6.6) $B \in \mathscr{B}(\mathbb{R})$ und $\varphi: S \to B$ bijektiv und bi-messbar; betrachte $X' := \varphi(X)$, konstruiere ein entsprechendes $f': S' \times [0,1] \to B$ und setze dann $f := \varphi^{-1} \circ f$.

Sei κ' ein stochastischer Kern von S' nach S gemäß Korollar 6.7. Setze

$$f(y,u) := \sup\{x \in \mathbb{R} \mid \kappa'(y, (-\infty, x]) < u\}, \quad y \in S', u \in [0,1].$$

f ist messbar, denn

$$f(y,u) = \sup\{x \in \mathbb{Q} \mid \kappa'(y, (-\infty, x]) < u\}$$

und

$$\begin{aligned} C_x &:= \{(y,u) \in S' \times [0,1] \mid \kappa'(y, (-\infty, x]) < u\} \\ &= \bigcup_{v \in \mathbb{Q}_+ \cap (0,1)} \{y \mid \kappa'(y, (-\infty, x]) < v\} \times [v, 1] \in \mathscr{A}' \otimes \mathscr{B}([0,1]). \end{aligned}$$

Für $x \in \mathbb{R}$ und $A \in \mathscr{A}$ ist

$$\begin{aligned} \mathbb{P}\left(\widetilde{Y} \in A, f(\widetilde{Y}, \widetilde{U}) > x\right) &= \mathbb{P}\left(\widetilde{Y} \in A, \kappa'(\widetilde{Y}, (-\infty, x]) < \widetilde{U}\right) \\ &= \int_{S' \times [0,1]} \mathbf{1}_A(y) \mathbf{1}_{\{\kappa'(y, (-\infty, x]) < u\}} \mathscr{L}(\widetilde{Y}, \widetilde{U})(\,dy,\, du) \\ &= \int_{S'} \mathbf{1}_A(y) \int_{[0,1]} \mathbf{1}_{\{\kappa'(y, (-\infty, x]) < u\}} \, du \, \mathscr{L}(\widetilde{Y})(\,dy) \\ &= \int_{S'} \mathbf{1}_A(y) \left(1 - \kappa'(y, (-\infty, x])\right) \mathscr{L}(\widetilde{Y})(\,dy) \\ &= \mathbb{E}\left[\mathbf{1}_{\{Y \in A\}} \left(1 - \kappa'(Y, (-\infty, x])\right)\right] \\ &= \mathbb{E}\left[\mathbf{1}_{\{Y \in A\}} \kappa'(Y, (x, \infty))\right] \\ &= \mathbb{E}\left[\mathbf{1}_{\{Y \in A\}} \mathbf{1}_{\{X > x\}}\right]. \end{aligned}$$

□

Im Folgenden sei I eine beliebige Indexmenge, $(S_i, \mathscr{A}_i)$, $i \in I$ seien Borel-Räume und für $J \subset I$ sei $\Omega_J := \times_{i \in J} S_i$ mit Produkt-σ-Algebra $\mathscr{A}_J := \bigotimes_{i \in J} \mathscr{A}_i$. Für $J' \subset J \subset I$ sei

$$\pi^J_{J'}: \Omega_J \to \Omega_{J'}, \quad (x_i)_{i \in J} \mapsto (x_i)_{i \in J'}$$

die kanonische Projektion. ($\pi^J_{J'}$ ist offenbar messbar.)

Definition 6.10. Für $J \subset I$ mit $0 < |J| < \infty$ sei P_J ein Wahrscheinlichkeitsmaß auf $(\Omega_J, \mathscr{A}_J)$. $\{P_J \mid J \subset I,\ 0 < |J| < \infty\}$ heißt *projektive Familie*, wenn gilt

$$P_{J'} = P_J \circ \left(\pi^J_{J'}\right)^{-1} \quad \text{für alle } J' \subset J.$$

Beispiel 6.11. Sei $I = \mathbb{N}_0$, $S_i = S$ eine höchstens abzählbare Menge, $\mu \in \mathcal{M}_1(S)$ und $(p_{xy})_{x,y\in S}$ eine stochastische Matrix. Sei $P_n := P_{\{0,1,\ldots,n\}} \in \mathcal{M}_1\left(S^{\{0,1,\ldots,n\}}\right)$ gegeben durch

$$P_n\left(\{x_0, x_1, \ldots, x_n\}\right) = \mu(\{x_0\}) \prod_{i=0}^{n-1} p_{x_i, x_{i+1}}$$

(und erhalte P_J für allgemeine $J \subset \mathbb{N}$, $0 < |J| < \infty$ durch „Aussummieren" der Koordinaten in $\{0, 1, \ldots, n\} \setminus J$ aus P_n, wobei $n = \max J$). Dann ist $\{P_J \mid J \subset I,\ 0 < |J| < \infty\}$ eine projektive Familie.

(P_n beschreibt die Verteilung der ersten n Schritte einer Markovkette mit Startverteilung μ und Übergangsmatrix p.)

Satz 6.12 (Kolmogorovs Erweiterungssatz[1]). *Zu einer projektiven Familie* $\{P_J \mid J \subset I,\ 0 < |J| < \infty\}$ *auf einem Produkt von Borel-Räumen gibt es genau ein Wahrscheinlichkeitsmaß* P *auf* $(\Omega, \mathscr{A}) := (\Omega_I, \mathscr{A}_I)$ *mit*

$$P_J = P \circ (\pi^I_J)^{-1} \quad \textit{für jedes} \quad J \subset I \textit{ mit } 0 < |J| < \infty.$$

P *heißt* projektiver Limes der $\{P_J\}$, *auch geschrieben* $P = \varprojlim_{J \nearrow I} P_J$.

Beweis. Zur Eindeutigkeit: Angenommen P und P' seien projektive Limiten der $\{P_J\}$. Seien

$$Z_J = \left\{A \in \mathscr{A} \mid A = (\pi^I_J)^{-1}(B) \text{ für ein } B \in \mathscr{A}_J\right\}$$

die „Zylindermengen" mit Basis J. $\bigcup_{J \subset I, |J| < \infty} Z_J$ ist ein schnittstabiler Erzeuger von $\mathscr{A}$, auf dem P und P' wegen $P \circ (\pi^I_J)^{-1} = P_J = P' \circ (\pi^I_J)^{-1}$ übereinstimmen. Nach dem Eindeutigkeitssatz für Maße folgt daher $P = P'$.

Zur Existenz: Betrachte dazu zunächst den Fall I abzählbar, sei ohne Einschränkung $I = \mathbb{N}_0$ und $P_n := P_{\{0,1,\ldots,n\}}$.

Nach Lemma 6.9 gibt es eine messbare Funktion

$$f_n : S_0 \times \ldots \times S_n \times [0,1] \to S_{n+1},$$

so dass gilt: Ist $(\tilde{X}_0, \ldots \tilde{X}_n) \sim P_n$ und $\tilde{U}_n \sim \mathrm{unif}([0,1])$ unabhängig von $(\tilde{X}_0, \ldots \tilde{X}_n)$, so ist

$$\left(\tilde{X}_0, \ldots, \tilde{X}_n, f((\tilde{X}_0, \ldots \tilde{X}_n), \tilde{U}_n)\right) \sim P_{n+1}$$

(sei $(X_0, \ldots, X_{n+1}) \sim P_{n+1}$ und lese $X := X_{n+1}$, $Y := (X_0, \ldots, X_n)$ in Lemma 6.9).

Sei $X_0 \sim P_0$ und seien $U_0, U_1, \ldots \sim \mathrm{unif}([0,1])$ unabhängig von X_0 (für die Existenz von $U_0, U_1, \ldots$ vgl. [St1, Kor. 5.10] oder [Kl, Kor. 14.36]). Konstruiere $X_1, X_2, \ldots$ via

$$X_{n+1} := f_n((X_0, \ldots, X_n), U_n),$$

[1] Andrej N. Kolmogorov, 1903–1987

dann gilt induktiv

$$(X_0,\dots,X_n)\sim P_n.$$

$P:=\mathscr{L}(X_0,X_1,\dots)$ ist somit projektiver Limes der $\{P_J\}$.

Sei nun I überabzählbar. Sei

$$\mathcal{C}:=\bigcup_{\substack{I'\subset I\\ I' \text{ abzählbar}}} Z_{I'},$$

offenbar ist $\mathcal{C}\subset\mathscr{A}$ und alle Koordinatenprojektionen sind $\mathcal{C}$-messbar.

$\mathcal{C}$ ist σ-Algebra, denn

i) $\Omega=(\pi^I_{I'})^{-1}(\Omega_{I'})\in\mathcal{C}$

ii) Sei $A\in\mathcal{C}$, das heißt es gibt ein abzählbares $I'\subset I$ mit $A\in Z_{I'}$. Dann ist aber auch $A^{\mathrm{c}}\in Z_{I'}$, also $A^{\mathrm{c}}\in\mathcal{C}$.

iii) Seien $A_1,A_2,\dots\in\mathcal{C}$, also $A_n\in Z_{I'_n}$ für ein abzählbares $I'_n\subset I$. Dann ist auch $I':=\bigcup_{n\in\mathbb{N}}I'_n\subset I$ abzählbar und es gilt $\bigcup_{n\in\mathbb{N}}A_n\in Z_{I'}\subset\mathcal{C}$.

Zudem gilt $\mathscr{A}\subset\mathcal{C}$, da alle Koordinatenprojektionen $\mathcal{C}$-messbar sind. Zusammen folgt $\mathscr{A}=\mathcal{C}$.

Für ein abzählbares $J\subset I$ gibt es nach obigem genau ein Wahrscheinlichkeitsmaß P_J auf $(\Omega_J,\mathscr{A}_J)$ mit

$$P_K=P_J\circ(\pi^J_K)^{-1}\quad\text{für jedes endliche } K\subset J.$$

Fasse dies via

$$\widetilde{P}_J\left((\pi^I_J)^{-1}(B)\right):=P_J(B)\quad\text{für } B\in\mathscr{A}_J$$

als Wahrscheinlichkeitsmaß $\widetilde{P}_J$ auf (Ω,Z_J) auf.

Für $A\in\mathscr{A}=\mathcal{C}$ setze

$$P(A):=\widetilde{P}_J(A),\qquad\text{falls } A\in Z_J.$$

P ist wohldefiniert, denn ist $B\in Z_J\cap Z_{J'}$ eine Zylindermenge mit endlicher Basis K ($\subset J\cap J'$), d.h. $B=(\pi^I_K)^{-1}(B')$ für ein $B'\in\mathscr{A}_K$, dann ist

$$\widetilde{P}_J(B)=P_K(B')=\widetilde{P}_{J'}(B)$$

und

$$\left\{(\pi^I_K)^{-1}(B)\mid K\subset J\cap J',\ |K|<\infty,\ B\in\mathscr{A}_K\right\}$$

ist ein schnittstabiler Erzeuger von $Z_J\cap Z_{J'}$. Also gilt $\widetilde{P}_J=\widetilde{P}_{J'}$ auf $Z_J\cap Z_{J'}$.

Es bleibt nur noch zu zeigen, dass P ein Wahrscheinlichkeitsmaß auf $(\Omega,\mathscr{A})$ ist. Es gilt $P(\varnothing)=0$ und $P(\Omega)=1$, denn $\varnothing,\Omega\in Z_K$ für jedes K. Seien $A_1,A_2,\dots\in\mathscr{A}$ paarweise disjunkt, das heißt $A_n\in Z_{J_n}$ für ein abzählbares $J_n\subset I$. Dann ist auch $J:=\bigcup_{n\in\mathbb{N}}J_n$ abzählbar und es gilt $A_1,A_2,\dots,\bigcup_{n\in\mathbb{N}}A_n\in Z_J$, also folgt

$$P\left(\bigcup_{n\in\mathbb{N}}A_n\right)=\widetilde{P}_J\left(\bigcup_{n\in\mathbb{N}}A_n\right)=\sum_{n=1}^\infty\widetilde{P}_J(A_n)=\sum_{n=1}^\infty P(A_n).$$

$\square$

6.2 Markov-Prozesse und Markov-Halbgruppen

Es sei E ein polnischer Raum, $I = \mathbb{N}_0$ oder $I = [0,\infty)$ (oder allgemein $I \subset \mathbb{R}$ mit der Interpretation I als „Zeitindexmenge"). Sei $X = (X_t)_{t\in I}$ ein adaptierter, stochastischer Prozess mit Werten in E (d.h. für jedes $t \in I$ ist X_t ist eine E-wertige, $\mathscr{F}_t$-messbare Zufallsvariable) definiert auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, (\mathscr{F}_t)_{t\in I})$. Sei weiter $(P_x)_{x\in E}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$.

Definition 6.13. X heißt *Markov-Prozess mit Verteilungen* $(P_x)_{x\in E}$, wenn gilt:

i) Für $x \in E$ gilt $P_x(X_0 = x) = 1$.

ii) $\kappa(x,B) := P_x(X \in B)$ für $x \in E, B \in \mathscr{B}(E)^{\otimes I}$ ist ein stochastischer Kern.

iii) X besitzt die *schwache Markov-Eigenschaft*: Für $x \in E$, $A \in \mathscr{B}(E)$ und $s,t \in I$ gilt

$$P_x(X_{t+s} \in A \mid \mathscr{F}_s) = \kappa_t(X_s, A) \quad P_x\text{-f.s.}$$

mit $\kappa_t(x,A) := \kappa(x, \{y = (y_v)_{v\in I} \in E^I \mid y_t \in A\}) = P_x(X_t \in A)$.

Falls E abzählbar ist, so heißt X auch *diskreter Markov-Prozess*. Falls $I = \mathbb{N}_0$, so heißt X auch *Markov-Kette*.

Wir schreiben $\mathbb{E}_x[\ldots]$ für Erwartungswerte unter P_x, $\mathscr{L}_x(X)$ für die Verteilung von X unter P_x, analog $\mathscr{L}_x(X \mid \mathscr{G})$, etc.

Wir betrachten in diesem Kapitel nur den zeitlich homogenen Fall: für $t_1 < t_2$ hängt die bedingte Verteilung von X_{t_2} gegeben $\mathcal{F}_{t_1}$ nur von X_{t_1} und $t_2 - t_1$ ab, nicht aber explizit von t_1 und t_2, vgl. Definition 6.13, iii).

Bemerkung 6.14. Ein Markov-Prozess besitzt die *elementare Markov-Eigenschaft*: Unter jedem P_x gilt für $u \le t$

$$P_x(X_t \in A \mid \mathscr{F}_u) = P_x(X_t \in A \mid X_u) \quad f.s.$$

Dies folgt unmittelbar aus Definition 6.13 $iii)$, verlangt aber im Gegensatz zu Definition 6.13 nicht die zeitliche Homogenität der Dynamik.

Definition 6.15. Sei $I \subset [0,\infty)$ abgeschlossen unter Addition. Eine Familie $(\kappa_t)_{t\in I}$ von (sub-) stochastischen Kernen von E nach E heißt *(sub-) stochastische Halbgruppe*, wenn für alle $s,t \in I$ gilt:

$$\kappa_s \circ \kappa_t = \kappa_{s+t}.$$

Dies sind die sogenannten *Chapman-Kolmogorov-Gleichungen*[2].

Satz 6.16. *Ist* $((X_t)_{t\in I}, (P_x)_{x\in E})$ *ein Markov-Prozess, so definiert*

$$\kappa_t(x,A) := P_x(X_t \in A), \quad x \in E,\ A \in \mathscr{B}(E),\ t \in I \tag{6.4}$$

[2]nach Andrej Nikolaevich Kolmogorov, 1903–1987; S. Chapman, 1888–1970

eine Markov-Halbgruppe und die endlich-dimensionalen Verteilungen von $(X_t)_{t\in I}$ sind durch $(\kappa_t)_{t\in I}$ festgelegt. Insbesondere gilt für $n \in \mathbb{N}$, $0 = t_0 < t_1 < \ldots < t_n$ und $f_1, \ldots, f_n\colon E \to \mathbb{R}$ beschränkt und messbar

$$\mathbb{E}_x\left[\prod_{j=1}^n f_j(X_{t_j})\right] = \int \kappa_{t_1-t_0}(x, dx_1) f_1(x_1) \cdots \int \kappa_{t_n-t_{n-1}}(x_{n-1}, dx_n) f_n(x_n). \qquad (6.5)$$

Umgekehrt gibt es zu jeder Markov-Halbgruppe $(\kappa_t)_{t\in I}$ einen Markov-Prozess $(X_t)_{t\in I}$, sodass (6.4) und (6.5) gelten.

Beweis. Sei $((X_t)_{t\in I}, (P_x)_{x\in E})$ ein Markov-Prozess. Nach Definition 6.13 $ii)$ definiert (6.4) einen stochastischen Kern. Weiter gilt für $s, t \in I$ und $A \in \mathscr{B}(E)$

$$\begin{aligned}\kappa_{s+t}(x, A) = P_x(X_{s+t} \in A) &= \mathbb{E}_x\left[P_x(X_{s+t} \in A \mid \mathscr{F}_s)\right] = \mathbb{E}_x[\kappa_t(X_s, A)] \\ &= \int \kappa_t(y, A)\, \kappa_s(x, dy) = (\kappa_s \circ \kappa_t)(x, A).\end{aligned}$$

Damit ist $(\kappa_t)_{t\in I}$ eine Markov-Halbgruppe.

Zeige (6.5) induktiv. Sei $f_1 = \mathbf{1}_A$, $A \in \mathscr{B}(R)$. Es ist

$$\mathbb{E}_x[f_1(X_{t_1})] = P_x(X_{t_1} \in A) = \kappa_{t_1}(x, A),$$

also gilt (6.5) für $n = 1$ für Linearkombination von Indikatorfunktionen und somit mit den „üblichen Approximationsargumenten“ auch für allgemeine f. Es sei nun (6.5) wahr für ein $n \in \mathbb{N}$. Sei $\tilde{f}(X_{t_n}) := \int f_{n+1}(y) \kappa_{t_{n+1}-t_n}(X_{t_n}, dy)$. Dann ist

$$\begin{aligned}\mathbb{E}_x\left[\prod_{j=1}^{n+1} f_j(X_{t_j})\right] &= \mathbb{E}_x\left[\mathbb{E}_x\left[\prod_{j=1}^{n+1} f_j(X_{t_j}) \,\middle|\, \mathscr{F}_{t_n}\right]\right] = \mathbb{E}_x\left[\prod_{j=1}^{n} f_j(X_{t_j}) \mathbb{E}_x[f_{n+1}(X_{t_{n+1}}) \mid \mathscr{F}_{t_n}]\right] \\ &= \mathbb{E}_x\left[\prod_{j=1}^{n} f_j(X_{t_j}) \int f_{n+1}(y) \kappa_{t_{n+1}-t_n}(X_{t_n}, dy)\right] = \mathbb{E}_x\left[\prod_{j=1}^{n} f_j(X_{t_j})\, \tilde{f}(X_{t_n})\right].\end{aligned}$$

Wende nun (6.5) an auf $f_1, f_2, \ldots, f_{n-1}, f_n \tilde{f}$.

Sei umgekehrt $(\kappa_t)_{t\in I}$ eine Markov-Halbgruppe. Für ein endliches $J \subset I$, $J = \{t_0, t_1, \ldots, t_n\}$ mit $0 = t_0 < t_1 < \ldots < t_n$ definiert (6.5) ein Wahrscheinlichkeitsmaß $P_{x,J}$ auf $E^{|J|}$:

$$P_{x,J}(A) = \int \kappa_{t_1-t_0}(x, dx_1) \ldots \int \kappa_{t_n-t_{n-1}}(x_{n-1}, dx_n) \cdot \mathbf{1}_A(x, x_1, \ldots, x_n), \quad A \in E^{|J|}.$$

Zeige: $\{P_{x,J} \mid J \subset I, |J| < \infty\}$ ist eine projektive Familie. Sei J wie oben gegeben, $J' := J \setminus \{t_l\}$ für ein $l \in \{1, \ldots, n\}$ und $A_i \in \mathscr{B}(E)$. Dann gilt

$$\begin{aligned}P_{x,J}\left((\pi^J_{J'})^{-1}(A_0 \times \ldots \times A_{l-1} \times A_{l+1} \times \ldots \times A_n)\right) &= P_{x,J}\left(\ldots \times A_{l-1} \times E \times A_{l+1} \times \ldots\right) \\ &= \ldots \int \kappa_{t_l-t_{l-1}}(x_{l-1}, dx_l) \mathbf{1}_E(x_l) \int \kappa_{t_{l+1}-t_l}(x_l, dx_{l+1}) \mathbf{1}_{A_{l+1}}(x_{l+1}) \ldots \\ = \ldots \int \kappa_{t_l-t_{l-1}} \circ \kappa_{t_{l+1}-t_l}(x_{l-1}, dx_{l+1}) \ldots &= \ldots \int \kappa_{t_{l+1}-t_{l-1}}(x_{l-1}, dx_{l+1}) \ldots \\ &= P_{x,J'}\left(A_0 \times \ldots \times A_{l-1} \times A_{l+1} \times \ldots \times A_n\right),\end{aligned}$$

das heißt $P_{x,J} \circ (\pi^J_{J'})^{-1} = P_{x,J'}$. Mit Satz 6.12 folgt die Existenz eines Wahrscheinlichkeitsmaßes P_x auf $\Omega = E^I$, $\mathscr{A} = \mathscr{B}(E)^{\otimes I}$, sodass für die t-te Koordinatenprojektion

$$X_t\colon \Omega \to E, \quad \mathscr{F}_t = \sigma(X_s \mid s \le t)$$

die Formel (6.5) gilt. $((X_t)_{t\in I}, (P_x)_{x\in E})$ leistet das Gewünschte.

Zeige die (schwache) Markov-Eigenschaft, das heißt $P_x(X_{t+s} \in A \mid \mathscr{F}_s) = \kappa_t(X_s, A)$. Mengen B der Form

$$B = \{X_{t_0} \in A_0, \dots, X_{t_{n-1}} \in A_{n-1}\}, \quad 0 = t_0 < \ldots < t_n = s,\ n \in \mathbb{N},\ A_i \in \mathscr{B}(E)$$

sind ein schnittstabiler Erzeuger von $\mathscr{F}_s$. Es reicht also,

$$\mathbb{E}_x\left[\kappa_t(X_s, A)\mathbf{1}_B\right] = \mathbb{E}_x\left[\mathbf{1}_{\{X_{t+s}\in A\}}\mathbf{1}_B\right]$$

zu zeigen. Betrachte dazu

$$\begin{aligned}
&P_x(X_{t_0} \in A_0, \dots, X_{t_n} \in A_n) \\
&= \int P_x(X_{t_0} \in A_0, \dots, X_{t_{n-2}} \in A_{n-2}, X_{t_{n-1}} \in dx_{n-1})\mathbf{1}_{A_{n-1}}(x_{n-1})\kappa_{t_n - t_{n-1}}(x_{n-1}, A_n) \\
&= \mathbb{E}_x\left[\mathbf{1}_{\{X_{t_0}\in A_0,\dots,X_{t_{n-1}}\in A_{n-1}\}}\kappa_{t_n-t_{n-1}}(X_{n-1}, A_n)\right],
\end{aligned}$$

demnach also $P_x(X_{t_n} \in A_n \mid \mathscr{F}_{t_{n-1}}) = \kappa_{t_n - t_{n-1}}(X_{t_{n-1}}, A_n)$ (f.s.). □

Bemerkung 6.17. Alternativ (und äquivalent zu Definition 6.15) betrachtet man die sogenannte Übergangshalbgruppe

$$T_t f(x) = \int_E f(y)\,\kappa_t(x, dy), \quad t \ge 0, x \in E$$

die zumindest für $f : E \to \mathbb{R}$ beschränkt und messbar wohldefiniert ist; $T_t f$ ist dann wiederum eine beschränkte, messbare Funktion auf E. Offenbar gewinnt man mit Wahl $f = \mathbf{1}_A$, $A \in \mathscr{B}(E)$ aus $T_t\mathbf{1}_A(x) = \kappa_t(x, A)$ den Kern κ_t zurück. Analog zur Interpretation $\kappa_t(x, A) = P_x(X_t \in A)$ in (6.4) ist $T_t f(x) = \mathbb{E}_x[f(X_t)]$ für den Markovprozess $X = (X_t)_{t\in I}$ aus Satz 6.16.

Die Halbgruppeneigenschaft von $(\kappa_t)_{t\in I}$ bezüglich Verkettung aus Definition 6.15 ist äquivalent zur Halbgruppeneigenschaft der Familie $(T_t)_{t\in I}$ bezüglich Hintereinanderausführung (als lineare Operatoren), d.h.

$$T_t T_s f := (T_t \circ T_s) f = T_t(T_s f) = T_{t+s} f$$

Offensichtlich gilt $T_t f \ge 0$ für $f \ge 0$ (T_t ist *positiv*) und mit $\|f\| := \sup_{x\in E}|f(x)|$ auch $\|T_t f\| \le \|f\|$ (T_t ist eine *Kontraktion*) und $T_t\mathbf{1} = \mathbf{1}$ (bzw. $T_t\mathbf{1} \le \mathbf{1}$, falls (κ_t) sub-Markov ist).

Beispiel 6.18. Sei E abzählbar und $p = (p(x,y))_{x,y\in E}$ eine stochastische Matrix. Sei

$$\kappa_n(x, A) = \sum_{y\in A} p^n(x, y).$$

$(\kappa_n)_{n\in\mathbb{N}_0}$ ist eine Markov-Halbgruppe. Satz 6.16 liefert eine Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$ mit Übergangsmatrix p.

Beispiel 6.19 (Faltungshalbgruppen und Markovprozesse mit unabhängigen stationären Zuwächsen). Sei $(\nu_t)_{t\in[0,\infty)} \subset \mathcal{M}_1(\mathbb{R})$ eine Faltungsgruppe, das heißt $\nu_s * \nu_t = \nu_{s+t}$ für $s,t \geq 0$.

$$\kappa_t(x,\cdot) = \delta_x * \nu_t$$

definiert eine Markov-Halbgruppe auf $\mathbb{R}$, denn für eine beschränkte, messbare Funktion $f\colon\mathbb{R} \to \mathbb{R}$ gilt

$$\begin{aligned}\int f(y)\,(\kappa_s \circ \kappa_t)(x,dy) &= \int\int f(y)\,\kappa_t(z,dy)\,\kappa_s(x,dz) = \int\int f\,d(\delta_z * \nu_t)\,\kappa_s(x,dz)\\ &= \int\int f(z+y')\,\nu_t(dy')\,\kappa_s(x,dz) = \int\int f(x+z'+y')\,\nu_s(dz')\,\nu_t(dy')\\ &= \int f(x+x')\,(\nu_s * \nu_t)(dx') = \int f(x')\,\kappa_{s+t}(x,dx').\end{aligned}$$

Sei $X_t : \mathbb{R}^{[0,\infty)} \to \mathbb{R}$ die t-te Koordinatenprojektion. Nach Satz 6.16 (und seinem Beweis) gibt es auf $\Omega := \mathbb{R}^{[0,\infty)}$, $\mathscr{A} := \mathscr{B}(\mathbb{R})^{[0,\infty)}$ eine Familie von Wahrscheinlichkeitsmaßen $(P_x)_{x\in\mathbb{R}}$ mit

$$P_x \circ (X_{t_0},\dots,X_{t_n})^{-1} = \delta_x \otimes \bigotimes_{j=1}^{n} \kappa_{t_j - t_{j-1}}, \qquad n \in \mathbb{N},\ 0 = t_0 < \ldots < t_n.$$

Es gilt

$$P_x(X_{t_0} \in B_0, X_{t_1} - X_{t_0} \in B_1, \dots, X_{t_n} - X_{t_{n-1}} \in B_n) = \delta_x(B_0) \prod_{j=1}^{n} \nu_{t_j - t_{j-1}}(B_j).$$

Man sagt: $(X_t)_{t\in[0,\infty)}$ hat unabhängige, stationäre Zuwächse.

Insbesondere: Sei $\nu \in \mathcal{M}_1(\mathbb{R})$ unendlich teilbar mit $\varphi_\nu(u) = \exp(\psi(u))$ mit ψ gemäß der Lévy-Khinchin-Formel aus Satz 5.9. Ist $\nu_t, t \geq 0$ das Wahrscheinlichkeitsmaß auf $\mathbb{R}$ mit $\varphi_{\nu_t}(u) = \exp(t\psi(u))$, so ist

$$\varphi_{\nu_s * \nu_t}(u) = \varphi_{\nu_s} \cdot \varphi_{\nu_t} = e^{s\psi} \cdot e^{t\psi} = e^{(s+t)\psi} = \varphi_{\nu_{s+t}}(u),$$

das heißt $(\nu_t)_{t\geq 0}$ ist eine Faltungshalbgruppe.

Bericht 6.20. Den zur Wahl $\nu_t = \mathcal{N}(0,t)$ in Bsp. 6.19 gehörigen Prozess (nämlich die Brownsche Bewegung) hatten wir bereits explizit in Satz 2.2 konstruiert. Wir haben dort auch gezeigt, dass der zugehörige Prozess so gewählt werden kann, dass er stetige Pfade besitzt, was nicht aus dem Argument in Bsp. 6.19 folgt (Bsp. 6.19 garantiert nur eine „rohe“ Version ohne Pfadeigenschaften, d.h. ein W'maß auf dem Produktraum $\mathbb{R}^{[0,\infty)}$, dessen endlich-dimensionale Verteilungen mit denen aus Satz 2.2 übereinstimmen).

Im Allgemeinen kann man für die Prozesse aus Bsp. 6.19 keine Version mit stetigen Pfaden finden, es ist gibt aber immer eine Version mit rechtsstetigen Pfaden, die linksseitige Limiten besitzen. Dies sind die sogenannten Lévy-Prozesse. (Ein einfaches Beispiel sehen wir in Bem. 6.22 unten, nämlich Poissonprozesse auf $\mathbb{R}_+$.)

Beispiel 6.21. Sei E endlich, $(Q(x,y))_{x,y\in E}$ eine Matrix mit $Q(x,y) \geq 0$ für $x \neq y$ und $\sum_{y\in E} Q(x,y) = 0$ für alle $x \in E$, wir betrachten Zeitindexmenge $I = \mathbb{R}_+$. Sei

$$T_t := e^{tQ} = \sum_{n=0}^{\infty} \frac{1}{n!} t^n Q^n, \quad P(0) = I. \tag{6.6}$$

Die Reihe konvergiert, denn $\max_{x,y\in E} |Q^n(x,y)| \leq (|E| \max_{x,y\in E} |Q(x,y)|)^n$. Es gilt $T_t T_s = T_{t+s}$, denn für kommutierende Matrizen A, B (d.h. $AB = BA$) ist $e^A e^B = e^{A+B}$. Die linearen Operatoren $(T_t)_{t\geq 0}$ (die als Matrizen dargestellt sind , da $|E| < \infty$) bilden eine Markov-Übergangshalbgruppe wie in Bemerkung 6.17. Insbesondere gilt

$$T_{t+s}(x,y) = \sum_{z\in E} T_t(x,z) T_s(z,y)$$

$\kappa_t(x,A) := \sum_{y\in A} T_t(x,y)$ bildet eine Markov-Halbgruppe, das heißt es gibt einen Markov-Prozess $(X_t)_{t\geq 0}$ mit

$$P_{x_0}(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) = \prod_{j=1}^{n} T_{t_j - t_{j-1}}(x_{j-1}, x_j)$$

Sei $\varrho := \max_{x\in E}\{-Q(x,x)\}$, $\widehat{p}(x,y) := \frac{Q(x,y)}{\varrho}$ für $x \neq y$ und $\widehat{p}(x,x) := 1 + \frac{Q(x,x)}{\varrho}$. $(\widehat{p}(x,y))_{x,y\in E}$ ist eine stochastische Matrix. Es gilt $Q = \varrho\widehat{p} - \varrho I$ und

$$T_t = \sum_{n=0}^{\infty} \frac{t^n}{n!} \sum_{k=0}^{n} \binom{n}{k} (\varrho\widehat{p})^k (-\varrho)^{n-k} = \sum_{k=0}^{\infty} \frac{(\varrho t)^k}{k!} \widehat{p}^k \underbrace{\sum_{n\geq k} \frac{(-t\varrho)^{n-k}}{(n-k)!}}_{=e^{-t\varrho}} = \sum_{k=0}^{\infty} e^{-t\varrho} \frac{(t\varrho)^k}{k!} \widehat{p}^k,$$

insbesondere ist also T_t eine stochastische Matrix und es gilt

$$\begin{aligned} P_{x_0}(X_{t+h} = y \mid X_t = x) &= \frac{T_t(x_0,x) T_h(x,y)}{T_t(x_0,x)} \\ &= T_h(x,y) = \left(e^{hQ}\right)(x,y) = \delta_{xy} + hQ(x,y) + o(h), \quad h \searrow 0. \end{aligned} \tag{6.7}$$

Die Matrix Q heißt die *Generatormatrix* oder auch die *Q-Matrix* der Halbgruppe $(T_t)_{t\geq 0}$.

Analytisch gesehen folgt aus (6.6), dass gilt (beachte: Q und $\exp(tQ)$ kommutieren)

$$\frac{\partial}{\partial t} \exp(tQ) = Q \exp(tQ) = \exp(tQ) Q, \tag{6.8}$$

wie man sich durch gliedweise Differentiation der Reihe in (6.6) überzeugt. Demnach gilt für $t \geq 0$

$$\begin{aligned} \frac{\partial}{\partial t} T_t = QT_t, \text{ d.h. } \forall\, x,y \in E : \ \frac{\partial}{\partial t} T_t(x,y) &= \sum_z Q(x,z) T_t(z,y) \\ &= \sum_z Q(x,z)\big(T_t(z,y) - T_t(x,y)\big), \end{aligned} \tag{6.9}$$

$$\begin{aligned} \frac{\partial}{\partial t} T_t = T_t Q, \text{ d.h. } \forall\, x,y \in E : \ \frac{\partial}{\partial t} T_t(x,y) &= \sum_z T_t(x,z) Q(z,y) \\ &= T_t(x,y) Q(y,y) + \sum_{z\neq y} T_t(x,z) Q(z,y) \end{aligned} \tag{6.10}$$

Stochastisch: Gleichung (6.9) heißt Kolmogorovs *Rückwärtsgleichung*, denn gemäß (6.7) gilt

$$\begin{aligned}
\frac{T_{t+h}(x,y) - T_t(x,y)}{h} &= \frac{1}{h}\big(P_x(X_{t+h} = y) - P_x(X_t = y)\big) \\
&= \frac{1}{h}\Big(\sum\nolimits_z P_x(X_{t+h} = y \,|\, X_h = z) P_x(X_h = z) - P_x(X_t = y)\Big) \\
&= \frac{1}{h}\Big(\sum\nolimits_z T_t(z,y)\big(1_{\{x=z\}} + hQ(x,z) + o(h)\big) - T_t(x,y)\Big) \\
&= \sum_z Q(x,z) T_t(z,y) + o(1)
\end{aligned}$$

für $h \downarrow 0$; man leitet sie also aus (6.7) her durch „Rückwärtszerlegung" des Prozesses X im Intervall $[0, t+h]$ gemäß dem Verhalten am Anfang des Intervalls. Analog heißt (6.10) Kolmogorovs *Vorwärtsgleichung*, sie entsteht durch Zerlegung gemäß dem Wert bei t:

$$\begin{aligned}
\frac{T_{t+h}(x,y) - T_t(x,y)}{h} &= \frac{1}{h}\Big(\sum\nolimits_z P_x(X_{t+h} = y \,|\, X_t = z) P_x(X_t = z) - P_x(X_t = y)\Big) \\
&= \frac{1}{h}\Big(\sum\nolimits_z T_t(x,z)\big(1_{\{z=y\}} + hQ(z,h) + o(h)\big) - T_t(x,y)\Big) \\
&= \sum_z T_t(x,z) Q(z,y) + o(1)
\end{aligned}$$

Sowohl (6.9) als auch (6.10) sind (im Fall $|E| < \infty$) eindeutig lösbar und beide bestimmen die Halbgruppe (es sind beides endliche Systeme linearer Differentialgleichungen mit konstanten Koeffizienten).

Bemerkung 6.22. In der Situation von Beispiel 6.21 kann man eine „Version" von $(X_t)_{t\geq 0}$ konstruieren, deren Pfade rechtsstetig mit linken Limiten sind. Seien dazu $\tau_1, \tau_2, \ldots$ unabhängig und identisch exponentialverteilt mit Parameter ϱ,

$$S_0 := 0,\ S_n = \sum_{i=1}^n \tau_i \text{ für } n \in \mathbb{N} \text{ und } N_t := \sum_{n=1}^\infty \mathbf{1}_{\{S_n \leq t\}},\ t \geq 0.$$

$(N_t)_{t\geq 0}$ ist ein Poisson-Prozess (mit Rate ϱ): Es gilt für $0 = t_0 < t_1 < \cdots < t_n$:

$$\mathscr{L}\big(N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}\big) = \bigotimes_{i=1}^n \mathrm{Poi}(\varrho(t_i - t_{i-1}))$$

(Übung), $(N_t)_{t\geq 0}$ ist ein Markovprozess mit unabhängigen, stationären Zuwächsen wie in Bsp. 6.19 (zur Faltungshalbgruppe $\nu_t = \mathrm{Poi}(\varrho t)$).

Sei $(\widehat{X}_k)_{k\in\mathbb{N}}$ eine zeitdiskrete Markov-Kette mit Übergangsmatrix $\widehat{p}$, dann leistet $X_t := \widehat{X}_{N_t}, t \geq 0$ das Gewünschte.

Satz 6.23. *Ein E-wertiger stochastischer Prozess $(X_t)_{t\in I}$ mit Wahrscheinlichkeitsmaßen $(P_x)_{x\in E}$ ist genau dann ein Markov-Prozess, wenn es einen stochastischen Kern $\kappa\colon E \times \mathscr{B}(E)^{\otimes I} \to [0,1]$ gibt, sodass für alle $x \in E,\ s \in I$ und alle beschränkten, messbaren Funktionen $f : E^I \to \mathbb{R}$ gilt*

$$\mathbb{E}_x\left[f((X_{t+s})_{t\in I}) \mid \mathscr{F}_s\right] = \mathbb{E}_{X_s}\left[f((X_u)_{u\in I})\right] = \int_{E^I} f(y)\, \kappa(X_s, dy) \quad \textit{f.s.} \tag{6.11}$$

(Zudem gelte $P_x(X_0 = x) = \kappa(x, \{y \in E^I : y_0 = x\}) = 1$, um auch Definition 6.13 i) zu erfüllen.)

Beweis. Es gelte zunächst (6.11). Für $t \in I$, $A \in \mathscr{B}(E)$ setze $f\colon E^I \to \mathbb{R}$, $y \mapsto \mathbf{1}_A(y_t)$ in (6.11) ein:

$$P_x(X_{t+s} \in A \mid \mathscr{F}_s) = \mathbb{E}_{X_s}[f(X)] = \kappa_t(X_s, A),$$

also gilt die schwache Markov-Eigenschaft (vgl. Definition 6.13 iii)).

Sei nun umgekehrt $(X_t)_{t\in I}$ ein Markov-Prozess. Betrachte zunächst f von der Form

$$f(y) = f_1(y_{t_1})\cdots f_n(y_{t_n})$$

für $n \in \mathbb{N}$, $t_1 < \ldots < t_n \in I$ und gewisse beschränkte, messbare Funktionen f_i. Für $n = 1$ folgt (6.11) aus der schwachen Markov-Eigenschaft von $(X_t)_{t\in I}$. Nehme also an, dass (6.11) für Funktionen f obiger Form mit einem festen $n \in \mathbb{N}$ erfüllt sei. Sei $f(y) = \mathbf{1}_{B_1}(y_{t_1})\cdots\mathbf{1}_{B_{n+1}}(y_{t_{n+1}})$. Dann gilt

$$\begin{aligned}
\mathbb{E}_x\left[f((X_{t+s})_{t\in I}) \mid \mathscr{F}_s\right] &= \mathbb{E}_x\left[\mathbf{1}_{B_1}(X_{s+t_1})\cdots\mathbf{1}_{B_{n+1}}(X_{s+t_{n+1}}) \mid \mathscr{F}_s\right] \\
&= \mathbb{E}_x\left[\mathbb{E}_x\left[\mathbf{1}_{B_1}(X_{s+t_1})\cdots\mathbf{1}_{B_{n+1}}(X_{s+t_{n+1}}) \mid \mathscr{F}_{t_n+s}\right] \mid \mathscr{F}_s\right] \\
&= \mathbb{E}_x\left[\mathbf{1}_{B_1}(X_{s+t_1})\cdots\mathbf{1}_{B_n}(X_{s+t_n}) \cdot P_x(X_{s+t_{n+1}} \in B_{n+1} \mid \mathscr{F}_{s+t_n}) \mid \mathscr{F}_s\right] \\
&= \mathbb{E}_x\left[\mathbf{1}_{B_1}(X_{s+t_1})\cdots\mathbf{1}_{B_n}(X_{s+t_n}) \cdot P_{X_{s+t_n}}(X_{t_{n+1}-t_n} \in B_{n+1}) \mid \mathscr{F}_s\right] \\
&= \mathbb{E}_{X_s}\left[\mathbf{1}_{B_1}(X_{t_1})\cdots\mathbf{1}_{B_n}(X_{t_n}) \cdot P_{X_{t_n}}(X_{t_{n+1}-t_n} \in B_{n+1})\right] \\
&= \mathbb{E}_{X_s}\left[\mathbf{1}_{B_1}(X_{t_1})\cdots\mathbf{1}_{B_n}(X_{t_n}) \cdot P_{X_s}(X_{t_{n+1}} \in B_{n+1} \mid \mathscr{F}_{t_n})\right] \\
&= \mathbb{E}_{X_s}\left[\mathbb{E}_{X_s}\left[\mathbf{1}_{B_1}(X_{t_1})\cdots\mathbf{1}_{B_n}(X_{t_n})\mathbf{1}_{B_{n+1}}(X_{t_{n+1}}) \mid \mathscr{F}_{t_n}\right]\right] \\
&= \mathbb{E}_{X_s}\left[f(X)\right],
\end{aligned} \tag{6.12}$$

also gilt (6.11) für Linearkombinationen solcher „Zylinderfunktionen“ und somit (mit den „üblichen Approximationsargumenten“) auch für allgemeine f mit $n+1$ Faktoren.

Ein weiteres „übliches“ Approximationsargument zeigt dann, dass (6.11) für beliebige, beschränkte, messbare $f : E^I \to \mathbb{R}$ gilt (beachte, dass Zylindermengen mit endlicher Basis einen schnittstabilen Erzeuger von $\mathscr{B}(E)^{\otimes I}$ bilden). □

Bemerkung. Im Beweis von Satz 6.23, speziell in Formel (6.12), spielt das Symbol $X = (X_t)_{t\in I}$ gewissermaßen mehrere Rollen, speziell beim Übergang von Zeile 4 zu Zeile 5: X_s wird „außen“ als „fixierte“ Zufallsvariable betrachtet, andererseits wird bei der Bildung des Erwartungswerts innerhalb von $\mathbb{E}_{X_s}[\cdots]$ über $X_{t_1}, \ldots, X_{t_{n+1}}$ mit dem Maß P_{X_s} gemittelt. Man könnte dies notationell betonen und eine „Kopie“ $\widetilde{X} = (\widetilde{X}_t)_{t\in I}$ des Prozesses X mit derselben Dynamik einführen. Damit schreibt sich (6.12) ab Zeile 5 als

$$\begin{aligned}
\ldots &= \mathbb{E}_{X_s}\left[\mathbf{1}_{B_1}(\widetilde{X}_{t_1})\cdots\mathbf{1}_{B_n}(\widetilde{X}_{t_n}) \cdot P_{\widetilde{X}_{t_n}}(\widetilde{X}_{t_{n+1}-t_n} \in B_{n+1})\right] \\
&= \mathbb{E}_{X_s}\left[\mathbf{1}_{B_1}(\widetilde{X}_{t_1})\cdots\mathbf{1}_{B_n}(\widetilde{X}_{t_n}) \cdot P_{X_s}(\widetilde{X}_{t_{n+1}} \in B_{n+1} \mid \mathscr{F}_{t_n})\right] \\
&= \mathbb{E}_{X_s}\left[\mathbb{E}_{X_s}\left[\mathbf{1}_{B_1}(\widetilde{X}_{t_1})\cdots\mathbf{1}_{B_n}(\widetilde{X}_{t_n})\mathbf{1}_{B_{n+1}}(\widetilde{X}_{t_{n+1}}) \mid \mathscr{F}_{t_n}\right]\right] \\
&= \mathbb{E}_{X_s}\left[f(\widetilde{X})\right]
\end{aligned}$$

Bemerkung 6.24. Sei $I = \mathbb{N}_0$.

i) $(X_n)_{n\in\mathbb{N}_0}$ ist genau dann eine Markov-Kette, wenn für alle $x \in E$ und $k \in \mathbb{N}_0$ gilt

$$\mathscr{L}_x\left((X_{n+k})_{n\in\mathbb{N}_0} \mid \mathscr{F}_k\right) = \mathscr{L}_{X_k}\left((X_n)_{n\in\mathbb{N}_0}\right).$$

ii) Sei $(X_n)_{n\in\mathbb{N}_0}$ ein stochastischer Prozess mit Verteilungen $(P_x)_{x\in E}$, wobei $P_x(X_0 = x) = 1$, und es gebe einen stochastischen Kern $\kappa_1\colon E \times \mathscr{B}(E) \to [0,1]$ mit

$$P_x(X_{s+1} \in A \mid \mathscr{F}_s) = \kappa_1(X_s, A) \quad f.s.$$

für alle $s \in \mathbb{N}_0$ und $A \in \mathscr{B}(E)$. Dann ist $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette und die n-Schritt-Übergangswahrscheinlichkeiten sind gegeben durch $\kappa_n = \kappa_{n-1} \circ \kappa_1$, $n \in \mathbb{N}$. $(\kappa_n)_{n\in\mathbb{N}}$ ist eine Markov-Halbgruppe und die Verteilung von X ist durch κ_1 festgelegt.

iii) In der Situation von $ii)$ sei E höchstens abzählbar. Dann ist $(X_n)_{n\in\mathbb{N}_0}$ genau dann eine Markov-Kette bezüglich $\mathscr{F}_n = \sigma(X_0, \dots, X_n)$, wenn es eine stochastische Matrix $(p(x,y))_{x,y\in E}$ gibt, sodass für alle $x_0, \dots, x_n, y \in E$ mit $P_{x_0}(X_0 = x_0, \dots, X_n = x_n) > 0$ gilt

$$P_{x_0}(X_{n+1} = y \mid X_0 = x_0, \dots, X_n = x_n) = p(x_n, y)$$

Beweis. i) Ist (E,d) polnisch, so ist $E^{\mathbb{N}_0}$ auch polnisch, zum Beispiel mit Metrik

$$d_{E^{\mathbb{N}_0}}((x_n),(y_n)) = \sum_{n=1}^{\infty} 2^{-n}(d(x_n,y_n) \wedge 1).$$

Dann ist $\mathscr{B}(E)^{\otimes \mathbb{N}_0} = \mathscr{B}(E^{\mathbb{N}_0})$, demnach gibt es eine Version der bedingten Verteilung von $(X_{n+k})_{n\in\mathbb{N}_0}$ bedingt auf $\mathscr{F}_k$ (siehe Satz 6.6).

ii) Lese den Beweis von Satz 6.23 erneut mit $t_i = i$.

iii) Ist E diskret, so ist $p_{xy} := \kappa(x, \{y\})$ eine stochastische Matrix.

□

Bemerkung 6.25 (Formulierung der Markov-Eigenschaft mittels Shifts)**.** Auf $(E^I, \mathscr{B}(E)^{\otimes I})$ definiert

$$(\Theta_t x)_u := x_{t+u}$$

eine messbare Selbstabbildung und es gilt $\Theta_{t+s} = \Theta_s \circ \Theta_t$. Wenn X auf kanonische Weise auf $(E^I, \mathscr{B}(E)^{\otimes I})$ definiert ist, so formuliert man (6.11) auch als

$$\mathbb{E}_x\left[f(X \circ \Theta_s) \mid \mathscr{F}_s\right] = \mathbb{E}_{X_s}\left[f(X)\right].$$

6.3 Zur starken Markov-Eigenschaft

Definition 6.26. Ein Markov-Prozess $X = (X_t)_{t\in I}$ mit Verteilungen $(P_x)_{x\in E}$ hat die *starke Markov-Eigenschaft*, falls für jede fast sicher endliche Stoppzeit τ, jedes $x \in E$ und jede beschränkte, messbare Funktion $f\colon E^I \to \mathbb{R}$ gilt

$$\mathbb{E}_x\left[f((X_{\tau+t})_{t\in I}) \mid \mathscr{F}_\tau\right] = \mathbb{E}_{X_\tau}\left[f(X)\right] \quad P_x\text{-f.s.}$$

Satz 6.27. *Im Fall $I = \mathbb{N}_0$ besitzt jeder Markov-Prozess die starke Markov-Eigenschaft.*

Beweis. Für $A \in \mathscr{F}_\tau$ und $s \in \mathbb{N}_0$ gilt $A \cap \{\tau = s\} \in \mathscr{F}_s$ und somit

$$\begin{aligned}\mathbb{E}_x\left[f((X_{\tau+t})_{t\in I})\mathbf{1}_A\right] &= \sum_{s\in\mathbb{N}_0} \mathbb{E}_x\left[f((X_{\tau+t})_{t\in I})\mathbf{1}_{A\cap\{\tau=s\}}\right] = \sum_{s\in\mathbb{N}_0} \mathbb{E}_x\left[f((X_{s+t})_{t\in I})\mathbf{1}_{A\cap\{\tau=s\}}\right] \\ &= \sum_{s\in\mathbb{N}_0} \mathbb{E}_x\left[\mathbb{E}_{X_s}[f(X)]\mathbf{1}_{A\cap\{\tau=s\}}\right] = \sum_{s\in\mathbb{N}_0} \mathbb{E}_x\left[\mathbb{E}_{X_\tau}[f(X)]\mathbf{1}_{A\cap\{\tau=s\}}\right] \\ &= \mathbb{E}_x\left[\mathbb{E}_{X_\tau}[f(X)]\mathbf{1}_A\right],\end{aligned}$$

wobei im vorletzten Schritt die schwache Markov-Eigenschaft bei Zeit s eingeht. □

Satz 6.28 ("Spiegelungsprinzip"). *Seien $Y_1, Y_2, \ldots$ unabhängige und identisch verteilte reelle Zufallsvariablen mit $\mathscr{L}(Y_1) = \mathscr{L}(-Y_1)$. Sei $X_0 := 0$ und $X_k := \sum_{i=1}^k Y_i$. Dann gilt für $n \in \mathbb{N}$ und $a > 0$*

$$\mathbb{P}\left(\max_{m\le n} X_m \ge a\right) \le 2\mathbb{P}(X_n \ge a) - \mathbb{P}(X_n = a).$$

Falls $\mathbb{P}(Y_1 \in \{-1, 0, 1\}) = 1$ und $a \in \mathbb{N}$, so gilt Gleichheit.

Beweis. $(X_k)_{k\in\mathbb{N}_0}$ ist eine Markov-Kette (wir betrachten $E = \mathbb{R}$ und $P_x = \mathscr{L}((X_k + x)_{k\in\mathbb{N}_0})$ für $x \in \mathbb{R}$, um wörtlich an Definition 6.13 anzuknüpfen). Sei $a > 0$, $n \in \mathbb{N}$ und

$$\tau := \inf\{m \in \mathbb{N}_0 : X_m \ge a\} \wedge (n+1).$$

Setze $f(m, X) := \mathbf{1}_{\{m\le n\}}\left(\mathbf{1}_{\{X_{n-m}>a\}} + \frac{1}{2}\mathbf{1}_{\{X_{n-m}=a\}}\right)$ und $\varphi(m, z) := \mathbb{E}_z[f(m, (X_k)_{k\in\mathbb{N}_0})]$. Dann gilt aufgrund der symmetrischen Verteilung von Y_1

$$\varphi(m,z) \begin{cases} \ge \frac{1}{2}, & \text{falls } m \le n \text{ und } z > a \\ = \frac{1}{2}, & \text{falls } m \le n \text{ und } z = a \\ = 0, & \text{falls } m > n \end{cases}$$

und

$$f(\tau, (X_{\tau+k})_{k\in\mathbb{N}_0}) = \mathbf{1}_{\{\tau\le n\}}\left(\mathbf{1}_{\{X_n>a\}} + \frac{1}{2}\mathbf{1}_{\{X_n=a\}}\right).$$

Wegen der starken Markov-Eigenschaft gilt $\mathbb{E}_0[f(\tau, (X_{\tau+k})_{k\in\mathbb{N}_0}) \mid \mathscr{F}_\tau] = \varphi(\tau, X_\tau)$. Weiter ist

$$\{\tau \le n\} = \{\tau \le n\} \cap \{X_\tau \ge a\} \subset \{\varphi(\tau, X_\tau) \ge \frac{1}{2}\} \cap \{\tau \le n\} = \{\varphi(\tau, X_\tau) > 0\} \cap \{\tau \le n\},$$

also

$$\mathbb{P}(X_n > a) + \frac{1}{2}\mathbb{P}(X_n = a) = \mathbb{E}_0[f(\tau, (X_{\tau+k})_{k\in\mathbb{N}_0})] \ge \frac{1}{2}\mathbb{P}(\tau \le n) = \frac{1}{2}\mathbb{P}\left(\max_{m\le n} X_m \ge a\right). \quad (6.13)$$

Falls Y_i nur Werte in $\{-1, 0, 1\}$ annimmt und $a \in \mathbb{N}$, dann ist $\{\tau \le n\} = \{X_\tau = a\}$, also

$$\{\varphi(\tau, X_\tau) > 0\} \cap \{\tau \le n\} = \{\varphi(\tau, X_\tau) = \frac{1}{2}\} \cap \{\tau \le n\},$$

das heißt es gilt die Gleichheit in (6.13). □

Beispiel 6.29 (Eine Situation, in der die starke Markov-Eigenschaft nicht gilt). Im Fall $I = [0,\infty)$ fordert man, dass die Filtration $(\mathscr{F}_t)_{t\geq 0}$ rechtsstetig ist, das heißt

$$\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s \quad \text{für } t \geq 0$$

(andernfalls betrachtet man die rechtsstetige Version $(\mathscr{F}_t^+)_{t\geq 0}$ mit $\mathscr{F}_t^+ := \bigcap_{s>t} \mathscr{F}_s$). Sei $E = [0,\infty)$ und eine Markov-Halbgruppe $(\kappa_t)_{t\geq 0}$ gegeben durch

$$\kappa_t(x,A) := \begin{cases} \delta_{x+t}(A)\,, & x > 0 \\ e^{-t}\delta_0(A) + \int_0^t e^{-s}\delta_{t-s}(A)\,ds\,, & x = 0 \end{cases}$$

$(X_t)_{t\geq 0}$ kann folgendermaßen dargestellt werden: Sei $T \sim \mathrm{Exp}(1)$ und für $t \geq 0$

$$X_t := (t - T\mathbf{1}_{\{X_0=0\}})_+ + \mathbf{1}_{\{X_0\neq 0\}}X_0.$$

Sei $\tau := \inf\{t \geq 0 \mid X_t > 0\} = T\mathbf{1}_{\{X_0=0\}}$ (dies *ist* eine Stoppzeit, da $(\mathscr{F}_t)_t$ rechtsstetig ist). Dann gilt (für $f : [0,\infty) \to \mathbb{R}$, $t > 0$)

$$\mathbb{E}_0\left[f(X_{T+t}) \mid \mathscr{F}_T^+\right] = f(t) \neq \int f(y)\,\kappa_t(0,dy) = e^{-t}f(0) + \int_0^t e^{-s}f(t-s)\,ds,$$

das heißt die starke Markov-Eigenschaft gilt nicht für dieses τ.

6.4 Ergänzung: Feller-Halbgruppen und Generatoren

Eine wichtige Klasse von Markovprozessen in stetiger Zeit bilden die sogenannten Feller-Prozesse, deren Übergangshalbgruppen $(T_t)_{t\geq 0}$ (vgl. Bemerkung 6.17) stetige Funktionen in stetige Funktionen überführen (siehe Definition 6.30 unten). Fellerprozesse besitzen stets die starke Markoveigenschaft (skizziert in Satz 6.40 unten). Zudem besitzen sie Versionen, deren Pfade rechtsstetig sind und an jeder Stelle linke Limiten besitzen (sogenannte càdlàg-Pfade, Akronym für frz. continue à droite, limite à gauche), man kann daher für einen Fellerprozess $X = (X_t)_{t\geq 0}$ als Zufallsvariable anstelle des allgemeinen und „unhandlichen" Wertebereichs $E^{[0,\infty)}$ (vgl. Satz 6.16) die Menge $D([0,\infty),E)$ aller càdlàg-Funktionen von $[0,\infty)$ nach E verwenden, siehe Bericht 6.39 unten. Wir behandeln in diesem Unterkapitel (nur) einen etwas skizzenhaften Überblick über die zugehörige Theorie, Details finden sich beispielsweise in [Ka, Ch. 17], [EK, Ch. 1] oder [KW, Kap. 5.4–5.4].

Wir nehmen im Folgenden an: Der Wertebereich ist E polnisch und lokalkompakt, wir betrachten kontinuierliche Zeit, d.h. in Definition 6.13 bzw. Definition 6.15 betrachten wir Zeitindexmenge $I = \mathbb{R}_+$. Wir betrachten als „natürliche" Familie von Testfunktionen auf E

$$C_0(E) = \left\{f : E \to \mathbb{R} \,:\, f \text{ stetig mit } \lim_{x\to\infty} f(x) = 0\right\}$$

(Umgebungen von „∞" sind die Komplemente kompakter Mengen); man überzeugt sich: ausgestattet mit Supremumsnorm $\|f\| := \sup_{x\in E}|f(x)|$ ist $C_0(E)$ ein Banachraum.

Definition 6.30. Eine Markov-Übergangshalbgruppe $(T_t)_{t\geq 0}$ wie in Bemerkung 6.17 heißt eine *Feller-Halbgruppe*, falls gilt

1. $T_t f \in C_0(E)$ für $f \in C_0(E), t \geq 0$

2. $T_t f(x) \underset{t \downarrow 0}{\longrightarrow} f(x), \quad x \in E,\ f \in C_0(E)$

Ein Markovprozess $X = (X_t)_t$, dessen Übergangshalbgruppe diese Eigenschaften hat, heißt ein *Feller-Prozess*.

(Wir werden sehen, dass die Konvergenz in 2. in Definition 6.30 dann tatsächlich gleichmäßig in x gilt, siehe Satz 6.36 unten.)

Beispiel 6.31 (Pseudo-Poisson-Prozess). $\kappa : E \times \mathscr{B}(E) \to [0,1]$ stochastischer Kern, $\lambda \geq 0$, so definiert

$$Af(x) := \lambda \int_E f(y) - f(x)\, \kappa(x, dy), \quad x \in E$$

einen beschränkten linearen Operator auf $C_0(E)$ (beachte $\|Af\| \leq 2\lambda \|f\|$) und es ist

$$A^n f(x) = \lambda^n \int_E f(y) - f(x)\, \kappa^n(x, dy)$$

mit $\kappa^n = n$-fache Verkettung von κ mit sich selbst, vgl. Bemerkung und Definition 6.3, ii).

$T_t := e^{tA}, t \geq 0$ mit

$$T_t f(x) = e^{tA} f(x) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \int_E f(y) - f(x)\, \kappa^n(x, dy), \quad x \in E$$

ist eine (stark stetige) Markov-Halbgruppe (d.h. $\|T_{t+s} f - T_t f\| \to 0$ für $s \to 0$).

Stochastische Interpretation: $(N_t)_{t \geq 0}$ Poissonprozess mit Rate λ, $(X_n)_{n \in \mathbb{N}_0}$ zeitdiskrete Markovkette auf E mit ein-Schritt-Übergangswahrscheinlich gemäß κ (vgl. Bemerkung 6.24), so ist $T_t f(x) = \mathbb{E}_x[f(X_{N_t})]$.

Bemerkung (ein „elementarer" Funktionalanalysis-Blickwinkel). A beschränkter linearer Operator auf $C_0(E)$, d.h. $\|A\| := \sup_{f \in C_0(E), \|f\|=1} \|Af\| < \infty$, so definiert analog zu (6.6) in Beispiel 6.21 die Formel $T_t := e^{tA} := \sum_{n=0}^{\infty} \frac{1}{n!} t^n A^n, t \geq 0$ eine (stark stetige) Halbgruppe beschränkter linearer Operatoren auf $C_0(E)$. Man gewinnt in diesem Fall den „Generator" A aus der so erzeugten Halbgruppe $(T_t)_{t \geq 0}$ zurück durch die Formel

$$Af = \lim_{t \downarrow 0} \frac{1}{t} \big(T_t f - f \big)$$

(der Grenzwert im Sinne der Supremumsnorm existiert in diesem Fall tatsächlich für jedes $f \in C_0(E)$).

Ist umgekehrt eine Feller-Halbgruppe $(T_t)_{t \geq 0}$ gegeben, so ist zumindest auf den ersten Blick nicht klar, ob es wie oben einen solchen „Generator" überhaupt gibt und ob der oben betrachtete Grenzwert existiert. (Tatsächlich muss man f dafür aus einer geeigneten Teilmenge $\mathcal{D} \subset C_0(E)$ wählen, i.A. $\mathcal{D} \subsetneq C_0(E)$, siehe Sätze 6.34 und 6.36.)

Technisch geht man dazu den Weg über die sogenannte Resolvente, siehe die folgende Definition 6.32 unten. Zur Intuition dazu folgende elementare Beobachtung: Für $g \in \mathbb{R}$ ($g < 0$, sagen wir) bildet $b_t = e^{tg}$, $t \geq 0$ eine (multiplikative) stetige Halbgruppe $\subset \mathbb{R}_+$, offenbar ist $b_s b_t = b_{s+t}$. Wir können g aus $(b_t)_{t\geq 0}$ zurückgewinnen durch die Formel

$$g = \lim_{t\downarrow 0} \frac{1}{t}\big(b_t - b_0\big) = \lim_{t\downarrow 0} \frac{1}{t}\big(e^{tg} - 1\big)$$

Andererseits ist aber auch für jedes $\lambda > 0$

$$\int_0^\infty e^{-\lambda t} e^{gt}\, dt = (\lambda - g)^{-1}$$

und wir könnten prinzipiell aus dieser Formel g extrahieren (ohne eine Ableitung bilden zu müssen).

Definition und Beobachtung 6.32 (Resolvente)**.** Für $\lambda > 0$

$$R_\lambda f(x) := \int_0^\infty e^{-\lambda t} T_t f(x)\, dt \tag{6.14}$$

heißt R_λ die (λ-)Resolvente (von $(T_t)_{t\geq 0}$, diese ist zumindest wohldefiniert, wenn f beschränkt und messbar).

Für $f \in C_0(E)$ gilt $R_\lambda f \in C_0(E)$ (verwende dominerte Konvergenz, um zu sehen, dass $x \mapsto \int_0^\infty e^{-\lambda t} T_t f(x)\, dt$ stetig ist und für $x \to \infty$ gegen 0 konvergiert).

$R_\lambda : C_0(E) \to C_0(E)$ ist ein beschränkter linearer Operator: es ist für $x \in E$

$$|R_\lambda f(x)| \leq \int_0^\infty e^{-\lambda t} \|T_t f\|\, dt \leq \int_0^\infty e^{-\lambda t} \|f\|\, dt = \frac{1}{\lambda}\|f\|$$

(denn jedes T_t ist eine Kontraktion auf $C_0(E)$) und somit

$$\|R_\lambda\| = \sup_{f\in C_0(E), f\not\equiv 0} \frac{\|R_\lambda f\|}{\|f\|} \leq \frac{1}{\lambda}$$

Lemma 6.33 (Resolventengleichung)**.** *Für* $\lambda, \mu > 0$ *ist*

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu \tag{6.15}$$

Beweis. Sei $\lambda \neq \mu$ (für $\lambda = \mu$ gilt (6.15) trivialerweise), $f \in C_0(E)$

$$\begin{aligned}
R_\lambda R_\mu f(x) &= \int_0^\infty e^{-\lambda t} T_t(R_\mu f)(x)\, dt = \int_0^\infty e^{-\lambda t} T_t\Big(\textstyle\int_0^\infty e^{-\mu s} T_s f\, ds\Big)(x)\, dt \\
&= \int_0^\infty \int_0^\infty e^{-\lambda t} e^{-\mu s} T_{t+s} f(x)\, ds\, dt = \int_0^\infty \int_0^\infty e^{-(\mu-\lambda)s} e^{-\lambda(t+s)} T_{t+s} f(x)\, dt\, ds \\
&= \int_0^\infty e^{-(\mu-\lambda)s} \int_s^\infty e^{-\lambda u} T_u f(x)\, du\, ds = \int_0^\infty e^{-\lambda u} T_u f(x) \int_0^u e^{-(\mu-\lambda)s}\, ds\, du \\
&= \int_0^\infty e^{-\lambda u} T_u f(x) \frac{1 - e^{-(\mu-\lambda)u}}{\mu - \lambda}\, du \\
&= \frac{1}{\mu - \lambda} \int_0^\infty e^{-\lambda u} T_u f(x)\, du - \frac{1}{\mu - \lambda} \int_0^\infty e^{-\mu u} T_u f(x)\, du
\end{aligned}$$

$\square$

Satz 6.34. *1. λR_λ is ein injektiver Kontraktionsoperator auf $C_0(E)$ und*

$$\|\lambda R_\lambda f - f\| \underset{\lambda\to\infty}{\longrightarrow} 0 \quad \textit{für } f \in C_0(E)$$

2. *Das Bild $\mathcal{D} = \{R_\lambda f : f \in C_0(E)\} \subset C_0(E)$ liegt dicht in $C_0(E)$ und hängt nicht von $\lambda > 0$ ab.*

3. *Es gibt einen Operator A mit Definitionsbereich $\mathcal{D}$, so dass*

$$R_\lambda^{-1} = \lambda I - A \quad \textit{für } \lambda > 0$$

4. *A und T_t kommutieren für $t \geq 0$.*

Man nennt den Operator A mit Definitionsbereich $\mathcal{D}$ aus Satz 6.34, 3. den *Generator* oder *Erzeuger* (auch: *infinitesimaler Erzeuger*) der Übergangshalbgruppe $(T_t)_{t\geq 0}$.

Beweisskizze. Wir beobachten zunächst:

a) Es ist

$$\|\lambda R_\lambda f\| \leq \int_0^\infty \lambda e^{-\lambda t}\|T_t f\|\, dt \leq \int_0^\infty \lambda e^{-\lambda t}\|f\|\, dt = \|f\|$$

d.h. λR_λ ist Kontraktion.

b) Die Resolventengleichung (6.15) zeigt, dass R_λ und R_μ kommutieren, und dass

$$R_\lambda(C_0(E)) = R_\mu(C_0(E)) =: \mathcal{D}$$

unabhängig von $\lambda > 0$ ist.

c) Für $f = R_1 g$ mit $g \in C_0(E)$ ist mit (6.15)

$$\|\lambda R_\lambda f - f\| = \left\|(\lambda R_\lambda - I)R_1 g\right\| = \left\|(R_1 - I)R_\lambda g\right\| \leq \|R_1 - I\|\,\|R_\lambda g\| \leq \|R_1 - I\|\frac{1}{\lambda}\|g\| \underset{\lambda\to\infty}{\longrightarrow} 0$$

Demnach gilt

$$\|\lambda R_\lambda f - f\| \underset{\lambda\to\infty}{\longrightarrow} 0 \quad \text{für } f \in \mathcal{D}$$

Approximation bezüglich der Supremumsnorm liefert, dass dies auch für $f \in \overline{\mathcal{D}}$ (Abschluss bezüglich der Supremumsnorm) gilt.

1. Nach a) ist λR_λ eine Kontraktion und nach c) gilt die behauptete Formel.

Zur Injektivität von R_λ: Sei $f \in C_0(E)$ und $\lambda_0 > 0$ mit $R_{\lambda_0} f = 0$. Mit der Resolventengleichung (6.15) folgt dann $R_\lambda f - 0 = (\lambda - \lambda_0)R_\lambda R_{\lambda_0} f = 0$, d.h. $R_\lambda f = 0$ für alle $\lambda > 0$. Wegen $\|\lambda R_\lambda f - f\|_{\lambda\to\infty} \to 0$ folgt $f \equiv 0$. Somit ist R_λ injektiv und es gibt $R_\lambda^{-1} : \mathcal{D} \to C_0(E)$ mit $R_\lambda^{-1}R_\lambda = I$ (auf $C_0(E)$) und $R_\lambda R_\lambda^{-1} = I$ (auf $\mathcal{D}$)

2. Nach b) hängt $\mathcal{D} = R_\lambda(C_0(E))$ nicht von $\lambda > 0$ ab. Dichtheit des Bilds, d.h. $\overline{\mathcal{D}} = C_0(E)$ benötigt etwas „Funktionalanalysis-Maschinerie", vgl. z.B. [Ka, S. 371]. Falls $\overline{\mathcal{D}} \neq C_0(E)$, so gäbe es (mit Satz von Hahn-Banach und dem Rieszschen Darstellungssatz für lineare Funktionale auf

$C_0(E)$) ein (nicht-triviales) signiertes Maß μ auf E (strenggenommen muss man ggf. zur ein-Punkt-Kompaktifizierung von E übergehen) mit $\int_E R_\lambda f \, d\mu = 0$ für alle $f \in C_0(E)$ und $\lambda > 0$ sowie eine Funktion $f \in C_0(E)$ mit $\int f \, d\mu \neq 0$. Dann wäre aber

$$0 = \int_E \lambda R_\lambda f \, d\mu = \int_E \int_0^\infty \lambda e^{-\lambda t} T_f f(x) \, dt \, \mu(dx) \underset{\lambda\to\infty}{\longrightarrow} \int_E f(x) \, \mu(dx) \neq 0$$

ein Widerspruch.

3. Die Resolventengleichung (6.15) liefert $R_\lambda^{-1} f - R_\mu^{-1} f = (\mu - \lambda) f$ für $f \in \mathcal{D}$ (wende auf (6.15) von links R_λ^{-1} und von rechts R_μ^{-1} an), somit ist

$$A := \lambda I - R_\lambda^{-1} = \mu I - R_\mu^{-1}$$

unabhängig von λ auf $\mathcal{D}$ (wohl-)definiert.

4. T_t und R_λ kommutieren für $t \geq 0$, $\lambda > 0$, daher ist

$$T_t(\lambda I - A) R_\lambda = T_t = (\lambda I - A) R_\lambda T_t = (\lambda I - A) T_t R_\lambda \qquad \text{und somit} \qquad T_t A R_\lambda = A T_t R_\lambda$$

□

Beobachtung 6.35. $(T_t)_{t\geq 0}$ ist durch A eindeutig bestimmt: $R_\lambda = (\lambda I - A)^{-1}$ ist durch $(A, \mathcal{D})$ festgelegt und somit für nicht-negatives $f \in C_0(E)$ und $x \in E$ das Maß $\mu_{f,x}$ auf $\mathbb{R}_+$ mit $\mu_{f,x}(dt) = T_t f(x) \, dt$, da seine Laplace-Transformierte gerade $R_\lambda f(x)$ ist. Da $t \mapsto T_t f(x)$, die Dichte von $\mu_{f,x}$, stetig ist, ist sie durch $\mu_{f,x}$ festgelegt: $T_t f(x) = \lim_{\varepsilon\downarrow 0} \mu_{f,x}([t, t+\varepsilon))/\varepsilon$.

Der folgende Satz verallgemeinert die Vorwärts- und Rückwärtsgleichungen aus dem Kontext von Beispiel 6.21 mit endlichem E (vgl. (6.10) und (6.9)) auf den allgemeinen Fall.

Satz 6.36. *1.* (T_t) *ist stark stetig, d.h.* $\|T_t f - f\| \to 0$ *für* $t \downarrow 0$*, es ist*

$$T_t f - f = \int_0^t T_u A f \, du, \quad f \in \mathcal{D}, \ t \geq 0$$

2. $T_t f$ *ist differenzierbar in* $t = 0$ *für* $f \in \mathcal{D}$ *und*

$$\frac{d}{dt}(T_t f) = T_t A f = A T_t f, \quad f \in \mathcal{D}, \ t \geq 0$$

Beweisskizze. Ein entscheidendes Werkzeug ist die „Yosida-Approximation" (für $\lambda > 0$)

$$A^{(\lambda)} := \lambda A R_\lambda = \lambda(\lambda I - R_\lambda^{-1}) R_\lambda = \lambda(\lambda \mathbb{R}_\lambda - I) \tag{6.16}$$

von A. Offenbar ist $A^{(\lambda)}$ auf ganz $C_0(E)$ definiert, und $A^{(\lambda)}$ ist der Generator eines Pseudo-Poisson-Prozesses in Sinne von Beispiel 6.31. Der zugehörige Prozess führt mit Rate λ jeweils Sprünge gemäß T_τ aus, wobei $\tau \sim \mathrm{Exp}(\lambda)$ (in der Notation von Beispiel 6.31 wähle dort $\kappa(x, A) = \int_0^\infty \lambda e^{-\lambda t} T_t \mathbf{1}_A(x) \, dt$).

Sei $(T_t^{(\lambda)})_{t\geq 0}$ die zugehörige Übergangshalbgruppe ($T_t^{(\lambda)} = e^{tA^{(\lambda)}}$ wie in Beispiel 6.31), es gilt

$$\frac{1}{h}\big(T_h^{(\lambda)} - I\big) \underset{h\downarrow 0}{\longrightarrow} A^{(\lambda)} \tag{6.17}$$

(Konvergenz im Sinne der Operatornorm), siehe explizite die Konstruktion in Beispiel 6.31.

a) Mit Satz 6.34, 1. folgt für $f \in \mathcal{D}$

$$A^{(\lambda)}f = \lambda R_\lambda(Af) \xrightarrow[\lambda\to\infty]{} Af \tag{6.18}$$

b) Zeige: für $f \in C_0(E), t \geq 0$ ist $T_t^{(\lambda)}f, \lambda \in (0,\infty)$ eine Cauchy-Folge (bezüglich Supremumsnorm) für $\lambda \to \infty$.

Sei zunächst $f \in \mathcal{D}, \mu, \lambda, t > 0$:

$$T_t^{(\lambda)}f - T_t^{(\mu)}f = (T_{t/n}^{(\lambda)})^n f - (T_{t/n}^{(\mu)})^n f = \sum_{j=0}^{n-1} \left((T_{t/n}^{(\lambda)})^{n-1-j}(T_{t/n}^{(\mu)})^j\right)\left(T_{t/n}^{(\lambda)}f - T_{t/n}^{(\mu)}f\right)$$

(beachte: $T_{t/n}^{(\lambda)}$ und $T_{t/n}^{(\mu)}$ kommutieren), daher gilt

$$\begin{aligned}
\|T_t^{(\lambda)}f - T_t^{(\mu)}f\| &\leq \Big\|\sum_{j=0}^{n-1} \left((T_{t/n}^{(\lambda)})^{n-1-j}(T_{t/n}^{(\mu)})^j\right)\Big\| \|T_{t/n}^{(\lambda)}f - T_{t/n}^{(\mu)}f\| \\
&\leq n\|T_{t/n}^{(\lambda)}f - T_{t/n}^{(\mu)}f\| \\
&= t\left\|\frac{T_{t/n}^{(\lambda)}f - f}{t/n} - \frac{T_{t/n}^{(\mu)}f - f}{t/n}\right\| \xrightarrow[n\to\infty]{} t\|A^{(\lambda)}f - A^{(\mu)}f\|
\end{aligned} \tag{6.19}$$

(beachte: $T_{t/n}^{(\lambda)}$ und $T_{t/n}^{(\mu)}$ sind Kontraktionsoperatoren; wir verwenden (6.17) im letzten Schritt). Somit bildet wegen (6.18) tatsächlich $T_t^{(\lambda)}f, \lambda > 0$ eine Cauchy-Folge für $\lambda \to \infty$ und wir halten fest: die Konvergenz ist gleichmäßig für beschränkte t.

Da $\mathcal{D}$ dicht liegt in $C_0(E)$ (und alle beteiligten linearen Operatoren beschränkt und somit stetig sind), gilt diese Konvergenz auch für beliebiges $f \in C_0(E)$. Wir bezeichnen

$$\widetilde{T}_t f := \lim_{\lambda\to\infty} T_t^{(\lambda)}f, \quad f \in C_0(E) \tag{6.20}$$

$(\widetilde{T}_t)_{t\geq 0}$ „erbt“ die Kontraktions- und Halbgruppeneigenschaften der $(T_t^{(\lambda)})$.

c) Wir prüfen, dass $T_t = \widetilde{T}_t$ für $t \geq 0$ gilt, indem wir verifizieren, dass beide Halbgruppen dieselbe Resolvente $(R_\lambda)_{\lambda>0}$ besitzen; dann folgt die Behauptung aus Beobachtung 6.35.

Es ist

$$\begin{aligned}
\left\|T_t^{(\mu)}\mu R_\mu f - \widetilde{T}_t f\right\| &= \left\|T_t^{(\mu)}(\mu R_\mu f - f) + T_t^{(\mu)}f - \widetilde{T}_t f\right\| \\
&\leq \left\|\mu R_\mu f - f\right\| + \left\|T_t^{(\mu)}f - \widetilde{T}_t f\right\| \xrightarrow[\mu\to\infty]{} 0
\end{aligned}$$

(beachte $T_t^{(\mu)}$ ist Kontraktion) und für $\lambda > 0$

$$\begin{aligned}
\int_0^\infty e^{-\lambda t}T_t^{(\mu)}\mu R_\mu f\,dt &= \mu(\lambda I - A^{(\mu)})^{-1}R_\mu f \\
&= \mu\left((\lambda+\mu)I - \mu^2 R_\mu\right)^{-1}R_\mu f = \frac{\mu}{\lambda+\mu}R_\nu f
\end{aligned} \tag{6.21}$$

mit $\nu = \nu(\lambda, \mu) = (\lambda\mu)/(\lambda + \mu)$. Für das letzte Gleichheitszeichen beachte

$$\begin{aligned}\big((\lambda+\mu)I - \mu^2 R_\mu\big)R_\nu &= (\lambda+\mu)R_\nu - \mu^2 R_\mu R_\nu \\ &= (\lambda+\mu)R_\nu - \mu^2 \frac{R_\mu - R_\nu}{\nu - \mu} = (\lambda+\mu)R_\mu\end{aligned}$$

gemäß Resolventengleichung (siehe Lemma 6.33). $\mu \to \infty$ in (6.21) liefert somit

$$\int_0^\infty e^{-\lambda t}\widetilde{T}_t f\, dt = R_\lambda f$$

für $\lambda > 0$ wie behauptet (beachte $\nu = \nu(\lambda, \mu) \to \lambda$ für $\mu \to \infty$).

d) Es gilt

$$\|T_t^{(\lambda)} f - f\| \xrightarrow[t\downarrow 0]{} 0 \quad \text{für jedes } \lambda > 0$$

(via explizite stochastische Konstruktion wie in Beispiel 6.31: die Anzahl Sprünge bis t ist $\mathrm{Pois}(\lambda t)$, etc.), mit (6.20) oben gilt daher auch $\|T_t f - f\| \to 0$ für $t \downarrow 0$. Demnach ist $(T_t)_{t\ge 0}$ stark stetig wie in 1. behauptet.

e) (6.17) oben liefert $(T_h^{(\lambda)} - I)/h \to A^{(\lambda)}$ mit $h \downarrow 0$ für jedes $\lambda > 0$ und allgemein gilt für $t > 0$

$$\frac{d}{dt}T_t^{(\lambda)} = A^{(\lambda)}T_t^{(\lambda)} = T_t^{(\lambda)}A^{(\lambda)}$$

(verwende Halbgruppen-Eigenschaft und starke Stetigkeit von $(T_t^{(\lambda)})$; $T_t^{(\lambda)}$ und $A^{(\lambda)}$ kommutieren nach Satz 6.34) somit auch

$$T_t^{(\lambda)} f - f = \int_0^t T_u^{(\lambda)} A^{(\lambda)} f\, du$$

für $f \in \mathcal{D}$. Mit Approximation von T_t durch $T_t^{(\lambda)}$ und von A durch $A^{(\lambda)}$ (siehe (6.20) und (6.18)) ist

$$\begin{aligned}\|T_u^{(\lambda)}A^{(\lambda)}f - T_u A f\| &\le \|T_u^{(\lambda)}A^{(\lambda)}f - T_u A^{(\lambda)}f\| + \|T_u(A^{(\lambda)} - A)f\| \\ &\le \|T_u^{(\lambda)}A^{(\lambda)}f - T_u A^{(\lambda)}f\| + \|(A^{(\lambda)} - A)f\| \xrightarrow[\lambda\to\infty]{} 0\end{aligned}$$

gleichmäßig in $u \in [0, T]$, damit folgt aus obigem auch

$$T_t f - f = \int_0^t T_u A f\, du \qquad \text{bzw.} \qquad T_{t+h} f - T_t f = \int_t^{t+h} T_u A f\, du$$

Differentation liefert $\frac{d}{dt}T_t = T_t A$ und es gilt $T_t A = A T_t$ nach Satz 6.34. □

Bemerkung 6.37. Tatsächlich charakterisisiert die Eigenschaft 2.) aus Satz 6.36 den Definitionsbereich $\mathcal{D}$ des Generators, d.h. es gilt

$$\mathcal{D} = \left\{ f \in C_0(E) \,:\, \lim_{h\downarrow 0} \frac{1}{h}\big(T_h f - f\big) \text{ existiert in } C_0(E) \right\}$$

Siehe dazu z.B. [Ka, Beweis von Thm. 17.6].

Beobachtung 6.38. Sei $(X_t)_{t\geq 0}$ Markov-Prozess (gemäß Satz 6.16) zur Feller-Übergangshalbgruppe $(T_t)_t$ mit Generator A. Dann ist für $f \in \mathcal{D}$

$$M_t := f(X_t) - f(X_0) - \int_0^t Af(X_s)\, ds, \quad t \geq 0$$

ein Martingal (bezüglich $(\mathscr{F}_t)_t$ mit $\mathscr{F}_t = \sigma(X_s, s \leq t)$):

$$\begin{aligned} \mathbb{E}_x[M_{t+h} - M_t \,|\, \mathscr{F}_t] &= \mathbb{E}_x[f(X_{t+h}) - f(X_t) \,|\, \mathscr{F}_t] - \int_t^{t+h} \mathbb{E}_x[Af(X_s) \,|\, \mathscr{F}_t]\, ds \\ &= T_h f(X_t) - f(X_t) - \int_t^{t+h} T_s Af(X_t)\, ds = 0 \end{aligned}$$

nach Satz 6.36.

Weiter ist für $f \in C_0(E)$, $f \geq 0$ der Prozess $Y_t := e^{-t} R_1 f(X_t)$, $t \geq 0$ ein (nicht-negatives) Supermartingal:

$$\begin{aligned} \mathbb{E}_x[Y_{t+h} \,|\, \mathscr{F}_t] &= e^{-t(t+h)} T_h R_1 f(X_t) = e^{-t(t+h)} \int_0^\infty e^{-u} T_{h+u} f(X_t)\, du \\ &= e^{-t} \int_h^\infty e^{-s} T_s f(X_t)\, ds \leq e^{-t} R_1 f(X_t) \end{aligned}$$

Bericht 6.39. Jeder Feller-Prozess besitzt eine Version mit càdlàg-Pfaden. Die entscheidende Beobachtung ist, dass Beobachtung 6.38 für eine dicht liegende Menge von Testfunktionen Supermartingale bereitstellt; man greift dann auf eine sehr allgemeine Regularitätseigenschaft zeitkontinuierlicher Supermartingale zurück. Siehe z.B. [Ka, Theorem 17.15] für Details.

Die Markov-Eigenschaft gilt auch, wenn man wie in Beispiel 6.29 die Filtration rechtsstetig vervollständigt, d.h. zu $\mathscr{F}_t^+ := \bigcap_{s>t} \mathscr{F}_s$ übergeht (was wir ggf. im Folgenden stillschweigend tun werden).

Satz 6.40. *Jeder Feller-Prozess $(X_t)_{t\geq 0}$ (mit càdlàg-Pfaden) besitzt die starke Markov-Eigenschaft.*

Beweisskizze. Sei τ f.s. endliche Stoppzeit. $\tau_n := \frac{1}{n}\lceil n\tau \rceil$ ist ebenfalls Stoppzeit und $\tau_n \searrow \tau$ für $n \to \infty$. Für $A \in \mathscr{F}_\tau^+$ ($\subset \mathscr{F}_{\tau_n}^+$ für jedes n) und $f \in C_0(E)$ ist

$$\mathbb{E}_x\big[\mathbf{1}_A f(X_{\tau+h})\big] = \lim_{n\to\infty} \mathbb{E}_x\big[\mathbf{1}_A f(X_{\tau_n+h})\big] = \lim_{n\to\infty} \mathbb{E}_x\big[\mathbf{1}_A \mathbb{E}_{X_{\tau_n}}[f(X_h)]\big] = \mathbb{E}_x\big[\mathbf{1}_A \mathbb{E}_{X_\tau}[f(X_h)]\big]$$

(Wir verwenden für die mittlere Gleichung Satz 6.27: da τ_n nur Werte auf $\frac{1}{n}\mathbb{N}_0$ annimmt, können wir hier zum zeitdiskreten Prozess $(X_{k/n})_{k\in\mathbb{N}_0}$ übergehen; die beiden anderen Gleichungen verwenden die Rechtsstetigkeit der Pfade von X, die Feller-Eigenschaft und dominierte Konvergenz.) □

Analog zum Fall zeitdiskreter Markovketten heißt ein W'maß $\mu \in \mathcal{M}_1(E)$ eine invariante Verteilung für den Feller-Prozess $X = (X_t)_{t\geq 0}$, wenn $\mathbb{E}_\mu[f(X_t)] = \mathbb{E}_\mu[f(X_0)]$ für jedes $t \geq 0$ und jedes $f \in C_0(E)$ gilt. Invariante Verteilungen für diskrete Markovketten lösen bekanntlich ein Gleichungssystem, das auf den Übergangswahrscheinlichkeiten basiert (siehe z.B. Abschnitt 6.5.3 unten). Das Analogon für Feller-Prozesse ist die folgende Charakterisierung mittels des Generators.

Beobachtung 6.41. $\mu \in \mathcal{M}_1(E)$ ist genau dann eine invariante Verteilung für X, wenn

$$\int_E Ag(x)\,\mu(dx) = 0 \quad \text{für alle } g \in \mathcal{D}$$

gilt.

Falls obige Bedingung erfüllt ist, so zeigt Beobachtung 6.38, dass zumindest für $f \in \mathcal{D}$

$$\begin{aligned}
0 &= \mathbb{E}_\mu[f(X_t)] - \mathbb{E}_\mu[f(X_0)] - \int_0^t \mathbb{E}_\mu[Af(X_s)]\,ds \\
&= \mathbb{E}_\mu[f(X_t)] - \mathbb{E}_\mu[f(X_0)] - \int_0^t \int_E T_s Af(x)\,\mu(dx)\,ds \\
&= \mathbb{E}_\mu[f(X_t)] - \mathbb{E}_\mu[f(X_0)] - \int_0^t \int_E AT_s f(x)\,\mu(dx)\,ds = \mathbb{E}_\mu[f(X_t)] - \mathbb{E}_\mu[f(X_0)]
\end{aligned}$$

(wobei wir im dritten Schritt $T_s A = AT_s$ nach Satz 6.34 verwenden). Dichtheit von $\mathcal{D} \subset C_0(E)$ zeigt die Invarianz von μ.

Ist andererseits $\mu \in \mathcal{M}_1(E)$ invariant, so ist für $g \in \mathcal{D}$

$$\begin{aligned}
\int_E Ag(x)\,\mu(dx) &= \int_E \lim_{h\downarrow 0} \frac{1}{h}\big(T_h g(x) - g(x)\big)\,\mu(dx) \\
&= \lim_{h\downarrow 0} \frac{1}{h} \int_E \big(T_h g(x) - g(x)\big)\,\mu(dx) = \lim_{h\downarrow 0} \frac{1}{h}\big(\mathbb{E}_\mu[g(X_h)] - \mathbb{E}_\mu[g(X_0)]\big) = 0
\end{aligned}$$

nach Satz 6.36.

Manchmal ist der volle Definitionsbereich $\mathcal{D}$ des Generators A „unhandlich“ bzw. schwer explizit zu machen. Dann ist es zum Argumentieren angenehm, sich auf eine kleinere Menge $\widetilde{\mathcal{D}} \subset \mathcal{D}$ von Testfunktionen, einen sogenannten “core”, zurückziehen zu können (z.B. eine Menge von Funktionen mit „guten Glattheitseigenschaften“), die aber immer noch reichhaltig genug ist, um den Generator A zu charakterisieren (und damit auch die zugehörige Übergangshalbgruppe $(T_t)_{t\geq 0}$).

Definition 6.42. Ein linearer Operator A mit Definitionsbereich (englisch “domain”) $\mathcal{D} \subset C_0(E)$ heißt *abgeschlossen*, wenn sein Graph $G_A := \{(f, Af) : f \in \mathcal{D}\} \subset C_0(E)^2$ abgeschlossen in $C_0(E)^2$ ist. A heißt *abschließbar*, wenn $\overline{G_A} \subset C_0(E)^2$ einen (eindeutigen, engl. “single-valued”) Operator $\overline{A}$ definiert (d.h. $(f_1, g), (f_2, g) \in \overline{G_A} \Rightarrow f_1 = f_2$).

Da A linear ist, gilt:

$$A \text{ abschließbar g.d.w.} \quad \mathcal{D} \ni f_n \to 0 \text{ und } Af_n \to g \;\;\Rightarrow\;\; g = 0$$

Für abgeschlossenes A heißt $\widetilde{\mathcal{D}} \subset \mathcal{D}$ ein core, wenn der Abschluss von $A|_{\widetilde{\mathcal{D}}}$ gleich A ist.

Bericht 6.43. Für den Generator A einer Feller-Übergangshalbgruppe wie in Satz 6.34 und beliebiges $\lambda > 0$ gilt

1. $(A, \mathcal{D})$ *ist* abgeschlossen.
2. $\widetilde{\mathcal{D}}$ core für $A \iff (\lambda I - A)\widetilde{\mathcal{D}}$ dicht in $C_0(E)$

3. $\widetilde{\mathcal{D}} \subset \mathcal{D}$ dichter linearer Unterraum und $T_t\widetilde{\mathcal{D}} \subset \widetilde{\mathcal{D}}$ für alle $t \geq 0$ ($\widetilde{\mathcal{D}}$ ist „invariant“)
$\Rightarrow$ $\widetilde{\mathcal{D}}$ ist ein core für A

Siehe z.B. [Ka, Lemma 17.8 und Prop. 17.9].

Die Frage, welche linearen Operatoren überhaupt Generatoren von Feller-Halbgruppen sind, klärt (wenigstens prinzipiell) der folgende Satz.

Satz 6.44 (Hille-Yosida-Theorem[3]). *Ein linearer Operator $A : \widetilde{\mathcal{D}} \to C_0(E)$ auf $\widetilde{\mathcal{D}} \subset C_0(E)$ ist abschließbar und sein Abschluss $\overline{A}$ erzeugt eine (möglicherweise sub-Markov-)Feller-Halbgruppe $(T_t)_{t\geq 0}$ auf $C_0(E)$ g.d.w.*

1. *$\widetilde{\mathcal{D}}$ ist dicht in $C_0(E)$*
2. *das Bild $\mathcal{R}(\lambda_0 I - A) := \{\lambda_0 f - Af : f \in \widetilde{\mathcal{D}}\}$ von $\lambda_0 I - A$ ist dicht in $C_0(E)$ für ein $\lambda_0 > 0$*
3. *A erfüllt das positive-Maximum-Prinzip: Für alle $x \in E$ und $f \in \widetilde{\mathcal{D}}$ gilt*

$$0 \leq f(x) = \max_{y \in E} f(y) \implies Af(x) \leq 0$$

Beweisskizze. “$\Rightarrow$“: Falls A eine Feller-Halbgruppe erzeugt und $0 \leq f(x) = \max_{y\in E} f(y)$ gilt, so gilt $T_t f(z) \leq \max_{y\in E} f(y) = f(x)$ für jedes $z \in E$ und somit $Af(x) = \lim_{h\downarrow 0}(T_h f(x) - f(x))/h \leq 0$; die beiden anderen Eigenschaften gelten nach Satz 6.34.

“$\Leftarrow$“: A erfülle 1., 2. und 3. oben.

Schritt i) Entscheidende Beobachtung: A erfüllt das positive-Maximum-Prinzip $\Longrightarrow$ A is *dissipativ*, d.h. $\|(\lambda I - A)f\| \geq \lambda\|f\|$ für $f \in \widetilde{\mathcal{D}}$, $\lambda > 0$. Sei dazu $f \in \widetilde{\mathcal{D}}$, wähle $x \in E$ mit $|f(x)| = \|f\|$, setze $g := (\mathrm{sign} f(x))\, f$. Es ist $g \in \widetilde{\mathcal{D}}$ und $0 \leq g(x) = \|g\|$, somit $Ag(x) \geq 0$ und

$$\|(\lambda I - A)f\| \geq \lambda g(x) - Ag(x) \geq \lambda g(x) = \lambda\|f\| \tag{6.22}$$

Zeige: A ist abschließbar. Seien dazu $f_n \in \widetilde{\mathcal{D}}$ mit $f_n \to 0$ und $Af_n \to g \in C_0(E)$ gegeben. Wähle (gemäß Voraussetzung 1.) eine Folge $(g_m)_{m\in\mathbb{N}} \subset \widetilde{\mathcal{D}}$ mit $g_m \to g$, es ist nach (6.22)

$$\left\|(\lambda I - A)(g_m + \lambda f_n)\right\| \geq \lambda\|g_m + \lambda f_n\| \quad \lambda > 0, m, n \in \mathbb{N}$$

mit $n \to \infty$ (und Division durch λ) folgt $\|(I - \lambda^{-1}A)g_m - g\| \geq \|g_m\|$, mit $\lambda \to \infty$ dann $\|g_m - g\| \geq \|g_m\|$, mit $m \to \infty$ somit $0 = \|0\| \geq \|g\|$ wie behauptet.

Der Abschluss $\overline{A}$ (mit Definitionsbereich $\mathrm{dom}(\overline{A})$) von A ist ebenfalls dissipativ: Wir approximieren ein beliebiges $f \in C_0(E)$ gemäß Voraussetzung 1. durch eine Folge $(f_n)_n \subset \widetilde{\mathcal{D}}$ und erhalten aus (6.22)

$$\|(\lambda I - \overline{A})f\| \geq \lambda\|f\|, \quad f \in C_0(E), \lambda > 0 \tag{6.23}$$

[3]nach Carl Einar Hille, 1894–1980 und Kōsaku Yosida, 1909–1990. E. Hille, Functional analysis and semi-groups. American Mathematical Society Colloquium Publications, 31, (1948); K. Yosida, On the differentiability and the representation of one-parameter semi-group of linear operators, J. Math. Soc. Japan 1, No. 1, 15-21 (1948).

Schritt ii) Zeige:

$$\Lambda := \left\{\lambda > 0 : \text{Bild von } \lambda I - \overline{A} \text{ ist gleich } C_0(E)\right\} = (0, \infty)$$

Dazu zeige:

ii), a) Λ ist abgeschlossen. Seien $\Lambda \ni \lambda_n \underset{n\to\infty}{\longrightarrow} \lambda \in (0,\infty)$, $g \in C_0(E)$. Dann liegt

$$g_n := (\lambda I - \overline{A})(\lambda_n I - \overline{A})^{-1} g$$

im Bild $\mathcal{R}(\lambda I - \overline{A})$ und

$$\|g_n - g\| = \left\|(\lambda - \lambda_n)(\lambda_n I - \overline{A})^{-1} g\right\| \le \frac{|\lambda - \lambda_n|}{\lambda_n}\|g\| \underset{n\to\infty}{\longrightarrow} 0$$

(wende (6.23) an mit $f = (\lambda_n I - \overline{A})^{-1} g$), d.h. $\mathcal{R}(\lambda I - \overline{A})$ liegt dicht in $C_0(E)$.

Da $\overline{A}$ ein abgeschlossener Operator (nach Definition) und dissipativ ist, ist $\mathcal{R}(\lambda I - \overline{A}) \subset C_0(E)$ auch abgeschlossen: Seien $\mathrm{dom}(\overline{A}) \ni f_n$ mit $\lim_{n\to\infty}(\lambda I - \overline{A}) f_n = h \in C_0(E)$ für $n \to \infty$, so ist mit (6.23)

$$\|f_n - f_m\| \le \frac{1}{\lambda}\left\|(\lambda I - \overline{A}) f_n - (\lambda I - \overline{A}) f_m\right\| \underset{m,n\to\infty}{\longrightarrow} 0$$

Demnach ist $(f_n)_n$ Cauchy-Folge und es gibt $f \in C_0(E)$ mit $f_n \to f$ für $n \to \infty$. Da $\overline{A}$ abgeschlossen ist, ist $(\lambda I - \overline{A}) f = \lim_{n\to\infty}(\lambda I - \overline{A}) f_n = h$, d.h. $h \in \mathcal{R}(\lambda I - \overline{A})$.

Ingesamt folgt: $\mathcal{R}(\lambda I - \overline{A}) = C_0(E)$, d.h. $\lambda \in \Lambda$.

Zudem zeigt obiges: $R_\lambda := (\lambda I - \overline{A})^{-1}$ ist für $\lambda \in \Lambda$ definiert und ein beschränkter linearer Operator auf $C_0(E)$ mit $\|R_\lambda\| \le 1/\lambda$:

$$\|R_\lambda f\| = \left\|(\lambda I - \overline{A})^{-1} f\right\| \le \frac{1}{\lambda}\|f\| \tag{6.24}$$

ii), b) Λ ist offen. Sei $\lambda \in \Lambda$, für $\mu > 0$ mit $|\mu - \lambda|\|R_\lambda\| < 1$ ist (vgl. (6.24))

$$\widetilde{R}_\mu := \sum_{n=0}^{\infty} (\lambda - \mu)^n R_\lambda^{n+1}$$

ein beschränkter linearer Operator auf $C_0(E)$ und für beliebiges $f \in C_0(E)$

$$(\mu - \overline{A})\widetilde{R}_\mu f = (\lambda - \overline{A})\widetilde{R}_\mu f - (\lambda - \mu)\widetilde{R}_\mu f = \sum_{n=0}^{\infty}(\lambda - \mu)^n R_\lambda^n f - \sum_{n=0}^{\infty}(\lambda - \mu)^{n+1} R_\lambda^{n+1} f = f$$

d.h. $\mu \in \Lambda$. Somit ist $(\lambda - 1/\|R_\lambda\|, \lambda + 1/\|R_\lambda\|) \subset \Lambda$.

ii), c) Wegen Voraussetzung 2. ist $\Lambda \neq \emptyset$, daher muss $\Lambda = (0, \infty)$ gelten.

Schritt iii) Zeige: Die R_λ aus Schritt *ii)* haben die Eigenschaften einer Resolvente wie in Lemma 6.33 (Resolventengleichung, siehe (6.15)) Satz 6.34.

Nach obigem ist R_λ für jedes $\lambda > 0$ ein beschränkter linearer Operator auf $C_0(E)$ mit $\|R_\lambda\| \le 1/\lambda$.

iii), a) Es ist (nach Definition) $(\lambda I - \overline{A})R_\lambda = (\mu I - \overline{A})R_\mu$ für $\lambda, \mu > 0$ und somit

$$(\lambda I - \overline{A})(R_\lambda - R_\mu) = (\mu - \lambda)R_\mu \quad \Longrightarrow \quad R_\lambda - R_\mu = (\mu - \lambda)R_\lambda R_\mu$$

(wende R_λ von links auf die erste Identität an, um die zweite zu erhalten). Demnach kommutieren die R_λ, $\lambda > 0$ und erfüllen die Resolventengleichung (6.15).
iii), b) Zeige:

$$\|\lambda R_\lambda f - f\| \underset{\lambda\to\infty}{\longrightarrow} 0 \quad \text{für alle } f \in C_0(E)$$

Für $f \in \operatorname{dom}(\overline{A})$ ist $R_\lambda(\lambda I - \overline{A})f = f$ nach Definition und somit $\|\lambda R_\lambda f - f\| = \|R_\lambda \overline{A} f\| \le \lambda^{-1}\|\overline{A}f\| \to 0$ für $\lambda \to \infty$; wegen $\operatorname{dom}(\overline{A}) \supset \widetilde{\mathcal{D}}$ und Voraussetzung 1. überträgt sich dies auf ganz $C_0(E)$.
iii), b) Zeige: R_λ ist ein positiver Operator (d.h. $R_\lambda g \ge 0$ für $C_0(E) \ni g \ge 0$).
Sei $g \in C_0(E)$ mit $g \ge 0$, $f := R_\lambda g$ und somit $g = (\lambda I - \overline{A})f$. Angenommen, es gäbe $x \in E$ mit $f(x) < 0$. Nach Definition von $\overline{A}$ gibt es $(f_n)_n \subset \widetilde{\mathcal{D}}$ mit $f_n \to f$ und $Af_n \to \overline{A}f$. Demnach gibt es (zumindest für n genügend groß) $x_n \in E$ und $\delta > 0$ mit $f_n(x_n) = \min_{y\in E} f_n(y) \le -\delta$. Nach Voraussetzung 3. (angewendet auf $-f_n$) ist $Af_n(x_n) \ge 0$ und daher

$$\inf_{y\in E} (\lambda I - A)f_n(y) \le (\lambda I - A)f_n(x_n) \le \lambda f_n(x_n) = \lambda \inf_{y\in E} f_n(y) \le -\lambda\delta$$

Andererseits ist

$$\begin{aligned} 0 \le \inf_{y\in E} g(y) &= \inf_{y\in E} (\lambda I - \overline{A})f(y) = \inf_{y\in E} \lim_{n\to\infty} (\lambda I - A)f_n(y) \\ &\le \liminf_{n\to\infty} \inf_{y\in E} (\lambda I - A)f_n(y) \le \liminf_{n\to\infty} (\lambda I - A)f_n(x_n) \le -\lambda\delta \end{aligned}$$

ein Widerspruch.

Schritt iv) Analog zum Beweis von Satz 6.36 verwenden wir die Yosida-Approximation (vgl. (6.16)) um aus $(R_\lambda)_{\lambda>0}$ eine Halbgruppe zu gewinnen. Setze

$$A^{(\lambda)} := \lambda(\lambda\mathbb{R}_\lambda - I) \qquad \text{und} \qquad T_t^{(\lambda)} = e^{tA^{(\lambda)}}$$

Da R_λ positive Operatoren sind, gilt dies auch für $T_t^{(\lambda)} = e^{tA^{(\lambda)}}$. Wie im Beweis von Satz 6.36 folgt $T_t^{(\lambda)} f \to T_t f$ für $f \in C_0(E)$ mit $\lambda \to \infty$ gleichmäßig für t aus beschränkten Intervallen, wobei die Limesobjekte T_t für $t \ge 0$ stark stetige, positive Kontraktionsoperatoren sind mit $R_\lambda f = \int_0^\infty e^{-\lambda u} T_u f \, du$ für $f \in C_0(E)$. Auch die Halbgruppeneigenschaft überträgt sich von $(T_t^{(\lambda)})$ im Grenzwert $\lambda \to \infty$ auf (T_t):

$$\big(T_{s+t} - T_s T_t\big)f = \big(T_{s+t} - T_{s+t}^{(\lambda)}\big)f + T_s^{(\lambda)}\big(T_t^{(\lambda)} - T_t\big)f + \big(T_s^{(\lambda)} - T_s\big)T_t f \underset{\lambda\to\infty}{\longrightarrow} 0$$

Da nach Konstruktion $(\lambda I - \overline{A})^{-1} = R_\lambda$ und $(R_\lambda)_{\lambda>0}$ nach obigem die Resolvente zu $(T_t)_{t\ge 0}$ ist, folgt die Behauptung nun aus Satz 6.34 und Beobachtung 6.35. □

Bemerkung 6.45. 1. Der Beweis von Satz 6.44 zeigt auch: das Bild von $\lambda I - A$ liegt dann für jedes $\lambda > 0$ dicht in $C_0(E)$.

2. Die Konstruktion aus Satz 6.44 garantiert i.A. nicht, dass $T_t\mathbf{1}(x) \equiv 1$ gilt, d.h. im Allgemeinen erhält man eine Halbgruppe von sub-Markov-Übergangsoperatoren. Man kann dann dem Zustandsraum einen E zusätzlichen Punkt Δ als sogenannten „Friedhofszustand" hinzufügen und $T_t\mathbf{1}_{\{\Delta\}}(x) = 1 - T_t\mathbf{1}(x)$ sowie $T_t\mathbf{1}_{\{\Delta\}}(\Delta) = 1$ setzen, um auf $E\dot\cup\{\Delta\}$ eine Markov-Übergangshalbgruppe zu gewinnen.

Falls $A\mathbf{1} = 0$ gilt, so erfüllt die zugehörige Übergangshalbgruppe $T_t\mathbf{1}(x) \equiv 1$. Es ist dann nämlich $(\lambda I - \overline{A})\mathbf{1} = \lambda\mathbf{1}$, folglich $R_\lambda\mathbf{1} = \lambda^{-1}\mathbf{1}$ und $A^{(\lambda)}\mathbf{1} = \lambda(\lambda\mathbb{R}_\lambda - I)\mathbf{1} = 0$, somit $T_t^{(\lambda)}\mathbf{1} = \mathbf{1}$ für jedes $\lambda > 0$ und $T_t\mathbf{1} = \lim_{\lambda\to\infty} T_t^{(\lambda)}\mathbf{1} = \mathbf{1}$.

Beispiele

Beispiel 6.46. Wir können die Beispiele aus Abschnitt 6.2 im Licht obiger Ergebnisse erneut betrachten:

1. (Beispiel 6.19 erneut): Sei $E = \mathbb{R}$, $(\nu_t)_{t\in[0,\infty)} \subset \mathcal{M}_1(\mathbb{R})$ eine Faltungshalbgruppe und

$$\log\varphi_{\nu_1}(t) = ibu - \frac{1}{2}\sigma^2u^2 + \int\left(e^{iux} - 1 - \mathbf{1}_{\{|x|<1\}}(x)\cdot iux\right)\nu(dx)$$

mit kanonischem Tripel (b, σ^2, ν) die zugehörige Lévy-Khinchin-Darstellung (vgl. Satz 5.9). Die zugehörige Übergangshalbgruppe aus Beispiel 6.19, $T_tf(x) = \int_{\mathbb{R}} f(x+y)\,\nu_t(dy)$ hat Generator A

$$\begin{aligned}Af(x) &= \lim_{t\downarrow 0}\frac{T_tf(x) - T_0f(x)}{t}\\ &= bf'(x) + \frac{1}{2}\sigma^2f''(x) + \int_{\mathbb{R}}\left(f(x+y) - f(x) - y\mathbf{1}_{\{|y|<1\}}f'(x)\right)\nu(dh), \quad x\in\mathbb{R}\end{aligned}$$

für

$$f\in\mathcal{D} = \{f\in C_0(E) : \ f \text{ zweimal stetig differenzierbar mit } f''\in C_0(E)\}$$

(Man kann eine Darstellung wie in Beobachtung 5.7 der Vorlesungsnotizen verwenden und dann Taylorentwicklung von f in x bis zur zweiten Ordnung betrachten. Details: Übung) Man kann auch (mit Bericht 6.43, 3.) zeigen, dass z.B. $\widetilde{\mathcal{D}} = \{f\in\mathcal{D} : f \text{ ist unendlich oft differenzierbar}\}$ ein core für A ist.

Insbesondere ist der Generator der Übergangshalbgruppe $(T_t)_{t\geq 0}$ mit

$$T_tf(x) = \int_{\mathbb{R}}(2\pi t)^{-1/2}f(x+y)\exp(-y^2/(2t))\,dy$$

der standard-Brownbewegung $(B_t)_{t\geq 0}$ (vgl. Kapitel 2, speziell Definition 2.1) der Operator $Af = \frac{1}{2}f''$ mit Definitionsbereich $\mathcal{D}$ wie oben.

Analog besitzt $\sigma B_t + \mu t$, $t\geq 0$, die Brownbewegung mit Varianzparameter $\sigma^2 > 0$ und Driftparameter $\mu\in\mathbb{R}$ den Generator $f\mapsto\frac{1}{2}\sigma^2f'' + \mu f'$ mit Definitionsbereich $\mathcal{D}$.

2. (Beispiel 6.21 erneut): Für eine Markovkette $(X_t)_{t\geq 0}$ mit endlichem Zustandsraum E und Sprungratenmatrix $Q = (Q(x,y))_{x,y\in E}$ ist der Generator (natürlich) $Qf(x) = \sum_{y\in E}Q(x,y)f(y)$ und der Definitionsbereich sind (natürlich) alle Funktionen $f : E\to\mathbb{R}$.

Etwas allgemeiner ist für einen Pseudo-Poisson-Prozess aus Beispiel 6.31 der Generator A mit Definitionsbereich $\mathcal{D} = C_0(E)$ durch die Formel dort gegeben.

Beispiel 6.47 (Reflektierte bzw. absorbierte Brownbewegung im $d = 1$)**.** Sei $(B_t)_{t\geq 0}$ standard-Brownbewegung auf $\mathbb{R}^1$.

1. $X_t := |x + B_t|$, $t \geq 0$ ist eine reflektierte Brownbewegung mit Start in $x \in E := [0,\infty)$, die Übergangshalbgruppe ist gegeben durch

$$\begin{aligned} T_t f(x) = \mathbb{E}\big[f(|x+B_t|)\big] &= \mathbb{E}\big[f(x+B_t)\mathbf{1}_{\{B_t > -x\}}\big] + \mathbb{E}\big[f(-x-B_t)\mathbf{1}_{\{B_t < -x\}}\big] \\ &= \mathbb{E}_x\big[f(B_t)\mathbf{1}_{\{B_t \geq 0\}}\big] + \mathbb{E}_{-x}\big[f(B_t)\mathbf{1}_{\{B_t \geq 0\}}\big] \end{aligned}$$

und man prüft, dass der Generator gegeben ist durch

$$Af(x) = \lim_{t\downarrow 0} \frac{T_t f(x) - T_0 f(x)}{t} = \frac{1}{2} f''(x), \quad x \in [0,\infty)$$

mit Definitionsbereich

$$\mathcal{D}_1 = \{f \in C_0([0,\infty)) : f \text{ zweimal stetig differenzierbar mit } f'' \in C_0([0,\infty)) \text{ und } f'(0) = 0\}$$

2. $X_t := x + B_{t\wedge\tau}$ mit $\tau = \inf\{t \geq 0 : B_t = -x\}$ ist eine in 0 absorbierte Brownbewegung mit Start in $x \in E := [0,\infty)$, die Übergangshalbgruppe ist gegeben durch

$$\begin{aligned} T_t f(x) = \mathbb{E}\big[f(x+B_{t\wedge\tau})\big] &= \mathbb{E}\big[f(x+B_t)\mathbf{1}_{\{t<\tau\}}\big] + \mathbb{E}\big[f(0)\mathbf{1}_{\{\tau\leq t\}}\big] \\ &= \mathbb{E}_x\big[f(B_t)\mathbf{1}_{\{B_t\geq 0\}}\mathbf{1}_{\{\inf\{B_u : u\leq t\}>0\}}\big] + \mathbb{E}_x\big[f(0)\mathbf{1}_{\{\inf\{B_u:u\leq t\}\leq 0\}}\big] \\ &= \mathbb{E}_x\big[f(B_t)\mathbf{1}_{\{B_t\geq 0\}}\big] - \mathbb{E}_{-x}\big[f(B_t)\mathbf{1}_{\{B_t\geq 0\}}\big] + f(0)\mathbb{P}_0(\sup\{B_u : u\leq t\} \geq x) \end{aligned}$$

(wir verwenden hier das Analogon von Satz 6.28, das Spiegelungsprinzip für die Brownsche Bewegung). Man prüft, dass der Generator hier gegeben ist durch

$$Af(x) = \lim_{t\downarrow 0} \frac{T_t f(x) - T_0 f(x)}{t} = \frac{1}{2} f''(x), \quad x \in [0,\infty)$$

nun mit Definitionsbereich

$$\mathcal{D}_2 = \{f \in C_0([0,\infty)) : f \text{ zweimal stetig differenzierbar mit } f'' \in C_0([0,\infty)) \text{ und } f''(0) = 0\}$$

Insbesondere sieht man, dass es zur Charakterisierung einer Übergangshalbgruppe durch ihren Generator $(A,\mathcal{D})$ nicht nur auf die „formale Struktur“ von A sondern auch auf seinen Definitionsbereich ankommt.

Zudem zeigen Beispiel 6.47, 1. und 2. zusammen, dass man in Bericht 6.43, 3. auf die Forderung, dass $\widetilde{\mathcal{D}}$ invariant unter der Halbgruppe (T_t) ist, nicht verzichten kann:

$$\mathcal{D}_1 \cap \mathcal{D}_2 = \{f \in C_0([0,\infty)) : f \text{ zweimal ste. diff'bar mit } f'' \in C_0([0,\infty)) \text{ und } f'(0) = f''(0) = 0\}$$

liegt zwar dicht in $C_0([0,\infty))$, aber der Abschluss der Einschränkung von $A : f \mapsto \frac{1}{2}f''$ auf $\mathcal{D}_1 \cap \mathcal{D}_2$ ist kein Erzeuger einer Feller-Halbgruppe. Andernfalls müsste der zugehörige stochastische Prozess zugleich eine reflektierte und eine absorbierte Brownbewgung sein, siehe z.B. [L10, Remark 3.57] für mehr Details.

Beispiel 6.48. Sei $E = \mathbb{R}$, $a : E \to (0,\infty)$ und $b : E \to \mathbb{R}$ seien stetig und

$$Af(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x), \quad x \in \mathbb{R}$$

für $f \in \mathcal{D}$ mit $\mathcal{D}$ aus Beispiel 6.46, 1.

$\mathcal{D}$ liegt dicht in $C_0(\mathbb{R})$ und wegen $a(\cdot) > 0$ erfüllt A das positive-Maximum-Prinzip (Satz 6.44, Voraussetzung 3.). Unter geeigneten Zusatzbedingungen (z.B. $a(x) \le K(1+x^2)$, $|b(x)| \le K(1+|x|)$ für ein $K < \infty$) erfüllt A auch Voraussetzung 2. in Satz 6.44, siehe z.B. [EK, Ch. 8.1]. Demnach gibt es einen Markovprozess $X = (X_t)_{t\ge 0}$ auf $\mathbb{R}$, dessen Übergangshalbgruppe Generator $(A,\mathcal{D})$ hat.

Vergleich mit Beispiel 6.46, 1. zeigt: „lokal“, d.h. in der Nähe des Punkts x über kurze Zeiträume betrachtet, verhält sich X wie eine Brownbewegung mit Varianzparameter $a(x)$ und Driftparameter $b(x)$.

(Für $a(x) = \sigma^2, b(x) = -cx$ mit $\sigma^2, c > 0$ handelt es sich um einen Ornstein-Uhlenbeck-Prozess.)

Beispiel 6.49 (Ein stochastisches Ising-Modell auf $\mathbb{Z}^d$)**.** $d \in \mathbb{N}$, $E = \{-1,+1\}^{\mathbb{Z}^d} \ni \eta = (\eta(x))_{x\in\mathbb{Z}^d}$ mit Metrik $d(\eta,\eta') = \sum_{x\in\mathbb{Z}^d} e^{-\|x\|}|\eta(x) - \eta'(x)|$ (die die Produkttopologie erzeugt). Für $x \in \mathbb{Z}^d$, $\eta \in E$ sei $\eta^{(x)} \in E$ gegeben durch

$$\eta^{(x)}(y) = \begin{cases} -\eta(x), & y = x \\ \eta(x), & y \ne x \end{cases}$$

Setze

$$\widetilde{\mathcal{D}} = \big\{ f : E \to E \,:\, f \text{ hängt nur von endlich vielen Koordinaten } \eta(x) \text{ ab} \big\} \quad \big(\subset C_0(E)\big)$$

und

$$Af(\eta) = \sum_{x\in\mathbb{Z}^d} c(x,\eta)\big(f(\eta^{(x)}) - f(\eta)\big), \quad f \in \widetilde{\mathcal{D}}$$

mit

$$c(x,\eta) = \left(1 + \exp\Big(2\beta \sum_{\substack{y\in\mathbb{Z}^d:\\ \|x-y\|=1}} \eta(x)\eta(y) - h\eta(x)\Big)\right)^{-1} \tag{6.25}$$

wobei $\beta > 0$ („inverse Temperatur“) und $h \in \mathbb{R}$ („externe Feldstärke“) Parameter sind, vgl. z.B. [L85, Ch. IV]. A erfüllt das positive-Maximum-Prinzip (Satz 6.44, Voraussetzung 3.); $\widetilde{\mathcal{D}}$ liegt dicht in $C_0(E)$, die ist nicht allzu schwer zu beweisen via Satz von Stone-Weierstraß (Satz 3.23, siehe auch [L85, S. 21f] (beachte: E is kompakt); Voraussetzung 2.) in Satz 6.44 ist etwas aufwendiger zu verifizieren, siehe z.B. [L85, Beweis von Theorem I.3.9].

6.5 Ergänzung: Zu diskreten Markov-Ketten

6.5.1 Grundlegendes Szenario

Es sei E eine abzählbare Menge, $p = (p_{xy})_{x,y\in E}$ eine stochastische Matrix und $X = (X_n)_{n\in\mathbb{N}_0}$ eine (diskrete) p-Markov-Kette, das heißt

$$\mathbb{P}_{x_0}(X_{n+1} = y \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x) = p_{xy}$$

für alle $n \in \mathbb{N}$ und $x_0, \dots, x_{n-1}, x, y \in E$ mit $\mathbb{P}_{x_0}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) > 0$. Wir notieren die n-ten Potenzen der Übergangsmatrix als $p^n_{xy} = \mathbb{P}_x(X_n = y)$.

Beispiel 6.50 (Erneuerungskette)**.** Sei $E = \mathbb{N}_0$, $\nu = (\nu_k)_{k\in\mathbb{N}} \in \mathcal{M}_1(\mathbb{N})$ und p gegeben durch

$$p_{i,j} = \begin{cases} \nu_{j+1}, & i = 0 \\ 1, & j = i - 1 \\ 0, & \text{sonst} \end{cases}$$

Eine Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$ mit Übergangsmatrix p und Start in $X_0 = x_0$ kann folgendermaßen dargestellt werden: Seien $\xi_1, \xi_2, \dots \sim \nu$ unabhängig, $T_0 := x_0$, $T_m := T_0 + \sum_{k=1}^m \xi_k$ für $m \in \mathbb{N}$ und

$$X_n := \inf\{T_k - n \mid k \in \mathbb{N}_0, T_k \geq n\}.$$

Interpretation: Die T_k sind die „Erneuerungszeitpunkte" — man denke an Zeitpunkte, zu denen jeweils eine defekte Glühbirne ausgetauscht wird — X_n ist dann die „Restlebensdauer" der zum Zeitpunkt n brennenden Birne.

6.5.2 Rekurrenz und Transienz

Definition 6.51. Für $x \in E$ sei $T_x^{(1)} := \inf\{n > 0 \mid X_n = x\}$ und für $k \in \mathbb{N}$, $k \geq 2$

$$T_x^{(k)} := \inf\{n > T_x^{(k-1)} \mid X_n = x\}$$

(mit Setzung $T_x^{(k)} = \infty$, falls $T_x^{(k-1)} = \infty$). $T_x^{(k)}$ heißt die *k-te Eintrittszeit* (auch *Rückkehrzeit*) von X in x.

Bemerkung. i) Nach Konvention ist $T_x^{(1)} > 0$, auch bei Start in x.

ii) Die $T_x^{(k)}$ sind Stoppzeiten bezüglich $(\mathscr{F}_n)_n$ mit $\mathscr{F}_n = \sigma(X_0, \dots, X_n)$.

Definition 6.52. Für $x, y \in E$ sei

$$F(x,y) := \mathbb{P}_x(T_y^{(1)} < \infty) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\{X_n = y\}\right)$$

die Wahrscheinlichkeit, bei Start in x jemals y zu erreichen, beziehungsweise für $y = x$ die Wahrscheinlichkeit, bei Start in x jemals nach x zurückzukehren.

Lemma 6.53. *Für $x, y \in E$ und $k \in \mathbb{N}$ gilt $P_x(T_y^{(k)} < \infty) = F(x,y)F(y,y)^{k-1}$.*

Intuitiv: Damit $T_y^{(k)} < \infty$ gilt, muss die Kette zunächst y erreichen und dann noch $(k-1)$-mal zurückkehren.

Beweis. Durch Induktion über $k \in \mathbb{N}$. Für $k = 1$ gilt die Aussage nach Definition von $F(x,y)$. Sei die Behauptung also für $k \in \mathbb{N}$ erfüllt. Es gilt

$$\begin{aligned}
\mathbb{P}_x(T_y^{(k+1)} < \infty) &= \mathbb{E}_x\Big[\mathbb{P}_x\big(T_y^{(k+1)} < \infty \,|\, \mathcal{F}_{T_y^{(k)}}\big)\mathbf{1}_{\{T_y^{(k)}<\infty\}}\Big] \\
&= \mathbb{E}_x\Big[\mathbb{P}_x\big(\inf\{n > 0 : X_{T_y^{(k)}+n} = y\} < \infty \,|\, \mathcal{F}_{T_y^{(k)}}\big)\mathbf{1}_{\{T_y^{(k)}<\infty\}}\Big] \\
&= \mathbb{E}_x\Big[\mathbb{P}_{X_{T_y^{(k)}}}\big(T_y^{(1)} < \infty\big)\mathbf{1}_{\{T_y^{(k)}<\infty\}}\Big] \\
&= \mathbb{E}_x\Big[\mathbb{P}_y\big(T_y^{(k)} < \infty\big)\mathbf{1}_{\{T_y^{(k)}<\infty\}}\Big] \\
&= F(y,y)\mathbb{P}_x(T_y^{(k)} < \infty) = F(y,y)F(x,y)F(y,y)^{k-1} \\
&= F(x,y)F(y,y)^k,
\end{aligned}$$

wobei wir in der vierten Zeile die starke Markov-Eigenschaft und in der fünften Zeile die Induktionsannahme verwendet haben. □

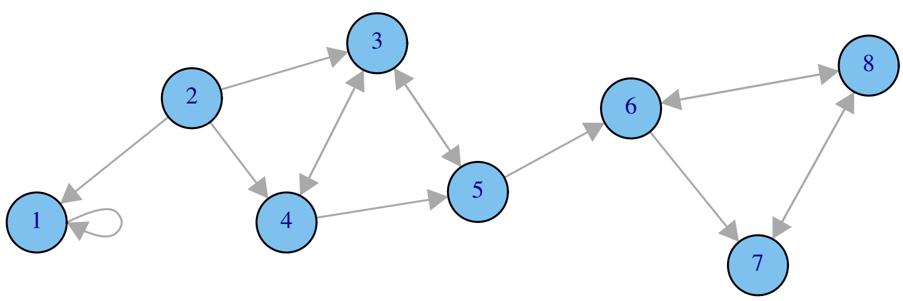
Definition 6.54. Ein Zustand $x \in E$ heißt (bezüglich p)

- *rekurrent*, falls $F(x,x) = 1$,
- *positiv rekurrent*, falls $\mathbb{E}_x\big[T_x^{(1)}\big] < \infty$ (also insbesondere $F(x,x) = 1$),
- *nullrekurrent*, falls $\mathbb{E}_x\big[T_x^{(1)}\big] = \infty$ und $F(x,x) = 1$,
- *transient*, falls $F(x,x) < 1$,
- *absorbierend*, falls $p_{xx} = 1$.

Die Markov-Kette X heißt *(positiv / null-) rekurrent*, wenn dies für jeden Zustand gilt. X heißt *transient*, wenn jeder rekurrente Zustand absorbierend ist.

Bemerkung. Es gilt: x absorbierend $\Rightarrow$ x positiv rekurrent $\Rightarrow$ x rekurrent .

Beispiel. Betrachte folgende Markov-Kette auf $E = \{1, 2, \dots, 8\}$, wobei die Pfeile positiven Übergangswahrscheinlichkeiten entsprechen:



Zustand 1 ist absorbierend, Zustände 2,3,4,5 sind transient und Zustände 6,7,8 sind (positiv) rekurrent.

Definition 6.55. Sei $N(y) := \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=y\}}$ die Anzahl der Besuche in $y \in E$. Sei

$$G(x,y) := \mathbb{E}_x[N(y)] = \sum_{n=0}^{\infty} p_{x,y}^n.$$

G heißt *Greenfunktion*[4] *von* X.

Satz 6.56. *Für* $x, y \in E$ *gilt*

$$G(x,y) = \begin{cases} \frac{F(x,y)}{1-F(y,y)}, & y \neq x \\ \frac{1}{1-F(y,y)}, & y = x \end{cases}$$

mit Konvention $1/0 = \infty$. *Insbesondere ist* $G(x,y) = F(x,y)G(y,y) + \mathbf{1}_{\{x=y\}}$ *und* y *ist genau dann rekurrent, wenn* $G(y,y) = \infty$.

Beweis.

$$\begin{aligned} G(x,y) = \mathbb{E}_x[N(y)] &= \sum_{k=1}^{\infty} \mathbb{P}_x(N(y) \geq k) = \mathbf{1}_{\{x=y\}} + \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^{(k)} < \infty) \\ &= \mathbf{1}_{\{x=y\}} + \sum_{k=1}^{\infty} F(x,y)F(y,y)^{k-1} = \mathbf{1}_{\{x=y\}} + \frac{F(x,y)}{1-F(y,y)} \end{aligned}$$

□

Satz 6.57. *Sei* $x \in E$ *rekurrent und es gelte* $F(x,y) > 0$ *für ein* $y \in E$. *Dann ist auch* y *rekurrent und es gilt* $F(x,y) = F(y,x) = 1$. *Falls* x *positiv rekurrent ist, so auch* y *und es gilt* $\mathbb{E}_x[T_y^{(1)}], \mathbb{E}_y[T_x^{(1)}] < \infty$.

Beweis. Sei ohne Einschränkung $y \neq x$. Wegen $F(x,y) > 0$ gibt es ein $k \in \mathbb{N}$ und $x_1, \ldots, x_k \in E$ mit $x_k = y$, $x_i \neq x$ für $i = 1, \ldots, k$ und $\mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k) > 0$. Insbesondere ist $p_{xy}^k > 0$. Es gilt

$$\begin{aligned} 0 = 1 - F(x,x) &= \mathbb{P}_x(T_x^{(1)} = \infty) \geq \mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k, T_x^{(1)} = \infty) \\ &= \mathbb{E}_x\big[\mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k, T_x^{(1)} = \infty \,|\, \mathcal{F}_k)\big] \\ &= \mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k)\mathbb{P}_y(T_x^{(1)} = \infty), \end{aligned}$$

also $\mathbb{P}_y(T_x^{(1)} = \infty) = 0$ und damit $F(y,x) = 1 - \mathbb{P}_y(T_x^{(1)} = \infty) = 1$. Insbesondere gilt $p_{yx}^{\ell} > 0$ für ein $\ell \in \mathbb{N}$, somit

$$G(y,y) = \textstyle\sum_{n=0}^{\infty} p_{yy}^n \geq \sum_{m=0}^{\infty} p_{yx}^{\ell} p_{xx}^m p_{xy}^k = p_{yx}^{\ell} p_{xy}^k G(x,x) = \infty$$

nach Voraussetzung, das heißt y ist rekurrent. Vertauschung der Rollen von x und y zeigt nun auch $F(x,y) = 1$.

Sei nun x positiv rekurrent und $x_1, \ldots, x_k$ wie zuvor, dann gilt

$$\begin{aligned} \infty > \mathbb{E}_x[T_x^{(1)}] &\geq \mathbb{E}_x\big[\mathbf{1}_{\{X_1=x_1,\ldots,X_k=x_k\}} T_x^{(1)}\big] = \mathbb{E}_x\big[\mathbb{E}_x[\mathbf{1}_{\{X_1=x_1,\ldots,X_k=x_k\}} T_x^{(1)} \,|\, \mathcal{F}_k]\big] \\ &= \mathbb{E}_x\big[\mathbf{1}_{\{X_1=x_1,\ldots,X_k=x_k\}} \mathbb{E}_y[T_x^{(1)}]\big] = \mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k)\mathbb{E}_y[T_x^{(1)}], \end{aligned}$$

demnach ist $\mathbb{E}_y[T_x^{(1)}] < \infty$. Nach Vertauschung der Rollen von x und y erhält man nun auch $\mathbb{E}_x[T_y^{(1)}] < \infty$. □

[4] benannt nach George Green (1793-1841), der ein analoges Objekt in der Theorie der partiellen Differentialgleichungen eingeführt hat

Definition 6.58. Eine diskrete Markov-Kette heißt *irreduzibel*, wenn

$$F(x,y) > 0 \quad \forall\, x,y \in E.$$

Bemerkung 6.59. i) Eine irreduzible Kette X besitzt entweder nur rekurrente oder nur transiente Zustände. Falls $|E| > 1$, so gibt es keine absorbierenden Zustände.

ii) Ist $|E| < \infty$ und X irreduzibel, so ist X rekurrent.

Beweis. i) Die Aussage folgt aus Satz 6.57.

ii) Sei $x \in E$. Es gilt

$$\sum_{y\in E} G(x,y) = \sum_{n=0}^{\infty}\sum_{y\in E} p_{xy}^n = \sum_{n=0}^{\infty} 1 = \infty,$$

das heißt es gibt ein y mit $G(x,y) = \infty$. Wegen der Irreduzibilität von X existiert ein $k \in \mathbb{N}$ mit $p_{yx}^k > 0$, somit gilt auch

$$G(x,x) \geq \sum_{m=0}^{\infty} p_{xy}^m p_{yx}^k = G(x,y)p_{yx}^k = \infty.$$

□

Beispiel 6.60 (Irrfahrten auf $\mathbb{Z}^d$)**.** Seien $Y_1, Y_2, \dots$ unabhängige und identisch verteilte, $\mathbb{Z}^d$-wertige Zufallsvariablen. Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette mit Übergangsmatrix $p_{xy} = \mathbb{P}(Y_1 = y - x)$, das heißt unter $\mathbb{P}_x$ ist

$$(X_n)_{n\in\mathbb{N}_0} \overset{d}{=} (x + Y_1 + \dots + Y_n).$$

i) Sei $\mu := \mathbb{E}[Y_1] \neq \mathbf{0}$. Dann gilt für jedes $x \in \mathbb{Z}^d$

$$\frac{X_n}{n} \to \mu \quad \mathbb{P}_x\text{-f.s.},$$

insbesondere ist $\#\{n \in \mathbb{N}_0 \mid X_n = x\} < \infty$ $\mathbb{P}_x$-f.s., das heißt X ist transient.

ii) Sei $\mu := \mathbb{E}[Y_1] = \mathbf{0}$, $\mathbb{E}[\|Y_1\|^2] < \infty$, die Kovarianzmatrix $C = (\mathrm{Cov}[Y_{1_i}, Y_{1,j}])_{i,j=1,\dots,d}$ sei invertierbar und die von $\{y \mid \mathbb{P}(Y_1 = y) > 0\}$ erzeugte Gruppe sei $\mathbb{Z}^d$. Dann gilt

$$p_{\mathbf{0},\mathbf{0}}^n = \mathbb{P}_0(X_n = \mathbf{0}) = \mathbb{P}(Y_1 + \dots + Y_n = \mathbf{0}) \sim \frac{1}{(2\pi)^{d/2}|\det(C)|^{1/2}n^{d/2}}$$

für $n \to \infty$ (lokaler Zentraler Grenzwertsatz, zum Beispiel via Fourier-Inversion). Somit ist

$$G(\mathbf{0},\mathbf{0}) = \sum_{n=0}^{\infty} p_{\mathbf{0},\mathbf{0}}^n \begin{cases} = \infty, & d = 1,2, \\ < \infty, & d \geq 3, \end{cases}$$

das heißt eine zentrierte Irrfahrt mit endlicher Varianz ist genau dann rekurrent, wenn $d \leq 2$.

6.5.3 Invariante Verteilungen

Für ein Maß μ auf E sei $\mu p(\{x\}) = \sum_{y\in E} \mu(\{y\})p_{yx}$ (Transport von μ durch den Übergangskern p).

Definition 6.61. Ein (σ-endliches) Maß μ heißt (p-)*invariant*, falls $\mu p = \mu$. Ist μ zudem ein Wahrscheinlichkeitsmaß, so heißt es auch eine *invariante Verteilung* (auch: *Gleichgewichtsverteilung*) von p.

Lemma 6.62. *Sei jeder Zustand transient. Dann gibt es keine invariante Verteilung für p.*

Beweis. Nach Voraussetzung und Satz 6.56 ist $G(x,y) = \sum_n p_{xy}^n < \infty$, insbesondere gilt $p_{xy}^n \to 0$ für $n \to \infty$ für alle $x, y \in E$.

Angenommen μ wäre eine invariante Verteilung. Dann gibt es ein $y \in E$ mit $\mu(\{y\}) > 0$, ein endliches $E' \subset E$ mit $\mu(E \setminus E') \le \frac{1}{4}\mu(\{y\})$ und für genügend großes n ist

$$\max_{x\in E'} p_{xy}^n < \frac{\mu(\{y\})}{4\mu(E')}.$$

Somit folgt

$$\begin{aligned}\mu(\{y\}) = \mu p(\{y\}) = \cdots = \mu p^n(\{y\}) &= \sum_{x\in E'} \mu(\{x\})p_{xy}^n + \sum_{x\in E\setminus E'} \mu(\{x\})p_{xy}^n \\ &\le \frac{1}{4}\mu(\{y\}) + \frac{1}{4}\mu(\{y\}) = \frac{1}{2}\mu(\{y\}),\end{aligned}$$

also ein Widerspruch. □

Satz 6.63. *Sei x rekurrenter Zustand, dann definiert*

$$\mu_x(\{y\}) := \mathbb{E}_x\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x^{(1)} > x), \quad y \in E$$

ein invariantes Maß μ_x.

Diese Konstruktion wird auch der „Zyklus-Trick“ genannt. Intuitiv ist $\mu_x(\{y\})$ die erwartete Anzahl Besuche in y während $\{0, 1, \dots, T_x^{(1)} - 1\}$ und $\mu_x p(\{y\})$ die erwartete Anzahl Besuche in y während $\{1, 2, \dots, T_x^{(1)}\}$, diese Anzahlen sind aber wegen $X_{T_x^{(1)}} = x$ gleich.

Beweis. Zeige zunächst $\mu_x(\{y\}) < \infty$ für alle $y \in E$. Es gilt $\mu_x(x) = 1$ und $\mu_x(y) = 0$, falls $F(x,y) = 0$. Falls $F(x,y) > 0$ für $y \neq x$, so setze $\hat{F}(x,y) := \mathbb{P}_x(T_y^{(1)} < T_x^{(1)})$. Dann ist $\hat{F}(x,y) > 0$ und

$\hat{F}(y,x) > 0$ und es gilt

$$\begin{aligned}
\mathbb{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right] &= 1 + \mathbb{E}_y\left[\sum_{n=T_y^{(1)}}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\mathbf{1}_{\{T_y^{(1)}<T_x^{(1)}\}}\right] \\
&= 1 + \mathbb{E}_y\left[\mathbb{E}_y\left[\sum_{n=T_y^{(1)}}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\mathbf{1}_{\{T_y^{(1)}<T_x^{(1)}\}} \mid \mathscr{F}_{T_y^{(1)}}\right]\right] \\
&= 1 + \mathbb{E}_y\left[\mathbf{1}_{\{T_y^{(1)}<T_x^{(1)}\}}\mathbb{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right]\right] \\
&= 1 + \big(1-\hat{F}(y,x)\big)\mathbb{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right],
\end{aligned}$$

das heißt

$$\mathbb{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right] = \frac{1}{\hat{F}(y,x)}.$$

Damit folgt

$$\begin{aligned}
\mu_x(\{y\}) = \mathbb{E}_x\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right] &= \mathbb{E}_x\left[\mathbf{1}_{\{T_y^{(1)}<T_x^{(1)}\}}\sum_{n=T_y^{(1)}}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right] \\
&= \mathbb{P}_x(T_y^{(1)} < T_x^{(1)})\mathbb{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_n=y\}}\right] = \frac{\hat{F}(x,y)}{\hat{F}(y,x)} < \infty
\end{aligned}$$

Setze nun $\overline{p}_n(x,y) := \mathbb{P}_x(X_n = y, T_x^{(1)} > n)$, dann gilt

$$\mu_x p(\{z\}) = \sum_{y\in E} \mu_x(\{y\})p_{yz} = \sum_{n=0}^{\infty}\sum_{y\in E} \overline{p}_n(x,y)p_{yz}.$$

1. Fall: $x \neq z$. Es gilt

$$\sum_{y\in E} \overline{p}_n(x,y) = \sum_{y\in E} \underbrace{\mathbb{P}_x(X_n = y,\ T_x^{(1)} > n,\ X_{n+1} = z)}_{=\mathbb{P}_x(X_n=y,\ X_{n+1}=z,\ T_x^{(1)}>n+1)} = \mathbb{P}_x(X_{n+1} = z,\ T_x^{(1)} > n+1)$$

und somit wegen $\overline{p}_0(x,z) = 0$ für $x \neq z$

$$\mu_x p(z) = \sum_{n=0}^{\infty} \overline{p}_{n+1}(x,y) \sum_{n=0}^{\infty} \overline{p}_n(x,y) = \mu_x(\{z\}).$$

2. Fall: $x = z$. Es gilt

$$\sum_{y\in E} \overline{p}_n(x,y)p_{yx} = \sum_{y\in E} \mathbb{P}_x(X_n = y,\ T_x^{(1)} > n,\ X_{n+1} = x) = \mathbb{P}_x(T_x^{(1)} = n+1),$$

das heißt

$$\mu_x p(\{x\}) = \sum_{n=0}^{\infty}\sum_{y\in E} \overline{p}_n(x,y)p_{yz} = \sum_{n=0}^{\infty} \mathbb{P}_x(T_x^{(1)} = n+1) = 1 = \mu_x(\{x\}).$$

Damit ist μ_x ein invariantes Maß. □

Korollar 6.64. *Ist x positiv rekurrent, so definiert*

$$\pi := \frac{1}{\mathbb{E}_x\big[T_x^{(1)}\big]}\mu_x$$

eine invariante Verteilung.

Satz 6.65. *Eine irreduzible Markov-Kette besitzt höchstens eine invariante Verteilung.*

Beweis. Seien $\pi, \nu \in \mathcal{M}_1(E)$ invariante Verteilungen. $\tilde{p}_{xy} := \sum_{n=1}^{\infty} 2^{-n} p_{xy}^n$ ist eine stochastische Matrix und aufgrund der Irreduzibilität gilt $\tilde{p}_{xy} > 0$ für alle $x, y \in E$.

$\mu := \pi - \nu$ ist ein signiertes Maß mit $\mu\tilde{p} = \mu$ und $\mu(E) = 0$. Angenommen μ ist nicht das Nullmaß, dann gibt es $x_1, x_2 \in E$ mit $\mu(\{x_1\}) < 0 < \mu(\{x_2\})$, somit

$$\left|\mu(\{x_1\})\tilde{p}_{x_1 y} + \mu(\{x_2\})\tilde{p}_{x_2 y}\right| < |\mu(\{x_1\})\tilde{p}_{x_1 y}| + |\mu(\{x_2\})\tilde{p}_{x_2 y}| \quad \forall\, y \in E$$

und

$$\|\mu\tilde{p}\|_{\mathrm{TV}} = \sum_y \Big|\sum_x \mu(\{x\})\tilde{p}_{xy}\Big| < \sum_y \sum_x |\mu(\{x\})|\tilde{p}_{xy} = \sum_x |\mu(\{x\})| = \|\mu\|_{\mathrm{TV}} = \|\mu\tilde{p}\|_{\mathrm{TV}}\,.$$

Dies ist ein Widerspruch, μ muss also das Nullmaß sein und somit gilt $\pi = \nu$. □

Satz 6.66. *Eine irreduzible Markov-Kette X ist genau dann positiv rekurrent, wenn sie eine (notwendigerweise eindeutige) invariante Verteilung π besitzt. Diese ist dann gegeben durch*

$$\pi(\{x\}) = \frac{1}{\mathbb{E}_x\big[T_x^{(1)}\big]} > 0, \qquad x \in E.$$

Beweis. Wenn x positiv rekurrent ist, so gibt es nach Korollar 6.64 eine invariante Verteilung, welche nach Satz 6.65 eindeutig ist.

Sei nun π eine invariante Verteilung. Wegen $\pi p = \pi$ und der Irreduzibilität von X gilt $\pi(\{x\}) > 0$ für jedes $x \in E$. Setze $\mathbb{P}_\pi := \sum_{x\in E} \pi(\{x\})\mathbb{P}_x$, das heißt unter $\mathbb{P}_\pi$ ist $X_0 \sim \pi$. Für $x \in E$ und $n \in \mathbb{N}$ sei $\sigma_x^{(n)} := \sup\{m \le n : X_m = x\}$ mit Werten in $\{0, 1, \ldots, n\} \cup \{-\infty\}$. Für $k \le n$ ist

$$\begin{aligned}
\mathbb{P}_\pi(\sigma_x^{(n)} = k) &= \mathbb{P}_\pi(X_k = x, X_{k+1} \ne x, \ldots, X_n \ne x) \\
&= \mathbb{P}_\pi(X_k = x)\mathbb{P}_x(X_1 \ne x, \ldots, X_{n-k} \ne x) \\
&= \pi(\{x\})P_x(T_x^{(1)} \ge n - k + 1),
\end{aligned}$$

wobei wir in der zweiten Zeile die Markov-Eigenschaft und in der dritten Zeile die Invarianz von π ausnutzen. Somit folgt

$$\begin{aligned}
1 &= \textstyle\sum_{k=0}^n \mathbb{P}_\pi(\sigma_x^{(n)} = k) + \mathbb{P}_\pi(\sigma_x^{(n)} = -\infty) \\
&\overset{(\dagger)}{=} \pi(\{x\}) \textstyle\sum_{k=0}^n P_x(T_x^{(1)} \ge n - k + 1) + \mathbb{P}_\pi(T_x^{(1)} \ge n + 1).
\end{aligned}$$

Weiter gilt wegen der Irreduzibilität und mit monotoner Konvergenz

$$\mathbb{P}_\pi(T_x^{(1)} \ge n + 1) = \sum_y \pi(\{y\})\mathbb{P}_y(T_x^{(1)} \ge n + 1) \xrightarrow[n\to\infty]{} 0,$$

das heißt mit $n \to \infty$ folgt aus $(\dagger)$

$$1 = \pi(\{x\}) \textstyle\sum_{\ell=1}^\infty P_x(T_x^{(1)} \ge \ell) = \pi(\{x\})\mathbb{E}_x\big[T_x^{(1)}\big].$$

Damit ist $\mathbb{E}_x\big[T_x^{(1)}\big] < \infty$ und folglich x positiv rekurrent. □

6.5.4 Konvergenz ins Gleichgewicht

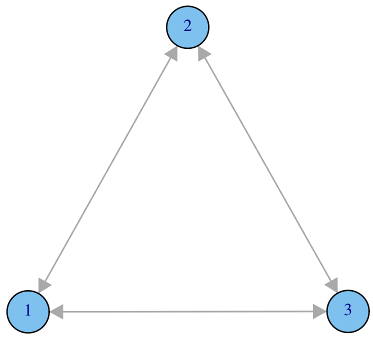
Definition 6.67. Für $x, y \in E$ sei $N(x,y) := \{n \in \mathbb{N} \mid p^n_{xy} > 0\}$.

i) $d_x := \text{ggT}\big(N(x,x)\big)$ heißt *Periode* des Zustands x.

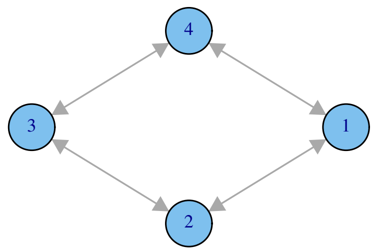
ii) Falls $d_x = d_y = d$ für alle $x, y \in E$, so heißt d die *Periode* der Markov-Kette X.

iii) X heißt *aperiodisch*, wenn $d = d_x = 1$ für alle $x \in E$.

Beispiel.



aperiodisch



Periode $d = 2$

Lemma 6.68. *i) Für $x \in E$ gibt es ein $n_x \in \mathbb{N}$ mit $p^{nd_x}_{xx} > 0$ für alle $n \geq n_x$.*

ii) Im irreduziblen Fall gilt $d_x = d_y$ für alle $x, y \in E$.

Beweis. i) Setze $\tilde{N} := \frac{N(x,x)}{d_x}$, dann ist $\tilde{N} \subset \mathbb{N}_0$ mit $\text{ggt}(\tilde{N}) = 1$ und $\tilde{N}$ ist abgeschlossen unter Addition.

Zeige: Es gibt ein $n' \in \mathbb{N}_0$ mit $n', n' + 1 \in \tilde{N}$. Seien dazu $n_0, n_0 + k \in \tilde{N}$. Falls $k = 1$, so ist $n' = n_0$. Ist $k > 1$, so gibt es ein $n_1 \in \tilde{N}$ mit $k \nmid n_1$, das heißt $n_1 = mk + r$ mit $m \in \mathbb{N}$ und $r \in \{1, \ldots, k-1\}$. Es gilt $\tilde{N} \ni (m+1)(n_0 + k) > (m+1)n_0 + n_1 \in \tilde{N}$ mit Differenz $(m+1)k - n_1 = k - r < k$. Iteriere dieses Argument maximal k mal und erhalte n'.

Sei nun $n_x := (n')^2$. Für $n \geq n_x$ schreibe $n = (n')^2 + n - (n'^2) = (n')^2 + kn' + r$ mit $r \in \{0, 1, \ldots, n' - 1\}$ und $k \in \mathbb{N}_0$, demnach ist $n = r(n'+1) + (n' - r + k)n' \in \tilde{N}$.

ii) Es gilt $N(x,y) + N(y,z) \subset N(x,z)$ für alle $x, y, z \in E$, denn $p^{m+n}_{xz} \geq p^m_{xy} p^n_{yz}$.

Wegen der Irreduzibilität existieren $m \in N(x,y)$ und $n \in N(y,x)$. Sei $k \geq n_y$ (das heißt $kd_y \in N(y,y)$), dann folgt $m + kd_y \in N(x,y)$ und $m + n + kd_y \in N(x,x)$, also

$$d_x \mid \underbrace{(m+n)}_{\in N(x,x)} + kd_y \text{ für alle } k \geq n_y.$$

Demnach gilt $d_x \mid d_y$ und mit vertauschten Rollen von x und y auch $d_y \mid d_x$, das heißt $d_x = d_y$. □

Bericht 6.69. Im irreduziblen Fall mit Periode $d > 1$ kann man $E = E_0 \uplus E_1 \uplus \cdots \uplus E_{d-1}$ disjunkt zerlegen, so dass

$$p_{xy} > 0, x \in E_i \;\Rightarrow\; y \in E_{(i+1) \bmod d}$$

gilt (vgl. [Kl, Satz 18.4]).

Satz 6.70. *Sei X eine aperiodische, irreduzible Markov-Kette mit invarianter Verteilung π. Dann gilt für jedes $x \in E$*

$$\sum_{y\in E} \left|\mathbb{P}_x(X_n = y) - \pi(\{y\})\right| \xrightarrow[n\to\infty]{} 0.$$

Beweis. Wir konstruieren eine *Kopplung* von $\mathbb{P}_x$ und $\mathbb{P}_\pi$.

$\overline{p}_{(x_1,x_2),(y_1,y_2)} := p_{x_1,y_1} p_{x_2,y_2}$ ist eine irreduzible stochastische Matrix auf $E \times E$ (die Irreduzibilität verwendet Lemma 6.68), $\overline{\pi}(\{x,y\}) := \pi(\{x\})\pi(\{y\})$ ist zugehörige invariante Verteilung. Sei $(X_n, Y_n)_n$ eine $\overline{p}$-Markov-Kette mit Startverteilung $\delta_x \otimes \pi$. Nach Satz 6.66 ist (X, Y) rekurrent, insbesondere ist $T := \inf\{n \in \mathbb{N}_0 : X_n = Y_n\} < \infty$ f.s. und es gilt

$$\begin{aligned}
\mathbb{P}(X_n = y, T \le n) &= \sum_{m=0}^{n} \sum_{x} \underbrace{\mathbb{P}(T = m, X_m = x, X_n = y)}_{=\mathbb{P}(T=m,X_m=x)p_{xy}^{n-m}=\mathbb{P}(T=m,Y_m=x)p_{xy}^{n-m}} \\
&= \sum_{m=0}^{n} \sum_{x} \mathbb{P}(T = m, Y_m = x, Y_n = y) = \mathbb{P}(Y_n = y, T \le n), \\
\mathbb{P}(X_n = y) &= \mathbb{P}(Y_n = y, T \le n) + \mathbb{P}(X_n = y, T > n) \le \mathbb{P}(Y_n = y) + \mathbb{P}(X_n = y, T > n)
\end{aligned}$$

und analog $\mathbb{P}(Y_n = y) \le \mathbb{P}(X_n = y) + \mathbb{P}(Y_n = y, T > n)$, somit

$$\sum_{y\in E} \left|\mathbb{P}_x(X_n = y) - \pi(\{y\})\right| = \sum_{y\in E} \left|\mathbb{P}(X_n = y) - \mathbb{P}(Y_n = y)\right| \le 2\mathbb{P}(T > n) \xrightarrow[n\to\infty]{} 0.$$

□

Beispiel (Erneuerungskette, vgl. Beispiel 6.50)**.** Sei $E = \mathbb{N}_0$, $\nu \in \mathcal{M}_1(\mathbb{N})$ und X eine Markov-Kette mit Übergangsmatrix $p_{0,j} = \nu(\{j+1\})$ für $j \in \mathbb{N}_0$, $p_{i,i-1} = 1$ für $i \in \mathbb{N}$ und $p_{ij} = 0$ sonst.

Sei $\mu := \sum_{x\in\mathbb{N}} x\nu(\{x\}) < \infty$ und $\mathrm{ggT}(\{x : \nu(\{x\}) > 0\}) = 1$, dann ist X aperiodisch, irreduzibel und positiv rekurrent mit (eindeutiger) invarianter Verteilung π, gegeben durch

$$\pi(\{x\}) = \frac{1}{\mu}\nu\big(\{x+1, x+2, \dots\}\big), \quad x \in \mathbb{N}_0$$

(falls $m := \sup\{x : \nu(\{x\}) > 0\} < \infty$ so muss man wörtlich auf $E' := \{0, 1, \dots, m+1\}$ einschränken).

Beweis. Die Irreduzibilität und Aperiodizität von X sind klar. Weiter gilt

$$\pi(\{j+1\})p_{j+1,j} + \pi(\{0\})p_{0,j} = \frac{1}{\mu}\big(\nu(\{j+2,\dots\}) \cdot 1 + 1 \cdot \nu(\{j+1\})\big) = \pi(\{j\}),$$

somit ist π die invariante Verteilung. □

Satz 6.71 (Diskreter Erneuerungssatz). *Mit der Darstellung*

$$X_n := \inf\left\{T_k - n \mid k \in \mathbb{N}_0, T_k \geq n\right\}$$

mit $\xi_1, \xi_2, \dots \sim \nu$ *unabhängig,* $\mathbb{E}[\xi_1] = \mu$, $T_0 := x_0$ *und* $T_m := x_0 + \xi_1 + \dots + \xi_m$ *für* $m \in \mathbb{N}$ *ergibt sich aus Satz 6.70*

$$\mathbb{P}\big(\exists\, k \in \mathbb{N} : T_k = n\big) = \mathbb{P}_{x_0}(X_n = 0) \xrightarrow[n\to\infty]{} \pi(\{0\}) = \frac{1}{\mu}.$$

6.5.5 Markov-Ketten und Randwertprobleme

Definition 6.72. Sei $f\colon E \to \mathbb{R}$, $pf(x) := \sum_{y\in E} p_{xy} f(y)$ (f sei derart, dass die Summe existiert). f heißt *harmonisch (für p)*, wenn $pf = f$. f heißt *subharmonisch*, wenn $pf \geq f$ und *superharmonisch*, falls $pf \leq f$.

Bemerkung. Ist f harmonisch und $(X_n)_n$ eine p-Markov-Kette, so ist $(f(X_n))_n$ ein Martingal.

Definition 6.73. Sei $\varnothing \neq A \subset E$ und $g : A \to \mathbb{R}$ beschränkt. f löst das *Dirichlet-Problem* (zu $p - I$) auf $E \setminus A$ mit Randwerten g (auf A), wenn gilt

$$\begin{aligned} (pf - f)(x) &= 0, & x \in E \setminus A, \\ f(x) &= g(x), & x \in A. \end{aligned}$$

Beobachtung 6.74. Sei $(X_n)_n$ p-Markovkette, $\tau_A := \inf\{n \in \mathbb{N}_0 : X_n \in A\}$ (die Treffzeit von $A \subset E$) und es gelte $\mathbb{P}_x(\tau_A < \infty) = 1$ für alle $x \in E$. Dann löst $f(x) := \mathbb{E}_x[g(X_{\tau_A})]$ das Dirichlet-Problem mit Randwerten g.

Beweis. Offenbar ist $f(x) = g(x)$ für $x \in A$, denn dann ist $\mathbb{P}_x(\tau_A = 0) = 1$. Für $x \notin A$ gilt wegen der Markov-Eigenschaft

$$\mathbb{E}_x[g(X_{\tau_A})|X_1 = y] = \begin{cases} g(y) = f(y), & y \in A, \\ \mathbb{E}_y[g(X_{\tau_A})] = f(y), & y \notin A. \end{cases}$$

Demnach ist für $x \notin A$

$$f(x) = \sum_y \mathbb{P}_x(X_1 = y)\mathbb{E}_x[g(X_{\tau_A})|X_1 = y] = \sum_y p_{xy} f(y) = pf(x).$$

$\square$

Bemerkung. Auch $\tilde{f}(x) = \mathbb{E}_x\big[g(X_{\tau_A})\mathbf{1}_{\{\tau_A<\infty\}}\big]$ löst das Dirichlet-Problem. Ist $\mathbb{P}_x(\tau_A = \infty) > 0$, so kann es verschiedene Lösungen geben.

Sei $\tilde{p}_{xy} := \mathbf{1}_{E\setminus A}(x)p_{xy} + \mathbf{1}_A \delta_{xy}$ ($\tilde{p}$ ist die Übergangsmatrix von $\tilde{X}_n := X_{n\wedge\tau_A}$, der in A absorbierten Kette). Wir treffen folgende *Annahme:*

$$\forall\, x \in E \setminus A, y \in E : \exists\, n \in \mathbb{N}_0 : \tilde{p}^n_{xy} > 0 \quad \text{und} \quad \inf_{x\in E} \mathbb{P}_x(\tau_A < \infty) = 1 \qquad (\mathrm{Irr}_A)$$

Satz 6.75 (Maximumprinzip). *Es gelte (Irr$_A$). Sei $f: E \to \mathbb{R}$ mit $(pf - f)(x) = 0$ für $x \in E \setminus A$ und es gebe $x^* \in E \setminus A$ mit $f(x^*) = \sup_{x \in E} f(x)$. Dann ist f konstant.*

Beweis. Es gilt $f(x) = \tilde{p} f(x) = \cdots = \tilde{p}^n f(x)$, insbesondere

$$f(x^*) = \tilde{p}^n f(x^*) = \sum_y \tilde{p}^n_{x^* y} f(y) \leq f(x^*),$$

also gilt $f(y) = f(x^*)$ für alle $y \in \{z \in E \mid \tilde{p}^n_{x^* z} > 0\} =: A_n$. Nach Annahme ist $\bigcup_{n=1}^\infty A_n = E$. □

Satz 6.76. *Unter der Annahme (Irr$_A$) ist für jede beschränkte Funktion $g: A \to \mathbb{R}$ das Dirichlet-Problem mit Randwerten g eindeutig lösbar durch $f(x) = \mathbb{E}_x[g(X_{\tau_A})]$.*

Beweis. Seien f_1 und f_2 Lösungen. Da g beschränkt ist, sind gemäß Satz 6.75 (Maximumprinzip) auch f_1 und f_2 beschränkt mit $\|f_1\|_\infty, \|f_2\|_\infty \leq \|g\|_\infty$.

$\hat{f} := f_1 - f_2$ erfüllt $(p\hat{f} - \hat{f})|_{E \setminus A} \equiv 0$ und $\hat{f}|_A \equiv 0$, also muss $\hat{f} \equiv 0$ sein, denn ein $x^* \in E \setminus A$ mit $f(x^*) > 0$ oder $-f(x^*) > 0$ ergäbe einen Widerspruch zum Maximumprinzip. □

Beispiel 6.77 (Einfache Irrfahrt auf $\mathbb{Z}$ mit Drift). Sei $E = \mathbb{Z}$, $r \in (\frac{1}{2}, 1)$ und

$$p_{xy} = r\delta_{y,x+1} + (1-r)\delta_{y,x-1}.$$

$\varphi(x) := \left(\frac{1-r}{r}\right)^x$ ist harmonisch. Sei $A := \big((-\infty, a] \cup [b, \infty)\big) \cap \mathbb{Z}$ mit $-a, b \in \mathbb{N}$ und betrachte $g(a) = 1$, $g(b) = 0$. Für $a < x < b$ ist

$$\mathbb{P}_x(\tau_{\{a\}} < \tau_{\{b\}}) = \mathbb{E}_x[g(X_{\tau_A})] = \frac{\left(\frac{1-r}{r}\right)^b - \left(\frac{1-r}{r}\right)^x}{\left(\frac{1-r}{r}\right)^b - \left(\frac{1-r}{r}\right)^a}.$$

6.5.6 Beispiel Gibbs-Sampler und Isingmodell

Das Ising-Modell[5] ist ein einfaches thermodynamisches und quantenmechanisches Modell für (Ferro-) Magnetismus von Kristallen.

- Die Atome sitzen auf den Knoten des Gitters, wir denken an $\Lambda = \{0, 1, \ldots, L-1\}^2$ für ein $L \in \mathbb{N}$.
- Jedes Atom $(i,j) \in \Lambda$ besitzt ein magnetisches Moment $x_{(i,j)} \in \{-1, +1\}$ (*Spin*).
- Atome wechselwirken (nur) mit ihren direkten Nachbarn auf dem Gitter, und sie bevorzugen dieselbe (Spin-) Orientierung wie ihre Nachbarn zu haben.
- Die *Energiefunktion* eines Zustands $x \in \{\pm 1\}^\Lambda$ ist gegeben durch

  $$H(x) = -\sum_{(i,j) \sim (i',j')} x_{(i,j)} x_{(i',j')},$$

  wobei $(i,j) \sim (i',j')$ bedeutet, dass (i,j) und (i',j') benachbarte Gitterpunkte sind. Demnach ist die Energie eines Zustands umso kleiner, je mehr „gleichsinnige" Nachbarpaare (+/+ oder -/-) es gibt.

[5] Ernst Ising, 1924

- Wegen thermischer Fluktuationen ist der („mikroskopische") Zustand eines Systems bei Temperatur $T > 0$ zufällig.

Bei Temperatur $T > 0$ hat ein Zustand x die Wahrscheinlichkeit

$$\mu_\beta(x) := \frac{1}{Z_\beta} \exp\big(-\beta H(x)\big)$$

mit $\beta = \frac{1}{T}$ (*inverse Temperatur*) und Normierungskonstante (*Zustandssumme*)

$$Z_\beta := \sum_{x\in\{\pm1\}^\Lambda} \exp\big(-\beta H(x)\big).$$

μ_β ist ein Wahrscheinlichkeitsmaß auf $\{\pm1\}^\Lambda$, die *Gibbs*[6]*-Verteilung* (oft auch *Boltzmann*[7]*-Verteilung*).

Um „typische" Konfgurationen im thermischen Gleichgewicht zu beschreiben, müssen wir Erwartungswerte bezüglich μ_β ausrechnen. Da Z_β eine Summe über $2^{|\Lambda|}$ Terme ist, kann man $\mu_\beta(x)$ aber nur sehr schwer ausrechnen (zum Beispiel für ein Gitter der Größe 100×100 sind dies $2^{10000} \approx 2 \cdot 10^{3010}$ Summanden).

Lösungsvorschlag: Finde eine Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$, die $\mu_\beta(x)$ als einziges Gleichgewicht besitzt, denn dann hat (für $n \gg 1$) X_n (ungefähr) Verteilung μ_β.

Gibbs-Sampler. Für $(i,j) \in \Lambda$, $x \in \{\pm1\}^\Lambda$ sei $x^{(i,j),+} \in \{\pm1\}^\Lambda$ gegeben durch

$$x^{(i,j),+}_{(i',j')} = \begin{cases} +1, & (i',j') = (i,j) \\ x_{(i',j')} & (i',j') \neq (i,j) \end{cases}$$

und analog $x^{(i,j),-}$. Definiere $(A(x,y))_{x,y\in\{\pm1\}^\Lambda}$ durch

$$\begin{aligned} A(x, x^{(i,j),+}) &:= \frac{1}{|\Lambda|} \frac{\mu_\beta(x^{(i,j),+})}{\mu_\beta(x^{(i,j),+}) + \mu_\beta(x^{(i,j),-})}, \\ A(x, x^{(i,j),-}) &:= \frac{1}{|\Lambda|} \frac{\mu_\beta(x^{(i,j),-})}{\mu_\beta(x^{(i,j),+}) + \mu_\beta(x^{(i,j),-})}, \\ A(x,y) &:= 0, \qquad \text{falls } y \notin \cup_{(i,j)}\{x^{(i,j),+}, x^{(i,j),-}\} \end{aligned}$$

Beachte: A ist eine (irreduzible und aperiodische) stochastische Matrix und man braucht Z_β nicht zu kennen, um A zu bestimmen. Interpretation von A: Wähle zufällig einen Gitterpunkt, flippe den Spin dort gemäß μ_β, bedingt auf alle anderen Spins. Es gilt

$$\mu_\beta(y)A(y,z) = \mu_\beta(z)A(z,y) \quad \text{für alle } y,z \in \{\pm1\}^\Lambda$$

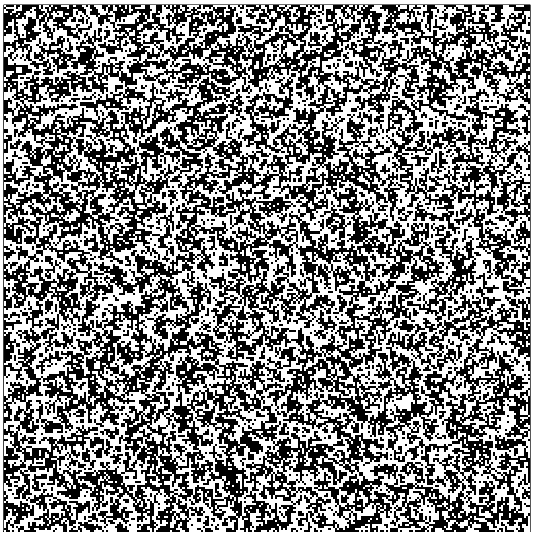
(es genügt hier, dies für $y = x^{(i,j),+}$, $z = x^{(i,j),-}$ für alle $x \in \{\pm1\}^\Lambda$, $(i,j) \in \Lambda$ zu prüfen) und somit

$$\sum_z \mu_\beta(z)A(z,y) = \mu_\beta(y) \quad \text{für alle } y \in \{\pm1\}^\Lambda.$$

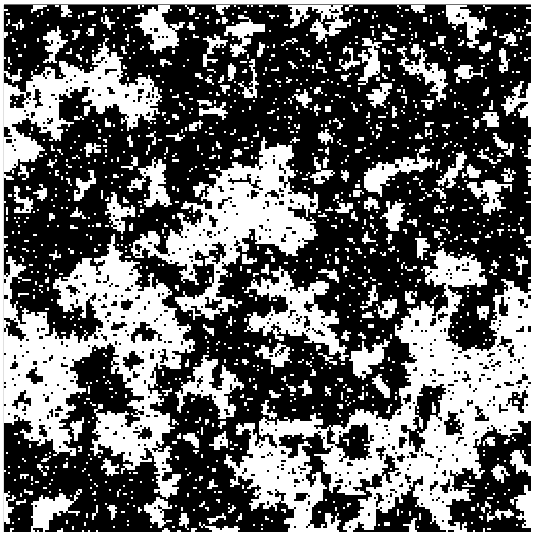
μ_β ist ein (reversibles) Gleichgewicht für A.

[6]Josiah Willard Gibbs (1839-1903)

[7]Ludwig Boltzmann, 1844-1906



(a) $\beta = 0.2$



(b) $\beta = 0.43$

Abbildung 6.1: Simulationen eines Zustands gemäß μ_β für $L = 256$

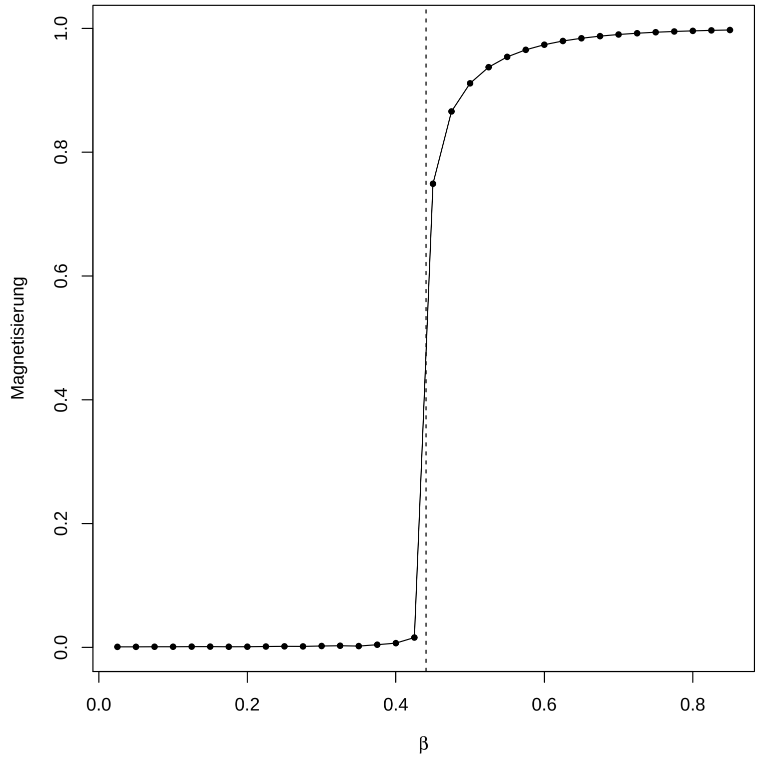
Für einen „Mikro“-Zustand $x\{\pm 1\}^\Lambda$ ist

$$m(x) := \frac{1}{|\Lambda|} \sum_{(i,j)\in\Lambda} x_{(i,j)}$$

die Magnetisierung (pro Spin) und

$$m_\beta := \sum_{x\{\pm 1\}^\Lambda} \mu_\beta(x)\,|m(x)|$$

die mittlere (absolute) Magnetisierung bei inverser Temperatur β.



mittlere (absolute) Magnetisierung als Funktion von β
(basierend auf einer Simulation für $L = 1000$)

Kapitel 7

(Etwas) Ergodentheorie

Definition 7.1. Ein stochastischer Prozess $(X_n)_{n\in\mathbb{N}_0}$ mit Werten in einem polnischen Raum E heißt *stationär*, wenn für alle $m \in \mathbb{N}$ gilt

$$\mathbb{P}((X_n)_{n\in\mathbb{N}_0} \in \cdot\,) = \mathbb{P}((X_{n+m})_{n\in\mathbb{N}_0} \in \cdot\,).$$

Beispiel. i) Sind $X_1, X_2, X_3, \dots$ unabhängig und identisch verteilt, so ist $(X_n)_{n\in\mathbb{N}_0}$ stationär.

ii) Ist $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette mit invarianter Verteilung π und $X_0 \sim \pi$, so ist $(X_n)_{n\in\mathbb{N}_0}$ stationär.

iii) Für $k \in \mathbb{Z}$ seien Y_k unabhängige und identisch verteilte, reelle Zufallsvariablen. Seien $m \in \mathbb{N}_0$, $c_0, c_1, \dots, c_m \in \mathbb{R}$ und $X_n := \sum_{j=0}^{m} c_j Y_{n-j}$ („gleitendes Mittel"). Dann ist $(X_n)_{n\in\mathbb{N}_0}$ stationär.

Definition 7.2. Sei $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\tau : \Omega \to \Omega$ messbar.

i) $A \in \mathscr{A}$ heißt *invariant*, falls $\tau^{-1}(A) = A$.

$A \in \mathscr{A}$ heißt *quasi-invariant*, falls $\mathbb{P}(\tau^{-1}(A) \triangle A) = 0$.

$\mathcal{I} := \{A \in \mathscr{A} \mid A \text{ invariant}\}$ heißt die *σ-Algebra der invarianten Ereignisse* (auch *invariante σ-Algebra*).

ii) τ heißt *maßtreu* (auch *maßerhaltend*), falls $\mathbb{P}(\tau^{-1}(A)) = \mathbb{P}(A)$ für alle $A \in \mathscr{A}$.

$(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ heißt dann *(maßerhaltendes) dynamisches System*.

iii) $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ heißt *ergodisch*[1], wenn $\mathcal{I}$ $\mathbb{P}$-trivial ist, das heißt $\mathbb{P}(A) \in \{0, 1\}$ für alle $A \in \mathcal{I}$.

Bemerkung 7.3. Sei $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ ein maßerhaltendes dynamisches System.

i) Sei $f : \Omega \to \mathbb{R}$ messbar. f ist genau dann $\mathcal{I}$-messbar, wenn $f = f \circ \tau$.

ii) Ist $A \in \mathcal{I}$ mit $\mathbb{P}(A) \in (0, 1)$, so ist $(A, \mathscr{A}_{|A}, \mathbb{P}(\,\cdot \mid A), \tau_{|A})$ ein maßerhaltendes dynamisches System ($\mathscr{A}_{|A}$ bezeichnet dabei die Spur-σ-Algebra über A).

[1] Zur Etymologie des Worts „ergodisch" vgl. auch G. Gallavotti, J. Stat. Phys. 78, 1571-1589 (1995).

Beispiel 7.4 (Rotation des Einheitskreises). Sei $\Omega = [0,1)$ (mit „periodischem Rand"), $\mathscr{A} = \mathscr{B}(\Omega)$, $\mathbb{P}$ das Lebesgue-Maß auf Ω und $\tau_r : x \mapsto x + r \bmod 1$ (alternative Parametrisierung via $x \mapsto e^{2\pi i x}$).

Sei zunächst $r = \frac{p}{q} \in (0,1)$ mit $p, q \in \mathbb{N}$ teilerfremd. Dann gilt

$$\tau_r^q(x) = (x + qr) \bmod 1 = (x + p) \bmod 1 = x \bmod 1 = x,$$

das heißt τ_r hat periodische Orbits. Sei $A_0 := [0, \frac{1}{2q})$ und $A := \bigcup_{n=0}^{q-1} \tau_r^n(A_0)$. Dann ist $\tau_r^{-1}(A) = A$ und $\mathbb{P}(A) = \frac{1}{2} \notin \{0,1\}$, also ist $(\Omega, \mathscr{A}, \mathbb{P}, \tau_r)$ in diesem Fall nicht ergodisch.

Sei nun $r \in (0,1) \setminus \mathbb{Q}$. Zeige: Die Orbits liegen dicht.

Setze dazu $x_n := \tau_r^n(0)$. Es gilt $x_n \neq x_m$ für $n \neq m$, denn sonst wäre $rm = rn + k$, also $r = \frac{k}{m-n} \in \mathbb{Q}$, was zu einem Widerspruch führen würde. Für alle $N \in \mathbb{N}$ existieren $m, n \in \mathbb{N}_0$ mit $|x_n - x_m| < \frac{1}{N}$ („Schubfachprinzip"), also existiert auch ein $k \in \mathbb{N}$ mit $0 < x_k \leq \frac{1}{N}$ (wähle zum Beispiel $k = |n - m|$). Für $L := \lceil \frac{1}{x_k} \rceil$ haben $0 < x_k < x_{2k} < \ldots < x_{Lk} < 1$ jeweils einen Abstand kleiner gleich $\frac{1}{N}$, folglich liegen die Orbits dicht.

Sei $A \in \mathcal{I}$ mit $\mathbb{P}(A) > 0$. Zunächst: Falls es ein $x \in A$ und ein $\varepsilon > 0$ gibt mit $(x - \varepsilon, x + \varepsilon) \subset A$, so folgt

$$[0,1) = \bigcup_{n=0}^{\infty} \tau_r^n((x - \varepsilon, x + \varepsilon)) \subset A,$$

also $A = [0,1)$ und somit $\mathbb{P}(A) = 1$.

Für den allgemeinen Fall benutzen wir den Dichtesatz von Lebesgue:

Bericht 7.5 (Dichtesatz von Lebesgue[2]). Sei λ das Lebesgue-Maß auf $\mathbb{R}$ und $A \in \mathscr{B}(\mathbb{R})$ mit $\lambda(A) > 0$. Dann gilt für $\lambda(\,\cdot \cap A)$-fast alle x und $B_\varepsilon(x) := (x - \varepsilon, x + \varepsilon)$

$$\lim_{\varepsilon \searrow 0} \frac{\lambda(A \cap B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} = 1,$$

das heißt

$$\lambda\left(\left\{x \in A \,\middle|\, \liminf_{\varepsilon \searrow 0} \frac{\lambda(A \cap B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} < 1\right\}\right) = 0.$$

Es gibt ein $x \in A$ mit $\frac{\mathbb{P}(A \cap B_\varepsilon(x))}{\mathbb{P}(B_\varepsilon(x))} > \frac{3}{4}$ für $0 < \varepsilon < \varepsilon_0$. Falls $A^c = \varnothing$, so ist $\mathbb{P}(A) = 1$. Andernfalls sei $y \in A^c$. Zu $\varepsilon > 0$ gibt es ein $n \in \mathbb{N}$ mit $B_\varepsilon(\tau_r^n(x)) \subset B_{2\varepsilon}(y)$, demnach

$$\begin{aligned}\frac{\mathbb{P}(A^c \cap B_{2\varepsilon}(y))}{\mathbb{P}(B_{2\varepsilon}(y))} &= \underbrace{\frac{\mathbb{P}(A^c \cap \tau_r^n(B_\varepsilon(x)))}{4\varepsilon}}_{= \frac{1}{4\varepsilon}\mathbb{P}(\tau_r^n(A^c \cap B_\varepsilon(x)))} + \underbrace{\frac{\mathbb{P}(A^c \cap (B_{2\varepsilon}(y) \setminus \tau_r^n(B_\varepsilon(x))))}{4\varepsilon}}_{\leq \frac{1}{4\varepsilon}\mathbb{P}(B_{2\varepsilon}(y) \setminus \tau_r^n(B_\varepsilon(x))) = \frac{2\varepsilon}{4\varepsilon} = \frac{1}{2}} \\ &\leq \frac{1}{4\varepsilon} \underbrace{\mathbb{P}(A^c \cap B_\varepsilon(x))}_{\leq \frac{1}{4}\mathbb{P}(B_\varepsilon(x))} + \frac{1}{2} \leq \frac{1}{4\varepsilon}\frac{2\varepsilon}{4} + \frac{2\varepsilon}{4\varepsilon} = \frac{5}{8}.\end{aligned}$$

Damit gilt $\liminf_{\varepsilon \searrow 0} \frac{\mathbb{P}(A^c \cap B_\varepsilon(y))}{\mathbb{P}(B_\varepsilon(y))} < 1$ für jedes $y \in A^c$, wonach mit dem Dichtesatz von Lebesgue $\mathbb{P}(A^c) = 0$ folgt.

Somit ist $(\Omega, \mathscr{A}, \mathbb{P}, \tau_r)$ für $r \in (0,1) \setminus \mathbb{Q}$ ergodisch.

[2] Siehe z.B. J. Elstrodt, *Maß- und Integrationstheorie*, Satz VII.4.9

Beispiel und Definition 7.6. Sei $X = (X_n)_{n\in\mathbb{N}_0}$ ein stochastischer Prozess mit Werten in E, realisiert als kanonischer Prozess auf $(\Omega, \mathscr{A}, \mathbb{P})$ mit $\Omega = E^{\mathbb{N}_0}$, $\mathscr{A} = \mathscr{B}(E)^{\otimes\mathbb{N}_0}$, $\mathbb{P} = \mathscr{L}((X_n)_{n\in\mathbb{N}_0})$ mit $X_i\colon \Omega \to E,\ (x_0, x_1, \dots) \mapsto x_i$ und

$$\tau\colon \Omega \to \Omega,\ (x_0, x_1, \dots) \mapsto (x_1, x_2, \dots).$$

Insbesondere gilt $X_n(\omega) = X_0(\tau^n(\omega))$. X ist genau dann stationär, wenn $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ ein maßerhaltendes dynamisches System ist. Ein stationärer stochastischer Prozess X heißt *ergodisch*, wenn dies für $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ gilt.

Beispiel.

i) Sind $X_1, X_2, \dots$ unabhängig und identisch verteilt, so ist $(X_n)_n$ ergodisch. (Dies folgt z.B. mittels Kolmogorovs-0-1-Gesetz.)

ii) Wir werden sehen (siehe Satz 7.13 unten): Aperiodische, irreduzible Markov-Ketten mit invarianter Verteilung sind ergodisch.

Satz 7.7 (Ergodensatz). *Sei $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ ein maßerhaltendes dynamisches System, $f\colon \Omega \to \mathbb{R}$ messbar, $X_n := f \circ \tau^n$ und $S_n := \sum_{j=0}^{n-1} X_j$. Falls $f \in \mathcal{L}^p(\mathbb{P})$ für ein $p \geq 1$, so gilt*

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{j=0}^{n-1} f \circ \tau^j \xrightarrow[n\to\infty]{} \mathbb{E}[X_0 \mid \mathcal{I}] \quad \text{f.s. und in } \mathcal{L}^p.$$

Die f.s.-Konvergenz-Aussage in Satz 7.7 heißt in der Literatur auch „Birkhoffs[3] individueller Ergodensatz“, die $\mathcal{L}^p$-Konvergenz „von Neumanns[4] statistischer Ergodensatz.“

Lemma 7.8 (Hopfs[5] Maximal-Ergodenlemma). *In der Situation von Satz 7.7 sei $X_0 \in \mathcal{L}^1$ und $M_n := \max\{0, S_1, S_2, \dots, S_n\}$. Dann gilt $\mathbb{E}[X_0 \mathbf{1}_{\{M_n>0\}}] \geq 0$.*

Beweis. Für $1 \leq k \leq n$ gilt

$$X_0 + M_n \circ \tau \geq X_0 + S_k \circ \tau = S_{k+1},$$

sowie $X_0 \geq S_1 - M_n \circ \tau$. Damit ist

$$X_0 \geq \max\{S_1, \dots, S_n\} - M_n \circ \tau.$$

Weiterhin ist

$$\{M_n > 0\}^c = \{M_n = 0\} \cap \{M_n \circ \tau \geq 0\} \subset \{M_n - M_n \circ \tau \leq 0\},$$

also

$$\begin{aligned} \mathbb{E}\big[X_0 \mathbf{1}_{\{M_n>0\}}\big] &\geq \mathbb{E}\big[(\max\{S_1, \dots, S_n\} - M_n \circ \tau)\mathbf{1}_{\{M_n>0\}}\big] \\ &= \mathbb{E}\big[(M_n - M_n \circ \tau)\mathbf{1}_{\{M_n>0\}}\big] \\ &\geq \mathbb{E}[M_n - M_n \circ \tau] = \mathbb{E}[M_n] - \mathbb{E}[M_n] = 0. \end{aligned}$$

□

[3] George David Birkhoff 1884-1944; 1931
[4] John von Neumann 1903-1957; 1931
[5] Eberhard Hopf 1902-1983

Beweis von Satz 7.7. Sei ohne Einschränkung $\mathbb{E}[X_0 \mid \mathcal{I}] = 0$ (ansonsten betrachte $\widetilde{X}_n := X_n - \mathbb{E}[X_n \mid \mathcal{I}] = X_n - \mathbb{E}[X_0 \mid \mathcal{I}]$). Sei

$$Z := \limsup_n \tfrac{1}{n} S_n,$$

$\varepsilon > 0$ und $F := \{Z > \varepsilon\} \in \mathcal{I}$. Weiter sei

$$X_n^\varepsilon := (X_n - \varepsilon)\mathbf{1}_F, \quad S_n^\varepsilon := \sum_{k=0}^{n-1} X_k^\varepsilon \text{ und } M_n^\varepsilon := \max\{0, S_1^\varepsilon, \dots, S_n^\varepsilon\}.$$

Für $F_n := \{M_n^\varepsilon > 0\}$ gilt $F_1 \subset F_2 \subset \dots$ und

$$\bigcup_{n=1}^\infty F_n = \left\{\sup_k \frac{1}{k} S_k^\varepsilon > 0\right\} = \left\{\sup_k \frac{1}{k} S_k^\varepsilon > 0\right\} \cap F = F,$$

daher folgt mit Lemma 7.8 und monotoner Konvergenz

$$0 \le \mathbb{E}[X_0^\varepsilon \mathbf{1}_{F_n}] \xrightarrow[n\to\infty]{} \mathbb{E}[X_0^\varepsilon \mathbf{1}_F] = \mathbb{E}[X_0^\varepsilon].$$

Demnach gilt

$$0 \le \mathbb{E}[X_0^\varepsilon] = \mathbb{E}[(X_0 - \varepsilon)\mathbf{1}_F] = \mathbb{E}\big[\mathbb{E}[(X_0 - \varepsilon)\mathbf{1}_F \mid \mathcal{I}]\big] = \mathbb{E}\big[\mathbf{1}_F(\mathbb{E}[X_0 \mid \mathcal{I}] - \varepsilon)\big] = -\varepsilon\mathbb{P}(F),$$

das heißt $\mathbb{P}(F) = 0$ und mit $\varepsilon \searrow 0$ folgt $\limsup_n \frac{1}{n} S_n \le 0$ f.s. Ersetze nun X_n durch $-X_n$ und erhalte

$$\frac{1}{n} S_n \xrightarrow[n\to\infty]{} 0 \text{ f.s.}$$

Sei nun $p \ge 1$ und $X_0 \in \mathcal{L}^p$. Zeige: $\{|\frac{1}{n} S_n|^p \mid n \in \mathbb{N}\}$ ist gleichgradig integrierbar. $\{|X_0|^p\}$ ist gleichgradig integrierbar, also existiert nach Erinnerung B.23 eine monoton wachsende, konvexe Funktion $\varphi\colon [0,\infty) \to [0,\infty)$ mit $\frac{\varphi(x)}{x} \to \infty$, $x \to \infty$ und $\mathbb{E}[\varphi(|X_0|^p)] < \infty$. Nach der Jensen-Ungleichung gilt $|\frac{1}{n} S_n|^p \le \frac{1}{n} \sum_{j=0}^{n-1} |X_j|^p$, also folgt wegen der Monotonie und Konvexität von φ

$$\varphi\left(|\frac{1}{n} S_n|^p\right) \le \varphi\left(\frac{1}{n}\sum_{j=0}^{n-1} |X_j|^p\right) \le \frac{1}{n}\sum_{j=0}^{n-1} \varphi(|X_j|^p)$$

und somit

$$\sup_n \mathbb{E}\left[\varphi\left(|\frac{1}{n} S_n|^p\right)\right] \le \sup_n \frac{1}{n} n \mathbb{E}\big[\varphi(|X_0|^p)\big] = \mathbb{E}\big[\varphi(|X_0|^p)\big] < \infty.$$

Demnach ist $\{|\frac{1}{n} S_n|^p \mid n \in \mathbb{N}\}$ gleichgradig integrierbar. Dies zusammen mit $\frac{1}{n} S_n \to 0, n \to \infty$ f.s. impliziert dies auch $\frac{1}{n} S_n \to 0, n \to \infty$ in $\mathcal{L}^p$. □

Beispiel 7.9. Sei X eine positiv rekurrente, irreduzible Markovkette auf einer abzählbaren Menge E. Sei π die eindeutige invariante Verteilung und $\mathbb{P}_\pi := \sum_{x \in E} \pi(\{x\}) \mathbb{P}_x$. Dann ist X ein stationärer Prozess (zum Beispiel auf $\Omega = E^{\mathbb{N}_0}$ mit $\tau : (x_n)_n \mapsto (x_{n+1})_n$) und $(\Omega, (2^E)^{\otimes \mathbb{N}_0}, \mathbb{P}_\pi, \tau)$ ist ergodisch.

Beweis. Sei $A \in \mathcal{I} \subset \mathcal{T} = \bigcap_n \sigma(X_n, X_{n+1}, \ldots)$. Es gilt $\mathbb{P}_\pi(X \in A \mid \mathscr{F}_\eta) = \mathbb{P}_{X_\eta}(X \in A)$ für jede f.s. endliche Stoppzeit η, denn für $B \in \mathscr{F}_\eta$ gilt

$$\begin{aligned}\mathbb{E}_\pi[\mathbf{1}_{\{X\in B\}}\mathbf{1}_{\{X\in A\}}] &= \sum_{n=0}^{\infty}\sum_{x\in E} \mathbb{P}_\pi(X \in B, \eta = n, X_n = x, \underbrace{X \in A}_{= X\circ\tau^n \in A}) \\ &= \sum_{n=0}^{\infty}\sum_{x\in E} \mathbb{P}_\pi(X \in B, \eta = n, X_n = x)\mathbb{P}_x(X \in A) \\ &= \mathbb{E}_\pi\big[\mathbf{1}_{\{X\in B\}}\mathbb{P}_{X_\eta}(X \in A)\big].\end{aligned}$$

Insbesondere ist mit $T_x^{(1)} := \inf\{n > 0 \mid X_n = x\}$ ($< \infty$ f.s., da X irreduzibel und rekurrent) und der Markov-Eigenschaft

$$\mathbb{P}_\pi(X \in A) = \mathbb{E}_\pi\big[\mathbb{P}_\pi(X \in A \mid \mathscr{F}_{T_x^{(1)}})\big] = \mathbb{E}_\pi\big[\mathbb{P}_{X_{T_x^{(1)}}}(X \in A)\big] = \mathbb{E}_\pi\big[\mathbb{P}_x(X \in A)\big] = \mathbb{P}_x(X \in A)$$

für jedes $x \in E$. Somit ist

$$\begin{aligned}\mathbb{P}_\pi(X \in A) &= \mathbb{P}_{X_n}(X \in A) \\ &= \mathbb{P}_\pi(\underbrace{X \in A}_{= X\circ\tau^n \in A} \mid X_0, \ldots, X_n) \xrightarrow[n\to\infty]{} \mathbb{P}_\pi(X \in A \mid \sigma(X_0, X_1, \ldots)) = \mathbf{1}_{\{X\in A\}}.\end{aligned}$$

Für beliebiges $A \in \mathcal{I}$ gilt demnach $\mathbb{P}_\pi(X \in A) \in \{0, 1\}$, das heißt $(\Omega, (2^E)^{\otimes\mathbb{N}_0}, \mathbb{P}_\pi, \tau)$ ist ergodisch. □

Bemerkung 7.10. Insbesondere gilt in der Situation von Beispiel 7.9 für jedes $f \in \mathcal{L}^1(\pi)$

$$\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \xrightarrow[n\to\infty]{} \mathbb{E}_\pi(f(X_0)) = \sum_{x\in E} \pi(\{x\})f(x) \quad \text{f.s.}$$

Definition 7.11. Ein maßerhaltendes dynamisches System $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ heißt *mischend*, wenn für alle $A, B \in \mathscr{A}$ gilt

$$\lim_{n\to\infty} \mathbb{P}(A \cap \tau^{-n}(B)) = \mathbb{P}(A)\mathbb{P}(B).$$

Ein stochastischer Prozess $X = (X_n)_{n\in\mathbb{N}_0}$ mit Werten in E heißt *mischend*, wenn dies für seine Darstellung als kanonischer Prozess auf $E^{\mathbb{N}_0}$ gilt, das heißt

$$\lim_{n\to\infty} \mathbb{P}\big((X_m)_{m\in\mathbb{N}_0} \in A, (X_{m+n})_{m\in\mathbb{N}_0} \in B\big) = \mathbb{P}(X \in A)\mathbb{P}(X \in B)$$

(für alle messbaren Teilmengen $A, B \subset E^{\mathbb{N}_0}$)

Beobachtung 7.12. Ist $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ mischend, so ist es auch ergodisch.

Beweis. Sei $(\Omega, \mathscr{A}, \mathbb{P}, \tau)$ mischend und $A \in \mathcal{I}$. Für alle $n \in \mathbb{N}$ gilt $\mathbb{P}(A) = \mathbb{P}(A \cap \tau^{-n}(A))$, demnach gilt

$$\mathbb{P}(A) = \mathbb{P}(A \cap \tau^{-n}(A)) \xrightarrow[n\to\infty]{} \mathbb{P}(A)\mathbb{P}(A) = \mathbb{P}(A)^2$$

und somit $\mathbb{P}(A) \in \{0, 1\}$. □

Bemerkung. Die irrationale Rotation aus Beispiel 7.4 ist ergodisch, aber nicht mischend. Sei $\tau\colon x \mapsto x + r \bmod 1$ für $x \in [0,1)$ und $r \in (0,1) \setminus \mathbb{Q}$, so ist $\tau^{k_n}(0) \in (\frac{1}{4}, \frac{3}{4})$ für eine Folge $k_n \nearrow \infty$. Für $A := [0, \frac{1}{4})$ ist $A \cap \tau^{-k_n}(A) = \varnothing$, also $\liminf_n \mathbb{P}(A \cap \tau^{-n}(A)) = 0 \neq \frac{1}{16} = \mathbb{P}(A)^2$.

Satz 7.13. *Ist $X = (X_n)_{n\in\mathbb{N}_0}$ eine positiv rekurrente, irreduzible und aperiodische Markov-Kette (auf einer diskreten Menge E) mit invarianter Verteilung π, so ist X mischend.*

Beweis. Stelle X dar als kanonischer Prozess auf $E^{\mathbb{N}_0}$. Seien $A, B \in (2^E)^{\mathbb{N}_0}$ und $\varepsilon > 0$. Nach dem Approximationssatz für Maße gibt es ein $N \in \mathbb{N}$ und ein $\widetilde{A} \in E^{\{0,1,\dots,N\}}$, sodass mit $A_\varepsilon := \widetilde{A} \times E^{\mathbb{N}\setminus\{0,1,\dots,N\}}$ gilt $\mathbb{P}_\pi(A \triangle A_\varepsilon) < \varepsilon$.

Weiter gilt (für $n > N$)

$$\begin{aligned}
\mathbb{P}_\pi(A_\varepsilon \cap \tau^{-n}(B)) &= \mathbb{P}_\pi((X_0,\dots,X_N) \in \widetilde{A}, (X_{m+n})_{m\in\mathbb{N}_0} \in B) \\
&= \sum_{x,y\in E} \mathbb{E}_\pi\big[\mathbf{1}_{A_\varepsilon}\mathbf{1}_{\{X_N=x\}}\mathbf{1}_{\{X_n=y\}}\mathbf{1}_B(X_n, X_{n+1},\dots)\big] \\
&= \sum_{x,y\in E} \mathbb{E}_\pi\big[\mathbf{1}_{A_\varepsilon}\mathbf{1}_{\{X_N=x\}}\big] p_{xy}^{n-N}\mathbb{P}_y(X \in B).
\end{aligned}$$

Nach Satz 6.70 gilt $p_{xy}^{n-N} \to \pi(\{y\})$ für $n \to \infty$, demnach ist

$$\lim_{n\to\infty} \mathbb{P}_\pi(A_\varepsilon \cap \tau^{-n}(B)) = \sum_{x,y\in E} \mathbb{E}_\pi\big[\mathbf{1}_{A_\varepsilon}\mathbf{1}_{\{X_N=x\}}\big]\pi(\{y\})\mathbb{P}_y(X \in B) = \mathbb{P}_\pi(A_\varepsilon)\mathbb{P}_\pi(B).$$

Wegen $|\mathbb{P}_\pi(A_\varepsilon \cap \tau^{-n}(B)) - \mathbb{P}_\pi(A \cap \tau^{-n}(B))| \le \mathbb{P}_\pi(A_\varepsilon \triangle A) < \varepsilon$ folgt

$$\mathbb{P}_\pi(A_\varepsilon)\mathbb{P}_\pi(B) - \varepsilon \le \liminf_n \mathbb{P}_\pi(A \cap \tau^{-n}(B)) \le \limsup_n \mathbb{P}_\pi(A \cap \tau^{-n}(B)) \le \mathbb{P}_\pi(A_\varepsilon)\mathbb{P}_\pi(B) + \varepsilon$$

und $\mathbb{P}_\pi(A_\varepsilon) \to \mathbb{P}_\pi(A)$ für $\varepsilon \searrow 0$, also folgt mit $\varepsilon \searrow 0$ die Behauptung. □

Bemerkung. Wenn X periodisch ist mit Periode $d > 0$, so ist X nicht mischend: Wenn man X_0 kennt, so weiß man, in welcher „Periodenklasse“ (vgl. Bericht 6.69) der Pfad zu jedem Zeitpunkt ist.

Satz 7.14. *Sei $(X_n)_{n\in\mathbb{N}_0}$ ein stationärer Prozess mit Werten in $\mathbb{R}^d$. Sei $S_0 := 0$ und $S_n := \sum_{k=1}^n X_n$ für $n \in \mathbb{N}$. Ist $R_n := |\{S_1,\dots,S_n\}|$ (die „Größe des Range“) und $A = \{S_n \neq 0 \;\forall\, n \in \mathbb{N}\}$, dann gilt*

$$\frac{1}{n}R_n \xrightarrow[n\to\infty]{} \mathbb{P}(A \mid \mathcal{I}) \quad \textit{f.s.}$$

Beweis. Wir realisieren $(X_n)_{n\in\mathbb{N}_0}$ als kanonischen Prozess auf dem Produktraum, also $X_n = X_0 \circ \tau^n$. Es gilt

$$R_n = \Big|\Big\{1 \le k \le n \mid S_l \neq S_k \;\forall\, l \in \{k+1,\dots,n\}\Big\}\Big| \ge |\{1 \le k \le n \mid S_l \neq S_k \;\forall\, l > k\}| = \sum_{k=1}^n \mathbf{1}_A \circ \tau^k.$$

Nach Satz 7.7 folgt demnach

$$\liminf_{n\to\infty} \frac{1}{n}R_n \ge \liminf_{n\to\infty} \sum_{k=1}^n \mathbf{1}_A \circ \tau^k = \mathbb{P}(A \mid \mathcal{I}).$$

Sei $A_m := \{S_l \neq 0 \;\forall\, l = 1, \ldots, m\}$ für $m < n$, dann gilt

$$R_n \leq m + \left|\left\{k \leq n - m \mid S_l \neq S_k \;\forall\, l \in \{k+1, \ldots, k+m\}\right\}\right| = \sum_{k=1}^{n-m} \mathbf{1}_{A_m} \circ \tau^k,$$

somit folgt wieder nach Satz 7.7

$$\limsup_{n\to\infty} \frac{1}{n} R_n \leq 0 + \mathbb{P}(A_m \mid \mathcal{I}).$$

Es gilt $A_m \searrow A$, das heißt $\mathbb{P}(A \mid \mathcal{I}) = \lim_{m\to\infty} \mathbb{P}(A_m \mid \mathcal{I})$, also folgt mit $m \to \infty$ die Behauptung. $\square$

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

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

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 $r < r'$, $(r, r' \in \mathbb{Q})$, $\lim_{n\to\infty} F_{r+\frac{1}{n}} = F_r$ für $r \in \mathbb{Q}$, $\lim_{n\to\infty} F_n = 1$, $\lim_{n\to-\infty} F_n = 0$.

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

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

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}$$

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

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

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

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

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

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

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

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

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

Bemerkung B.6. Induktiv folgt für ein Martingal X

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

(bzw. „$\geq$“ für ein Sub- und „$\leq$“ 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 $|C_n| \leq c_n$ 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

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

□

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

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

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

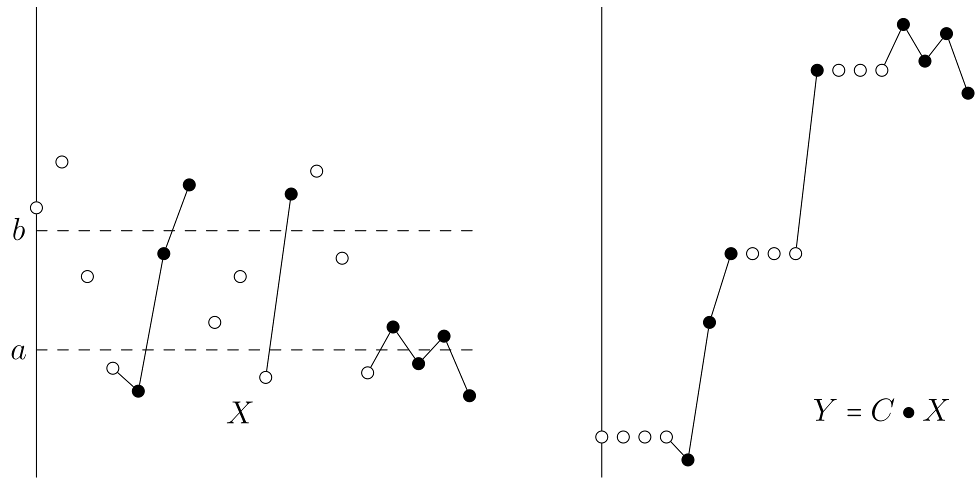


Abbildung B.1: Aufkreuzungen

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)

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

$(C_n)_n$ 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

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

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

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

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

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

□

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

$$\mathbb{E}[X_\infty^+] = \mathbb{E}[\liminf_{n\to\infty} X_n^+] \le \liminf_{n\to\infty} \mathbb{E}[X_n^+] = \liminf_{n\to\infty}(\mathbb{E}[X_n^-] + \mathbb{E}[X_n]) \le \mathbb{E}[X_0] + \sup_n \mathbb{E}[X_n^-] < \infty.$$

□

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 $\mathbb{E}[X_\infty] \ne \lim_{n\to\infty} \mathbb{E}[X_n]$ 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*

$$\mathbb{E}[X_\infty \mid \mathscr{F}_n] \leq X_n \textit{ f.s. für alle } n \in \mathbb{N}.$$

(Analog gilt $\mathbb{E}[X_\infty \mid \mathscr{F}_n] \geq X_n$, falls $(X_n)_n$ ein gleichgradig integrierbares Submartingal, und $\mathbb{E}[X_\infty \mid \mathscr{F}_n] = X_n$, 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 $n \geq m$:

$$\begin{aligned}\mathbb{E}\big[|\mathbb{E}[X_\infty \mid \mathscr{F}_m] - \mathbb{E}[X_n \mid \mathscr{F}_m]|\big] = \mathbb{E}\big[|\mathbb{E}[X_\infty - X_n \mid \mathscr{F}_m]|\big] &\leq \mathbb{E}\big[\mathbb{E}[|X_\infty - X_n| \mid \mathscr{F}_m]\big] \\ &= \mathbb{E}[|X_\infty - X_n|] \xrightarrow[n\to\infty]{} 0\end{aligned}$$

und damit ist

$$\mathbb{E}\big[(\mathbb{E}[X_\infty \mid \mathscr{F}_m] - X_m)^+\big] \leq \underbrace{\mathbb{E}\big[(\mathbb{E}[X_\infty \mid \mathscr{F}_m] - \mathbb{E}[X_n \mid \mathscr{F}_m])^+\big]}_{\to 0} + \mathbb{E}\big[(\underbrace{\mathbb{E}[X_n \mid \mathscr{F}_m]}_{\leq X_m \text{ für } n>m} - X_m)^+\big].$$

Also gilt $\mathbb{E}[X_\infty \mid \mathscr{F}_m] \leq X_m$ f.s. □

Satz B.25. *Sei $Y \in \mathcal{L}^1(\mathbb{P})$ und $(\mathscr{F}_n)_n$ eine Filtration. Sei $\mathscr{F}_\infty := \sigma(\mathscr{F}_n, n \in \mathbb{N})$ und $X_n := \mathbb{E}[Y \mid \mathscr{F}_n]$. Dann ist $(X_n)_n$ ein gleichgradig integrierbares Martingal und es gilt*

$$X_n \xrightarrow[n\to\infty]{} X_\infty := \mathbb{E}[Y \mid \mathscr{F}_\infty] \quad \textit{f.s. und in } \mathcal{L}^1(\mathbb{P}).$$

Lemma B.26. *$Y \in \mathcal{L}^1(\mathbb{P})$ und $\mathscr{F}_n \subset \mathscr{F}$ σ-Algebren, dann ist $(\mathbb{E}[Y \mid \mathscr{F}_n])_{n\in\mathbb{N}}$ 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)

$$\sup_n \mathbb{E}\big[h(|\mathbb{E}[Y \mid \mathscr{F}_n]|)\big] \leq \sup_n \mathbb{E}\big[\mathbb{E}[h(|Y|) \mid \mathscr{F}_n]\big] = \mathbb{E}[h(|Y|)] < \infty.$$

□

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 $\mathscr{F}_\infty$-messbar.

Es bleibt $X_\infty = \mathbb{E}[Y \mid \mathscr{F}_\infty]$ 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

$$\mu_1(A) = \mathbb{E}[X_\infty \mathbf{1}_A] = \lim_{n\to\infty} \mathbb{E}[X_n \mathbf{1}_A] \overset{n\geq m}{=} \mathbb{E}[Y\mathbf{1}_A] = \mu_2(A).$$

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]). $\square$

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

$$\mathbb{E}[X_m \mathbf{1}_A] = \sum_{n=0}^m \mathbb{E}\big[X_m \mathbf{1}_{\underbrace{A\cap\{T=n\}}_{\in\mathscr{F}_n}}\big] \leq \sum_{n=0}^m \mathbb{E}\big[X_n \mathbf{1}_{A\cap\{T=n\}}\big] = \mathbb{E}\left[\left(\sum_{n=0}^m X_n \mathbf{1}_{\{T=n\}}\right)\mathbf{1}_A\right] = \mathbb{E}[X_T \mathbf{1}_A].$$

$\square$

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

$$\begin{aligned}\mathbb{E}\big[h(|X_T|)\mathbf{1}_{\{T\leq n\}}\big] &= \mathbb{E}\big[h(|X_{T\wedge n}|)\mathbf{1}_{\{T\leq n\}}\big] \overset{B.28}{=} \mathbb{E}\big[h(|\mathbb{E}[X_n \mid \mathscr{F}_{T\wedge n}]|)\mathbf{1}_{\{T\leq n\}}\big] \\ &\leq \mathbb{E}\big[h(\mathbb{E}[|X_n| \mid \mathscr{F}_{T\wedge n}])\mathbf{1}_{\{T\leq n\}}\big] = \mathbb{E}\big[\mathbb{E}[h(|X_n|)\mathbf{1}_{\{T\leq n\}} \mid \mathscr{F}_{T\wedge n}]\big] \\ &\leq M.\end{aligned}$$

Mit $n \to \infty$ folgt $\mathbb{E}\big[h(|X_T|)\mathbf{1}_{\{T<\infty\}}\big] \leq M$, das heißt

$$\sup_{T \text{ Stoppzeit}} \mathbb{E}[h(|X_T|)] \leq 2M < \infty.$$

$\square$

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

$$\begin{aligned}\mathbb{E}\big[X_\infty \mathbf{1}_{A\cap\{T\le m\}}\big] &= \mathbb{E}\big[\mathbb{E}[X_\infty \mathbf{1}_{A\cap\{T\le m\}} \mid \mathscr{F}_{T\wedge m}]\big] \\ &= \mathbb{E}\big[X_{T\wedge m} \mathbf{1}_{A\cap\{T\le m\}}\big] = \mathbb{E}\big[X_T \mathbf{1}_{A\cap\{T\le m\}}\big].\end{aligned} \tag{B.2}$$

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. nicht-fallend 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

$$\sup_n \mathbb{E}[X_n^2] < \infty \qquad \Longleftrightarrow \qquad \sup_n \mathbb{E}[\langle X\rangle_n] < \infty.$$

Satz B.36. *Sei $(M_n)_n$ ein $\mathcal{L}^2$-Martingal. Dann gilt:*

i) $\{\langle M\rangle_\infty < \infty\} \overset{f.s.}{\subset} \left\{\lim\limits_{n\to\infty} M_n \text{ existiert}\right\}$.

ii) Wenn $|M_n - M_{n-1}| \le c$ für alle n für ein $c < \infty$, so gilt auch

$$\left\{\lim_{n\to\infty} M_n \text{ existiert}\right\} \overset{f.s.}{\subset} \{\langle M\rangle_\infty < \infty\}.$$

iii) $\{\langle M\rangle_\infty = \infty\} \overset{f.s.}{\subset} \left\{\dfrac{M_n}{\langle M\rangle_n} \xrightarrow[n\to\infty]{} 0\right\}$.

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 $\operatorname{Var}[Y_1] < \infty$.

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 $0 \le b_n \nearrow \infty$, 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 $\varepsilon > 0$*

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{n}(\log n)^{1/2+\varepsilon}} = 0 \quad \textit{f.s.}$$

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 $S = \dot\bigcup_{k=1}^{m_n} C_{n,k}$, $C_{n,k} \neq \emptyset$), 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 $n \in \mathbb{N}$

$$P(X_n > M) \leq \frac{1}{M}\mathbb{E}_P[X_n] = \delta,$$

mit $(*)$ also

$$\mathbb{E}_P\big[X_n \mathbf{1}_{\{X_n \geq M\}}\big] = Q(X_n > M) \leq \varepsilon.$$

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`

[KW] G. Kersting, A. Wakolbinger, Stochastische Prozesse, Birkhäuser, 2014.

[EK] S.N. Ethier, T.G. Kurtz, Markov processes: characterization and convergence, Wiley, 1986.

[L85] T.G. Liggett, Interacting particle systems, Springer, 1985.

[L10] T.G. Liggett, Continuous time Markov processes: an introduction, AMS, 2010.