Stochastik II

Wintersemester 14/15

Institut für Mathematik, FB 08
Johannes Gutenberg-Universität Mainz

gehalten von Matthias Birkner

Vorlesungsskript mitgeschrieben und in LaTeX gesetzt von Matthias Muth

Version vom 11. Juli 2015

Hinweise auf Fehler, Korrektur- und Verbesserungsvorschläge gerne per Email an

birkner@mathematik.uni-mainz.de

Inhaltsverzeichnis

1 Martingale

Hinweis: 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`.

Beispiel 1.1. Betrachte einen fairen Münzwurf, d.h. seien $W_1, W_2, \dots$ unabhängig und identisch uniform verteilt auf $\{K, Z\}$. Sei

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

$R + 3$ ist eine sog. Stoppzeit. Aber was ist $\mathbf{E}[R]$?
Betrachte ein faires Casino: Setze vor dem i-ten Wurf x Euro, erhalte $2x$ Euro oder 0 Euro je nach Ausgang. 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. Sei weiter $X_n := \sum_{i=1}^{\infty} X_{i,n}$ der Gesamtgewinn aller Spieler nach Runde n. Aufgrund der "Fairness" gilt

$$0 = \mathbf{E}[X_0] = \mathbf{E}[X_n] = \mathbf{E}[X_{R+3}]. \tag{1.1}$$

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

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

(1.1) liefert

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

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

1.1 Grundlegendes

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

Definition 1.2. *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. $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0,1,\dots}, \mathbf{P})$ *heißt filtrierter Wahrscheinlichkeitsraum.*

Bemerkung 1.3. 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 1.4. *Es sei* $X = (Xn)_n$ *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 1.5. *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* $\mathbf{P}$*), wenn gilt:*

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

ii) $X_n \in \mathcal{L}^1(\mathbf{P})$.

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

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

Bemerkung 1.6. Induktiv folgt für ein Martingal X

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

Beispiel 1.7. i) Seien $Y_1, Y_2, \dots$ unabhängige, reelle Zufallsvariablen mit $Y_n \in \mathcal{L}^1(\mathbf{P})$ und $\mathbf{E}[Y_n] = 0$ für alle $n \in \mathbb{N}$. Sei $S_0 := 0$ und $S_n := Y_1 + Y_2 + \dots + Y_n = S_{n-1} + Y_n$ für $n \in \mathbb{N}$. Dann ist $(S_n)_n$ ein Martingal bezgl. $(F_n)_n$ mit $\mathcal{F}_n = \sigma(S_1, \dots S_n) = \sigma(Y_1, \dots Y_n)$, denn $S_n \in \mathcal{L}^1(\mathbf{P})$ als Summe von $\mathcal{L}^1(\mathbf{P})$-Variablen und es gilt

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

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

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

iii) 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. $\mathcal{F}_n = \sigma(A_0, \dots, A_n)$, denn auf $\{X_n = k\}$ gilt

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

Definition 1.8. *Sei $(\mathcal{F}_n)_n$ eine Filtration. Ein stochastischer Prozess $(C_n)_n$ heißt* previsibel *(bzgl. $(\mathcal{F}_n)_n$), auch vorhersagbar, wenn C_n $\mathcal{F}_{n-1}$-messbar ist für jedes $n \in \mathbb{N}$ (C_0 spielt hier keine Rolle).*

Definition 1.9. *Sei $(X_n)_n$ adaptiert und $(C_n)_n$ previsibel bzgl. $(\mathcal{F}_n)_n$. Sei*

$$(C \bullet X)_0 := 0, \quad (C \bullet X)_n := \sum_{m=1}^{n} C_m (X_m - X_{m-1}), \quad n \in \mathbb{N}. \tag{1.2}$$

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

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

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

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

Dann ist $C \bullet X$ ein Martingal. Ist $C_n \ge 0$ 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 $C_m(X_m - X_{m-1}) \in \mathcal{L}^1(\mathbf{P})$, denn für iii) gilt mit der Cauchy-Schwarz-Ungleichung

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

Also gilt

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

□

Definition 1.11. *Sei $(\mathcal{F}_n)_n$ eine Filtration. Eine Zufallsvariable T mit Werten in $\mathbb{N}_0 \cup \{\infty\}$ heißt eine ($(\mathcal{F}_n)_n$-)* Stoppzeit, *wenn $\{T \leq n\} \in \mathcal{F}_n$ für alle $n \in \mathbb{N}_0$ gilt. Für eine Stoppzeit T ist*

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

eine σ-Algebra, sie heißt die (σ-Algebra der) T-Vergangenheit.
Interpretation: $\mathcal{F}_T$ enthält diejenigen Ereignisse, die sich zu dem (zufälligen) Zeitpunkt T entscheiden lassen.

Bemerkung 1.12. T ist genau dann eine Stoppzeit, wenn $\{T = n\} \in \mathcal{F}_n$ für alle $n \in \mathbb{N}_0$, denn $\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^\mathsf{c}$.

Beispiel 1.13. i) Jede Konstante t_0 ist eine Stoppzeit.

ii) Es sei $(X_n)_n$ ein adaptierter stochastischer Prozess mit Werten in $(E, \mathcal{B})$ und $K \in \mathcal{B}$. Dann ist
$$T := \inf \{n \in \mathbb{N}_0 \mid X_n \in K\}$$
eine Stoppzeit, denn $\{T \leq n\} = \bigcup_{m=0}^n \{X_m \in K\} \in \mathcal{F}_n$.

Bemerkung 1.14. $L := \sup \{n \in \mathbb{N}_0 \mid X_n \in K\}$ ist im Allgemeinen keine Stoppzeit!

Lemma 1.15. *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

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

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 $m \le n$: $\{\sigma \wedge n \le m\}, \{\tau \wedge n \le m\} \in \mathcal{F}_m \subset \mathcal{F}_n$. Dann sind

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

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

$$\{\sigma + \tau \le n\} = \left\{\sigma' + \tau' \le n\right\} \in \mathcal{F}_n,$$

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

Bemerkung 1.16. $\sigma - \tau$ ist im Allgemeinen keine Stoppzeit!

Lemma 1.17. *Sind σ, τ Stoppzeiten mit $\sigma \le \tau$, dann gilt $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.*

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

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

□

Beobachtung 1.18. Sei $(F_n)_n$ eine Filtration, T eine Stoppzeit mit $T < \infty$ f.s. und $(X_n)_n$ ein adaptierter stochastischer Prozess mit Werten in $(E, \mathcal{B})$. Dann ist $X_T = X_{T(\omega)}(\omega)$ eine $\mathcal{F}_T$-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 $B \in \mathcal{B}$. Dann ist

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

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

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

also ist X_T $\mathcal{F}_T$-messbar. Weiter ist $X_n^{(T)} = X_{T \wedge n} \in \mathcal{F}_{T \wedge n} \subset \mathcal{F}_n$, das heißt $(X_n^{(T)})_n$ ist adaptiert. □

Bemerkung 1.19. $(X_n^{(T)})_n$ ist auch adaptiert an $\mathcal{F}^{(T)} := (\mathcal{F}_{T \wedge n})_n$.

Lemma 1.20. *Sei T eine Stoppzeit. Ist $(X_n)_n$ ein (Sub- / Super-) Martingal, so auch $(X_n^{(T)})_n$.*

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

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

Damit folgt die Behauptung aus Lemma 1.10. □

Korollar 1.21. *Sei X ein Supermartingal und T eine Stoppzeit. Es gelte mindestens eine der folgenden beiden Bedingungen*

i) *T ist beschränkt.*

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

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

Beweis. Angenommen i) gilt, dann existiert ein $m \in \mathbb{N}$ mit $T \leq m$. Nach Lemma 1.20 ist $(X_{T \wedge n})_n$ ein Supermartingal, also gilt

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

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

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

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

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

□

1.2 Martingalkonvergenzsatz

Sei $(X_n)_n$ ein adaptierter, reellwertiger Prozess und $-\infty < a < b < \infty$.
Setze $C_1 := \mathbb{1}_{\{X_0 < a\}}$ und für $n > 1$ rekursiv (siehe auch Abbildung 1.1)

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

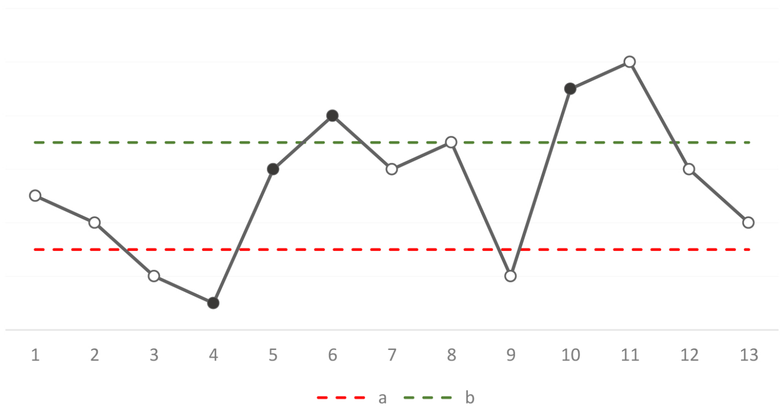


Abbildung 1.1: Aufkreuzungen

$(C_n)_n$ ist previsibel. Sei weiter

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

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

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

Lemma 1.22 (Doobs[1] Aufkreuzungslemma)**.** *Sei X ein Supermartingal. Dann gilt für alle $n \in \mathbb{N}$*

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

Beweis. Nach Lemma 1.10 ist $Y = C \bullet X$ ein Supermartingal. Also gilt

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

□

Satz 1.23 (Doobs (Super-) Martingalkonvergenzsatz)**.** *Ist $(X_n)_n$ ein Supermartingal mit $\sup_n \mathbf{E}[X_n^-] < \infty$, dann gibt es ein $X_\infty \in \mathcal{L}^1(\mathbf{P})$ mit $X_n \to X_\infty$ f.s.*

[1] Joseph L. Doob (1910-2004)

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

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

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

$$\mathbf{P}(\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n) = \mathbf{P}\left(\bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} O_{a,b}\right) = 0.$$

Mit $X_\infty := \limsup_n X_n$ gilt also $X_n \to X_\infty$ f.s. Es bleibt $X_\infty \in \mathcal{L}^1(\mathbf{P})$ zu zeigen. Es gilt:

$$\mathbf{E}[X_\infty^-] = \mathbf{E}[\liminf_{n\to\infty} X_n^-] \overset{\text{Fatou}}{\le} \liminf_{n\to\infty} \mathbf{E}[X_n^-] < \infty$$

nach Voraussetzung und

$$\mathbf{E}[X_\infty^+] = \mathbf{E}[\liminf_{n\to\infty} X_n^+] \le \liminf_{n\to\infty} \mathbf{E}[X_n^+] = \liminf_{n\to\infty}(\mathbf{E}[X_n^-] + \mathbf{E}[X_n]) \le \mathbf{E}[X_0] + \sup_n \mathbf{E}[X_n^-] < \infty.$$

□

Bemerkung 1.24. Die analoge Aussage von Satz 1.23 gilt für ein Submartingal $(X_n)_n$ mit $\sup_n \mathbf{E}[X_n^+] < \infty$.

Bemerkung 1.25. Falls $X_n \ge c > -\infty$ für ein festes c, so gilt $\mathbf{E}[X_\infty] \le \lim_n \mathbf{E}[X_n]$. Im Allgemeinen gilt $\mathbf{E}[X_\infty] \ne \lim_n \mathbf{E}[X_n]$, betrachte zum Beispiel die symmetrische gewöhnliche Irrfahrt startend in 1, gestoppt bei Erreichen der 0.

1.3 Gleichgradig integrierbare Martingale und optionales Stoppen

Erinnerung 1.26. Eine Familie reeller Zufallsvariablen $(X_n)_n$ heißt gleichgradig integrierbar, falls

$$\lim_{k\to\infty} \sup_n \mathbf{E}\big[|X_n| \cdot \mathbb{1}_{\{|X_n|\ge k\}}\big] = 0.$$

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 $\sup_n \mathbf{E}[h(|X_n|)] < \infty$. (Man kann annehmen, dass h monoton wachsend und konvex ist, vgl. [Kle13, Satz 6.19])

ii) Sei $X_n \xrightarrow{\mathbf{P}} X_\infty$. Dann gilt $X_n \xrightarrow{\mathcal{L}^1(\mathbf{P})} X_\infty$ genau dann, wenn $(X_n)_n$ gleichgradig integrierbar ist.

Satz 1.27. *Sei $(X_n)_n$ ein gleichgradig integrierbares Supermartingal. Dann existiert $X_\infty \in \mathcal{L}^1(\mathbf{P})$ mit $X_n \to X_\infty$ f.s. und in $\mathcal{L}^1(\mathbf{P})$. Es gilt*

$$\mathbf{E}[X_\infty \mid \mathcal{F}_n] \leq X_n \text{ f.s. für alle } n \in \mathbb{N}.$$

(Analog gilt $\mathbf{E}[X_\infty \mid \mathcal{F}_n] \geq X_n$, falls $(X_n)_n$ ein gleichgradig integrierbares Submartingal, und $\mathbf{E}[X_\infty \mid \mathcal{F}_n] = X_n$, falls $(X_n)_n$ ein gleichgradig integrierbares Martingal.)

Beweis. Die Existenz von X_∞ mit $X_n \to X_\infty$ f.s. folgt aus Satz 1.23. Aufgrund der gleichgradigen Integrierbarkeit gilt $\mathbf{E}[|X_n - X_\infty|] \xrightarrow[n\to\infty]{} 0$. Weiter gilt für $n \geq m$:

$$\begin{aligned}\mathbf{E}\big[|\mathbf{E}[X_\infty \mid \mathcal{F}_m] - \mathbf{E}[X_n \mid \mathcal{F}_m]|\big] = \mathbf{E}\big[|\mathbf{E}[X_\infty - X_n \mid \mathcal{F}_m]|\big] &\leq \mathbf{E}\big[\mathbf{E}[|X_\infty - X_n| \mid \mathcal{F}_m]\big] \\ &= \mathbf{E}[|X_\infty - X_n|] \xrightarrow[n\to\infty]{} 0\end{aligned}$$

und damit ist

$$\mathbf{E}\big[(\mathbf{E}[X_\infty \mid \mathcal{F}_m] - X_m)^+\big] \leq \underbrace{\mathbf{E}\big[(\mathbf{E}[X_\infty \mid \mathcal{F}_m] - \mathbf{E}[X_n \mid \mathcal{F}_m])^+\big]}_{\to 0} + \mathbf{E}\big[(\underbrace{\mathbf{E}[X_n \mid \mathcal{F}_m]}_{\leq X_m \text{ für } n>m} - X_m)^+\big].$$

Also gilt $\mathbf{E}[X_\infty \mid \mathcal{F}_m] \leq X_m$ f.s. □

Satz 1.28. *Sei $Y \in \mathcal{L}^1(\mathbf{P})$ und $(\mathcal{F}_n)_n$ eine Filtration. Sei $\mathcal{F}_\infty := \sigma(\mathcal{F}_n \mid n \in \mathbb{N})$ und $X_n := \mathbf{E}[Y \mid \mathcal{F}_n]$. Dann ist $(X_n)_n$ ein gleichgradig integrierbares Martingal und es gilt*

$$X_n \xrightarrow[n\to\infty]{} X_\infty := \mathbf{E}[Y \mid \mathcal{F}_\infty] \text{ f.s. und in } \mathcal{L}^1(\mathbf{P}).$$

Lemma 1.29. *Ist $Y \in \mathcal{L}^1(\mathbf{P})$ und $\mathcal{F}_n \subset \mathcal{F}$ σ-Algebren. Dann ist $(\mathbf{E}[Y \mid \mathcal{F}_n])_{n\in\mathbb{N}}$ gleichgradig integrierbar.*

Beweis. Es existiert ein $h\colon [0,\infty) \to [0,\infty)$ konvex und monoton wachsend mit $\frac{h(x)}{x} \xrightarrow[x\to\infty]{} \infty$ und $\mathbf{E}[h(|Y|)] < \infty$. Es gilt mit der Jensen-Ungleichung

$$\sup_n \mathbf{E}\big[h(|\mathbf{E}[Y \mid \mathcal{F}_n]|)\big] \leq \sup_n \mathbf{E}\big[\mathbf{E}[h(|Y|) \mid \mathcal{F}_n]\big] = \mathbf{E}[h(|Y|)] < \infty.$$

□

Beweis von Satz 1.28. $(X_n)_n$ ist ein Martingal und nach Lemma 1.29 gleichgradig integrierbar. Nach Satz 1.27 gilt $X_n \to X_\infty := \limsup_{n\to\infty} X_n$ f.s. und in $\mathcal{L}^1(\mathbf{P})$. X_∞ ist $\mathcal{F}_\infty$-messbar. Es bleibt $X_\infty = \mathbf{E}[Y \mid \mathcal{F}_\infty]$ zu zeigen. Ohne Einschränkung können wir $Y \geq 0$ annehmen (ansonsten betrachte Y^+, Y^- separat). Insbesondere ist dann $X_\infty \geq 0$ f.s. Für $A \in \mathcal{F}_\infty$ sind $\mu_1(A) := \mathbf{E}[X_\infty \mathbb{1}_A]$ und $\mu_2(A) := \mathbf{E}[Y\mathbb{1}_A]$ endliche Maße auf $(\Omega, \mathcal{F}_\infty, \mathbf{P})$. Sei $A \in \mathcal{F}_m \subset \mathcal{F}_\infty$. Dann gilt

$$\mu_1(A) = \mathbf{E}[X_\infty \mathbb{1}_A] = \lim_{n\to\infty} \mathbf{E}[X_n \mathbb{1}_A] \overset{n\geq m}{=} \mathbf{E}[Y\mathbb{1}_A] = \mu_2(A).$$

Da $\bigcup_{m\in\mathbb{N}} F_m$ ein schnittstabiler Erzeuger von $\mathcal{F}_\infty$ ist, folgt $\mu_1 = \mu_2$ mit [Dep14, Satz 1.37]. □

Bemerkung 1.30. Für ein gleichgradig integrierbares Martingal $(X_n)_n$ gilt

$$\mathbf{E}[X_\infty \mid \mathcal{F}_n] = X_n \text{ für alle } n \in \mathbb{N}.$$

Solche Martingale heißen *Doobsche Martingale.*

Lemma 1.31. *Sei $(X_n)_n$ ein Supermartingal und T eine Stoppzeit mit $T \leq m$ f.s. für ein $m \in \mathbb{N}$. Dann ist $X_T \in \mathcal{L}^1(\mathbf{P})$ und es gilt $\mathbf{E}[X_m \mid \mathcal{F}_T] \leq X_T$ f.s. Im Falle eines Martingals gilt Gleichheit.*

Beweis. Nach Lemma 1.20 ist $(X_{T\wedge n})_n$ ein Supermartingal. Es gilt $X_T \in \mathcal{L}^1(\mathbf{P})$, denn $X_T = X_{T\wedge m}$ f.s. Für $A \in \mathcal{F}_T$ gilt

$$\mathbf{E}[X_m \mathbb{1}_A] = \sum_{n=0}^{m} \mathbf{E}\big[X_m \mathbb{1}_{\underbrace{A\cap\{T=n\}}_{\in\mathcal{F}_n}}\big] \leq \sum_{n=0}^{m} \mathbf{E}\big[X_n \mathbb{1}_{A\cap\{T=n\}}\big] = \mathbf{E}\left[\left(\sum_{n=0}^{m} X_n \mathbb{1}_{\{T=n\}}\right)\mathbb{1}_A\right] = \mathbf{E}[X_T \mathbb{1}_A].$$

□

Lemma 1.32. *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 ein $h\colon [0,\infty) \to [0,\infty)$ konvex und monoton wachsend mit $\frac{h(x)}{x} \xrightarrow[x\to\infty]{} \infty$ und $\mathbf{E}[h(|X_n|)] =: M < \infty$. Sei T eine Stoppzeit und $n \in \mathbb{N}$. Es gilt

$$\begin{aligned}
\mathbf{E}\big[h(|X_T|)\mathbb{1}_{\{T\leq n\}}\big] &= \mathbf{E}\big[h(|X_{T\wedge n}|)\mathbb{1}_{\{T\leq n\}}\big] \overset{1.31}{=} \mathbf{E}\big[h(|\mathbf{E}[X_n \mid \mathcal{F}_{T\wedge n}]|)\mathbb{1}_{\{T\leq n\}}\big] \\
&\leq \mathbf{E}\big[h(\mathbf{E}[|X_n| \mid \mathcal{F}_{T\wedge n}])\mathbb{1}_{\{T\leq n\}}\big] = \mathbf{E}\big[\mathbf{E}[h(|X_n|)\mathbb{1}_{\{T\leq n\}} \mid \mathcal{F}_{T\wedge n}]\big] \\
&\leq M.
\end{aligned}$$

Mit $n \to \infty$ folgt $\mathbf{E}\big[h(|X_T|)\mathbb{1}_{\{T<\infty\}}\big] \le M$, das heißt

$$\sup_{T \text{ Stoppzeit}} \mathbf{E}[h(|X_T|)] \le 2M < \infty.$$

□

Satz 1.33 (optional-sampling-Theorem)**.** *Es sei* $(X_n)_n$ *ein gleichgradig integrierbares Martingal,* $X_\infty := \lim X_n$ *und* T *eine Stoppzeit. Dann ist* $X_T \in \mathcal{L}^1(\mathbf{P})$ *und es gilt* $\mathbf{E}[X_\infty \mid \mathcal{F}_T] = X_T$ *f.s. Insbesondere gilt* $\mathbf{E}[X_T] = \mathbf{E}[X_\infty] = \mathbf{E}[X_0]$*. Ist* S *eine Stoppzeit mit* $S \le T$*, so gilt* $\mathbf{E}[X_T \mid \mathcal{F}_S] = X_S$ *f.s.*

Beweis. Nach Satz 1.28 gilt $\mathbf{E}[X_\infty \mid \mathcal{F}_m] = X_m$ f.s. und nach Lemma 1.31 ist $\mathbf{E}[X_m \mid \mathcal{F}_{T\wedge m}] = X_{T\wedge m}$ f.s. Also ist $\mathbf{E}[X_\infty \mid \mathcal{F}_{T\wedge m}] = \mathbf{E}[\mathbf{E}[X_\infty \mid \mathcal{F}_m] \mid \mathcal{F}_{T\wedge m}] = X_{T\wedge m}$ f.s.
Sei nun $A \in \mathcal{F}_T$. Dann ist $A \cap \{T \le m\} \in \mathcal{F}_{T\wedge m}$, denn für $n \in N_0$ gilt $A \cap \{T \le m\} \cap \{T \le n\} = A \cap \{T \le m \wedge n\} \in \mathcal{F}_{m\wedge n} \subset \mathcal{F}_n$. Somit gilt

$$\mathbf{E}\big[X_\infty \mathbb{1}_{A\cap\{T\le m\}}\big] = \mathbf{E}[\mathbf{E}[X_\infty \mathbb{1}_{A\cap\{T\le m\}} \mid \mathcal{F}_{T\wedge m}]\big] = \mathbf{E}\big[X_{T\wedge m}\mathbb{1}_{A\cap\{T\le m\}}\big] = \mathbf{E}\big[X_T \mathbb{1}_{A\cap\{T\le m\}}\big]. \tag{1.3}$$

Sei ohne Einschränkung $X_\infty \ge 0$ (sonst betrachte X_∞^+, X_∞^- separat). $m \to \infty$ in (1.3) mit monotoner Konvergenz liefert $\mathbf{E}\big[X_\infty \mathbb{1}_{A\cap\{T<\infty\}}\big] = \mathbf{E}\big[X_T \mathbb{1}_{A\cap\{T<\infty\}}\big]$. Nach Definition gilt aber auch $\mathbf{E}\big[X_\infty \mathbb{1}_{A\cap\{T=\infty\}}\big] = \mathbf{E}\big[X_T \mathbb{1}_{A\cap\{T=\infty\}}\big]$, d.h. $\mathbf{E}[X_\infty \mathbb{1}_A] = \mathbf{E}[X_T \mathbb{1}_A]$.

Sei $S \le T$ eine Stoppzeit. Dann ist $\mathcal{F}_S \subset \mathcal{F}_T$, also gilt

$$\mathbf{E}[X_T \mid \mathcal{F}_S] = \mathbf{E}\big[\mathbf{E}[X_\infty \mid \mathcal{F}_T] \mid \mathcal{F}_S\big] = \mathbf{E}[X_\infty \mid \mathcal{F}_S] = X_S \quad \text{f.s.}$$

□

Bemerkung und Definition 1.34. Sei $(X_n)_n$ ein adaptierter Prozess mit $X_n \in \mathcal{L}^1(\mathbf{P})$ für $n \in \mathbb{N}_0$. Dann ist $X_n = M_n + A_n$ mit

$$M_0 := X_0, \quad M_n := X_0 + \sum_{k=1}^{n} \big(X_k - \mathbf{E}[X_k \mid \mathcal{F}_{k-1}]\big),$$

$$A_0 := 0, \quad A_n := \sum_{k=1}^{n} \big(\mathbf{E}[X_k \mid \mathcal{F}_{k-1}] - X_{k-1}\big).$$

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 $\tilde{M}_n \equiv \tilde{M}_0 \equiv 0$, denn $\tilde{M}_{n-1} = \mathbf{E}[\tilde{M}_n \mid \mathcal{F}_{n-1}] = \tilde{M}_n$ f.s. □

Satz 1.35. *Sei $(X_n)_n$ ein gleichgradig integrierbares Supermartingal und seien S, T Stoppzeiten mit $S \leq T$. Dann gilt $\mathbf{E}[X_T \mid \mathcal{F}_S] \leq X_S$ f.s.*

Beweis. Sei $X_n = M_n + A_n$ die Doob-Zerlegung. Dann gilt $A_n \searrow A_\infty \leq 0$. Es ist

$$\mathbf{E}[|A_n|] = \mathbf{E}[-A_n] = \mathbf{E}[M_n - X_n] = \mathbf{E}[M_n - M_0 + X_0 - X_n] \leq \mathbf{E}[|X_n| + \mathbf{E}[|X_0|]] \leq M$$

für alle n für ein geeignetes M. Somit ist $(A_n)_n$ gleichgradig integrierbar und damit auch $(M_n)_n = (X_n - A_n)_n$. Also gilt

$$\mathbf{E}[X_T \mid \mathcal{F}_S] = \underbrace{\mathbf{E}[M_T \mid \mathcal{F}_S]}_{=M_S \text{ f.s.}} + \mathbf{E}[A_T \mid \mathcal{F}_S] \leq M_S + \mathbf{E}[A_S \mid \mathcal{F}_S] = M_S + A_S = X_S.$$

□

1.4 $\mathcal{L}^2$-Martingale

Bemerkung 1.36. Sei $(X_n)_n$ ein Martingal und $\varphi\colon \mathbb{R} \to \mathbb{R}$ konvex, sodass $\mathbf{E}[\varphi(X_n)]$ für alle n existiert. Dann ist $(\varphi(X_n))_n$ ein Submartingal, denn

$$\mathbf{E}[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geq \varphi\left(\mathbf{E}[X_{n+1} \mid \mathcal{F}_n]\right) = \varphi(X_n).$$

Die Aussage gilt ebenso, wenn $(X_n)_n$ ein Submartingal und φ konvex und nicht fallend ist.

Bemerkung 1.37. Eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ heißt *harmonisch*, wenn

$$f(x) = \frac{1}{2a} \int_{x-a}^{x+a} f(y)\,\mathrm{d}y \quad \forall x \in \mathbb{R}, a > 0.$$

$f\colon \mathbb{R} \to \mathbb{R}$ heißt *subharmonisch*, wenn $f(x) \leq \frac{1}{2a} \int_{x-a}^{x+a} f(y)\,\mathrm{d}y$ und *superharmonisch*, wenn $f(x) \geq \frac{1}{2a} \int_{x-a}^{x+a} f(y)\,\mathrm{d}y$ für alle $x \in \mathbb{R}, a > 0$. Konvexe Funktionen sind subharmonisch, zusammen mit Bemerkung 1.36 motiviert dies den Namen Submartingal (und entsprechend auch den Namen Supermartingal).

Beobachtung 1.38. Sei $(X_n)_n$ ein quadratintegrables Martingal, d.h. $X_n \in \mathcal{L}^2(\mathbf{P})$ für alle

$n \in \mathbb{N}$. Dann gilt $\mathbf{E}[(X_k - X_l)(X_l - X_m)] = 0$ für alle $0 \leq m \leq l \leq k$, denn

$$\begin{aligned}\mathbf{E}[(X_k - X_l)(X_l - X_m)] = \mathbf{E}\big[\mathbf{E}[(X_k - X_l)(X_l - X_m) \mid \mathcal{F}_l]\big] &= \mathbf{E}\big[(X_l - X_m)(\mathbf{E}[X_k \mid \mathcal{F}_l] - X_l)\big] \\ &= \mathbf{E}[(X_l - X_m)(X_l - X_l)] = \mathbf{E}[0] = 0.\end{aligned}$$

Bemerkung 1.39. $\|X\|_2 := \sqrt{\mathbf{E}[X^2]}$ ist eine Norm und $\langle X, Y\rangle := \mathbf{E}[XY]$ ist ein Skalarpodukt, $\mathcal{L}^2(\mathbf{P})$ ist ein Hilbertraum.

Man sagt auch: Martingalinkremente über disjunkte Zeitintervalle sind orthogonal.

Lemma und Definition 1.40. *Sei $(X_n)_n$ ein quadratintegrables Martingal.*

$$A_n := \sum_{k=1}^{n} \mathbf{E}[(X_k - X_{k-1})^2 \mid \mathcal{F}_{k-1}]$$

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:

$$\begin{aligned}\mathbf{E}[X_{n+1}^2 - A_{n+1} \mid \mathcal{F}_n] &= \mathbf{E}[(X_{n+1} - X_n)^2 + 2(X_{n+1} - X_n)X_n + X_n^2 - A_{n+1} \mid \mathcal{F}_n] \\ &= \underbrace{\mathbf{E}[(X_{n+1} - X_n)^2 \mid \mathcal{F}_n] - A_{n+1}}_{=-A_n} + X_n^2 + 2X_n \underbrace{\mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n]}_{=0} \\ &= X_n^2 - A_n.\end{aligned}$$

Die Eindeutigkeit folgt aus der Eindeutigkeit der Doob-Zerlegung 1.34. □

Insbesondere gilt also:

$$\mathbf{E}[X_n^2] = \mathbf{E}[X_0^2] + \mathbf{E}[\langle X\rangle_n] = \mathbf{E}[X_0^2] + \sum_{k=1}^{n} \mathbf{E}[(X_k - X_{k-1})^2]$$

und

$$\sup_n \mathbf{E}[X_n^2] < \infty \quad \Leftrightarrow \quad \sup_n \mathbf{E}[\langle X\rangle_n] < \infty.$$

Satz 1.41. *Sei $(M_n)_n$ ein $\mathcal{L}^2$-Martingal. Dann gilt:*

i) $\{\langle M\rangle_\infty < \infty\} \overset{f.s.}{\subset} \left\{\lim_{n\to\infty} M_n \text{ existiert}\right\}$.

ii) Wenn $|M_n - M_{n-1}| \leq c$ für alle n für ein $c < \infty$, so gilt auch $\left\{\lim_{n\to\infty} M_n \text{ existiert}\right\} \overset{f.s.}{\subset}$

$\{\langle M\rangle_\infty < \infty\}$.

iii) $\{\langle M\rangle_\infty = \infty\} \overset{f.s.}{\subset} \left\{\frac{M_n}{\langle M\rangle_n} \xrightarrow[n\to\infty]{} 0\right\}$.

Bemerkung 1.42. iii) impliziert das Starke Gesetz der großen Zahlen für $M_n = Y_1 + \ldots + Y_n$ mit Y_i unabhängig und identisch verteilt mit $\mathbf{E}[Y_1] = 0$ und $\mathbf{Var}[Y_1] < \infty$.

Lemma 1.43 (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 1.41. Sei $k \in \mathbb{R}^+$. $S_k := \inf\{n \in \mathbb{N}_0 \mid \{M\}_{n+1} > k\}$ ist eine Stoppzeit. Nach Lemma 1.20 und Lemma 1.40 ist $(M^2_{n\wedge S_k} - \langle M\rangle_{n\wedge S_k})_n$ ein Martingal, also gilt

$$\sup_n \mathbf{E}[M^2_{n\wedge S_k}] = \mathbf{E}[M_0] + \underbrace{\sup_n \mathbf{E}[\langle M\rangle_{n\wedge S_k}]}_{\le k} \le \infty,$$

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 $\{\langle M\rangle_\infty < \infty\} = \bigcup\limits_{k\in\mathbb{N}} \{S_k = \infty\}$ und $\bigcup\limits_{k\in\mathbb{N}} \{S_k = \infty, \lim_n M_{n\wedge S_k} \text{ existiert}\} \subset \{\lim_n M_n \text{ existiert}\}$, damit gilt $i)$.

Sei $K > 0$. $T_K := \inf\{n \in \mathbb{N}_0 \mid |M_n| > K\}$ ist eine Stoppzeit und es gilt

$$\mathbf{E}[\underbrace{M^2_{n\wedge T_K}}_{\le (K+c)^2} - \langle M\rangle_{n\wedge T_K}] = \mathbf{E}[M_0^2].$$

Also folgt mit monotoner Konvergenz

$$\mathbf{E}[\langle M\rangle_{T_K}] = \sup_n \mathbf{E}[\langle M\rangle_{n\wedge T_K}] \le \mathbf{E}[M_0^2] + (K+c)^2 < \infty.$$

Demnach ist $\langle M\rangle_{T_K} < \infty$ f.s. und es gilt

$$\mathbf{P}\left(\underbrace{\left\{\lim_{n\to\infty} M_n \text{ existiert}\right\}}_{\subset \bigcup_{k\in\mathbb{N}}\{T_k=\infty\}} \cap \{\langle M\rangle_\infty = \infty\}\right)$$

$$\le \mathbf{P}\left(\bigcup_{K\in\mathbb{N}} \{T_K = \infty\} \cap \{\langle M\rangle_\infty = \infty\} \cap \{\langle M\rangle_{T_K} < \infty\}\right) = 0.$$

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 1.10 ist $(W_n)_n$ ein

Martingal und es gilt

$$\begin{aligned}\mathbf{E}[(W_n - W_{n-1})^2 \mid \mathcal{F}_{n-1}] &= \frac{1}{(1+\langle M\rangle_n)^2}\mathbf{E}[(M_n - M_{n-1})^2 \mid \mathcal{F}_{n-1}] = \frac{\langle M\rangle_n - \langle M\rangle_{n-1}}{(1+\langle M\rangle_n)^2} \\ &\leq \frac{(\langle M\rangle_n + 1) - (\langle M\rangle_{n-1}+1)}{(1+\langle M\rangle_n)(1+\langle M\rangle_{n-1})} = \frac{1}{1+\langle M\rangle_{n-1}} - \frac{1}{1+\langle M\rangle_n}.\end{aligned}$$

Also gilt

$$\langle W\rangle_\infty = \lim_{N\to\infty}\langle W\rangle_N = \lim_{N\to\infty}\left(\frac{1}{1+\langle M\rangle_{n-1}} - \frac{1}{1+\langle M\rangle_n}\right) \leq 1$$

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 1.43, 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. □

Im Folgenden sei $(X_n)_n$ ein stochastischer Prozess mit Werten in $\mathbb{R}$ und

$$X_n^* := \max_{0\leq k\leq n} X_k, \quad |X|_n^* := \max_{0\leq k\leq n}|X_k|.$$

Lemma 1.44. *Sei $(X_n)_n$ ein Submartingal und $\lambda \geq 0$. Dann gilt*

$$\lambda\mathbf{P}(X_n^* \geq \lambda) \leq \mathbf{E}\big[X_n \mathbb{1}_{\{X_n^*\geq\lambda\}}\big] \leq \mathbf{E}\big[|X_n|\mathbb{1}_{\{X_n^*\geq\lambda\}}\big] \ (\leq \mathbf{E}[|X_n|]).$$

Beweis. Sei $T := \inf\{k \in \mathbb{N}_0 \mid X_k \geq \lambda\} \wedge n$. T ist eine beschränkte Stoppzeit und es gilt mit Korollar 1.21

$$\begin{aligned}\mathbf{E}\big[X_n\mathbb{1}_{\{X_n^*\geq\lambda\}}\big] + \mathbf{E}\big[X_n\mathbb{1}_{\{X_n^*<\lambda\}}\big] = \mathbf{E}[X_n] \overset{1.21}{\geq} \mathbf{E}[X_T] &= \mathbf{E}\big[X_T\mathbb{1}_{\{X_n^*\geq\lambda\}}\big] + \mathbf{E}\big[X_T\mathbb{1}_{\{X_n^*<\lambda\}}\big] \\ &\geq \lambda\mathbf{P}(X_n^*\geq\lambda) + \mathbf{E}\big[X_n\mathbb{1}_{\{X_n^*<\lambda\}}\big].\end{aligned}$$

□

Satz 1.45 (Doobs $\mathcal{L}^p$-Ungleichungen)**.** *Sei $(X_n)_n$ ein Martingal oder ein nichtnegatives Submartingal.*

i) Für $p \geq 1$ und $\lambda > 0$ gilt $\lambda^p\mathbf{P}(|X|_n^* \geq \lambda) \leq \mathbf{E}[|X_n|^p]$.

ii) Für $p > 1$ gilt $\mathbf{E}[|X_n|^p] \leq \mathbf{E}[(|X|_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p\mathbf{E}[|X_n|^p]$.

Bemerkung 1.46. Für ein $\mathcal{L}^2$-Martingal gestattet dies, $\mathbf{E}[(|X|_n^*)^2]$ durch $\mathbf{E}[\langle X\rangle_n]$ zu kontrollieren, denn $\mathbf{E}[X_n^2] = \mathbf{E}[X_0^2] + \mathbf{E}[\langle X\rangle_n]$.

Beweis von Satz 1.45. i) $(|X_n|^p)_n$ ist ein Submartingal, also folgt die Aussage durch Anwenden von Lemma 1.44 auf $(|X_n|^p)_n$.

ii) Sei $c > 0$. Es gilt:

$$\begin{aligned}
\mathbf{E}[(|X|_n^* \wedge c)^p] &= \mathbf{E}\left[\int_0^{|X|_n^* \wedge c} p\lambda^{p-1} \mathrm{d}\lambda\right] = \mathbf{E}\left[\int_0^c p\lambda^{p-1} \mathbb{1}_{\{|X|_n^* \geq \lambda\}} \,\mathrm{d}\lambda\right] \\
&\overset{\text{Fubini}}{=} \int_0^c p\lambda^{p-1} \mathbf{P}(|X|_n^* \geq \lambda) \,\mathrm{d}\lambda \leq \int_0^c p\lambda^{p-1} \frac{1}{\lambda} \mathbf{E}[|X_n| \mathbb{1}_{\{|X|_n^* \geq \lambda\}}] \,\mathrm{d}\lambda \\
&\overset{\text{Fubini}}{=} p\mathbf{E}\left[|X_n| \int_0^{c \wedge |X|_n^*} \lambda^{p-2} \,\mathrm{d}\lambda\right] = \frac{p}{p-1} \mathbf{E}[|X_n| (|X|_n^* \wedge c)^{p-1}] \\
&\overset{\text{Hölder}}{\leq} \frac{p}{p-1} E[(|X|_n^* \wedge c)^p]^{\frac{p-1}{p}} \mathbf{E}[|X_n|^p]^{\frac{1}{p}}.
\end{aligned}$$

Für $c \to \infty$ folgt mit monotoner Konvergenz

$$\mathbf{E}[(|X|_n^*)^p] \leq \frac{p}{p-1} \mathbf{E}[(|X|_n^*)^p]^{\frac{p-1}{p}} \mathbf{E}[|X_n|^p]^{\frac{1}{p}}$$

und durch Umstellen und Potenzieren dieser Ungleichung erhält man

$$\mathbf{E}[(|X|_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_n|^p].$$

□

Bemerkung 1.47. Für $p = 2$ impliziert i) die Kolmogorovsche Ungleichung. ii) legt für $p = 2$ nahe: Ist $X_n = Y_1 + \ldots + Y_n$, wobei Y_i unabhängig und identisch verteilt mit $\mathbf{E}[Y_1] = 0$ und $\sigma^2 := \mathbf{E}[Y_1^2] \leq \infty$, so ist typischerweise $\max_{k \leq n} |X_k| = \mathcal{O}(\sqrt{n})$.

2 Austauschbarkeit

2.1 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 2.1. *$(X_i)_{i\in I}$ heißt* austauschbar*, wenn $\mathcal{L}((X_i)_{i\in I}) = \mathcal{L}((X_{\pi(i)})_{i\in I})$ für jede endliche Permutation $\pi: I \to I$ (das heißt π ist bijektiv und $|\{i \mid \pi(i) \neq i\}| < \infty$) gilt.*

Bemerkung 2.2. $(X_i)_{i\in I}$ ist genau dann austauschbar, wenn für alle $n \in \mathbb{N}$, $i_1, \ldots i_n, j_1, \ldots, j_n \in I$ gilt: $\mathcal{L}((X_{i_1}, \ldots, X_{i_n})) = \mathcal{L}((X_{j_1}, \ldots, X_{j_n}))$.

Beweis. Ist $(X_i)_{i\in I}$ austauschbar, so gilt $\mathcal{L}((X_{i_1}, \ldots, X_{i_n})) = \mathcal{L}((X_{j_1}, \ldots, X_{j_n}))$ nach Definition mit $\pi(i_k) = j_k$. Andererseits ist $\mathcal{L}((X_i)_{i\in I})$ festgelegt durch $\{\mathcal{L}((X_j)_{j\in J}) \mid J \subset I \text{ endlich}\}$, 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 2.3. 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 $X_i := \mathbb{1}_{\{i\text{-te Kugel ist schwarz}\}}$. Für $x_1, \ldots, x_N \in \{0,1\}$ mit $x_1 + \ldots + x_N = M$ gilt

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

also sind die X_i austauschbar.

iv) (Münzwurf mit zufälliger Erfolgswahrscheinlichkeit) Sei Y eine Zufallsvariable mit Werten in $[0,1]$. Gegeben $Y = y$ seien $X_1, X_2, \ldots \sim \mathbf{Ber}(y)$ unabhängig und identisch verteilt (zum Beispiel realisierbar als $X_i = \mathbb{1}_{\{U_i \le Y\}}$ mit $U_1, U_2, \ldots \sim \mathbf{Unif}([0,1])$ u.i.v. und unabhängig von Y). Für $n \in \mathbb{N}$ und $x_1, \ldots, x_n \in \{0,1\}$ mit $x_1 + \ldots + x_n = s$ gilt:

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

Wir benutzen folgende Sprechweisen: Ein $\pi \in S_n$ fassen wir auch auf als (endliche) Permutation von $\mathbb{N}$ via $\pi(j) = j$ für $j > n$.
Ist $x = (x_1, \ldots, x_n) \in E^n$, so definieren wir $x^\pi := (x_{\pi(1)}, \ldots, x_{\pi(n)})$.
Ist $x = (x_1, x_2, \ldots) \in E^{\mathbb{N}}$, so definieren wir $x^\pi := (x_{\pi(1)}, x_{\pi(2)}, \ldots)$.
Für $f: E^n \to E$ definieren wir $f^\pi((x_1, \ldots, x_n)) := f(x^\pi)$.

Definition 2.4. *$f: E^n \to E'$ heißt (n-) symmetrisch, wenn $f = f^\pi$ für alle $\pi \in S_n$ gilt. $f: E^{\mathbb{N}} \to E'$ heißt n-symmetrisch, wenn $f = f^\pi$ für alle $\pi \in S_n$. f heißt symmetrisch, falls f n-symmetrisch ist für jedes $n \in \mathbb{N}$.*

Beispiel 2.5. i) Ist $E = \mathbb{R}$, $E' = \overline{\mathbb{R}}$, so sind die Funktionen $f((x_1, x_2, \ldots)) := \limsup_{n\to\infty} x_n$ und $f((x_1, x_2, \ldots)) := \limsup_{n\to\infty} \frac{x_1 + \ldots + x_n}{n}$ symmetrisch.

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

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

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

Beispiel 2.6. Sei $k \in N$, $\varphi: E^k \to R$ und $A_n(\varphi): E^{\mathbb{N}} \to \mathbb{R}$ mit

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

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

Definition 2.7. *Sei $X = (X_n)_n$ ein stochastischer Prozess mit Werten in E.*

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

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

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

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

Bemerkung 2.8. $\mathcal{E} = \big\{X^{-1}(B) \mid B \in \mathcal{B}(E)^{\otimes\mathbb{N}}$ mit $B^\pi = B$ für alle $\pi \in S_n, n \in \mathbb{N}\big\}$, wobei $B^\pi = \{x^\pi \mid x \in B\}$.

Beobachtung 2.9. Es sei $\mathcal{T} = \bigcap_{n\in\mathbb{N}} \sigma(X_n, X_{n+1}, \dots)$ die terminale σ-Algebra für X. Dann gilt $\mathcal{T} \subsetneq \mathcal{E}$.

Beweis. Es gilt $\sigma(X_n, X_{n+1}, \dots) \subset \mathcal{E}_{n-1}$, also $\mathcal{T} \subset \mathcal{E}$. Betrachte $|E| > 1$ und wähle $B \in \mathcal{B}(E) \setminus \{\varnothing, E\}$. $S := \sum_{n=1}^\infty \mathbb{1}_B(X_n)$ ist $\mathcal{E}$-messbar, aber $\{S = s\} \notin \mathcal{T}$ für $s \in \mathbb{N}_0$. □

Lemma 2.10. *Es sei $X = (X_n)_n$ eine Folge austauschbarer Zufallsvariablen mit Werten in E und $\varphi\colon E^k \to \mathbb{R}$ messbar mit $\mathbf{E}[|\varphi(X)|] < \infty$. Dann gilt für alle $n \geq k$ und $\pi \in S_n$:*

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

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

Beweis. Betrachte zunächst $A \in \mathcal{E}_n$ 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 $B_1, \dots B_k \in \mathcal{B}(\mathbb{R})$, $k \in \mathbb{N}$.
Sei $\pi \in S_n$. Dann ist

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

Solche As bilden einen $\cap$-stabilen Erzeuger von $\mathcal{E}_n$, daher gilt $\mathbf{E}[\varphi(X) \mid \mathcal{E}_n] = \mathbf{E}[\varphi(X^\pi) \mid \mathcal{E}_n]$

f.s. Damit gilt auch (f.s.)

$$\mathbf{E}[\varphi(X) \mid \mathcal{E}_n] = \frac{1}{n!}\sum_{\pi\in S_n}\mathbf{E}[\varphi(X^\pi) \mid \mathcal{E}_n] = \mathbf{E}\left[\frac{1}{n!}\sum_{\pi\in S_n}\varphi(X^\pi) \mid \mathcal{E}_n\right] = \mathbf{E}[A_n(\varphi)(X) \mid \mathcal{E}_n]$$
$$= A_n(\varphi)(X),$$

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

2.2 Rückwärtsmartingale

Definition 2.11. *Sei* $\mathcal{F} = (\mathcal{F}_{-n})_{n\in\mathbb{N}_0}$ *eine absteigende Filtration, d.h.* $\mathcal{F}_0 \supset F_{-1} \supset \mathcal{F}_{-2} \supset \dots$ $X = (X_{-n})_{n\in\mathbb{N}_0}$ *heißt ein* $\mathcal{F}$-Rückwärtsmartingal (unter $\mathbf{P}$), *wenn*

i) $\mathbf{E}[|X_i|] < \infty$ *für alle* $i \in \mathbb{N}_0$.

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

Wir nehmen implizit an, dass X_{-n} $\mathcal{F}_{-n}$*-messbar ist für alle* $n \in \mathbb{N}_0$.

Beispiel 2.12. Sei $X = (X_1, X_2, \dots)$ eine austauschbare Folge von Zufallsvariablen mit Werten in $\mathbb{R}$ und $\mathbf{E}[|X_1|] < \infty$. Sei $\mathcal{F}_{-n} := \mathcal{E}_n$ und $Y_{-n} := \frac{1}{n}\sum_{i=1}^n X_i$, $n \in \mathbb{N}$. Es gilt mit Lemma 2.10:

$$\mathbf{E}[Y_{-n+1} \mid \mathcal{F}_{-n}] = \mathbf{E}\left[\frac{1}{n-1}\sum_{i=1}^{n-1} X_i \mid \mathcal{E}_n\right] = \frac{1}{n-1}\sum_{i=1}^{n-1}\mathbf{E}[X_i \mid \mathcal{E}_n] = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{1}{n!}\sum_{\pi\in S_n} X_{\pi(i)}$$
$$= \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{1}{n!}(n-1)!\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} X_i = Y_{-n}.$$

Bemerkung 2.13. Wegen $X_{-n} = \mathbf{E}[X_0 \mid \mathcal{F}_{-n}]$, $n \in \mathbb{N}$ ist ein Rückwärtsmartingal stets gleichgradig integrierbar (vgl. Lemma 1.29).

Satz 2.14. *Sei* $(X_{-n})_n$ *ein Rückwärtsmartingal bezüglich* $(\mathcal{F}_{-n})_n$*. Dann existiert ein* $X_{-\infty}$ *mit* $X_{-n} \to X_{-\infty}$ *f.s. und in* $\mathcal{L}^1(\mathbf{P})$*. Es gilt* $X_{-\infty} = \mathbf{E}[X_0 \mid \mathcal{F}_{-\infty}]$ *mit* $\mathcal{F}_{-\infty} := \bigcap_n \mathcal{F}_{-n}$.

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 1.22 gilt:

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

Also existiert $U^{(a,b)} := \lim_n U_{-n}^{(a,b)}$ mit $\mathbf{E}\big[U^{(a,b)}\big] < \infty$. Analog zum Beweis von Satz 1.23 existiert damit auch $X_{-\infty} := \lim_n X_{-n}$ f.s. Mit Bemerkung 2.13 folgt $X_{-n} \to X_{-\infty}$ in $\mathcal{L}^1(\mathbf{P})$.

$X_{-\infty}$ ist $\mathcal{F}_{-\infty}$-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 $\mathcal{F}_{-k}$-messbar). Sei $A \in \mathcal{F}_{-\infty}$. Dann gilt

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

□

Korollar 2.15. *Sei $X = (X_1, X_2, \ldots)$ eine austauschbare Folge von Zufallsvariablen mit Werten in $\mathbb{R}$ und $\mathbf{E}[|X_1|] < \infty$. Sei $\mathcal{F}_{-n} := \mathcal{E}_n$, $\mathcal{T} = \bigcap_n \sigma(X_n, X_{n+1}, \ldots)$ die terminale σ-Algebra und $\mathcal{E} = \bigcap_n \mathcal{F}_{-n}$ die austauschbare σ-Algebra. Dann gilt*

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

Beweis. $Y_{-n} := \frac{1}{n}\sum_{i=1}^n X_i$ ist ein Rückwärtsmartingal. Es gilt $Y_{-n} \to Y_{-\infty} := \mathbf{E}[X_1 \mid \mathcal{E}]$ f.s. und in $\mathcal{L}^1(\mathbf{P})$ nach Satz 2.14 (wobei wir aus Notationsbequemlichkeit den Index um 1 verschoben haben). $Y_{-\infty}$ ist $\mathcal{T}$-messbar, denn es ist ein Grenzwert. Also gilt

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

□

Korollar 2.16 (Starkes Gesetz der großen Zahlen)**.** *Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte, reelle Zufallsvariablen mit $\mathbf{E}[|X_1|] < \infty$. Dann gilt*

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

denn gemäß Kolmogorovs 0-1-Gesetz gilt $P(A) \in \{0,1\}$ für alle $A \in \mathcal{T}$, demnach $\mathbf{E}[X_1 \mid \mathcal{T}] = \mathbf{E}[X_1]$ f.s. (vgl. [Dep14, Satz 3.48]).

Satz 2.17. *Sei $(X_n)_n$ austauschbar mit Werten in E und $\varphi\colon E^k \to \mathbb{R}$ eine messbare Funktion mit $\mathbf{E}[|\varphi(X_1, \ldots, X_k)|] < \infty$. Dann gilt*

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

Beweis. Nach Lemma 2.10 gilt $A_n(\varphi)(X) = \mathbf{E}[\varphi(X) \mid \mathcal{E}]$. Es gilt $\mathcal{E}_1 \supset \mathcal{E}_2 \supset \ldots \supset \mathcal{E} = \bigcap_n \mathcal{E}_n$, also ist $(A_n(\varphi)(X))_n$ ein Rückwärtsmartingal. Dann gilt nach Satz 2.14: $\lim_n A_n(\varphi)(X) = \mathbf{E}[\varphi(X) \mid \mathcal{E}]$.
Zeige, dass $\lim_n A_n(\varphi)(X)$ $\mathcal{T}$-messbar ist. Sei $l \in \mathbb{N}$ und $S_{n,l} := \{\pi \in S_n \mid \pi(1), \ldots, \pi(k) \geq l\}$. Dann ist $S_n \setminus S_{n,l} = \bigcup_{i=1}^k \{\pi \in S_n \mid \pi(i) < l\}$ und es gilt $|S_n \setminus S_{n,l}| \leq k(l-1)(n-1)!$. Weiter sei

$A_{n,l}(\varphi) := \frac{1}{n!}\sum_{\pi\in S_{n,l}}\varphi^\pi$. Dann gilt

$$\mathbf{E}[|A_{n,l}(\varphi)(X) - A_n(\varphi)(X)|] \le \frac{1}{n!}|S_n \setminus S_{n,l}|\cdot\mathbf{E}[|\varphi(X)|] \xrightarrow[n\to\infty]{} 0,$$

das heißt $A_{n,l}(\varphi)(X) - A_n(\varphi)(X) \to 0$ in $\mathcal{L}^1(\mathbf{P})$ und damit auch stochastisch. Wähle eine Teilfolge $n_m \nearrow \infty$ mit $A_{n,l}(\varphi)(X) - A_n(\varphi)(X) \to 0$ 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 $\mathcal{T}$-messbar. Mit der Turmeigenschaft folgt

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

□

Korollar 2.18. *Sei $(X_n)_n$ austauschbar. Dann gibt es für alle $A \in \mathcal{E}$ ein $B \in \mathcal{T}$ mit* $\mathbf{P}(A\triangle B) = 0$.

Beweis. Sei $A \in \mathcal{E} \subset \sigma(X_1, X_2, \ldots)$. Wähle $A_k \in \sigma(X_1, \ldots, X_k)$ mit $P(A \triangle A_k) \to 0$ für $k \to \infty$. Dies ist möglich nach dem Approximationssatz für Maße (vgl. [Dep14, Satz 1.54]). Sei $C_k \subset E^k$ messbar, sodass $A_k = \{(X_1, \ldots, X_k) \in C_k\}$ und sei $\varphi_k := \mathbb{1}_{C_k}$. Dann gilt $\varphi_k(X) \to \mathbb{1}_A$ (ggf. wähle eine Teilfolge k_m mit $\sum_m \mathbf{P}(A \triangle A_{k_m}) < \infty$). Dann gilt mit der dominierten Konvergenz für bedingte Erwartungen:

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

ψ ist $\mathcal{T}$-messbar mit $\psi = \mathbb{1}_A$ f.s. Sei $N := \{\psi \ne \mathbb{1}\}$. Dann ist $\mathbf{P}(N) = 0$ und es folgt $P(\{\psi = 1\} \triangle A) \le \mathbf{P}(N) = 0$. □

Korollar 2.19 (0-1-Gesetz von Hewitt und Savage)**.** *Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt, so gilt $\mathbf{P}(A) \in \{0,1\}$ für alle $A \in \mathcal{E}$.*

Beweis. Der Beweis folgt aus Korollar 2.18 und aus Kolmogorovs 0-1-Gesetz. □

2.3 Struktur unendlicher austauschbarer Familien

Eine Heuristik: Sei $E = \{1, 2, \ldots, k\}$ und seien $X_1, X_2, \ldots$ austauschbare, E-wertige Zufallsvariablen. 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 der Notation $(a)_n := a(a-1)\ldots(a-n+1)$

gilt

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

wenn wir für den Moment annehmen, dass sich die empirische Verteilung für $N \to \infty$ stabilisiert mit Grenzwert ξ_∞ (was aus Satz 2.22 folgen wird).

Zum Beispiel für $k = 2$, $j_1 + j_2 = N$ ist

$$\mathbf{P}(X_1 = 1, X_2 = 1, X_3 = 2) = \frac{j_1}{N} \cdot \frac{j_1 - 1}{N - 1} \cdot \frac{j_2}{N - 2} = \frac{(j_1)_2 (j_2)_1}{(N)_2}.$$

Definition 2.20. *Es sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Seien $\mathcal{G}, \mathcal{G}_i \subset \mathcal{F}, i \in I$ Teil-σ-Algebren. Die Familie $(\mathcal{G}_i)_{i\in I}$ heißt* unabhängig gegeben $\mathcal{G}$, *wenn für alle endlichen $J \subset I$, $A_j \in \mathcal{G}_j, j \in J$ gilt:*

$$\mathbf{P}\left(\bigcap_{j\in J} A_j \mid \mathcal{G}\right) = \prod_{j\in J} \mathbf{P}(A_j \mid \mathcal{G}) \quad \textit{f.s.}$$

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

Bemerkung 2.21. i) $(\mathcal{G}_i)$ ist stets unabhängig gegeben $\mathcal{G}$, wenn $\mathcal{G}_i \subset \mathcal{G}$ für alle $i \in I$.

ii) Unabhängigkeit gegeben $\{\emptyset, \Omega\}$ ist die „gewöhnliche" Unabhängigkeit.

iii) Sind $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$ σ-Algebren und ist $(\mathcal{G}_i)$ unabhängig gegeben $\mathcal{F}_1$ und unabhängig gegeben $\mathcal{F}_3$, so folgt nicht notwendigerweise die Unabhängigkeit gegeben $\mathcal{F}_2$, betrachte zum Beispiel X_1, X_2 unabhängige, reelle Zufallsvariablen, $\mathcal{G}_i = \sigma(X_i)$, $\mathcal{F}_1 = \{\emptyset, \Omega\}$, $\mathcal{F}_2 = \sigma(X_1 + X_2)$ und $\mathcal{F}_3 = \sigma(X_1, X_2)$.

Satz 2.22 (Satz von de Finetti)**.** *Sei $(X_n)_n$ eine Folge von Zufallsvariablen mit Werten in E. $(X_n)_n$ ist genau dann austauschbar, wenn es eine σ-Algebra $\mathcal{G}$ gibt, sodass $(X_n)_n$ unabhängig und identisch verteilt gegeben $\mathcal{G}$ ist. In diesem Fall kann man $\mathcal{G} = \mathcal{E}$ oder $\mathcal{G} = \mathcal{T}$ wählen.*

Beweis. Sei zunächst $(X_n)_n$ austauschbar. Sei $\mathcal{G} = \mathcal{E}$ oder $\mathcal{G} = \mathcal{T}$, $f_n : E \to \mathbb{R}$ beschränkt und messbar und sei $\varphi_k(X) = \prod_{i=1}^k f_i(X_i)$. Gilt

$$\mathbf{E}[\varphi_k(X) \mid \mathcal{G}] = \mathbf{E}[\varphi_{k-1}(X) \mid \mathcal{G}] \cdot \mathbf{E}[f_k(X_k) \mid \mathcal{G}], \tag{2.1}$$

so folgt induktiv $\mathbf{E}\big[\prod_{i=1}^k f_i(X_i) \mid \mathcal{G}\big] = \prod_{i=1}^k \mathbf{E}[f_i(X_i) \mid \mathcal{G}]$, d.h. $(X_n)_n$ sind bedingt unabhängig gegeben $\mathcal{G}$ (lese $f_i = \mathbb{1}_{B_i}$, um wörtlich an Definition 2.20 anzuschließen). Zeige also (2.1): Betrachte

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

mit $|R_{n,k}(X)| \leq \frac{k-1}{n} \|f_1\|_\infty \cdots \|f_k\|_\infty \xrightarrow[n\to\infty]{} 0$, also gilt (2.1).

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

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

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

3 Schwache Konvergenz und charakteristische Funktionen

3.1 Vorbemerkungen zur mengentheoretischen Topologie

Sei E ein topologischer (meist metrischer) Raum. Folgende Begriffe werden in diesem Kapitel vorausgesetzt: kompakt, relativkompakt, folgenkompakt, lokalkompakt, totalbeschränkt. Es bezeichne

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

Es gilt: $\mathcal{C}_c(E) \subset \mathcal{C}_b(E) \subset \mathcal{C}(E)$.

Definition 3.1. *Eine Teilmenge $A \subset E$ heißt σ-kompakt, wenn es kompakte Mengen $B_n \subset E$ gibt mit $\bigcup_n B_n = A$.*

Definition 3.2. *Sei μ ein σ-endliches Maß auf $(E, \mathcal{B}(E))$.*

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 $A \in \mathcal{B}(E)$.*

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

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

Es bezeichne

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

- $\mathcal{M}_{\le 1}(E)$ die Menge der Subwahrscheinlichkeitsmaße auf E.

Beispiel 3.3. i) Ist $f\colon \mathbb{R}^d \to [0,\infty]$ messbar mit $f \in \mathcal{L}^1_{\mathrm{loc}}(\mathbb{R}^d)$ und ist λ das Lebesguemaß auf $\mathbb{R}^d$, so ist $\mu = f\lambda$ ein Radon-Maß mit $\mu(B) = \int \mathbb{1}_B(x) f(x) \lambda(\mathrm{d}x)$.

ii) Ist $E = \mathbb{R}$, dann ist $\mu(\mathrm{d}x) = \mathbb{1}_{\mathbb{R}\setminus\{0\}}(x)\frac{1}{|x|}\lambda(\mathrm{d}x) + \delta_0(\mathrm{d}x)$ 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 3.4. 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 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 $\mu\left(\bigcup_{i=1}^{k_n} B_{\frac{1}{n}}(x_i^{(n)})\right) \ge \mu(E) - \frac{\varepsilon}{2^n}$. 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

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

□

Definition 3.5. *Sei $\mathfrak{F} \subset \mathcal{M}(E)$ und sei $\mathcal{C}$ eine Menge von messbaren Funktionen von E nach $\mathbb{R}$. $\mathcal{C}$ heißt* trennende Familie *(kurz* trennend*) für $\mathfrak{F}$, falls für alle $\mu, \nu \in \mathfrak{F}$ gilt:*

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

Beispiel 3.6. $f\colon E \to \mathbb{R}$ heißt *lipschitz-stetig*, wenn eine Konstante $c_f < \infty$ existiert, sodass für alle $x, y \in E$ gilt: $|f(x) - f(y)| \le c_f \cdot d(x,y)$. $L_f := \sup\limits_{x \ne y \in E} \dfrac{|f(x) - f(y)|}{d(x,y)}$ ist die Lipschitz-Konstante von f. Es bezeichne $\mathrm{Lip}(E)$ die lipschitz-stetigen Funktionen auf E und $\mathrm{Lip}_1(E) = \left\{f \in \mathrm{Lip}(E) \mid L_f \le 1\right\}$. Dann ist $\mathrm{Lip}_1(E)$ trennend für $\mathcal{M}_f(E)$ — falls E lokalkompakt auch $\mathrm{Lip}_1(E) \cap \mathcal{C}_c(E)$.

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 μ_1, μ_2 mit $\int f\,\mathrm{d}\mu_1 = \int f\,\mathrm{d}\mu_2$ für alle $f \in \mathrm{Lip}_1(E) \cap \mathcal{L}^1(\mu_1) \cap \mathcal{L}^1(\mu_2)$. Wir zeigen $\mu_1(K) = \mu_2(K)$ für alle kompakten $K \subset E$, denn dann folgt $\mu_1 = \mu_2$ 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, $K \subset U := \bigcup_{i=1}^n U_{x_i}$ für geeignete $x_1, \dots, x_n$. Dann ist $\mu_1(U) < \infty$ und $\mu_2(U) < \infty$ und es gilt $\delta := d(U, K^{\complement}) > 0$. Falls $\varepsilon < \delta$, dann gilt $\mathbb{1}_K \leq \rho_{K,\varepsilon} \leq \mathbb{1}_U$ und $\rho_{K,\varepsilon} \to \mathbb{1}_k$ punktweise für $\varepsilon \searrow 0$, also folgt mit der dominierten Konvergenz auch $\int \rho_{K,\varepsilon} \,\mathrm{d}\mu_i \to \mu_i(K)$ für $\varepsilon \searrow 0$. Wegen $\int \rho_{K,\varepsilon} \,\mathrm{d}\mu_1 = \frac{1}{\varepsilon} \int \varepsilon\rho_{K,\varepsilon} \,\mathrm{d}\mu_1 = \frac{1}{\varepsilon} \int \varepsilon\rho_{K,\varepsilon} \,\mathrm{d}\mu_2 = \int \rho_{K,\varepsilon} \,\mathrm{d}\mu_2$ folgt $\mu_1(K) = \mu_2(K)$. □

3.2 Schwache und vage Konvergenz

Definition 3.7. *Es sei E ein metrischer Raum.*

i) Seien $\mu_1, \mu_2, \ldots, \mu \in \mathcal{M}_f(E)$. Man sagt, μ_n konvergiert schwach *(engl.* weakly*) gegen μ, wenn für alle $f \in \mathcal{C}_b(E)$ gilt: $\int f \,d\mu_n \to \int f \,d\mu$ für $n \to \infty$. In diesem Fall schreibt man $\mu_n \to \mu$ schwach oder $\mu_n \xrightarrow{w} \mu$ oder $\mu = w - \lim \mu_n$.*

ii) Seien $\mu_1, \mu_2, \ldots, \mu \in \mathcal{M}(E)$. Man sagt, μ_n konvergiert vage *(auch* vag*, engl.* vaguely*) gegen μ, wenn für alle $f \in \mathcal{C}_c(E)$ gilt: $\int f \,d\mu_n \to \int f \,d\mu$ für $n \to \infty$. In diesem Fall schreibt man $\mu_n \to \mu$ vage oder $\mu_n \xrightarrow{v} \mu$ oder $\mu = v - \lim \mu_n$.*

Bemerkung 3.8. 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 (verwende Beispiel 3.6).

Beispiel 3.9. Es sei $E = \mathbb{R}$. Dann gilt $\delta_{\frac{1}{n}} \xrightarrow[n\to\infty]{\mathrm{w}} \delta_0$, aber $\delta_{\frac{1}{n}}\big((0,\infty)\big) = 1 \nrightarrow 0 = \delta_0\big((0,\infty)\big)$. $\delta_n \xrightarrow{\mathrm{v}} 0$-Maß, aber $(\delta_n)_n$ konvergiert nicht schwach.

Beobachtung 3.10. Sei E lokalkompakt und polnisch. Ist $(\mu_n) \subset \mathcal{M}_f(E)$ mit $\mu_n \xrightarrow{\mathrm{v}} \mu$, so gilt $\mu(E) \leq \liminf_n \mu_n(E)$.

Beweis. Wähle $f_N \in \mathcal{C}_c(E)$ mit $f_N \nearrow 1$. Dann gilt

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

□

Satz 3.11 (Portmanteau-Theorem)**.** *Sei E ein metrischer Raum und $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\leq 1}(E)$. Dann sind äquivalent:*

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

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

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

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

vi) $\mu_n(A) \to \mu(A)$ *für alle* $A \in \mathcal{B}(E)$ *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. Die Implikationen $iv) \Leftrightarrow v), iv), v) \Rightarrow vi), iii) \Rightarrow i) \Rightarrow ii)$ sind klar, ebenso $i) \Rightarrow vii)$ und $vii) \Rightarrow viii)$.

Zeige $ii) \Rightarrow iv)$: Bemerke zunächst: $f \equiv 1 \in \mathrm{Lip}(E) \cap \mathcal{C}_b(E)$, also $\mu(E) = \int 1 \,\mathrm{d}\mu = \lim \mu_n(E)$. Sei nun $F \subset E$ abgeschlossen und $\rho_{F,\varepsilon}(x) := 1 - \left(\frac{1}{\varepsilon} d(x,F) \wedge 1\right) \in \mathcal{C}_b(E) \cap \mathrm{Lip}(E)$. Dann gilt $\mathbb{1}_F \leq \mathbb{1}_{\rho_{F,\varepsilon}}$ und mit monotoner Konvergenz

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

Zeige $vi) \Rightarrow iii)$. Sei f beschränkt und messbar mit $\mu(U_f) = 0$. Behauptung: Für alle $D \in \mathcal{B}(\mathbb{R})$ gilt $\partial f^{-1}(D) \subset f^{-1}(\partial D) \cup U_f$. Sei $x \in \partial f^{-1}(D)$ und f stetig in x. Zu $\delta > 0$ existiert ein $\varepsilon > 0$ mit $f(B_\varepsilon(x)) \subset B_\delta(f(x))$. Es existiert ein $y \in f^{-1}(D) \cap B_\varepsilon(x)$ und es existiert 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 $f(y) \in B_\delta(f(x)) \cap D$ und $f(z) \in B_\delta(f(x)) \cap D^{\mathsf{c}}$. Sei nun $A = \left\{y \colon \mu(f^{-1}(\{y\})) > 0\right\}$. Die Atome von $\mu \circ f^{-1}$ sind höchstens abzählbar, also gibt es ein $N \in \mathbb{N}$ mit $y_0 \leq -\|f\|_\infty < y_1 < y_2 < \ldots < y_{N-1} < \|f\|_\infty < y_N$ mit $y_i \notin A$ 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

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

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

Zeige $viii) \Rightarrow vii)$. Es gelte $\mu = \mathrm{v}{-}\lim \mu_n$ und $\mu(E) \geq \limsup \mu_n(E)$. Nach Beobachtung 3.10 gilt aber auch $\mu(E) \leq \liminf_n \mu_n(E)$, d.h. $\mu(E) = \lim \mu_n(E)$.

Zeige $vii) \Rightarrow v)$. Sei $G \subset E$ offen und $\varepsilon > 0$. Wähle gemäß Beobachtung 3.4 ein $K \subset G$ kompakt mit $\mu(G \setminus K) < \varepsilon$. Da E lokalkompakt ist, gibt es eine kompakte Menge L mit $K \subset L^\circ \subset L \subset G$. 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 $\mathbb{1}_K \leq \rho_{K,\delta} \leq \mathbb{1}_L \leq \mathbb{1}_G$ und $\rho_{K,\delta} \in \mathcal{C}_c(E)$. Damit folgt

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

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

Definition 3.12. *Seien $X, X_1, X_2, \dots$ Zufallsvariablen mit Werten im metrischen Raum E.* X_n konvergiert in Verteilung gegen X, *wenn die Verteilungen $\mathcal{L}(X_n) \in \mathcal{M}_1(E)$ schwach gegen $\mathcal{L}(X)$ konvergieren. In diesem Fall schreibt man $X_n \xrightarrow{D} X$ oder $X_n \Rightarrow X$ für $n \to \infty$. Gelegentlich schreibt man $X_n \xrightarrow{D} \mu$ bzw. $X_n \Rightarrow \mu$, wenn $\mu = \mathcal{L}(X)$ und X unspezifiziert bleibt. Konvergiert X_n in Verteilung gegen X, so gilt für alle $f \in \mathcal{C}_b(E)$: $\mathbf{E}[f(X_n)] \to \mathbf{E}[f(X)]$.*

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

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

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

Die Umkehrung gilt i.A. nicht.

Beispiel 3.13. Seien $X, X_1, X_2, \dots$ unabhängig und identisch standardnormalverteilt. Dann gilt $X_n \Rightarrow X$, aber nicht $X_n \xrightarrow{\mathbf{P}} X$.

Beobachtung 3.14 (Lemma von Slutsky)**.** Seien $X, X_1, X_2, \dots, Y_1, Y_2, \dots$ Zufallsvariablen mit Werten im separablen metrischen Raum E. Gilt $X_n \Rightarrow X$ und $d(X_n, Y_n) \to 0$ stochastisch, dann gilt auch $Y_n \Rightarrow X$.

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

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

$\square$

Lemma 3.15 (Continuous mapping theorem)**.** *Seien (E_1, d_1) und (E_2, d_2) metrische Räume, $\varphi: E_1 \to E_2$ messbar und U_φ die Menge der Unstetigkeitsstellen von φ.*

i) Sind $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_{\le 1}(E_1)$ 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 $X, X_1, X_2, \dots$ E_1-wertige Zufallsvariablen mit $\mathbf{P}(X \in U_\varphi) = 0$ und $X_n \Rightarrow X$, so gilt $\varphi(X_n) \Rightarrow \varphi(X)$.

Beweis. i) Sei $f \in \mathcal{C}_b(E_2)$, 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 3.11 *iii)*

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

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

□

Bemerkung 3.16 (Der Fall $E = \mathbb{R}$). $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$ ist durch seine Verteilungsfunktion $F_\mu(x) = \mu((-\infty, x])$ festgelegt. Es gilt $F_{\mu_n} \Rightarrow F_\mu$, wenn $F_{\mu_n}(x) \to F_\mu(x)$ für alle Stetigkeitsstellen x von F_μ. Dies ist äquivalent zu $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ für alle $f \in \mathcal{C}_b(\mathbb{R})$ (vgl. [Dep14, Definition 4.1, Satz 4.10]).

3.3 Straffheit

Definition 3.17. *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 3.18. i) Ist E polnisch, so ist $\{\mu\}$ für jedes $\mu \in \mathcal{M}_f(E)$ straff nach Beobachtung 3.4.

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

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

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

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

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

Satz 3.19 (Satz von Prohorov)**.** *Sei E ein metrischer Raum und $\mathcal{K} \subset \mathcal{M}_{\leq 1}(E)$. Dann gilt:*

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

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.

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

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

$\bigcup_{j=1}^{\infty} K_j$ ist σ-kompakt und es gilt $\mu\left(E \setminus \bigcup_{j=1}^{\infty} K_j\right) = 0$ für alle $\mu \in \mathcal{K}$. Nehme also 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.

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

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

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

für alle $H \in \mathcal{H}$ existiert (möglich durch Diagonalargument). Für $G \subset E$ offen sei

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

für $B \subset E$ beliebig sei

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

Dadurch ist μ^* wohldefiniert.

Zeige nun: μ^* ist ein äußeres Maß. Es gilt $\mu^*(\varnothing) = 0$ nach Definition und $\mu^*(B) \leq \mu^*(B')$ für $B \subset B'$ nach Konstruktion. Es bleibt also nur noch die σ-Subadditivität zu zeigen. Seien dazu $G_1, G_2, \ldots \subset E$ offene Mengen, $\mathcal{H} \ni H \subset \bigcup_{n=1}^{\infty} G_n$. H ist kompakt, d.h. es 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 $H \subset \bigcup_{i=1}^{r} B_{\varepsilon_i}(x_{j_i})$ und $\overline{B_{\varepsilon_i}(x_{j_i})} \subset G_{m_i}$ für $i = 1, \ldots, r$. Weiter ist $H \subset K_{j_0}$ für ein genügend großes $j_0 \in \mathbb{N}$. Sei

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

Dann ist $H_m \subset G_m$ und $H \subset H_{m_1} \cup \ldots \cup H_{m_r}$. Also gilt

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

Nehme nun das Supremum über alle $H \subset G$, dann 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 $\mu^*(B_n) \geq \mu^*(G_n) - \frac{\varepsilon}{2^n}$ 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 $\varepsilon > 0$. Sei $O \supset B$ offene Obermenge mit $\mu^*(O) \leq \mu^*(B) + \varepsilon$. Seien $H_1 \in \mathcal{H}$ 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 der Monotonie von μ^* folgt nun

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

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 Caratheodory (vgl. [Dep14, Satz 1.50]) ist 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

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

Nehme nun das Supremum über $H \subset G, H \in \mathcal{H}$, dann folgt $\mu^*(G) \leq \liminf \mu_{n_k}(G)$. Mit Satz 3.11 $v)$ folgt nun die Behauptung.

Zeige nun $ii)$. Sei E polnisch und $x_1, x_2, \ldots$ eine Aufzählung einer dichten Teilmenge von E. Sei

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

Dann gilt $A_{n,N} \nearrow_{N\to\infty} E$ für jedes $n \in \mathbb{N}$. Sei $\mathcal{K} \subset \mathcal{M}_{\leq 1}(E)$ schwach relativ (folgen-) kompakt und sei

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

Zeige, dass $\delta = 0$. Sei n so groß, dass es für jedes $N \in \mathbb{N}$ ein $\mu_N \in \mathcal{K}$ gibt mit $\mu_N(A_{n,N}^c) \geq \frac{\delta}{2}$. Sei N_k eine Teilfolge mit $N_k \nearrow \infty$ und $\mu_{N_k} \xrightarrow{\mathrm{w}} \tilde{\mu} \in \mathcal{M}_{\leq 1}(E)$. Da $A_{n,N}^c$ abgeschlossen ist, gilt mit Satz 3.11 *iv)*

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

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

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

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

also ist $\mathcal{K}$ straff. □

Korollar 3.20. *Ist E kompakt, so sind $\mathcal{M}_{\leq 1}(E)$ und $\mathcal{M}_1(E)$ schwach (folgen-) kompakt.*

Beobachtung 3.21. Sei E ein polnischer Raum und seien $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\leq 1}(E)$. 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 $\mathcal{C} \subset \mathcal{C}_b(E)$ gilt $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ für alle $f \in \mathcal{C}$.

Beweis. *i)* ⇒ *ii)*. Sei $\mu = \mathrm{w} - \lim \mu_n$. Nach Definition gilt $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ für alle $f \in \mathcal{C}_b(E)$, insbesondere also auch für alle $f \in \mathcal{C}$ für jede trennende Familie $\mathcal{C} \subset \mathcal{C}_b(E)$. Die Straffheit von $\{\mu_n \mid n \in \mathbb{N}\}$ folgt mit Satz 3.19 *ii)*.

ii) ⇒ *i)*. Angenommen $\mu \neq \mathrm{w} - \lim \mu_n$. Dann gibt es ein $g \in \mathcal{C}_b(E)$, ein $\varepsilon > 0$ und eine Teilfolge $(n_k)_k$ mit

$$\left| \int g \mathrm{d}\mu_{n_k} - \int g \mathrm{d}\mu \right| \geq \varepsilon.$$

Nach Satz 3.19 *i)* gibt es eine Teilfolge $(n_{k_j})_j$ und ein $\nu \in \mathcal{M}_1(E)$ mit $\mu_{n_{k_j}} \xrightarrow[j\to\infty]{\mathrm{w}} \nu$. Dann gilt für alle $f \in \mathcal{C}$

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

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

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

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

Bemerkung 3.22. Sei E lokalkompakt und polnisch und seien $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_f(E)$. Dann sind äquivalent:

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

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

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

Beweis. Die Äquivalenz von $i)$ und $ii)$ gilt nach Satz 3.11, $i) \Rightarrow iii)$ ist klar.

$iii) \Rightarrow i)$. Sei $L \subset E$ kompakt mit $\sup_n \mu_n(E \setminus L) \leq 1$. Sei $h \in \mathcal{C}_c(E)$ mit $h \geq \mathbb{1}_L$. Dann gilt

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

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 $\mu'_n \xrightarrow{\mathrm{v}} \mu'$. Da E lokalkompakt ist, ist $C_c(E)$ trennend für $\mathcal{M}_{\leq 1}(E)$, also folgt $\mu' = \mathrm{w} - \lim \mu'_n$ mit Beobachtung 3.21 und damit auch $\mu = \mathrm{w} - \lim \mu_n$. □

3.4 Charakteristische Funktionen

Vorbemerkung. Sei E ein messbarer Raum. Eine Funktion

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

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

$$|\operatorname{Re} f|, |\operatorname{Im} f| \leq |f| = \left((\operatorname{Re} f)^2 + (\operatorname{Im} f)^2\right)^{\frac{1}{2}} \leq \sqrt{2}\,(|\operatorname{Re} f| + |\operatorname{Im} f|)\,.$$

Man setzt

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

Das Integral ist $\mathbb{C}$-linear und es gilt $\left|\int f \mathrm{d}\mu\right| \leq \int |f| \mathrm{d}\mu$.

Definition 3.23. *Sei $\mu \in \mathcal{M}_f(\mathbb{R}^d)$. Die Funktion*

$$\varphi_\mu: \mathbb{R}^d \to \mathbb{C}, \quad t \mapsto \int e^{i\langle t,x\rangle} \mu(dx) \tag{3.1}$$

heißt die charakteristische Funktion *von μ. Dabei ist $\langle t, x\rangle = \sum_{i=1}^d t_j x_i$. Für eine $\mathbb{R}^d$-wertige*

Zufallsvariable X schreiben wir

$$\varphi_X(t) = \mathbf{E}\left[e^{i\langle t,X\rangle}\right] = \varphi_{\mathcal{L}(X)}(t). \tag{3.2}$$

Bemerkung 3.24. In der Analysis ist auch die Bezeichnung *Fourier-Transformierte* üblich, zum Teil mit anderen Vorzeichenkonventionen.

Beispiel 3.25. i) Ist X diskret, so gilt $\varphi_X(t) = \sum_x \mathbf{P}(X = x)e^{i\langle t,x\rangle}$, insbesondere ist

$$\varphi_{\mathbf{Ber}(p)}(t) = pe^{it} + 1 - p,$$

$$\varphi_{\mathbf{Bin}(n,p)}(t) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{it} = (pe^{it} + 1 - p)^n,$$

$$\varphi_{\mathbf{Poi}(\lambda)}(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk} = e^{\lambda(e^{it}-1)}.$$

ii) Es ist $\varphi_{\mathbb{N}(\mu,\sigma^2)}(t) = e^{it\mu - \frac{1}{2}\sigma^2 t}$. Betrachte ohne Einschränkung (vgl. Lemma 3.26, *ii)*) den Fall $\mu = 0, \sigma^2 = 1$. Es gilt

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

iii) $\varphi_{\mathbf{Unif}([-1,1])}(t) = \frac{1}{2}\int_{-1}^{1} e^{itx} \mathrm{d}x = \frac{\sin(t)}{t}$.

Lemma 3.26. *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) $0 \le 1 - \operatorname{Re}\varphi_X(2t) \le 4(1 - \operatorname{Re}\varphi_X(t))$.

Beweis. i) Zeige die gleichmäßige Stetigkeit von φ_X. Sei $\varepsilon > 0$ und K so groß, dass

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

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

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

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

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

iv) Da X und Y unabhängig sind, gilt:
$$\varphi_{X+Y} = \mathbf{E}\left[e^{i\langle t,X+Y\rangle}\right] = \mathbf{E}\left[e^{i\langle t,X\rangle}e^{i\langle t,Y\rangle}\right] = \mathbf{E}\left[e^{i\langle t,X\rangle}\right] \cdot \mathbf{E}\left[e^{i\langle t,Y\rangle}\right] = \varphi_X(t)\varphi_Y(t).$$

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 $\operatorname{Re}\varphi_X(t) = \mathbf{E}[\cos(\langle t,X\rangle)]$, folgt damit die Behauptung.

□

Definition 3.27. *Sei $\mathbb{K} \in \{\mathbb{R},\mathbb{C}\}$. $\mathcal{C} \subset \mathcal{C}_b(E,\mathbb{K})$ 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 3.28 (Satz von Stone-Weierstraß)**.** *Sei E ein kompakter, topologischer Raum, $\mathbb{K} \in \{\mathbb{R},\mathbb{C}\}$ und $\mathcal{C} \subset \mathcal{C}_b(E,\mathbb{K})$ eine Punkte trennende Algebra. Falls $\mathbb{K} = \mathbb{C}$ sei $\mathcal{C}$ abgeschlossen unter komplexer Konjugation. Dann liegt $\mathcal{C}$ bezüglich der Supremumsnorm dicht in $\mathbb{C}_b(E,\mathbb{K})$.*

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

Zeige zuerst, dass für $f, g \in \overline{\mathcal{C}}$ auch $f \wedge g$, $f \vee g \in \overline{\mathcal{C}}$. 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

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

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

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

$$g_{x,\varepsilon}(x) = f(x) \quad \text{und} \quad g_{x,e} \leq f + \varepsilon. \tag{3.3}$$

Da $\mathcal{C}$ Punkte trennend ist, gibt es zu jedem $z \in E$, $z \neq x$ ein $H_z \in \mathcal{C}$ 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 für alle $z \in E$: $h_z(x) = f(x)$ 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 $z_1, \ldots, z_n$. Dann erfüllt

$$g_{x,\varepsilon} := \min\{h_{z_1}, \ldots, h_{z_n}\} \tag{3.4}$$

die geforderten Bedingungen in (3.3).

Sei nun $x \in E$ und $\varepsilon > 0$. Wähle $g_{x,\varepsilon}$ gemäß (3.4) und wähle eine offene Umgebung V_x von x mit $g_{x,\varepsilon}(y) \geq f(y) - \varepsilon$ 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

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

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

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

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

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

Korollar 3.29. *Sind $\mu, \nu \in \mathcal{M}_f(\mathbb{R}^d)$ mit $\varphi_\mu = \varphi_\nu$, so gilt $\mu = \nu$.*

Beweis. Sei $f \in \mathcal{C}_c(\mathbb{R}^d)$. Zeige $\int f \, \mathrm{d}\mu = \int f \, \mathrm{d}\nu$, denn falls dies für jedes solche f gilt, so ist $\mu = \nu$

nach Beispiel 3.6. Ist $f \equiv 0$, so sind die Integrale gleich. Betrachte also den Fall $f \not\equiv 0$. Sei $\varepsilon > 0$ und K so groß, dass $\operatorname{supp}(f) \subset [-K,K]^d$, $\mu\left(\mathbb{R}^d \setminus [-K,K]^d\right) \leq \frac{\varepsilon}{2\|f\|_\infty}$ und $\nu\left(\mathbb{R}^d \setminus [-K,K]^d\right) \leq \frac{\varepsilon}{2\|f\|_\infty}$. Für $m = (m_1,\dots,m_d) \in \mathbb{Z}^d$ und $x = (x_1,\dots,x_d) \in \mathbb{R}^d$ sei $f_{m,k}(x) = e^{i\frac{\pi}{K}\langle m,x\rangle}$. Die von $\left\{f_{m,k} \mid m \in \mathbb{Z}^d\right\}$ erzeugte Algebra $\mathcal{C}$ trennt Punkte von $\mathbb{R}^d/2K\mathbb{Z}^d$. Nach Satz 3.28 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 $\|g\|_\infty = \sup_{x\in[-K,K]^d} |g(x)|$)

$$\int |f-g| \mathrm{d}\mu \leq \varepsilon\mu([-K,K]^d) + 2\,\|f\|_\infty\,\mu\left(\mathbb{R}^d \setminus [-K,K]^d\right) \leq \varepsilon\mu([-K,K]^d) + \varepsilon \leq \varepsilon\mu(\mathbb{R}^d) + \varepsilon.$$

Die analoge Abschätzung gilt ebenso für ν. Zusammen folgt daher

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

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

Korollar 3.30. $\mu \in \mathcal{M}_f([0,\infty))$ *ist durch seine Laplace-Transformierte*

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

eindeutig festgelegt.

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

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

$\mathcal{C} = \{\sum_{i=1}^n a_i f_{\lambda_i} \mid n \in \mathbb{N},\ a_i \in \mathbb{R},\ \lambda_i \geq 0\}$ ist eine Punkte trennende Algebra, also folgt die Behauptung mit Satz 3.28. □

Korollar 3.31. *Sei X eine Zufallsvariable mit Werten in $[a,b]$. Dann ist ihre Verteilung durch ihre Momente $m_n := \mathbf{E}[X^n], n = 0,1,\dots$ festgelegt, denn die Polynome liegen dicht in $\mathcal{C}_b([a,b])$.*

Bemerke: Die Aussage des Korollars gilt nicht, wenn der Wertebereich von X unbeschränkt ist. Betrachte dazu folgendes Gegenbeispiel: Sei $X \sim \mathbb{N}(0,1)$ 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:

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

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

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

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

Satz 3.32 (Lévys Stetigkeitssatz)**.** *Seien $P, P_1, P_2, \ldots \in \mathcal{M}_1\left(\mathbb{R}^d\right)$ 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 $Q \in \mathcal{M}_1\left(\mathbb{R}^d\right)$ mit $f = \varphi_Q$ und $Q = w-\lim P_n$ (vgl. [Dep14, Satz 4.33] im Fall $d = 1$).

Lemma 3.33. *Sei $\mathfrak{F} \subset \mathcal{M}_1\left(\mathbb{R}^d\right)$ straff. Dann ist $\{\varphi_\mu \mid \mu \in \mathfrak{F}\}$ 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*

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

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

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

Wähle K so groß, dass $\sup_{\mu\in\mathfrak{F}} \mu\left(\mathbb{R}^d \setminus [-K,K]^d\right) < \frac{\varepsilon^2}{6}$ und wähle δ so klein, dass für alle $|u| < \delta$ gilt $\sup_{x\in[-K,K]^d} \left|1 - e^{i\langle u,x\rangle}\right| < \frac{\varepsilon^2}{6}$. Dann gilt für alle $\mu \in \mathfrak{F}$, $t, t' \in \mathbb{R}^d$ mit $|t - t'| < \delta$:

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

□

Beweis von Satz 3.32. Die punktweise Konvergenz $\varphi_n \to \varphi$ in $i)$ ist klar und die lokal gleichmäßige Konvergenz folgt damit aus Lemma 3.33 (vgl. z.B. [Kle13, Lemma 15.22] für weitere Details).

Zu $ii)$ zeigen wir zunächst, dass $\{P_n \mid n \in \mathbb{N}\}$ straff ist. Dafür 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, eine straffe Familie sind, denn es gilt $P_n\left(\mathbb{R}^d \setminus [-K,K]^d\right) \le \sum_{k=1}^d P_n^{(k)}([-K,K]^\mathsf{c})$. 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 $f(0 \cdot e_k) = 1$. Sei

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

und $a := \inf\{h(x) \mid |x| \ge 1\} = 1 - \sin(1) > 0$. Dann gilt für $\varepsilon > 0$ und M genügend groß mit dem Satz von Fubini:

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

Damit folgt:

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

also ist $\{P_n \mid n\in\mathbb{N}\}$ straff. Nach Satz 3.19 gibt es eine Teilfolge $(P_{n_j})_j$ mit $P_{n_j} \xrightarrow[j\to\infty]{\mathrm{w}} Q$ und mit $i)$ gilt $f=\varphi_Q$. Somit folgt $Q=Q'$ für jede konvergente Teilfolge $(P_{n_k})_k$ mit $P_{n_k} \xrightarrow{\mathrm{w}} Q'$. □

Erinnerung 3.34. Sei X eine reelle Zufallsvariable mit $\mathbf{E}[|X|^n]<\infty$. Dann ist φ_X n-mal stetig differenzierbar mit

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

(vgl. [Dep14, Lemma 4.34]), denn die Restgliedabschätzung der Taylorentwicklung liefert

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

und somit

$$\mathbf{E}\left[\left|e^{i(t+h)X}-e^{itX}\sum_{k=0}^n\frac{(ih)^kX^k}{k!}\right|\right] \le \mathbf{E}\left[\frac{|h|^{n+1}|X|^{n+1}}{(n+1)!}\wedge\frac{2|h|^n|X|^n}{n!}\right].$$

4 Zentrale Grenzwertsätze

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

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

Dann gilt $S_n^* \Rightarrow \mathbf{N}(0,1),\ n \to \infty$, denn es gilt mit der Taylorentwicklung

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

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

$$\mathbf{P}\left(a \le S_n^* \le b\right) \xrightarrow[n\to\infty]{} \int_a^b \frac{1}{2\pi} e^{-\frac{x^2}{2}} \mathrm{d}x$$

folgt aus $\mathcal{L}(S_n^*) \Rightarrow \mathbf{N}(0,1)$ zusammen mit Satz 3.11 *vi*).

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

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

Es gelte die Lindeberg-Bedingung

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

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

Bemerkung 4.3. i) Falls $(X_{n,l})$ die *Lyapunov-Bedingung*

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

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

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

ii) Die Lindeberg-Bedingung (4.1) impliziert

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

denn

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

Man sagt auch: Das Schema ist „asymptotisch vernachlässigbar“. Es gilt auch die Umkehrung: Gilt (4.3) und $S_n\Rightarrow\mathbf{N}(0,1)$, so gilt auch (4.1) (vgl. [Kal02, Theorem 5.12]).

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

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

denn für $n=2$ gilt

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

und der allgemeine Fall folgt induktiv.

Sei $c_{n,l}:=\mathbf{E}[X_{n,l}^2]=\mathbf{Var}[X_{n,l}]$ und $\varphi_{n,l}:=\mathbf{E}\left[e^{itX_{n,l}}\right]$. Es gilt

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

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

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

Damit folgt

$$\left|\mathbf{E}\left[e^{itS_n}\right] - e^{-\frac{1}{2}t^2}\right| = \left|\prod_{l=1}^{k_n} \varphi_{n,l}(t) - \prod_{l=1}^{k_n} e^{-\frac{1}{2}c_{n,l}t^2}\right| \le \sum_{l=1}^{k_n} \left|\varphi_{n,l}(t) - e^{-\frac{1}{2}c_{n,l}t^2}\right|$$
$$\le \sum_{l=1}^{k_n} \left|\varphi_{n,l}(t) - 1 + \frac{1}{2}c_{n,l}t^2\right| + \sum_{l=1}^{k_n} \left|e^{-\frac{1}{2}c_{n,l}t^2} - 1 + \frac{1}{2}c_{n,l}t^2\right|.$$

Mit Erinnerung 3.34 lässt sich das Argument der ersten Summe durch das Restglied der Taylorentwicklung abschätzen:

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

Weiterhin gilt

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

Zusammen ergibt sich

$$\left|\mathbf{E}\left[e^{itS_n}\right] - e^{-\frac{1}{2}t^2}\right| \le \sum_{l=1}^{k_n} \left|\varphi_{n,l}(t) - 1 + \frac{1}{2}c_{n,l}t^2\right| + \sum_{l=1}^{k_n} \left|e^{-\frac{1}{2}c_{n,l}t^2} - 1 + \frac{1}{2}c_{n,l}t^2\right|$$
$$\le \varepsilon \sum_{l=1}^{k_n} c_{n,l} + L_n(\varepsilon) + \frac{t^4}{8}(\max_{1\le l\le k_n} c_{n,l}) \sum_{l=1}^{k_n} c_{n,l}$$
$$= \varepsilon + L_n(\varepsilon) + \frac{t^4}{8}(\max_{1\le l\le k_n} c_{n,l}) \xrightarrow[n\to\infty]{} \varepsilon + 0 + 0 = \varepsilon.$$

Also gilt

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

Mit $\varepsilon \searrow 0$ und Levys Stetigkeitssatz (Satz 3.32) 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 $c_{ij} = \mathbf{Cov}[X_i, X_j]$ ist symmetrisch und positiv definit, denn $c_{ij} = \mathbf{Cov}[X_i, X_j] = \mathbf{Cov}[X_j, X_i] = c_{ji}$ und für $a = (a_1, \dots, a_d)^{\mathsf{t}} \in \mathbb{R}^d$ ist

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

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

Dichte

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

$\mathcal{L}(Z)$ 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) $\mathbf{E}[X] = (\mathbf{E}[X_1], \dots, \mathbf{E}[X_d]) = \mu$ und die Kovarianzmatrix $C := AA^{\mathrm{t}}$, denn

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

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 $\det C = \det(AA^{\mathrm{t}}) = (\det A)^2$:

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

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 $u \in \mathbb{R}^d$:

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

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

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

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

Lemma 4.7 (Cramér-Wold device)**.** *Für* $n \in \mathbb{N} \cup \{\infty\}$ *seien* $X_n = (X_{n,1}, \dots, X_{n,n})^{\mathsf{t}}$ *Zufallsvariablen. 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 $t \in \mathbb{R},\ \lambda \in \mathbb{R}^d$

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

Somit folgt $ii)$ aus Satz 3.32 (Lévys Stetigkeitssatz).

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

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

und somit folgt $i)$ mit Satz 3.32. □

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

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 $S_n^\lambda \Rightarrow \mathbf{N}(0, \mathbf{Var}[X_1^\lambda]) = \mathbf{N}(0, \langle \lambda, C\lambda\rangle)$, also gilt

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

Somit folgt die Behauptung mit Lemma 4.7. □

5 Unendlich teilbare Verteilungen

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 $X \overset{d}{=} X_{n,1} + \dots + X_{n,n}$.*

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

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 5.2. i) δ_x ist unendlich teilbar, denn $\delta_x = \delta_{\frac{x}{n}}^n$.

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

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

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

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

Beispiel und Definition 5.3. Zu $\nu \in \mathcal{M}_f(\mathbb{R}) \setminus \{0\}$ heißt

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

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

Sind $X_1, X_2, \dots$ unabhängige und identisch verteilte Zufallsvariablen mit $X_i \sim \tilde{\nu} := \frac{\nu}{\nu(\mathbb{R})}$ für alle $i \in \mathbb{N}$ und ist $N \sim \mathbf{Poi}(\nu(\mathbb{R}))$ unabhängig von $X_1, X_2, \dots$, so hat $S := \sum_{j=1}^N X_j$ die Verteilung $\mathbf{CPoi}(\nu)$ und die charakteristische Funktion ist

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

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

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

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

Für die charakteristische Funktion betrachte

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

$\square$

Satz 5.4. *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 $\mathbf{CPoi}(\nu_n) \xrightarrow[n\to\infty]{w} \nu$.*

Lemma 5.5. *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 $\varphi = e^{\psi}$ 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

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

Zeige $ii) \Rightarrow i)$. 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 $t \in \mathbb{R}$.

Zeige nun $i) \Rightarrow ii)$. 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 3.32 (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 3.32. 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 3.26 $v)$ 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.

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

und man zeigt leicht, dass dies (5.2) impliziert. □

Korollar 5.6. *i) Unter den Voraussetzungen von Lemma 5.5 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 $\varphi\colon \mathbb{R} \to \mathbb{C}$ 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 5.5 und μ_n mit $\varphi_{\mu_n} = \varphi_n$. Es ist $e^{rn(\varphi_n-1)}$ die charakteristische Funktion von $\mathbf{CPoi}(rn\mu_n)$, also ist $\varphi^r = \left(\lim_{n\to\infty} e^{n(\varphi_n-1)}\right)^r = e^{r\psi}$.

ii) Sei φ die charakteristische Funktion eines Wahrscheinlichkeitsmaßes und unendlich teilbar. Setze $\varphi_n = \varphi^{\frac{1}{n}}$. Dann gilt $\varphi_n^n(t) = \varphi(t) \to \varphi(t)$ für alle $t \in \mathbb{R}$.

Die Rückrichtung folgt aus Lemma 5.5.

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 5.4. Sei zunächst $(\nu_n) \subset \mathcal{M}_f(\mathbb{R})$ mit $\mathbf{CPoi}(\nu_n) \xrightarrow{\text{w}} \mu$. Jedes $\mathbf{CPoi}(\nu_n)$ ist unendlich teilbar, also ist auch μ unendlich teilbar nach Korollar 5.6 *iii*).

Sei nun $\mu \in \mathcal{M}_1(\mathbb{R})$ unendlich teilbar, $\mu = \mu_n^{*n}$ für $n \in \mathbb{N}$. Sei $\varphi_n = \varphi_{\mu_n}$ und $\varphi = \varphi_\mu$. Es ist $e^{n(\varphi_n-1)} = \varphi_{\mathbf{CPoi}(n\mu_n)}$ und $e^{n(\varphi_n(t)-1)} \xrightarrow[n\to\infty]{} \varphi(t)$, also gilt $\mathbf{CPoi}(n\mu_n) \xrightarrow{\text{w}} \mu$. □

Bemerkung 5.7. Sei $\nu \in \mathcal{M}_f(\mathbb{R})$ mit $\int_{\mathbb{R}} x^2\,\nu(\mathrm{d}x) < \infty$ und $X \sim \mathbf{CPoi}(\nu)$. Dann ist

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

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

$$\varphi_X'(t) = \int_{\mathbb{R}} ixe^{itx}\,\nu(\mathrm{d}x)\cdot\varphi_X(t),$$

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

Also gilt

$$\mathbf{E}[X] = -i\varphi_X'(0) = \int_{\mathbb{R}} x\,\nu(\mathrm{d}x),$$

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = -\varphi_X''(0) - \left(\int_{\mathbb{R}} x\,\nu(\mathrm{d}x)\right)^2 = \int_{\mathbb{R}} x^2\,\nu(\mathrm{d}x).$$

$\square$

Beobachtung 5.8. Sei $b \in \mathbb{R}$, $\sigma^2 \geq 0$ und $\nu_k \in \mathcal{M}_f(\mathbb{R})$ mit

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

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

$$\alpha_k := \int x\,\nu_k(\mathrm{d}x), \quad k = 1, 2, \ldots$$

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

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

Sofern $M_n := \sum_{k=1}^{n} (X_k - \alpha_k)$ f.s. konvergiert, ist X unendlich teilbar (nach Korollar 5.6, *iii)*, denn alle „Bauteile“ sind unendlich teilbar). Ist

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

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

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

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

$$\int \left(x^2 \wedge 1\right)\,\nu(\mathrm{d}x) < \infty$$

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

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

Definition 5.9. *Ein σ-endliches Maß ν auf $\mathbb{R}$ mit $\nu(\{0\}) = 0$ und $\int (x^2 \wedge 1)\, \nu(dx) < \infty$ heißt* kanonisches Maß. *(b, σ^2, ν) mit $b \in \mathbb{R}$, $\sigma^2 \in [0, \infty)$ und einem kanonischen Maß ν heißt* kanonisches Tripel.

Satz 5.10 (Lévy-Khinchin-Formel)**.** *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*

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

Das kanonische Tripel ist dabei eindeutig festgelegt.

Beweis. Sei zunächst ein solches kanonisches Tripel (b, σ^2, ν) gegeben, sodass (5.3) gilt. Dann folgt aus Beobachtung 5.8, dass μ unendlich teilbar ist. Zeige nun, dass ψ das kanonische Tripel festlegt. Sei $g_t(x) := e^{itx} - 1 - \mathbb{1}_{\{|x|<1\}}(x) \cdot itx$. 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:

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

Also ist σ^2 festgelegt. Sei nun ohne Einschränkung $\sigma^2 = 0$, ansonsten betrachte $\tilde{\psi}(t) = \psi(t) + \frac{1}{2}\sigma^2 t$. Sei

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

Dann gilt:

$$\overline{\psi}(t) := \psi(t) - \frac{1}{2}\int_{t-1}^{t+1} \psi(s)\mathrm{d}s = \int e^{itx}\left(1 - \frac{1}{2}\int_{-1}^{1} e^{isx}\mathrm{d}s\right) \nu(\mathrm{d}x) = \int e^{itx} h(x)\, \nu(\mathrm{d}x)$$

Demnach ist also $\overline{\psi}$ die charakteristische Funktion zu $\overline{\nu} := h\nu$. Es sind also $\overline{\nu}$ und ν durch ψ festgelegt und damit auch b.

Sei nun umgekehrt μ unendlich teilbar und $\psi = \log \varphi_\mu$. Dann ist $\mathrm{Re}(\psi) \leq 0$ und $t \mapsto \mathrm{Im}(\psi(t))$ ist ungerade. Daher ist $\overline{\psi}(0)$ reell und $\overline{\psi}(0) \geq 0$. Angenommen $\overline{\psi}(0) = 0$, dann wäre $\mu = \delta_b$ 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 damit (falls μ nicht trivial) $\left|\int e^{itx}\, \mu(\mathrm{d}x)\right| < \int \left|e^{itx}\right| \mu(\mathrm{d}x) = 1$ (denn eine nicht-triviale Konvexkombination von komplexen Zahlen vom Betrag 1 liegt strikt im Inneren des Einheitskreises), ein Widerspruch. Also ist $\overline{\psi}(0) > 0$. Wähle gemäß Satz 5.4

$(\nu_n)_n \subset \mathcal{M}_1(\mathbb{R} \setminus \{0\})$ mit $\mathbf{CPoi}(\nu_n) \xrightarrow{\mathrm{w}} \mu$. Sei

$$b_n := \int x \cdot \mathbb{1}_{\{|x|<1\}}(x)\, \nu_n(\mathrm{d}x),$$

$$\psi_n(t) = \log \varphi_{\mathbf{CPoi}(\nu_n)}(t) = \int \left(e^{itx} - 1\right) \nu_n(\mathrm{d}x) = \int g_t \mathrm{d}\nu_n + itb_n,$$

$$\overline{\psi}_n(t) = \psi_n(t) - \frac{1}{2}\int_{t-1}^{t+1} \psi_n(s)\mathrm{d}s = \int e^{itx} h(x)\, \nu_n(\mathrm{d}x).$$

Dann gilt nach Satz 3.32 $\psi_n \to \psi$ lokal gleichmäßig und ψ ist stetig, also gilt $\overline{\psi}_n(t) \to \overline{\psi}(t)$ für alle $t \in \mathbb{R}$. Für n genügend groß ist $\overline{\psi}_n(0) > 0$ und somit $\tilde{\nu}_n(\mathrm{d}x) = \frac{1}{\overline{\psi}_n(0)} h(x)\nu_n(\mathrm{d}x) \in \mathcal{M}_1(\mathbb{R})$ und $\int e^{itx}\, \tilde{\nu}_n(\mathrm{d}x) \xrightarrow[n\to\infty]{} \frac{\overline{\psi}(t)}{\overline{\psi}(0)}$ ist die charakteristische Funktion eines Wahrscheinlichkeitsmaßes. Es gibt also ein $\tilde{\nu} \in \mathcal{M}_1(\mathbb{R})$ mit $\tilde{\nu}_n \xrightarrow{\mathrm{w}} \tilde{\nu}$ und $\overline{\psi}(t) = \overline{\psi}(0) \int e^{itx}\, \tilde{\nu}(\mathrm{d}x)$. Sei $\sigma^2 := 6\overline{\psi}(0)\tilde{\nu}(\{0\})$, $\nu(\mathrm{d}x) := \frac{\overline{\psi}(0)}{h(x)} \cdot \mathbb{1}_{\{x \neq 0\}}(x)\tilde{\nu}(\mathrm{d}x)$ (dies ist ein kanonisches Maß). Sei weiter

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

f_t ist stetig und beschränkt, daher gilt

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

und damit

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

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

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

□

Bemerkung 5.11 (Uneindeutigkeit des komplexen Logarithmus). Bekanntlich gilt $\exp(z) = \exp(z + 2\pi i)$ für alle $z \in \mathbb{C}$. Ist $z = re^{i\alpha}$, so kann der Logarithmus durch $\log(z) := \log(r) + i\alpha$ als stetige Funkion auf jeder geschlitzten Ebene definiert werden, es gibt aber keine auf ganz $\mathbb{C} \setminus \{0\}$ stetige Version.

Andererseits ist in der Situation von Satz 5.10 die Abbildung $\mathbb{R} \ni t \mapsto \varphi(t) := \int e^{itx}\, \mu(\mathrm{d}x) \in$

$\mathbb{C} \setminus \{0\}$ (natürlich) stetig, und $\psi(t)$ ist eine stetige Version von $\log \varphi(t)$.

Beispiel 5.12. Ist $X \sim \mathbf{N}(1, \sigma^2)$, so ist $\varphi_X(t) = \exp(it - \frac{1}{2}\sigma^2 t)$.

Bemerkung 5.13. Anstelle der Abschneidefunktion $x \cdot \mathbb{1}_{\{|x|<1\}}$ in (5.3) kann prinzipiell jede Funktion $\tilde{f}$ mit $\tilde{f}(x) \sim x$ für $x \to 0$ und $\int |\tilde{f}(x) - x \cdot \mathbb{1}_{\{|x|<1\}}| \nu(\mathrm{d}x) < \infty$ gewählt werden. In der Literatur üblich sind auch $\tilde{f}(x) = \sin(x)$ (vgl. Taylorentwicklung von $\sin(x)$ ist x) oder $\tilde{f}(x) = \frac{x}{1+x^2}$. In der Lévy-Khinchin-Formel ändert sich dann b zu $\tilde{b} = b + \int \tilde{f}(x) - x \cdot \mathbb{1}_{\{|x|<1\}} \nu(\mathrm{d}x)$.

Beispiel 5.14. $\mathbf{N}(\mu, \sigma^2)$ hat das kanonische Tripel $(\mu, \sigma^2, 0)$.

Beobachtung 5.15. Sei μ ein unendlich teilbares Wahrscheinlichkeitsmaß auf $\mathbb{R}$. Dann ist $\nu := \mathrm{v} - \lim_{n\to\infty} n\mu^{\frac{1}{n}} |_{\mathbb{R}\setminus\{0\}}$.

Beispiel 5.16. Sei $a > 0$. $\mathbf{Cau}(a)$ 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 $a = 1$:

$$n\mathbf{Cau}\left(\frac{1}{n}\right)(A) = \frac{1}{\pi} \int_A \frac{n^2}{1 + n^2 x^2} \mathrm{d}x \xrightarrow[n\to\infty]{} \frac{1}{\pi} \int_A \frac{1}{x^2} \mathrm{d}x.$$

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

5.1 Ein Bericht über stabile Verteilungen

Beispiel 5.17. Sei $\alpha \in (0, 2)$ und $\nu_\alpha(\mathrm{d}x) := \frac{1}{\theta_\alpha} \mathbb{1}_{\{x\neq 0\}} |x|^{-\alpha-1} \mathrm{d}x$ mit

$$\theta_\alpha := \int (1 - \cos(x))|x|^{-\alpha-1} \mathrm{d}x = \begin{cases} -2\Gamma(-\alpha)\cos(\frac{\alpha\pi}{2}) & , \alpha \neq 1 \\ \pi & , \alpha = 1 \end{cases}$$

ν_α ist ein kanonisches Maß (vgl. [Fel71, S. 568-569]). Sei μ_α die unendlich teilbare Verteilung mit kanonischem Tripel $(0, 0, \nu_\alpha)$. μ_α heißt *(standard-) symmetrisch stabile Verteilung von Index α*. Es gilt

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

Der Imaginärteil des Integranden ist aus Symmetriegründen 0, daher folgt

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

Insbesondere gilt also: Sind $X_1, \dots, X_n \sim \mu_\alpha$ unabhängig, so ist $\frac{X_1 + \dots + X_n}{n^{\frac{1}{\alpha}}} \overset{\mathrm{d}}{=} X_1$, denn $\left(\varphi_{\mu_\alpha}\left(\frac{t}{n^{1/\alpha}}\right)\right)^n = \left(e^{-\frac{|t|^\alpha}{n}}\right)^n = \varphi_{\mu_\alpha}(t)$.

Definition 5.18. *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*

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

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

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

Beobachtung 5.19. Sei X unendlich teilbar mit kanonischem Tripel (b, σ^2, ν) und sei $a > 0$. Dann gilt:

$$\log \varphi_{aX}(t) = \log \varphi_X(at) = ibat - \frac{1}{2}\sigma^2 a^2 t^2 + \int e^{itax} - 1 - \mathbb{1}_{\{|x|<1\}}\, \nu(\mathrm{d}x)$$

$$= ibat - \frac{1}{2}\sigma^2 a^2 t^2 + \int e^{itax} - 1 - \mathbb{1}_{\{|ax|<1\}}\, \nu(\mathrm{d}x) + it \int ax \left(\mathbb{1}_{\{|ax|<1\}} - \mathbb{1}_{\{|x|<1\}}\right) \nu(\mathrm{d}x).$$

Also hat aX das kanonische Tripel

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

mit $f_a(x) = ax$.

Satz 5.20. *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*

$$\nu(dx) = c_- \mathbb{1}_{\{x<0\}} |x|^{-\alpha-1}\, dx + c_+ \mathbb{1}_{\{x>0\}} |x|^{-\alpha-1}\, dx. \tag{5.4}$$

Bericht 5.21. Die stabilen X aus Satz 5.20 haben

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

mit $c \in \mathbb{R}$, $d > 0$, $\theta = \frac{c_+ - c_-}{c_+ + c_-} \in [-1, 1]$ (vgl. [Bre68, Theorem 9.32]).

Bericht 5.22. Ist X stabil mit Index α, so ist

$$\mathbf{E}\left[|X|^\beta\right] \begin{cases} < \infty & , 0 \leq \beta < \alpha \\ \infty & , \beta \geq \alpha \end{cases}$$

Bericht 5.23 (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

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \xrightarrow[n\to\infty]{\mathrm{d}} Y \tag{5.5}$$

gilt, so ist Y stabil.

ii) Genau dann existieren $(a_n)_n$ und $(b_n)_n$ mit (5.5), sodass Y stabil ist mit Index $\alpha \in (0, 2)$, wenn es $c_+, c_- \geq 0$ 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 $\xi > 0$: $\lim\limits_{x\to\infty} \dfrac{\mathbf{P}(X > \xi x)}{\mathbf{P}(X > x)} = \dfrac{1}{\xi^\alpha}$, und falls $c_- > 0$, so gilt für alle $\xi > 0$: $\lim\limits_{x\to\infty} \dfrac{\mathbf{P}(X < -\xi x)}{\mathbf{P}(X < -x)} = \dfrac{1}{\xi^\alpha}$.

Bemerkung 5.24. Das Paradebeispiel dazu ist eine Zufallsvariable X mit Dichte

$$f(x) = c_1 \mathbb{1}_{\{x\geq 0\}} x^{-\alpha-1} + c_2 \mathbb{1}_{\{x<0\}} (-x)^{-\alpha-1}.$$

6 Markovprozesse

6.1 Grundlegendes: Stochastische Kerne, projektive Familien

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

i) Für alle $A_2 \in \mathcal{A}_2$ *gilt:* $S_1 \ni x \mapsto \kappa(x, A_2)$ *ist* $(\mathcal{A}_1 - \mathcal{B}(\mathbb{R}))$*-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* $\kappa(x, \cdot) \in \mathcal{M}_{\le 1}(S_2)$*.*

Beispiel 6.2. i) Sei $S_1 = S_2 = S$ höchstens abzählbar, $\mathcal{A}_1 = \mathcal{A}_2 = 2^S$ 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 $S_1 = S_2 = \mathbb{R}$, $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{B}(\mathbb{R})$ 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 6.3 (Produkt σ-Algebra)**.** Seien $(S_i, \mathcal{A}_i)$, $i \in I$ messbare Räume und sei $S = \bigtimes_{i \in I} S_i$. Dann ist $\mathcal{A} := \bigotimes_{i \in I} \mathcal{A}_i$ die kleinste σ-Algebra, bezüglich der alle kanonischen Projektionen $\pi_i: S \to S_i$ messbar sind.

Bemerkung und Definition 6.4. i) Seien $(S_i, \mathcal{A}_i)$, $i = 0, 1, 2$ 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

$$\begin{aligned} \kappa_1 \otimes \kappa_2: S_2 \times (\mathcal{A}_1 \otimes \mathcal{A}_2) &\to [0,1] \\ (x_0, A) &\mapsto \int_{S_1} \int_{S_2} \mathbb{1}_A(x_1, x_2)\, \kappa_2((x_0, x_1), \mathrm{d}x_2)\, \kappa_1(x_0, \mathrm{d}x_1) \end{aligned}$$

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

ii) Seien $(S_i, \mathcal{A}_i)$, $i = 0, 1, 2$ 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

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

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

Definition 6.5. *Ein messbarer Raum* $(S,\mathcal{A})$ *heißt* Borel-Raum *(oder auch* Standard-Borel-Raum*), wenn es ein* $B \in \mathcal{B}(\mathbb{R})$ *und eine Bijektion* $\varphi\colon S \to B$ *gibt, sodass* φ *und* φ^{-1} *messbar sind.*

Satz 6.6. *Jeder polnische Raum (ausgestattet mit seiner Borel-*σ*-Algebra) ist ein Borel-Raum.*

Beweisskizze. Betrachte zunächst $[0,1]^\infty$ mit Metrik $d((x_i),(y_i)) := \sum_{i=1}^\infty 2^{-i}|x_i - y_i|$. Für $x = (x_1,x_2,\dots) \in [0,1]^\infty$ sei $x_i = \sum_{j=1}^\infty 2^{-j}x_{i,j}$ mit $x_{i,j} \in \{0,1\}$ die Binärdarstellung. Sei $(a(n),b(n))_{n\in\mathbb{N}}$ eine Aufzählung von $\mathbb{N}^2$ und sei

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

ψ ist bijektiv und messbar, denn $y \mapsto k$-te Ziffer der Binärentwicklung ist messbar. Also ist $[0,1]^\infty$ ein Borel-Raum (vgl. [Dep14, Satz 3.30]).

Zeige nun: Ist E ein polnischer Raum, so gibt es ein messbares $S \subset [0,1]^\infty$ und eine bi-messbare Bijektion $\varphi\colon E \to S$. Sei dazu $x_1,x_2,\dots$ eine dichte Folge in E und definiere $\varphi(x) := (d(x,x_1)\wedge 1,\ d(x,x_2)\wedge 1,\dots) \in [0,1]^\infty$. Zeige, dass $y_n \xrightarrow[n\to\infty]{} y$ in E genau dann, wenn $d(y_n,x_m)\wedge 1 \xrightarrow[n\to\infty]{} d(y,x_m)\wedge 1$ für alle $m \in \mathbb{N}$ gilt. 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 $d(y_n,y)\wedge 1 \le (d(y_n,x_m)\wedge 1) + (d(y,x_m)\wedge 1)$, also ist $\limsup_{n\to\infty}(d(y_n,y)\wedge 1) \le 2\,(d(y,x_m)\wedge 1)$. 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 $\varphi^{-1}\colon \varphi(E) \to E$ ist stetig.

Zeige weiter: $S := \varphi(E) \subset [0,1]^\infty$ ist messbar. 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, 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 $V_n \subset [0,1]^\infty$ offen mit $x_n \in V_n \subset V_{n-1}$ und $\operatorname{diam}(\varphi^{-1}(V\cap S)) < \frac{1}{n}$ sowie $x_n' \in V_n \cap S$ mit $x_n' \to x$. $y_n' := \varphi^{-1}(x_n')$ ist eine Cauchyfolge in E. Da E vollständig ist, existiert ein $y \in E$ mit $y'_n \to y$. Es 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 $S = \bigcap_n U_n$ messbar, denn U_n ist relativ offen in $\overline{S}$ und damit messbar (vgl. [Wil00, Ch. II.82], [Bre68, Appendix 7]). □

Satz 6.7 (Existenz einer regulären Version der bedingten Wahrscheinlichkeiten für Borel-Wertebereiche)**.** *Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{G} \subset \mathcal{F}$ eine Teil-σ-Algebra und X eine Zufallsvariable mit Werten im Borel-Raum $(S, \mathcal{A})$. Dann gibt es einen stochastischen Kern κ von $(\Omega, \mathcal{G})$ nach $(S, \mathcal{A})$ mit*

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

für jedes $B \in \mathcal{A}$.

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

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

$$F_r \leq F_{r'}, \qquad \lim_{n\to\infty} F_n = 1, \qquad \lim_{n\to\infty} F_{-n} = 0, \qquad \lim_{n\to\infty} F_{r+\frac{1}{n}} = F_r \tag{6.1}$$

P-fast sicher, d.h. es gibt ein $N \in \mathcal{F}$ mit $\mathbf{P}(N) = 0$, sodass die Ereignisse aus (6.1) auf $\Omega \setminus N$ gelten. Ist

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

so ist $\tilde{F}_\cdot$ 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 $\kappa(\omega, (-\infty, r]) = F_r$ $\mathcal{G}$-messbar nach Konstruktion. Demnach ist $\omega \mapsto \kappa(\omega, B)$ $\mathcal{G}$-messbar für alle $B \in \mathcal{B}(\mathbb{R})$, denn die Menge aller B mit dieser Eigenschaft ist ein Dynkin-System, das den schnittstabilen Erzeuger $\{(-\infty, r] \mid r \in \mathbb{Q}\}$ von $\mathcal{B}(\mathbb{R})$ umfasst (vgl. [Dep14, Satz 1.36]).

Zeige nun, dass $\kappa(\cdot, B)$ für jedes $B \in \mathcal{B}(\mathbb{R})$ eine Version von $\mathbf{E}[\mathbb{1}_B(X) \mid \mathcal{G}]$ ist. Sei dazu $A \in \mathcal{G}$, $\nu_1(B) := \mathbf{E}[\mathbb{1}_A \kappa(\cdot, B)]$ und $\nu_2(B) := \mathbf{E}[\mathbb{1}_A \mathbb{1}_B(X)]$. 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. [Dep14, Satz 1.37]) folgt damit $\nu_1(B) = \nu_2(B)$ für alle $B \in \mathcal{B}(\mathbb{R})$, also gilt $\mathbf{E}[\mathbb{1}_A \kappa(\cdot, B)] = \mathbf{E}[\mathbb{1}_A \mathbb{1}_B(X)]$ für jedes $A \in \mathcal{G}$. □

Lemma 6.8 (Faktorisierungslemma)**.** *Sei Y eine Zufallsvariable mit Werten im messbaren Raum $(S', \mathcal{A}')$ und sei $f: \Omega \to \mathbb{R}$ $\sigma(Y)$-messbar. Dann gibt es eine messbare Funktion $g: S' \to \mathbb{R}$ mit $f = g \circ Y$ (vgl. [Kle13, Korollar 1.97]).*

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

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

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

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 Funtkionen gilt, folgt sie damit auch für f. □

Korollar 6.9. *In der Situation von Satz 6.7 sei $\mathcal{G} = \sigma(Y)$ für eine Zufallsvariable Y mit Werten im messbaren Raum $(S', \mathcal{A}')$. Dann gibt es einen stochastischen Kern κ' von S' nach S mit $\kappa'(Y, B) = \mathbf{E}[\mathbb{1}_B(X) \mid Y]$ fast sicher für alle $B \in \mathcal{A}$.*

Beweis. Die F_r, $r \in \mathbb{Q}$ aus dem Beweis von Satz 6.7 sind $\sigma(Y)$-messbar, also gibt es nach Lemma 6.8 eine messbare Funktion $F_r' \colon S' \to \mathbb{R}$ mit $F_r = F_r' \circ Y$. Führe die Konstruktion aus dem Beweis von Satz 6.7 für die F_r' durch und erhalte $\kappa'(y, \cdot)$ als das Wahrscheinlichkeitsmaß der Verteilungsfunktion $F_\cdot'(y)$, $y \in S$. □

Lemma 6.10. *Sei $(S, \mathcal{A})$ ein Borel-Raum, $(S', \mathcal{A}')$ ein messbarer Raum, X eine S-wertige und Y eine S'-wertige Zufallsvariable auf demselben Wahrscheinlichkeitsraum. Dann gibt es eine messbare Funktion $f \colon S' \times [0,1] \to S$, sodass gilt: Ist $\tilde{Y}$ eine Zufallsvariable mit $\tilde{Y} \stackrel{d}{=} Y$ und $\tilde{U} \sim \mathbf{Unif}([0,1])$ unabhängig von $\tilde{Y}$, so ist $\big(f(\tilde{Y}, \tilde{U}), \tilde{Y}\big) \stackrel{d}{=} (X, Y)$.*

Beweis. Sei ohne Einschränkung $S \subset \mathbb{R}$, ansonsten wähle $\varphi \colon S \to \hat{S} \subset \mathbb{R}$ bijektiv und bi-messbar wie im Beweis von Satz 6.7. Sei κ' ein stochastischer Kern von S' nach S gemäß Korollar 6.9. Sei

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

f ist messbar, denn $f(y,u) := \sup\{x \in \mathbb{Q} \mid \kappa'(y, (-\infty, x]) < u\}$ und es gilt $\{(y,u) \in S' \times [0,1] \mid \kappa'(y, (-\infty, x]) < u\} = \bigcup_{v \in \mathbb{Q}_+ \cap (0,1)} \{y \mid \kappa'(y, (-\infty, x]) < v\} \times [v, 1]$.

Es gilt für $x \in \mathbb{R}$ und $A \in \mathcal{A}$:

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

□

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

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

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

Definition 6.11. *Für $J \subset I$ mit $0 < |J| < \infty$ sei P_J ein Wahrscheinlichkeitsmaß auf $(\Omega_J, \mathcal{A}_J)$. $\{P_J \mid J \subset I,\ 0 < |J| < \infty\}$ heißt* projektive Familie*, wenn gilt*

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

Beispiel 6.12. 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 $P_n := P_{\{0,1,\dots,n\}} \in \mathcal{M}_1\left(S^{\{0,1,\dots,n\}}\right)$ gegeben durch

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

(und erhalte P_J für allgemeine $J \subset \mathbb{N}$, $0 < |J| < \infty$ durch „Aussummieren" der Koordinaten in $\{0, 1, \dots, n\} \setminus J$ 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 Übergansgmatrix p.)

Satz 6.13 (Kolmogorovs Erweiterungssatz)**.** *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* $(\Omega, \mathcal{A}) := (\Omega_I, \mathcal{A}_I)$ *mit* $P_J = P \circ (\pi^I_J)^{-1}$ *für jedes* $J \subset I$ *mit* $0 < |J| < \infty$. P *heißt* projektiver Limes der $\{P_J\}$, *auch geschrieben* $P = \underset{\overleftarrow{J \nearrow I}}{\lim} P_J$.

Beweis. Zeige zunächst die Eindeutigkeit. Angenommen P und P' seien projektiver Limes der $\{P_J\}$. Seien

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

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

Zeige nun die Existenz. Betrachte dazu zunächst den Fall, dass I abzählbar ist. Sei ohne Einschränkung $I = \mathbb{N}_0$ und $P_n := P_{\{0,1,\dots,n\}}$. Nach Lemma 6.10 gibt es eine messbare Funktion $f_n\colon S_0 \times \ldots \times S_n \times [0,1] \to S_{n+1}$, sodass gilt: Ist $(\tilde{X}_0, \ldots \tilde{X}_n) \sim P_n$ und $\tilde{U}_n \sim \mathbf{Unif}([0,1])$ 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 6.10). Sei $X_0 \sim P_0$ und seien $U_0, U_1, \ldots \sim \mathbf{Unif}([0,1])$ unabhängig von X_0 (für die Existenz von $U_0, U_1, \ldots$ vgl. [Dep14, Satz 3.30], [Kle13, Satz 1.64]). Konstruiere $X_1, X_2, \ldots$ via $X_{n+1} := f_n((X_0, \ldots, X_n), U_n)$, dann gilt induktiv $(X_0, \ldots, X_n) \sim P_n$. $P := \mathcal{L}(X_0, X_1, \ldots)$ 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 $\mathcal{C} \subset \mathcal{A}$. Es gilt $\Omega \in \mathcal{C}$, denn $\Omega = (\pi^I_{I'})^{-1}(\Omega_{I'})$. Sei weiter $A \in \mathcal{C}$, das heißt es gibt ein abzählbares $I' \subset I$ mit $A \in Z_{I'}$. Dann ist aber auch $A^c \in Z_{I'}$, also $A^c \in \mathcal{C}$. Seien $A_1, A_2, \ldots \in \mathcal{C}$, 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 $\bigcup_{n \in \mathbb{N}} A_n \in Z_{I'} \subset \mathcal{C}$. Damit ist also $\mathcal{C}$ eine σ-Algebra. Zudem gilt $\mathcal{A} \subset \mathcal{C}$, da alle Koordinatenprojektionen $\mathcal{C}$-messbar sind. Zusammen folgt $\mathcal{A} = \mathcal{C}$.

Für ein abzählbares $J \subset I$ gibt es nach dem ersten Teil des Beweises genau ein Wahrscheinlichkeitsmaß P_J auf $(\Omega_J, \mathcal{A}_J)$ mit $P_K = P_J \circ (\pi^J_K)^{-1}$ für jedes endliche $K \subset J$. Für $B \in \mathcal{A}_J$ sei $\tilde{P}_J\left((\pi^I_J)^{-1}(B)\right) := P_J(B)$, dann ist $\tilde{P}_J$ ein Maß auf Z_J. Für ein $A \in \mathcal{A} = \mathcal{C}$ setze

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

P ist wohldefiniert, denn ist $A \in Z_J \cap Z_{J'}$ eine Zylindermenge mit endlicher Basis K, dann ist $\tilde{P}_J(A) = P_K(A) = \tilde{P}_{J'}(A)$ und $\left\{(\pi^I_K)^{-1}(B) \mid K \subset J \cap J',\ |K| < \infty,\ B \in \mathcal{A}_K\right\}$ ist ein schnittstabiler Erzeuger von $Z_J \cap Z_{J'}$. Also gilt $\tilde{P}_J = \tilde{P}_{J'}$ auf $Z_J \cap Z_{J'}$.

Es bleibt nur noch zu zeigen, dass P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ ist. Es gilt $P(\varnothing) = 0$ und $P(\Omega) = 1$, denn $\varnothing, \Omega \in Z_K$ für jedes K. Seien $A_1, A_2, \ldots \in \mathcal{A}$ 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

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

□

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 (das heißt X_t ist eine E-wertige, $\mathcal{F}_t$-messbare Zufallsvariable) definiert auf einem filtrierten Raum $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\in I})$. Sei weiter $(P_x)_{x\in E}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{A})$.

Definition 6.14. *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 $x \in E$, $B \in \mathcal{B}(E)^{\otimes I}$ ist ein stochastischer Kern.

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

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

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 $E_x[\dots]$ für Erwartungswerte unter P_x, $\mathcal{L}_x(X)$ für die Verteilung von X unter P_x, analog $\mathcal{L}_x(X \mid \mathcal{G})$, etc.

Bemerkung 6.15. Ein Markov-Prozess besitzt die *elementare Markov-Eigenschaft*: Unter jedem P_x gilt für $u \le t$

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

Dies folgt unmittelbar aus Definition 6.14 *iii*), verlangt aber im Gegensatz zu Definition 6.14 nicht die zeitliche Homogenität der Dynamik.

Definition 6.16. *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.

Satz 6.17. *Ist* $((X_t)_{t\in I}, (P_x)_{x\in E})$ *ein Markov-Prozess, so definiert*

$$\kappa_t(x,A) := P_x(X_t \in A), \quad x \in E,\ A \in \mathcal{B}(E),\ t \in I \tag{6.2}$$

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* $f_1, \ldots, f_n : E \to \mathbb{R}$ *beschränkt und messbar*

$$\mathbf{E}_x\left[\prod_{j=1}^n f_j(X_{t_j})\right] = \int \kappa_{t_1-t_0}(x, dx_1) f_1(x_1) \cdots \int \kappa_{t_n-t_{n-1}}(x_{n-1}, dx_n) f_n(x_n). \tag{6.3}$$

Umgekehrt gibt es zu jeder Markov-Halbgruppe $(\kappa_t)_{t\in I}$ *einen Markov-Prozess* $(X_t)_{t\in I}$*, sodass (6.2) und (6.3) gelten.*

Beweis. Sei $((X_t)_{t\in I}, (P_x)_{x\in E})$ ein Markov-Prozess. Nach Definition 6.14 *ii)* definiert (6.2) einen stochastischen Kern. Weiter gilt für $s, t \in I$ und $A \in \mathcal{B}(E)$

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

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

Zeige (6.3) induktiv. Sei $f_1 = \mathbb{1}_A$, $A \in \mathcal{B}(R)$. Es ist

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

also gilt (6.3) für $n = 1$ für Linearkombination von Indikatorfunktionen und somit mit den „üblichen Approximationsargumenten" auch für allgemeine f. Es sei nun (6.3) wahr für ein $n \in \mathbb{N}$. Sei $\tilde{f}(X_{t_n}) := \int f_{n+1}(y) \kappa_{t_{n+1}-t_n}(X_{t_n}, \mathrm{d}y)$. Dann ist

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

Wende nun (6.3) 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 (6.3) ein Wahrscheinlichkeitsmaß $P_{x,J}$ auf $E^{|J|}$:

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

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 $A_i \in \mathcal{B}(E)$. Dann gilt

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

das heißt $P_{x,J} \circ (\pi^J_{J'})^{-1} = P_{x,J'}$. Mit Satz 6.13 folgt die Existenz eines Wahrscheinlichkeitsmaßes P_x auf $\Omega = E^I$, $\mathcal{A} = \mathcal{B}(E)^{\otimes I}$, sodass für die t-te Koordinatenprojektion $X_t : \Omega \to E$, $\mathcal{F}_t = \sigma(X_s \mid s \leq t)$ die Formel (6.3) gilt. $((X_t)_{t \in I}, (P_x)_{x \in E})$ leistet das Gewünschte.

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

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

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

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

zu zeigen. Betrachte dazu

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

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

Beispiel 6.18. Sei E abzählbar und $p = (p_{xy})_{x,y \in E}$ 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 6.17 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

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

Sei $X_t\colon\mathbb{R}^{[0,\infty)}\to\mathbb{R}$ die t-te Koordinatenprojektion. Nach Satz 6.17 gibt es auf $\Omega := \mathbb{R}^{[0,\infty)}$, $\mathcal{A} := \mathcal{B}(\mathbb{R})^{[0,\infty)}$ eine Familie von Wahrscheinlichkeitsmaßen $(P_x)_{x\in\mathbb{R}}$ mit

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

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 5.10. 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 $(\nu_t)_{t\geq0}$ ist eine Faltungshalbgruppe.

Speziell mit der Wahl $\nu = \mathbf{N}(0,1)$ haben wir somit eine „rohe" Version der Brownschen Bewegung konstruiert.

Beispiel 6.20. Sei E endlich, $(Q_{xy})_{x,y\in E}$ eine Matrix mit $Q_{xy}\geq 0$ für $x\neq y$ und $\sum_{y\in E} Q_{xy} = 0$

für alle $x \in E$. Sei

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

Die Reihe konvergiert, denn $\max_{x,y\in E} |Q^n_{xy}| \le (|E| \max_{x,y\in E} |Q_{x,y}|)^n$. Es gilt $P(t)P(s) = P(t+s)$, denn ist $AB = BA$ für Matrizen A, B, so ist $e^A e^B = e^{A+B}$. Weiter gilt

$$P_{xy}(t+s) = \sum_{z\in E} P_{xz}(t) P_{zy}(s).$$

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

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

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

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

insbesondere ist also $P(t)$ eine stochastische Matrix und es gilt

$$P_{x_0}(X_{t+n} = y \mid X_t = x) = \frac{P_{xx_0}(t) P_{xy}(h)}{P_{x_0x}(t)} = P_{xy}(h) = \left(e^{hQ}\right)_{xy} = \delta_{xy} + hQ_{xy} + o(h), \quad h \searrow 0$$

Bemerkung 6.21. In der Situation von Beispiel 6.20 kann man eine „Version“ von $(X_t)_{t\ge0}$ konstruieren, dessen Pfade rechtsstetig mit linken Limiten sind. Seien dazu $\tau_1, \tau_2, \dots$ unabhängig und identisch exponentialverteilt mit Parameter ϱ, $T_0 := 0$, $T_n = \sum_{i=1}^{n} \tau_i$ und $N_t := \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n \le t\}}$. $(N_t)_{t\ge0}$ ist ein Poisson-Prozess. Sei $(\hat{X}_k)_{k\in\mathbb{N}}$ eine zeitdiskrete Markov-Kette mit Übergangsmatrix $\hat{p}$, dann leistet $X_t := \hat{X}_{N_t}$, $t \ge 0$ das Gewünschte.

Satz 6.22. *$(X_t)_{t\in I}$ mit Wahrscheinlichkeitsmaßen $(P_x)_{x\in E}$ ist genau dann ein Markov-Prozess, wenn es einen stochastischen Kern $\kappa\colon E \times \mathcal{B}(E)^{\otimes I} \to [0,1]$ gibt, sodass für alle $x \in E$, $s \in I$ und alle beschränkten, messbaren Funktionen $f\colon E^I \to \mathbb{R}$ gilt*

$$\mathbf{E}_x\left[f((X_t)_{t\in I}) \mid \mathcal{F}_s\right] = \mathbf{E}_{X_s}\left[f((X_u)_{u\in I})\right] = \int_{E^I} f(y)\,\kappa(X_s, dy) \quad f.s. \tag{6.4}$$

Beweis. Es gelte zunächst (6.4). Für $t \in I$, $A \in \mathcal{B}(E)$ setze $f\colon E^I \to \mathbb{R}$, $y \mapsto \mathbb{1}_A(y_t)$ in (6.4) ein:

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

also gilt die schwache Markov-Eigenschaft (vgl. Definition 6.14 *iii*)).

Sei nun umgekehrt $(X_t)_{t\in I}$ ein Markov-Prozess. Sei f zunächst von der Form $f(y) = \mathbb{1}_{B_1}(y_{t_1})\cdots\mathbb{1}_{B_n}(y_{t_n})$ für $n \in \mathbb{N}$, $t_1 < \ldots < t_n \in I$ und $B_1, \ldots, B_n \in \mathcal{B}(E)$. Für $n = 1$ folgt (6.4) aus der schwachen Markov-Eigenschaft von $(X_t)_{t\in I}$. Nehme also an, dass (6.4) für ein festes $n \in \mathbb{N}$ erfüllt sei. Sei $f(y) = \mathbb{1}_{B_1}(y_{t_1})\cdots\mathbb{1}_{B_{n+1}}(y_{t_{n+1}})$. Dann gilt

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

also gilt (6.4) für Linearkombinationen solcher „Zylinderfunktionen“ und somit mit den üblichen Approximationsargumenten auch für allgemeine f. □

Bemerkung 6.23. 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

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

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 $\kappa_1\colon E \times \mathcal{B}(E) \to [0,1]$ mit

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

für alle $s \in \mathbb{N}_0$ und $A \in \mathcal{B}(E)$. 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 $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, wenn es eine stochastische Matrix

$(p_{xy})_{x,y\in E}$ 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 $\mathcal{B}(E)^{\otimes\mathbb{N}_0}=\mathcal{B}(E^{\mathbb{N}_0})$, demnach gibt es eine Version der bedingten Verteilung von $(X_{n+k})_{n\in\mathbb{N}_0}$ bedingt auf $\mathcal{F}_k$.

ii) Lese den Beweis von Satz 6.22 erneut mit $t_i=i$.

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

□

Bemerkung 6.24 (Formulierung der Markov-Eigenschaft mittels Shifts). Auf $(E^I,\mathcal{B}(E)^{\otimes I})$ 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 $(E^I,\mathcal{B}(E)^{\otimes I})$ definiert ist, so formuliert man (6.4) auch als

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

6.3 Die starke Markov-Eigenschaft

Definition 6.25. *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*

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

Satz 6.26. *Im Fall $I=\mathbb{N}_0$ besitzt jeder Markov-Prozess die starke Markov-Eigenschaft.*

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

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

wobei im vorletzten Schritt die schwache Markov-Eigenschaft bei Zeit s eingeht. $\square$

Satz 6.27 ("Spiegelungsprinzip"). *Seien $Y_1, Y_2, \ldots$ unabhängige und identisch verteilte reelle Zufallsvariablen mit $\mathcal{L}(Y_1) = \mathcal{L}(-Y_1)$. Sei $X_0 := 0$ und $X_k := \sum_{i=1}^k Y_i$. Dann gilt für $n \in \mathbb{N}$ und $a > 0$*

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

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

Beweis. $(X_k)_{k \in \mathbb{N}_0}$ ist eine Markov-Kette. Sei $a > 0$, $n \in \mathbb{N}$ und

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

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

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

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

$$\mathbf{P}(X_n > a) + \frac{1}{2}\mathbf{P}(X_n = a) = \mathbf{E}_0[f(\tau, (X_{\tau+k})_{k \in \mathbb{N}_0})] \ge \frac{1}{2}\mathbf{P}(\tau \le n) = \frac{1}{2}\mathbf{P}\left(\max_{m \le n} X_m \ge a\right). \tag{6.5}$$

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 (6.5). $\square$

Beispiel 6.28 (Eine Situation, in der die starke Markov-Eigenschaft nicht gilt). Im Fall $I = [0, \infty)$ fordert man, dass die Filtration $(\mathcal{F}_t)_{t \ge 0}$ rechtsstetig ist, das heißt $\mathcal{F}_t^+ := \bigcap_{s > t} F_s = \mathcal{F}_t$ für $t \ge 0$.

Sei $E = [0, \infty)$ und eine Markov-Halbgruppe $(\kappa_t)_{t \ge 0}$ gegeben durch

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

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

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

Sei $\tau := \inf\{t \geq 0 \mid X_t > 0\} = T\mathbb{1}_{\{X_0=0\}}$. Dann gilt

$$\mathbf{E}_0\left[f((X_{T+t})_{t\geq 0}) \mid \mathcal{F}_T^+\right] = f(t) \neq \int f(y)\,\kappa_t(0,\mathrm{d}y) = e^{-t}f(0) + \int_0^t e^{-s}f(t-s)\mathrm{d}s,$$

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

6.4 Diskrete Markov-Ketten

6.4.1 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

$$\mathbf{P}_{x_0}(X_{n+1} = y \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) = p_{xy}$$

für alle $n \in \mathbb{N}$ und $x_0, \dots, x_{n-1}, x, y \in E$ mit $\mathbf{P}_{x_0}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) > 0$. Wir notieren die n-ten Potenzen der Übergangsmatrix als $p_{xy}^n = \mathbf{P}_x(X_n = y)$.

Beispiel 6.29 (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.

6.4.2 Rekurrenz und Transienz

Definition 6.30. *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 6.31. i) Nach Konvention ist $T_x^{(1)} > 0$, auch bei Start in x.

ii) Die $T_x^{(k)}$ sind Stoppzeiten bezüglich $(\mathcal{F}_n)_n$ mit $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$.

Definition 6.32. *Für* $x, y \in E$ *sei*

$$F(x,y) := \mathbf{P}_x(T_y^{(1)} < \infty) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \{X_n = y\}\right)$$

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 6.33. *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

$$\begin{aligned}
\mathbf{P}_x(T_y^{(k+1)} < \infty) &= \mathbf{E}_x\Big[\mathbf{P}_x\big(T_y^{(k+1)} < \infty \,|\, \mathcal{F}_{T_y^{(k)}}\big)\mathbb{1}_{\{T_y^{(k)} < \infty\}}\Big] \\
&= \mathbf{E}_x\Big[\mathbf{P}_x\big(\inf\{n > 0 : X_{T_y^{(k)}+n} = y\} < \infty \,|\, \mathcal{F}_{T_y^{(k)}}\big)\mathbb{1}_{\{T_y^{(k)} < \infty\}}\Big] \\
&= \mathbf{E}_x\Big[\mathbf{P}_{X_{T_y^{(k)}}}\big(T_y^{(1)} < \infty\big)\mathbb{1}_{\{T_y^{(k)} < \infty\}}\Big] \\
&= \mathbf{E}_x\Big[\mathbf{P}_y\big(T_y^{(k)} < \infty\big)\mathbb{1}_{\{T_y^{(k)} < \infty\}}\Big] \\
&= F(y,y)\mathbf{P}_x(T_y^{(k)} < \infty) = F(y,y)F(x,y)F(y,y)^{k-1} \\
&= F(x,y)F(y,y)^k,
\end{aligned}$$

wobei wir in der vierten Zeile die starke Markov-Eigenschaft und in der fünften Zeile die Induktionsannahme verwendet haben. ☐

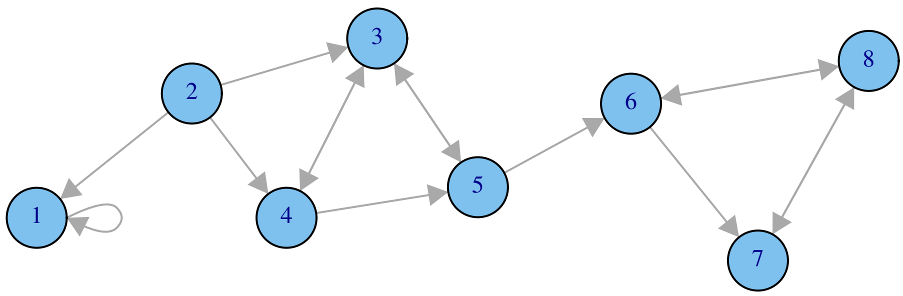
Definition 6.34. *Ein Zustand* $x \in E$ *heißt (bezüglich* p*)*

- rekurrent*, falls* $F(x,x) = 1$*,*
- positiv rekurrent*, falls* $\mathbf{E}_x\big[T_x^{(1)}\big] < \infty$ *(also insbesondere* $F(x,x) = 1$*),*
- nullrekurrent*, falls* $\mathbf{E}_x\big[T_x^{(1)}\big] = \infty$ *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 6.35. Es gilt: x absorbierend $\Rightarrow$ x positiv rekurrent $\Rightarrow$ x rekurrent .

Beispiel 6.36. 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 6.37. *Sei* $N(y) := \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = y\}}$ *die Anzahl der Besuche in* $y \in E$*. Sei*

$$G(x,y) := \mathbf{E}_x[N(y)] = \sum_{n=0}^{\infty} p_{x,y}^n.$$

G *heißt* Greenfunktion[1] von X*.*

Satz 6.38. *Für* $x, y \in E$ *gilt*

$$G(x,y) = \begin{cases} \frac{F(x,y)}{1-F(y,y)}, & y \neq x \\ \frac{1}{1-F(y,y)}, & y = x \end{cases}$$

[1] benannt nach George Green (1793-1841), der ein analoges Objekt in der Theorie der partiellen Differentialgleichungen eingeführt hat

mit Konvention $1/0 = \infty$. Insbesondere ist $G(x,y) = F(x,y)G(y,y) + \mathbb{1}_{\{x=y\}}$ und y ist genau dann rekurrent, wenn $G(y,y) = \infty$.

Beweis.

$$\begin{aligned} G(x,y) &= \mathbf{E}_x[N(y)] = \sum_{k=1}^{\infty} \mathbf{P}_x(N(y) \geq k) = \mathbb{1}_{\{x=y\}} + \sum_{k=1}^{\infty} \mathbf{P}_x(T_y^{(k)} < \infty) \\ &= \mathbb{1}_{\{x=y\}} + \sum_{k=1}^{\infty} F(x,y)F(y,y)^{k-1} = \mathbb{1}_{\{x=y\}} + \frac{F(x,y)}{1 - F(y,y)} \end{aligned}$$

□

Satz 6.39. *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 $\mathbf{E}_x[T_y^{(1)}], \mathbf{E}_y[T_x^{(1)}] < \infty$.*

Beweis. Sei ohne Einschränkung $y \neq x$. Wegen $F(x,y) > 0$ gibt es ein $k \in \mathbb{N}$ und $x_1, \dots, x_k \in E$ mit $x_k = y$, $x_i \neq x$ für $i = 1, \dots, k$ und $\mathbf{P}_x(X_1 = x_1, \dots, X_k = x_k) > 0$. Insbesondere ist $p_{xy}^k > 0$. Es gilt

$$\begin{aligned} 0 = 1 - F(x,x) = \mathbf{P}_x(T_x^{(1)} = \infty) &\geq \mathbf{P}_x(X_1 = x_1, \dots, X_k = x_k, T_x^{(1)} = \infty) \\ &= \mathbf{E}_x\big[\mathbf{P}_x(X_1 = x_1, \dots, X_k = x_k, T_x^{(1)} = \infty \mid \mathcal{F}_k)\big] \\ &= \mathbf{P}_x(X_1 = x_1, \dots, X_k = x_k)\mathbf{P}_y(T_x^{(1)} = \infty), \end{aligned}$$

also $\mathbf{P}_y(T_x^{(1)} = \infty) = 0$ und damit $F(y,x) = 1 - \mathbf{P}_y(T_x^{(1)} = \infty) = 1$. 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 $x_1, \dots, x_k$ wie zuvor, dann gilt

$$\begin{aligned} \infty > \mathbf{E}_x[T_x^{(1)}] &\geq \mathbf{E}_x\big[\mathbb{1}_{\{X_1=x_1,\dots,X_k=x_k\}} T_x^{(1)}\big] = \mathbf{E}_x\big[\mathbf{E}_x[\mathbb{1}_{\{X_1=x_1,\dots,X_k=x_k\}} T_x^{(1)} \mid \mathcal{F}_k]\big] \\ &= \mathbf{E}_x\big[\mathbb{1}_{\{X_1=x_1,\dots,X_k=x_k\}} \mathbf{E}_y[T_x^{(1)}]\big] = \mathbf{P}_x(X_1 = x_1, \dots, X_k = x_k)\mathbf{E}_y[T_x^{(1)}], \end{aligned}$$

demnach ist $\mathbf{E}_y[T_x^{(1)}] < \infty$. Nach Vertauschung der Rollen von x und y erhält man nun auch $\mathbf{E}_x[T_y^{(1)}] < \infty$. □

Definition 6.40. *Eine diskrete Markov-Kette heißt* irreduzibel*, wenn*

$$F(x,y) > 0 \quad \forall\, x,y \in E.$$

Bemerkung 6.41. 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 6.39.

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 6.42 (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 $p_{xy} = \mathbf{P}(Y_1 = y - x)$, das heißt unter $\mathbf{P}_x$ ist

$$(X_n)_{n \in \mathbb{N}_0} \overset{d}{=} (x + Y_1 + \dots + Y_n).$$

i) Sei $\mu := \mathbf{E}[Y_1] \neq \mathbf{0}$. Dann gilt für jedes $x \in \mathbb{Z}^d$

$$\frac{X_n}{n} \to \mu \quad \mathbf{P}_x\text{-f.s.},$$

insbesondere ist $\#\{n \in \mathbb{N}_0 \mid X_n = x\} < \infty$ $\mathbf{P}_x$-f.s., das heißt X ist transient.

ii) Sei $\mu := \mathbf{E}[Y_1] = \mathbf{0}$, $\mathbf{E}[\|Y_1\|^2] < \infty$, die Kovarianzmatrix $C = (\mathbf{Cov}[Y_{1_i}, Y_{1,j}])_{i,j=1,\dots,d}$ sei invertierbar und die von $\{y \mid \mathbf{P}(Y_1 = y) > 0\}$ erzeugte Gruppe sei $\mathbb{Z}^d$. Dann gilt

$$p_{\mathbf{0},\mathbf{0}}^n = \mathbf{P}_0(X_n = \mathbf{0}) = \mathbf{P}(Y_1 + \dots + Y_n = \mathbf{0}) \sim \frac{1}{(2\pi)^{d/2} |\det(C)|^{1/2} n^{d/2}}$$

für $n \to \infty$ (lokaler Zentraler Grenzwertsatz, zum Beispiel via Fourier-Inversion). Somit ist

$$G(\mathbf{0},\mathbf{0}) = \sum_{n=0}^{\infty} p^n_{\mathbf{0},\mathbf{0}} \begin{cases} = \infty, & d = 1,2, \\ < \infty, & d \geq 3, \end{cases}$$

das heißt eine zentrierte Irrfahrt mit endlicher Varianz ist genau dann rekurrent, wenn $d \leq 2$.

6.4.3 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 6.43. *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 6.44. *Sei jeder Zustand transient. Dann gibt es keine invariante Verteilung für p.*

Beweis. Nach Voraussetzung und Satz 6.38 ist $G(x,y) = \sum_n p^n_{xy} < \infty$, insbesondere gilt $p^n_{xy} \to 0$ 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 $\mu(E \setminus E') \leq \frac{1}{4}\mu(\{y\})$ und für genügend großes n ist

$$\max_{x\in E'} p^n_{xy} < \frac{\mu(\{y\})}{4\mu(E')}.$$

Somit folgt

$$\begin{aligned} \mu(\{y\}) = \mu p(\{y\}) = \cdots = \mu p^n(\{y\}) &= \sum_{x\in E'} \mu(\{x\}) p^n_{xy} + \sum_{x\in E\setminus E'} \mu(\{x\}) p^n_{xy} \\ &\leq \frac{1}{4}\mu(\{y\}) + \frac{1}{4}\mu(\{y\}) = \frac{1}{2}\mu(\{y\}), \end{aligned}$$

also ein Widerspruch. □

Satz 6.45. *Sei x rekurrenter Zustand, dann definiert*

$$\mu_x(\{y\}) := \mathbf{E}_x\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n = y\}}\right] = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x^{(1)} > x), \quad y \in E$$

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 $\hat{F}(x,y) := \mathbf{P}_x(T_y^{(1)} < T_x^{(1)})$. Dann ist $\hat{F}(x,y) > 0$ und $\hat{F}(y,x) > 0$ und es gilt

$$\begin{aligned}
\mathbf{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}}\right] &= 1 + \mathbf{E}_y\left[\sum_{n=T_y^{(1)}}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}} \mathbb{1}_{\{T_y^{(1)}<T_x^{(1)}\}}\right] \\
&= 1 + \mathbf{E}_y\left[\mathbf{E}_y\left[\sum_{n=T_y^{(1)}}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}} \mathbb{1}_{\{T_y^{(1)}<T_x^{(1)}\}} \mid \mathcal{F}_{T_y^{(1)}}\right]\right] \\
&= 1 + \mathbf{E}_y\left[\mathbb{1}_{\{T_y^{(1)}<T_x^{(1)}\}} \mathbf{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}}\right]\right] \\
&= 1 + \big(1-\hat{F}(y,x)\big)\mathbf{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}}\right],
\end{aligned}$$

das heißt

$$\mathbf{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}}\right] = \frac{1}{\hat{F}(y,x)}.$$

Damit folgt

$$\begin{aligned}
\mu_x(\{y\}) = \mathbf{E}_x\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}}\right] &= \mathbf{E}_x\left[\mathbb{1}_{\{T_y^{(1)}<T_x^{(1)}\}} \sum_{n=T_y^{(1)}}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}}\right] \\
&= \mathbf{P}_x(T_y^{(1)} < T_x^{(1)})\mathbf{E}_y\left[\sum_{n=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_n=y\}}\right] = \frac{\hat{F}(x,y)}{\hat{F}(y,x)} < \infty
\end{aligned}$$

Setze nun $\overline{p}_n(x,y) := \mathbf{P}_x(X_n = y, T_x^{(1)} > n)$, 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

$$\sum_{y\in E} \overline{p}_n(x,y) = \sum_{y \in E} \underbrace{\mathbf{P}_x(X_n = y,\ T_x^{(1)} > n,\ X_{n+1} = z)}_{=\mathbf{P}_x(X_n=y,\ X_{n+1}=z,\ T_x^{(1)}>n+1)} = \mathbf{P}_x(X_{n+1} = z,\ T_x^{(1)} > n+1)$$

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

$$\sum_{y \in E} \overline{p}_n(x,y) p_{yx} = \sum_{y \in E} \mathbf{P}_x(X_n = y,\ T_x^{(1)} > n,\ X_{n+1} = x) = \mathbf{P}_x(T_x^{(1)} = n+1),$$

das heißt

$$\mu_x p(\{x\}) = \sum_{n=0}^{\infty} \sum_{y \in E} \overline{p}_n(x,y) p_{yz} = \sum_{n=0}^{\infty} \mathbf{P}_x(T_x^{(1)} = n+1) = 1 = \mu_x(\{x\}).$$

Damit ist μ_x ein invariantes Maß. □

Korollar 6.46. *Ist x positiv rekurrent, so definiert*

$$\pi := \frac{1}{\mathbf{E}_x\big[T_x^{(1)}\big]} \mu_x$$

eine invariante Verteilung.

Satz 6.47. *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 6.48. *Eine irreduzible Markov-Kette X ist genau dann positiv rekurrent, wenn sie eine* *(notwendigerweise eindeutige) invariante Verteilung π besitzt. Diese ist dann gegeben durch*

$$\pi(\{x\}) = \frac{1}{\mathbf{E}_x\big[T_x^{(1)}\big]} > 0, \quad x \in E.$$

Beweis. Wenn x positiv rekurrent ist, so gibt es nach Korollar 6.46 eine invariante Verteilung, welche nach Satz 6.47 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 $\mathbf{P}_\pi := \sum_{x\in E} \pi(\{x\})\mathbf{P}_x$, das heißt unter $\mathbf{P}_\pi$ 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 $\{0, 1, \dots, n\} \cup \{-\infty\}$. Für $k \le n$ ist

$$\begin{aligned}
\mathbf{P}_\pi(\sigma_x^{(n)} = k) &= \mathbf{P}_\pi(X_k = x, X_{k+1} \ne x, \dots, X_n \ne x)\\
&= \mathbf{P}_\pi(X_k = x)\mathbf{P}_x(X_1 \ne x, \dots, X_{n-k} \ne x)\\
&= \pi(\{x\})P_x(T_x^{(1)} \ge n - k + 1),
\end{aligned}$$

wobei wir in der zweiten Zeile die Markov-Eigenschaft und in der dritten Zeile die Invarianz von π ausnutzen. Somit folgt

$$\begin{aligned}
1 &= \textstyle\sum_{k=0}^n \mathbf{P}_\pi(\sigma_x^{(n)} = k) + \mathbf{P}_\pi(\sigma_x^{(n)} = -\infty)\\
&\overset{(\dagger)}{=} \pi(\{x\}) \textstyle\sum_{k=0}^n P_x(T_x^{(1)} \ge n - k + 1) + \mathbf{P}_\pi(T_x^{(1)} \ge n + 1).
\end{aligned}$$

Weiter gilt wegen der Irreduzibilität und mit monotoner Konvergenz

$$\mathbf{P}_\pi(T_x^{(1)} \ge n + 1) = \sum_y \pi(\{y\})\mathbf{P}_y(T_x^{(1)} \ge n + 1) \xrightarrow[n\to\infty]{} 0,$$

das heißt mit $n \to \infty$ folgt aus $(\dagger)$

$$1 = \pi(\{x\}) \textstyle\sum_{\ell=1}^\infty P_x(T_x^{(1)} \ge \ell) = \pi(\{x\})\mathbf{E}_x\big[T_x^{(1)}\big].$$

Damit ist $\mathbf{E}_x[T_x^{(1)}] < \infty$ und folglich x positiv rekurrent. □

6.4.4 Konvergenz ins Gleichgewicht

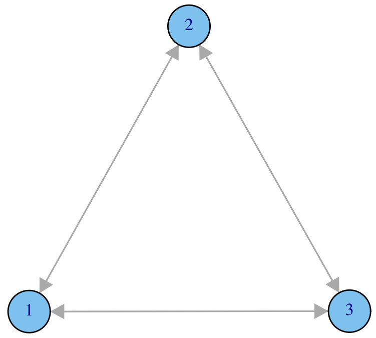
Definition 6.49. *Für $x, y \in E$ sei $N(x, y) := \{n \in \mathbb{N} \mid p_{xy}^n > 0\}$.*

i) $d_x := ggT\big(N(x, x)\big)$ heißt Periode *des Zustands x.*

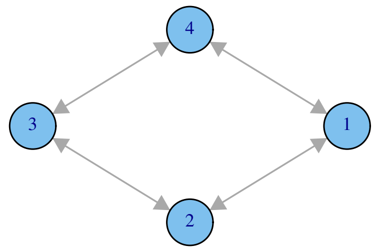
ii) Falls $d_x = d_y = d$ für alle $x, y \in E$, so heißt d die Periode *der Markov-Kette X.*

iii) X heißt aperiodisch*, wenn $d = d_x = 1$ für alle $x \in E$.*

Beispiel 6.50.



aperiodisch



Periode $d = 2$

Lemma 6.51. *i) Für $x \in E$ gibt es ein $n_x \in \mathbb{N}$ mit $p_{xx}^{nd_x} > 0$ 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 $\operatorname{ggt}(\tilde{N}) = 1$ 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 $r \in \{1, \dots, k-1\}$. 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 $r \in \{0, 1, \dots, n' - 1\}$ 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 $p_{xz}^{m+n} \geq p_{xy}^m p_{yz}^n$.

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 6.52. Im irreduziblen Fall mit Periode $d > 1$ kann man $E = E_0 \cup E_1 \cup \dots \cup E_{d-1}$ disjunkt zerlegen, so dass

$$p_{xy} > 0, x \in E_i \;\Rightarrow\; y \in E_{(i+1) \bmod d}$$

gilt (vgl. [Kle13, Satz 18.4]).

Satz 6.53. *Sei X eine aperiodische, irreduzible Markov-Kette mit invarianter Verteilung π. Dann gilt für jedes $x \in E$*

$$\sum_{y \in E} \left| \mathbf{P}_x(X_n = y) - \pi(\{y\}) \right| \xrightarrow[n\to\infty]{} 0.$$

Beweis. Wir konstruieren eine *Kopplung* von $\mathbf{P}_x$ und $\mathbf{P}_\pi$.

$\overline{p}_{(x_1,x_2),(y_1,y_2)} := p_{x_1,y_1} p_{x_2,y_2}$ ist eine irreduzible stochastische Matrix auf $E \times E$ (die Irreduzibilität verwendet Lemma 6.51), $\overline{\pi}(\{x,y\}) := \pi(\{x\})\pi(\{y\})$ ist zugehörige invariante Verteilung. Sei $(X_n, Y_n)_n$ eine $\overline{p}$-Markov-Kette mit Startverteilung $\delta_x \otimes \pi$. Nach Satz 6.48 ist (X,Y) rekurrent, insbesondere ist $T := \inf\{n \in \mathbb{N}_0 : X_n = Y_n\} < \infty$ f.s. und es gilt

$$\begin{aligned}
\mathbf{P}(X_n = y, T \le n) &= \sum_{m=0}^{n} \sum_x \underbrace{\mathbf{P}(T = m, X_m = x, X_n = y)}_{=\mathbf{P}(T=m,X_m=x)p_{xy}^{n-m}=\mathbf{P}(T=m,Y_m=x)p_{xy}^{n-m}} \\
&= \sum_{m=0}^{n} \sum_x \mathbf{P}(T = m, Y_m = x, Y_n = y) = \mathbf{P}(Y_n = y, T \le n), \\
\mathbf{P}(X_N = y) &= \mathbf{P}(Y_n = y, T \le n) + \mathbf{P}(X_N = y, T > n) \le \mathbf{P}(Y_N = y) + \mathbf{P}(X_N = y, T > n)
\end{aligned}$$

und analog $\mathbf{P}(Y_N = y) \le \mathbf{P}(X_N = y) + \mathbf{P}(Y_N = y, T > n)$, somit

$$\sum_{y \in E} \left| \mathbf{P}_x(X_n = y) - \pi(\{y\}) \right| = \sum_{y \in E} \left| \mathbf{P}(X_n = y) - \mathbf{P}(Y_n = y) \right| \le 2\mathbf{P}(T > n) \xrightarrow[n\to\infty]{} 0.$$

□

Beispiel 6.54 (Erneuerungskette)**.** 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. $\square$

Satz 6.55 (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,* $\mathbf{E}[\xi_1] = \mu$, $T_0 := x_0$ *und* $T_m := x_0 + \xi_1 + \dots + \xi_m$ *für* $m \in \mathbb{N}$ *ergibt sich aus Satz 6.53*

$$\mathbf{P}\big(\exists\, k \in \mathbb{N} : T_k = n\big) = \mathbf{P}_{x_0}(X_n = 0) \xrightarrow[n\to\infty]{} \pi(\{0\}) = \frac{1}{\mu}.$$

6.4.5 Markov-Ketten und Randwertprobleme

Definition 6.56. *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 6.57. Ist f harmonisch und $(X_n)_n$ eine p-Markov-Kette, so ist $(f(X_n))_n$ ein Martingal.

Definition 6.58. *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*

$$\begin{aligned}(pf - f)(x) &= 0, \quad x \in E \setminus A,\\ f(x) &= g(x), \quad x \in A.\end{aligned}$$

Beobachtung 6.59. 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 $\mathbf{P}_x(\tau_A < \infty) = 1$ für alle $x \in E$. Dann löst $f(x) := \mathbf{E}_x[g(X_{\tau_A})]$ das Dirichlet-Problem mit Randwerten g.

Beweis. Offenbar ist $f(x) = g(x)$ für $x \in A$, denn dann ist $\mathbf{P}_x(\tau_A = 0) = 1$. Für $x \notin A$ gilt wegen der Markov-Eigenschaft

$$\mathbf{E}_x[g(X_{\tau_A})|X_1 = y] = \begin{cases} g(y) = f(y), & y \in A,\\ \mathbf{E}_y[g(X_{\tau_A})] = f(y), & y \notin A.\end{cases}$$

Demnach ist für $x \notin A$

$$f(x) = \sum_y \mathbf{P}_x(X_1 = y)\mathbf{E}_x[g(X_{\tau_A})|X_1 = y] = \sum_y p_{xy} f(y) = pf(x).$$

$\square$

Bemerkung 6.60. Auch $\tilde{f}(x) = \mathbf{E}_x\big[g(X_{\tau_A})\mathbb{1}_{\{\tau_A<\infty\}}\big]$ löst das Dirichlet-Problem. Ist $\mathbf{P}_x(\tau_A = \infty) > 0$, so kann es verschiedene Lösungen geben.

Sei $\tilde{p}_{xy} := \mathbb{1}_{E\setminus A}(x)p_{xy} + \mathbb{1}_A\delta_{xy}$ ($\tilde{p}$ ist die Übergangsmatrix von $\tilde{X}_n := X_{n\wedge\tau_A}$, der in A absorbierten Kette). Wir treffen folgende *Annahme:*

$$\forall\, x \in E \setminus A, y \in E \,:\, \exists\, n \in \mathbb{N}_0 \,:\, \tilde{p}^n_{xy} > 0 \quad \text{und} \quad \inf_{x\in E} \mathbf{P}_x(\tau_A < \infty) = 1 \tag{Irr$_A$}$$

Satz 6.61 (Maximumprinzip)**.** *Es gelte (Irr$_A$). Sei $f\colon E \to \mathbb{R}$ 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 $y \in \{z \in E \mid \tilde{p}^n_{x^*z} > 0\}$. Nach Annahme ist $\cup_{n=1}^\infty A_n = E$. □

Satz 6.62. *Unter der Annahme (Irr$_A$) ist für jede beschränkte Funktion $g\colon A \to \mathbb{R}$ das Dirichlet-Problem mit Randwerten g eindeutig lösbar durch $f(x) = \mathbf{E}_x[g(X_{\tau_A})]$.*

Beweis. Seien f_1 und f_2 Lösungen. Da g beschränkt ist, sind gemäß Satz 6.61 (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 6.63 (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

$$\mathbf{P}_x(\tau_{\{a\}} < \tau_{\{b\}}) = \mathbf{E}_x[g(\tau_A)] = \frac{\left(\frac{1-r}{r}\right)^b - \left(\frac{1-r}{r}\right)^x}{\left(\frac{1-r}{r}\right)^b - \left(\frac{1-r}{r}\right)^a}.$$

6.4.6 Beispiel Gibbs-Sampler und Isingmodell

Das Ising-Modell[2] ist ein einfaches thermodynamisches und quantenmechanisches Modell für (Ferro-) Magnetismus von Kristallen.

[2]Ernst Ising, 1924

- 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.
- 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*)

$$Z_\beta := \sum_{y \in \{\pm 1\}^\Lambda} \exp\big(-\beta H(x)\big).$$

μ_β ist ein Wahrscheinlichkeitsmaß auf $\{\pm 1\}^\Lambda$, die *Gibbs*[3]*-Verteilung* (oft auch *Boltzmann*[4]*-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 $(i, j) \in \Lambda$, $x \in \{\pm 1\}^\Lambda$ sei $x^{(i,j),+} \in \{\pm 1\}^\Lambda$ gegeben durch

$$x^{(i,j),+}_{(i',j')} = \begin{cases} +1, & (i', j') = (i, j) \\ x_{(i',j')} & (i', j') \neq (i, j) \end{cases}$$

[3] Josiah Willard Gibbs (1839-1903)
[4] Ludwig Boltzmann, 1844-1906

und analog $x^{(i,j),-}$. Definiere $(A(x,y))_{x,y\in\{\pm 1\}^\Lambda}$ durch

$$\begin{aligned}
A(x, x^{(i,j),+}) &:= \frac{1}{|\Lambda|} \frac{\mu_\beta(x^{(i,j),+})}{\mu_\beta(x^{(i,j),+}) + \mu_\beta(x^{(i,j),-})}, \\
A(x, x^{(i,j),-}) &:= \frac{1}{|\Lambda|} \frac{\mu_\beta(x^{(i,j),-})}{\mu_\beta(x^{(i,j),+}) + \mu_\beta(x^{(i,j),-})}, \\
A(x,y) &:= 0, \qquad \text{falls } y \notin \cup_{(i,j)} \{x^{(i,j),+}, x^{(i,j),-}\}
\end{aligned}$$

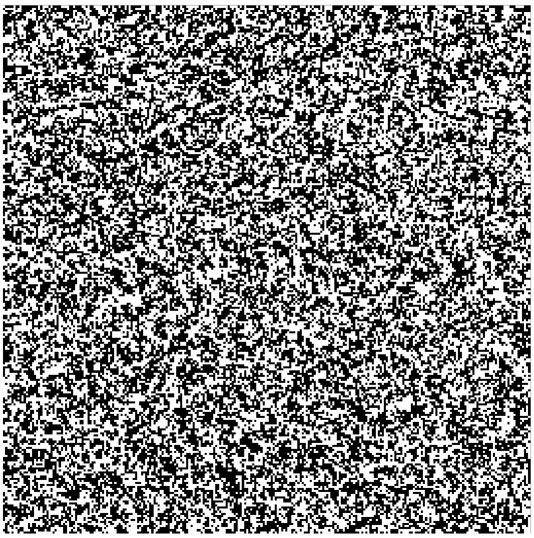
Beachte: A ist eine (irreduzible und aperiodische) stochastische Matrix und man braucht Z_β nicht zu kennen, um A zu bestimmen. Interpretation von A: Wähle zufällig einen Gitterpunkt, flippe den Spin dort gemäß μ_β, bedingt auf alle anderen Spins. Es gilt

$$\mu_\beta(y) A(y,z) = \mu_\beta(z) A(z,y) \quad \text{für alle } y, z \in \{\pm 1\}^\Lambda$$

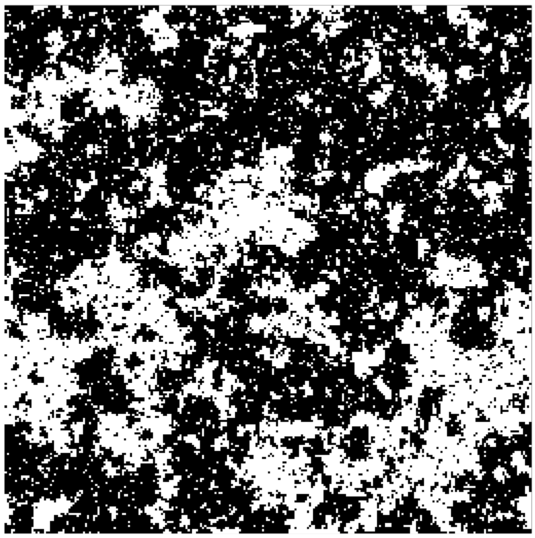
(es genügt hier, dies für $y = x^{(i,j),+}$, $z = x^{(i,j),-}$ für alle $x \in \{\pm 1\}^\Lambda$, $(i,j) \in \Lambda$ zu prüfen) und somit

$$\sum_z \mu_\beta(z) A(z,y) = \mu_\beta(y) \quad \text{für alle } y \in \{\pm 1\}^\Lambda.$$

μ_β ist ein (reversibles) Gleichgewicht für A.



(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.

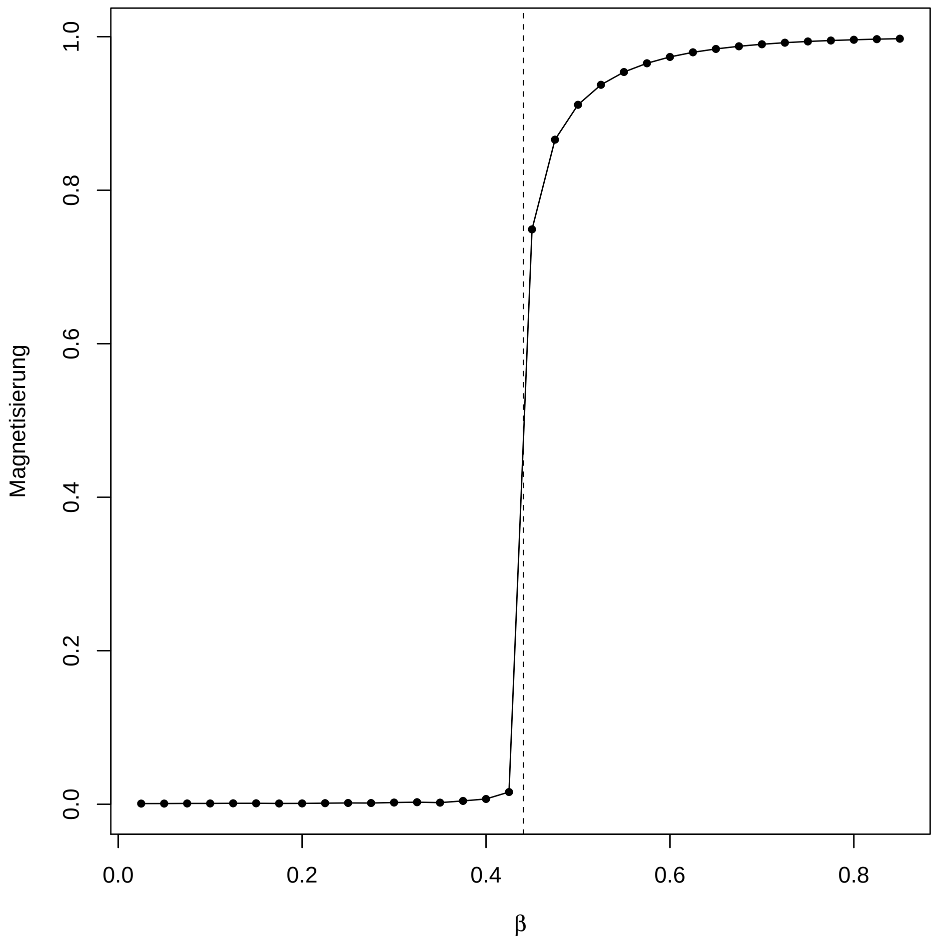


Abbildung 6.2: mittlere (absolute) Magnetisierung als Funktion von β (basierend auf einer Simulation für $L = 1000$)

7 (Etwas) Ergodentheorie

Zur Etymologie des Worts „ergodisch“ vgl. auch G. Gallavotti, J. Stat. Phys. 78, 1571-1589 (1995).

Definition 7.1. *Ein stochastischer Prozess* $(X_n)_{n\in\mathbb{N}_0}$ *mit Werten auf einem polnischen Raum* E *heißt* stationär, *wenn für alle* $m \in \mathbb{N}$ *gilt*

$$\mathbf{P}((X_n)_{n\in\mathbb{N}_0} \in \cdot\,) = \mathbf{P}((X_{n+m})_{n\in\mathbb{N}_0} \in \cdot\,).$$

Beispiel 7.2. i) Sind $X_1, X_2, X_3, \ldots$ 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 $m \in \mathbb{N}_0$, $c_0, c_1, \ldots, c_m \in \mathbb{R}$ 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 7.3. *Sei* $(\Omega, \mathcal{A}, \mathbf{P})$ *ein Wahrscheinlichkeitsraum und* $\tau\colon \Omega \to \Omega$ *messbar.*

i) $A \in \mathcal{A}$ *heißt* invariant, *falls* $\tau^{-1}(A) = A$.

$A \in \mathcal{A}$ *heißt* quasi-invariant, *falls* $\mathbf{P}(\tau^{-1}(A) \triangle A) = 0$.

$\mathcal{I} := \{A \in \mathcal{A} \mid A \text{ invariant}\}$ *heißt* σ-Algebra der invarianten Ereignisse *(auch* invariante σ-Algebra*).*

ii) τ *heißt* maßtreu *(auch* maßerhaltend*), falls* $\mathbf{P}(\tau^{-1}(A)) = \mathbf{P}(A)$ *für alle* $A \in \mathcal{A}$.

$(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ *heißt dann* (maßerhaltendes) dynamisches System.

iii) $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ *heißt* ergodisch, *wenn* $\mathcal{I}$ $\mathbf{P}$*-trivial ist, das heißt* $\mathbf{P}(A) \in \{0,1\}$ *für alle* $A \in \mathcal{I}$.

Bemerkung 7.4. i) Sei $f\colon \Omega \to \mathbb{R}$ messbar. f ist genau dann $\mathcal{I}$-messbar, wenn $f = f \circ \tau$.

ii) Ist $A \in \mathcal{I}$ mit $\mathbf{P}(A) \in (0,1)$, so ist $(A, \mathcal{A}_{|A}, \mathbf{P}(\,\cdot \mid A), \tau_{|A})$ ein maßerhaltendes dynamisches System ($\mathcal{A}_{|A}$ bezeichnet dabei die Spur-σ-Algebra über A).

Beispiel 7.5 (Rotation des Einheitskreises)**.** Sei $\Omega = [0,1)$ (mit „periodischem Rand"), $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbf{P}$ das Lebesgue-Maß auf Ω und $\tau_r \colon x \mapsto x + r \bmod 1$ (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 $\mathbf{P}(A) = \frac{1}{2} \notin \{0,1\}$, also ist $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ 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 $\mathbf{P}(A) > 0$. 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 $\mathbf{P}(A) = 1$. Andernfalls benutze den Dichtesatz von Lebesgue:

Bericht 7.6 (Dichtesatz von Lebesgue)**.** Sei λ das Lebesgue-Maß auf $\mathbb{R}$ und $A \in \mathcal{B}(\mathbb{R})$ 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

$$\lambda\left(\left\{x \in \mathcal{A} \mid \liminf_{\varepsilon \searrow 0} \frac{\lambda(A \cap B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} < 1\right\}\right) = 0.$$

Es gibt ein $x \in A$ mit $\frac{\mathbf{P}(A \cap B_\varepsilon(x))}{\mathbf{P}(B_\varepsilon(x))} > \frac{3}{4}$ für $0 < \varepsilon < \varepsilon_0$. Falls $A^\mathsf{c} = \varnothing$, so ist $\mathbf{P}(A) = 1$. Andernfalls sei $y \in A^\mathsf{c}$. Zu $\varepsilon > 0$ gibt es ein $n \in \mathbb{N}$ mit $B_\varepsilon(\tau_r^n(x)) \subset B_{2\varepsilon}(y)$, demnach

$$\begin{aligned}\frac{\mathbf{P}(A^\mathsf{c} \cap B_{2\varepsilon}(y))}{\mathbf{P}(B_{2\varepsilon}(y))} &= \underbrace{\frac{\mathbf{P}(A^\mathsf{c} \cap \tau_r^n(B_\varepsilon(x)))}{4\varepsilon}}_{= \frac{1}{4\varepsilon}\mathbf{P}(\tau_r^n(A^\mathsf{c} \cap B_\varepsilon(x)))} + \underbrace{\frac{\mathbf{P}(A^\mathsf{c} \cap (B_{2\varepsilon}(y) \setminus \tau_r^n(B_\varepsilon(x))))}{4\varepsilon}}_{\leq \frac{1}{4\varepsilon}\mathbf{P}(B_{2\varepsilon}(y) \setminus \tau_r^n(B_\varepsilon(x))) = \frac{2\varepsilon}{4\varepsilon} = \frac{1}{2}} \\ &\leq \frac{1}{4\varepsilon} \underbrace{\mathbf{P}(A^\mathsf{c} \cap B_\varepsilon(x))}_{\leq \frac{1}{4}\mathbf{P}(B_\varepsilon(x))} + \frac{1}{2} \leq \frac{1}{4\varepsilon}\frac{2\varepsilon}{4} + \frac{2\varepsilon}{4\varepsilon} = \frac{5}{8}.\end{aligned}$$

Damit gilt $\liminf_{\varepsilon \searrow 0} \frac{\mathbf{P}(A^\mathsf{c} \cap B_\varepsilon(y))}{\mathbf{P}(B_\varepsilon(y))} < 1$ für jedes $y \in A^\mathsf{c}$, wonach mit dem Dichtesatz von Lebesgue $\mathbf{P}(A^\mathsf{c}) = 0$ folgt. In diesem Fall ist $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ somit ergodisch.

Beispiel und Definition 7.7. Sei $X = (X_n)_{n\in\mathbb{N}_0}$ ein stochastischer Prozess mit Werten in E, realisiert als kanonischer Prozess auf $\Omega = E^{\mathbb{N}_0}$, $\mathcal{A} = \mathcal{B}(E)^{\otimes\mathbb{N}_0}$, $\mathbf{P} = \mathcal{L}((X_n)_{n\in\mathbb{N}_0})$ mit $X_i\colon\Omega \to E$, $(x_0, x_1, \ldots) \mapsto x_i$ und $\tau\colon\Omega \to \Omega$, $(x_0, x_1, \ldots) \mapsto (x_1, x_2, \ldots)$. Insbesondere gilt $X_n(\omega) = X_0(\tau^n(\omega))$. X ist genau dann stationär, wenn $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ein maßerhaltendes dynamisches System ist. Ein stationärer stochastischer Prozess X heißt *ergodisch*, wenn dies für $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ gilt.

Beispiel 7.8. i) Sind $X_1, X_2, \ldots$ unabhängig und identisch verteilt, so ist $(X_n)_n$ ergodisch.

ii) Aperiodische, irreduzible Markov-Ketten mit invarianter Verteilung sind ergodisch.

Satz 7.9 (Ergodensatz)**.** *Sei $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ 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 $f \in \mathcal{L}^p(\mathbf{P})$ für ein $p \geq 1$, so gilt*

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{j=0}^{n-1} f \circ \tau^j \xrightarrow[n\to\infty]{} \mathbf{E}[X_0 \mid \mathcal{I}] \quad \textit{f.s. und in } \mathcal{L}^p.$$

Die f.s.-Konvergenz in Satz 7.9 heißt traditionell „Birkhoffs[1] individueller Ergodensatz“, die $\mathcal{L}^p$-Konvergenz „von Neumanns[2] statistischer Ergodensatz.“

Lemma 7.10 (Hopfs[3] Maximal-Ergodenlemma)**.** *In der Situation von Satz 7.9 sei $X_0 \in \mathcal{L}^1$ und $M_n := \max\{0, S_1, S_2, \ldots, S_n\}$. Dann gilt $\mathbf{E}[X_0 \mathbb{1}_{\{M_n>0\}}] \geq 0$.*

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 $X_0 \geq \max\{S_1, \ldots, S_n\} - M_n \circ \tau$. 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

$$\begin{aligned}
\mathbf{E}\big[X_0 \mathbb{1}_{\{M_n>0\}}\big] &\geq \mathbf{E}\big[(\max\{S_1, \ldots, S_n\} - M_n \circ \tau)\mathbb{1}_{\{M_n>0\}}\big] \\
&= \mathbf{E}\big[(M_n - M_n \circ \tau)\mathbb{1}_{\{M_n>0\}}\big] \\
&\geq \mathbf{E}[M_n - M_n \circ \tau] \\
&= \mathbf{E}[M_n] - \mathbf{E}[M_n] \\
&= 0.
\end{aligned}$$

□

Beweis von Satz 7.9. Sei ohne Einschränkung $\mathbf{E}[X_0 \mid \mathcal{I}] = 0$ (ansonsten betrachte $\tilde{X}_n := X_n - \mathbf{E}[X_n \mid \mathcal{I}] = X_n - \mathbf{E}[X_0 \mid \mathcal{I}]$). Sei $Z := \limsup_n \frac{1}{n} S_n$, $\varepsilon > 0$ und $F := \{Z > \varepsilon\} \in \mathcal{I}$. Weiter

[1] George David Birkhoff 1884-1944; 1931
[2] John von Neumann 1903-1957; 1931
[3] Eberhard Hopf 1902-1983

sei $X_n^\varepsilon := (X_n - \varepsilon)\mathbb{1}_F$, $S_n^\varepsilon := \sum_{k=0}^{n-1} X_k^\varepsilon$ und $M_n^\varepsilon := \max\{0, S_1^\varepsilon, \ldots, S_n^\varepsilon\}$. Für $F_n := \{M_n^\varepsilon > 0\}$ gilt $F_1 \subset F_2 \subset \ldots$ 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 7.10 und monotoner Konvergenz

$$0 \leq \mathbf{E}[X_0^\varepsilon \mathbb{1}_{F_n}] \xrightarrow[n\to\infty]{} \mathbf{E}[X_0^\varepsilon \mathbb{1}_F] = \mathbf{E}[X_0^\varepsilon].$$

Demnach gilt

$$0 \leq \mathbf{E}[X_0^\varepsilon] = \mathbf{E}[(X_0 - \varepsilon)\mathbb{1}_F] = \mathbf{E}\big[\mathbf{E}[(X_0 - \varepsilon)\mathbb{1}_F \mid \mathcal{I}]\big] = \mathbf{E}\big[\mathbb{1}_F(\mathbf{E}[X_0 \mid \mathcal{I}] - \varepsilon)\big] = -\varepsilon \mathbf{P}(F),$$

das heißt $\mathbf{P}(F) = 0$ und mit $\varepsilon \searrow 0$ folgt $\limsup_n \frac{1}{n} S_n \leq 0$ f.s. Ersetze nun X_n durch $-X_n$ und erhalte $\frac{1}{n} S_n \to 0$, $n \to \infty$ f.s.

Sei nun $p \geq 1$ 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 1.26 eine monoton wachsende, konvexe Funktion $\varphi\colon [0,\infty) \to [0,\infty)$ mit $\frac{\varphi(x)}{x} \to \infty$, $x \to \infty$ und $\mathbf{E}[\varphi(|X_0|^p)] < \infty$. Nach der Jensen-Ungleichung gilt $|\frac{1}{n} S_n|^p \leq \frac{1}{n} \sum_{j=0}^{n-1} |X_j|^p$, also folgt wegen der Monotonie und Konvexität von φ

$$\varphi\left(|\frac{1}{n} S_n|^p\right) \leq \varphi\left(\frac{1}{n} \sum_{j=0}^{n-1} |X_j|^p\right) \leq \frac{1}{n} \sum_{j=0}^{n-1} \varphi(|X_j|^p)$$

und somit

$$\sup_n \mathbf{E}\left[\varphi\left(|\frac{1}{n} S_n|^p\right)\right] \leq \sup_n \frac{1}{n} n \mathbf{E}[\varphi(|X_0|^p] = \mathbf{E}[\varphi(|X_0|^p] < \infty.$$

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 auch $\frac{1}{n} S_n \to 0$, $n \to \infty$ in $\mathcal{L}^p$. □

Beispiel 7.11. Sei X eine positiv rekurrente, irreduzible Markov-kette auf einer abzählbaren Menge E. Sei π die eindeutige invariante Verteilung und $\mathbf{P}_\pi := \sum_{x \in E} \pi(\{x\}) \mathbf{P}_x$. Dann ist X ein stationärer Prozess (zum Beispiel auf $\Omega = E^{\mathbb{N}_0}$ mit $\tau\colon (x_n)_n \mapsto (x_{n+1})_n$) und $\big(\Omega, (2^E)^{\otimes \mathbb{N}_0}, \mathbf{P}_\pi, \tau\big)$ ist ergodisch.

Beweis. Sei $A \in \mathcal{I} \subset \mathcal{T} = \bigcap_n \sigma(X_n, X_n + 1, \ldots)$. Es gilt $\mathbf{P}_\pi(X \in A \mid \mathcal{F}_\eta) = \mathbf{P}_{X_\eta}(X \in A)$ für jede f.s. endliche Stoppzeit η, denn für $B \in \mathcal{F}_\eta$ gilt

$$\begin{aligned}\mathbf{E}_\pi[\mathbb{1}_{\{X\in B\}}\mathbb{1}_{\{X\in A\}}] &= \sum_{n=0}^{\infty}\sum_{x\in E}\mathbf{P}_\pi(X\in B, \eta = n, X_n = x, \underbrace{X\in A}_{=X\circ\tau^n\in A})\\ &= \sum_{n=0}^{\infty}\sum_{x\in E}\mathbf{P}_\pi(X\in B, \eta = n, X_n = x)\mathbf{P}_x(X\in A)\\ &= \mathbf{E}_\pi\big[\mathbb{1}_{\{X\in B\}}\mathbf{P}_{X_\eta}(X\in A)\big].\end{aligned}$$

Insbesondere ist mit $T_x^{(1)} \coloneqq \inf\{n > 0 \mid X_n = x\}$ ($< \infty$ f.s., da X irreduzibel und rekurrent) und der Markov-Eigenschaft

$$\mathbf{P}_\pi(X\in A) = \mathbf{E}_\pi\big[\mathbf{P}_\pi(X\in A\mid \mathcal{F}_{T_x^{(1)}})\big] = \mathbf{E}_\pi\big[\mathbf{P}_{X_{T_x^{(1)}}}(X\in A)\big] = \mathbf{E}_\pi\big[\mathbf{P}_x(X\in A)\big] = \mathbf{P}_x(X\in A)$$

für jedes $x \in E$. Somit ist

$$\begin{aligned}\mathbf{P}_\pi(X\in A) &= \mathbf{P}_{X_n}(X\in A)\\ &= \mathbf{P}_\pi(\underbrace{X\in A}_{=X\circ\tau^n\in A}\mid X_0,\dots,X_n) \xrightarrow[n\to\infty]{} \mathbf{P}_\pi(X\in A\mid\sigma(X_0,X_1,\dots)) = \mathbb{1}_{\{X\in A\}}.\end{aligned}$$

Für beliebiges $A \in \mathcal{I}$ gilt demnach $\mathbf{P}_\pi(X\in A)\in\{0,1\}$, das heißt $\big(\Omega,(2^E)^{\otimes\mathbb{N}_0},\mathbf{P}_\pi,\tau\big)$ ist ergodisch. □

Bemerkung 7.12. Insbesondere gilt für jedes $f \in \mathcal{L}^1(\pi)$

$$\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \xrightarrow[n\to\infty]{} \mathbf{E}_\pi(f(X_0)) = \sum_{x\in E}\pi(\{x\})f(x) \quad \text{f.s.}$$

Definition 7.13. *Ein maßerhaltendes dynamisches System* $(\Omega,\mathcal{A},\mathbf{P},\tau)$ *heißt* mischend*, wenn für alle* $A, B \in A$ *gilt*

$$\lim_{n\to\infty}\mathbf{P}(A\cap\tau^{-n}(B)) = \mathbf{P}(A)\mathbf{P}(B).$$

Ein stochastischer Prozess $X = (X_n)_{n\in\mathbb{N}_0}$ *heißt* mischend*, wenn dies für seine Darstellung als kanonischer Prozess auf* $E^{\mathbb{N}_0}$ *gilt, das heißt*

$$\lim_{n\to\infty}\mathbf{P}((X_m)_{m\in\mathbb{N}_0}\in A, (X_{m+n})_{m\in\mathbb{N}_0}\in B) = \mathbf{P}(X\in A)\mathbf{P}(X\in B).$$

Beobachtung 7.14. Ist $(\Omega,\mathcal{A},\mathbf{P},\tau)$ mischend, so ist es auch ergodisch.

Beweis. Sei $(\Omega,\mathcal{A},\mathbf{P},\tau)$ mischend und $A \in \mathcal{I}$. Für alle $n \in \mathbb{N}$ gilt $\mathbf{P}(A) = \mathbf{P}(A\cap\tau^{-n}(A))$, demnach gilt

$$\mathbf{P}(A) = \mathbf{P}(A \cap \tau^{-n}(A)) \xrightarrow[n\to\infty]{} \mathbf{P}(A)\mathbf{P}(A) = \mathbf{P}(A)^2$$

und somit $\mathbf{P}(A) \in \{0,1\}$. □

Bemerkung 7.15. Die irrationale Rotation aus Beispiel 7.5 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 $\liminf_n \mathbf{P}(A \cap \tau^{-n}(A)) = 0 \neq \frac{1}{16} = \mathbf{P}(A)^2$.

Satz 7.16. *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 $\tilde{A} \in E^{\{0,1,\ldots,N\}}$, sodass mit $A_\varepsilon := \tilde{A} \times E^{\mathbb{N}\setminus\{0,1,\ldots,N\}}$ gilt $\mathbf{P}_\pi(A \triangle A_\varepsilon) < \varepsilon$. Weiter gilt

$$\begin{aligned}\mathbf{P}_\pi(A_\varepsilon \cap \tau^{-n}(B)) &= \mathbf{P}_\pi((X_0,\ldots,X_N) \in \tilde{A}, (X_{m+n})_{m\in\mathbb{N}_0} \in B)\\ &= \sum_{x,y\in E} \mathbf{E}_\pi\big[\mathbb{1}_{A_\varepsilon}\mathbb{1}_{\{X_N=x\}}\mathbb{1}_{\{X_n=y\}}\mathbb{1}_B(X_n, X_{n+1},\ldots)\big]\\ &= \sum_{x,y\in E} \mathbf{E}_\pi\big[\mathbb{1}_{A_\varepsilon}\mathbb{1}_{\{X_N=x\}}\big]p_{xy}^{n-N}\mathbf{P}_y(X \in B).\end{aligned}$$

Nach Satz 6.53 gilt $p_{xy}^{n-N} \to \pi(\{y\}),\ n \to \infty$, demnach ist

$$\lim_{n\to\infty} \mathbf{P}_\pi(A_\varepsilon \cap \tau^{-n}(B)) = \sum_{x,y\in E} \mathbf{E}_\pi\big[\mathbb{1}_{A_\varepsilon}\mathbb{1}_{\{X_N=x\}}\big]\pi(\{y\})\mathbf{P}_y(X \in B) = \mathbf{P}_\pi(A_\varepsilon)\mathbf{P}_\pi(B).$$

Wegen $|\mathbf{P}_\pi(A_\varepsilon \cap \tau^{-n}(B)) - \mathbf{P}_\pi(A \cap \tau^{-n}(B))| \le \mathbf{P}_\pi(A_\varepsilon \triangle A) < \varepsilon$ folgt

$$\mathbf{P}_\pi(A_\varepsilon)\mathbf{P}_\pi(B) - \varepsilon \le \liminf_n \mathbf{P}_\pi(A \cap \tau^{-n}(B)) \le \limsup_n \mathbf{P}_\pi(A \cap \tau^{-n}(B)) \le \mathbf{P}_\pi(A_\varepsilon)\mathbf{P}_\pi(B) + \varepsilon$$

und $\mathbf{P}_\pi(A_\varepsilon) \to \mathbf{P}_\pi(A)$ für $\varepsilon \searrow 0$, also folgt mit $\varepsilon \searrow 0$ die Behauptung. □

Satz 7.17. *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 $R_n := |\{S_1,\ldots,S_n\}|$ (die „Größe des Range“) und $A = \{S_n \neq 0\ \forall\ n \in \mathbb{N}\}$, dann gilt*

$$\frac{1}{n}R_n \xrightarrow[n\to\infty]{} \mathbf{P}(A \mid \mathcal{I}) \quad \textit{f.s.}$$

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

$$R_n = \left|\left\{1 \le k \le n \mid S_l \ne S_k \;\; \forall \;\, l \in \{k+1, \dots, n\}\right\}\right| \ge |\{1 \le k \le n \mid S_l \ne S_k \;\; \forall \;\, l > k\}| = \sum_{k=1}^{n} \mathbb{1}_A \circ \tau^k.$$

Nach Satz 7.9 folgt demnach

$$\liminf_{n\to\infty} \frac{1}{n} R_n \ge \liminf_{n\to\infty} \sum_{k=1}^{n} \mathbb{1}_A \circ \tau^k = \mathbf{P}(A \mid \mathcal{I}).$$

Sei $A_m := \{S_l \ne 0 \;\, \forall \;\, l = 1, \dots, m\}$ für $m < n$, dann gilt

$$R_n \le m + \left|\left\{k \le n - m \mid S_l \ne S_k \;\; \forall \;\, l \in \{k+1, \dots, k+m\}\right\}\right| = \sum_{k=1}^{n-m} \mathbb{1}_{A_m} \circ \tau^k,$$

somit folgt wieder nach Satz 7.9

$$\limsup_{n\to\infty} \frac{1}{n} R_n \le 0 + \mathbf{P}(A_m \mid \mathcal{I}).$$

Es gilt $A_m \searrow A$, das heißt $\mathbf{P}(A \mid \mathcal{I}) = \lim_{m\to\infty} \mathbf{P}(A_m \mid \mathcal{I})$, also folgt mit $m \to \infty$ die Behauptung.

$\square$

8 Brownsche Bewegung

Definition 8.1. *Ein stochastischer Prozess* $(B_t)_{t\geq 0}$ *mit Werten in* $\mathbb{R}$ *heißt* Brownsche[1] Bewegung, *wenn gilt:*

i) Für alle $n \in \mathbb{N}$ *und* $0 = t_0 < \ldots < t_n$ *sind* $B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ *unabhängig und es gilt* $B_{t_i} - B_{t_{i-1}} \sim \mathbf{N}(0, t_i - t_{i-1})$ *für alle* $i = 1, \ldots, n$.

ii) $t \mapsto B_t$ *ist f.s. stetig.*

Bemerkung 8.2. Einen Prozess, der $i)$ erfüllt, haben wir in Beispiel 6.19 konstruiert (als kanonischen Prozess auf $\mathbb{R}^{[0,\infty)}$).

Satz 8.3. *Die Brownsche Bewegung im Sinne von Definition 8.1 existiert.*

Beweis via Lévy-Konstruktion der Brownschen Bewegung. Betrachte zunächst nur $t \in [0,1]$ und $B_0 = 0$. Sei $\mathcal{D}_n := \left\{\frac{k}{2^n} \mid k \in \mathbb{N}_0, k \leq 2^n\right\}$ und $\mathcal{D} := \bigcup_{n=1}^{\infty} \mathcal{D}_n$. Für $t \in \mathcal{D}$ seien $Z_t \sim \mathbf{N}(0,1)$ 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 $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$

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

Es gilt (nach Konstruktion)

$$\{B_d \mid d \in \mathcal{D}_n\} \text{ und } \{Z_t \mid t \in \mathcal{D} \setminus \mathcal{D}_n\} \text{ sind unabhängig.} \tag{8.1}$$

Zeige:

$$B(d) - B(d - 2^{-n}),\ d \in \mathcal{D}_n \setminus \{0\} \text{ sind u.i.v. mit } B(d) - B(d - 2^{-n}) \sim \mathbf{N}(0, 2^{-n}) \tag{8.2}$$

durch Induktion über n. Für $n = 0$ ist die Behauptung klar. Sei also (8.2) für $n - 1$ erfüllt. Betrachte $d = \frac{k}{2^n}$ für k ungerade, das heißt $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$. Setze

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

[1] Robert Brown (1773-1858)

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

Nach (8.1) 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

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

Demnach ist

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

das heißt (8.2) gilt auch für n.

Setze nun

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

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 $d \in \mathcal{D}_n$

$$B(d) = \sum_{i=0}^{n} F_i(d) = \sum_{i=0}^{\infty} F_i(d) \tag{8.3}$$

durch Induktion über n. Für $n = 0$ ist die Aussage klar. Sei also (8.3) für $n-1$ erfüllt. Für $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ gilt nach Induktionsvoraussetzung

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

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 (8.3) gilt auch für n.

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

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

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

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

Somit existiert nach Borel-Cantelli ein (zufälliges) N_0, sodass für alle $n \geq N_0$ gilt

$$\|F_n\|_\infty := \sup_{t \in [0,1]} |F_n(t)| \leq c\sqrt{n} 2^{-\frac{n+1}{2}},$$

das heißt $\sum_{n=0}^{\infty} \|F_n\|_\infty < \infty$. Demnach ist $B(t) = \sum_{n=0}^{\infty} F_n(t),\ t \in [0,1]$ als Grenzwert einer gleichmäßig konvergenten Reihe von stetigen Funktionen selbst stetig.

Wegen (8.2) gilt für $t_0 < \ldots < t_n \in \mathcal{D}$

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

Der Fall allgemeiner t_i folgt durch Approximation mit $t'_{i_k} \in \mathcal{D}$ (vgl. [Per10]).

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

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

□

Beobachtung 8.4. $(B_t)_{t \geq 0}$ ist genau dann eine (Standard-) Brownsche Bewegung (das heißt eine Brownsche Bewegung mit $B_0 = 0$), wenn $(B_t)_{t \geq 0}$ ein zentrierter Gaußscher Prozess mit $\mathbf{Cov}[B_s, B_t] = s \wedge t$ und stetigen Pfaden ist.

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

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

Die Umkehrung folgt aus der Tatsache, dass die endlich dimensionalen Verteilungen eines (zentrierten) Gaußschen Prozesses durch die Kovarianzen festgelegt sind (vgl. Beispiel 4.5).

□

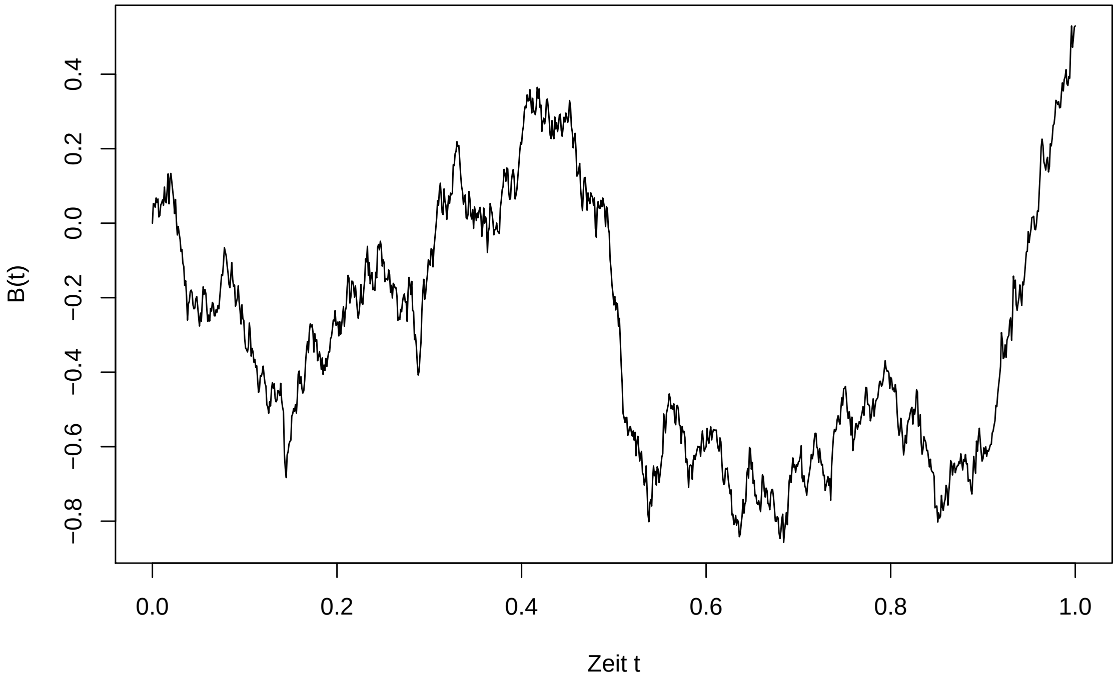


Abbildung 8.1: Simulation einer (Standard-) Brownschen Bewegung mit Gitterweite $\frac{1}{N}$, $N = 1000$

Korollar 8.5. *Ist $(B_t)_{t\geq 0}$ eine (Standard-) Brownsche Bewegung und $c \neq 0$, so ist auch $\widetilde{B}(t) := \frac{1}{c}B_{c^2 t},\ t \geq 0$ eine (Standard-) Brownsche Bewegung.*

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

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

□

Beobachtung 8.6. Für alle $\varepsilon > 0$ gilt $B_\varepsilon \overset{\mathrm{d}}{=} \sqrt{\varepsilon} B_1$, was nahe legt, dass für kleines h die „typische“ Größe von $B_{t+h} - B_t \approx \sqrt{h}$ ist. Genauere Auskunft gibt der folgende Satz.

Satz 8.7. *Es gibt ein $c < \infty$ und ein (zufälliges) $h_0 > 0$, sodass*

$$|B(t+h) - B(t)| \le c\sqrt{h \log \frac{1}{h}}$$

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

Beweis. Schreibe $B(t) = \sum_{n=0}^{\infty} F_n(t)$ wie im Beweis von Satz 8.3. F_n ist differenzierbar bis auf endlich viele „Knickstellen", es gilt

$$\left\|F_n'\right\|_\infty \le \frac{\|F_n\|_\infty}{2^{-n}} \le c\sqrt{n} 2^{\frac{n-1}{2}}$$

für alle $n \ge N_0$ mit N_0 aus dem Beweis von Satz 8.3. Damit folgt für $l > N_0$

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

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

$$S_2 \le c_2' \sqrt{l 2^l} \le c_2'' \sqrt{\frac{1}{h} \log \frac{1}{h}}$$

und

$$S_3 \le c_3' \sqrt{l 2^{-l}} \le c_3'' \sqrt{h \log \frac{1}{h}}$$

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

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

□

Bericht 8.8 (Lévys Stetigkeitsmodul der Brownschen Bewegung)**.** Es gilt

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

Literaturverzeichnis

[Bre68] Leo Breiman. *Probability.* Wiley, 1968.

[Dep14] Andrej Depperschmidt. Stochastik 1. Vorlesungsskript, 2014.

[Fel71] William Feller. *An Introduction to Probability Theory and Its Applications, Vol. 2.* Wiley, 2. edition, 1971.

[Kal02] Olav Kallenberg. *Foundations of Modern Probability.* Springer New York, 2. edition, 2002.

[Kle13] Achim Klenke. *Wahrscheinlichkeitstheorie.* Springer Spektrum, 3. edition, 2013.

[Per10] Peter Mörters; Yuval Peres. *Brownian Motion.* Cambridge University Press, 1. edition, 2010.

[Wil00] Leonard C.G. Rogers; David Williams. *Diffusions, Markov processes and martingales, Vol. 1.* Cambridge University Press, 2. edition, 2000.