Stochastische Prozesse

Wintersemester 2019/20

Mathias Trabs*
Universität Hamburg

17. Oktober 2019

Es handelt sich hierbei um eine vorläufige Version des Vorlesungsskriptes, welche im Laufe des Semesters überarbeitet wird. Verbesserungsvorschläge und insbesondere Fehler, die Sie im Skript finden, nehme ich dankend entgegen.

Inhaltsverzeichnis

*Email: mathias.trabs@uni-hamburg.de

1 Allgemeine Theorie stochastischer Prozesse

In Ihrer bisherigen Stochastikausbildung haben Sie bereits die mathematische Modellierung und Analyse von Zufallsgrößen kennengelernt. Im Rahmen dieser Vorlesung werden wir die stochastische Modellierung um eine zeitliche Komponente ergänzen. Wir werden also zeitlich abhängige, zufällige Phänomene studieren, wie sie in zahlreichen Anwendungen in den Natur-, Sozial- und Wirtschaftswissenschaften auftauchen.

1.1 Grundbegriffe und Beispiele

Zunächst sollten wir definieren, was wir unter einem stochastischen Prozess verstehen und ein paar Grundbegriffe einführen, die anschließend in Beispielen illustriert werden.

Definition 1.1. Eine Familie $X = (X_t, t \in T)$ von Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt stochastischer Prozess. Wir sprechen von diskreter Zeit, falls $T = \mathbb{N}_0$, und von stetiger Zeit, falls $T = [0, \infty)$. Nehmen alle X_t Werte in $(S, \mathcal{S})$ an, so heißt $(S, \mathcal{S})$ Zustandsraum von X. Für jedes $\omega \in \Omega$ heißt die Abbildung

$$T \ni t \mapsto X_t(\omega) \in S$$

Pfad, Trajektorie oder Realisierung von X. Die Wahrscheinlichkeitsverteilungen der Vektoren $(X_{t_1}, \dots, X_{t_n})$ für $t_1, \dots, t_n \in T, n \in \mathbb{N}$ nennen wir endlichdimensionale Verteilungen von X. Wir schreiben $\mathbb{P}_{\{t_1,\dots,t_n\}} = \mathbb{P}^{(X_{t_1},\dots,X_{t_n})}$ (man beachte, dass trotz dieser Notation die Reihenfolge der Zeitpunkte eine Rolle spielt).

Statt wie in Ihrer bisherigen Stochastikausbildung Zufallszahlen oder Zufallsvektoren zu betrachten, beschäftigen wir uns hier also mit zufälligen Funktionen.

Lemma 1.2. *Für einen stochastischen Prozess $X = (X_t, t \in T)$ mit Zustandsraum $(S, \mathcal{S})$ ist die Abbildung $\overline{X} \colon \Omega \to S^T$ mit $\overline{X}(\omega)(t) := X_t(\omega)$ eine $(S^T, \mathcal{S}^{\otimes T})$-wertige Zufallsvariable.*

Beweis. Wir müssen die Messbarkeit von $\overline{X}$ zeigen. Da $\mathcal{S}^{\otimes T}$ von den Projektionen $\pi_t \colon S^T \to S$ auf die t. Koordinate für $t \in T$ erzeugt wird, genügt es, die Messbarkeit der Verknüpfungen $\pi_t \circ \overline{X} \colon \Omega \to S$ für alle $t \in T$ zu zeigen. Diese folgt, da $\pi_t \circ \overline{X}(\omega) = X_t(\omega), \omega \in \Omega, t \in T$, eine Zufallsvariable also messbar ist. □

Bemerkung 1.3. Wir werden auch kleinere Funktionenräume als S^T betrachten, wie z.B. $C([0,1])$ statt $\mathbb{R}^{[0,1]}$ (Übung ☑). In diesem Fall kann es deutlich schwerer sein, $\overline{X}$ als Zufallsvariable im entsprechenden Funktionenraum zu etablieren.

Eine erste wichtige Klasse von stochastischen Prozessen kennen Sie bereits aus den „Maßtheoretischen Konzepten der Stochastik“: die Martingale. Sie dienen der Formalisierung von fairen Spielen und sind bspw. in der Finanzmathematik zentral, um den Handel mit Aktien an (vollständigen) Märkten zu modellieren. Zur Erinnerung (und Verallgemeinerung auf stetige Zeit):

Definition 1.4. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $I \subseteq \mathbb{R}$ und $(\mathcal{F}_t)_{t \in I}$ eine Filtration (also eine Familie von σ-Algebren mit $\mathcal{F}_s \subseteq \mathcal{F}_t$ für $s \leqslant t$). Ein stochastischer Prozess $X = (X_t, t \in I)$ heißt Martingal (bzw. Submartingal oder Supermartingal) bzgl. $(\mathcal{F}_t)_{t \in I}$, falls gilt:

(i) $X_t \in L^1(\mathbb{P})$ für alle $t \in I$,

(ii) X_t ist $\mathcal{F}_t$-messbar für jedes $t \in I$ (X ist adaptiert an $(\mathcal{F}_t)_{t \in I}$) und

(iii) $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$ $\mathbb{P}$-f.s. für alle $s, t \in I, t \geqslant s$ (bzw. $\mathbb{E}[X_t | \mathcal{F}_s] \geqslant X_s$ für Sub- oder $\mathbb{E}[X_t | \mathcal{F}_s] \leqslant X_s$ für Supermartingale).

Beispiel 1.5.

(i) Sind $(X_i)_{i\in\mathbb{N}}$ unabhängige, integrierbare Zufallsvariablen mit $\mathbb{E}[X_i]=0$ (bzw. $\geqslant 0$ oder $\leqslant 0$) für alle $i\in\mathbb{N}$, so ist $(S_n)_{n\geqslant 0}$ mit $S_n := \sum_{i=1}^n X_i, n\in\mathbb{N}$, und $S_0 := 0$ ein Martingal (bzw. Sub- oder Supermartingal) bzgl. seiner natürlichen Filtration $\mathcal{F}_n := \sigma(S_m : m\leqslant n)$. Im Spezialfall $\mathbb{P}(X_i=1)=\mathbb{P}(X_i=-1)=\frac{1}{2}$ sprechen wir von einer *einfachen, symmetrischen Irrfahrt.*

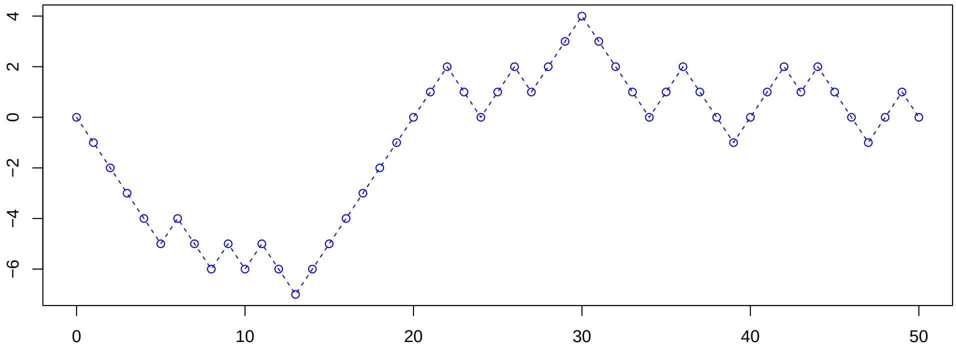

Abbildung 1: Eine Realisierung der einfachen symmetrische Irrfahrt auf $\{0,1,\dots,50\}$.

(ii) Wir betrachten zwei Vermögenswerte $(S_t^{(0)}, S_t^{(1)})_{t\in\mathbb{N}_0}$, wobei $S_t^{(0)} = (1+r)^t$ eine risikolose Anleihe mit fester Zinsrate $r>0$ ist und $(S_t^{(1)})_{t\in\mathbb{N}_0}$ ein zufälliger Aktienkurs auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in\mathbb{N}_0},\mathbb{P})$. Der *Fundamentalsatz der Preistheorie* besagt, dass dieser Aktienmarkt genau dann fair (d.h. arbitrage-frei) ist, wenn der diskontierte Preisprozess $X_t = S_t^{(1)}/S_t^{(0)}$ ein Martingal auf $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in\mathbb{N}_0},\mathbb{Q})$ für ein sogenanntes äquivalentes Martingalmaß $\mathbb{Q}\sim\mathbb{P}$ ist. Für den erwarteten Gewinn (bzw. Verlust) zur Zeit $t+1$ auf Grundlage aller Information bis zur Zeit t gilt dann $\mathbb{E}_{\mathbb{Q}}[X_{t+1}-X_t|\mathcal{F}_t]=0$ (das Kapital bleibt also im Durchschnitt konstant). Ein Submartingal wäre für den Investor vorteilhaft, wohingegen ein Supermartingal im Mittel zu Verlusten führt.

Eine weitere wichtige Klasse sind Markovprozesse. Während wir uns später allgemeine Zustandsräume ansehen werden, beschränken wir uns hier auf den abzählbaren Fall.

Definition 1.6. Sei $T=\mathbb{N}_0$ oder $T=[0,\infty)$ und S ein abzählbarer (Zustands-)Raum. Dann besitzt eine Familie $X=(X_t, t\in T)$ von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},\mathbb{P})$ mit Werten in $(S,\mathscr{P}(S))$ die (schwache) Markoveigenschaft, falls für alle $n\in\mathbb{N}, t_1<t_2<\dots<t_{n+1}, s_1,\dots,s_{n+1}\in S$ mit $\mathbb{P}(X_{t_1}=s_1,\dots,X_{t_n}=s_n)>0$ gilt:

$$\mathbb{P}\big(X_{t_{n+1}}=s_{n+1}\big|X_{t_1}=s_1,\dots,X_{t_n}=s_n\big)=\mathbb{P}\big(X_{t_{n+1}}=s_{n+1}\big|X_{t_n}=s_n\big).$$

Auf geeigneten Wahrscheinlichkeitsräumen werden wir stochastische Prozesse, welche die Markoveigenschaft erfüllen, Markovprozesse bzw., im Fall $T=\mathbb{N}_0$, Markovketten nennen.

Beispiel 1.7. Die Irrfahrt $S=(S_n, n\in\mathbb{N}_0)$ mit $S_n := \sum_{k=1}^n X_k, n\geqslant 1$, und $S_0 := 0$ für X_k i.i.d. $\mathbb{Z}$-wertigen Zufallsvariablen ist ein stochastischer Prozess mit $T=\mathbb{N}_0$ und Zustandsraum $\mathbb{Z}$. S erfüllt die Markoveigenschaft aufgrund der Unabhängigkeit der Inkremente $S_{t_k}-S_{t_{k-1}} = \sum_{n=t_{k-1}+1}^{t_k} X_n$ für $n\in\mathbb{N}, t_1<t_2<\dots<t_{n+1}$: Für $s_0=0$ und $s_1,\dots,s_{n+1}\in S$ gilt

$$\begin{aligned}\mathbb{P}\big(S_{t_{n+1}}=s_{n+1}\big|S_{t_1}=s_1,\dots,S_{t_n}=s_n\big) &= \frac{\mathbb{P}\big(S_{t_1}=s_1,\dots,S_{t_{n+1}}=s_{n+1}\big)}{\mathbb{P}\big(S_{t_1}=s_1,\dots,S_{t_n}=s_n\big)}\\ &= \frac{\prod_{k=1}^{n+1}\mathbb{P}(S_{t_k}-S_{t_{k-1}}=s_k-s_{k-1})}{\prod_{k=1}^{n}\mathbb{P}(S_{t_k}-S_{t_{k-1}}=s_k-s_{k-1})}\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}(S_{t_{n+1}} - S_{t_n} = s_{n+1} - s_n) \\
&= \frac{\mathbb{P}(S_{t_n} = s_n, S_{t_{n+1}} - S_{t_n} = s_{n+1} - s_n)}{\mathbb{P}(S_{t_n} = s_n)} \\
&= \mathbb{P}(S_{t_{n+1}} = s_{n+1} | S_{t_n} = s_n).
\end{aligned}
$$

Definition 1.8. Für einen Prozess X mit der schwachen Markoveigenschaft und $t_1 \leqslant t_2$, $i, j \in S$, sind die Übergangswahrscheinlichkeiten vom Zustand i zum Zeitpunkt t_1 in den Zustand j zur Zeit t_2 definiert als

$$
p_{ij}(t_1, t_2) := \begin{cases} \mathbb{P}(X_{t_2} = j | X_{t_1} = i), & \text{falls } \mathbb{P}(X_{t_1} = i) > 0, \\ 0, & \text{sonst.} \end{cases}
$$

Die Übergangsmatrix ist gegeben durch

$$
P(t_1, t_2) := \big(p_{ij}(t_1, t_2)\big)_{i,j \in S}.
$$

Die Übergangsmatrix und der zugehörige (Markov-)Prozess heißen (zeit-)homogen, falls $P(t_1, t_2) = P(0, t_2 - t_1) =: P(t_2 - t_1)$ für alle $0 \leqslant t_1 \leqslant t_2$ gilt.

Satz 1.9. *Ist $P(t_1, t_2)$ die Übergangsmatrix eines stochastischen Prozesses mit der Markoveigenschaft auf abzählbarem Zustandsraum, so gilt die Chapman-Kolmogorov-Gleichung*

$$
P(t_1, t_3) = P(t_1, t_2) P(t_2, t_3) \quad \textit{für alle } t_1 \leqslant t_2 \leqslant t_3
$$

(Matrixmultiplikation). Im zeitlich homogenen Fall ergibt sich die Halbgruppeneigenschaft

$$
P(t + s) = P(s) P(t) \quad \textit{für alle } s, t \geqslant 0
$$

und insbesondere $P(n) = P(1)^n$ für $n \in \mathbb{N}$.

Beweis. Nach Definition ergibt sich aus der Markoveigenschaft für alle $i, j \in S$:

$$
\begin{aligned}
P(t_1, t_3)_{ij} &= \mathbb{P}(X_{t_3} = j | X_{t_1} = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{t_3} = j, X_{t_2} = k | X_{t_1} = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{t_3} = j | X_{t_2} = k, X_{t_1} = i) \mathbb{P}(X_{t_2} = k | X_{t_1} = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{t_3} = j | X_{t_2} = k) \mathbb{P}(X_{t_2} = k | X_{t_1} = i) \\
&= \sum_{k \in S} P(t_2, t_3)_{kj} P(t_1, t_2)_{ik} \\
&= \big(P(t_1, t_2) P(t_2, t_3)\big)_{i,j}.
\end{aligned}
$$

Im zeitlich homogenen Falle reduziert sich diese Gleichheit auf $P(t_3 - t_1) = P(t_2 - t_1) P(t_3 - t_2)$, was die Behauptung für $s = t_2 - t_1$ und $t = t_3 - t_2$ impliziert. □

Für einen Markovprozess $X = (X_t, t \in T)$ mit diskretem Zustandsraum S haben die endlichdimensionalen Verteilungen eine besonders einfache Form, denn deren Zähldichte ist für Zeitpunkte $0 = t_0 < t_1 \leqslant \ldots \leqslant t_n$ und Zustände $j_0, \ldots, j_n \in S$ mit $\mathbb{P}_{j_0}(X_{t_1} = j_1, \ldots, X_{t_n} = j_n) > 0$ gegeben durch

$$
\begin{aligned}
\mathbb{P}^{(X_{t_0}, \ldots, X_{t_n})}(\{j_0, \ldots, j_n\}) &= \mathbb{P}(X_{t_0} = j_0, \ldots, X_{t_n} = j_n) \\
&= \mathbb{P}(X_0 = j_0) \prod_{i=1}^{n} \mathbb{P}(X_{t_i} = j_i | X_{t_0} = j_0, \ldots, X_{t_{i-1}} = j_{i-1})
\end{aligned}
$$

$$= \mathbb{P}(X_0 = j_0) \prod_{i=1}^{n} \mathbb{P}_{j_0}(X_{t_i} = j_i | X_{t_{i-1}} = j_{i-1})$$
$$= \mathbb{P}(X_0 = j_0) \prod_{i=1}^{n} P(t_{i-1}, t_i)_{j_{i-1} j_i}.$$

Insbesondere beschreiben die Übergangsmatrizen die endlichdimensionalen Verteilungen eindeutig.

Beispiel 1.10 (Ehrenfest-Modell)**.** Wir betrachten folgendes sehr einfaches Modell zum Stoffaustausch zwischen zwei verbunden Behältern: Es gibt insgesamt N Partikel verteilt auf beide Behälter. Zu jedem Zeitpunkt $n \in \mathbb{N}$ (z.B. in Minuten) wandert ein zufällig gleichverteilt ausgewähltes Teilchen in den jeweils anderen Behälter. Wir setzen diesen Vorgang beliebig lange fort und bezeichnen mit X_n die Partikelanzahl im ersten Behälter zum Zeitpunkt n. Es ergeben sich folgende Übergangswahrscheinlichkeiten

$$\begin{aligned}
\mathbb{P}(X_{n+1} = i+1 | X_n = i) &= \frac{N-i}{N}, \quad i \in \{0, \dots, N-1\}, \\
\mathbb{P}(X_{n+1} = i-1 | X_n = i) &= \frac{i}{N}, \qquad i \in \{1, \dots, N\}, \\
\mathbb{P}(X_{n+1} = j | X_n = i) &= 0, \quad i = 0, \dots, N, j \notin \{i-1, i+1\}.
\end{aligned}$$

Die zugehörige Ein-Schritt-Übergangsmatrix ist

$$P(1) = \begin{pmatrix} 0 & \frac{N}{N} & & & 0 \\ \frac{1}{N} & & \frac{N-1}{N} & & \\ & \frac{2}{N} & \ddots & \ddots & \\ & & \ddots & & \frac{1}{N} \\ 0 & & & \frac{N}{N} & 0 \end{pmatrix}.$$

Wir haben also für jeden Anfangswert X_0 eine positive Wahrscheinlichkeit, jede andere Verteilung der Partikel irgendwann in der Zukunft zu erreichen. Frage: Gibt es ein asymptotisches Verhalten? Antwort: Ergodentheorie (Kapitel 4).

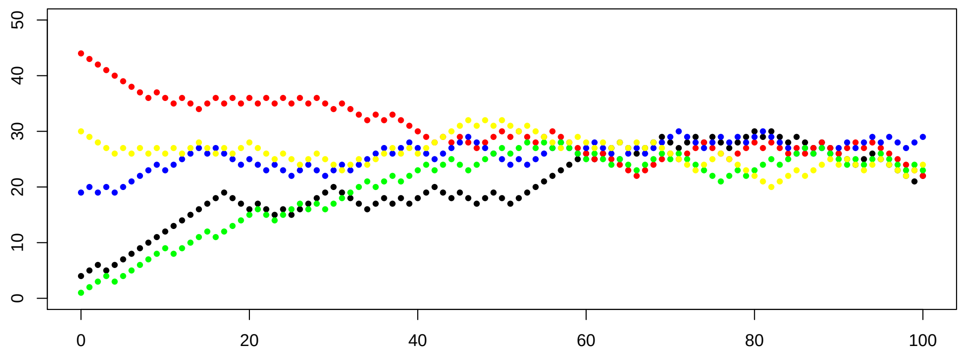

Abbildung 2: Fünf Trajektorien des Ehrenfestmodels mit $N = 50$ und uniform auf $\{0, \dots, 50\}$ verteiltem Startwert.

Unsere bisherigen konkreten Beispiele stochastischer Prozesse in den Klassen der Martingale und der Markovprozesse waren alle in diskreter Zeit definiert. Zum Abschluss dieses ersten Ausblicks soll ein zeitstetiger Prozess von grundlegender Bedeutung studiert werden. Wir beginnen mit einer physikalischen Motivation:

Beispiel 1.11. Wir zählen die Anzahl N_t der Klicks eines Geigerzählers (=Anzahl der Emissionen einer radioaktiven Substanz) im Zeitintervall $[0, t]$. Die Werte N_t sind zufällig und werden durch einen Zählprozess modelliert.

Definition 1.12. Seien $(S_k)_{k\geqslant 1}$ $\mathbb{R}_+$-wertige Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $0 \leqslant S_1(\omega) \leqslant S_2(\omega) \leqslant \dots$ für alle $\omega \in \Omega$. Dann heißt $N = (N_t, t \geqslant 0)$ mit $N_0 = 0$ und

$$N_t := \sum_{k\geqslant 1} \mathbb{1}_{\{S_k \leqslant t\}}, \quad t > 0,$$

Zählprozess mit Sprungzeiten (S_k).

Beispiel 1.11 (Fortsetzung). Die Anzahl der Klicks in einem kurzen Zeitintervall $[t, t+h]$ sollte entweder 0 oder 1 und die Wahrscheinlichkeit eines Klicks sollte (approximativ) proportional zu h sein. Außerdem sollten die Klickanzahlen in den Intervallen $(0, t_1], \dots, (t_{n-1}, t_n]$ für Zeiten $0 < t_1 < \dots < t_n$ unabhängig sein und ihre Verteilung sollte nur von der Intervalllänge abhängen.

Definition 1.13. Ein Zählprozess N heißt Poissonprozess mit Intensität $\lambda > 0$, falls

(a) $\mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)$ für $h \downarrow 0$;

(b) $\mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h)$ für $h \downarrow 0$;

(c) $(N_{t_i} - N_{t_{i-1}})_{1\leqslant i\leqslant n}$ sind für beliebige Zeitpunkte $0 = t_0 < t_1 < \dots < t_n$ unabhängig (unabhängige Inkremente);

(d) $N_t - N_s \stackrel{d}{=} N_{t-s}$ für alle $t \geqslant s \geqslant 0$ (stationäre Inkremente)

Wir verwenden hier folgenden *Notationen:*

- $X \stackrel{d}{=} Y$ für Zufallsvariablen X, Y heißt, dass X und Y die gleiche Verteilung haben: $P^X = P^Y$.
- $A(h) = o(h)$ für $h \downarrow 0$ heißt $\lim_{h\downarrow 0} \frac{A(h)}{h} = 0$.

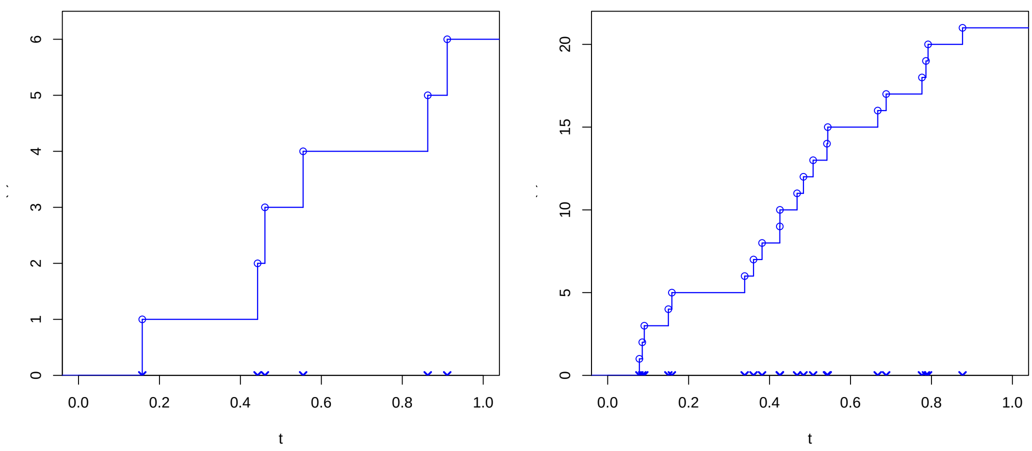

Abbildung 3: Zufällige Pfade eines Poissonprozesses mit $\lambda = 5$ (links) und $\lambda = 20$ (rechts) sowie markierte Sprungstellen auf der horizontalen Achse.

Satz 1.14. *Für einen Zählprozess N mit Sprungzeiten (S_k) sind folgende Eigenschaften äquivalent:*

(i) N *ist ein Poissonprozess mit Intensität* $\lambda > 0$.

(ii) N *erfüllt die Eigenschaften (c) und (d) eines Poissonprozesses und es gilt* $N_t \sim Poiss(\lambda t)$ *für alle* $t > 0$.

(iii) $T_1 := S_1, T_k := S_k - S_{k-1}, k \geqslant 2$, *sind unabhängig und* $Exp(\lambda)$*-verteilt.*

(iv) $N_t \sim Poiss(\lambda t)$ *gilt für alle* $t > 0$ *und die Verteilung von* $(S_1, \dots, S_n)$ *gegeben* $\{N_t = n\}$ *besitzt die Dichte*

$$f(x_1, \dots, x_n) = \frac{n!}{t^n} \mathbb{1}_{\{0 \leqslant x_1 \leqslant \dots \leqslant x_n \leqslant t\}}. \tag{1}$$

Bemerkung 1.15.

- Wir schlussfolgern die Poisson-Verteilung aus den Bedingung (a) und (b) ohne weitere Verteilungsannahmen.
- Die Existenz des Poissonprozesses folgt aus (iii): Für unabhängige und $Exp(\lambda)$-verteilte $(T_k)_{k \in \mathbb{N}}$ ist $N_t = \max\{n \in \mathbb{N} : \sum_{k=1}^n T_k \leqslant t\}$ ein Poissonprozess.
- Die Dichte f aus (1) ist auch die Dichte der Ordnungsstatistiken $U_{(1)}, \dots, U_{(n)}$ von unabhängigen $\mathcal{U}([0,t])$-verteilten Zufallsvariablen $U_1, \dots, U_n$ (Übung ☑).
- Die beiden vorangegangen Bemerkungen liefern uns insbesondere zwei explizite Konstruktionsmöglichkeiten um Poissonprozesse zu simulieren, vgl. Abbildung 3.

Beweis. *(i)⇒(ii)* Wir setzen $p_n(t) := \mathbb{P}(N_t = n)$. Aus der Definition des Poissonprozesses folgt für $h > 0$

$$\begin{aligned} p_0(t+h) = \mathbb{P}(N_t = 0, N_{t+h} - N_t = 0) &= \mathbb{P}(N_t = 0)\mathbb{P}(N_h = 0) \\ &= p_0(t)p_0(h) = p_0(t)\big(1 - \lambda h + o(h)\big). \end{aligned}$$

Daraus folgen zum einen die (lokale) Lipschitzstetigkeit von $t \mapsto p_0(t)$ und zum anderen die rechtsseitigen (und damit dann auch die linksseitigen) Grenzwerte

$$\frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t) + \frac{o(h)}{h} \xrightarrow{h \downarrow 0} -\lambda p_0(t) = p_0'(t).$$

Zusammen mit $p_0(0) = 1$ ergibt sich $p_0(t) = e^{-\lambda t}$.

Für $n \geqslant 1$ erhalten wir analog

$$\begin{aligned} p_n(t+h) &= \mathbb{P}\big(\{N_{t+h} = n\} \cap (\{N_t \leqslant n-2\} \cup \{N_t = n-1\} \cup \{N_t = n\})\big) \\ &= o(h) + p_{n-1}(t)\big(\lambda h + o(h)\big) + p_n(t)\big(1 - \lambda h + o(h)\big) \end{aligned}$$

und somit

$$\begin{aligned} p_n'(t) &= \lim_{h \downarrow 0} \frac{p_n(t+h) - p_n(t)}{h} \\ &= \lim_{h \downarrow 0} \frac{o(h) + p_{n-1}(t)\big(\lambda h + o(h)\big) + p_n(t)\big(-\lambda h + o(h)\big)}{h} \\ &= \lambda p_{n-1}(t) - \lambda p_n(t). \end{aligned}$$

Wegen $p_n(0) = 0$, folgt induktiv $p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$ also $N_t \sim Poiss(\lambda t)$.

(ii) ⇒ *(iii)* Für $0 = b_0 \leqslant a_1 < b_1 \leqslant \dots \leqslant a_n < b_n$ berechnen wir

$$\mathbb{P}\Big(\bigcap_{k=1}^n \{a_k \leqslant S_k \leqslant b_k\}\Big)$$

$$
\begin{aligned}
&= \mathbb{P}\Big(\bigcap_{k=1}^{n-1}\{N_{a_k}-N_{b_{k-1}}=0, N_{b_k}-N_{a_k}=1\}\cap\{N_{a_n}-N_{b_{n-1}}=0, N_{b_n}-N_{a_n}\geqslant 1\}\Big)\\
&= \Big(\prod_{k=1}^{n-1}\mathbb{P}(N_{a_k-b_{k-1}}=0)\mathbb{P}(N_{b_k-a_k}=1)\Big)\mathbb{P}(N_{a_n-b_{n-1}}=0)\mathbb{P}(N_{b_n-a_n}\geqslant 1)\\
&= \Big(\prod_{k=1}^{n-1}\lambda(b_k-a_k)e^{-\lambda(b_k-a_k)-\lambda(a_k-b_{k-1})}\Big)e^{-\lambda(a_n-b_{n-1})}(1-e^{-\lambda(b_n-a_n)})\\
&= (e^{-\lambda a_n}-e^{-\lambda b_n})\lambda^{n-1}\prod_{k=1}^{n-1}(b_k-a_k)\\
&= \int_{a_1}^{b_1}\int_{a_2-x_1}^{b_2-x_1}\cdots\int_{a_n-x_1-\cdots-x_{n-1}}^{b_n-x_1-\cdots-x_{n-1}}\lambda^n e^{-\lambda(x_1+\cdots+x_n)}\mathrm{d}x_n\cdots\mathrm{d}x_1.
\end{aligned}
$$

Damit besitzt $(T_1,\dots,T_n)=(S_1,S_2-S_1,\dots,S_n-S_{n-1})$ die Dichte $\lambda^n e^{-\lambda(x_1+\cdots+x_n)}$ für $x_i\geqslant 0$. Da dies eine Produktdichte ist, folgern wir, dass T_i unabhängig und $Exp(\lambda)$ verteilt sind.

(iii) ⇒ (iv) Es gilt $\mathbb{P}(N_t=0)=\mathbb{P}(S_1>t)=e^{-\lambda t}$ und

$$\mathbb{P}(N_t=n)=\mathbb{P}(N_t\geqslant n)-\mathbb{P}(N_t\geqslant n+1)=\mathbb{P}(S_n\leqslant t)-\mathbb{P}(S_{n+1}\leqslant t).$$

Da $S_n=T_1+\cdots+T_n\sim\Gamma(\lambda,n)$, folgt

$$\mathbb{P}(N_t=n)=\int_0^t\Big(\frac{\lambda^n x^{n-1}}{(n-1)!}-\frac{\lambda^{n+1}x^n}{n!}\Big)e^{-\lambda x}\mathrm{d}x=\frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

und somit $N_t\sim Poiss(\lambda t)$. Nach dem Dichtetransformationssatz ist die Dichte von $(S_1,\dots,S_{n+1})$ für $s_1\leqslant s_2\leqslant\dots$ gegeben durch

$$f^{S_1,\dots,S_{n+1}}(s_1,\dots,s_{n+1})=\prod_{k=1}^{n+1}\lambda e^{-\lambda(s_k-s_{k-1})}=\lambda^{n+1}e^{-\lambda s_{n+1}}.$$

Da $\{N_t=n\}=\{S_n\leqslant t,S_{n+1}>t\}$, berechnen wir für $0\leqslant a_1<b_1\leqslant\dots\leqslant a_n<b_n\leqslant t$ die bedingte Wahrscheinlichkeit

$$
\begin{aligned}
\mathbb{P}\big(S_1\in[a_1,b_1],\dots,S_n\in[a_n,b_n]\big|N_t=n\big) &= \frac{\mathbb{P}\big(S_1\in[a_1,b_1],\dots,S_n\in[a_n,b_n],S_{n+1}>t\big)}{\frac{(\lambda t)^n}{n!}e^{-\lambda t}}\\
&= \int_{a_1}^{b_1}\cdots\int_{a_n}^{b_n}\frac{n!}{t^n}\mathbb{1}_{\{0\leqslant s_1\leqslant\cdots\leqslant s_n\leqslant t\}}\mathrm{d}s_n\cdots\mathrm{d}s_1,
\end{aligned}
$$

was den Integranden als die bedingte Dichte identifiziert.

(iv) ⇒ (i) Wir weisen die Eigenschaften des Poissonprozesses nach:

(c) Für $0=t_0<t_1<\cdots<t_n=t$ und $k_1,\dots,k_n\in\mathbb{N}$ betrachten wir mit $K:=\sum_{l=1}^n k_l$:

$$
\begin{aligned}
&\mathbb{P}\big(\forall l=1,\dots,n: N_{t_l}-N_{t_{l-1}}=k_l\big)\\
&\quad= \mathbb{P}(N_{t_n}=K)P\big(\forall l=1,\dots,n: N_{t_l}-N_{t_{l-1}}=k_l\big|N_{t_n}=K\big)\\
&\quad= \frac{(\lambda t_n)^K}{K!}e^{-\lambda t_n}P\big(S_{k_1}\leqslant t_1<S_{k_1+1},\dots,S_K\leqslant t_n<S_{K+1}\big|N_{t_n}=K\big)\\
&\quad= \frac{(\lambda t_n)^K}{K!}e^{-\lambda t_n}\frac{K!}{t_n^K}\prod_{l=1}^n\frac{(t_l-t_{l-1})^{k_l}}{k_l!}\\
&\quad= \prod_{l=1}^n\mathbb{P}(N_{t_l}-N_{t_{l-1}}=k_l).
\end{aligned}
$$

Damit sind $(N_{t_l} - N_{t_{l-1}})_l$ unabhängig.

(d) Für $t, h > 0$, $k \in \mathbb{N}_0$ und $K \geqslant k$ gilt die Verschiebungsinvarianz

$$\begin{aligned}\mathbb{P}(N_{t+h} - N_h = k | N_{t+h} = K) &= \mathbb{P}(N_h = K - k, N_{t+h} - N_h = k | N_{t+h} = K) \\ &= \frac{K!}{(t+h)^K} \frac{(t+h-h)^k}{k!} \cdot \frac{h^{K-k}}{(K-k)!} \\ &= \mathbb{P}(N_{t+h} - N_t = K - k, N_t = k | N_{t+h} = K) \\ &= \mathbb{P}(N_t = k | N_{t+h} = K).\end{aligned}$$

Durch Summieren über alle $K \geqslant k$ erhalten wir $N_{t+h} - N_h \stackrel{d}{=} N_t$, also Stationarität der Zuwächse.

(a)+(b) Weiterhin ergibt sich für $0 < h < 1$

$$p(h) := \mathbb{P}(N_h = 0) = \sum_{k=0}^{\infty} \mathbb{P}(N_1 = k)\mathbb{P}(N_1 - N_h = k | N_1 = k) = \sum_{k=0}^{\infty} \mathbb{P}(N_1 = k)(1-h)^k.$$

Da $\sum_{k \geqslant 0} \mathbb{P}(N_1 = k)k = \mathbb{E}[N_1] = \lambda < \infty$ ist $p(h)$ differenzierbar auf $[0,1]$ mit $p'(0) = -\lambda$. Wir erhalten

$$\mathbb{P}(N_h = 0) = \mathbb{P}(N_0 = 0) - \lambda h + o(h).$$

Analog folgt

$$\begin{aligned}\mathbb{P}(N_h = 1) &= \sum_{k=1}^{\infty} \mathbb{P}(N_1 = k)\mathbb{P}(N_1 - N_h = k - 1 | N_1 = k). \\ &= \sum_{k=1}^{\infty} \mathbb{P}(N_1 = k) k h (1-h)^{k-1},\end{aligned}$$

sodass $\mathbb{P}(N_h = 1) = \lambda h + o(h)$. □

Aufgrund seiner unabhängigen Inkremente ist ein Poissonprozess $N = (N_t, t \geqslant 0)$ auch ein Markovprozess in stetiger Zeit, vgl. Beispiel 1.7. Ein Martingal ist N allerdings nicht, denn $\mathbb{E}[N_t] = \lambda t \neq 0$ für jedes $t > 0$. Jedoch ist der zentrierte Prozess $M = (M_t, t \geqslant 0)$ mit

$$M_t := N_t - \lambda t$$

ein zeitstetiges Martingal, denn

$$\begin{aligned}\mathbb{E}[M_t - M_s | M_s] &= \mathbb{E}[N_t - N_s - (t-s)\lambda | N_s] \\ &= \mathbb{E}[N_t - N_s] - (t-s)\lambda \\ &= \mathbb{E}[N_{t-s}] - (t-s)\lambda = 0, \qquad \text{für alle } 0 \leqslant s < t.\end{aligned}$$

Definition 1.16. Zwei stochastische Prozesse $(X_t, t \in T)$ und $(Y_t, t \in T)$ auf $(\Omega, \mathcal{F}, \mathbb{P})$ heißen

(i) Versionen oder Modifikationen voneinander, falls $\forall t \in T : \mathbb{P}(X_t = Y_t) = 1$;

(ii) ununterscheidbar, falls $A \subseteq \{\forall t \in T : X_t = Y_t\}$ für ein $A \in \mathcal{F}$ mit $\mathbb{P}(A) = 1$ gilt.

Bemerkung 1.17. Ist X eine Version von Y, so haben X und Y die gleichen endlichdimensionalen Verteilungen: Für alle $t_1, \ldots, t_n \in T$ folgt aus $\mathbb{P}(X_{t_i} = Y_{t_i}) = 1, i = 1, \ldots, n$, dass $\mathbb{P}(X_{t_1} = Y_{t_1}, \ldots, X_{t_n} = Y_{t_n}) = 1$ also $(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (Y_{t_1}, \ldots, Y_{t_n})$. Offensichtlich sind ununterscheidbare Prozesse Versionen voneinander. Das Gegenteil gilt nicht, wie das nächste Beispiel zeigt.

Beispiel 1.18. Für ein $\lambda > 0$ seien $(T_k)_{k \geqslant 1}$ unabhängige *Exp*(λ)-verteilte Zufallsvariablen und $S_n := \sum_{k=1}^n T_k, n \geqslant 1$. Dann ist $N_t := \sum_{n \geqslant 1} \mathbb{1}_{\{S_n \leqslant t\}}$ ein Poissonprozess mit Intensität λ. Der Prozess $X_t := \sum_{n \geqslant 1} \mathbb{1}_{\{S_n < t\}}, t \geqslant 0$, hat die gleichen endlichdimensionalen Verteilungen wie N:

$$\mathbb{P}(N_{t_1} \in A_1, \ldots, N_{t_n} \in A_n) = \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n)$$

für Borel-Mengen $A_1, \dots, A_n$, denn $\mathbb{P}(\exists k \geqslant 1 : S_k = t_i) = 0$ für alle $i = 1, \dots, n$. Andererseits haben N und X völlig unterschiedliche Pfade:

$$\mathbb{P}\big(t \mapsto N_t \text{ ist rechts-stetig}\big) = 1, \quad \text{aber} \quad \mathbb{P}\big(t \mapsto X_t \text{ ist rechts-stetig}\big) = 0.$$

Da alle Sprungzeiten S_n stetig verteilt sind, gilt $\mathbb{P}(\exists n \geqslant 1 : S_n = t) = 0$ für alle $t > 0$. Folglich sind die Prozesse N und X Versionen voneinander. Sie sind jedoch nicht ununterscheidbar, denn für ein beliebiges $k \geqslant 1$ gilt

$$\mathbb{P}(\forall t \geqslant 0 : N_t = X_t) < \mathbb{P}(N_{S_k} = X_{S_k}) = 0.$$

Dieses Beispiel macht zudem deutlich, dass die endlichdimensionalen Verteilungen zwar die Verteilungseigenschaften eines Prozesses beschreiben, jedoch keine Pfadeigenschaften.

Definition 1.19. Ein stochastischer Prozess $X = (X_t, t \in T)$ heißt stetig, falls alle Trajektorien stetig sind. X heißt stochastisch stetig, falls $t_n \to t$ die stochastische Konvergenz $X_{t_n} \xrightarrow{\mathbb{P}} X_t$ impliziert.

Bemerkung 1.20. Jeder stetige Prozess ist auch stochastisch stetig, da fast sichere Konvergenz die stochastische Konvergenz impliziert. Die umgekehrte Richtung gilt nicht, da bspw. der Poissonprozess nicht stetig, aber stochastisch stetig ist:

$$\forall \varepsilon \in (0,1) : \lim_{t_n \to t} \mathbb{P}(|N_t - N_{t_n}| > \varepsilon) = \lim_{t_n \to t} (1 - e^{-\lambda |t - t_n|}) = 0.$$

1.2 Erweiterungssatz von Kolmogorov

Ziel dieses Abschnittes ist ein allgemeines Existenzresultat für stochastische Prozesse. Genauer wollen wir die Frage untersuchen, ob für eine vorgegebene Familie von endlichdimensionalen Verteilungen ein zugehöriger stochastischer Prozess existiert. Zunächst finden wir folgende notwendige Bedingung an die Verteilungsfamilie:

Lemma 1.21. *Es sei $(X_t, t \in T)$ ein stochastischer Prozess mit Zustandsraum $(S, \mathcal{S})$. Für $I \subseteq J \subseteq T$ definieren wir die Koordinatenprojektion*

$$\pi_{J,I} : \quad S^J \to S^I, \quad (s_j)_{j \in J} \mapsto (s_j)_{j \in I}.$$

Dann erfüllen die endlichdimensionalen Verteilungen von X folgende Konsistenzbedingung

$$\forall I \subseteq J \subseteq T \text{ mit } I, J \text{ endlich}, \forall A \in \mathcal{S}^{\otimes I} : \mathbb{P}_J(\pi_{J,I}^{-1}(A)) = \mathbb{P}_I(A).$$

Beweis. Wir schreiben für $A \in \mathcal{S}^{\otimes I}$

$$\begin{aligned} \mathbb{P}_I(A) &= \mathbb{P}\big((X_t)_{t \in I} \in A\big) = \mathbb{P}\big(\overline{X} \in \pi_{T,I}^{-1}(A)\big) \\ &= \mathbb{P}\big(\overline{X} \in (\pi_{J,I} \circ \pi_{T,J})^{-1}(A)\big) = \mathbb{P}\big((X_t)_{t \in J} \in \pi_{J,I}^{-1}(A)\big) = \mathbb{P}_J(\pi_{J,I}^{-1}(A)). \end{aligned}$$

□

Definition 1.22. Es sei $T \neq \emptyset$ eine Indexmenge und $(S, \mathcal{S})$ ein messbarer Raum. Für jede Teilmenge $J \subseteq T$ sei ein Wahrscheinlichkeitsmaß $\mathbb{P}_J$ auf dem Produktraum $(S^J, \mathcal{S}^{\otimes J})$ gegeben. $(\mathbb{P}_J)_{J \subseteq T}$ heißt projektive Familie, falls für alle endlichen $I \subseteq J \subseteq T$ gilt, dass

$$\forall A \in \mathcal{S}^{\otimes I} : \quad \mathbb{P}_I(A) = \mathbb{P}_J\big(\pi_{J,I}^{-1}(A)\big).$$

Es wird sich herausstellen, dass unter einer recht allgemeinen Strukturannahme an den Zustandsraum diese notwendige Bedingung bereits genügt, um für jede projektive Familie die Existenz eines zugehörigen Prozesses zu beweisen.

Definition 1.23. Ein metrischer Raum (S, d) heißt polnischer Raum, falls er separabel und vollständig ist. Als kanonische σ-Algebra auf S wählen wir die Borel-σ-Algebra $\mathscr{B}_S$.

Zur Erinnerung: S heißt vollständig, falls jede Cauchy-Folge konvergiert und S heißt separabel, falls es eine abzählbare, dichte Teilfolge in S gibt, d.h. jede nichtleere, offene Teilmenge aus S enthält mindestens ein Element dieser Folge.

Bemerkung 1.24. In polnischen Räumen existiert eine abzählbare Basis (der Topologie), d.h. es existiert eine abzählbare Folge offener Mengen derart, dass jede offene Teilmenge von S als die Vereinigung von Elementen dieser Folge dargestellt werden kann. Ist $(s_n)_{n\in\mathbb{N}}$ eine dichte Folge in S, so kann diese Basis als $\{B_\varepsilon(s_n) : n \in \mathbb{N}, \varepsilon \in \mathbb{Q}_+\}$ gewählt werden, wobei $B_\varepsilon(s_n) := \{y \in S : d(s_n, y) < \varepsilon\}$.

Beispiel 1.25.

(i) Der euklidische Raum $\mathbb{R}^d$ jeder Dimension $d \geqslant 1$ ist polnisch.

(ii) Für endlich viele polnische Räume $(S_k, d_k), k = 1, \ldots, n$ ist das Produkt $\times_{k=1}^n S_k$ versehen mit der Metrik $d((s_k), (t_k)) := \sum_{k=1}^n d_k(s_k, t_k)$ wieder polnisch.

(iii) Abgeschlossene Teilmengen A eines polnischen Raumes (S, d) sind wieder polnisch: Ist $(s_i)_{i\geqslant 1}$ eine dichte Teilmenge von S, so wählen wir für jedes $n \geqslant 1$ eine Abzählbare Menge $A_n \subseteq A$, sodass für alle s_i mit $A \cap \{s : d(s, s_i) \leqslant \frac{1}{n}\} \neq \emptyset$ ein $a_i \in A_n$ existiert mit $d(a_i, s_i) \leqslant \frac{1}{n}$. Dann ist $\bigcup_n A_n$ eine abzählbare, dichte Teilmenge von A.

(iv) Der Raum der stetigen Funktionen $C([a, b]) := \{f : [a, b] \to \mathbb{R}\}$ auf dem Intervall $[a, b], a, b \in \mathbb{R}$, ausgestattet mit der Supremumsnorm $\|f\|_\infty := \sup_{x\in[a,b]} |f(x)|$ ist vollständig (die Gleichmäßige Konvergenz garantiert die Stetigkeit) und separabel (Weierstraß' Theorem: die Polynome liegen dicht). Damit ist $(C([a, b]), \|\cdot\|_\infty)$ polnisch.

Lemma 1.26. *Seien $S_1, \ldots, S_n$ polnische Räume, dann gilt für die Borel-σ-Algebra des Produktraumes $\mathscr{B}_{\times_{k=1}^n S_k} = \bigotimes_{k=1}^n \mathscr{B}_{S_k}$.*

Beweis. $\bigotimes_k \mathscr{B}_{S_k}$ ist die kleinste σ-Algebra, sodass die Koordinatenprojektionen $\pi_i \colon \times_{k=1}^n S_k \to S_i, i \geqslant 1$, messbar sind. Andererseits ist die Produkttopologie die raueste Topologie, sodass alle π_i stetig sind. Folglich sind alle π_i $\mathscr{B}_{\times_k S_k}$-messbar, was $\mathscr{B}_{\times_k S_k} \supseteq \bigotimes_k \mathscr{B}_{S_k}$ impliziert.

Aufgrund der Separabilität kann jede offene Menge $O \subseteq \times_k S_k$ als abzählbare Vereinigung offener Mengen der Form $\bigcap_{i=1}^n \pi_i^{-1}(O_i)$ mit offenen Mengen O_i aus der (topologischen) Basis von S_i, dargestellt werden. Letztere sind Elemente von $\bigotimes_k \mathscr{B}_{S_k}$, sodass $\mathscr{B}_{\times_k S_k} \subseteq \bigotimes_k \mathscr{B}_{S_k}$. □

Der Beweis zeigt, dass die $\supseteq$-Relation für alle topologischen Räume und Produkte beliebiger Kardinalität gilt. Die $\subseteq$-Richtung kann jedoch bereits mit zwei nicht-polnischen Räumen fehlschlagen.

Definition 1.27. Ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf einem metrischen Raum $(S, \mathscr{B}_S)$ heißt

(i) <u>straff</u>, falls für alle $\varepsilon > 0$ eine kompakte Menge $K \subseteq S$ existiert, sodass $\mathbb{P}(K) > 1 - \varepsilon$.

(ii) <u>regulär</u>, falls für alle $\varepsilon > 0, B \in \mathscr{B}_S$ eine kompakte Menge $K \subseteq B$ existiert mit $\mathbb{P}(B \setminus K) \leqslant \varepsilon$ und eine offene Menge $O \supseteq B$ existiert mit $\mathbb{P}(O \setminus B) \leqslant \varepsilon$.

Lemma 1.28. *Jedes Wahrscheinlichkeitsmaß auf einem polnischen Raum ist straff.*

Beweis. Sei $(s_n)_{n\geqslant 1}$ eine dichte Folge in S. Wir betrachten für einen Radius $\rho > 0$ die abgeschlossenen Bälle $B_\rho(s_n)$ um s_n. Dann gilt $S = \bigcup_{n\geqslant 1} B_\rho(s_n)$ und die σ-Stetigkeit liefert

$$\lim_{N\to\infty} \mathbb{P}\Big(\bigcup_{n=1}^N B_\rho(s_n)\Big) = 1.$$

Für jedes $\varepsilon > 0$ und $\rho = 1/k$ gibt es also ein N_k, sodass

$$\mathbb{P}\Big(\bigcup_{n=1}^{N_k} B_{1/k}(s_n)\Big) \geqslant 1 - \varepsilon 2^{-k}.$$

Dann ist $K := \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{N_k} B_{1/k}(s_n)$ eine abgeschlossene Teilmenge für die

$$\mathbb{P}(S \setminus K) = \mathbb{P}\Big(\bigcup_{k=1}^{\infty} \Big(\bigcup_{n=1}^{N_k} B_{1/k}(s_n)\Big)^c\Big) \leqslant \sum_{k \geqslant 1} \varepsilon 2^{-k} = \varepsilon.$$

Es bleibt also zu zeigen, dass die abgeschlossene Menge K sogar kompakt ist. Da für jedes $\delta > 0$ eine endliche Überdeckung aus Bällen $B_{1/k}(s_n)$ mit Durchmesser kleiner als δ existiert, ist K total-beschränkt und abgeschlossen, also wegen der Vollständigkeit des metrischen Raumes auch kompakt. □

Lemma 1.29. *Jedes Wahrscheinlichkeitsmaß auf einem polnischen Raum (S, d) ist regulär.*

Beweis. Wir betrachten folgende Menge

$$\mathcal{D} := \Big\{B \in \mathscr{B}_S : \mathbb{P}(B) = \sup_{K \subseteq B \text{ kompakt}} \mathbb{P}(K) = \inf_{O \supseteq B \text{ offen}} \mathbb{P}(O)\Big\}.$$

Wir zeigen nun, dass jede abgeschlossene Menge F in $\mathcal{D}$ liegt. Aufgrund der Straffheit von $\mathbb{P}$ gibt es für jedes $\varepsilon > 0$ eine kompakte Menge K_ε, sodass $\mathbb{P}(K_\varepsilon) \geqslant 1 - \varepsilon$ gilt. Dann ist $F \cap K_\varepsilon \subseteq F$ ebenfalls eine kompakte Menge mit

$$\mathbb{P}\big(F \setminus (F \cap K_\varepsilon)\big) \leqslant \mathbb{P}(K_\varepsilon^c) \leqslant \varepsilon.$$

Daraus folgt $\mathbb{P}(F) = \sup\{\mathbb{P}(K) : K \subseteq F \text{ kompakt}\}$. Andererseits gilt $F = \bigcap_{n \geqslant 1} O_n$ für die offenen Mengen $O_n := \{s \in S : \inf_{x \in F} d(s, x) < 1/n\}$. Da $\mathbb{P}$ σ-stetig ist folgt

$$\mathbb{P}(F) = \inf_{N \geqslant 1} \mathbb{P}\Big(\underbrace{\bigcap_{n=1}^{N} O_n}_{\text{offen}}\Big).$$

Damit haben wir $F \in \mathcal{D}$ gezeigt.

Weiterhin ist $\mathcal{D}$ ein Dynkin-System (Übung ☑). Da die Menge der abgeschlossenen Mengen $\cap$-stabil ist und $\mathscr{B}_S$ erzeugt, folgt $\mathcal{D} = \mathscr{B}_S$. □

Nun sind wir bereit für das Hauptresultat dieses Abschnittes:

Satz 1.30 (Kolmogorovs Erweiterungssatz)**.** *Ist $(S, \mathscr{B}_S)$ ein polnischer Raum und T eine beliebige nichtleere Indexmenge, so existiert zu jeder projektiven Familie $(\mathbb{P}_J)_{J \subseteq T \text{endlich}}$ von Wahrscheinlichkeitsmaßen über S^J genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf dem Produktraum $(S^T, \mathscr{B}_S^{\otimes T})$, sodass für alle endlichen $J \subseteq T$ und $B \in \mathscr{B}_S^{\otimes J}$*

$$\mathbb{P}_J(B) = \mathbb{P}(\pi_{T,J}^{-1}(B))$$

gilt. Das Maß $\mathbb{P}$ wird auch <u>projektiver Limes</u> der Familie $(\mathbb{P}_J)$ genannt.

Beweis. Wir betrachten die Algebra der Zylindermengen

$$\mathcal{A} := \bigcup_{J \subseteq T \text{ endlich}} \pi_{T,J}^{-1}(\mathscr{B}_S^{\otimes J}).$$

Da $\mathcal{A}$ $\cap$-stabil ist und $\mathscr{B}_S^{\otimes T} = \sigma(\mathcal{A})$ gilt, wird $\mathbb{P}$ eindeutig durch seine Werte auf $\mathcal{A}$ festgelegt.

Die Existenz von $\mathbb{P}$ folgt aus Caratheodorys Fortsetzungssatz, falls $\mathbb{P}$ auf $\mathcal{A}$ ein Prämaß ist. Die Konsistenzbedingung an $(\mathbb{P}_J)$ garantiert, dass $\mathbb{P}$ auf $\mathcal{A}$ wohldefiniert und additiv ist: Für disjunkte Mengen $A, B \in \mathcal{A}$ gibt es ein endliches $J \subseteq T$ und $A', B' \in \mathscr{B}_S^{\otimes J}$, sodass $A = \pi_{T,J}^{-1}(A')$ und $B = \pi_{T,J}^{-1}(B')$. Da $\mathbb{P}_J$ ein Maß ist, folgt

$$\mathbb{P}(A \cup B) = \mathbb{P}_J(A' \cup B') = \mathbb{P}_J(A') + \mathbb{P}_J(B') = \mathbb{P}(A) + \mathbb{P}(B).$$

Offensichtlich gilt außerdem $\mathbb{P}(S^T) = \mathbb{P}_J(S^J) = 1$ für jedes endliche $J \subseteq T$. Es bleibt, die σ-Additivität von $\mathbb{P}$ auf $\mathcal{A}$ zu zeigen oder äquivalent die Stetigkeit in $\emptyset$: Für jede monoton fallende Folge $(B_n)_{n\geqslant 1} \subseteq \mathcal{A}$ mit $B_n \downarrow \emptyset$ muss $\mathbb{P}(B_n) \to 0$ gelten.

O.B.d.A. können wir $B_n = \pi_{T,J_n}^{-1}(A_n)$ für endliche $J_n \subseteq J_{n+1} \subseteq T$ und $A_n \in \mathscr{B}_S^{\otimes J_n}$ schreiben. Sei $\varepsilon > 0$. Aufgrund der Regularität von $\mathbb{P}_{J_n}$ gibt es für alle $n \in \mathbb{N}$ kompakte Mengen $K_n \subseteq A_n$ mit $\mathbb{P}_{J_n}(A_n \setminus K_n) \leqslant \varepsilon 2^{-n}$. Dann ist auch

$$K_n' = \bigcap_{l=1}^{n-1} \pi_{J_n,J_l}^{-1}(K_l) \cap K_n \subseteq S^{J_n}$$

als abgeschlossene Teilmenge einer kompakten Menge kompakt. Für $C_n := \pi_{T,J_n}^{-1}(K_n') = \bigcap_{l=1}^n \pi_{T,J_l}^{-1}(K_l) \subseteq B_n$ gilt dann ebenfalls $C_n \downarrow \emptyset$. Wir werden abschließend zeigen, dass ein $n_0 \in \mathbb{N}$ mit $C_{n_0} = \emptyset$ existiert. Daraus folgt

$$\limsup_{n\to\infty} \mathbb{P}(B_n) \leqslant \mathbb{P}(B_{n_0}) = \mathbb{P}(B_{n_0} \setminus C_{n_0}) \leqslant \sum_{l=1}^{n_0} \mathbb{P}_{J_l}(A_l \setminus K_l) \leqslant \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, zeigt dies $\mathbb{P}(B_n) \to 0$.

Um die Existenz von n_0 zu zeigen, führen wir einen Widerspruchsbeweis und nehmen an, dass für alle $n \geqslant 1$ ein $y_n \in C_n$ existiert. Da $K_n' \subseteq S^{J_n}$ kompakt ist, existiert eine Teilfolge $(n_l^{(1)})_l$, sodass $(\pi_{T,J_1}(y_{n_l^{(1)}}))_{l\geqslant 1}$ in K_1' konvergiert. Für eine weitere Teilteilfolge $(n_l^{(2)})_l$ konvergiert $(\pi_{T,J_2}(y_{n_l^{(2)}}))_{l\geqslant 1}$ in K_2' und so weiter. Aufgrund der Abzählbarkeit von $J := \bigcup_{n\geqslant 1} J_n = \{i_1, i_2, \dots\}$, können wir S^J mit der Produktmetrik $d((s_i)_{i\in J}, (t_i)_{i\in J}) := \sum_{k\geqslant 1} 2^{-k}(d(s_{i_k}, t_{i_k}) \wedge 1)$ versehen. Da die Diagonalfolge $(\pi_{T,J_m}(y_{n_l^{(l)}}))_{l\geqslant 1}$ für jedes $m \geqslant 1$ in K_m' konvergiert, konvergiert auch $(\pi_{T,J}(y_{n_l^{(l)}}))_{l\geqslant 1}$ in der Produktmetrik gegen ein $z \in S^J$. Wir wählen nun ein beliebiges Element $z' \in \pi_{T,J}^{-1}(\{z\})$, sodass

$$\pi_{T,J_m}(z') = \pi_{J,J_m}(z) = \lim_{l\to\infty} \pi_{T,J_m}(y_{n_l^{(l)}}) \in K_m'$$

aufgrund der Kompaktheit von K_m' für alle m gilt. Wir erhalten also $z' \in \pi_{T,J_m}^{-1}(K_m') = C_m$ für alle $m \geqslant 1$, was $\bigcap_{m\geqslant 1} C_m = \emptyset$ widerspricht. □

Korollar 1.31. *Ist S ein polnischer Raum und $T \neq \emptyset$ eine beliebige Indexmenge, so existiert zu jeder projektiven Familie $(\mathbb{P}_J)_{J\subseteq T \text{endlich}}$ ein stochastischer Prozess $X = (X_t, t \in T)$ mit Zustandsraum $(S, \mathscr{B}_S)$, dessen endlichdimensionale Verteilungen durch $(\mathbb{P}_J)$ gegeben sind.*

Beweis. Nach Kolmogorovs Erweiterungssatz existiert ein Maß $\mathbb{P}$ auf $(S^T, \mathscr{B}_S^{\otimes T})$ mit

$$\mathbb{P}\big(\pi_{T,\{t_1,\dots,t_n\}}^{-1}(A)\big) = \mathbb{P}_{t_1,\dots,t_n}(A)$$

für alle $n \in \mathbb{N}, A \in \mathscr{B}_S^{\otimes n}$ und $t_1, \dots, t_n \in T$. Wir definieren X als den Koordinatenprozess auf $(S^T, \mathscr{B}_S^{\otimes T}, \mathbb{P})$ via

$$X_t((s_r)_{r\in T}) := s_t.$$

Dann ist X_t für alle $t \in T$ messbar und es gilt

$$\mathbb{P}\big((X_{t_1}, \dots, X_{t_n}) \in A\big) = \mathbb{P}\big(\pi_{T,\{t_1,\dots,t_n\}}^{-1}(A)\big) = \mathbb{P}_{t_1,\dots,t_n}(A) \quad \text{für alle } A \in \mathscr{B}_S^{\otimes n}.$$ □

Bemerkung 1.32. Es gibt Gegenbeispiele, die zeigen, dass Kolmogorovs Erweiterungssatz für nicht-polnische Räume im Allgmeinen nicht gilt. Andererseits liefert der Satz von Ionescu-Tulcea (vgl. Maßtheorie) die Existenz von allgemeinen Maßräumen für abzählbare Indexmengen und unter einer Markovschen Abhängigkeitsstruktur. Im Spezialfall des Produktmaßes kann man die Existenz sogar für beliebige Indexmengen und ohne Annahme an den Zustandsraum zeigen (vgl. Bauer (1992)). Für polnische Räume folgt die Existenz des Produktmaßes und damit die Existenz beliebig vieler unabhängiger Zufallsvariablen bereits aus Kolmogorovs Erweiterungssatz (Übung ☑).

Beispiel 1.33 (Markovketten mit diskretem Zustandsraum)**.** Sei $(S, \mathscr{P}(S))$ ein abzählbarer Zustandsraum. Dieser ist polnisch, wenn S mit der diskreten Metrik $d(s,t) = \mathbb{1}_{s\neq t}$ versehen wird. Nehmen wir ein Wahrscheinlichkeitsmaß $\mu^{(0)}$ auf $(S, \mathscr{P}(S))$ als gegeben an (die Anfangsverteilung) und betrachten so genannte *stochastische Matrizen*

$$P^{(n)} = (p_{i,j}^{(n)})_{i,j\in S} \quad \text{mit} \quad p_{i,j}^{(n)} \in [0,1], \sum_{j\in S} p_{i,j}^{(n)} = 1 \text{ für alle } i,j \in S.$$

Dann folgt die Existenz einer Markovkette $(X_n, n \geqslant 0)$ mit $\mathbb{P}^{X_0} = \mu^{(0)}$ und $\mathbb{P}(X_n = j | X_{n-1} = i) = p_{i,j}^{(n)}$ aus Kolomogorovs Erweiterungssatz. Wir weisen zunächst Konsistenz der Verteilungen

$$\mu_n(A) := \sum_{i_0\in S}\sum_{i_1\in S}\cdots\sum_{i_n\in S} \mathbb{1}_A(i_0,\ldots,i_n)\mu_{i_0}^{(0)} P_{i_0,i_1}^{(1)}\cdots P_{i_{n-1},i_n}^{(n)} \quad \text{für alle } A \subseteq S^{n+1}$$

nach. Dies folgt induktiv aus der Konsistenz von μ_n und μ_{n+1}, für die gilt

$$\begin{aligned}
\mu_{n+1}(\pi^{-1}_{\{0,\ldots,n+1\},\{0,\ldots,n\}}(A)) &= \sum_{i_0\in S}\sum_{i_1\in S}\cdots\sum_{i_{n+1}\in S} \mathbb{1}_{A\times S}(i_0,\ldots,i_{n+1})\mu_{i_0}^{(0)} P_{i_0,i_1}^{(1)}\cdots P_{i_n,i_{n+1}}^{(n+1)}\\
&= \sum_{i_0\in S}\sum_{i_1\in S}\cdots\sum_{i_n\in S} \mathbb{1}_A(i_0,\ldots,i_n)\mu_{i_0}^{(0)} P_{i_0,i_1}^{(1)}\cdots P_{i_{n-1},i_n}^{(n)}\Big(\sum_{i_{n+1}\in S} P_{i_n,i_{n+1}}^{(n+1)}\Big)\\
&= \mu_n(A).
\end{aligned}$$

Für beliebige endliche $J \subseteq \mathbb{N}_0$ setzen wir

$$\mu_J(A) := \mu_{\max J}(\pi^{-1}_{\{0,\ldots,\max J\},J}(A)), \quad A \subseteq S^J.$$

Dann ist $(\mu_J)_{J\subseteq\mathbb{N}_0}$ tatsächlich eine projektive Familie, denn für alle endlichen $J \subseteq J'$ und $A \subseteq S^J$ folgt

$$\begin{aligned}
\mu_{J'}(\pi^{-1}_{J',J}(A)) &= \mu_{\max J'}(\pi^{-1}_{\{0,\ldots,\max J'\},J}(A))\\
&= \mu_{\max J'}(\pi^{-1}_{\{0,\ldots,\max J'\},\{0,\ldots,\max J\}}(\pi^{-1}_{\{0,\ldots,\max J\},J}(A)))\\
&= \mu_{\max J}(\pi^{-1}_{\{0,\ldots,\max J\},J}(A)) = \mu_J(A).
\end{aligned}$$

Somit kann die Markovkette als Koordinatenprozess auf $(S^{\mathbb{N}_0}, \mathscr{P}(S)^{\otimes\mathbb{N}_0})$ mit einem geeigneten Wahrscheinlichkeitsmaß $\mathbb{P}$ konstruiert werden.

2 Markovprozesse

2.1 Existenz von Markovprozessen

Nachdem wir mit Kolmogorovs Erweiterungssatz ein sehr starkes Resultat zur Hand haben, können wir das Existenzresultat aus dem letzten Beispiel deutlich verallgemeinern. Lassen wir neben $T = \mathbb{N}_0$ (diskrete Zeit) auch $T = \mathbb{R}_+$ (stetige Zeit) zu, so sprechen wir von einem Markovprozess. Zudem werden wir nicht annehmen, dass der Zustandsraum abzählbar ist. Um die Übergangswahrscheinlichkeit rigoros beschreiben zu können, benötigen wir die aus der Maßtheorie bekannten Markovkerne.

Wiederholung zu Markovkernen: Es seien $(E_i, \mathcal{A}_i), i = 0, 1, 2$, drei messbare Räume.

- Eine Abbildung $\kappa\colon E_1 \times \mathcal{A}_2 \to [0,1]$ heißt *Markovkern* von $(E_1, \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$, falls
  - (i) $x_1 \mapsto \kappa(x_1, A_2)$ eine $\mathcal{A}_1$-messbare Abbildung für jedes $A_2 \in \mathcal{A}_2$ ist und
  - (ii) $A_2 \mapsto \kappa(x_1, A_2)$ ein Wahrscheinlichkeitsmaß auf $\mathcal{A}_2$ für jedes $x_1 \in E_1$ ist.
- Ist κ_1 ein Markovkern von $(E_0, \mathcal{A}_0)$ nach $(E_1, \mathcal{A}_1)$ sowie κ_2 ein Markovkern von $(E_0 \times E_1, \mathcal{A}_0 \otimes \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$, dann ist die Abbildung
$$\begin{aligned}\kappa_1 \otimes \kappa_2\colon E_0 \times (\mathcal{A}_1 \otimes \mathcal{A}_2) &\to [0,1]\\ (x_0, A) &\mapsto \int_{E_1} \Big(\int_{E_2} \mathbb{1}_A(x_1, x_2) \kappa_2((x_0, x_1), \mathrm{d}x_2) \Big) \kappa_1(x_0, \mathrm{d}x_1)\end{aligned}$$
ein Markovkern von $(E_0, \mathcal{A}_0)$ nach $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ und heißt der *Produktkern* von κ_1 und κ_2.
- Ist κ_2 ein Kern von $(E_1, \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$, so definieren wir das Produkt $\kappa_1 \otimes \kappa_2$ analog, wobei wir κ_2 als Kern von $(E_0 \times E_1, \mathcal{A}_0 \otimes \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$ auffassen, der nicht von der E_0-Koordinate abhängt.
- Ist μ ein Wahrscheinlichkeitsmaß auf $(E_1, \mathcal{A}_1)$ und κ ein Markovkern von $(E_1, \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$, dann definiert $\mu \otimes \kappa(A) := \int_{E_1} \Big(\int_{E_2} \mathbb{1}_A(x_1, x_2) \kappa_1(x_1, \mathrm{d}x_2) \Big) \mu(\mathrm{d}x_1), A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, ein Wahrscheinlichkeitsmaß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ und heißt *Kopplung*.
- Ist κ_1 ein Markovkern von $(E_0, \mathcal{A}_0)$ nach $(E_1, \mathcal{A}_1)$ sowie κ_2 ein Markovkern von $(E_1, \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$, dann heißt
$$\kappa_1 \cdot \kappa_2\colon \quad E_0 \times \mathcal{A}_2 \to [0,1], \quad (x_0, A) \mapsto \int_{E_1} \kappa_2(x_1, A_2) \kappa_1(x_0, \mathrm{d}x_1)$$
Verkettung von κ_1 und κ_2 und ist ein Markovkern $(E_0, \mathcal{A}_0)$ nach $(E_2, \mathcal{A}_2)$, da $(\kappa_1 \cdot \kappa_2)(x_0, A) = (\kappa_1 \otimes \kappa_2)(x_0, \pi_2^{-1}(A))$.
- Sind X, Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in einem polnischen Raum E, dann existiert ein Markovkern $\kappa_{X,Y}$ von $(E, \mathscr{B}_E)$ nach $(E, \mathscr{B}_E)$ mit
$$\kappa_{X,Y}(y, A) = \mathbb{E}\big[\mathbb{1}_{X^{-1}(A)} | Y = y\big].$$
$\kappa_{X,Y}$ wird *reguläre Version der bedingten Wahrscheinlichkeit* genannt.

Definition 2.1. Sei $T = \mathbb{N}_0$ *(diskrete Zeit)* oder $T = [0, \infty)$ *(stetige Zeit)* und E ein polnischer Raum. Dann heißt eine $(E, \mathscr{B}_E)$-wertige Familie $X = (X_t, t \in T)$ von Zufallsvariablen mit Verteilungen $(\mathbb{P}_x)_{x \in E}$ (homogener) Markovprozess, falls gilt:

(i) Für jedes $x \in E$ ist X ein stochastischer Prozess auf $(\Omega, \mathcal{F}, \mathbb{P}_x)$ mit $\mathbb{P}_x(X_0 = x) = 1$.

(ii) Die Abbildung $\kappa\colon E \times \mathscr{B}_E^{\otimes T} \to [0,1], (x, A) \mapsto \mathbb{P}_x(X \in A) = \mathbb{P}(X \in A | X_0 = x)$ ist ein Markovkern.

(iii) Für jedes $s, t \in T, x \in E$ und $A \in \mathscr{B}_E$ gilt die (schwache) Markoveigenschaft
$$\mathbb{P}_x\big(X_{t+s} \in A \big| \mathcal{F}_s\big) = \kappa_t(X_s, A), \quad \mathbb{P}_x\text{-f.s.},$$
wobei $\mathcal{F}_s := \sigma(X_r, r \leqslant s)$ und
$$\kappa_t(x, A) := \kappa(x, \pi_t^{-1}(A)) = \mathbb{P}(X_t \in A | X_0 = x)$$
die Übergangswahrscheinlichkeit von X zur Zeitdifferenz t beschreibt.

Bemerkung 2.2. Eigenschaft (i) besagt, dass unter dem Maß $\mathbb{P}_x$ der Markovprozess im Punkt $X_0 = x$ beginnt. Wir schreiben $\mathbb{E}_x$ für den Erwartungswert unter $\mathbb{P}_x$ und $\mathbb{P}_Y := \kappa(Y, \cdot)$ für eine E-wertige Zufallsvariable Y. Die Markoveigenschaft besagt also

$$\mathbb{P}_x\big(X_{t+s} \in A \big| \mathcal{F}_s\big) = \mathbb{P}_x\big(X_{t+s} \in A \big| X_s\big) = \mathbb{P}_{X_s}(X_t \in A).$$

Die Verteilung von X_{t+s} gegeben der Vergangenheit bis zum Zeitpunkt s hängt also nur von X_s (Markoveigenschaft, wie oben) und der Zeitdifferenz t (Homogenität) ab.

Beispiel 2.3. Ist $(X_n, n \in \mathbb{N}_0)$ eine (homogene) Markovkette mit abzählbarem Zustandsraum $(S, \mathcal{S})$ und 1-Schritt-Übergangsmatrix $P = (p_{ij})_{i,j \in S}$, $p_{ij} = \mathbb{P}(X_1 = j | X_0 = i) = \mathbb{P}_i(X_1 = j)$, dann sind die zugehörigen Markovkerne gegeben durch

$$\kappa_n(i, A) = \mathbb{P}(X_n \in A | X_0 = i) = \sum_{j \in S} \mathbb{1}_A(j)(P^n)_{ij} = \Big(\sum_{j \in S}(P^n)_{ij} \cdot \delta_j\Big)(A), \quad n \in \mathbb{N}, i \in S, A \in \mathcal{S}.$$

Unser Ziel ist es, die Existenz von Markovprozessen nachzuweisen. Wir werden also insbesondere Kolmogorovs Erweiterungssatz anwenden. Im Spezialfall von Markov-Ketten in diskreter Zeit haben wir die Existenz bereits in Beispiel 1.33 nachgewiesen. Auch in stetiger Zeit und für polnische Zustandsräume ist die Struktur der endlichdimensionalen Verteilungen durch die Markoveigenschaft bestimmt und relativ einfach.

Definition 2.4. Es seien E ein polnischer Raum und $T = \mathbb{N}_0$ oder $T = [0, \infty)$. Eine Familie $(\kappa_t, t \in T)$ von Markovkernen von E nach E heißt Markov'sche Halbgruppe, falls $\kappa_0(x, \cdot) = \delta_x$ für alle $x \in E$ und die Chapman-Kolmogorov-Gleichung erfüllt ist:

$$\kappa_s \cdot \kappa_t = \kappa_{s+t} \quad \text{für alle } s, t \in T.$$

Beispiel 2.5. Betrachte die Familie $(\kappa_t, t \geqslant 0)$ von Markovkernen auf $\mathbb{R}$, wobei $\kappa_0(x, \cdot) = \delta_x$ und $\kappa_t(x, \cdot)$ das Wahrscheinlichkeitsmaß der $\mathcal{N}(x, t)$-Verteilung für $t > 0$ sei. Dann gilt für unabhängige Zufallsvariablen $X \sim \mathcal{N}(0, s), Y \sim \mathcal{N}(0, t)$

$$\begin{aligned}
\kappa_s \cdot \kappa_t(x_0, A) &= \int_{\mathbb{R}} \kappa_t(x, A)\kappa_s(x_0, \mathrm{d}x) = \int_{\mathbb{R}} \mathbb{P}(x + Y \in A)\kappa_s(x_0, \mathrm{d}x) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_A(x + y)\frac{1}{2\pi} e^{-y^2/2} e^{-(x-x_0)^2/2} \mathrm{d}y\mathrm{d}x \\
&= \mathbb{P}(x_0 + X + Y \in A) = \kappa_{s+t}(x_0, A),
\end{aligned}$$

da $X + Y \sim \mathcal{N}(0, s + t)$. Also ist $(\kappa_t, t \geqslant 0)$ eine Markov'sche Halbgruppe.

Lemma 2.6. *Für jeden Markovprozess $X = (X_t, t \in T)$ in einem polnischen Raum E mit Übergangswahrscheinlichkeiten*

$$\mathbb{P}_x(X_t \in A) = \kappa_t(x, A) \quad \textit{für alle } x \in E, A \in \mathscr{B}_E, t \in T, \tag{2}$$

ist $(\kappa_t, t \in T)$ eine Markov'sche Halbgruppe. Die Verteilung von X ist durch (2) eindeutig bestimmt.

Beweis. Zum Zeitpunkt 0 gilt $\kappa_0(x, \cdot) = \delta_x$ nach Bedingung (i). Aufgrund der Markoveigenschaft gilt weiter

$$\begin{aligned}
\kappa_{s+t}(x, A) &= \mathbb{P}_x(X_{s+t} \in A) \\
&= \mathbb{E}_x\big[\mathbb{P}_x(X_{s+t} \in A | X_s)\big] \\
&= \int_E \kappa_t(y, A)\kappa_s(x, \mathrm{d}y) = \kappa_s \cdot \kappa_t(A).
\end{aligned}$$

Die Eindeutigkeit folgt, da die endlichdimensionalen Randverteilungen durch $(\kappa_t, t \in T)$ festgelegt sind. □

Andersherum lässt sich zeigen, dass jede Markov'sche Halbgruppe einen Markovprozess definiert.

Satz 2.7. *Es seien E ein polnischer Raum, $T = \mathbb{N}_0$ oder $T = [0, \infty)$ und $(\kappa_t, t \in T)$ eine Markov'sche Halbgruppe von Kernen auf E. Dann existiert ein Markovkern κ von $(E, \mathscr{B}_E)$ nach $(E^T, \mathscr{B}_E^{\otimes T})$, sodass für jedes $x \in E$ und beliebige $J := \{t_0, \dots, t_n\} \subseteq T$ mit $0 = t_0 < t_1 < \dots < t_n$ und $n \in \mathbb{N}$ gilt*

$$\kappa\big(x, \pi_{T,J}^{-1}(A)\big) = \big(\delta_x \otimes \bigotimes_{k=1}^{n} \kappa_{t_k - t_{k-1}}\big)(A) \quad \text{für alle } A \in \mathscr{B}_E^{\otimes J}. \tag{3}$$

Beweis. Wir zeigen zunächst, dass (3) für jedes feste $x \in E$ ein Wahrscheinlichkeitsmaß auf $(E^T, \mathscr{B}_E^{\otimes T})$ definiert. Nach Kolmogorovs Erweiterungssatz genügt es, zu zeigen, dass die Familie $\{\mathbb{P}_J : J \subseteq T \text{ endlich}, 0 \in J\}$ definiert durch

$$\mathbb{P}_J := \delta_x \otimes \bigotimes_{k=1}^{n} \kappa_{t_k - t_{k-1}}$$

konsistent ist, d.h. für alle $0 \in I \subseteq J \subseteq T$ mit I, J endlich gilt

$$\forall A \in \mathscr{B}_E^{\otimes I}: \quad \mathbb{P}_J(\pi_{J,I}^{-1}(A)) = \mathbb{P}_I(A).$$

Da die Rechteckmengen ein $\cap$-stabiler Erzeuger von $\mathscr{B}_E^{\otimes I}$ sind, genügt es, die Gleichheit für $A := \times_{i \in I} A_i$ mit $A_j \in \mathscr{B}_E, j \in I$, zu zeigen. Wir können außerdem annehmen, dass $I = J \setminus \{t_i\}$ für ein $i = 1, \dots, n$ ist, woraus der allgemeine Fall induktiv folgt.

Ist $i = n$, dann gilt

$$\begin{aligned}
\mathbb{P}_J(\pi_{J,I}^{-1}(A)) &= \mathbb{P}_J(A \times E) = \mathbb{P}_I \otimes \kappa_{t_n - t_{n-1}}(A \times E) \\
&= \int_A \kappa_{t_n - t_{n-1}}(x_{n-1}, E) \mathbb{P}_I\big(\mathrm{d}(x_0, \dots, x_{n-1})\big) = \mathbb{P}_I(A).
\end{aligned}$$

Sei nun $i \in \{1, \dots, n-1\}$ und setze $A_i = E$. Wir definieren für $j = 0, \dots, n-1$

$$f_j(x_j) := \Big(\bigotimes_{k=j}^{n-1} \kappa_{t_{k+1} - t_k}\Big)(x_j, A_{j+1} \times \dots \times A_n).$$

Nach dem Satz von Fubini und der Chapman-Kolmogorov-Gleichung gilt

$$\begin{aligned}
f_{i-1}(x_{i-1}) &= \int_E \int_{A_{i+1}} f_{i+1}(x_{i+1}) \kappa_{t_{i+1} - t_i}(x_i, \mathrm{d}x_{i+1}) \kappa_{t_i - t_{i-1}}(x_{i-1}, \mathrm{d}x_i) \\
&= \int_{A_{i+1}} f_{i+1}(x_{i+1}) \kappa_{t_{i+1} - t_{i-1}}(x_{i-1}, \mathrm{d}x_{i+1}).
\end{aligned}$$

Daraus folgt

$$\begin{aligned}
\mathbb{P}_J(\pi_{J,I}^{-1}(A)) &= \int_{A_0 \times \dots \times A_{i-1}} f_{i-1}(x_{i-1}) \mathbb{P}_{\{t_0, \dots, t_{i-1}\}}(\mathrm{d}(x_0, \dots, x_{i-1})) \\
&= \int_{A_0 \times \dots \times A_{i-1}} \int_{A_{i+1}} f_{i+1}(x_{i+1}) \kappa_{t_{i+1} - t_{i-1}}(x_{i-1}, \mathrm{d}x_{i+1}) \mathbb{P}_{\{t_0, \dots, t_{i-1}\}}(\mathrm{d}(x_0, \dots, x_{i-1})) \\
&= \mathbb{P}_I(A).
\end{aligned}$$

Es bleibt zu zeigen, dass für jedes $A \in \mathscr{B}_E^{\otimes T}$ die Abbildung $x \mapsto \kappa(x, A)$ $(\mathscr{B}_E, \mathscr{B}_{[0,1]})$-messbar ist. Hierfür sei zunächst A wieder eine Rechteckzylindermenge. Dann folgt die Messbarkeit aus der Darstellung

$$\kappa\big(x, \pi_{T,J}^{-1}(A_0 \times \dots \times A_n)\big) = \big(\delta_x \otimes \bigotimes_{k=1}^{n} \kappa_{t_k - t_{k-1}}\big)(A_0 \times \dots \times A_n)$$

$$= \Big(\bigotimes_{k=1}^{n} \kappa_{t_k - t_{k-1}}\Big)(x, A_1 \times \cdots \times A_n) \mathbb{1}_{A_0}(x)$$

für $J = \{t_0, \dots, t_n\} \subseteq T$ und $A_0, \dots, A_n \in \mathscr{B}_E$. Nun ist

$$\mathcal{D} := \big\{A \in \mathscr{B}_E^{\otimes T} : x \mapsto \kappa(x, A) \text{ ist } \mathscr{B}_E\text{-messbar}\big\}$$

ein Dynkinsystem und die Rechteckzylindermengen ein $\cap$-stabiler Erzeuger von $\mathscr{B}_E^{\otimes T}$, sodass $\mathcal{D} = \mathscr{B}_E^{\otimes T}$ folgt. □

Korollar 2.8. *Für einen polnischen Raum E, $T = \mathbb{N}_0$ oder $T = [0, \infty)$ und eine Markov'sche Halbgruppe $(\kappa_t, t \in T)$ von Markovkernen von E nach E existieren ein messbarer Raum $(\Omega, \mathcal{F})$ und ein Markovprozess $X = (X_t, t \in T)$ mit Verteilungen $(\mathbb{P}_x)_{x \in E}$ auf $(\Omega, \mathcal{F})$ mit Übergangswahrscheinlichkeiten*

$$\mathbb{P}_x(X_t \in A) = \kappa_t(x, A) \quad \textit{für alle } x \in E, A \in \mathscr{B}_E, t \in T.$$

Beweis. Wir konstruieren X als kanonischen Prozess und setzen daher $\Omega := E^T, \mathcal{F} := \mathscr{B}_E^{\otimes T}$ und $X_t = \pi_t$ die Projektion auf die t-te Koordinate. Für den Markovkern κ aus Satz 2.7 und für jedes $x \in E$ definieren wir $\mathbb{P}_x := \kappa(x, \cdot)$. Nach Konstruktion gilt dann für endlich viele Zeitpunkte $0 = t_0 < t_1 < \cdots < t_n$ und $A \in \mathscr{B}_E^{\otimes(n+1)}$

$$\mathbb{P}_x\big((X_{t_0}, \dots, X_{t_n}) \in A\big) = \mathbb{P}_x(\pi^{-1}_{T,\{t_0,\dots,t_n\}}(A)) = \delta_x \otimes \bigotimes_{k=1}^{n} \kappa_{t_k - t_{k-1}}(A).$$

Insbesondere gilt für $A \in \mathscr{B}_E$ und alle nichtnegativen, messbare Funktionen $f \colon E^n \to \mathbb{R}_+$

$$\begin{aligned}
&\mathbb{E}_x\big[f(X_{t_0}, \dots, X_{t_{n-1}}) \mathbb{1}_A(X_{t_n})\big] \\
&\quad = \int_{E^n} f(x) \kappa_{t_n - t_{n-1}}(x_{n-1}, A_n) \big(\delta_x \otimes \bigotimes_{k=1}^{n} \kappa_{t_k - t_{k-1}}\big)(\mathrm{d}x) \\
&\quad = \mathbb{E}_x\big[f(X_{t_0}, \dots, X_{t_{n-1}}) \kappa_{t_n - t_{n-1}}(X_{t_{n-1}}, A_n)\big],
\end{aligned}$$

also $\mathbb{P}_x(X_{t_n} \in A | \mathcal{F}_{t_{n-1}}) = \kappa_{t_n - t_{n-1}}(X_{t_{n-1}}, A_n)$. Damit ist X tatsächlich ein Markovprozess und es gilt

$$\mathbb{P}_x(X_t \in A) = (\delta_x \cdot \kappa_t)(A) = \kappa_t(x, A), \quad A \in \mathscr{B}_E.$$ □

Beispiel 2.9. Für den Markov-Prozess $X = (X_t, t \geqslant 0)$, welcher durch die Markov'sche Halbgruppe der Normalverteilungen aus Beispiel 2.5 gegeben ist, gilt

$$\mathbb{P}_0(X_t - X_s \in A | X_s) = \kappa_{t-s}(X_s, A + X_s) = \kappa_{t-s}(0, A) \quad \forall A \in \mathscr{B}_{\mathbb{R}}.$$

Daher ist $\mathbb{P}_0(X_t - X_s \in A) = \mathbb{E}_0[\mathbb{P}_0(X_t - X_s \in A | X_s)] = \kappa_{t-s}(0, A)$, also sind die Zuwächse $X_t - X_s \overset{d}{=} X_{t-s} \sim \mathcal{N}(0, t-s)$ stationär und normalverteilt. Zudem sind die Zuwächse unabhängig: Für $0 = t_0 < t_1 < \cdots < t_n$, $x_0 = 0$ und $A_1, \dots, A_n \in \mathscr{B}_{\mathbb{R}}$ gilt

$$\begin{aligned}
\mathbb{P}_0(\forall i = 1, \dots, n : X_{t_i} - X_{t_{i-1}} \in A_i) &= \int_{\mathbb{R}^n} \prod_{i=1}^{n} \mathbb{1}_{A_i}(x_i - x_{i-1}) \kappa_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n) \cdots \kappa_{t_1 - t_0}(0, \mathrm{d}x_1) \\
&= \prod_{i=1}^{n} \kappa_{t_i - t_{i-1}}(0, A_i) = \prod_{i=1}^{n} \mathbb{P}_0(X_{t_i} - X_{t_{i-1}} \in A_i).
\end{aligned}$$

Wir haben damit fast die (standard) Brownsche Bewegung (siehe Kapitel 5) konstruiert, für die wir allerdings noch stetige Pfade fordern werden. Ob es eine Version von X mit stetigen Pfaden gibt, ist allerdings zu diesem Zeitpunkt noch ungeklärt.

Sofern keine Verwechslung vorkommen kann, werden wir die Verteilungen $(\mathbb{P}_x)_{x \in E}$ nicht mehr explizit erwähnen. Die schwache Markoveigenschaft kann auch wie folgt charakterisiert werden.

Lemma 2.10. *Ein E-wertiger, stochastischer Prozess $(X_t, t \in T)$ mit $T = \mathbb{R}_+$ oder $T = \mathbb{N}_0$ ist genau dann ein Markovprozess, wenn es einen Markovkern $\kappa\colon E \times \mathscr{B}_E^{\otimes T} \to [0,1]$ gibt, sodass für alle messbaren und beschränkten Funktionen $f\colon E^T \to \mathbb{R}$ und alle $s \in T, x \in E$ gilt:*

$$\mathbb{E}_x\big[f((X_{t+s})_{t\in T})|\mathcal{F}_s\big] = \mathbb{E}_{X_s}\big[f(X)\big] := \int_{E^T} f(y)\,\kappa(X_s, \mathrm{d}y). \tag{4}$$

Beweis. „$\Longleftarrow$“ Die Markoveigenschaft folgt aus (4) mit $f(y) := \mathbb{1}_A(y_t)$, denn $\mathbb{P}_x(X_{t+s} \in A|\mathcal{F}_s) = \mathbb{P}_{X_s}(X_t \in A) = \kappa_t(X_s, A)$.

„$\Longrightarrow$“ Wir verwenden maßtheoretische Induktion. Da die Menge der Rechteckzylindermengen ein $\cap$-stabiler Erzeuger von $\mathscr{B}_E^{\otimes T}$ ist, genügt es, Indikatorfunktionen der Form

$$f_m((x_t)_{t\in T}) = \mathbb{1}_{\{A_0\times\dots\times A_m\}}((x_{t_0}, \dots, x_{t_m})) = \prod_{n=0}^{m} \mathbb{1}_{A_n}(x_{t_n})$$

für $m \in \mathbb{N}, t_0 = 0, t_1, \dots, t_m \in T$ und $A_0, \dots, A_m \in \mathscr{B}_E$ zu betrachten. Wir zeigen nun (4) induktiv.

Für $m = 1$ ist (4) äquivalent zu der (schwachen) Markoveigenschaft

$$\mathbb{E}_x[f_1(X_s, X_{s+t_1})|\mathcal{F}_s] = \mathbb{P}_x(X_{s+t_1} \in A_1|\mathcal{F}_s)\mathbb{1}_{A_0}(X_s) = \mathbb{P}_{X_s}(X_{t_1} \in A_1)\mathbb{1}_{A_0}(X_s).$$

Mittels maßtheoretischer Induktion folgt die Gleichheit für alle beschränkten, messbaren Funktionen, die nur von (x_0, x_{t_1}) abhängen.

Nun schließen wir induktiv

$$\begin{aligned}
\mathbb{E}_x\big[f_{m+1}\big((X_{t+s})_{t\in T}\big)\big|\mathcal{F}_s\big] &= \mathbb{E}_x\big[\mathbb{E}_x\big[f_{m+1}\big((X_{t+s})_{t\in T}\big)\big|\mathcal{F}_{s+t_m}\big]\big|\mathcal{F}_s\big] \\
&= \mathbb{E}_x\big[\mathbb{E}_x\big[\mathbb{1}_{A_{m+1}}(X_{s+t_{m+1}})\big|\mathcal{F}_{s+t_m}\big]\prod_{n=0}^{m}\mathbb{1}_{A_n}(X_{s+t_n})\big|\mathcal{F}_s\big] \\
&\overset{Markov}{=} \mathbb{E}_x\big[\mathbb{E}_{X_{s+t_m}}\big[\mathbb{1}_{A_{m+1}}(X_{t_{m+1}-t_m})\big]\prod_{n=0}^{m}\mathbb{1}_{A_n}(X_{s+t_n})\big|\mathcal{F}_s\big] \\
&\overset{IV}{=} \mathbb{E}_{X_s}\big[\mathbb{E}_{X_{t_m}}\big[\mathbb{1}_{A_{m+1}}(X_{t_{m+1}-t_m})\big]\prod_{n=0}^{m}\mathbb{1}_{A_n}(X_{t_n})\big] \\
&\overset{Markov}{=} \mathbb{E}_{X_s}\big[\mathbb{E}_{X_s}\big[\mathbb{1}_{A_{m+1}}(X_{t_{m+1}})|\mathcal{F}_{t_m}\big]\prod_{n=0}^{m}\mathbb{1}_{A_n}(X_{t_n})\big] \\
&= \mathbb{E}_{X_s}\big[\mathbb{E}_{X_s}\big[\prod_{n=0}^{m+1}\mathbb{1}_{A_n}(X_{t_n})|\mathcal{F}_{t_m}\big]\big] = \mathbb{E}_{X_s}[f((X_n)_{n\in T})].
\end{aligned}$$

Erneut folgt hieraus die Behauptung für alle beschränkten, messbaren Funktionen, die von $(x_{t_0}, \dots, x_{m+1})$ abhängen. □

In diskreter Zeit gilt sogar automatisch eine stärkere Form der Markoveigenschaft, bei der der Zeitpunkt auf den wir bedingen auch zufällig sein kann. Die Definition und wesentliche Eigenschaften von Stoppzeiten (in diskreter Zeit) werden in Anhang B wiederholt.

Definition 2.11. Für $n \geqslant 0$ definieren wir den Verschiebungs- oder Shift-Operator

$$\vartheta_n: \quad E^{\mathbb{N}_0} \to E^{\mathbb{N}_0}, \quad (x_k)_{k\geqslant 0} \mapsto (x_{n+k})_{k\geqslant 0}.$$

Satz 2.12. *Jede Markovkette $X = (X_n, n \in \mathbb{N}_0)$ erfüllt die starke Markoveigenschaft: Für jede Stoppzeit τ und jede $(\mathscr{B}_E^{\otimes\mathbb{N}_0}, \mathscr{B}_\mathbb{R})$-messbare, beschränkte Funktion $f\colon E^{\mathbb{N}_0} \to \mathbb{R}$ sowie jedes $x \in E$ gilt*

$$\mathbb{E}_x[f \circ \vartheta_\tau(X))|\mathcal{F}_\tau] = \mathbb{E}_{X_\tau}[f(X)] = \int_{E^T} f(y)\kappa(X_\tau, \mathrm{d}y) \quad f.s.$$

Beweis. Aus (4) folgt für $A \in \mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n \forall n \in \mathbb{N}_0\}$

$$\begin{aligned}\mathbb{E}_x[f((X_{\tau+n})_{n\geqslant 0})\mathbb{1}_A] &= \sum_{k\geqslant 0} \mathbb{E}_x[f((X_{\tau+n})_{n\geqslant 0})\mathbb{1}_{A\cap\{\tau=k\}}] \\ &= \sum_{k\geqslant 0} \mathbb{E}_x\left[\mathbb{E}_x[f((X_{k+n})_{n\geqslant 0})|\mathcal{F}_k]\mathbb{1}_{A\cap\{\tau=k\}}\right] \\ &= \sum_{k\geqslant 0} \mathbb{E}_x\left[\mathbb{E}_{X_k}[f((X_n)_{n\geqslant 0})]\mathbb{1}_{A\cap\{\tau=k\}}\right] \\ &= \mathbb{E}_x\left[\mathbb{E}_{X_\tau}[f((X_n)_{n\geqslant 0})]\mathbb{1}_A\right]. \end{aligned}$$

□

2.2 Markovketten in diskreten Räumen: Rekurrenz und Transienz

Im Folgenden sei $X = (X_n, n \in \mathbb{N}_0)$ eine homogene Markovkette mit diskretem Zustandsraum S, Übergangswahrscheinlichkeiten $p_{ij} = \mathbb{P}(X_1 = j|X_0 = i) = \mathbb{P}_i(X_1 = j)$ und Übergangsmatrix $P = (p_{ij})_{i,j\in S}$. Wie aus Kolmogorovs Erweiterungssatz gefolgert, kann X als kanonischer Prozess auf $(\Omega, \mathcal{F}, \mathbb{P}_x)$ mit $\Omega = S^{\mathbb{N}_0}$ und $\mathcal{F} = \mathscr{P}(S)^{\otimes\mathbb{N}_0}$ definiert werden, sodass $X_n(\omega) = \omega_n$. Als Filtration verwenden wir die natürliche Filtration von X.

Definition 2.13. Für jedes $y \in S$ sei $T_y^0 = 0$ und

$$T_y^k := \inf\left\{n > T_y^{k-1} : X_n = y\right\}, \quad k \in \mathbb{N}.$$

T_y^k heißt k-te Eintrittszeit von X in y oder Rückkehrzeit nach y. Für $x, y \in S$ ist

$$\rho_{x,y} := \mathbb{P}_x(T_y^1 < \infty)$$

die Wahrscheinlichkeit, jemals von x nach y zu gehen.

Bemerkung 2.14. T_y^k ist eine Stoppzeit. $\rho_{x,x}$ ist insbesondere die Rückkehrwahrscheinlichkeit von x nach x. Beachte, dass $T_x^1 > 0$ selbst beim Start in x gilt.

Satz 2.15. *Für alle $x, y \in S$ und $k \in \mathbb{N}$ gilt $\mathbb{P}_x(T_y^k < \infty) = \rho_{x,y} \cdot \rho_{y,y}^{k-1}$.*

Beweis. Die Behauptung folgt induktiv. Für $k = 1$ ist nichts zu zeigen.

Sein nun $k \geqslant 2$. Für $Y := \mathbb{1}_{\{T_y^1<\infty\}}$ gilt $\{T_y^k < \infty\} = \{T_y^1 \circ \vartheta_{T_y^{k-1}} < \infty\} \cap \{T_y^{k-1} < \infty\} = \{Y \circ \vartheta_{T_y^{k-1}} = 1\} \cap \{T_y^{k-1} < \infty\}$. Aus der starken Markoveigenschaft folgt

$$\begin{aligned}\mathbb{P}_x(T_y^k < \infty) =&\mathbb{E}_x\left[\mathbb{1}_{\{T_y^k<\infty\}}\mathbb{1}_{\{T_y^{k-1}<\infty\}}\right] \\ =&\mathbb{E}_x\left[\mathbb{E}_x\left[Y \circ \vartheta_{T_y^{k-1}}|\mathcal{F}_{T_y^{k-1}}\right]\mathbb{1}_{\{T_y^{k-1}<\infty\}}\right] \\ =&\mathbb{E}_x\left[\mathbb{E}_{X_{T_y^{k-1}}}[Y]\mathbb{1}_{\{T_y^{k-1}<\infty\}}\right] \\ =&\mathbb{E}_x\left[\mathbb{P}_y(T_y^1 < \infty)\mathbb{1}_{\{T_y^{k-1}<\infty\}}\right] = \rho_{y,y}\mathbb{P}_x(T_y^{k-1} < \infty).\end{aligned}$$

□

Definition 2.16. Ein Zustand $y \in S$ heißt rekurrent, falls $\rho_{y,y} = 1$ und er heißt transient, falls $\rho_{y,y} < 1$. Die Anzahl der Besuche von X in y bezeichnen wir mit

$$N_y := \sum_{n\geqslant 1} \mathbb{1}_{\{X_n=y\}}.$$

Satz 2.17.

(i) Ist $y \in S$ transient, dann gilt $\mathbb{E}_x[N_y] = \frac{\rho_{x,y}}{1-\rho_{y,y}}$.

(ii) Sind $x, y \in S$ mit $\rho_{x,y} > 0$, so ist y genau dann rekurrent, wenn $\mathbb{E}_x[N_y] = \infty$.

Beweis. (i) Es gilt

$$\begin{aligned}\mathbb{E}_x[N_y] &= \sum_{k=1}^{\infty} \mathbb{P}_x(N_y \geqslant k) = \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < \infty) \\ &= \sum_{k=1}^{\infty} \rho_{x,y} \cdot \rho_{y,y}^{k-1} = \frac{\rho_{x,y}}{1-\rho_{y,y}}.\end{aligned}$$

(ii) $\mathbb{E}_x[N_y] = \sum_{k=1}^{\infty} \rho_{x,y} \cdot \rho_{y,y}^{k-1} = \infty \Leftrightarrow \rho_{y,y} = 1$. □

Satz 2.18. *Ist $x \in S$ rekurrent und $\rho_{x,y} > 0$ für ein $y \in S$, dann ist auch y rekurrent und $\rho_{x,y} = \rho_{y,x} = 1$.*

Beweis. (i) Aus $\rho_{x,x} = 1, \rho_{x,y} > 0$ folgt, dass $\rho_{y,x} = 1$ gilt: Da $\rho_{x,y} > 0$ existiert ein $k \in \mathbb{N}$ und $x_1, \ldots, x_k \in S \setminus \{x\}$ mit $x_k = y$ und

$$\mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k) > 0.$$

Die Markoveigenschaft ergibt dann:

$$\begin{aligned}0 = 1 - \rho_{x,x} = \mathbb{P}_x(T_x^1 = \infty) &\geqslant \mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k, T_x^1 \circ \vartheta_k = \infty) \\ &= \mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k)\mathbb{P}_y(T_x^1 = \infty) \\ &= \mathbb{P}_x(X_1 = x_1, \ldots, X_k = x_k)(1 - \rho_{y,x}).\end{aligned}$$

Da $\rho_{x,y} > 0$, muss also $\rho_{y,x} = 1$ gelten.

(ii) y ist rekurrent: Da $\rho_{x,y} > 0, \rho_{y,x} > 0$, existieren $k_1, k_2 \in \mathbb{N}$ mit $(P^{k_1})_{y,x}, (P^{k_2})_{x,y} > 0$. Dann gilt

$$\begin{aligned}\mathbb{E}_y[N_y] = \sum_{n=1}^{\infty} \mathbb{P}_y(X_n = y) &\geqslant \sum_{n=1}^{\infty} \mathbb{P}_y(X_{k_1} = x, X_{k_1+n} = x, X_{k_1+k_2+n} = y) \\ &= \sum_{n=1}^{\infty} (P^{k_1})_{y,x}(P^n)_{x,x}(P^{k_2})_{x,y} \\ &= (P^{k_1})_{y,x}(P^{k_2})_{x,y} \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = x) \\ &= (P^{k_1})_{y,x}(P^{k_2})_{x,y}\mathbb{E}_x[N_x] = \infty.\end{aligned}$$

Aus Symmetrie folgt, dass auch $\rho_{x,y} = 1$ gilt. □

Definition 2.19. Die Markovkette $X = (X_n, n \in \mathbb{N}_0)$ heißt <u>irreduzibel</u>, *falls $\rho_{x,y} > 0$ für alle $x, y \in S$.*

Aus Satz 2.18 folgt direkt, dass in einer irreduziblen Markovkette entweder alle Zustände rekurrent oder alle Zustände transient sind. Ist der Zustandsraum endlich, schließt sich sogar zusätzlich die zweite Möglichkeit aus.

Satz 2.20. *Für eine irreduzible homogene Markovkette in einem endlichen Zustandsraum S ist jeder Zustand rekurrent.*

Beweis. Nehmen wir an, alle Zustände sind transient. Dann folgt aus Satz 2.17 und dem Satz von Fubini:

$$\begin{aligned}\infty > \sum_{y \in S} \mathbb{E}_x[N_y] &= \sum_{y \in S} \sum_{n \geqslant 1} \mathbb{P}_x(X_n = y) \\ &= \sum_{n \geqslant 1} \underbrace{\sum_{y \in S} \mathbb{P}_x(X_n = y)}_{=1} = \infty.\end{aligned}$$

Dieser Widerspruch zeigt, dass alle Zustände rekurrent sein müssen. □

Beispiel 2.21.

(i) Für *endliche Zustandsräume* können wir Rekurrenz/Transienz und Irreduzibilität leicht anhand der Übergangswahrscheinlichkeiten ablesen.

(ii) Für die *einfache Irrfahrt* $S_n = \sum_{k=1}^{n} X_k$ mit i.i.d. $(X_k)_{k\geqslant 1}$ und $\mathbb{P}(X_1 = 1) = p \in (0,1)$ und $\mathbb{P}(X_1 = -1) = 1 - p =: q$ gilt (Übung ☑):

$$\mathbb{P}_0(T_b^1 < \infty) = \min\big(1, \frac{p}{q}\big)^b \quad \text{für alle } b \in \mathbb{N}.$$

(a) *Symmetrischer Fall* $p = \frac{1}{2}$: Dann gilt $\mathbb{P}_0(T_b^1 < \infty) = 1$ für alle $b \in \mathbb{Z} \setminus \{0\}$ wegen Symmetrie. Aufgrund der Homogenität des Zustandsraumes folgt auch $\rho_{x,y} = \mathbb{P}_x(T_y^1 < \infty) = 1$ für alle $x, y \in \mathbb{Z}$ mit $x \neq y$. Aus der Markoveigenschaft folgt nun

$$\begin{aligned}\rho_{x,x} = \mathbb{P}_x(T_x^1 < \infty) &= \mathbb{P}_x(X_1 = x+1, T_x^1 < \infty) + \mathbb{P}_x(X_1 = x-1, T_x^1 < \infty)\\ &= \frac{1}{2}\mathbb{P}_{x+1}(T_x^1 < \infty) + \frac{1}{2}\mathbb{P}_{x-1}(T_x^1 < \infty)\\ &= \frac{1}{2}(\rho_{x+1,x} + \rho_{x-1,x}).\end{aligned}$$

Damit ist die symmetrische einfache Irrfahrt irreduzibel und rekurrent.

(b) *Asymmetrischer Fall* $p < 1/2$: Aus $\mathbb{P}_0(T_b^1 < \infty) < 1$ und Symmetrie/Homogenität folgt $\rho_{x,y} = (\frac{p}{q})^{y-x} < 1$ für $x < y$ und $\rho_{x,y} = 1$ für $x > y$. Die Markoveigenschaft ergibt wieder

$$\rho_{x,x} = p\rho_{x+1,x} + q\rho_{x-1,x} = p \cdot 1 + q\big(\frac{p}{q}\big)^1 = 2p < 1.$$

Damit ist die asymmetrische einfache Irrfahrt irreduzibel und transient.

Höhere Dimensionen? Die symmetrische einfache Irrfahrt in $\mathbb{Z}^d$ mit $d \geqslant 1$ ist genau dann rekurrent, falls $d \leqslant 2$ (Satz von Polya, vgl. Klenke (2006, Satz 17.39)).

2.3 Zeitstetige Markovprozesse mit abzählbarem Zustandsraum

Es sei $X = (X_t, t \geqslant 0)$ ein homogener Markovprozess mit abzählbarem Zustandsraum S und Übergangswahrscheinlichkeiten $p_{ij}(t) = \mathbb{P}_i(X_t = j), i, j \in S$, und Übergangsmatrizen $P_t := (p_{ij}(t))_{i,j\in S}$. Da wir im zeitstetigen Modell keine kleinste Zeiteinheit haben, werden wir die Übungsmatrizen über einen infinitesimalen Ansatz beschreiben.

Definition 2.22. Gilt für die Halbgruppe $(P_t, t \geqslant 0)$, dass

(i) der Grenzwert $q_{ij} := \lim_{t\downarrow 0} \frac{1}{t}\mathbb{P}_i(X_t = j)$ für alle $i, j \in S$ mit $i \neq j$ existiert und

(ii) $|q_{ii}| < \infty$ für $q_{ii} := -\sum_{j\in S, j\neq i} q_{ij}$ und alle $i \in S$ gilt,

dann heißt die Matrix $Q = (q_{ij})_{i,j\in S} \in \mathbb{R}^{S\times S}$ <u>Generator</u> der Halbgruppe.

Aus dieser Definition und unter der Annahme, dass wir die Summe und die Ableitung vertauschen dürfen, folgt die Gleichheit

$$\lim_{t\downarrow 0} \frac{1}{t}\big(\mathbb{P}_i(X_t = j) - \mathbb{1}_{\{i=j\}}\big) = q_{ij} \quad \forall i, j \in S.$$

Beispiel 2.23. Ist N ein Poissonprozess mit Intensität $\lambda > 0$, so ist sein Generator gegeben durch $Q = (q_{ij})_{i,j\in\mathbb{N}_0}$ mit $q_{ij} = \lambda(\mathbb{1}_{j=i+1} - \mathbb{1}_{\{i=j\}})$.

Satz 2.24. *Erfüllt eine Matrix $Q = (q_{ij})_{i,j\in S} \in \mathbb{R}^{S\times S}$ die Bedingungen $q_{ij} \geqslant 0$ für alle $i \neq j$, $q_{ii} = -\sum_{j\neq i} q_{ij}$ und für alle $i \in S$ sowie $\lambda := \sup_{i\in S} |q_{ii}| < \infty$, dann ist Q der Generator eines eindeutig bestimmten homogenen Markovprozesses $X = (X_t, t \geqslant 0)$. In diesem Fall gilt für die Übergangsmatrizen*

$$P_t = e^{tQ} := \sum_{n\geqslant 0} \frac{(tQ)^n}{n!} \quad \textit{für alle } t > 0.$$

Bemerkung 2.25. Jede stochastische Matrix $P = (p_{ij})_{i,j\in S}$, d.h. $p_{ij} \geqslant 0$ und $\sum_j p_{ij} = 1$ für alle $i, j \in S$, kann als linearer Operator

$$P\colon \ell_\infty(S) \to \ell_\infty(S), x \mapsto \Big(\sum_{j\in S} p_{ij} x_j\Big)_{i\in S}$$

auf dem Raum der beschränkten Folgen $\ell_\infty(S) := \{x \in \mathbb{R}^S |\, \|x\|_\infty := \sup_{i\in S} |x_i| < \infty\}$ verstanden werden. Für dessen Operatornorm gilt

$$\|P\| := \sup_{x\in\mathbb{R}^S} \frac{\|Px\|_\infty}{\|x\|_\infty} = \sup_{x\in\mathbb{R}^S} \frac{\sup_i |\sum_j p_{ij} x_j|}{\|x\|_\infty} \leqslant \sup_i \sum_j p_{ij} = 1.$$

Folglich ist der Operator P beschränkt (und stetig). Für die Folge $e = (e_i)_{i\in S}$ mit $e_i = 1$ für alle $i \in S$ gilt $Pe = e$, sodass sich $\|P\| = 1$ ergibt.

Unter den Annahmen des Satzes ist $P = \frac{1}{\lambda} Q + I$ mit der Einheitsmatrix $I \in \mathbb{R}^{S\times S}$ eine stochastische Matrix. Damit gilt $\|Q\| \leqslant \lambda(\|P\| + \|I\|) = 2\lambda$, sodass Q ein beschränkter Operator auf $\ell_\infty(S)$ ist. Insbesondere konvergiert $\big\|\sum_{n=0}^N \frac{(tQ)^n}{n!}\big\| \leqslant \sum_{n=0}^N \frac{t^n\|Q\|^n}{n!}$ für $N \to \infty$ und das Matrixexponential e^{tQ} (im Sinne obiger Reihe) ist wohldefiniert.

Beweis von Satz 2.24. Existenz: Ohne Einschränkung sei $S \subseteq \mathbb{N}$. Wir betrachten die stochastische Matrix

$$P := (p_{ij})_{i,j\in S} := \frac{1}{\lambda} Q + I.$$

Sei $Y = (Y_n, n \geqslant 0)$ eine homogene Markovkette (in diskreter Zeit) mit Ein-Schritt-Übergangsmatrix P und Startwert i unter $\mathbb{P}_i$ sowie $N = (N_t, t \geqslant 0)$ ein unabhängiger Poissonprozess mit Intensität λ. Dann ist $X = (X_t, t \geqslant 0)$ mit $X_t := Y_{N_t}$ ein Markovprozess mit Übergangswahrscheinlichkeiten

$$p_{ij}(t) = \mathbb{P}_i(X_t = j) = \sum_{n\geqslant 0} \mathbb{P}_i(N_t = n, Y_n = j) = e^{-\lambda t} \sum_{n\geqslant 0} \frac{(\lambda t)^n}{n!} (P^n)_{ij}, \quad i, j \in S.$$

Die Potenzreihe in t ist überall konvergent, da P als linearer Operator auf $\mathbb{R}^S$ eine endliche Norm hat (Bemerkung 2.25). Damit erhalten wir für $P_t = (p_{ij}(t))_{i,j\in S}$

$$P_t = e^{-\lambda t} e^{t\lambda P} = e^{t\lambda(P-I)} = e^{tQ}.$$

Durch summandenweises Ableiten erhalten wir $\frac{d}{dt} p_{ij}(t)\big|_{t=0} = q_{ij}$. Somit ist X der gewünschte Markov-Prozess.

Eindeutigkeit: Nehmen wir an, dass $\widetilde{P}_t = (\widetilde{p}_{ij}(t))_{ij}, t \geqslant 0$, die Übergangswahrscheinlichkeiten eines weiteren Markovprozesses $\widetilde{X}$ mit dem selben Generator Q sind. Dann ist (nachrechnen!)

$$\lim_{s\downarrow 0} \frac{p_{ij}(t+s) - p_{ij}(t)}{s} = (Q \cdot P_t)_{ij}$$

und analog für $\widetilde{P}_t$, sodass

$$p_{ij}(t) - \widetilde{p}_{ij}(t) = \int_0^t \big(Q(P_s - \widetilde{P}_s)\big)_{ij}\, \mathrm{d}s.$$

Wir erhalten

$$\sup_{s\leqslant t}\left\|P_s-\widetilde{P}_s\right\|_2\leqslant \sup_{s\leqslant t}\int_0^s \left\|Q(P_r-\widetilde{P}_r)\right\|_2 \mathrm{d}r\leqslant \|Q\|_2\sup_{s\leqslant t}\|P_s-\widetilde{P}_s\|_2 t\leqslant 2\lambda t\sup_{s\leqslant t}\|P_s-\widetilde{P}_s\|_2.$$

Für $t<(2\lambda)^{-1}$ folgt hieraus $P_t=\widetilde{P}_t$. Für ein allgemeines $t>0$ wählen wir $n\in\mathbb{N}$, sodass $t/n<(2\lambda)^{-1}$. Dann ist $\widetilde{P}_t=(\widetilde{P}_{t/n})^n=(P_{t/n})^n=P_t$. □

Bemerkung 2.26. Ist der Zustandsraum S endlich, dann ist die Annahme $\sup_{i\in S}|q_{ii}|<\infty$ immer erfüllt und wir haben eine Bijektion zwischen zeitstetigen Markovprozessen auf endlichen Zustandsräumen (mit in 0 differenzierbaren Übergangswahrscheinlichkeiten) und Q-Matrizen. Für unendliche Zustandsräume gilt diese Beziehung nur unter Zusatzannahmen, wie $\sup_{i\in S}|q_{ii}|<\infty$. Letztere Bedingung ist jedoch relativ stark und kann abgeschwächt. An dieser Stelle sei auf Liggett (2010) verwiesen.

Beispiel 2.27. Wir betrachten die zeitliche Entwicklung einer Population bei der jedes Individuum (oder jeder Partikel) mit einer Rate τ (d.h. nach einer $Exp(\tau)$-Wartezeit) einen Nachkommen zeugt. Mutter und Kind folgen dieser Dynamik als neue unabhängige Individuen. $X=(X_t,t\geqslant 0)$ sei die Populationsgröße zur Zeit $t\geqslant 0$ auf dem Zustandsraum $S=\mathbb{N}_0$. Die beschriebene Entwicklung der Population kann durch die Q-Matrix $Q=(q_{i,j})_{i,j\in S}$ mit

$$q_{ij}=\begin{cases}\tau i, & j=i+1,\\ -\tau i, & j=i,\\ 0, & \text{sonst}\end{cases}$$

beschrieben werden. Insbesondere gilt $\sup_i|q_{ii}|=\infty$. Dennoch können wir den Markovprozess X über die Übergangswahrscheinlichkeiten $P_t=(p_{i,j}(t))_{i,j\in S}$

$$p_{i,j}(t)=\binom{j-1}{j-i}e^{-i\tau t}(1-e^{-\tau t})^{j-i}\mathbb{1}_{\{i\leqslant j\}},\qquad t\geqslant 0, i,j\in\mathbb{N}_0$$

realisieren. Man prüfe, dass Q der Generator der Halbgruppe $(P_t,t\geqslant 0)$ ist.

3 Ergodentheorie

3.1 Stationäre und ergodische Prozesse

Erinnern wir uns, dass $X\overset{d}{=}Y$ für zwei Zufallsvariablen X,Y heißt, dass $\mathbb{P}^X=\mathbb{P}^Y$ (oder $\mathcal{L}(X)=\mathcal{L}(Y)$).

Definition 3.1. Ein stochastischer Prozess $(X_t,t\in T)$ mit $T\in\{\mathbb{R}_+,\mathbb{N}_0\}$ heißt stationär, falls

$$(X_{t_1},\dots,X_{t_n})\overset{d}{=}(X_{t_1+s},\dots,X_{t_n+s})$$

für alle $n\geqslant 1$ und $t_1,\dots,t_n,s\in T$ gilt.

Beispiel 3.2 (*Markovketten*)**.** Sei $X=(X_n,n\in\mathbb{N}_0)$ eine homogene Markovkette mit abzählbarem Zustandsraum $(S,\mathscr{P}(S))$ und Verteilungsfamilie $(\mathbb{P}_i)_{i\in S}$. Für eine Verteilung μ von X_0 betrachten wir das Maß

$$\mathbb{P}_\mu:=\sum_{i\in S}\mathbb{P}_i\mu(\{i\})=\mathbb{E}[\mathbb{P}_{X_0}(\cdot)]\quad \text{mit } X_0\sim\mu.$$

Die Anfangsverteilung μ heißt invariant, falls

$$\mathbb{P}_\mu(X_1=i)=\mathbb{P}_\mu(X_0=i)=\mu(\{i\})\quad\forall i\in S.$$

Ist $P = (p_{ij})_{i,j\in S}$ die (1-Schritt-)Übergangsmatrix und bezeichnen wir mit $\mu = (\mu_i)_{i\in S} = (\mu(\{i\}))_{i\in S}$ den Anfangsverteilungsvektor, dann besagt die Invarianz von μ gerade

$$\forall i \in S: \quad \sum_{j\in S} \mu_j p_{j,i} = \mu_i \qquad \Longleftrightarrow \qquad \mu \cdot P = \mu,$$

d.h. μ ist ein linker Eigenvektor der Matrix P zum Eigenwert 1.

Die Markoveigenschaft impliziert nun, dass X unter einem invarianten Maß stationär ist: Für alle $s \geqslant 1$ und $i_0, \dots, i_n \in S$ gilt

$$\begin{aligned}\mathbb{P}_\mu(X_{0+s} = i_0, \dots, X_{n+s} = i_n) = &\sum_{j_0,\dots,j_{s-1}\in S} \underbrace{\mu_{j_0} p_{j_0,j_1}}_{\sum_{j_0} \mu_{j_0} p_{j_0,j_1} = \mu_{j_1}} \cdots p_{j_{s-2}j_{s-1}} p_{j_{s-1}i_0} p_{i_0 i_1} \cdots p_{i_{n-1}i_n}\\ =&\mu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1}i_n} = \mathbb{P}_\mu(X_0 = i_0, \dots, X_n = i_n).\end{aligned}$$

Ein einfaches konkretes Beispiel mit $S = \{1, 2\}$ ist $P = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$ für $p \in [0,1], q = 1-p$ und $\mu = (\frac{1}{2}, \frac{1}{2})$.

Definition 3.3. Eine messbare Abbildung $T\colon \Omega \to \Omega$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt maßerhaltend oder maßtreu, falls $\mathbb{P}^T = \mathbb{P}$ gilt, d.h. $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$ für alle $A \in \mathcal{F}$.

Beispiel 3.4. Betrachte $\Omega = \{z \in \mathbb{C} : |z| = 1\}$ mit $\mathcal{F} = \mathscr{B}_\Omega$ sowie die Ω-wertige Zufallsvariable $Z = e^{iU}$ mit $U \sim \mathcal{U}([0, 2\pi])$. Dann ist für jedes $r \in \mathbb{R}$ die Rotation $T(z) = z \cdot e^{ir}, z \in \Omega$, maßerhaltend auf $(\Omega, \mathcal{F}, \mathbb{P}^Z)$.

Wir betrachten im Folgenden den zeitdiskreten Fall $T = \mathbb{N}_0$.

Lemma 3.5.

(i) Jeder $(S, \mathcal{S})$-wertige stationäre Prozess $X = (X_n, n \in \mathbb{N}_0)$ induziert eine maßerhaltende Abbildung $T = \vartheta_1$ auf dem Pfadraum $(S^{\mathbb{N}_0}, \mathcal{S}^{\otimes \mathbb{N}_0}, \mathbb{P}^X)$ durch den (Links-)Shift

$$T\big((X_n)_{n\geqslant 0}\big) = (X_n)_{n\geqslant 1}.$$

In diesem Fall gilt $X_n = \pi_0 \circ T^n(X)$, wobei π_0 die Projektion auf die 0-te Koordinate und T^n die n-fache Verknüfung von T mit sich selbst bezeichnet.

(ii) Ist Y eine Zufallsvariable und T maßerhaltend auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, dann definiert

$$X_n(\omega) := Y(T^n(\omega)), \qquad n \geqslant 0,$$

einen stationären Prozess $(X_n, n \in \mathbb{N}_0)$.

Beweis. (i) Für Zylindermengen $A = \pi^{-1}_{\mathbb{N}_0, \{0,\dots,n\}}(B_n), n \geqslant 0, B_n \in \mathcal{S}^{\otimes(n+1)}$ gilt aufgrund der Stationärität

$$\begin{aligned}\mathbb{P}^X(T^{-1}(A)) &= \mathbb{P}\big(X \in T^{-1}(A)\big)\\ &= \mathbb{P}(T \circ X \in A)\\ &= \mathbb{P}\big((X_1, \dots, X_{n+1}) \in B_n\big)\\ &= \mathbb{P}\big((X_0, \dots, X_n) \in B_n\big) = \mathbb{P}(X \in A) = \mathbb{P}^X(A).\end{aligned}$$

Also stimmen $\mathbb{P}^X(T^{-1}(\cdot))$ und $\mathbb{P}^X$ auf einem $\cap$-stabilen Erzeuger von $\mathcal{S}^{\otimes \mathbb{N}_0}$ überein und sind damit identisch.

(ii) Für $A \in \mathcal{S}^{\otimes(n+1)}$ folgt aus der Maßtreue von T

$$\mathbb{P}\big((X_m, X_{m+1}, \dots, X_{m+n}) \in A\big) =\mathbb{P}\big(\{\omega \in \Omega : (Y \circ T^m(\omega), \dots, Y \circ T^{m+n}(\omega)) \in A\}\big)$$

$$=\mathbb{P}\big(T^{-m}(\{\omega \in \Omega : (Y \circ T^0(\omega), \dots, Y \circ T^n(\omega)) \in A\})\big)$$
$$=\mathbb{P}\big(\{\omega \in \Omega : (Y \circ T^0(\omega), \dots, Y \circ T^n(\omega)) \in A\}\big)$$
$$=\mathbb{P}\big((X_0, X_1, \dots, X_n) \in A\big).$$

Damit ist $(X_n, n \in \mathbb{N}_0)$ stationär. □

Falls $(X_n, n \geqslant 0)$ stationär ist, dann sind insbesondere die Verteilungen $\mathbb{P}^{X_n}$ für alle n gleich. Die X_n sind also gleich verteilt, aber nicht notwendigerweise unabhängig. Wann können wir dennoch auf ein Gesetz der großen Zahlen hoffen?

Beispiel 3.6. Betrachte $\Omega = \{a, b, c, d, e\}$, $\mathcal{F} = \mathscr{P}(\Omega)$ und $T\colon \Omega \to \Omega$ mit

$$a \overset{T}{\mapsto} b \overset{T}{\mapsto} c \overset{T}{\mapsto} a, \quad d \overset{T}{\mapsto} e \overset{T}{\mapsto} d.$$

Dann ist T für jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ mit $\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = \mathbb{P}(\{c\}) = p_1$ und $\mathbb{P}(\{d\}) = \mathbb{P}(\{e\}) = p_2$ maßerhaltend. Was ist der Grenzwert $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i$ für $X_n = \mathbb{1}_{\{a,d\}} \circ T^n$?

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n X_i(\omega) = \begin{cases} \frac{1}{3}, & \omega \in \{a, b, c\}, \\ \frac{1}{2}, & \omega \in \{d, e\}. \end{cases}$$

Andererseits ist $\mathbb{E}[X_i] = p_1 + p_2$ im Allgemeinen weder $\frac{1}{3}$ noch $\frac{1}{2}$.

Definition 3.7. Ein Ereignis $A \in \mathcal{F}$ heißt <u>(quasi-)invariant</u> bezüglich der maßerhaltenden Transformation T auf $(\Omega, \mathcal{F}, \mathbb{P})$, falls $\mathbb{1}_{T^{-1}(A)} = \mathbb{1}_A$ $\mathbb{P}$-f.s. (äquivalent $\mathbb{P}(T^{-1}(A)\Delta A) = 0$ mit $A\Delta B := \{B \setminus A\} \cup \{A \setminus B\}$). A heißt <u>strikt invariant</u>, falls $T^{-1}(A) = A$. Die Menge aller (quasi-)invarianten Ereignisse bilden eine σ-Algebra (!) $\mathscr{I}_T$. T heißt <u>ergodisch</u>, falls $\mathscr{I}_T$ trivial ist, d.h. $\mathbb{P}(A) \in \{0, 1\}$ für alle $A \in \mathscr{I}_T$.

Ein stationärer Prozess $X = (Y \circ T^n, n \geqslant 0)$ für eine Zufallsvariable Y und eine maßerhaltende Abbildung T auf $(\Omega, \mathcal{F}, \mathbb{P})$ heißt <u>ergodisch</u>, falls T ergodisch ist.

Man beachte, dass Nullmengen und Einsmengen immer in $\mathscr{I}_T$ liegen. Ist die Abbildung T aus dem vorangegangen Beispiel ergodisch?

Lemma 3.8.

(i) Eine reelle Zufallsvariable Y ist genau dann $\mathscr{I}_T$-messbar, wenn $Y \circ T = Y$ f.s. gilt. Insbesondere ist T genau dann ergodisch, wenn für alle beschränkten Zufallsvariablen die Invarianz $Y \circ T = Y$ f.s. impliziert, dass Y f.s. konstant ist.

(ii) Für jedes invariante Ereignis $A \in \mathscr{I}_T$ existiert ein strikt invariantes Ereignis B, sodass $\mathbb{P}(A\Delta B) = 0$.

Beweis. Übung ☑. □

Bemerkung 3.9. Aufgrund von (ii) ist T ergodisch, falls $T^{-1}(A) = A$ bereits $\mathbb{P}(A) \in \{0, 1\}$ impliziert.

Beispiel 3.10.

(i) *I.i.d. Zufallsvariablen* $(X_n)_{n\geqslant 0}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ bilden einen ergodischen Prozess: Realisieren wir $(X_n)_{n\geqslant 0}$ als kanonischen Prozess auf dem Produktraum, gilt $X_n(\omega) = X_0(T^n(\omega)), \omega \in \Omega$, für den Shift T. Für jedes invariante $A \in \mathcal{F}$ mit $T^{-1}(A) = A$ gilt dann:

$$A \in \sigma((X_{n+k})_{k\geqslant 0}) \quad \text{für jedes } n \in \mathbb{N}_0,$$

denn $\omega = (\omega_n)_{n\geqslant 0} \in T^{-1}(A)$ ist äquivalent zu $(\omega_n)_{n\geqslant 1} \in A$, sodass $A = T^{-1}(A) \in \sigma(X_1, X_2, \dots)$. Wir erhalten induktiv $A = T^{-n}(A) \in \sigma(X_n, X_{n+1}, \dots)$ für jedes n.

Damit liegt A in der terminalen σ-Algebra $\bigcap_{n\geqslant 0} \sigma((X_{n+k})_{k\geqslant 0})$ und Kolmogorovs 0-1-Gesetz impliziert (dank der Unabhängigkeit) $\mathbb{P}(A) \in \{0, 1\}$.

(ii) Wir betrachten die *Rotation* um einen festen Winkel. Sei hierzu $\Omega = [0,1]$, $\mathcal{F} = \mathscr{B}_\Omega$ und $\mathbb{P}$ die Gleichverteilung auf $[0,1]$. Wir betrachten $T(\omega) = (\omega + \vartheta) \mod 1$ für alle $\omega \in \Omega$ und ein festes $\vartheta \in \mathbb{R}$. Dann ist T maßerhaltend.

(a) Sei $\vartheta = \frac{p}{q} \in \mathbb{Q}, p, q \in \mathbb{N}$: Betrachte $A = \bigcup_{k=0}^{q-1}[\frac{k}{q}, \frac{k+1/2}{q})$. Dann gilt $A = (A + \frac{p}{q}) \mod 1$ und $\mathbb{P}(A) = \frac{1}{2}$. Folglich ist T nicht ergodisch für rationale ϑ.

(b) Sei ϑ irrational: Nehmen wir an $f = \mathbb{1}_A \in L^2([0,1])$ für ein $A \in \mathcal{F}$ ist invariant: $f \circ T = f$ f.s. Wir können f als Fourierreihe entwickeln:

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi i k x} \quad (\text{in } L^2), \qquad c_k = \int_0^1 f(x) e^{-2\pi i k x} \mathrm{d}x,$$

und damit

$$f(T(x)) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi i k (x+\vartheta)} = \sum_{k=-\infty}^{\infty} c_k e^{2\pi i k \vartheta} e^{2\pi i k x}.$$

Da die Fourierkoeffizienten eindeutig bestimmt sind, impliziert $f = f \circ T$, dass

$$c_k = c_k e^{2\pi i k \vartheta} \quad \forall k \in \mathbb{Z}.$$

Ist $\vartheta \notin \mathbb{Q}$, folgt aber $e^{2\pi i k \vartheta} \neq 1$ für alle $k \in \mathbb{Z} \setminus \{0\}$ und somit $f = c_0$ f.s. Somit ist $f = \mathbb{1}_A$ f.s. konstant und T ergodisch.

3.2 Ergodensätze

Wie oben gesehen ist $X = (X_n, n \in \mathbb{N}_0)$ mit $X_n = Y \circ T^n, n \geqslant 0$, für eine Zufallsvariable Y und eine maßerhaltende Abbildung T ein stationärer Prozess. Wir wollen nun Gesetze der großen Zahlen für Partialsummen $S_n = \sum_{i=0}^{n-1} X_n$ untersuchen. Vorbereitend benötigen wir folgendes Hilfsresultat:

Lemma 3.11 (Maximal-Ergodenlemma)**.** *Sei $Y \in \mathcal{L}^1(\mathbb{P})$ und T maßerhaltend. Setzen wir $S_n = \sum_{i=0}^{n-1} Y \circ T^i, S_0 = 0$ und $M_n = \max\{S_0, S_1, \ldots, S_n\}$, dann gilt für alle $n \in \mathbb{N}$*

$$\mathbb{E}[Y \mathbb{1}_{\{M_n > 0\}}] \geqslant 0.$$

Beweis. Für $1 \leqslant j \leqslant n$ gilt $M_{n-1}(T(\omega)) \geqslant S_{j-1}(T(\omega))$ und daher

$$Y + M_n \circ T \geqslant Y + M_{n-1} \circ T \geqslant Y + S_{j-1} \circ T = S_j.$$

Auf $\{M_n > 0\}$ ist $M_n = \max\{S_1, \ldots, S_n\}$ und somit

$$Y + M_n \circ T \geqslant M_n.$$

Wir erhalten

$$\mathbb{E}\big[(Y + M_n \circ T)\mathbb{1}_{\{M_n > 0\}}\big] \geqslant \mathbb{E}\big[M_n \mathbb{1}_{\{M_n > 0\}}\big] = \mathbb{E}\big[M_n\big],$$

Da $M_n \geqslant 0$, folgt

$$\mathbb{E}\big[Y \mathbb{1}_{\{M_n > 0\}} + M_n \circ T\big] \geqslant \mathbb{E}\big[M_n\big].$$

Da T maßerhaltend ist, gilt $\mathbb{E}[M_n \circ T] = \mathbb{E}[M_n]$ und deshalb $\mathbb{E}\big[Y \mathbb{1}_{\{M_n > 0\}}\big] \geqslant 0$. □

Satz 3.12 (Ergodensatz von Birkhoff)**.** *Für eine maßerhaltende Abbildung T und $X \in \mathcal{L}^1(\mathbb{P})$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ gilt*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i = \mathbb{E}[X | \mathscr{I}_T] \quad \textit{f.s. und in } L^1\textit{-Konvergenz.}$$

Ist T ergodisch, dann gilt

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i = \mathbb{E}[X] \quad \textit{f.s. und in } L^1\textit{-Konvergenz.}$$

Beweis. Schritt 1: Setze $R_n := \frac{1}{n}\sum_{i=0}^{n-1} X \circ T^i$. Wir zeigen:

$$\overline{X} := \limsup_{n\to\infty} R_n \quad \text{und} \quad \underline{X} := \liminf_{n\to\infty} R_n \quad \text{sind } T\text{-invariant.}$$

Es gilt $\frac{n+1}{n} R_{n+1} = R_n \circ T + \frac{1}{n} X$, wobei $\frac{X}{n} \to 0$ (ω-weise). Das impliziert

$$\overline{X} := \limsup_{n\to\infty} R_n = \limsup_{n\to\infty} \frac{n+1}{n} \cdot R_{n+1} = \limsup_{n\to\infty} R_n \circ T = \overline{X} \circ T,$$

sodass $\overline{X}$ invariant ist. Analog folgt die Behauptung für $\underline{X}$.

Schritt 2: Es gilt $\overline{X} = \underline{X}$ f.s.:

Für $a < b$ betrachten wir $\mathbb{P}(\underline{X} < a, \overline{X} > b)$. Wir wollen nun das Maximal-Ergodenlemma auf $Y = (X-b)\mathbb{1}_{\{\underline{X}<a,\overline{X}>b\}}$ anwenden. Es gilt für S_n, M_n wie oben

$$\begin{aligned}\bigcup_{n\geqslant 1}\{M_n > 0\} &= \Big\{\sup_{n\geqslant 1} \frac{1}{n} S_n > 0\Big\} \\ &= \Big\{\sup_{n\geqslant 1} \frac{1}{n}\sum_{i=0}^{n-1}(X\circ T^i - b)\mathbb{1}_{\{\underline{X}\circ T^i<a,\overline{X}\circ T^i>b\}} > 0\Big\} \\ &= \Big\{\sup_{n\geqslant 1}(R_n - b)\mathbb{1}_{\{\underline{X}<a,\overline{X}>b\}} > 0\Big\} \\ &= \{\underline{X} < a, \overline{X} > b\}.\end{aligned}$$

Lemma 3.11 und dominierte Konvergenz liefern also

$$0 \leqslant \mathbb{E}\big[(X-b)\mathbb{1}_{\{\underline{X}<a,\overline{X}>b\}}\mathbb{1}_{\{M_n>0\}}\big] \to \mathbb{E}\big[(X-b)\mathbb{1}_{\{\underline{X}<a,\overline{X}>b\}}\big] \quad n\to\infty.$$

Daraus folgt $\mathbb{E}\big[X\mathbb{1}_{\{\underline{X}<a,\overline{X}>b\}}\big] \geqslant b\mathbb{P}(\underline{X} < a, \overline{X} > b)$ und analog erhalten wir für $Y = (a - X)\mathbb{1}_{\{\underline{X}<a,\overline{X}>b\}}$, dass $\mathbb{E}\big[X\mathbb{1}_{\{\underline{X}<a,\overline{X}>b\}}\big] \leqslant a\mathbb{P}(\underline{X} < a, \overline{X} > b)$. Es folgt

$$0 \leqslant (a-b)\mathbb{P}(\underline{X} < a, \overline{X} > b).$$

Daher muss $\mathbb{P}(\underline{X} < a, \overline{X} > b) = 0$ für alle $a < b, a, b \in \mathbb{Q}$, gelten. Wir erhalten $\mathbb{P}(\underline{X} < \overline{X}) = 0$.

Schritt 3: $(R_n)_{n\geqslant 1}$ ist gleichgradig integrierbar (zur Erinnerung: $\lim_{C\to\infty}\sup_n \mathbb{E}[|R_n|\mathbb{1}_{\{|R_n|>C\}}] = 0$, Anhang A). Dies ist äquivalent dazu, dass $\sup_n \mathbb{E}[|R_n|] < \infty$ und

$$\forall \varepsilon > 0\, \exists \delta_\varepsilon > 0 : \forall B \in \mathcal{F} \text{ mit } \mathbb{P}(B) < \delta_\varepsilon \text{ gilt } \sup_n \mathbb{E}[|R_n|\mathbb{1}_B] < \varepsilon$$

(Übung ☑).

Wegen

$$\lim_{C\to\infty}\sup_{i\in\mathbb{N}_0}\mathbb{E}\big[|X\circ T^i|\mathbb{1}_{\{|X\circ T^i|>C\}}\big] = \lim_{C\to\infty}\sup_{i\in\mathbb{N}}\mathbb{E}\big[|X|\mathbb{1}_{\{|X|>C\}}\big] = 0,$$

existiert für jedes $\varepsilon > 0$ ein $\delta > 0$, sodass $\sup_i \mathbb{E}[|X\circ T^i|\mathbb{1}_B] < \varepsilon$ für alle $B \in \mathcal{F}$ mit $\mathbb{P}(B) < \delta$ gilt. Wir erhalten also für jedes $B \in \mathcal{F}$ mit $\mathbb{P}(B) < \delta$:

$$\sup_n \mathbb{E}[|R_n|\mathbb{1}_B] \leqslant \sup_n \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\big[|X\circ T^i|\mathbb{1}_B\big] < \varepsilon.$$

Außerdem folgt aus der Dreiecksungleichung und der Maßtreue von T, dass

$$\sup_n \mathbb{E}[|R_n|] \leqslant \sup_n \mathbb{E}[|X\circ T^n|] = \mathbb{E}[|X|] < \infty.$$

Schritt 4: Wir zeigen die f.s. und L^1-Konvergenz gegen $\underline{X} = \overline{X} = \mathbb{E}[X|\mathscr{I}_T]$ f.s.

Aus Schritt 2 und 3 folgt, dass $R_n \to \underline{X} = \overline{X}$ f.s. und in $L^1(\mathbb{P})$. Damit folgt entlang einer Teilfolge $(R_{n_m})_{m\geqslant 1}$

$$\mathbb{E}[R_{n_m}|\mathscr{I}_T] \xrightarrow[m\to\infty]{f.s.} \mathbb{E}[\overline{X}|\mathscr{I}_T] = \overline{X} \quad \text{f.s.}$$

Nun ist $\mathbb{E}[X\circ T|\mathscr{I}_T] = \mathbb{E}[X|\mathscr{I}_T]$, denn für alle $A \in \mathscr{I}_T$ gilt

$$\mathbb{E}\big[X\mathbb{1}_A\big] = \mathbb{E}\big[(X\circ T)(\mathbb{1}_A \circ T)\big] = \mathbb{E}\big[(X\circ T)\mathbb{1}_{T^{-1}(A)}\big] = \mathbb{E}\big[(X\circ T)\mathbb{1}_A\big].$$

Wir schließen hieraus

$$\mathbb{E}[R_n|\mathscr{I}_T] = \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[X\circ T^i|\mathscr{I}_T] = \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[X|\mathscr{I}_T] = \mathbb{E}[X|\mathscr{I}_T],$$

also folgt $\overline{X} = \mathbb{E}[X|\mathscr{I}_T]$ f.s.

Schritt 5: Ist T ergodisch, so ist jede $\mathscr{I}_T$-messbare Abbildung f.s. konstant, woraus $\mathbb{E}[X|\mathscr{I}_T] = \mathbb{E}[X]$ f.s. folgt. □

Satz 3.13 (L^p-Ergodensatz von Neumann)**.** *Für eine maßerhaltende Abbildung T und $X \in \mathcal{L}^p(\mathbb{P}), p \geqslant 1,$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ gilt*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} X\circ T^i = \mathbb{E}[X|\mathscr{I}_T] \quad \textit{f.s. und in } L^p\textit{-Konvergenz.}$$

Beweis. Übung ☑. □

Beispiel 3.14. Ist $(X_n, n\geqslant 0)$ ein ergodischer Prozess in $\mathcal{L}^1(\mathbb{P})$, dann gilt $\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}X_k = \mathbb{E}[X_0]$ f.s. und in L^1. Wir erhalten insbesondere Kolmogorovs starkes Gesetz der großen Zahlen für i.i.d. $(X_n)_{n\geqslant 0}$ aus $\mathcal{L}^1(\mathbb{P})$.

3.3 Anwendung auf Markovketten

Wann ist eine Markovkette stationär oder ergodisch? Wir betrachten wieder eine homogene (irreduzible) Markovkette $X = (X_n, n \in \mathbb{N}_0)$ mit abzählbarem Zustandsraum S mit Übergangsmatrix P. Dabei sei X als kanonischer Prozess auf dem Pfadraum $(S^{\mathbb{N}_0}, \mathcal{S}^{\otimes \mathbb{N}_0})$ gegeben, d.h. X_n ist gerade die Projektion auf die n-te Koordinate.

Da Markovketten stationär sind, falls sie in einer invarianten Verteilung μ ($\Leftrightarrow \mu P = \mu$ in Matrixnotation) starten, suchen wir also zunächst nach notwendigen und hinreichenden Bedingungen für die Existenz eines invarianten Maßes. Für jeden Zustand $y \in S$ erinnern wir uns an die Stoppzeit

$$T_y := T_y^1 := \inf\{n > 0 : X_n = y\}.$$

y hieß rekurrent, falls $\mathbb{P}_y(T_y < \infty) = 1$. Diese Eigenschaft verstärken wir nun:

Definition 3.15. Ein rekurrenter Zustand $x \in S$ heißt positiv rekurrent, falls $\mathbb{E}_x[T_x] < \infty$, und andernfalls nullrekurrent.

Beispiel 3.16. Auf $S = \{1,2\}$ sei $(X_n, n \geqslant 0)$ eine Markovkette mit Übergangsmatrix $\begin{pmatrix} p & q \\ q & p\end{pmatrix}$ mit $p, q > 0$ und $p + q = 1$. Dann sind beide Zustände positiv rekurrent:

$$\begin{aligned}\mathbb{E}_1[T_1] &= 1p + 2q^2 + 3pq^2 + \dots + (k+2)p^k q^2 + \dots \\ &= p + (1 - 2p + p^2)\sum_{k\geqslant 0}(k+2)p^k \\ &= 2p^0 + (1 - 4 + 3)p + \sum_{k\geqslant 2}(k + 2 - 2(k+1) + k)p^k = 2 < \infty.\end{aligned}$$

Satz 3.17. *Sei $x \in S$ positiv rekurrent. Dann existiert ein invariantes Maß μ gegeben durch*

$$\mu(\{y\}) := \frac{\mathbb{E}_x[\sum_{n=0}^{T_x-1} \mathbb{1}_{\{X_n=y\}}]}{\mathbb{E}_x[T_x]} = \frac{\sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x > n)}{\mathbb{E}_x[T_x]}, \quad y \in S.$$

μ ist also der erwartete Anteil der Besuche von y auf einer Exkursion von x nach x.

Beweis. Da x positiv rekurrent ist, gilt $\mathbb{E}_x[T_x] < \infty$ und

$$\sum_{y \in S} \mu(\{y\}) = \frac{\mathbb{E}_x[\sum_{n=0}^{T_x-1} \sum_{y\in S} \mathbb{1}_{\{X_n=y\}}]}{\mathbb{E}_x[T_x]} = 1.$$

Setzen wir nun $q_{n,x}(y) := \mathbb{P}_x(X_n = y, T_x > n)$, so ist $\mu P = \mu$ äquivalent zu

$$\forall y \in S : \quad \sum_{z\in S} \sum_{n \geqslant 0} q_{n,x}(z) p_{zy} = \sum_{n\geqslant 0} q_{n,x}(y).$$

Sei zunächst $y \neq x$, dann

$$\sum_{z\in S} q_{n,x}(z)p_{zy} = \sum_{z\in S} \mathbb{P}_x(\underbrace{X_n = z, T_x > n}_{\in \mathcal{F}_n^X}, X_{n+1} = y) = \mathbb{P}_x(T_x > n+1, X_{n+1} = y) = q_{n+1,x}(y).$$

Da $q_{0,x}(y) = 0$ wegen $y \neq x$, erhalten wir

$$\sum_{z\in S} \sum_{n\geqslant 0} q_{n,x}(z)p_{zy} = \sum_{n\geqslant 1} q_{n,x}(y) = \sum_{n\geqslant 0} q_{n,x}(y).$$

Andererseits gilt für $y = x$

$$\sum_{z\in S} q_{n,x}(z)p_{zx} = \sum_{z\in S} \mathbb{P}_x(X_n = z, T_x > n, X_{n+1} = x) = \mathbb{P}_x(T_x = n+1)$$

und damit

$$\sum_{z\in S} \sum_{n\geqslant 0} q_{n,x}(z)p_{zy} = \sum_{n\geqslant 1} \mathbb{P}_x(T_x = n) = 1 = \mathbb{E}_x\Big[\sum_{n=0}^{T_x-1} \mathbb{1}_{\{X_n=x\}}\Big] = \sum_{n\geqslant 0} q_{n,x}(x). \qquad \square$$

Lemma 3.18. *Ist X irreduzibel, so hat X höchstens eine invariante Verteilung.*

Beweis. Wir definieren die stochastische Matrix $\widetilde{P} = (\widetilde{p}_{xy})_{x,y\in S}$ via $\widetilde{p}_{xy} := \sum_{n=1}^{\infty} 2^{-n}(P^n)_{x,y}$ für alle $x, y \in S$. Für jede invariante Verteilung μ von X gilt dann $\mu\widetilde{P} = \mu$. Da X irreduzibel ist, folgt außerdem $\widetilde{p}_{xy} > 0$ für alle $x, y \in S$.

Wir führen nun einen Widerspruchsbeweis. Seien $\mu = (\mu_x)_{x\in S}$ und $\nu = (\nu_x)_{x\in S}$ zwei invariante Verteilungen mit $\mu \neq \nu$. Dann existieren zwei Zustände $x_1, x_2 \in S$ mit $\mu_{x_1} > \nu_{x_1}$ und $\mu_{x_2} < \nu_{x_2}$. Damit gilt

$$\big|(\mu_{x_1} - \nu_{x_1})\widetilde{p}_{x_1 y} + (\mu_{x_2} - \nu_{x_2})\widetilde{p}_{x_2 y}\big| < |\mu_{x_1} - \nu_{x_1}|\widetilde{p}_{x_1 y} + |\mu_{x_2} - \nu_{x_2}|\widetilde{p}_{x_2 y}$$

Da μ und ν invariant sind, folgt

$$\begin{aligned}
\sum_{y\in S} |\mu_y - \nu_y| &= \sum_{y\in S} \big|\big((\mu-\nu)\widetilde{P}\big)_y\big| \\
&= \sum_{y\in S} \big|\sum_{x\in S} (\mu_x - \nu_x)\widetilde{p}_{x,y}\big| \\
&< \sum_{y\in S}\sum_{x\in S} |\mu_x - \nu_x|\widetilde{p}_{x,y} = \sum_{x\in S} |\mu_x - \nu_x|.
\end{aligned}$$

Dies ist ein Widerspruch. $\square$

Haben wir eine invariante Verteilung, so liefert uns folgender Satz eine (erstaunlich schwache) Bedingung für die Ergodizität von X.

Satz 3.19. *Ist X eine irreduzible Markovkette mit einem rekurrenten Zustand und einer invarianten Verteilung μ, dann ist X unter $\mathbb{P}_\mu$ ergodisch und es gilt für alle $x \in S$*

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} \overset{n\to\infty}{\longrightarrow} \mu(\{x\}), \qquad \mathbb{P}_\mu\text{-f.s.}$$

Beweis. Unter $\mathbb{P}_\mu$ ist X stationär und es gilt $(X_k)_{k\geqslant n} = \vartheta^n(X)$ für den Shift ϑ. Für $A \in \mathscr{I}_\vartheta$ und jede endliche Stoppzeit τ gilt

$$\mathbb{P}_\mu(X \in A|\mathcal{F}_\tau) = \mathbb{P}_{X_\tau}(X \in A), \tag{5}$$

denn für jedes $B \in \mathcal{F}_\tau$ erhalten wir aus $\{X \in A\} = \{X \in \vartheta^{-n}(A)\}$ $\mathbb{P}$-f.s. und der starken Markoveigenschaft

$$\begin{aligned}
\mathbb{E}_\mu[\mathbb{1}_{\{X\in B\}}\mathbb{1}_{\{X\in A\}}] &= \sum_{n=0}^{\infty}\sum_{x\in S}\mathbb{P}_\mu\big(X \in B, \tau = n, X_n = x, X \in A\big)\\
&= \sum_{n=0}^{\infty}\sum_{x\in S}\mathbb{P}_\mu\big(\underbrace{X \in B, \tau = n}_{\in \mathcal{F}_n^X}, X_n = x, \vartheta^n(X) \in A\big)\\
&\overset{Markov}{=} \sum_{n=0}^{\infty}\sum_{x\in S}\mathbb{P}_\mu\big(X \in B, \tau = n, X_n = x\big)\mathbb{P}_x(X \in A)\\
&= \mathbb{E}_\mu\big[\mathbb{1}_{\{X\in B\}}\mathbb{P}_{X_\tau}(X \in A)\big].
\end{aligned}$$

Da X eine irreduzible Markovkette mit einem rekurrenten Zustand ist, ist jedes $x \in S$ rekurrent (Satz 2.18). Wir können also für jedes $x \in S$ die Stoppzeit $\tau = T_x$ wählen. Dann gilt $T_x < \infty$ f.s. und aus (5) ergibt sich

$$\mathbb{P}_\mu(X \in A) = \mathbb{E}_\mu[\mathbb{P}_\mu(X \in A|\mathcal{F}_{T_x})] = \mathbb{P}_x(X \in A) \qquad \forall x \in S.$$

Insbesondere ist $\mathbb{P}_{X_n}(X \in A) = \mathbb{P}_\mu(X \in A)$ f.s. und mit $\tau = n$ in (5)

$$\mathbb{P}_\mu(X \in A|X_0, \dots, X_n) = \mathbb{P}_{X_n}(X \in A) = \mathbb{P}_\mu(X \in A) \quad \text{f.s.}$$

Mit dem Martingalkonvergenzsatz ergibt sich für jedes $A \in \mathscr{I}_\vartheta \subseteq \sigma(X_k, k \geqslant 0)$

$$\mathbb{P}_\mu(X \in A|X_0, \dots, X_n) \overset{n\to\infty}{\longrightarrow} \mathbb{P}_\mu(X \in A|\sigma(X_k, k \geqslant 0)) = \mathbb{1}_{\{X\in A\}} \quad \mathbb{P}_\mu - \text{f.s.}$$

Damit folgt $\mathbb{P}_\mu(X \in A) \in \{0, 1\}$, sodass X unter $\mathbb{P}_\mu$ ergodisch ist (genauer ist $(\pi_0 \circ \vartheta^n)_{n\geqslant 1}$ ergodisch auf $(S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}}, \mathbb{P}_\mu^x)$). Dann ist auch $(\mathbb{1}_{\{x\}}(X_n))_{n\geqslant 0}$ ergodisch die Konvergenzaussage für $\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{1}_{\{X_k=x\}}$ ergibt sich aus Birkhoffs Ergodensatz. □

Unter den Voraussetzungen des vorherigen Satzes ist also $\mu(\{x\})$ die mittlere Aufenthaltsdauer von X in x. Im Vergleich zu Satz 3.17 ergibt sich, dass für einen ergodischen Markovprozess das Ortsmittel gleich dem Zeitmittel ist.

Satz 3.20. *Besitzt eine irreduzible Markovkette X eine invariante Verteilung μ, dann sind alle Zustände positiv rekurrent und es gilt:*

$$\mu(\{y\}) = \frac{1}{\mathbb{E}_y[T_y]} > 0, \qquad y \in S.$$

Beweis. Da μ ein Wahrscheinlichkeitsmaß ist, existiert ein $x \in S$ mit $\mu(\{x\}) > 0$. Da X irreduzibel ist, existiert zudem für jedes $y \in S$ ein $n = n(y) \in \mathbb{N}$ mit $(P^n)_{x,y} > 0$. Aus der Stationarität unter μ folgt dann für alle $y \in S$:

$$\mu(\{y\}) = \mathbb{P}_\mu(X_0 = y) = \mathbb{P}_\mu(X_n = y) = \sum_{x\in S} \mu(\{x\})(P^n)_{x,y} > 0.$$

Nun ist ein $y \in S$ genau dann rekurrent, wenn $\mathbb{E}_x[\sum_{k\geqslant 1} \mathbb{1}_{\{X_k=y\}}] = \infty$ (Satz 2.17) für ein $x \in S$. Gäbe es keinen rekurrenten Zustand, würde $\sum_{k\geqslant 0}(P^k)_{x,y}$ für alle $x, y \in S$ endlich sein, d.h. insbesondere würde $\lim_{k\to\infty}(P^k)_{x,y} = 0$ gelten. Mit dominierter Konvergenz folgt der Widerspruch:

$$\mu(y) = (\mu P^k)_y = \sum_{x\in S} \mu(\{x\})(P^k)_{x,y} \stackrel{k\to\infty}{\longrightarrow} 0 \neq \mu(y).$$

Die Markovkette ist somit nicht transient und es gibt (mindestens) einen rekurrenten Zustand. Aus Satz 3.19 folgt nun, dass X ergodisch ist.

Wir betrachten nun die Folge $(T_y \circ \vartheta_{T_y^l})_{l\geqslant 0}$, wobei $T_y^k = \sum_{l=0}^{k-1} T_y \circ \vartheta_{T_y^l}$ für die k-te Eintrittszeit und $\mathbb{P}_y(T_y^k < \infty) = 1$ (Satz 2.15) gelten. Somit erhalten wir für $k \to \infty$ (also auch $T_y^k \to \infty$)

$$\frac{1}{k}\sum_{l=0}^{k-1} T_y \circ \vartheta_{T_y^l} = \frac{T_y^k}{k} = \Big(\frac{1}{T_y^k}\sum_{n=1}^{T_y^k} \mathbb{1}_{\{y\}}(X_n)\Big)^{-1} \to \frac{1}{\mu(\{y\})} \quad \mathbb{P}_\mu\text{-f.s.}$$

Wegen $\mu(\{y\}) > 0$ gilt dann auch $\lim_{k\to\infty} \frac{1}{k}\sum_{l=0}^{k-1} T_y \circ \vartheta_{T_y^l} = \mu(\{y\})^{-1}$ $\mathbb{P}_y$-f.s. Da $(T_y \circ \vartheta_{T_y^l})_{l\geqslant 0}$ unter $\mathbb{P}_y$ i.i.d. ist, folgt aus dem starken Gesetz der großen Zahlen, dass $\mathbb{E}_y[T_y] < \infty$ (weil sonst $\lim_{k\to\infty} \frac{1}{k}\sum_{l=0}^{k-1} T_y \circ \vartheta_{T_y^l} = \infty$ f.s. gelten würde) für alle $y \in S$. Es sind also alle Zustände positiv rekurrent. Zudem gilt

$$\frac{1}{\mu(\{y\})} = \lim_{k\to\infty} \frac{1}{k}\sum_{l=0}^{k-1} T_y \circ \vartheta_{T_y^l} = \mathbb{E}_y[T_y] \quad \mathbb{P}_y - f.s.$$

Da die linke und die rechte Seite deterministisch sind, muss also $\mu(\{y\}) = \frac{1}{\mathbb{E}_y[T_y]}$ gelten. □

Fassen wir die vorangegangen Sätze zusammen, so erhalten wir:

Korollar 3.21. *Es sei $X = (X_n, n \geqslant 0)$ eine homogene, irreduzible Markovkette. Dann sind folgende Bedingungen äquivalent:*

(i) Es gibt einen positiv rekurrenten Zustand.

(ii) Alle Zustände sind positiv rekurrent.

(iii) Es existiert eine invariante Verteilung μ.

Ist eine und somit alle Bedingungen erfüllt, dann ist X ergodisch und die invariante Verteilung ist eindeutig.

Wir schließen dieses Kapitel mit einem Resultat ab, welches einen zusätzlichen Zusammenhang zwischen einer irreduziblen, positiv rekurrenten Markovkette und ihrer invarianten Verteilung offenbart: Unter der zusätzlichen Bedingung der Aperiodizität konvergieren die Verteilungen $\mathbb{P}_\mu^{X_n}$ für $n \to \infty$ gegen das invariante Maß π und zwar unabhängig von der Startverteilung μ. Den Abstand zweier Maße ν_1, ν_2 auf einem gemeinsamen messbaren Raum $(S, \mathcal{S})$ werden wir bzgl. des Totalvariationsabstands messen:

$$\|\nu_1 - \nu_2\|_{TV} := \sup_{A\in\mathcal{S}} |\nu_1(A) - \nu_2(A)| = \frac{1}{2}\sum_{y\in S} |\nu_1(\{y\}) - \nu_2(\{y\})|,$$

wobei die zweite Gleichheit nur für abzählbare S gilt. Auf abzählbaren Räumen kann die schwache Konvergenz durch den Abstand $d(\nu_1, \nu_2) := \|\nu_1 - \nu_2\|_{TV}$ metrisiert werden (Übung ☑).

Definition 3.22. Ist P die Übergangsmatrix einer homogenen Markovkette X auf S, so definieren wir durch

$$d_x = ggT(N_x) \quad \text{mit} \quad N_x := \{n \in \mathbb{N}_0 : (P^n)_{x,x} > 0\}$$

die Periode eines Punktes $x \in S$. Gilt $d_x = d_y$ für alle $x, y \in E$, heißt $d := d_x$ die Periode von X. Im Fall $d = 1$ heißt X aperiodisch.

Dass diese Definition tatsächlich eine Periodizität der Besuche im Zustand x impliziert, zeigt folgendes Lemma:

Lemma 3.23. *Für jedes $x \in S$ existiert ein $n_x \in \mathbb{N}$, sodass $(P^{nd_x})_{x,x} > 0$ für alle $n \geqslant n_x$ gilt.*

Beweis. Seien $k_1, \dots, k_r \in N_x$ mit $ggT(k_1, \dots, k_r) = d_x$. Dann existieren $\alpha_1, \dots, \alpha_r \in \mathbb{Z}$ mit $\sum_{i=1}^r \alpha_i k_i = d_x$ (Lemma von Bézout). Da N_x abgeschlossen unter Addition ist, gilt

$$N := \sum_{i=1}^r |\alpha_i| k_i \in N_x \quad \text{und} \quad N + d_x = \sum_{i=1}^r \big(|\alpha_i| + \alpha_i\big) k_i \in N_x.$$

Setzen wir nun $n_x := (\frac{N}{d_x})^2$, so gibt es für alle $n \geqslant n_x$ eine Zerlegung $n - (\frac{N}{d_x})^2 = \frac{kN}{d_x} + l$ mit $k, l \in \mathbb{N}_0$, $0 \leqslant l < \frac{N}{d_x}$. Wir erhalten:

$$nd_x = \frac{N^2}{d_x} + kN + d_x l = N(\frac{N}{d_x} + k - l) + (N + d_x) l \in N_x. \qquad \square$$

Satz 3.24 (Konvergenzsatz für Markovketten)**.** *Sei X eine homogene, irreduzible, positiv rekurrente Markovkette auf S. Dann existiert eine invariante Verteilung π und folgende Aussagen sind äquivalent:*

(i) X ist aperiodisch.

(ii) Für ein $x \in S$ gilt $\|\mathbb{P}_x^{X_n} - \pi\|_{TV} \to 0$ für $n \to \infty$

(iii) Für jedes Wahrscheinlichkeitsmaß μ auf S gilt $\|\mathbb{P}_\mu^{X_n} - \pi\|_{TV} \to 0$ für $n \to \infty$.

Beweis. (iii)$\Longrightarrow$(ii): Wähle $\mu = \delta_x$.

(ii)$\Longrightarrow$(i): Nehmen wir an, (i) gilt nicht und x hat eine Periode $d_x \geqslant 2$. Für ein $n \in \mathbb{N}$, welches kein Vielfaches von d ist, folgt dann

$$\|\mathbb{P}_x^{X_n} - \pi\|_{TV} \geqslant |(P^n)_{x,x} - \pi(\{x\})| = \pi(\{x\}) \quad \text{für } n \notin N_x.$$

Da nach Satz 3.20 $\pi(\{x\}) > 0$ für jedes $x \in S$ gilt, folgt $\limsup_{n\to\infty} \|\mathbb{P}_x^{X_n} - \pi\|_{TV} > 0$ im Widerspruch zu (ii). Somit gilt $d_x = 1$ und da für irreduzible Ketten $d_x = d$ für alle $x \in S$ gilt (Übung ☑) folgt (i).

(i)$\Longrightarrow$(iii): Da π eine invariante Verteilung ist, gilt (in Vektorschreibweise $\pi_y = \pi(\{y\})$)

$$\begin{aligned}
\|\mathbb{P}_\mu^{X_n} - \pi\|_{TV} = \|\mathbb{P}_\mu^{X_n} - \mathbb{P}_\pi^{X_n}\|_{TV} &= \Big\| \sum_x \mu_x \mathbb{P}_x^{X_n} - \sum_y \pi_y \mathbb{P}_y^{X_n} \Big\|_{TV} \\
&= \Big\| \sum_{x,y} \mu_x \pi_y \big(\mathbb{P}_x^{X_n} - \mathbb{P}_y^{X_n}\big) \Big\|_{TV} \leqslant \sum_{x,y \in S} \mu_x \pi_y \|\mathbb{P}_x^{X_n} - \mathbb{P}_y^{X_n}\|_{TV}.
\end{aligned}$$

Wegen dominierter Konvergenz genügt es also $\|\mathbb{P}_x^{X_n} - \mathbb{P}_y^{X_n}\|_{TV} \to 0$ für alle $x, y \in S$ zu zeigen. Weiter gilt für jeden bivariaten Prozess $Z := ((X_n, Y_n), n \in \mathbb{N}_0)$ derart, dass $(X_n, n \geqslant 1)$ und $(Y_n, n \geqslant 0)$ Markovketten auf S sind mit der selben Übergangsmatrix P und Startwerten $X_0 = x$ und $Y_0 = y$ unter $\mathbb{P}_{(x,y)}$:

$$\|\mathbb{P}_x^{X_n} - \mathbb{P}_y^{X_n}\|_{TV} = \sum_z \big|\mathbb{P}_{(x,y)}(X_n = z) - \mathbb{P}_{(x,y)}(Y_n = z)\big|$$

$$
\begin{aligned}
&= \sum_{z} \Big| \sum_{z'} \big(\mathbb{P}_{(x,y)}(X_n = z, Y_n = z') - \mathbb{P}_{(x,y)}(X_n = z', Y_n = z) \big) \Big| \\
&\leqslant 2 \sum_{z \neq z'} \mathbb{P}_{(x,y)}(X_n = z, Y_n = z') \\
&= 2\mathbb{P}_{(x,y)}(X_n \neq Y_n).
\end{aligned}
$$

Seien nun $\widetilde{X}$ und $\widetilde{Y}$ voneinander unabhängige, homogene Markovketten (beide) mit der Übergangsmatrix P (konstruiert auf einem möglicherweise erweiterten Wahrscheinlichkeitsraum). Wir definieren die bivariate Markovkette $\widetilde{Z} := ((\widetilde{X}_n, \widetilde{Y}_n), n \in \mathbb{N}_0)$ deren Übergangswahrscheinlichkeiten durch

$$
\widetilde{p}_{(x_1,y_1),(x_2,y_2)} = p_{x_1,x_2} \cdot p_{y_1,y_2}
$$

gegeben sind. Da $\widetilde{X}$ und $\widetilde{Y}$ aperiodisch sind, folgt aus Lemma 3.23, dass für alle $(x_1, x_2), (y_1, y_2) \in S^2$ ein $n_0 \geqslant 1$ existiert, sodass

$$
(P^n)_{x_1,x_2} > 0 \quad \text{und} \quad (P^n)_{y_1,y_2} > 0 \qquad \text{für alle } n \geqslant n_0.
$$

Damit gilt auch $(\widetilde{P}^n)_{(x_1,y_1),(x_2,y_2)} > 0$. Folglich ist $\widetilde{Z}$ irreduzibel. Zudem ist das Produktmaß $\pi \otimes \pi$ ein invariantes Maß für $\widetilde{Z}$. Aus Korollar 3.21 folgt somit, dass $\widetilde{Z}$ auch positiv rekurrent ist. Insbesondere sind alle Zustände rekurrent und für das erste Eintreffen von $\widetilde{Z}$ auf der Diagonalen $D = \{(x, x) : x \in S\}$

$$
\tau := \inf \big\{ n \in \mathbb{N}_0 : \widetilde{Z} \in D \big\} = \inf \big\{ n \in \mathbb{N}_0 : \widetilde{X}_n = \widetilde{Y}_n \big\}
$$

folgt $\mathbb{P}_{(x,y)}(\tau < \infty) \geqslant \mathbb{P}_{(x,y)}(T^1_{(z,z)} < \infty) = 1$ für ein beliebiges Element $z \in S$, vgl. Satz 2.18.

Wir definieren nun einen weiteren Prozess $Z = ((X_n, Y_n), n \in \mathbb{N}_0)$ von $\widetilde{X}$ und $\widetilde{Y}$ via

$$
X_n := \widetilde{X}_n \qquad \text{und} \qquad Y_n := \begin{cases} \widetilde{Y}_n, & \text{falls } n < \tau, \\ \widetilde{X}_n, & \text{falls } n \geqslant \tau. \end{cases}
$$

Dann sind auch $(X_n, n \geqslant 0)$ und $(Y_n, n \geqslant 0)$ homogene Markovketten mit der Übergangsmatrix P (Übung ☑) und es gilt wegen $\mathbb{P}_{(x,y)}(\tau < \infty) = 1$ für $n \to \infty$:

$$
\|\mathbb{P}_x^{X_n} - \mathbb{P}_y^{X_n}\|_{TV} \leqslant 2\mathbb{P}_{(x,y)}(X_n \neq Y_n) \leqslant 2\mathbb{P}_{(x,y)}(\tau > n) \to 0. \qquad \square
$$

Bemerkung 3.25. Obiger Satz, häufig ebenfalls als ein Ergodensatz bezeichnet, wird insbesondere zur Simulation von analytisch schwer zugänglichen Wahrscheinlichkeitsverteilungen genutzt, bspw. im Ising-Modell. Die sich daraus ergebenden numerischen Verfahren werden *Markovketten Monte Carlo Methoden* (engl. Markov Chain Monte Carlo, MCMC) genannt.

Beispiel 3.26. Wir betrachten erneut das Ehrenfestmodell aus Beispiel 1.10: Die homogene Markovkette $X = (X_n : n \geqslant 0)$ auf dem Zustandsraum $S = \{0, \ldots, N\}$ für ein festes $N \in \mathbb{N}$ ist gegeben durch die Übergangsmatrix $P = (p_{i,j})_{i,j=1,\ldots,N}$ mit $p_{i,j} = \frac{N-i}{N}\mathbb{1}_{j=i+1} + \frac{i}{N}\mathbb{1}_{j=i-1}$. Dann ist X irreduzibel, rekurrent und besitzt die invariante Verteilung $\pi = Bin(N, \frac{1}{2})$ (Übung ☑). Somit ist X sogar positiv rekurrent. Zudem besitzt X die Periode 2, da X_n in jedem Schritt zwischen einem gerade und einem ungeraden Zustand wechselt. Obiger Konvergenzsatz kann also nicht angewendet werden.

Wir modifizieren das Modell nun leicht und erlauben $X_n = X_{n+1}$ mit Wahrscheinlichkeit $\frac{1}{2}$. Wir betrachten also $Y = (Y_n : n \geqslant 0)$ mit Übergangsmatrix $Q = (q_{i,j})_{i,j=0,\ldots,N}$ mit $p_{i,j} = \frac{N-i}{2N}\mathbb{1}_{j=i+1} + \frac{i}{2N}\mathbb{1}_{j=i-1} + \frac{1}{2}\mathbb{1}_{j=i}$. Dann ist die Markovkette Y irreduzibel, positiv rekurrent, hat die Periode 1 und die selbe invariante Verteilung wie X. Dann liefert Satz 3.24 $\|\mathbb{P}_\mu^{X_n} - \pi\|_{TV} \to 0$ für $n \to \infty$, wie in Abbildung 4 deutlich wird.

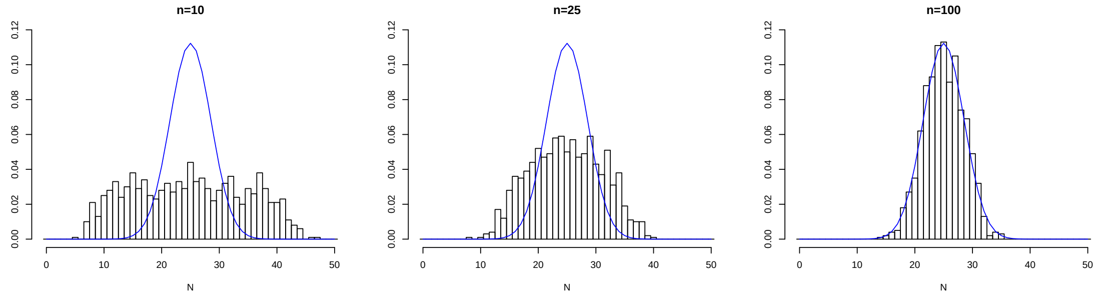

Abbildung 4: Relative Häufigkeiten (Histrogramme) aus 1000 unabhängigen Simulationsdurchläufen des modifizierten Ehrenfestmodells für die Zeitpunkte $n \in \{10, 25, 100\}$ und die uniforme Startverteilung $\mu = \mathcal{U}(\{0, \dots, N\})$ sowie die invariante Verteilung (in blau).

4 Brownsche Bewegung und Martingale in stetiger Zeit

4.1 Existenz der Brownschen Bewegung

In Beispiel 2.9 hatten wir bereits einen (Markov-)Prozess $X = (X_t, t \geqslant 0)$ diskutiert mit unabhängigen, stationären und normalverteilten Zuwächsen. Genauer galt $X_t - X_s \sim \mathcal{N}(0, t-s)$ für alle $0 \leqslant s < t$ (unter $\mathbb{P} = \mathbb{P}_0$). Häufig ist man aber nicht nur an Verteilungseigenschaften interessiert, sondern möchte auch das Verhalten der Pfade $t \mapsto X_t$ beschreiben, bspw. über das Funktional $F(X) := \sup_{t \in [0,1]} X_t$. A priori muss aber $F(X)$ nicht mal eine Zufallsvariable sein. Wir werden in diesem Kapitel zeigen, dass X eine Modifikation B (also $\mathbb{P}(X_t = B_t) = 1$ für alle $t \geqslant 0$) mit stetigen Pfaden besitzt, was insbesondere die Messbarkeit von $F(X)$ und anderen interessanten Funktionalen impliziert. Der resultierende Prozess ist das zentrale Objekt in der Theorie stochastischer Prozesse: die Brownsche Bewegung. Abbildung 5 zeigt fünf Pfade der Brownschen Bewegung.

Definition 4.1. Ein reellwertiger stochastischer Prozess $B = (B_t, t \geqslant 0)$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt (standard) Brownsche Bewegung, falls

(i) $B_0 = 0$ f.s.,

(ii) B besitzt unabhängige Zuwächse, d.h. für alle $n \in \mathbb{N}$, $0 \leqslant t_0 < \cdots < t_n$ sind $B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$ unabhängig,

(iii) B besitzt stationäre Zuwächse, d.h. für alle $0 \leqslant s < t$ gilt $B_t - B_s \overset{d}{=} B_{t-s}$,

(iv) $B_t \sim \mathcal{N}(0, t)$ für alle $t > 0$,

(v) Für $\mathbb{P}$-f.a. $\omega \in \Omega$ ist $t \mapsto B_t(\omega)$ stetig.

Wir haben bereits gesehen, dass ein Prozess mit den Eigenschaften (i)-(iv) existiert, wissen aber nicht, ob dieser eine stetige Modifikation besitzt. Um das zu beweisen, werden wir das Kolmogorov-Chentsov-Kriterium verwenden, welches uns über die Stetigkeit hinaus sogar Hölder-Regularität liefert.

Definition 4.2. Für $\gamma \in (0, 1]$ und $I \subseteq \mathbb{R}$ heißt eine Funktion $f : I \to \mathbb{R}^d$ lokal Hölder-stetig der Ordnung γ oder kurz lokal γ-Hölder-stetig, falls für alle $x \in I$ Konstanten $\varepsilon, C > 0$ existieren, sodass

$$\forall y_1, y_2 \in I \text{ mit } |x - y_1| \vee |x - y_2| \leqslant \varepsilon : \qquad |f(y_1) - f(y_2)| \leqslant C|y_1 - y_2|^\gamma.$$

f heißt Hölder-stetig *auf I von der Ordnung γ*, falls ε beliebig groß gewählt werden kann und C nicht von x abhängt.

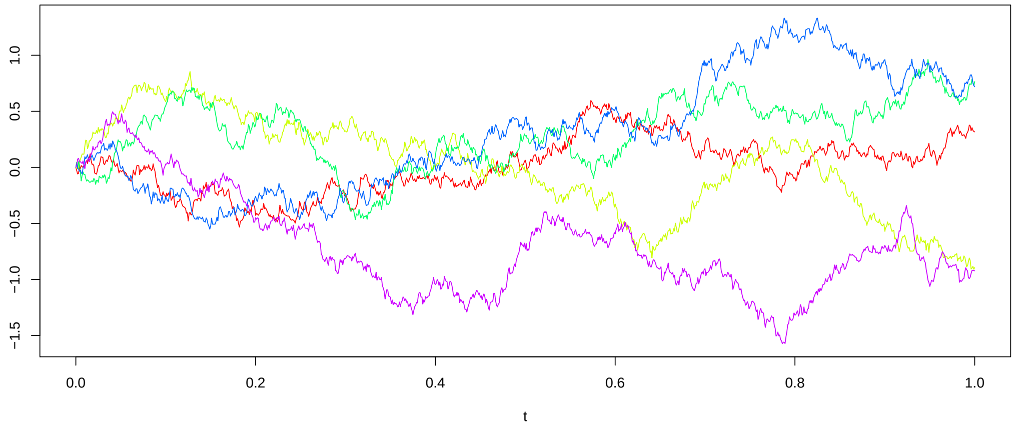

Abbildung 5: Fünf Trajektorien der standard Brownschen Bewegung.

Satz 4.3 (Kolmogorov-Chentsov)**.** *Es sei $X = (X_t, t \geqslant 0)$ ein reellwertiger Prozess. Für $\alpha, \beta > 0$ und jedes $T > 0$ gebe es Konstanten $\alpha, \beta, C > 0$ mit*

$$\forall s, t \in [0, T]: \qquad \mathbb{E}\big[|X_t - X_s|^\alpha\big] \leqslant C|t-s|^{1+\beta}.$$

Dann gilt:

(i) Es existiert eine Modifikation $Y = (Y_t, t \geqslant 0)$ von X, die lokal Hölder-stetige Pfade von jeder Ordnung $\gamma \in (0, \beta/\alpha)$ hat.

(ii) Ist $\gamma \in (0, \beta/\alpha)$, so existiert zu allen $\varepsilon, T > 0$ eine Zahl $K = K(\varepsilon, T, \alpha, \beta, C, \gamma) > 0$ (unabhängig von ω!) mit

$$\mathbb{P}\Big(\forall s, t \in [0, T] : |Y_s - Y_t| \leqslant K|t-s|^\gamma\Big) \geqslant 1 - \varepsilon.$$

Beweis. (i) Es reicht für jedes $T > 0$ zu zeigen, dass X auf $[0, T]$ eine Modifikation X^T besitzt, die γ-Hölder-stetig für jedes $\gamma \in (0, \beta/\alpha)$ ist. Für $S, T > 0$ sind dann die Prozesse X^S und X^T auf $[0, S \wedge T]$ ununterscheidbar (Übung ☑). Damit sind

$$\Omega_{S,T} := \big\{\exists t \in [0, S \wedge T] : \, X_t^S \neq X_t^T\big\} \quad \text{und} \quad \Omega_\infty := \bigcup_{S,T \in \mathbb{N}} \Omega_{S,T}$$

Nullmengen. Definieren wir nun

$$Y_t(\omega) := X_t^{\lceil t \rceil}(\omega), \quad t \geqslant 0, \omega \in \Omega \setminus \Omega_\infty,$$

(und $Y(\omega) = 0, \omega \in \Omega_\infty$) dann ist Y eine lokal Hölder-stetige Modifikation von X. O.B.d.A. sei nun $T = 1$. Wir werden zeigen, dass X eine auf $[0, 1]$ γ-Hölder-stetige Modifikation besitzt. Die Chebyshev-Ungleichung liefert für alle $\varepsilon > 0$

$$\mathbb{P}\big(|X_t - X_s| \geqslant \varepsilon\big) \leqslant \frac{C}{\varepsilon^\alpha}|t-s|^{1+\beta} \quad \forall s, t \geqslant 0. \tag{6}$$

Insbesondere gilt die stochastische Stetigkeit $X_s \xrightarrow{\mathbb{P}} X_t$ für $s \to t$.

Wir werden zunächst Y auf dyadischen Gittern konstruieren. Nach (6) gilt für $\gamma > 0, n \in \mathbb{N}$ und $k \in \{1, \ldots, 2^n\}$:

$$\mathbb{P}\big(|X_{k2^{-n}} - X_{(k-1)2^{-n}}| \geqslant 2^{-\gamma n}\big) \leqslant C 2^{-n(1+\beta-\alpha\gamma)}.$$

Wir setzen

$$A_{n,\gamma} := \left\{\omega : \exists k \in \{1, \dots, 2^n\} \text{ mit } |X_{k2^{-n}}(\omega) - X_{(k-1)2^{-n}}(\omega)| \geqslant 2^{-\gamma n}\right\}$$

sowie

$$B_{n,\gamma} := \bigcup_{m=n}^{\infty} A_{m,\gamma}, \qquad N := \limsup_{n\to\infty} A_{n,\gamma} = \bigcap_{n=1}^{\infty} B_{n,\gamma}.$$

Es folgt für jedes $n \in \mathbb{N}$

$$\mathbb{P}(A_{n,\gamma}) \leqslant \sum_{k=1}^{2^n} \mathbb{P}\big(|X_{k2^{-n}} - X_{(k-1)2^{-n}}| \geqslant 2^{-\gamma n}\big) \leqslant C2^{-n(\beta-\alpha\gamma)}.$$

Wählen wir nun ein $\gamma \in (0, \beta/\alpha)$, ergibt sich

$$\mathbb{P}(B_{n,\gamma}) \leqslant \sum_{m=n}^{\infty} \mathbb{P}(A_{m,\gamma}) \leqslant C\frac{2^{-(\beta-\alpha\gamma)n}}{1-2^{\alpha\gamma-\beta}} \to 0, \quad n \to \infty.$$

Das Lemma von Borel-Cantelli liefert $\mathbb{P}(N) = 0$. Sei nun $\omega \in N^c$ fest und $n_0 = n_0(\omega)$ so, dass $\omega \notin \bigcup_{n=n_0}^{\infty} A_{n,\gamma}$. Also gilt

$$|X_{k2^{-n}}(\omega) - X_{(k-1)2^{-n}}(\omega)| < 2^{-\gamma n} \quad \forall k \in \{1, \dots, 2^n\}, n \geqslant n_0. \tag{7}$$

Wir definieren nun die Menge der dyadischen Zahlen auf $[0, 1]$

$$D_m := \{k2^{-m}, k = 0, \dots, 2^m\} \quad \text{und} \quad D := \bigcup_{m\in\mathbb{N}} D_m.$$

Jedes $t \in D_m$ besitzt eine eindeutige Binärdarstellung $t = \sum_{i=0}^{m} b_i(t)2^{-i}$ für gewisse $b_i(t) \in \{0, 1\}, i = 1, \dots, m$. Seien $m \geqslant n \geqslant n_0$ sowie $s, t \in D_m$ mit $s \leqslant t$ und $|s - t| \leqslant 2^{-n}$. Für $u := \max(D_n \cap [0, s])$ ist dann

$$u \leqslant s < u + 2^{-n} \quad \text{und} \quad u \leqslant t < u + 2^{1-n}.$$

Damit gilt auch $b_i(t - u) = b_i(s - u) = 0$ für $i < n$. Setzen wir

$$t_l := u + \sum_{i=n}^{l} b_i(t-u)2^{-i} \quad \text{für } l = n-1, \dots, m,$$

erhalten wir $t_{n-1} = u, t_m = t, t_l \in D_l$ sowie

$$t_l - t_{l-1} \leqslant 2^{-l} \quad \text{für } l = n, \dots, m.$$

Also ist nach (7)

$$|X_t(\omega) - X_u(\omega)| \leqslant \sum_{l=n}^{m} \big|X_{t_l}(\omega) - X_{t_{l-1}}(\omega)\big| \leqslant \sum_{l=n}^{m} 2^{-\gamma l} \leqslant \frac{2^{-\gamma n}}{1-2^{-\gamma}}$$

und analog $|X_s(\omega) - X_u(\omega)| \leqslant \frac{2^{-\gamma n}}{1-2^{-\gamma}}$. Wir erhalten

$$|X_t(\omega) - X_s(\omega)| \leqslant 2\frac{2^{-\gamma n}}{1-2^{-\gamma}}, \qquad \text{für alle } n \geqslant n_0.$$

Wir setzen $C_0 := 2^{1+\gamma}/(1 + 2^{-\gamma})$. Für alle $s, t \in D$ mit $s \neq t, |s - t| \leqslant 2^{-n_0}$ erhalten wir nun für $n \geqslant n_0$ mit der Eigenschaft $2^{-n-1} \leqslant |t - s| \leqslant 2^{-n}$, dass

$$|X_t(\omega) - X_s(\omega)| \leqslant C_0 2^{-\gamma(n+1)} \leqslant C_0|t - s|^{\gamma}.$$

Für beliebige $s, t \in D$ mit $s < t$ erhalten wir mit $r_k := s + (t-s)k2^{-n_0}, k = 0, \dots 2^{n_0}$

$$|X_t(\omega) - X_s(\omega)| \leqslant \sum_{k=1}^{2^{n_0}} \big|X_{r_k}(\omega) - X_{r_{k-1}}(\omega)\big| \leqslant C_0 2^{n_0} \Big(\frac{t-s}{2^{n_0}}\Big)^\gamma = \underbrace{C_0 2^{(1-\gamma)n_0}}_{=:K} |t-s|^\gamma. \tag{8}$$

$X(\omega)$ ist also auf D γ-Hölderstetig mit einer globalen Konstante und wir können es eindeutig stetig auf $[0,1]$ fortsetzen: Für $t \in D$ setzen wir $Y_t := X_t$. Für $t \in [0,1] \setminus D$ und eine Folge $(s_n)_{n \in \mathbb{N}} \subseteq D$ mit $s_n \to t$ ist $(X_{s_n}(\omega))_n$ eine Cauchyfolge und es existiert der Grenzwert

$$Y_t(\omega) := \lim_{D \ni s \to t} X_s(\omega).$$

Es gilt damit für beliebige $s, t \in [0,1]$

$$|Y_t(\omega) - Y_s(\omega)| \leqslant K|t-s|^\gamma,$$

sodass Y Hölder-stetig von der Ordnung γ ist. Aufgrund der Konvergenzen $X_s \xrightarrow{\mathbb{P}} X_t$ und $Y_s(\omega) \to Y_t(\omega)$ für alle $s \to t$ und $\mathbb{P}$-f.a. $\omega \in \Omega$ gilt für eine Folge $(s_k) \subseteq D$ mit $s \to t$:

$$\begin{aligned}\mathbb{P}(X_t \neq Y_t) &= \mathbb{P}(\lim_{k\to\infty} |X_t - X_{s_k}| > 0) \leqslant \sum_{n \in \mathbb{N}} \mathbb{P}\Big(\liminf_{k\to\infty} \{|X_t - X_{s_k}| > 1/n\}\Big) \\ &= \sum_{n\in\mathbb{N}} \mathbb{E}\Big[\liminf_{k\to\infty} \mathbb{1}_{\{|X_t - X_{s_k}|>1/n\}}\Big] \overset{\text{Fatou}}{\leqslant} \sum_{n\in\mathbb{N}} \liminf_{k\to\infty} \mathbb{P}(|X_t - X_{s_k}| > 1/n) = 0.\end{aligned}$$

Also ist Y tatsächlich eine Modifikation von X.

(ii) Sei $\varepsilon > 0$ und $n_0 \in \mathbb{N}$ so groß, dass

$$\mathbb{P}(B_{n_0,\gamma}) \leqslant C \frac{2^{-(\beta-\alpha\gamma)n_0}}{1 - 2^{\alpha\gamma-\beta}} < \varepsilon.$$

Für $\omega \notin B_{n_0,\gamma}$ gilt nach dem oben gezeigten (8), was gerade die Behauptung mit $T = 1$ ist. Die Aussage für beliebige T folgt analog. □

Korollar 4.4. *Es existieren ein Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbb{P})$ *und eine Brownsche Bewegung* B *auf* $(\Omega, \mathcal{F}, \mathbb{P})$. *Die Pfade von* B *sind f.s. lokal* γ*-Hölder-stetig für jedes* $\gamma \in (0, 1/2)$.

Beweis. Wie in Beispiel 2.9 gesehen, existiert ein Prozess X mit den Eigenschaften (i)-(iv) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Für alle $0 \leqslant s < t$ gilt

$$X_t - X_s \overset{d}{=} \sqrt{t-s} X_1 \sim \mathcal{N}(0, t-s)$$

und daher für jedes $n \in \mathbb{N}$ und $C_n := \mathbb{E}[X_1^{2n}] = \frac{(2n)!}{2^n n!} < \infty$:

$$\mathbb{E}\big[(X_t - X_s)^{2n}\big] = \mathbb{E}\big[\big(\sqrt{t-s} X_1\big)^{2n}\big] = C_n |t-s|^n.$$

Für jedes beliebige $\gamma \in (0, 1/2)$ wählen wir nun n so, dass $\gamma < \frac{n-1}{2n}$ und erhalten aus Satz 4.3 die Existenz einer Version B von X mit f.s. γ-Hölder-stetigen Pfaden. Da alle stetigen Versionen ununterscheidbar sind, sind die Pfade von B γ-Hölder-stetig für jedes $\gamma \in (0, 1/2)$. □

Bemerkung 4.5. Umgekehrt besagt der Satz von Paley-Wiener-Zygmund, dass für jedes $\gamma > \frac{1}{2}$ die Pfade der Brownschen Bewegung f.s. in keinem Punkt γ-Hölder-stetig sind. Betrachtet man die Totalvariation der Pfade, lässt sich bereits leicht zeigen, dass die Brownsche Bewegung in keinem Intervall differenzierbar ist (Übung ☑).

Nachdem wir Pfadeigenschaften der Brownschen Bewegung kennengelernt haben, widmen wir uns nun einigen Verteilungseigenschafen.

Definition 4.6. Ein stochastischer Prozess $X = (X_t, t \in I)$ für ein Intervall $I \subseteq \mathbb{R}$ heißt Gaußprozess, falls für jedes $n \in \mathbb{N}$ und alle $t_1, \dots, t_n \in I$ der Vektor $(X_{t_1}, \dots, X_{t_n})$ n-dimensional normalverteilt ist. X heißt zentriert, falls $\mathbb{E}[X_t] = 0$ für alle $t \geqslant 0$ gilt. Die Funktion

$$\Gamma(s,t) := \operatorname{Cov}(X_s, X_t), \quad s,t \in I,$$

heißt Kovarianzfunktion von X.

Bemerkung 4.7. Durch die Kovarianzfunktion sind die endlichdimensionalen Verteilungen eines zentrierten Gaußprozesses eindeutig festgelegt.

Offensichtlich ist die Brownsche Bewegung ein zentrierter Gaußprozess. Genauer gilt:

Lemma 4.8. *Für einen stochastischen Prozess $X = (X_t, t \geqslant 0)$ sind äquivalent:*

(i) X ist eine Brownsche Bewegung.

(ii) X ist ein stetiger, zentrierter Gaußprozess mit Kovarianzfunktion $\Gamma(s,t) = s \wedge t$ für alle $s, t \geqslant 0$.

Beweis. Nach obiger Bemerkung ist die Verteilung von X durch (ii) eindeutig bestimmt. Es genügt also zu zeigen, dass die Kovarianzfunktion der Brownschen Bewegung gerade $\Gamma(s,t) = s \wedge t$ ist. Es gilt für jedes $0 \leqslant s \leqslant t$ aufgrund der Unabhängigkeit von X_s und $X_t - X_s$, dass

$$\Gamma(s,t) = \operatorname{Cov}(X_s, X_t) = \operatorname{Cov}(X_s, X_t - X_s) + \operatorname{Cov}(X_s, X_s) = \operatorname{Var}(X_s) = s. \qquad \square$$

Aus dieser Äquivalenz folgt sofort folgende Skalierungseigenschaft:

Korollar 4.9 (Skalierungsinvarianz)**.** *Ist B eine Brownsche Bewegung, dann ist für jedes $\alpha \neq 0$ der Prozess $(\alpha^{-1} B_{\alpha^2 t}, t \geqslant 0)$ ebenfalls eine Brownsche Bewegung.*

Beispiel 4.10.

(i) Ein zentrierter, stetiger Gaußprozess $X = (X_t, t \in [0,1])$, welcher durch die Kovarianzfunktion $\Gamma(s,t) = s \wedge t - st$ für $s, t \in [0,1]$ definiert ist, heißt *Brownsche Brücke*. Ist B eine Brownsche Bewegung, so kann man X mittels

$$X_t := B_t - tB_1, \quad t \in [0,1],$$

konstruieren: Offenbar ist X dann ein stetiger, zentrierter Gaußprozess. Wir rechnen nach:

$$\begin{aligned}\Gamma(s,t) &= \operatorname{Cov}(X_s, X_t) = \operatorname{Cov}(B_s - sB_1, B_t - tB_1) \\ &= \operatorname{Cov}(B_s, B_t) - s \operatorname{Cov}(B_1, B_t) - t \operatorname{Cov}(B_s, B_1) + st \operatorname{Cov}(B_1, B_1) \\ &= s \wedge t - st.\end{aligned}$$

(ii) Ein zentrierter Gaußprozess $Y = (Y_t, t \in [0,1])$ mit der Kovarianzfunktion $\Gamma(s,t) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})$ für ein $H \in (0,1)$ besitzt ebenfalls eine stetige Modifikation (Kolmogorov-Chentsov anwenden, Übung ☑). Der resultierende stetige Gaußprozess heißt *fraktionale Brownsche Bewegung* mit *Hurst index* H. Im Spezialfall $H = \frac{1}{2}$ erhalten wir wieder die standard Brownsche Bewegung.

4.2 Stoppzeiten in stetiger Zeit

Zunächst erweitern wir den Begriff der Filtration und der Stoppzeit auf zeitstetige Modelle.

Definition 4.11. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Folge $(\mathcal{F}_t)_{t \geqslant 0}$ von Sub-σ-Algebren $\mathcal{F}_t \subseteq \mathcal{F}$ heißt Filtration, falls $\mathcal{F}_s \subseteq \mathcal{F}_t$ für alle $0 \leqslant s < t$ gilt. Eine $\mathbb{R}_+$-wertige Zufallsvariable τ heißt Stoppzeit (bzgl. $(\mathcal{F}_t)$), falls $\{\tau \leqslant t\} \in \mathcal{F}_t$ für alle $t \geqslant 0$ gilt. Die σ-Algebra der τ-Vergangenheit ist gegeben durch $\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leqslant t\} \in \mathcal{F}_t \,\forall t \geqslant 0\}$. Ein Prozess $X = (X_t)_{t \geqslant 0}$ heißt adaptiert, falls X_t $\mathcal{F}_t$-messbar für alle $t \geqslant 0$ ist.

Bemerkung 4.12. Wie in diskreter Zeit gilt für zwei Stoppzeiten σ und τ (Übung ☑):

(i) $\sigma \wedge \tau$, $\sigma \vee \tau$ und $\sigma + \tau$ sind ebenfalls Stoppzeiten.

(ii) Gilt $\sigma \leqslant \tau$, so folgt $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

(iii) τ ist $\mathcal{F}_\tau$-messbar.

Ein typisches Beispiel sind die Eintrittszeiten

$$\tau_A := \inf\{t \geqslant 0 | X_t \in A\}$$

für eine messbare Menge A und einen adaptierten Prozess X. In diskreter Zeit ist keine weitere Bedingung nötig, um nachzuweisen, dass τ_A eine Stoppzeit ist. In stetiger Zeit benötigen hingegen zusätzliche Voraussetzungen. Folgendes Lemma gibt uns hinreichende Bedingungen:

Lemma 4.13. *Auf dem filtrierten Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ *mit rechtsstetiger Filtration* $(\mathcal{F}_t)_{t\geqslant 0}$ *, d.h.* $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ *für alle* $t \geqslant 0$*, sei* X *ein reellwertiger, adaptierter Prozess mit rechtsstetigen Pfaden.*

(i) Ist $A \subseteq \mathbb{R}$ *offen, so ist* τ_A *eine Stoppzeit.*

(ii) Ist $A \subseteq \mathbb{R}$ *abgeschlossen und* X *stetig, so ist* τ_A *eine Stoppzeit.*

Beweis. (i) Für $t \geqslant 0$ ist $\{\tau_A \leqslant t\} \in \mathcal{F}_t$ zu zeigen. Aufgrund der rechtsstetigen Pfade und da A offen ist, gilt

$$\begin{aligned}\{\tau_A < t\} &= \{\omega | \exists s < t : X_s(\omega) \in A\} \\ &= \{\omega | \exists s < t : s \in \mathbb{Q}, X_s(\omega) \in A\} = \bigcup_{s\in\mathbb{Q}, s<t} \{X_s \in A\} \in \mathcal{F}_t.\end{aligned}$$

Wir erhalten aus der Rechtsstetigkeit der Filtration:

$$\{\tau_A \leqslant t\} = \bigcap_{n\geqslant 1} \underbrace{\{\tau_A < t + 1/n\}}_{\in \mathcal{F}_{t+1/n}} \in \mathcal{F}_t.$$

(ii) Für die offenen Mengen $A_n := \{y \in \mathbb{R} : d(y, A) < \frac{1}{n}\}$ gilt $A = \bigcap_{n\geqslant 1} A_n$. Nach (i) sind τ_{A_n}Stoppzeiten. Es genügt somit, zu zeigen, dass $\tau_{A_n} \uparrow \tau_A$ gilt, woraus folgt, dass τ_A ebenfalls eine Stoppzeit ist.

Es gilt $\tau_{A_n} \leqslant \tau_{A_{n+1}} \leqslant \tau_A$, $n \in \mathbb{N}$. Für $\sigma := \sup_{n\geqslant 1} \tau_{A_n}$ gilt damit $\sigma \leqslant \tau_A$. Andererseits folgt aus der Stetigkeit von X, dass $X_{\tau_{A_n}}$ im Abschluss von A_n liegt. Daraus folgt $X_\sigma \in A$ (auf $\{\sigma < \infty\}$ und damit $\tau_A \leqslant \sigma$. Wir erhalten $\sigma = \tau_A$. □

Dass tatsächlich zusätzliche Bedingungen nötig sind, zeigt folgendes Gegenbeispiel.

Beispiel 4.14. Sei $A \in \mathcal{F} \setminus \{\emptyset, \Omega\}$ und setze

$$X_t(\omega) = \begin{cases} (t-1)\mathbb{1}_{\{t>1\}}, & \omega \in A, \\ (1-t)\mathbb{1}_{\{t>1\}}, & \omega \in A^c \end{cases} \quad \text{sowie} \quad \mathcal{F}_t = \begin{cases} \{\emptyset, \Omega\} & t \in [0,1], \\ \{\emptyset, A, A^c, \Omega\}, & t > 1. \end{cases}$$

Offensichtlich ist X bzgl $(\mathcal{F}_t)$ adaptiert, aber $(\mathcal{F}_t)$ ist nicht rechtsstetig. Für $\tau := \inf\{t \geqslant 0 : X_t < 0\}$ gilt aber, dass $\{\tau \leqslant 1\} = \{\tau = 1\} = A^c \notin \mathcal{F}_1$.

Satz 4.15. *Es seien* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ *ein filtrierter Wahrscheinlichkeitsraum,* X *ein adaptierter Prozess und* τ *eine endliche Stoppzeit, d.h.* $\tau(\omega) < \infty$ *für alle* $\omega \in \Omega$*. Ist* X *links- oder rechtsstetig, so ist die Zufallsvariable* $X_\tau : \Omega \to \mathbb{R}, \omega \mapsto X_{\tau(\omega)}(\omega)$ $\mathcal{F}_\tau$*-messbar.*

Beweis. Schritt 1: Wir zeigen zunächst, dass X progressiv messbar ist, d.h. für alle $t > 0$ ist die Abbildung $[0,t] \times \Omega \ni (s,\omega) \mapsto X_s(\omega) \in \mathbb{R}$ $(\mathscr{B}_{[0,t]} \otimes \mathcal{F}_t, \mathscr{B}_\mathbb{R})$-messbar.

Sei X rechtsstetig (linksstetig analog). Für $n \geqslant 1$, und $s \in (0,t]$ definieren wir

$$X_s^{(n)}(\omega) := \sum_{k=0}^{2^n-1} X_{(k+1)t/2^n}(\omega) \mathbb{1}_{(kt/2^n,(k+1)t/2^n]}(s), \qquad X_0^{(n)}(\omega) = X_0(\omega).$$

Dann ist jedes $X^{(n)}$ nach Konstruktion $\mathscr{B}_{[0,t]} \otimes \mathcal{F}_t$-messbar und somit auch $X_s(\omega) = \lim_{n\to\infty} X_s^{(n)}(\omega)$.

Schritt 2: Für jedes $B \in \mathscr{B}_\mathbb{R}$ und jedes $t \geqslant 0$ müssen wir nun $\{X_\tau \in B\} \cap \{\tau \leqslant t\} \in \mathcal{F}_t$ zeigen. Wir betrachten die Abbildung

$$\sigma \colon \Omega \ni \omega \mapsto (\tau(\omega) \wedge t, \omega) \in [0,t] \times \Omega.$$

σ ist $(\mathcal{F}_t, \mathscr{B}_{[0,t]} \otimes \mathcal{F}_t)$-messbar, da für alle $s \in [0,t)$ und $A \in \mathscr{F}_t$ gilt

$$\sigma^{-1}([0,s] \times A) = \{\tau \leqslant s\} \cap A \in \mathcal{F}_t.$$

Zusammen mit Schritt 1 erhalten wir also

$$\{X_\tau \in B\} \cap \{\tau \leqslant t\} = \{X_{\tau \wedge t} \in B\} \cap \{\tau \leqslant t\} = \{(X_s)_{s\in[0,t]} \circ \sigma \in B\} \cap \{\tau \leqslant t\} \in \mathcal{F}_t,$$

denn $(X_s)_{s\in[0,t]} \circ \sigma$ ist $(\mathcal{F}_t, \mathscr{B}_\mathbb{R})$-messbar. □

Kommen wir zu einer interessanten Anwendung von Stoppzeiten. Aus unserer Konstruktion der Brownschen Bewegung wissen wir bereits, dass B ein Markovprozess ist: Wir definieren die Verteilungsfamilie $\mathbb{P}_x$ für $x \in \mathbb{R}$ so, dass auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P}_x)$ der Prozess $(B_t - x, t \geqslant 0)$ eine Brownsche Bewegung ist. Mit anderen Worten ist B unter $\mathbb{P}_x$ eine in x gestartete Brownsche Bewegung und die (schwache) Markoveigenschaft folgt aus der Definition. Tatsächlich erfüllt die Brownsche Bewegung sogar die starke Markoveigenschaft, vgl. Satz 2.12, wie folgender Satz zeigt.

Satz 4.16 (Starke Markoveigenschaft)**.** *Für die Brownsche Bewegung B auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit natürlicher Filtration $\mathcal{F}_t = \sigma(B_s, s \leqslant t)$ für $t \geqslant 0$ und für jede f.s.-endliche Stoppzeit τ bgzl. $(\mathcal{F}_t)_{t\geqslant 0}$ ist $\widetilde{B}_t := B_{t+\tau} - B_\tau, t \geqslant 0$, eine Brownsche Bewegung, die unabhängig von $\mathcal{F}_\tau$ ist.*

Beweis. Für deterministische $\tau = s \geqslant 0$ folgt direkt aus der Definition, dass $(B_{t+s} - B_s)_{t\geqslant 0}$ wieder eine Brownsche Bewegung unabhängig von $\mathcal{F}_s$ ist. Da die Verteilung von B durch die endlich-dimensionalen Verteilungen festgelegt ist, genügt es zu zeigen, dass für jede beschränkte $\mathcal{F}_\tau$-messbare Funktion $\varphi \colon \Omega \to \mathbb{R}$ und jede beschränkte, messbare Funktion $f \colon \mathbb{R}^{[0,\infty)} \to \mathbb{R}$, die nur von endlich vielen Koordinaten $t_1, \dots, t_n \geqslant 0, n \in \mathbb{N}$, abhängt,

$$\mathbb{E}\big[\varphi f\big(\widetilde{B}\big)\big] = \mathbb{E}[\varphi]\mathbb{E}\big[f\big(B\big)\big]$$

gilt. Mittels Approximationsargumenten können wir f als stetig annehmen (z.B. durch Glättung einer approximierenden Folge einfacher Funktionen).

Sei also $f(B) = g(B_{t_1}, \dots, B_{t_n})$ für eine stetige, beschränkte Funktion $g \colon \mathbb{R}^n \to \mathbb{R}$. Für jedes $m \in \mathbb{N}$ definieren wir $\tau^m := 2^{-m}\lfloor 2^m\tau + 1\rfloor$ (τ^m ist also kleinste Zahl auf dem dyadischen Gitter $\mathbb{N}2^{-m}$, die größer als τ ist), sodass wegen

$$\{\tau^m \leqslant t\} = \bigcup_{k\in\mathbb{N}:2^{-m}k\leqslant t} \Big\{\tau \in \Big[\frac{k-1}{2^m}, \frac{k}{2^m}\Big)\Big\} \in \mathcal{F}_t \qquad \text{für alle } t > 0$$

τ^m ebenfalls eine Stoppzeit ist mit $\tau^m \downarrow \tau$ für $m \to \infty$. Wir definieren den Prozess $\widetilde{B}_t^m := B_{t+\tau^m} - B_{\tau^m}, t \geqslant 0$. Dann gilt

$$\mathbb{E}\big[\varphi f\big(\widetilde{B}^m\big)\big] = \mathbb{E}\Big[\varphi g\big(\widetilde{B}_{t_1}^m, \dots, \widetilde{B}_{t_n}^m\big)\Big]$$

$$
\begin{aligned}
&= \sum_{k\geqslant 0} \mathbb{E}\Big[\underbrace{\varphi \mathbb{1}_{\{\tau^m = k2^{-m}\}}}_{\text{messbar bzgl. } \mathcal{F}_{k2-m}} \underbrace{g\big(B_{t_1+k2^{-m}} - B_{k2^{-m}}, \dots, B_{t_n+k2^{-m}} - B_{k2^{-m}}\big)}_{\text{unabhängig von } \mathcal{F}_{k2-m}} \Big] \\
&= \sum_{k\geqslant 0} \mathbb{E}\big[\varphi \mathbb{1}_{\{\tau^m = k2^{-m}\}}\big] \mathbb{E}[f(B)] \\
&= \mathbb{E}[\varphi]\mathbb{E}[f(B)].
\end{aligned}
$$

Aufgrund der (fast sicheren) Stetigkeit von f und $\widetilde{B}$ folgt $f(\widetilde{B}^m) \to f(\widetilde{B})$ f.s. Zusammen mit der Beschränktheit von φ und f ergibt sich

$$
\mathbb{E}\big[\varphi f\big(\widetilde{B}\big)\big] = \lim_{m\to\infty} \mathbb{E}\big[\varphi f\big(\widetilde{B}^m\big)\big] = \mathbb{E}[\varphi]\mathbb{E}[f(B)]. \qquad \square
$$

Ähnlich zur starken Markoveigenschaft können wir das so genannte *Reflexionsprinzip* beweisen, welches uns schon für Irrfahrten begegnet ist:

Satz 4.17 (Reflexionsprinzip)**.** *Für die Brownsche Bewegung B auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit natürlicher Filtration $\mathcal{F}_t = \sigma(B_s, s \leqslant t), t \geqslant 0$, und für jede f.s.-endliche Stoppzeit τ bgzl. $(\mathcal{F}_t)_{t\geqslant 0}$ ist der Prozess*

$$
Y_t := \begin{cases} B_t, & t \leqslant \tau, \\ B_\tau - (B_t - B_\tau), & t > \tau \end{cases}
$$

eine Brownsche Bewegung.

Beweis. Wir wählen τ^m wie im Beweis zuvor und setzen

$$
Y_t^m := \begin{cases} B_t, & t \leqslant \tau^m, \\ B_{\tau^m} - (B_t - B_{\tau^m}), & t > \tau^m \end{cases}
$$

Es seien nun $0 = t_1 < t_2 < \cdots < t_n, n \in \mathbb{N}$ und $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig und beschränkt. Dann gilt

$$
\begin{aligned}
&\mathbb{E}\big[f(Y_{t_1}^m, \dots, Y_{t_n}^m)\mathbb{1}_{\{\tau<\infty\}}\big] \\
&= \sum_{k\geqslant 0} \mathbb{E}\big[f(Y_{t_1}^m, \dots, Y_{t_n}^m)\mathbb{1}_{\{\tau^m = k2^{-m}\}}\big] \\
&= \sum_{k\geqslant 0} \sum_{l=1}^{n-1} \mathbb{E}\big[f(Y_{t_1}^m, \dots, Y_{t_n}^m)\mathbb{1}_{\{\tau^m = k2^{-m}\}} \mathbb{1}_{(t_l, t_{l+1}]}\big(\frac{k}{2^m}\big)\big] \\
&\qquad + \sum_{k\geqslant 0} \mathbb{E}\big[f(Y_{t_1}^m, \dots, Y_{t_n}^m)\mathbb{1}_{\{\tau^m = k2^{-m}\}} \mathbb{1}_{(t_n, \infty)}\big(\frac{k}{2^m}\big)\big] \\
&= \sum_{k=0}^{n-1} \sum_{l=1}^{n-1} \mathbb{E}\big[f(B_{t_1}, \dots, B_{t_l}, B_{k2^{-m}} - (B_{t_{l+1}} - B_{k2^{-m}}), \dots, B_{k2^{-m}} - (B_{t_n} - B_{k2^{-m}}))\mathbb{1}_{\{\tau^m = k2^{-m}\}}\big] \\
&\qquad \times \mathbb{1}_{(t_l, t_{l+1}]}\big(\frac{k}{2^m}\big) + \sum_{k\geqslant 0} \mathbb{E}\big[f(B_{t_1}, \dots, B_{t_n})\mathbb{1}_{\{\tau^m = k2^{-m}\}} \mathbb{1}_{(t_n, \infty)}\big(\frac{k}{2^m}\big)\big].
\end{aligned}
$$

Nun ist $((B_{t_{l+1}} - B_{k2^{-m}}), \dots, (B_{t_{n+1}} - B_{k2^{-m}}))$ unabhängig von $\mathcal{F}_{k2^{-m}}$ und symmetrisch um 0 verteilt, sodass

$$
\mathbb{E}\big[f(Y_{t_1}^m, \dots, Y_{t_n}^m)\mathbb{1}_{\{\tau<\infty\}}\big] = \mathbb{E}\big[f(B_{t_1}, \dots, B_{t_n})\mathbb{1}_{\{\tau<\infty\}}\big]
$$

folgt. Mit majorisierter Konvergenz erhalten wir aus der Stetigkeit von B und f:

$$
\mathbb{E}\big[f(Y_{t_1}, \dots, Y_{t_n})\mathbb{1}_{\{\tau<\infty\}}\big] = \mathbb{E}\big[f(B_{t_1}, \dots, B_{t_n})\mathbb{1}_{\{\tau<\infty\}}\big]. \qquad \square
$$

Korollar 4.18. *Setze $B_t^* = \max_{0\leqslant s\leqslant t} B_s$ für $t \geqslant 0$. Für $x, y > 0$ gilt dann für alle $t \geqslant 0$*

$$
\mathbb{P}(B_t^* \geqslant x, B_t < x - y) = \mathbb{P}(B_t > x + y).
$$

Insbesondere erhalten wir

$$
\mathbb{P}\Big(B_t^* \geqslant x\Big) = 2\mathbb{P}(B_t \geqslant x) = \mathbb{P}(|B_t| \geqslant x) \quad \text{für alle } t > 0.
$$

Beweis. Wir betrachten die Stoppzeit (!) $\tau_x := \inf\{s \geqslant 0 : B_s \geqslant x\}$. Sei Y der zu B und $\tau_x \wedge t$ gehörige Prozess aus Satz 4.17. Für $\sigma_x = \inf(t \geqslant 0 : Y_t \geqslant x)$ gilt $\sigma_x = \tau_x$ und wir erhalten mit dem Reflexionsprinzip

$$\begin{aligned}\mathbb{P}(B_t \geqslant x+y) =& \mathbb{P}(B_t^* \geqslant x, B_t \geqslant x+y)\\ =& \mathbb{P}(\tau_x \leqslant t, B_t \geqslant x+y)\\ \overset{RP}{=}& \mathbb{P}(\sigma_x \leqslant t, Y_t \geqslant x+y)\\ =& \mathbb{P}(\tau_x \leqslant t, B_t \leqslant x-y) = \mathbb{P}(B_t^* \geqslant x, B_t \leqslant x-y).\end{aligned}$$

Für $y = 0$ gilt $\mathbb{P}(B_t^* \geqslant x, B_t \geqslant x) = \mathbb{P}(B_t \geqslant x)$, sodass folgt

$$\begin{aligned}\mathbb{P}(B_t^* \geqslant x) =& \mathbb{P}(B_t^* \geqslant x, B_t \geqslant x) + \mathbb{P}(B_t^* \geqslant x, B_t \leqslant x)\\ =& 2\mathbb{P}(B_t \geqslant x) = \mathbb{P}(|B_t| \geqslant x).\end{aligned}$$

□

Korollar 4.19. *Für jedes $t > 0$ haben die Zufallsvariablen B_t^*, $|B_t|$ sowie $B_t^* - B_t$ die gleiche Verteilung.*

Beweis. $B_t^* \overset{d}{=} |B_t|$ folgt direkt aus Korollar 4.18. Für die dritte Zufallsvariable verwenden wir, dass $\widetilde{B}_s := B_{t-s} - B_t$ auf $s \in [0,t]$ wieder eine Brownsche Bewegung ist und damit gilt

$$B_t^* - B_t = \max_{0 \leqslant s \leqslant t}(B_s - B_t) = \max_{0 \leqslant u \leqslant t}(B_{t-u} - B_t) = \max_{0 \leqslant u \leqslant t} \widetilde{B}_u =: \widetilde{B}_t^*.$$

Wegen $\widetilde{B}_t^* \overset{d}{=} B_t^*$ erhalten wir $B_t^* - B_t \overset{d}{=} B_t^*$. □

Bemerkung 4.20. Als weitere Anwendung des Reflexionsprinzip kann man zeigen (Übung ☑), dass für eine Brownsche Bewegung B der Prozess $X = (X_t, t \geqslant 0)$ mit

$$X_t = \begin{cases} tB_{1/t}, & \text{falls } t > 0,\\ 0, & \text{sonst,}\end{cases}$$

ebenfalls eine Brownsche Bewegung ist (Zeitinvertierung).

4.3 Martingale in stetiger Zeit

Aus der Vorlesung „Maßtheoretische Konzepte der Stochastik“ kennen Sie bereits Martingale in diskreter Zeit. Dieses Konzept wollen wir nun auf zeitstetige Prozesse übertragen. Das wichtigste prototypische Beispiel, die Brownsche Bewegung, haben wir bereits kennen gelernt.

Definition 4.21. Auf dem filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ heißt ein adaptierter Prozess $X = (X_t)_{t\geqslant 0}$ <u>Martingal</u> (bzw. <u>Sub-</u> oder <u>Supermartingal</u>), falls $\mathbb{E}[|X_t|] < \infty$ für alle $t \geqslant 0$ sowie $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ (bzw. $\mathbb{E}[X_t|\mathcal{F}_s] \geqslant X_s$ oder $\mathbb{E}[X_t|\mathcal{F}_s] \leqslant X_s$) für alle $0 \leqslant s \leqslant t$ gilt.

Beispiel 4.22.

(i) Die Brownsche Bewegung $B = (B_t)_{t\geqslant 0}$ ist ein stetiges Martingal bzgl. ihrer natürlichen Filtration $\mathcal{F}_t^B := \sigma(B_s : s \leqslant t)$.

(ii) Für eine Brownsche Bewegung B sind $(B_t^2 - t, t \geqslant 0)$ und $(\exp(\vartheta B_t - \frac{\vartheta^2 t}{2}), t \geqslant 0)$ für jedes $\vartheta > 0$ stetige Martingale bzgl. $\mathcal{F}_t^B$ (Übung ☑).

(iii) Ist $N = (N_t)$ ein Poissonprozess mit Intensität $\lambda > 0$, so ist $M_t := N_t - \lambda t$ ein rechtsstetiges Martingal (mit linken Grenzwerten) bzgl. $\mathcal{F}_t^N := \sigma(N_s : s \leqslant t)$.

Wiederholung: Aus der Maßtheorievorlesung (bzw. Anhang B) wissen wir bereits einiges über Martingale in diskreter Zeit.

- *Doobs Martingalungleichungen:* Sei $M = (M_k, k \in \mathbb{N}_0)$ ein Martingal oder ein nichtnegatives Submartingal, $n \in \mathbb{N}$ und $|M|_n^* := \max_{k=0,\dots,n} |M_k|$. Für $\alpha > 0$ gilt

$$\mathbb{P}\big(|M|_n^* \geqslant \alpha\big) \leqslant \frac{1}{\alpha}\mathbb{E}[|M_n|].$$

  Falls $M_k \in L^p(\mathbb{P})$ für alle $k \in \mathbb{N}_0$ und ein $p > 1$ gilt weiterhin

$$\mathbb{E}\big[(|M|_n^*)^p\big] \leqslant \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|M_n|^p].$$

- *Optional-Sampling-Satz*: Es seien $M = (M_n, n \in \mathbb{N}_0)$ ein Martingal (bzw. Submartingal) und σ, τ Stoppzeiten (alle bzgl. derselben Filtration) mit $\sigma \leqslant \tau \leqslant C$ für eine Konstante $C > 0$. Dann gilt

$$\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = M_\sigma \quad (\text{bzw. } \mathbb{E}[M_\tau|\mathcal{F}_\sigma] \geqslant M_\sigma).$$

Diese Resultate wollen wir nun auf den zeitstetigen Fall übertragen.

Satz 4.23 (Doobs Martingalungleichungen)**.** *Auf dem filtrierten Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ *seien* $X = (X_t)_{t\geqslant 0}$ *ein rechtsstetiges, nichtnegatives Submartingal und* $t > 0$*. Dann gilt:*

(i) Für jedes $\alpha > 0$ *gilt*

$$\mathbb{P}\Big(\sup_{0\leqslant r\leqslant t} X_r \geqslant \alpha\Big) \leqslant \frac{1}{\alpha}\mathbb{E}[X_t].$$

(ii) Gilt außerdem $X_t \in L^p(\mathbb{P})$ *für ein* $p > 1$*, dann folgt*

$$\mathbb{E}\Big[\Big(\sup_{0\leqslant r\leqslant t} X_r\Big)^p\Big] \leqslant \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[X_t^p].$$

Beweis. (i) Sei $T \subseteq [0,t]$ eine endliche Teilmenge mit $t \in T$, dann folgt aus dem zeitdiskreten Fall $\mathbb{P}(\sup_{r\in T} X_r \geqslant \alpha) \leqslant \frac{1}{\alpha}\mathbb{E}[X_t]$. Betrachten wir eine wachsende Folge $T_n \subseteq T_{n+1}$ endlicher Mengen mit $S := \bigcup_{n\geqslant 1} T_n = ([0,t] \cap \mathbb{Q}) \cup \{t\}$ folgt aus der Stetigkeit des Maßes $\mathbb{P}$ und der Rechtsstetigkeit von X

$$\mathbb{P}\Big(\sup_{0\leqslant r\leqslant t} X_r \geqslant \alpha\Big) = \mathbb{P}\Big(\sup_{r\in S} X_r \geqslant \alpha\Big) \leqslant \frac{1}{\alpha}\mathbb{E}[X_t].$$

(ii) Wir verfahren hier analog, wobei wir monotone Konvergenz verwenden:

$$\mathbb{E}\Big[\Big(\sup_{0\leqslant r\leqslant t} X_r\Big)^p\Big] = \lim_{n\to\infty} \mathbb{E}\Big[\Big(\sup_{r\in T_n\cup\{0,t\}} X_r\Big)^p\Big] \leqslant \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|X_t|^p]. \qquad \square$$

Um das Optional-Sampling-Resultat zu verallgemeinern, erinnern wir uns zunächst an gleichgradige Integrierbarkeit, siehe Anhang A.

Satz 4.24 (Optional Sampling)**.** *Auf dem filtrierten Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ *sei* $X = (X_t)_{t\geqslant 0}$ *ein stetiges Martingal und* σ, τ *Stoppzeiten mit* $\sigma(\omega) \leqslant \tau(\omega) \leqslant T$ *für ein festes* $T > 0$ *und alle* $\omega \in \Omega$*. Dann gilt*

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] = X_\sigma \qquad \mathbb{P}\text{-f.s.}$$

Insbesondere gilt $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.

Beweis. Schritt 1: Wir betrachten die Folge von Stoppzeiten (!)

$$\sigma_n(\omega) = \sum_{k=1}^{\lceil T2^n \rceil} \frac{k}{2^n} \mathbb{1}_{\{(k-1)/2^n \leqslant \sigma(\omega) < k/2^n\}}$$

und analog τ_n. Für jedes feste n nehmen σ_n und τ_n nur Werte auf dem Gitter $(k2^{-n})_{k \geqslant 0}$ an und es gilt $\sigma_n \leqslant \tau_n$. Aus dem Optional-Sampling-Satz in diskreter Zeit erhalten wir also

$$\int_A X_{\sigma_n} \mathrm{d}\mathbb{P} = \int_A X_{\tau_n} \mathrm{d}\mathbb{P} \qquad \text{für alle } A \in \mathcal{F}_{\sigma_n}$$

und wegen $\sigma \leqslant \sigma_n$ insbesondere für alle $A \in \mathcal{F}_\sigma$.

Schritt 2: Wir zeigen ein Hilfsresultat: Für jede Zufallsvariable $Z \in L^1(\mathbb{P})$ existiert ein konvexes und monoton wachsendes $\Psi \colon \mathbb{R}_+ \to \mathbb{R}_+$ mit $\Psi(x)/x \to \infty$ für $x \to \infty$ und $\mathbb{E}[\Psi(|Z|)] < \infty$. Wegen

$$\mathbb{E}[|Z|] = \int_0^\infty \mathbb{P}(|Z| \geqslant x) \mathrm{d}x$$

existiert eine messbare, monoton wachsende Funktion $a \colon \mathbb{R}_+ \to [1, \infty)$ mit $a(x) \to \infty$ für $x \to \infty$ und

$$\int_0^\infty \mathbb{P}(|Z| \geqslant x) a(x) \mathrm{d}x < \infty.$$

Wähle z.B. $x_n = \inf\{y \geqslant 0 : \int_0^y \mathbb{P}(|Z| \geqslant x)\mathrm{d}x \geqslant \mathbb{E}[|Z|](1 - 2^{-n})\}$ für $n \in \mathbb{N}$ und beachte, dass . Dann gilt $x_n \to \infty$ (falls Z unbeschränkt ist, sonst wähle x_n beliebig wachsend außerhalb des Trägers von Z) und mit $a(x) = \sum_{n \geqslant 0} (n+1) \mathbb{1}_{[x_n, x_{n+1})}(x)$ folgt

$$\begin{aligned}\int_0^\infty \mathbb{P}(|Z| \geqslant x) a(x) \mathrm{d}x &= \sum_{n \geqslant 0} (n+1) \int_{x_n}^{x_{n+1}} \mathbb{P}(|Z| \geqslant x) \mathrm{d}x \\ &\leqslant \sum_{n \geqslant 0} (n+1) 2^{-(n+1)} \mathbb{E}[|Z|] < \infty.\end{aligned}$$

Wir wählen nun $\Psi(x) = \int_0^x a(y)\mathrm{d}y$. Dann ist Ψ konvex, da a wachsend ist, und aus dem Satz von l'Hospital folgt $\Psi(x)/x \to \infty$ wegen $a(x) \to \infty$ für $x \to \infty$. Außerdem gilt:

$$\mathbb{E}[\Psi(|Z|)] = \mathbb{E}\Big[\int_0^{|Z|} a(x)\mathrm{d}x\Big] = \mathbb{E}\Big[\int_0^\infty \mathbb{1}_{\{x \leqslant |Z|\}} a(x) \mathrm{d}x\Big] = \int_0^\infty \mathbb{P}(|Z| \geqslant x) a(x) \mathrm{d}x < \infty.$$

Schritt 3: Wir zeigen nun die gleichgradige Integrierbarkeit von $(X_{\sigma_n})_{n \geqslant 1}$ und $(X_{\tau_n})_{n \geqslant 1}$. Wir wählen das zu X_T gehörige Ψ aus Schritt 2. Aus der Konvexität und Monotonie von Ψ folgt, dass $(\Psi(|X_t|))_{t \geqslant 0}$ ein nichtnegatives Submartingal ist, sodass aus dem Optional-Sampling-Satz in diskreter Zeit folgt:

$$\mathbb{E}[\Psi(|X_{\sigma_n}|)] \leqslant \mathbb{E}[\Psi(|X_T|)] < \infty.$$

Damit ergibt sich die gleichgradige Integrierbarkeit von $(X_{\sigma_n})_{n \geqslant 1}$ und analog von $(X_{\tau_n})_{n \geqslant 1}$ nach dem Satz von de la Vallée-Poussin.

Schritt 4: Aufgrund der Rechtsstetigkeit gilt $\lim_{n \to \infty} X_{\sigma_n}(\omega) = X_\sigma(\omega)$ für alle $\omega \in \Omega$, sodass $X_{\sigma_n} \xrightarrow{\mathbb{P}} X_\sigma$ folgt. Analog konvergiert X_{τ_n} stochastisch gegen X_τ. Zusammen mit der gleichgradigen Integriebarkeit folgt die L^1-Konvergenz von $X_{\sigma_n} \to X_\sigma$ und $X_{\tau_n} \to X_\tau$ und damit

$$\int_A X_\sigma \mathrm{d}\mathbb{P} = \lim_{n \to \infty} \int_A X_{\sigma_n} \mathrm{d}\mathbb{P} = \lim_{n \to \infty} \int_A X_{\tau_n} \mathrm{d}\mathbb{P} = \int_A X_\tau \mathrm{d}\mathbb{P} = \int_A \mathbb{E}[X_\tau | \mathcal{F}_\sigma] \mathrm{d}\mathbb{P} \qquad \text{für alle } A \in \mathcal{F}_\sigma.$$

□

Beispiel 4.25. Betrachten wir eine Brownsche Bewegung B auf $(\Omega,\mathcal{F},\mathbb{P})$ mit natürlicher Filtration und die Austrittszeit $\tau_x=\inf(t\geqslant 0,|B_t|\geqslant x)=\inf(t\geqslant 0,|B_t|=x)$ für ein $x>0$. Dann folgt aus Optional Sampling für das Martingal $(B_t^2-t,t\geqslant 0)$

$$\mathbb{E}[B_{\tau_x\wedge T}^2-(\tau_x\wedge T)]=B_0^2-0=0$$

und damit

$$\mathbb{E}[\tau_x\wedge T]=\mathbb{E}[B_{\tau_x\wedge T}^2]=x^2\mathbb{P}(\tau_x\leqslant T)+\mathbb{E}[B_T^2\mathbb{1}_{\{\tau_x>T\}}]$$

Für $T\to\infty$ erhalten wir mit monotoner Konvergenz und wegen $\mathbb{E}[B_T^2\mathbb{1}_{\{\tau_x>T\}}]\leqslant x^2\mathbb{P}(\tau_x>T)\to 0$:

$$\mathbb{E}[\tau_x]=x^2.$$

4.4 Schwache Konvergenz und der Satz von Donsker

Wir wollen nun die Brownsche Bewegung als kanonischen Prozess auf dem Raum

$$\Omega=C([0,\infty))$$

der stetigen Pfade auf $\mathbb{R}_+$ betrachten. Wir erinnern uns, dass $C([0,T])$ bzw. $C([0,\infty))$ ausgestattet mit der Supremumsnorm $\|f\|_\infty=\sup_{t\in[0,T]}|f(t)|$ (⇝gleichmäßige Konvergenz) bzw. der Metrik $d(f,g):=\sum_{n\geqslant 1}2^{-n}(1\wedge\sup_{t\in[0,n]}|f(t)-g(t)|)$ (⇝gleichmäßige Konvergenz auf Kompakta) separable Banachräume (=polnischer Raum) sind, deren Borel-σ-Algebren durch die Koordinatenprojektionen erzeugt werden:

$$\mathscr{B}_{C([0,T])}=\sigma\big(\pi_t,t\in[0,T]\big).\qquad \mathscr{B}_{C([0,\infty))}=\sigma\big(\pi_t,t\geqslant 0\big),$$

wobei $\pi_t(f):=f(t)$ (Übung ☑). Insbesondere wird eine Verteilung $\mathbb{P}$ auf $C([0,\infty))$ durch die endlich-dimensionalen Verteilungen

$$\mathbb{P}(\pi_{\{t_1,\dots,t_n\}}^{-1}(B_n)),\qquad n\in\mathbb{N},B_n\in\mathscr{B}_{\mathbb{R}^n},t_1,\dots,t_n\geqslant 0$$

eindeutig beschrieben.

Definition 4.26. Sei $\mathbb{P}$ das Wahrscheinlichkeitsmaß auf $\Omega=C([0,\infty))$ mit $\mathcal{F}=\mathscr{B}_{C([0,\infty))}$, bzgl. dessen der kanonische Prozess $X=(X_t,t\geqslant 0)$ (d.h. $X_t=\pi_t$) eine Brownsche Bewegung ist. Dann heißt $\mathbb{P}$ Wiener-Maß und das Tripel $(\Omega,\mathcal{F},\mathbb{P})$ Wiener-Raum. Der Prozess X wird dann auch Wiener-Prozess genannt.

In diesem stochastischen Modell wird sofort klar, dass Funktionale wie das eingangs erwähnte $F(X):=\sup_{t\in[0,1]}X_t$ tatsächlich Zufallsvariablen auf dem Wiener-Raum sind.

Das Ziel dieses Abschnittes ist eine explizite Konstruktion des Wiener-Maßes. Das gibt uns insbesondere eine Möglichkeit, die Brownsche Bewegung besser zu verstehen und zu simulieren. Hierzu werden wir nachweisen, dass der Wiener-Prozess der schwache Grenzwert (d.h. Grenzwert bzgl. der schwachen Konvergenz = Konvergenz in Verteilung) von geeignet skalierten Irrfahren auf dem Raum der stetigen Funktionen ist, siehe Abbildung 6. Wir wiederholen (bzw. verallgemeinern) hierzu die Konvergenz in Verteilung.

Definition 4.27. Sei $(S,\mathscr{B}_S)$ ein metrischer Raum mit der Borel-σ-Algebra. Eine Folge von Wahrscheinlichkeitsmaßen $(\mathbb{P}_n)$ auf $(S,\mathscr{B}_S)$ konvergiert schwach gegen ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(S,\mathscr{B}_S)$, falls

$$\forall f\in C_b(S):\quad \int_S f\,\mathrm{d}\mathbb{P}_n\overset{n\to\infty}{\longrightarrow}\int_S f\,\mathrm{d}\mathbb{P},$$

wobei $C_b(S):=\{f\colon S\to\mathbb{R}|f\text{ stetig, beschränkt}\}$. Wir schreiben $\mathbb{P}_n\Rightarrow\mathbb{P}$ (alternativ $\mathbb{P}_n\overset{w}{\to}\mathbb{P}$ oder $\mathbb{P}_n\overset{d}{\to}\mathbb{P}$). $(S,\mathscr{B}_S)$-wertige Zufallsvariablen X_n konvergieren in Verteilung gegen eine $(S,\mathscr{B}_S)$-wertige Zufallsvariable X, falls $\mathbb{P}^{X_n}\Rightarrow\mathbb{P}^X$, d.h. falls

$$\forall f\in C_b(S):\quad \mathbb{E}[f(X_n)]\to\mathbb{E}[f(X)].$$

Wir schreiben auch $X_n\Rightarrow X$.

Bemerkung 4.28. Der schwache Grenzwert ist, falls existent, eindeutig. Dies folgt durch Approximation der Indikatorfunktionen über abgeschlossenen Mengen durch stetige Funktionen.

Beispiel 4.29.

(i) Aus der „Mathematischen Stochastik“ ist bekannt, dass $\mathbb{P}_n \Rightarrow \mathbb{P}$ auf $(\mathbb{R}, \mathscr{B}_\mathbb{R})$ äquivalent zur punktweisen Konvergenz der Verteilungsfunktionen $\mathbb{P}_n((-\infty, x]) \overset{n\to\infty}{\longrightarrow} F(x) := \mathbb{P}((-\infty, x])$ an allen Stetigkeitsstellen x von F ist.

(ii) Für eine Folge von Dirac-Maßen $(\delta_{y_n})_{n\geqslant 1}$ auf $(S, \mathscr{B}_S)$ mit $y_n \to y$ (bzgl. der Metrik auf S) für ein $y \in S$ gilt $\delta_{y_n} \Rightarrow \delta_y$, denn für alle $f \in C_b(S)$ gilt

$$\int f \,\mathrm{d}\delta_{y_n} = f(y_n) \to f(y) = \int f \,\mathrm{d}\delta_y, \qquad n \to \infty.$$

(iii) Aus der Konvergenz in Totalvariationsnorm $\|\mathbb{P}_n - \mathbb{P}\|_{TV} := \sup_{A\in\mathscr{B}_S} |\mathbb{P}_n(A) - \mathbb{P}(A)| \to 0$ folgt $\mathbb{P}_n \Rightarrow \mathbb{P}$, denn für alle $f \in C_b(S)$ gilt mit dem dominierenden Maß $\nu_n := \mathbb{P}_n - \mathbb{P}_n$ und Radon-Nikodym-Dichten $p_n := \frac{d\mathbb{P}_n}{d\nu}, p := \frac{d\mathbb{P}}{d\nu}$:

$$\begin{aligned}\Big|\int f \,\mathrm{d}\mathbb{P}_n - \int f \,\mathrm{d}\mathbb{P}\Big| &= \Big|\int_{p_n>p} f(p_n - p) d\nu_n + \int_{p_n<p} f(p_n - p) d\nu_n\Big| \\ &\leqslant \int_{p_n>p} |f|(p_n - p) d\nu_n + \int_{p_n<p} |f|(p - p_n) d\nu_n \\ &\leqslant \|f\|_\infty \big((\mathbb{P}_n - \mathbb{P})(p_n > p) + (\mathbb{P} - \mathbb{P}_n)(p_n < p)\big) \\ &\leqslant 2\|f\|_\infty \|\mathbb{P}_n - \mathbb{P}\|_{TV} \to 0, \qquad n \to \infty.\end{aligned}$$

Beachte, dass $\|\delta_{y_n} - \delta_y\|_{TV} = 1$ für alle $y_n \neq y$, sodass schwache Konvergenz tatsächlich schwächer ist als Konvergenz in Totalvariationsnorm.

Lemma 4.30. *Für Wahrscheinlichkeitsmaße $(\mathbb{P}_n)_{n\geqslant 0}$ und $\mathbb{P}$ auf $(S, \mathscr{B}_S)$ folgt $\mathbb{P}_n \Rightarrow \mathbb{P}$ bereits aus*

$$\int_S f \mathrm{d}\mathbb{P}_n \overset{n\to\infty}{\longrightarrow} \int_S f \mathrm{d}\mathbb{P}$$

für alle beschränkten, Lipschitz-stetigen Funktionen $f \colon S \to \mathbb{R}$, d.h. $\exists L > 0 : \forall x, y \in S : |f(x) - f(y)| \leqslant L d(x, y)$.

Beweis. Wir werden zeigen, dass für jedes $f \in C_b(S)$ Lipschitz-stetige Funktionen $(g_k)_{k\geqslant 1}$ existieren mit $g_k \uparrow f$ und $g_k \geqslant -\|f\|_\infty > -\infty$. Daraus folgt für alle $k \geqslant 1$

$$\liminf_{n\to\infty} \int_S f \mathrm{d}\mathbb{P}_n \geqslant \liminf_{n\to\infty} \int_S g_k \mathrm{d}\mathbb{P}_n = \int_S g_k \mathrm{d}\mathbb{P}.$$

und mit monotoner Konvergenz

$$\liminf_{n\to\infty} \int_S f \mathrm{d}\mathbb{P}_n \geqslant \lim_{k\to\infty} \int_S g_k \mathrm{d}\mathbb{P} = \int f \mathrm{d}\mathbb{P}.$$

Zusammen mit der analogen Aussage für $-f$, erhalten wir die Behauptung.

Es bleibt die Folge $(g_k)_{k\geqslant 1}$ zu konstruieren. Durch Betrachten von $\tilde{f} = f + \|f\|_\infty$ können können wir o.B.d.A. $f \geqslant 0$ annehmen. Für $r \in \mathbb{Q}, r > 0$ und $m \in \mathbb{N}$ setze

$$h_{m,r}(x) := \min\Big\{r, m \cdot \min_{y\in S: f(y)\leqslant r} d(x, y)\Big\} \in [0, r].$$

Dann folgt aus der umgekehrten Dreiecksungleichung für d, dass

$$|h_{m,r}(x_1) - h_{m,r}(x_2)| \leqslant m d(x_1, x_2)$$

und es gilt $h_{m,r}(x) \leqslant f(x)$, denn

$$\begin{aligned} f(x) \leqslant r \quad &\Rightarrow \quad h_{m,r}(x) = 0, \\ f(x) > r \quad &\Rightarrow \quad h_{m,r}(x) \leqslant r. \end{aligned}$$

Für jedes $x \in S$ und $\varepsilon > 0$ existiert ein $r \in \mathbb{Q}$ mit $f(x) - \varepsilon < r < f(x)$, sodass

$$\lim_{m\to\infty} h_{m,r}(x) = r > f(x) - \varepsilon$$

und damit

$$\sup\left\{h_{m,r}(x) : m \in \mathbb{N}, r \in \mathbb{Q}\right\} = f(x).$$

Wählen wir nun eine Abzählung $(p_l)_{l\geqslant 1}$ von $\mathbb{Q}_+ \times \mathbb{N}$, dann erfüllt die Folge

$$g_k := \max\{h_{p_1}, \dots, h_{p_k}\}, \quad k \geqslant 1,$$

die gewünschten Eigenschaften. □

Tragen wir einige Eigenschaften der schwachen Konvergenz zusammen:

Lemma 4.31. $(X_n)_{n\geqslant 1}$, $(Y_n)_{n\geqslant 1}$ *und* X *seien* $(S, \mathscr{B}_S)$*-wertige Zufallsvariablen. Es gilt:*

(i) Continuous mapping*: Für zwei metrische Räume* S, T *und eine stetige Abbildung* $g\colon S \to T$ *folgt aus* $X_n \Rightarrow X$ *die Konvergenz* $g(X_n) \Rightarrow g(X)$.

(ii) Slutskys Lemma*: Aus* $X_n \Rightarrow X$ *und* $d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$ *folgt* $Y_n \Rightarrow X$.

(iii) $d(Y_n, X) \xrightarrow{\mathbb{P}} 0$ *impliziert* $Y_n \Rightarrow X$.

Beweis. (i) Da $f \circ g \in C_b(S)$ für jedes $f \in C_b(T)$, folgt die Behauptung aus der Definition.

(ii) Gemäß Lemma 4.30 genügt es, eine beschränkte, Lipschitz-stetige Funktion mit $|f(x) - f(y)| \leqslant L d(x, y)$ für alle $x, y \in S$ und ein $L > 0$ zu betrachten. Dann gilt für alle $\varepsilon > 0$:

$$\begin{aligned} \limsup_{n\to\infty} |\mathbb{E}[f(X_n) - f(Y_n)]| \leqslant& \limsup_{n\to\infty} \mathbb{E}[|f(X_n) - f(Y_n)|] \\ \leqslant& L\varepsilon + \limsup_{n\to\infty} \mathbb{E}[|f(X_n) - f(Y_n)|\mathbb{1}_{\{d(X_n,Y_n)>\varepsilon\}}] \\ \leqslant& L\varepsilon + 2\|f\|_\infty \limsup_{n\to\infty} \mathbb{P}(d(X_n, Y_n) > \varepsilon) = L\varepsilon. \end{aligned}$$

Da ε beliebig war, folgt $\lim_{n\to\infty} \mathbb{E}[f(Y_n)] = \lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$.

(iii) folgt aus (ii) mit $X_n = X$. □

Kommen wir nun zur schwachen Konvergenz auf $(C([0,T]), \|\cdot\|_\infty)$ bzw. $(C([0,\infty)), \|\cdot\|_{c,\infty})$.

Definition 4.32. Seien $(\mathbb{P}_n)_{n\geqslant 1}$ und $\mathbb{P}$ Wahrscheinlichkeitsmaße auf $C([0,T])$ (bzw. $C([0,\infty))$). Wir sagen, dass die endlichdimensionalen Verteilungen (finite dimensional distributions) von $(\mathbb{P}_n)_{n\geqslant 1}$ gegen die von $\mathbb{P}$ konvergieren, falls für alle $k \in \mathbb{N}$ und $t_1, \dots, t_k \in [0,T]$ (bzw. $t_1, \dots, t_k \geqslant 0$)

$$\mathbb{P}_n^{\pi_{\{t_1,\dots,t_k\}}} \Longrightarrow \mathbb{P}^{\pi_{\{t_1,\dots,t_k\}}}$$

gilt. Wir schreiben kurz $\mathbb{P}_n \overset{\text{fdd}}{\Rightarrow} \mathbb{P}$.

Bemerkung 4.33. Gilt $\mathbb{P}_n \overset{\text{fdd}}{\Rightarrow} \mathbb{P}'$ und $\mathbb{P}_n \overset{\text{fdd}}{\Rightarrow} \mathbb{P}''$, folgt $\mathbb{P}' = \mathbb{P}''$, denn die Projektionen erzeugen $\mathscr{B}_{C([0,T])}$ bzw. $\mathscr{B}_{C([0,\infty))}$ und die (schwachen) Grenzwerte der endlichdimensionalen Verteilung sind eindeutig.

Aus dem Continuous-Mapping-Theorem folgt sofort, dass schwache Konvergenz auf $(C([0,\infty)), \|\cdot\|_{c,\infty})$ die Konvergenz der endlichdimensionalen Verteilungen impliziert. Die Umgegkehrte Richtung gilt nur unter einer Zusatzannahme.

Definition 4.34. Es sei S ein metrischer Raum mit Borel-σ-Algebra $\mathscr{B}_S$. Eine Familie $(\mathbb{P}_i)_{i\in I}$ von Wahrscheinlichkeitsmaßen auf $(S, \mathscr{B}_S)$ heißt (schwach) relativkompakt, falls jede Folge $(\mathbb{P}_{i_k})_{k\geqslant 1}$eine schwach konvergente Teilfolge $(\mathbb{P}_{i_{k_l}})_{l\geqslant 1}$ besitzt. Die Familie $(\mathbb{P}_i)_{i\in I}$ heißt (gleichmäßig) straff, falls für jedes $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subseteq S$ existiert, sodass für alle $i \in I$ $\mathbb{P}_i(K_\varepsilon) \geqslant 1 - \varepsilon$ gilt.

Beispiel 4.35.

(i) Für eine Familie von reellen Zufallsvariablen $X_i, i \in I$, sind die Verteilungen $(\mathbb{P}^{X_i})_{i\in I}$ straff, falls

 (a) $\lim_{R\to\infty} \sup_{i\in I} \mathbb{P}(|X_i| > R) = 0$ oder

 (b) $\sup_{i\in I} \mathbb{E}[|X_i|^p] < \infty$ für ein $p > 0$ (L^p-beschränkt).

(ii) Die Folge $(\delta_n)_{n\geqslant 1}$ von Dirac-Maßen besitzt weder schwach konvergente Teilfolgen noch ist sie straff.

Ein zentrales Resultat für die schwache Konvergenz ist folgender Satz, den wir hier ohne Beweis angeben:

Satz 4.36 (Prohorov)**.** *Auf einem polnischen Raum S ist eine Familie $(\mathbb{P}_i)_{i\in I}$ genau dann relativkompakt, wenn sie straff ist.*

Beweisskizze. Die Hinrichtung ist relativ einfach zu zeigen (unter Verwendung von Portemanteaus Satz). Wir werden hier nur die Rückrichtung im Spezialfall $S = \mathbb{R}$ zeigen. Für den allgemeinen Fall sei auf Klenke (2006, Satz 13.29) verwiesen.

Es sei also $(\mathbb{P}_i)_{i\in I}$ eine straffe Familie von Wahrscheinlichkeitsmaßen auf $\mathbb{R}$ und $(\mathbb{P}_n)_{n\geqslant 1}$ eine Teilfolge. Bezeichnen wir die Verteilungsfunktionen mit $F_n(x) := \mathbb{P}_n((-\infty, x]), x \in \mathbb{R}$, dann müssen wir die Existenz einer Teilfolge $(F_{n_k})_{k\geqslant 1}$ und einer Verteilungsfunktion F zeigen mit $F_{n_k}(x) \to F(x)$ für alle Stetigkeitsstellen $x \in \mathbb{R}$ von F.

Sei $\mathbb{Q} = \{q_1, q_2, \dots\}$ eine Abzählung von $\mathbb{Q}$. Nach dem Satz von Bolzano-Weierstraß („Jede beschränkte Folge aus $\mathbb{R}^n$ besitzt eine konvergente Teilfolge.“) besitzt die Folge $(F_n(q_1))_{n\geqslant 1}$ eine konvergente Teilfolge $(F_{n_k^1}(q_1))_{k\geqslant 1}$. Induktiv folgt für jedes $l \geqslant 1$ die Existenz einer Teilfolge $(n_k^{l+1})_{k\geqslant 1}$ von $(n_k^l)_{k\geqslant 1}$, sodass $(F_{n_k^{l+1}}(q_{l+1}))_k$ konvergiert. Damit konvergiert $(F_{n_k^l}(q_l))_{k\geqslant 1}$ für jedes $l \geqslant 1$. Für die Diagonalfolge $n_k := n_k^k, k \geqslant 1$, konvergiert $(F_{n_k}(q))_{k\geqslant 1}$ für jedes $q \in \mathbb{Q}$.

Wir definieren nun $\widetilde{F}(q) := \lim_{k\to\infty} F_{n_k}(q)$ für alle $q \in \mathbb{Q}$ und

$$F(x) := \inf \big\{\widetilde{F}(q) : q \in \mathbb{Q}, q > x\big\}. \tag{9}$$

Da $\widetilde{F}$ monoton wachsend ist, ist F rechtsstetig und monoton wachsend. Ist F an $x \in \mathbb{R}$ stetig, so existieren für jedes $\varepsilon > 0$ Zahlen $q_\pm \in \mathbb{Q}$ mit

$$q_- < x < q_+ \quad \text{und} \quad \widetilde{F}(q_+) - \varepsilon \leqslant F(x) \leqslant \widetilde{F}(q_-) + \varepsilon.$$

Damit gilt

$$\limsup_{k\to\infty} F_{n_k}(x) \leqslant \limsup_{k\to\infty} F_{n_k}(q_+) = \widetilde{F}(q_+) \leqslant F(x) + \varepsilon$$

also $\limsup_{k\to\infty} F_{n_k}(x) \leqslant F(x)$. Analog folgt aus $\liminf_{k\to\infty} F_{n_k}(x) \geqslant \widetilde{F}(q_-) \geqslant F(x) - \varepsilon$, dass $\liminf_{k\to\infty} F_{n_k}(x) \geqslant F(x)$. Wir haben also $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ gezeigt.

Es bleibt nur zu zeigen, dass $\lim_{x\to\infty} F(x) = 1$ (und analog $\lim_{x\to-\infty} F(x) = 0$), wobei hier $\leqslant 1$ klar ist. Um

$$\lim_{x\to\infty} F(x) \geqslant \limsup_{k\to\infty} \lim_{x\to\infty} F_{n_k}(x) = 1$$

nachzuweisen, verwenden wir die Straffheit. Diese impliziert, dass für jedes $\varepsilon > 0$ ein $K > 0$ existiert mit

$$1 - F_n(x) < \varepsilon \quad \forall n \geqslant 1, x > K.$$

Da F als monotone, beschränkte Funktion nur abzählbar viele Sprungstellen haben kann, existiert eine Stetigkeitsstelle $x^* > K$ von F, sodass

$$1 \leqslant \limsup_{k\to\infty} F_{n_k}(x^*) + \varepsilon = F(x^*) + \varepsilon \leqslant \lim_{x\to\infty} F(x) + \varepsilon. \qquad \square$$

Korollar 4.37. *Seien $(\mathbb{P}_n)_{n\geqslant 1}$ und $\mathbb{P}$ Wahrscheinlichkeitsmaße auf $C([0,\infty))$. Dann sind äquivalent:*

(i) $\mathbb{P}_n \overset{fdd}{\Rightarrow} \mathbb{P}$ *für $n\to\infty$ und $(\mathbb{P}_n)_{n\geqslant 1}$ ist straff.*

(ii) $\mathbb{P}_n \Rightarrow \mathbb{P}$.

Beweis. (ii)⇒(i): Aus dem Satz von Prohorov folgt direkt die Straffheit. Da $\pi_{\{t_1,\dots,t_k\}}$ stetig ist, folgt aus dem Continuous-Mapping-Satz und $\mathbb{P}_n \Rightarrow \mathbb{P}$ die Konvergenz der endlichdimensionalen Verteilungen.

(i)⇒(ii): Nach dem Satz von Prohorov ist $(\mathbb{P}_n)_{n\geqslant 1}$ relativkompakt, d.h. alle Teilfolgen haben konvergierende Teilteilfolgen. Nach Bemerkung 4.33 stimmen alle Grenzwerte (= alle Häufungspunkte) überein, sodass $(\mathbb{P}_n)_{n\geqslant 1}$ selbst schwach konvergiert. $\square$

Es folgt ein nützliches hinreichendes Kriterium für Relativkompaktheit:

Satz 4.38. *Für jedes $f \in C([0,T])$ und $\delta > 0$ definieren wir das Stetigkeitsmodul*

$$\omega_\delta(f) := \sup\big\{|f(s) - f(t)| : s,t \in [0,T], |s-t| \leqslant \delta\big\}.$$

Eine Folge von Wahrscheinlichkeitsmaßen $(\mathbb{P}_n)_{n\geqslant 1}$ auf $\mathscr{B}_{C([0,T])}$ ist genau dann straff, wenn

(i) $\lim_{R\to\infty} \sup_n \mathbb{P}_n\big(f \in C([0,T]) : |f(0)| > R\big) = 0$ *und*

(ii) $\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}_n\big(f \in C([0,T]) : \omega_\delta(f) > \varepsilon\big) = 0$ *für jedes $\varepsilon > 0$.*

Bedingung (ii) wird impliziert von

(ii') $\lim_{\delta\to 0} \limsup_{n\to\infty} \sup_{t\in[0,T-\delta]} \frac{1}{\delta}\mathbb{P}_n\big(f \in C([0,T]) : \sup_{s\in[t,t+\delta]} |f(s) - f(t)| \geqslant \varepsilon\big) = 0$ *für jedes $\varepsilon > 0$.*

Man beachte, dass $\{\omega_\delta(f) > \varepsilon\} \in \mathscr{B}_{C([0,T])}$ gilt, da $f \mapsto \omega_\delta(f)$ stetig ist.

Beweis. Die Charakterisierung von Straffheit mittels (i) und (ii) folgt aus dem **Satz von Arzelà-Ascoli** (vgl. Werner (2000, Satz II.3.4)): *Eine Teilmenge $A \subseteq C([0,T])$ ist relativkompakt (d.h. der Abschluss $\overline{A}$ ist kompakt), genau dann wenn* $\sup_{f\in A} |f(0)| < \infty$ *und* $\lim_{\delta\to 0} \sup_{f\in A} \omega_\delta(f) = 0$ *(gleichradige Stetigkeit).*

„$\Longrightarrow$“ Da $(\mathbb{P}_n)$ straff ist, existiert für alle $\eta > 0$ ein kompaktes $K_\eta \subseteq C([0,T])$ mit $\mathbb{P}_n(K_\eta) > 1-\eta$. Da K_η insbesondere relativ kompakt ist, folgt aus dem Satz von Arzelà-Ascoli

$$K_\eta \subseteq \{f : |f(0)| \leqslant R_\eta\} \cap \{f : \omega_{\delta_{\eta,\varepsilon}}(f) \leqslant \varepsilon\}$$

für R_η hinreichend groß und $\delta_{\eta,\varepsilon}$ hinreichend klein. Es bleibt $\eta \to 0$ zu betrachten.

„$\Longleftarrow$“ Wir wählen $R > 0$ und δ_k für $k \geqslant 1$ so, dass

$$\mathbb{P}_n(f : |f(0)| \leqslant R) \geqslant 1 - \frac{\eta}{2} \quad \text{und} \quad \limsup_{n\to\infty} \mathbb{P}_n\big(f : \omega_{\delta_k}(f) \leqslant \frac{1}{k}\big) \geqslant 1 - \eta 2^{-k-1}.$$

Nach Arzelà-Ascoli ist $K_\eta := \{f : |f(0)| \leqslant R\} \cap \bigcap_{k\geqslant 1}\{f : \omega_{\delta_k}(f) \leqslant \frac{1}{k}\}$ relativkompakt. Somit folgt die Straffheit aus der Kompaktheit von $\overline{K}_\eta$ und

$$\mathbb{P}_n(\overline{K}_\eta^c) \leqslant \mathbb{P}_n(K_\eta^c) \leqslant \frac{\eta}{2} + \eta \sum_{k\geqslant 1} 2^{-k-1} = \eta.$$

Die Implikation (ii')⇒(ii) folgt durch Zerlegung von $[0,T]$ in hinreichend kleine Teilintervalle: Seien $\varepsilon, \eta > 0$ beliebig. Wähle $m, n_0 \in \mathbb{N}$, sodass

$$\frac{1}{m^{-1}}\mathbb{P}_n\Big(f : \sup_{s\in[t,t+m^{-1}]} |f(s)-f(t)| \geqslant \frac{\varepsilon}{2}\Big) \leqslant \frac{\eta}{2}, \qquad \text{für alle } t \in [0, T-m^{-1}], n \geqslant n_0.$$

Falls $\omega_{(2m)^{-1}}(f) \geqslant \varepsilon$ für ein $f \in C([0,T])$, dann gibt es ein $t \leqslant s$ mit $|f(t)-f(s)| \geqslant \varepsilon$ und $|t-s| \leqslant \frac{1}{2m}$. Wir wählen nun $k \in \mathbb{N}_0$ mit $\frac{k}{2m} \leqslant t \leqslant s \leqslant \frac{k}{2m} + \frac{1}{m}$. Aus der Dreiecksungleichung folgt

$$|f(t) - f(\frac{k}{2m})| \geqslant \frac{\varepsilon}{2} \quad \text{oder} \quad |f(s) - f(\frac{k}{2m})| \geqslant \frac{\varepsilon}{2}.$$

Wir erhalten

$$\{f : \omega_{(2m)^{-1}}(f) \geqslant \varepsilon\} \subseteq \bigcup_{k=1}^{\lfloor 2mT\rfloor -1} \Big\{f : \sup_{\frac{k}{2m}\leqslant s\leqslant \frac{k}{2m.}+\frac{1}{m}} \big|f(s) - f(\frac{k}{2m})\big| \geqslant \varepsilon/2\Big\}$$

und somit folgt für alle $n \geqslant n_0$

$$\mathbb{P}_n\big(f : \omega_{(2m)^{-1}}(f) \geqslant \varepsilon\big) \leqslant (2mT)\cdot m^{-1} \cdot \frac{\eta}{2} \leqslant T\eta.$$

Da η beliebig war, folgt (ii). □

Bemerkung 4.39.

(i) Da $C([0,\infty))$ mit der gleichmäßigen Konvergenz auf Kompakta versehen ist, ist eine Folge $(\mathbb{P}_n)_{n\geqslant 1}$ auf $\mathscr{B}_{C([0,\infty))}$ straff, falls (i) und (ii') für jedes $T > 0$ erfüllt sind.

(ii) Sind $(X^n_t, t \in [0,T])$ stetige stochastische Prozesse, so finden wir mit Hilfe des Satzes von Kolmogorov-Chentsov folgendes *Momentenkriterium*: $(\mathbb{P}^{X^n})_{n\geqslant 1}$ ist auf $\mathscr{B}_{C([0,T])}$ straff, falls:

  (a) $\lim_{R\to\infty} \sup_{n\geqslant 1} \mathbb{P}(|X^n_0| > R) = 0$ und

  (b) Es existieren $\alpha, \beta, K > 0$, sodass für alle $n \geqslant 1, s, t \in [0,T]$:

$$\mathbb{E}\big[|X^n_s - X^n_t|^\alpha\big] \leqslant K \cdot |s-t|^{1+\beta}.$$

Wir haben jetzt alle nötigen Vorbereitungen und kommen zum Hauptresultat dieses Abschnittes. Betrachten wir i.i.d. $(X_k)_{k\geqslant 1}$ aus $L^2(\mathbb{P})$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Dann impliziert der zentrale Grenzwertsatz für die normierten Partialsummen

$$S_n := \sum_{k=1}^n \frac{X_k - \mu}{\sigma},$$

$\mu := \mathbb{E}[X_1], \sigma^2 := \mathrm{Var}(X_1)$, dass

$$n^{-1/2} S_n \Rightarrow \mathcal{N}(0,1).$$

Wir interpretieren nun den Prozess $(S_n, n \geqslant 0)$ mit $S_0 := 0$ als Irrfahrt. Um einen stetigen Prozess zu erhalten, interpolieren wir:

$$Y^n_t := \frac{1}{\sqrt{n}} S_{\lfloor nt\rfloor} + \frac{nt - \lfloor nt\rfloor}{\sqrt{n}} \frac{X_{\lfloor nt\rfloor+1} - \mu}{\sigma}, \qquad t \in [0,1].$$

Es gilt also $Y^n_{k/n} = n^{-1/2} S_k$ und Y^n ist linear auf jedem Intervall $[\frac{k}{n}, \frac{k+1}{n}]$, siehe Abbildung 6.

Lemma 4.40. *Seien $(X_k)_{k\geqslant 1} \subseteq L^2(\mathbb{P})$ i.i.d. mit $\mathbb{E}[X_1] = 0$ und $\mathrm{Var}(X_1) = 1$. Dann gilt $\mathbb{P}^{Y^n} \overset{fdd}{\Rightarrow} \mathbb{P}^B$ für eine Brownsche Bewegung B.*

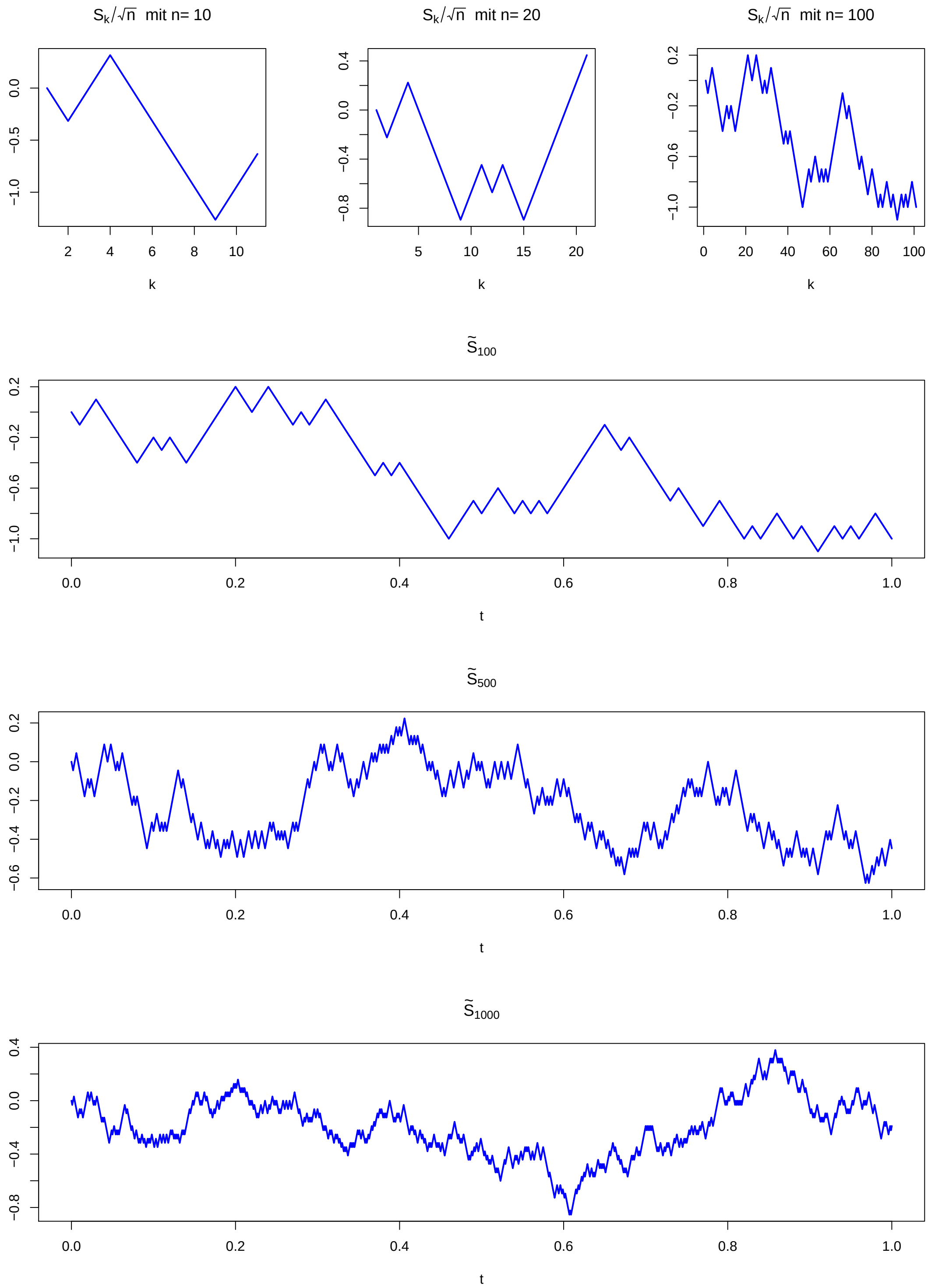

Abbildung 6: Approximation der Brownschen Bewegung durch eine Bernoulli-Irrfahrt. Die erste Zeile zeigt die (normierte und linear interpolierte) Irrfahrt $(\frac{S_k}{\sqrt{n}})_{k=1,\dots,n}$ für $n \in \{10, 20, 100\}$. Zeilen zwei bis vier zeigen den stetigen, reskalierten Prozess $\widetilde{S}_n = (Y_t^n : t \in [0, 1])$ für $n \in \{100, 500, 1000\}$. Alle Abbildungen wurden aufgrund einer Realisierung der Zufallsvariablen $(X_k)_{k=1,\dots,1000}$ erzeugt.

Beweis. Da $Y_0^n = B_0 = 0$, können wir $0 = t_0 < t_1 < \cdots < t_m \leqslant 1$ betrachten. Mit

$$Z_k^n := \sum_{i=\lfloor nt_{k-1}\rfloor+1}^{\lfloor nt_k\rfloor} \frac{X_i}{\sqrt{n}}, \quad k = 1, \dots, m,$$

erhalten wir die Darstellung

$$Y_{t_k}^n = \sum_{l=1}^{k} Z_l^n + \frac{nt_k - \lfloor nt_k\rfloor}{\sqrt{n}} X_{\lfloor nt_k\rfloor+1}.$$

Aus dem zentralen Grenzwertsatz folgt

$$\sum_{i=\lfloor nt_{k-1}\rfloor+1}^{\lfloor nt_k\rfloor} \frac{X_i}{\sqrt{\lfloor nt_k\rfloor - \lfloor nt_{k-1}\rfloor}} \Rightarrow \mathcal{N}(0,1).$$

Wegen $\frac{\lfloor nt_k\rfloor - \lfloor nt_{k-1}\rfloor}{n} \to t_k - t_{k-1}$ für $n \to \infty$, folgt aus Slutskys Lemma

$$Z_k^n \Rightarrow \mathcal{N}(0, t_k - t_{k-1}).$$

Da die Familie $(Z_k^n)_{k=1,\dots,m}$ unabhängig ist für jedes $n \geqslant 1$ erhalten wir den Grenzwert

$$(Z_1^n, \dots, Z_m^n) \Rightarrow \mathcal{N}\big(0, \operatorname{diag}(t_1 - t_0, \dots, t_m - t_{m-1})\big).$$

Continuous-mapping impliziert

$$(Z_1^n, Z_1^n + Z_2^n, \dots, Z_1^n + \cdots + Z_m^n) \Rightarrow \mathbb{P}^{(\overline{Z}_1, \overline{Z}_1 + \overline{Z}_2, \dots, \overline{Z}_1 + \cdots + \overline{Z}_m)}$$

für unabhängige $\overline{Z}_k \sim \mathcal{N}(0, t_k - t_{k-1})$. Insbesondere gilt $(\overline{Z}_1, \overline{Z}_1 + \overline{Z}_2, \dots, \overline{Z}_1 + \cdots + \overline{Z}_m) \overset{d}{=} (B_{t_1}, B_{t_1} + B_{t_2} - B_{t_1}, \dots, B_{t_m})$. Wegen

$$\underbrace{\frac{nt_k - \lfloor nt_k\rfloor}{\sqrt{n}}}_{\leqslant 1/\sqrt{n}} \underbrace{X_{\lfloor nt_k\rfloor+1}}_{=\mathcal{O}_P(1)} \xrightarrow{P} 0,$$

liefert Slutskys Lemma schließlich $(Y_{t_1}^n, \dots, Y_{t_m}^n) \Rightarrow (B_{t_1}, \dots, B_{t_m})$. □

Der Satz von Donsker besagt, dass nicht nur die endlichdimensionalen Verteilungen konvergieren, sondern auch die Folge der Bildmaße $\mathbb{P}^{Y^n}$ auf $(C([0,1]), \mathscr{B}_{C([0,1])})$ schwach gegen das Wiener-Maß konvergiert.

Satz 4.41 (Donsker, funktionaler zentraler Grenzwertsatz)**.** *Es sei* $(X_k)_{k\geqslant 1} \subseteq L^2(\mathbb{P})$ *eine Folge von i.i.d. Zufallsvariablen mit* $\mathbb{E}[X_k] = 0$ *und* $\operatorname{Var}(X_k) = 1$. *Für* $S_n := \sum_{k=1}^n X_k$ *und*

$$Y_t^n := \frac{1}{\sqrt{n}} S_{\lfloor nt\rfloor} + \frac{nt - \lfloor nt\rfloor}{\sqrt{n}} X_{\lfloor nt\rfloor+1}, \qquad t \in [0,1],$$

gilt

$$Y^n \Longrightarrow B \quad \textit{in } C([0,1])$$

für eine Brownsche Bewegung $B = (B_t, t \in [0,1])$.

Bevor wir den Satz von Donsker beweisen, benötigen wir noch ein Hilfsresultat.

Lemma 4.42. *In der Situation von Satz 4.41 gilt für alle* $\lambda \geqslant \sqrt{2}$ *und* $N \in \mathbb{N}$ *folgende Maximalungleichung:*

$$\mathbb{P}\big(\max_{1\leqslant n\leqslant N} |S_n| \geqslant \lambda\sqrt{N}\big) \leqslant 2\mathbb{P}\big(|S_N| \geqslant (\lambda - \sqrt{2})\sqrt{N}\big).$$

Beweis. Sei $\tau := \inf\{n \geqslant 0 : |S_n| \geqslant \lambda\sqrt{N}\} \wedge (N+1)$. Dann gilt

$$\begin{aligned}\mathbb{P}\big(\max_{1\leqslant n\leqslant N}|S_n| \geqslant \lambda\sqrt{N}\big) &= \mathbb{P}(\tau \leqslant N)\\ &\leqslant \mathbb{P}\big(|S_N| \geqslant (\lambda-\sqrt{2})\sqrt{N}\big) + \sum_{n=1}^{N}\mathbb{P}\big(\tau = n, |S_N| < (\lambda-\sqrt{2})\sqrt{N}\big)\\ &\leqslant \mathbb{P}\big(|S_N| \geqslant (\lambda-\sqrt{2})\sqrt{N}\big) + \sum_{n=1}^{N}\mathbb{P}\big(\tau = n, |S_N - S_n| > \sqrt{2N}\big).\end{aligned}$$

Aufgrund der Unabhängigkeit der Ereignisse $\{\tau = n\}$ und $\{|S_N - S_n| > \sqrt{2}\sqrt{N}\}$ erhalten wir zusammen mit Markovs Ungleichung

$$\begin{aligned}\mathbb{P}\big(\tau = n, |S_N - S_n| > \sqrt{2}\sqrt{N}\big) &= \mathbb{P}(\tau = n)\mathbb{P}\big(|S_N - S_n| > \sqrt{2N}\big)\\ &\leqslant \frac{\mathbb{P}(\tau = n)\mathbb{E}[(S_N - S_n)^2]}{2N} \leqslant \frac{1}{2}\mathbb{P}(\tau = n).\end{aligned}$$

Somit gilt

$$\mathbb{P}(\tau \leqslant N) \leqslant \mathbb{P}\big(|S_N| \geqslant (\lambda-\sqrt{2})\sqrt{N}\big) + \frac{1}{2}\mathbb{P}(\tau \leqslant N).$$

Subtrahieren von $\mathbb{P}(\tau \leqslant N)$ ergibt die Behauptung. □

Beweis von Satz 4.41. Es bleibt nur die Straffheit von $(\mathbb{P}^{Y^n})_{n\geqslant 1}$ zu zeigen. Hierzu verwenden wir Satz 4.38, wobei (i) wegen $Y_0^n = 0$ für alle $n \geqslant 1$ trivial ist. Es bleibt also (ii') nachzuweisen:

$$\forall \varepsilon > 0: \quad \lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{t\in[0,1-\delta]}\frac{1}{\delta}\mathbb{P}\big(\sup_{t\leqslant s\leqslant t+\delta}|Y_s^n - Y_t^n| \geqslant \varepsilon\big) = 0.$$

Da Y^n stückweise linear ist, gilt für $t = \frac{k}{n}$ und $t + \delta = \frac{j}{n}$:

$$\sup_{t\leqslant s\leqslant t+\delta}|Y_s^n - Y_t^n| = \max_{1\leqslant i\leqslant j-k}\frac{|S_{k+i} - S_k|}{\sqrt{n}}.$$

Im allgemeineren Fall erhalten wir für $n \geqslant \frac{1}{\delta}$ und $t \in [\frac{k}{n}, \frac{k+1}{n}], t + \delta \in [\frac{j-1}{n}, \frac{j}{n}]$

$$\begin{aligned}\sup_{t\leqslant s\leqslant t+\delta}|Y_s^n - Y_t^n| &\leqslant \sup_{\frac{k}{n}\leqslant s\leqslant \frac{j}{n}}\big\{|Y_s^n - Y_{k/n}^n| + |Y_t^n - Y_{k/n}^n|\big\}\\ &\leqslant \frac{2}{\sqrt{n}}\max_{1\leqslant i\leqslant j-k}|S_{k+i} - S_k|\\ &\leqslant \frac{2}{\sqrt{n}}\max_{1\leqslant i\leqslant 3n\delta}|S_{k+i} - S_k| \overset{d}{=} \frac{2}{\sqrt{n}}\max_{1\leqslant i\leqslant 3n\delta}|S_i|.\end{aligned}$$

Mit Lemma 4.42 ergibt sich

$$\begin{aligned}\limsup_{n\to\infty}\sup_{t\in[0,1-\delta]}\frac{1}{\delta}\mathbb{P}\Big(\sup_{t\leqslant s\leqslant t+\delta}|Y_s^n - Y_t^n| \geqslant \varepsilon\Big) &\leqslant \limsup_{n\to\infty}\frac{1}{\delta}\mathbb{P}\Big(\frac{2}{\sqrt{n}}\max_{1\leqslant i\leqslant 3n\delta}|S_i| \geqslant \varepsilon\Big)\\ &\leqslant \limsup_{n\to\infty}\frac{2}{\delta}\mathbb{P}\Big(|S_{\lfloor 3n\delta\rfloor}| \geqslant \big(\frac{\varepsilon}{2\sqrt{3\delta}} - \sqrt{2}\big)\sqrt{\lfloor 3n\delta\rfloor}\Big).\end{aligned}$$

Nach dem zentralen Grenzwertsatz konvergiert $(\lfloor 3n\delta\rfloor)^{-1/2}S_{\lfloor 3n\delta\rfloor}$ gegen eine Standardnormalverteilung. Bezeichnen wir mit $\Phi(y) := \int_{-\infty}^{y}(2\pi)^{-1/2}e^{-x^2/2}\mathrm{d}x$ die Verteilungsfunktion von $\mathcal{N}(0,1)$, ergibt sich

$$\limsup_{n\to\infty}\sup_{t\in[0,1-\delta]}\frac{1}{\delta}\mathbb{P}\Big(\sup_{t\leqslant s\leqslant t+\delta}|Y_s^n - Y_t^n| \geqslant \varepsilon\Big) \leqslant \frac{2}{\delta}\cdot 2\Big(1 - \Phi\big(\frac{\varepsilon}{2\sqrt{3\delta}} - \sqrt{2}\big)\Big).$$

Da $1 - \Phi(y)$ exponentiell gegen 0 konvergiert für $y \to \infty$, konvergiert die rechte Seite für $\delta \to 0$ gegen 0. □

Bemerkung 4.43. Der Satz von Donsker liefert uns insbesondere einen alternativen Beweis der Existenz der Brownschen Bewegung. Er wird auch *funktionaler zentraler Grenzwertsatz* genannt, da Y^n als Prozess auf $C([0,1])$ gegen die Brownsche Bewegung konvergiert. Da die Grenzverteilung nicht von der Verteilung der (X_k) abhängt, also universell ist, spricht man vom *Invarianzprinzip.*

A Anhang: Gleichgradige Integrierbarkeit

Definition A.1. Eine Familie von Zufallsvariablen $(X_i)_{i\in i}$ mit einer beliebigen Indexmenge I heißt gleichgradig integrierbar, falls

$$\lim_{R\to\infty} \sup_{i\in I} \mathbb{E}\big[|X_i| \mathbb{1}_{\{|X_i|>R\}}\big] = 0.$$

Beispiel A.2. Es sei $(X_i)_{i\in I}$ eine Familie von Zufallsvariablen mit einer beliebigen Indexmenge I.

(i) Ist $(X_i)_{i\in I}$ gleichgradig integrierbar, so gilt $\sup_{i\in I} \mathbb{E}[|X_i|] < \infty$.

(ii) Gilt $\sup_{i\in I} \mathbb{E}[|X_i|^p] < \infty$ für ein $p > 1$, so ist $(X_i)_{i\in I}$ gleichgradig integrierbar.

(iii) Existiert ein $Y \in L^1(\mathbb{P})$, sodass $|X_i| \leqslant Y$ f.s. für alle $i \in I$ gilt, dann ist $(X_i)_{i\in I}$ gleichgradig integrierbar.

Folgende Sätze geben zwei nützliche Charakterisierungen von gleichgradiger Integrierbarkeit an:

Satz A.3. *Eine Menge $(X_i)_{i\in I} \subseteq L^1(\mathbb{P})$ mit beliebiger Indexmenge I ist genau dann gleichgradig integrierbar, wenn $\sup_{i\in I} \mathbb{E}[|X_i|] < \infty$ gilt sowie*

$$\forall \varepsilon > 0 \exists \delta_\varepsilon > 0 : \forall B \in \mathcal{F} \text{ mit } \mathbb{P}(B) < \delta_\varepsilon \text{ gilt } \sup_{i\in I} \mathbb{E}[|X_i| \mathbb{1}_B] < \varepsilon.$$

Beweis. Übung ☑. □

Satz A.4 (de la Vallée-Poussin)**.** *Eine Menge $(X_i)_{i\in I} \subseteq L^1(\mathbb{P})$ mit beliebiger Indexmenge I ist genau dann gleichgradig integrierbar, wenn es eine Funktion $\Psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ gibt mit $\Psi(x)/x \to \infty$ für $x \to \infty$ und $\sup_{i\in I} \mathbb{E}[\Psi(|X_i|)] < \infty$. Ψ kann sogar monoton wachsend und konvex gewählt werden.*

Beweis. Klenke (2006, Satz 6.19). □

Satz A.5. *Es seien $(X_n)_{n\geqslant 1}$ Zufallsvariablen, die stochastisch gegen ein X konvergieren: $X_n \xrightarrow{\mathbb{P}} X$. Dann folgt aus gleichgradiger Intergierbarkeit von $(X_n)_{n\geqslant 1}$ die Konvergenz in $L^1(\mathbb{P})$: $\mathbb{E}[|X_n - X|] \to 0$.*

Beweis. Nehmen wir an, X_n konvergiert nicht gegen X in $L^1(\mathbb{P})$. Dann existiert für jedes $\varepsilon > 0$ eine Teilfolge $(X_{n_k})_{k\geqslant 1}$, sodass

$$\mathbb{E}\big[|X_{n_k} - X|\big] \geqslant \varepsilon \quad \text{für alle } k \in \mathbb{N}.$$

Da $X_n \xrightarrow{\mathbb{P}} X$, besitzt die (jede) Teilfolge (X_{n_k}) eine Teilteilfolge $(X_{n'_k})_k$, die f.s. gegen X konvergiert. Nach dem Lemma von Fatou und aufgrund der Beschränktheit in $L^1(\mathbb{P})$ gilt dann

$$\mathbb{E}[|X|] \leqslant \liminf_{k\to\infty} \mathbb{E}[|X_{n'_k}|] < \infty.$$

Also ist $X \in L^1(\mathbb{P})$. Wir setzen nun $\varphi_M(x) := (-M) \vee (x \wedge M)$ und wählen $M > 0$ so, dass

$$\sup_k \mathbb{E}[|X_{n'_k}| \mathbb{1}_{\{|X_{n'_k}|>M\}}] + \mathbb{E}[|X| \mathbb{1}_{\{|X|>M\}}] < \varepsilon/2$$

(möglich dank gleichgradiger Integrierbarkeit). Dominierte Konvergenz liefert ein $n_0 \in \mathbb{N}$ mit

$$\mathbb{E}[|\varphi_M(X_{n'_k}) - \varphi_M(X)|] < \varepsilon/2 \quad \text{für alle } k \geqslant n_0.$$

Somit folgt

$$\mathbb{E}[|X_{n'_k} - X|] \leqslant \mathbb{E}[|\varphi_M(X_{n'_k}) - \varphi_M(X)|] + \mathbb{E}[|X_{n'_k}|\mathbb{1}_{\{|X_{n'_k}|>M\}}] + \mathbb{E}[|X|\mathbb{1}_{\{|X|>M\}}] < \varepsilon.$$

Dies widerspricht $\mathbb{E}\big[|X_{n_k} - X|\big] \geqslant \varepsilon$. □

Bemerkung A.6. Liegt die stochastische Konvergenz vor, so ist die L^1-Konvergenz sogar äquivalent zur gleichgradigen Integrierbarkeit (Satz von Vitali).

B Anhang: Martingale in diskreter Zeit

Der wichtige Begriff des Martingals dient der Formalisierung von fairen Spielen. Er ist insbesondere in der Finanzmathematik zentral, um den Handel mit Aktien an (vollständigen) Märkten zu modellieren.

B.1 Definition und Doob-Zerlegung

Definition B.1. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $I \subseteq \mathbb{R}$. Eine Familie $(\mathcal{F}_t)_{t\in I}$ von σ-Algebren mit $\mathcal{F}_t \subseteq \mathcal{F}$ für jedes $t \in I$ heißt *Filtration*, falls $\mathcal{F}_s \subseteq \mathcal{F}_t$ für alle $s, t \in I$ mit $s \leqslant t$. Eine Familie von reellwertigen Zufallsvariablen $X = (X_t)_{t\in I}$ (auf $(\Omega, \mathcal{F}, P)$) heißt *stochastischer Prozess*. X heißt *adaptiert* an die Filtration $(\mathcal{F}_t)_{t\in I}$, falls X_t $\mathcal{F}_t$-messbar ist für jedes $t \in I$.

Definition B.2. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $I \subseteq \mathbb{R}$ und $(\mathcal{F}_t)_{t\in I}$ eine Filtration. Ein stochastischer Prozess $X = (X_t)_{t\in I}$ heißt *Martingal* (bzw. *Submartingal* oder *Supermartingal*) bzgl. $(\mathcal{F}_t)_{t\in I}$, falls

(i) $X_t \in L^1(P), t \in I$,

(ii) X adaptiert an $(\mathcal{F}_t)_{t\in I}$ ist und

(iii) $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ P-f.s. für alle $s, t \in I, t \geqslant s$ (bzw. $\mathbb{E}[X_t|\mathcal{F}_s] \geqslant X_s$ für Sub- oder $\mathbb{E}[X_t|\mathcal{F}_s] \leqslant X_s$ für Supermartingale).

Bemerkung B.3. (Übung ☑)

(i) Jeder stochastische Prozess ist adaptiert an seine *natürliche Filtration* $\mathcal{F}_t = \sigma(X_s, s \leqslant t)$ für jedes $t \in I$.

(ii) Ist $I = \mathbb{N} \cup \{0\}$, so folgt die Martingaleigenschaft $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ für alle $s, t \in I$ mit $s \leqslant t$ bereits aus der Bedingung $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ (analog für Sub- und Supermartingale).

(iii) Sind $(X_t)_{t\in I}$ und $(Y_t)_{t\in I}$ zwei Martingale (oder Supermartingale) bzgl. $(\mathcal{F}_t)_{t\in I}$, so gilt

  (a) $(-X_t)_{t\in I}$ ist ein Martingal (bzw. ein Submartingal) bzgl. $(\mathcal{F}_t)_{t\in I}$.

  (b) Für alle $\alpha \in \mathbb{R}$ ist $(\alpha X_t + Y_t)_{t\in I}$ ein Martingal (bzw. ein Supermartingal, falls $\alpha \geqslant 0$).

Beispiel B.4.

(i) Sind $(X_i)_{i\in\mathbb{N}}$ unabhängige, integrierbare Zufallsvariablen mit $\mathbb{E}[X_i] = 0$ (bzw. $\geqslant 0$ oder $\leqslant 0$) für alle $i \in \mathbb{N}$, so ist $(S_n)_{n\geqslant 0}$ mit $S_n := \sum_{i=1}^n X_i, n \in \mathbb{N}$, und $S_0 := 0$ ein Martingal (bzw. Sub- oder Supermartingal) bzgl. seiner natürlichen Filtration. Im Spezialfall $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$ sprechen wir von einer (symmetrischen) Irrfahrt.

(ii) Faires Spiel: Sei X_0 das Startkapital eines Spielers und X_n modelliere das Kapital nach n Runden. Die Filtration $\mathcal{F}_n$ interpretieren wir als die Information der Ausgänge der ersten $n \in \mathbb{N}$ Runden. Dann ist $\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n]$ die Vorhersage des Gewinns (bzw. Verlustes) in Runde $n+1$ gegeben aller Informationen bis zur "Zeit" n. Ist $(X_n)_{n\geqslant 0}$ ein Martingal, handelt es sich also um ein faires Spiel (das Kapital bleibt im Durchschnitt konstant). Ein Submartingal ist vorteilhaft für den Spieler, wohingegen ein Supermartingal im Mittel zu Verlusten führt.

(iii) Ist $Z \in L^1(P)$ und $(\mathcal{F}_t)_{t\in I}$, so definiert $X_t = \mathbb{E}[Z|\mathcal{F}_t], t \in I$, ein Martingal bzgl. $(\mathcal{F}_t)_{t\in I}$.

(iv) Ist $\varphi\colon \mathbb{R} \to \mathbb{R}$ konvex und $(X_t)_{t\in I}$ ein Martingal mit $\varphi(X_t) \in L^1(P)$, dann ist $(\varphi(X_t))_{t\in I}$ ein Submartingal (Jensens Ungleichung). Insbesondere ist $(X_t^2)_{t\in I}$ ein Submartingal für jedes Martingal aus $L^2(P)$.

Wir konzentrieren uns im Folgenden auf diskrete Martingale indiziert durch $\mathbb{N} \cup \{0\}$ und betrachteten stets einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit einer Filtration $(\mathcal{F}_n)_{n\geqslant 0}$.

Definition B.5. Ein stochastischer Prozess $(Y_n)_{n\geqslant 0}$ heißt *vorhersagbar* (oder *previsibel*) bzgl. der Filtration $(\mathcal{F}_n)_{n\geqslant 0}$, falls Y_0 konstant ist und Y_n $\mathcal{F}_{n-1}$-messbar ist für jedes $n \in \mathbb{N}$. Für einen vorhersagbaren Prozess $(Y_n)_{n\geqslant 0}$ und ein Martingal $(X_n)_{n\geqslant 0}$ (bzgl. der selben Filtration) ist die *Martingaltransformation* oder das *diskrete stochastische Integral* $(Y \cdot X)_{n\geqslant 0}$ definiert durch

$$(Y \cdot X)_0 = 0 \quad \text{und} \quad (Y \cdot X)_n = \sum_{k=1}^{n} Y_k(X_k - X_{k-1}), \quad n \geqslant 1.$$

Lemma B.6. *Ist $(Y_n)_{n\geqslant 0}$ ein beschränkter, vorhersagbarer Prozess und $(X_n)_{n\geqslant 0}$ ein Martingal bzgl. einer Filtration $(\mathcal{F}_n)_{n\geqslant 0}$, so ist $(Y \cdot X)_{n\geqslant 0}$ ebenfalls ein Martingal.*

Beweis. Es gilt $Y_n \leqslant C$ für eine Konstante $C > 0$ und alle $n \geqslant 0$. Wir prüfen die Martingalbedingungen: Für $n \geqslant 1$ gilt

$$\mathbb{E}[|(Y \cdot X)_n|] \leqslant \sum_{k=1}^{n} \mathbb{E}[|Y_k| \cdot |X_k - X_{k-1}|] \leqslant \sum_{k=1}^{n} C\big(\mathbb{E}[|X_k|] + \mathbb{E}[|X_{k-1}|]\big) < \infty.$$

Weiterhin ist $(Y \cdot X)_n$ $\mathcal{F}_n$-messbar, da Y_k und X_k $\mathcal{F}_n$-messbar sind für alle $k \leqslant n$. Schließlich gilt

$$\begin{aligned}\mathbb{E}\big[(Y \cdot X)_{n+1} - (Y \cdot X)_n\big|\mathcal{F}_n\big] &= \mathbb{E}\big[Y_{n+1}(X_{n+1} - X_n)\big|\mathcal{F}_n\big] \\ &= Y_{n+1}\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] = 0.\end{aligned}$$ □

Die Martingaltransformation hat folgende Interpretation: X_n sei der Gewinn eines fairen Spiels bei einem Einsatz von 1€. Ein Spieler wählt nun vor dem $(n+1)$. Spiel seinen Einsatz Y_{n+1} für die nächste Runde. Dann ist $Y_{n+1}(X_{n+1} - X_n)$ der Gewinn im $(n+1)$. Spiel und folglich $(Y \cdot X)_n$ das Kapital nach n Runden. Für jede Strategie erhält sich also die "Fairness"-Eigenschaft.

Satz B.7 (Doob-Zerlegung)**.** *Sei $X = (X_n)_{n\geqslant 0}$ ein bzgl. der Filtration $(\mathcal{F}_n)_{n\geqslant 0}$ adaptierter Prozess mit $X_n \in L^1(P), n \geqslant 0$. Dann existiert eine eindeutige Zerlegung*

$$X = M + A,$$

wobei $A = (A_n)_{n\geqslant 0}$ vorhersagbar ist mit $A_0 = 0$ und $M = (M_n)_{n\geqslant 0}$ ein Martingal. Diese Zerlegung heißt Doob-Zerlegung *und es gilt*

$$A_n = \sum_{k=1}^{n} \mathbb{E}[X_k - X_{k-1}|\mathcal{F}_{k-1}], \quad n \geqslant 1.$$

X ist ein Submartingal genau dann, wenn A monoton wachsend ist.

Beweis. Existenz: Per Konstruktion ist A vorhersagbar. Wir definieren den adaptierten Prozess

$$M_0 := X_0 \quad \text{und} \quad M_{n+1} := M_n + \big(X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n]\big).$$

Dann gilt

$$\begin{aligned} M_n + A_n &= X_0 + \sum_{k=1}^{n} \big(X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]\big) + \sum_{k=1}^{n} \mathbb{E}[X_k - X_{k-1}|\mathcal{F}_{k-1}] \\ &= X_0 + \sum_{k=1}^{n} \big(X_k - \mathbb{E}[X_{k-1}|\mathcal{F}_{k-1}]\big) = X_n. \end{aligned}$$

Weiterhin ist

$$\begin{aligned} \mathbb{E}[|M_n|] &\leqslant \mathbb{E}[|X_0|] + \sum_{k=1}^{n} \mathbb{E}\big[|X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]|\big] \\ &\leqslant \mathbb{E}[|X_0|] + \sum_{k=1}^{n} \big(\mathbb{E}[|X_k|] + \mathbb{E}\big[\mathbb{E}[|X_k||\mathcal{F}_{k-1}]\big] \leqslant 2\sum_{k=0}^{n} \mathbb{E}[|X_k|] < \infty. \end{aligned}$$

Schließlich gilt

$$\mathbb{E}\big[M_{n+1} - M_n\big|\mathcal{F}_n\big] = \mathbb{E}\big[X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n]\big|\mathcal{F}_n\big] = 0.$$

Eindeutigkeit: Seien $X = M + A = M' + A'$ zwei Zerlegungen mit den genannten Eigenschaften. Dann ist $M - M' = A' - A$ ein vorhersagbares Martingal. Somit gilt

$$M_n - M_n' = \mathbb{E}[M_{n+1} - M_{n+1}'|\mathcal{F}_n] = M_{n+1} - M_{n+1}^{'} \quad P\text{-f.s.}$$

Aus $M_0 - M_0' = 0$ folgt also $M_n - M_n' = 0$ und somit auch $A_n = A_n'$ für alle $n \geqslant 0$. □

B.2 Optional Stopping und Optional Sampling

Es sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum mit einer Filtration $(\mathcal{F}_n)_{n\geqslant 0}$.

Definition B.8. Eine Zufallsvariable τ mit Werten $\mathbb{N} \cup \{0, \infty\}$ heißt *Stoppzeit (bzgl. $(\mathcal{F}_n)_{n\geqslant 0}$), falls für jedes $n \geqslant 0$ gilt, dass $\{\tau \leqslant n\} \in \mathcal{F}_n$.*

Bemerkung B.9. τ ist genau dann eine Stoppzeit, wenn $\{\tau = n\} \in \mathcal{F}_n$. Dies folgt aus den Gleichheiten $\{\tau = n\} = \{\tau \leqslant n\} \setminus \{\tau \leqslant n-1\}$ und $\{\tau \leqslant n\} = \bigcup_{k=0}^{n}\{\tau = k\}$ für jedes $n \geqslant 0$.

Beispiel B.10.

(i) Jede deterministische Abbildung $\tau(\omega) = n_0$ für ein festes $n_0 \in \mathbb{N} \cup \{0, \infty\}$ ist eine Stoppzeit.

(ii) Sei $(X_n)_{n\geqslant 0}$ ein adaptierter Prozess und $B \in \mathscr{B}_{\mathbb{R}}$ eine Borelmenge. Dann ist die *Eintrittszeit* von $(X_n)_{n\geqslant 0}$ in B

$$\tau_B := \inf\{n \geqslant 0 : X_n \in B\} \qquad (\inf \emptyset := \infty)$$

eine Stoppzeit, denn

$$\{\tau_B \leqslant n\} = \bigcup_{k=0}^{n} \{X_k \in B\} \in \mathcal{F}_n, \qquad \text{für alle } n \geqslant 0.$$

(iii) Sind σ und τ Stoppzeiten, dann sind auch $\sigma \wedge \tau := \min(\sigma, \tau), \sigma \vee \tau := \max(\sigma, \tau)$ und $(\sigma + \tau)$ Stoppzeiten, da

$$\begin{aligned} \{\sigma \wedge \tau \leqslant n\} &= \{\sigma \leqslant n\} \cup \{\tau \leqslant n\} \in \mathcal{F}_n, \\ \{\sigma \vee \tau \leqslant n\} &= \{\sigma \leqslant n\} \cap \{\tau \leqslant n\} \in \mathcal{F}_n, \\ \{\sigma + \tau \leqslant n\} &= \bigcup_{k=0}^{n} \{\sigma \leqslant k\} \cap \{\tau \leqslant n-k\} \in \mathcal{F}_n. \end{aligned}$$

Satz B.11 (Optional Stopping). *Ist $X = (X_n)_{n \geqslant 0}$ ein (Sub-, Super-)Martingal und τ eine Stoppzeit (beide bzgl. $(\mathcal{F}_n)_{n \geqslant 0}$), dann ist $X_n^\tau := X_{\tau \wedge n}, n \geqslant 0$, wieder ein (Sub-, Super-)Martingal.*

Beweis. Sei X zunächst ein Submartingal. Für jedes $n \geqslant 0$ ist $X_n^\tau \in L^1(P)$, da

$$\mathbb{E}[|X_{\tau \wedge n}|] \leqslant \sum_{k=1}^{n} \mathbb{E}[|X_k|] < \infty.$$

Setzen wir nun $Y_n^\tau := \mathbb{1}_{\{n \leqslant \tau\}}$, ist $(Y_n^\tau)_{n \geqslant 1}$ vorhersagbar, da $\{n \leqslant \tau\} = \{\tau \leqslant n-1\}^c \in \mathcal{F}_{n-1}$. Es gilt

$$(Y^\tau \cdot X)_n = \sum_{k=1}^{n} Y_k^\tau (X_k - X_{k-1}) = \sum_{k=1}^{\tau \wedge n} (X_k - X_{k-1}) = X_{\tau \wedge n} - X_0$$

und

$$\mathbb{E}\big[(Y^\tau \cdot X)_{n+1} - (Y^\tau \cdot X)_n \big| \mathcal{F}_n\big] = \mathbb{E}\big[Y_{n+1}^\tau (X_{n+1} - X_n) | \mathcal{F}_n\big] = \underbrace{Y_{n+1}^\tau}_{\in \{0,1\}} \underbrace{\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n]}_{\geqslant 0} \geqslant 0.$$

Damit ist $X_n^\tau = (Y^\tau \cdot X)_n + X_0$ ein Submartingal.

Ist X ein Supermartingal, betrachte das Submartingal $(-X_n)_{n \geqslant 0}$. Ist X ein Martingal, folgt die Behauptung, da X dann ein Sub- und ein Supermartingal ist. □

Definition B.12. Für eine Stoppzeit τ ist die *σ-Algebra der τ-Vergangenheit* gegeben durch

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leqslant n\} \in \mathcal{F}_n \text{ für alle } n \geqslant 0\}.$$

Bemerkung B.13.

(i) Durch Nachprüfen der Eigenschaften von σ-Algebren sieht man leicht, dass $\mathcal{F}_\tau$ tatsächlich eine σ-Algebra ist.

(ii) τ ist $\mathcal{F}_\tau$-messbar, da für jedes $n \in \mathbb{N} \cup \{0, \infty\}$ und jedes $m \geqslant 0$ gilt

$$\{\tau = n\} \cap \{\tau \leqslant m\} = \begin{cases} \{\tau = n\} \in \mathcal{F}_n \subseteq \mathcal{F}_m, & m \geqslant n, \\ \emptyset \in \mathcal{F}_m, & m < n. \end{cases}$$

(iii) Für Stoppzeiten σ, τ mit $\sigma \leqslant \tau$ gilt $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$: Ist $A \in \mathcal{F}_\sigma$ folgt für alle $n \geqslant 0$ wegen $\sigma \leqslant \tau$

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

Lemma B.14. *Ist $(X_n)_{n \geqslant 0}$ ein adaptierter Prozess und τ eine endliche Stoppzeit, dann ist $X_\tau : \Omega \ni \omega \mapsto X_{\tau(\omega)}(\omega)$ $\mathcal{F}_\tau$-messbar.*

Beweis. Für $B \in \mathscr{B}_\mathbb{R}$ ist $\{X_\tau \in B\} \in \mathcal{F}_\tau$, da für alle $n \geqslant 0$ gilt

$$\begin{aligned} \{X_\tau \in B\} \cap \{\tau \leqslant n\} &= \bigcup_{k=0}^{n} \{X_\tau \in B\} \cap \{\tau = k\} \\ &= \bigcup_{k=0}^{n} \underbrace{\{X_k \in B\} \cap \{\tau = k\}}_{\in \mathcal{F}_k} \in \mathcal{F}_n. \end{aligned}$$

□

Satz B.15 (Optional Sampling). *Es seien $X = (X_n)_{n \geqslant 0}$ ein Martingal (bzw. Submartingal) und σ, τ Stoppzeiten mit $\sigma \leqslant \tau \leqslant C$ für eine Konstante $C > 0$. Dann gilt $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = X_\sigma$ (bzw. $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \geqslant X_\sigma$).*

Beweis. Schritt 1: Sei X ein Martingal. Da $\tau \leqslant C$ ist $X_\tau \in L^1(P)$ und in obigem Lemma haben wir gezeigt, dass X_σ $\mathcal{F}_\sigma$-messbar ist. Es genügt also zu zeigen, dass für jedes $A \in \mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$

$$\mathbb{E}[X_\tau \mathbb{1}_A] = \mathbb{E}[X_\sigma \mathbb{1}_A]$$

gilt. Die Abbildung $\rho := \sigma \mathbb{1}_A + \tau \mathbb{1}_{A^c}$ ist ebenfalls eine Stoppzeit, da

$$\{\rho \leqslant n\} = (A \cap \{\sigma \leqslant n\}) \cup (A^c \cap \{\tau \leqslant n\}) \in \mathcal{F}_n.$$

Der Optional-Stopping-Satz impliziert

$$\mathbb{E}[X_\rho] = \mathbb{E}[X_{\rho \wedge C}] = \mathbb{E}[X_0] = \mathbb{E}[X_\tau]$$

und somit $0 = \mathbb{E}[X_\rho - X_\tau] = \mathbb{E}[(X_\sigma - X_\tau)\mathbb{1}_A]$.

Schritt 2: Ist X ein Submartingal mit Doob-Zerlegung $X = A + M$ für einen monoton wachsenden Prozess A, so folgt aus der Monotonie, $\tau \geqslant \sigma$ und Schritt 1, dass

$$\begin{aligned}\mathbb{E}[X_\tau|\mathcal{F}_\sigma] &= \mathbb{E}[A_\tau|\mathcal{F}_\sigma] + \mathbb{E}[M_\tau|\mathcal{F}_\sigma] \\ &\geqslant \mathbb{E}[A_\sigma|\mathcal{F}_\sigma] + M_\sigma = A_\sigma + M_\sigma = X_\sigma.\end{aligned}$$

□

B.3 Ungleichungen und Konvergenzsätze

Wir erinnern uns an die Markov-Ungleichung: Für eine Zufallsvariable $M \in L^1(\mathbb{P})$ und $\alpha > 0$ gilt

$$\mathbb{P}(|M| \geqslant \alpha) \leqslant \frac{\mathbb{E}[|M|]}{\alpha}.$$

Wir betrachten nun das laufende Maximum

$$|M|_n^* := \max_{k=0,\dots,n} |M_k| \quad \text{für} \quad n \geqslant 1.$$

Die einfache Anwendung von Markovs Ungleichung liefert lediglich

$$\mathbb{P}\big(|M|_n^* \geqslant \alpha\big) \leqslant \alpha^{-1}\mathbb{E}[\max_{0 \leqslant k \leqslant n} |M_k|] \leqslant \alpha^{-1} \sum_{k=0}^n \mathbb{E}[|M_k|].$$

Die Martingaleigenschaft von M erlaubt eine deutlich bessere Abschätzung.

Satz B.16 (Doobs Maximalungleichung)**.** *Sei $M = (M_k, k \in \mathbb{N}_0)$ ein Martingal, $\alpha > 0$ und $n \in \mathbb{N}$. Dann gilt*

$$\mathbb{P}\big(|M|_n^* \geqslant \alpha\big) \leqslant \frac{1}{\alpha}\mathbb{E}\big[|M_n|\mathbb{1}_{\{|M|_n^* \geqslant \alpha\}}\big] \leqslant \frac{1}{\alpha}\mathbb{E}[|M_n|].$$

Beweis. Wir definieren $\tau := \inf\{n \geqslant 0 : |M_n| \geqslant \alpha\}$. Dann ist τ eine Stoppzeit und es gilt $\{|M|_n^* \geqslant \alpha\} = \{\tau \leqslant n\}$. Da $\mathbb{R} \ni x \mapsto |x|$ konvex ist, ist $(|M_k|, k \in \mathbb{N}_0)$ ein Submartingal (Jensens Ungleichung) und somit folgt aus dem Optional-Sampling-Satz

$$\begin{aligned}\mathbb{E}[|M_n|] \geqslant \mathbb{E}[|M_{\tau \wedge n}|] &= \mathbb{E}[\mathbb{1}_{\{\tau \leqslant n\}}|M_{\tau \wedge n}|] + \mathbb{E}[\mathbb{1}_{\{\tau > n\}}|M_{\tau \wedge n}|] \\ &\geqslant \alpha\mathbb{P}(|M|_n^* \geqslant \alpha) + \mathbb{E}[\mathbb{1}_{\{\tau > n\}}|M_n|].\end{aligned}$$

Also gilt $\alpha\mathbb{P}(|M|_n^* \geqslant \alpha) \leqslant \mathbb{E}[\mathbb{1}_{\{\tau \leqslant n\}}|M_n|]$. □

Satz B.17 (Doobs L^p-Martingalungleichung)**.** *Sei $M = (M_k, k \in \mathbb{N}_0)$ ein Martingal mit $M_k \in L^p(\mathbb{P})$ für alle $k \in \mathbb{N}_0$ und ein $p > 1$. Dann gilt*

$$\mathbb{E}\big[(|M|_n^*)^p\big] \leqslant \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|M_n|^p].$$

Beweis. Für jedes $K > 0$ folgt aus Doobs Maximalungleichung und dem Satz von Fubini

$$\begin{aligned}\mathbb{E}\big[(|M|_n^* \wedge K)^p\big] &= \mathbb{E}\Big[\int_0^{|M|_n^*\wedge K} p\alpha^{p-1}\mathrm{d}\alpha\Big]\\ &= \mathbb{E}\Big[\int_0^{K} p\alpha^{p-1}\mathbb{1}_{\{|M|_n^*\geqslant\alpha\}}\mathrm{d}\alpha\Big]\\ &= \int_0^K p\alpha^{p-1}\mathbb{P}(|M|_n^*\geqslant\alpha)\mathrm{d}\alpha\\ &\leqslant \int_0^K p\alpha^{p-2}\mathbb{E}\big[|M_n|\mathbb{1}_{\{|M|_n^*\geqslant\alpha\}}\big]\mathrm{d}\alpha\\ &= p\mathbb{E}\Big[|M_n|\int_0^{|M|_n^*\wedge K}\alpha^{p-2}\mathrm{d}\alpha\Big] = \frac{p}{p-1}\mathbb{E}\big[|M_n|\cdot(|M|_n^*\wedge K)^{p-1}\big].\end{aligned}$$

Hölders Ungleichung liefert nun

$$\mathbb{E}\big[(|M|_n^*\wedge K)^p\big] \leqslant \frac{p}{p-1}\mathbb{E}\big[|M_n|\cdot(|M|_n^*\wedge K)^{p-1}\big] \leqslant \frac{p}{p-1}\mathbb{E}\big[(|M|_n^*\wedge K)^p\big]^{(p-1)/p}\mathbb{E}\big[|M_n|^p\big]^{1/p}.$$

Da $\mathbb{E}[(|M|_n^*\wedge K)^p] < \infty$ folgt

$$\mathbb{E}\big[(|M|_n^*\wedge K)^p\big]^{1/p} \leqslant \frac{p}{p-1}\mathbb{E}\big[|M_n|^p\big]^{1/p}.$$

Schließlich können wir mit monotoner Konvergenz $K\to\infty$ betrachten. □

Wir wollen nun untersuchen, unter welchen Bedingungen ein Martingal $(M_n)_{n\geqslant 0}$ auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},(\mathcal{F}_n)_{n\geqslant 0},\mathbb{P})$ für $n\to\infty$ gegen einen Grenzwert konvergiert. Als Hilfsmittel benötigen noch eine weitere Ungleichung.

Definition B.18. Seien $(M_n)_{n\geqslant 0}$ Zufallsvariablen und $a<b$ reelle Zahlen. Die *Anzahl der Aufkreuzungen* (aufsteigende Überquerungen) von $[a,b]$ durch (M_n) bis zur Zeit n ist gegeben durch

$$U_n^{[a,b]} := \sup\{k\geqslant 1 | \tau_k \leqslant n\} \qquad (\text{mit } \sup\emptyset := 0),$$

wobei

$$\begin{aligned}\tau_0 &:= 0,\\ \sigma_{k+1} &:= \inf\{l\geqslant\tau_k : M_l\leqslant a\},\\ \tau_{k+1} &:= \inf\{l\geqslant\sigma_{k+1} : M_l\geqslant b\},\quad k\geqslant 0.\end{aligned}$$

Lemma B.19 (Doobs Aufkreuzungsungleichung)**.** *Für ein Submartingal* $(M_n, n\geqslant 0)$ *und reelle Zahlen* $a<b$ *gilt*

$$\mathbb{E}[U_n^{[a,b]}] \leqslant \frac{1}{b-a}\mathbb{E}\big[(M_n-a)_+\big],$$

wobei $A_+ := \max\{A,0\}$.

Beweis. Es gilt $(M_n-a)_+\in L^1(\mathbb{P})$. $X_n := (M_n-a)_+$ ist ein Submartingal, da $x\mapsto(x-a)_+$ konvex ist (Jensens Ungleichung). Da die Anzahl der Aufkreuzungen von $[0,b-a]$ durch X_n gleich $U_n^{[a,b]}$ sind, können wir o.B.d.A $a=0$ und $M_n\geqslant 0$ setzen. Mit der Konvention $\infty\wedge n := n$ gilt

$$\mathbb{E}[M_n] = \mathbb{E}[M_{\sigma_1\wedge n}] + \sum_{k=1}^n \mathbb{E}[M_{\tau_k\wedge n} - M_{\sigma_k\wedge n}] + \sum_{k=1}^n \mathbb{E}[M_{\sigma_{k+1}\wedge n} - M_{\tau_k\wedge n}].$$

Da (M_n) ein nichtnegatives Submartingal ist, impliziert der Optional-Sampling-Satz, dass alle Summanden nichtnegativ sind. Weiterhin gilt

$$\sum_{k=1}^n (M_{\tau_k\wedge n} - M_{\sigma_k\wedge n}) \geqslant bU_n^{[0,b]},$$

woraus $\mathbb{E}[M_n]\geqslant b\mathbb{E}[U_n^{[0,b]}]$ folgt. □

Satz B.20 (1. Martingalkonvergenzsatz)**.** *Sei* $(M_n, n \geqslant 0)$ *ein Sub- oder ein Supermartingal mit* $\sup_{n\geqslant 0} \mathbb{E}[|M_n|] < \infty$ *und definiere* $\mathcal{F}_\infty := \sigma(\mathcal{F}_n : n \geqslant 0)$. *Dann existiert der* $\mathcal{F}_\infty$*-messbare Grenzwert* $M_\infty := \lim_{n\to\infty} M_n$ *f.s. und es gilt* $M_\infty \in L^1(\mathbb{P})$.

Beweis. Nehmen wir an, (M_n) ist ein Submartingal (sonst betrachte $(-M_n)_n$). Da $(U_n^{[a,b]})_n$ eine monoton wachsende Folge ist, liefert der Satz über monotone Konvergenz für $U^{[a,b]} := \lim_{n\to\infty} U_n^{[a,b]}$ zusammen mit Doobs Aufkreuzungsungleichung:

$$\begin{aligned}\mathbb{E}[U^{[a,b]}] = \lim_{n\to\infty} \mathbb{E}[U_n^{[a,b]}] &\leqslant \sup_n \frac{1}{b-a}\mathbb{E}[(M_n-a)_+]\\ &\leqslant \frac{1}{b-a}\big(\sup_n \mathbb{E}[(M_n)_+] + |a|\big) < \infty.\end{aligned}$$

Somit gilt $\mathbb{P}(U^{[a,b]} < \infty) = 1$, d.h. (M_n) hat f.s. nur endlich viele Aufkreuzungen jedes beliebigen Intervalls $[a,b]$. Wir setzen nun

$$\Lambda_{a,b} := \big\{\omega \in \Omega : \limsup_{n\to\infty} M_n(\omega) \geqslant b, \liminf_{n\to\infty} M_n(\omega) \leqslant a\big\}.$$

Dann gilt $\mathbb{P}(\Lambda_{a,b}) = 0$ für alle $a < b$ und folglich $\mathbb{P}(\bigcup_{a,b\in\mathbb{Q}:a<b} \Lambda_{a,b}) = 0$. Wir erhalten also $\mathbb{P}\big(\limsup_{n\to\infty} M_n > \liminf_{n\to\infty} M_n\big) = 0$, d.h. (M_n) konvergiert f.s. und da alle M_n $\mathcal{F}_\infty$-messbar sind, ist auch M_∞ $\mathcal{F}_\infty$-messbar (modulo einer evtl. nötigen Abänderung auf einer Nullmenge). Schließlich liefert das Lemma von Fatou

$$\mathbb{E}[\lim_{n\to\infty} |M_n|] \leqslant \liminf_{n\to\infty} \mathbb{E}[|M_n|] \leqslant \sup_n \mathbb{E}[|M_n|] < \infty,$$

sodass $M_\infty := \lim_{n\to\infty} M_n \in L^1(\mathbb{P})$ (insbesondere ist M_∞ f.s. endlich). □

Bemerkung B.21. Die Konvergenz gilt im Allgemeinen nicht in $L^1(\mathbb{P})$. Ein Gegenbeispiel ist das Petersburger Spiel (Übung ☑).

Korollar B.22. *Jedes nichtnegative Supermartingal konvergiert f.s..*

Beweis. Es sei $(M_n, n \geqslant 0)$ ein nichtnegatives Supermartingal. $N_n := (-M_n)$ ist ein Submartingal mit $\mathbb{E}[(N_n)_+] = \mathbb{E}[0] = 0$. Weiterhin gilt $\mathbb{E}[(N_n)_-] = \mathbb{E}[(N_n)_+] - \mathbb{E}[N_n] \leqslant \mathbb{E}[(N_n)_+] - \mathbb{E}[N_0]$ und daher

$$\begin{aligned}\sup_{n\geqslant 0} \mathbb{E}[|M_n|] = \sup_{n\geqslant 0} \mathbb{E}[|N_n|] &\leqslant \sup_{n\geqslant 0} \mathbb{E}[(N_n)_+] + \sup_{n\geqslant 0} \mathbb{E}[(N_n)_-]\\ &\leqslant 2\sup_{n\geqslant 0} \mathbb{E}[(N_n)_+] + \mathbb{E}[-N_0] < \infty.\end{aligned}$$ □

Um das erste Martingalkonvergenzresultat auf L^1-Konvergenz zu verstärken, benötigen wir gleichgradige Integrierbarkeit.

Satz B.23 (2. Martingalkonvergenzsatz)**.** *Setze* $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \geqslant 0)$.

(i) Ist $(M_n, n \geqslant 0)$ *ein gleichgradig integrierbares Martingal, dann konvergiert* (M_n) *f.s. und in* L^1 *gegen ein* $\mathcal{F}_\infty$*-messbares* $M_\infty \in L^1(\mathbb{P})$ *und es gilt* $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$.

(ii) Ist $(M_n, n \geqslant 0)$ *ein Martingal mit* $M_n = \mathbb{E}[M|\mathcal{F}_n]$ *für ein* $M \in L^1(\mathbb{P})$, *dann ist* $(M_n)_{n\geqslant 0}$ *gleichgradig integrierbar und (i) gilt mit* $M_\infty = \mathbb{E}[M|\mathcal{F}_\infty]$.

Beweis. (i) Der erste Martingalkonvergenzsatz liefert zusammen mit Satz A.5 die Konvergenz in L^1. Weiterhin gilt für jedes $n \geqslant m \geqslant 0$ und alle $A \in \mathcal{F}_m$, dass $\mathbb{E}[M_m \mathbb{1}_A] = \mathbb{E}[M_n \mathbb{1}_A]$. Die Konvergenz $\|M_n - M_\infty\|_{L^1} \to \infty$ impliziert $\mathbb{E}[|M_n - M_\infty|\mathbb{1}_A] \to 0$. Zusammen erhalten wir $\mathbb{E}[M_m \mathbb{1}_A] = \mathbb{E}[M_\infty \mathbb{1}_A]$ also $M_m = \mathbb{E}[M_\infty|\mathcal{F}_m]$.

(ii) Wir bemerken zuerst, dass $\mathbb{E}[|M_n|] = \mathbb{E}[|\mathbb{E}[M|\mathcal{F}_n]|] \leqslant \mathbb{E}[|M|]$ nach Jensens Ungleichung. Für $R > 0$ ist $\mathbb{1}_{\{|M_n|>R\}}$ $\mathcal{F}_n$-messbar und damit

$$\begin{aligned}\mathbb{E}\big[|M_n|\mathbb{1}_{\{|M_n|>R\}}\big] &\leqslant \mathbb{E}\big[\mathbb{E}[|M|\,|\mathcal{F}_n]\mathbb{1}_{\{|M_n|>R\}}\big]\\ &= \mathbb{E}\big[|M|\mathbb{1}_{\{|M_n|>R\}}\big]\\ &\leqslant R^{1/2}\mathbb{P}(|M_n| > R) + \mathbb{E}\big[|M|\mathbb{1}_{\{|M|>R^{1/2}\}}\big].\end{aligned}$$

Der zweite Term konvergiert für $R \to \infty$ gegen 0 (dominierte Konvergenz). Für den ersten Term liefert die Markov-Ungleichung

$$R^{1/2}\mathbb{P}(|M_n| > R) \leqslant \frac{\mathbb{E}[|M_n|]}{R^{1/2}} \leqslant \frac{\mathbb{E}[|M|]}{R^{1/2}} \longrightarrow 0 \quad \text{für } R \to \infty.$$

Damit ist die gleichgradige Integrierbarkeit gezeigt und wir können (i) anwenden.

Es existiert also der f.s. und L^1-Grenzwert M_∞ mit $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n], n \geqslant 1$. Daraus folgt für alle $n \in \mathbb{N}$ und $B \in \mathcal{F}_n$, dass

$$\mathbb{E}[M_\infty \mathbb{1}_B] = \mathbb{E}[M_n \mathbb{1}_B] = \mathbb{E}[M\mathbb{1}_B].$$

Also stimmen die endlichen (signierten) Maße $\mu_1(B) := \mathbb{E}[M_\infty\mathbb{1}_B]$ und $\mu_2(B) := \mathbb{E}[M\mathbb{1}_B]$ auf $\bigcup_{n\geqslant 1}\mathcal{F}_n$ überein. Letzteres ist ein $\cap$-stabiler Erzeuger von $\mathcal{F}_\infty$, sodass μ_1 und μ_2 auf $\mathcal{F}_\infty$ identisch sind. Daraus folgt $M_\infty = \mathbb{E}[M|\mathcal{F}_\infty]$. □

B.4 Anwendungen

Satz B.24 (0-1-Gesetz von Kolmogorov)**.** *Für eine Folge $(X_n)_{n\in\mathbb{N}}$ von unabhängigen Zufallsvariablen heißt*

$$\mathcal{T} := \bigcap_{n\geqslant 1} \sigma\big(X_m : m \geqslant n\big)$$

die terminale σ-Algebra. *Für alle $A \in \mathcal{T}$ gilt $\mathbb{P}(A) \in \{0, 1\}$.*

Die terminale σ-Algebra enthält alle Ereignisse, deren Eintreten von jeder festen endlichen Teilfamilie der X_i nicht abhängt.

Beweis. Wir definieren die Filtration $\mathcal{F}_n := \sigma(X_m : m \leqslant n)$. Für $A \in \mathcal{T}$ betrachten wir das Martingal $M_n := \mathbb{E}[\mathbb{1}_A|\mathcal{F}_n], n \geqslant 1$. Aus dem 2. Martingalkonvergenzsatz folgt, dass M_n f.s. gegen $M_\infty = \mathbb{E}[\mathbb{1}_A|\mathcal{F}_\infty]$ konvergiert. Nun ist das Ereignis $A \in \mathcal{T} \subseteq \mathcal{F}_\infty$ unabhängig von $\mathcal{F}_n$ für alle $n \geqslant 1$, woraus folgt

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A|\mathcal{F}_n] \xrightarrow{f.s.} M_\infty = \mathbb{E}[\mathbb{1}_A|\mathcal{F}_\infty] = \mathbb{1}_A.$$

Damit haben wir $\mathbb{P}(A) = \mathbb{1}_A \in \{0, 1\}$ gezeigt. □

Die Konzepte der Filtration und des Martingals erfodern nicht, dass die Indexmenge aus $\mathbb{R}_+$ stammt. Wir können daher auch negative Indexmengen zulassen.

Definition B.25. Ein Prozess $(M_{-n})_{n\geqslant 0}$ heißt *Rückwärtsmartingal* bzgl. $(\mathcal{F}_{-n})_{n\geqslant 0}$ (σ-Algebren mit $\mathcal{F}_{-n} \subseteq \mathcal{F}_{-n+1}$), falls $M_{-n} \in L^1(\mathbb{P})$, M_{-n} $\mathcal{F}_{-n}$-messbar ist und $\mathbb{E}[M_{-n+1}|\mathcal{F}_{-n}] = M_{-n}$ für alle $n \geqslant 0$ gilt.

Satz B.26 (Rückwärtsmartingalkonvergenzsatz)**.** *Jedes Rückwärtsmartingal $(M_{-n})_{n\geqslant 0}$ konvergiert f.s. und in $L^1(\mathbb{P})$ für $n \to \infty$ gegen ein $M_{-\infty} \in L^1(\mathbb{P})$.*

Beweis. Wir bezeichnen mit $U_{-n}^{[a,b]}$ die Anzahl der Aufkreuzungen von $[a, b]$ durch $(M_{-n})_{n\geqslant 0}$ in der Zeit $[-n, 0]$. Dann gilt nach Lemma B.19 und $U^{[a,b]} := \lim_{n\to\infty} U_{-n}^{[a,b]}$

$$\mathbb{E}[U^{[a,b]}] = \lim_{n\to\infty} \mathbb{E}[U_{-n}^{[a,b]}] \leqslant \frac{1}{b-a}\mathbb{E}[(M_0 - a)_+] < \infty$$

wie im 1. Martingalkonvergenzsatz. Damit folgt $M_{-n} \to M_{-\infty}$ f.s.. Da $M_{-n} = \mathbb{E}[M_0|\mathcal{F}_{-n}]$ ist $(M_{-n})_{n\geqslant 0}$ nach dem 2. Martingalkonvergenzsatz gleichgradig integrierbar, sodass auch $M_{-n} \to M_{-\infty}$ in $L^1(\mathbb{P})$ folgt. □

Korollar B.27 (Starkes Gesetz der großen Zahlen)**.** *Es seien $(X_k)_{k\geqslant 1}$ reelle u.i.v. Zufallsvariablen in $L^1(\mathbb{P})$. Dann gilt*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X_1] \quad \textit{f.s.}$$

Beweis. Wir setzen $S_n := \sum_{k=1}^{n} X_k, S_0 = 0$ sowie $\mathcal{F}_{-n} := \sigma(S_k, k \geqslant n)$. Dann gilt $\mathcal{F}_{-n-1} \subseteq \mathcal{F}_{-n}$ und aus Symmetriegründen gilt

$$\mathbb{E}[X_j|\mathcal{F}_{-n}] = \mathbb{E}[X_1|\mathcal{F}_{-n}] \quad \text{für alle } j = 1, \ldots, n.$$

Da $\mathbb{E}[S_n|\mathcal{F}_{-n}] = S_n$, erhalten wir

$$M_{-n} := \mathbb{E}[X_1|\mathcal{F}_{-n}] = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[X_j|\mathcal{F}_{-n}] = \frac{S_n}{n}$$

und $(M_{-n})_{n\geqslant 0}$ ist ein Rückwärtsmartingal. Damit folgt $M_{-n} \to M_{-\infty}$ für $n \to \infty$ f.s. und in $L^1(\mathbb{P})$ sowie

$$\mathbb{E}[M_{-\infty}] = \lim_{n\to\infty} \mathbb{E}[M_{-n}] = \mathbb{E}[X_1].$$

Da für jedes $A \in \mathscr{B}_{\mathbb{R}}$ das Ereignis $\{M_{-\infty} \in A\}$ in der terminalen σ-Algebra der $(X_k)_{k\geqslant 1}$ folgt aus Kolmogorovs 0-1-Gesetz $\mathbb{P}(M_{-\infty} \in A) \in \{0, 1\}$. Also muss $M_{-\infty}$ konstant sein und es gilt insbesondere $M_{-\infty} \overset{f.s.}{=} \mathbb{E}[M_{-\infty}] = \mathbb{E}[X_1]$. □

Literaturempfehlung

Bauer, H. (1992). *Maß- und Integrationstheorie.* de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, second edition.

Bauer, H. (2002). *Wahrscheinlichkeitstheorie.* de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, fifth edition.

Georgii, H.-O. (2007). *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik.* de Gruyter, Berlin.

Jacod, J. and Protter, P. (2003). *Probability essentials.* Universitext. Springer-Verlag, Berlin, second edition.

Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition.

Klenke, A. (2006). *Wahrscheinlichkeitstheorie.* Springer.

Liggett, T. M. (2010). *Continuous time Markov processes*, volume 113 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI. An introduction.

Stroock, D. W. (2014). *An introduction to Markov processes*, volume 230 of *Graduate Texts in Mathematics.* Springer, Heidelberg, second edition.

Werner, D. (2000). *Funktionalanalysis.* Springer-Verlag, Berlin, extended edition.