Wahrscheinlichkeitstheorie,

Stochastische Prozesse,

Stochastische Integration und Finanzmathematik

von Peter Pfaffelhuber

Version: 26. November 2014

WARNUNG: Dieses Skript enthält noch viele Fehler. Es wurde teilweise in großer Eile geschrieben. Für alle Fehler bin ich selbst verantwortlich. Ich hoffe, in Zukunft eine Version mit weniger Fehlern bereit stellen zu können.

Vorbemerkung

Dieses Manuskript ist parallel zu Vorlesungen entstanden, die ich im Wintersemester 2010 (Wahrscheinlichkeitstheorie), Wintersemster 2011 (Stochastische Prozesse) und Sommersemester 2012 (Stochastische Integration und Finanzmathematik) an der Universität Freiburg gehalten habe. Es stellt eine Grundlage der modernen Stochastik dar.

Die Stochastik ist ein Gebiet der Mathematik, das sich bis heute ständig weiterentwickelt. Sie ist weder ein Gebiet der reinen noch der angewandten Mathematik, sondern beides. Auf der einen Seite stehen oftmals intuitive Ideen im Vordergrund, deren Formalisierung entscheidende Fortschritte bringt. Auf der anderen Seite führt der so entwickelte Formalismus ein Eigenleben, und ist unerlässlich im formalen Umgang mit der Stochastik.

Die moderne Wahrscheinlichkeitstheorie basiert auf der *Maßtheorie*, die wir zu Beginn des Skriptes wiederholen. Der Teil über *Wahrscheinlichkeitstheorie* umfasst die verschiedenen Konvergenzarten (fast sicher, stochastik, schwach, L^p), die in der Stochastik vorkommen. Weiter wird der Begriff der bedingten Erwartung eingeführt, der im Teil über *Stochastische Prozesse* die Behandlung von Martingalen einläutet. Zentral in diesem Abschnitt sind der Poisson-Prozess sowie die Brown'sche Bewegung, Markov-Prozesse und stationäre Prozesse. Im Teil über *Stochastische Integration* entwickeln wir as stochastische Integral bezüglich stetiger Semimartingale, und erhalten damit grundlegende Resultate über stochastische Differentialgleichungen. Die einzige konsequent behandelte Anwendung des erarbeiteten Stoffes stellt die *Finanzmathematik* dar. Neben einer Einführung im ersten Teil, in der wir ein zeitdiskretes Finanzmarkt-Modell behandeln, erfolgt im letzten Teil eine behandlung zeit-stetiger Finanzmärkte.

Es gibt viele Lehrbücher der Stochastik. Besonders erwähnen möchte ich hier allerdings jedoch nur ein paar, die mir beim Schreiben des Skriptes stets wertvolle Dienste erwiesen:

- Kallenberg, Olaf. Foundations of Modern Probability Theory. Springer, Second edition, 2010
- Klenke, Achim. Wahrscheinlichkeitstheorie, Springer, 2. Auflage, 2008
- Lamberton, Damien; Lapeyre, Bernard. Introduction to Stochastic Calculus Applied to Finance. Chapman and Hall/CRC Financial Mathematics. 2. Auflage, 2007
- Øksendal, Bernt. Stochastic Differential Equations: An Introduction with Applications. Springer, 6. Auflage, 2003
- Rogers, L.C.D.; Williams, David. Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus. Cambridge Mathematical Library, 2. Auflage, 2000

Inhaltsverzeichnis

Teil I
Maßtheorie

Die Maßtheorie war bereits Teil der Vorlesung *Analysis 3*. Es mag zunächst verwundern, dass ein Gebiet, in dem es um Messen von Teilmengen eines Grundraumes geht, etwas mit Stochastik zu tun hat. Insbesondere sind die in der Analysis betrachteten Grundräume meist Teilmengen von reellen Zahlen, Zufallsvariablen können jedoch Werte in ganz anderen Mengen annehmen. (Man denke etwa an die Menge $\{\text{Kopf}, \text{Zahl}\}$, die beim Münzwurf auftritt.) Wie sich jStochaedoch herausstellt, gelten für Flächenmessungen ähnliche Rechenregeln wie für Wahrscheinlichkeiten. Klar ist beispielsweise, dass man Flächen zweier disjunkter Mengen addieren muss, wenn man deren gemeinsamen Inhalt berechnen will. Genauso muss man die Wahrscheinlichkeit für zwei disjunkte Ereignisse addieren, wenn man die Wahrscheinlichkeit ausrechnen will, dass eines der beiden Ereignisse eintritt.

In diesem Kapitel bezeichnet $\Omega \neq \emptyset$ (irgend)eine Menge und alle aufgeführten Mengensysteme[1] sind Teilmengen der Potenzmenge von Ω, die wir mit 2^Ω bezeichnen. Alle betrachteten Mengensysteme seien o.E. nicht leer. Zentrale Begriffe werden *σ-Algebra* und *(Wahrscheinlichkeits-)Maß* sein.

1 Wiederholung Topologie

Topologien werden in der Mathematik immer dann gebraucht, wenn ein Konvergenzbegriff eingeführt wird. Auch wenn Topologien in den bisherigen Vorlesungen nur am Rande behandelt wurden, sind doch einige Konvergenzbegriffe bekannt. Auch zwischen Maßtheorie und Topologie gibt es viele Verbindungen. Insbesondere werden wir Maße auf topologischen Räumen betrachten. Dabei werden die zugrunde liegenden σ-Algebren von Topologie erzeugt (siehe Definition 2.7). Deshalb wiederholen wir zunächst Grundbegriffe der Topologie.

1.1 Grundlagen

Unter einer *Topologie* versteht man eine Familie offener Teilmengen eines Grundraumes Ω. Was offene Mengen sind, ist etwa schon aus der Analysis bekannt. In metrischen Räumen nennt man eine Menge A genau dann offen, wenn für jedes $\omega \in A$ auch ein offener Ball $B_\varepsilon(\omega) \subseteq A$ für ein $\varepsilon > 0$. Dieser Fall von metrischen Räumen ist in der Praxis auch am wichtigsten. In der Maßtheorie kommt dem Fall von separablen Topologien, die von vollständigen Metriken erzeugt werden, eine besondere Bedeutung zu. Solche Räume heißen *polnisch*.

Definition 1.1. *1. Eine Funktion $r : \Omega \times \Omega \to \mathbb{R}_+$ heißt* Metrik*, falls (i) $r(\omega, \omega') \neq 0$ für $\omega \neq \omega'$, (ii) $r(\omega, \omega') = r(\omega', \omega)$ für alle $\omega, \omega' \in \Omega$ und (iii) $r(\omega, \omega'') \leq r(\omega, \omega') + r(\omega', \omega'')$ für alle $\omega, \omega', \omega'' \in \Omega$. Das Paar (Ω, r) ist ein* metrischer Raum*.*

Für $\omega \in \Omega$ und $\varepsilon > 0$ bezeichnen wir mit $B_\varepsilon(\omega) := \{\omega' \in \Omega : r(\omega, \omega') < \varepsilon\}$ den offenen Ball *um ω mit Radius ε.*

[1] Als *Mengensystem* bezeichnet man eine Menge, deren Elemente wieder Mengen sind.

2. *Eine Metrik r auf Ω heißt* vollständig, *wenn jede Cauchy-Folge konvergiert. Ist also $\omega_1, \omega_2, \ldots \in \Omega$ mit*
$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n, m \geq N: \ \ r(\omega_n, \omega_m) < \varepsilon,$$
so gibt es ein $\omega \in \Omega$ mit $r(\omega_n, \omega) \xrightarrow{n\to\infty} 0$.

3. *Ein Mengensystem $\mathcal{O} \subseteq 2^\Omega$ heißt* Topologie, *falls (i) $\emptyset, \Omega \in \mathcal{O}$, (ii) ist $A, B \in \mathcal{O}$, so ist auch $A \cap B \in \mathcal{O}$ (iii) ist I beliebig und ist $A_i \in \mathcal{O}, i \in I$, so ist auch $\bigcup_{i\in I} A_i \in \mathcal{O}$. Das Paar $(\Omega, \mathcal{O})$ heißt* topologischer Raum. *Mengen $A \in \mathcal{O}$ heißen* offen, *Mengen $A \subseteq \Omega$ mit $A^c \in \mathcal{O}$ heißen* abgeschlossen.

4. *Ist $(\Omega, \mathcal{O})$ ein topologischer Raum und $A \subseteq \Omega$. Dann heißt*
$$A^\circ := \bigcup \{O \subseteq A : O \in \mathcal{O}\}$$
das Innere von *A und*
$$\overline{A} := \bigcap \{F \supseteq A : F^c \in \mathcal{O}\}$$
den Abschluss von A.

5. *Ein topologischer Raum $(\Omega, \mathcal{O})$ heißt* separabel, *wenn es eine abzählbare Menge $\Omega' \subseteq \Omega$ gibt mit $\overline{\Omega}' = \Omega$.*

6. *Sei $(\Omega, \mathcal{O})$ ein topologischer Raum und $\mathcal{B} \subseteq \mathcal{O}$. Dann heißt $\mathcal{B}$ eine* Basis *von $\mathcal{O}$, falls*
$$\forall A \in \mathcal{O} \ \ \forall \omega \in A \ \ \exists B \in \mathcal{B} : \omega \in B \subseteq A.$$
Dies ist genau dann der Fall, wenn
$$\mathcal{O} = \{A \subseteq \Omega : \ \ \forall \omega \in A \ \ \exists B \in \mathcal{B} : \omega \in B \subseteq A\}. \tag{1.1}$$
oder (äquivalent dazu)
$$\mathcal{O} = \Big\{ \bigcup_{i\in I} B_i : B_i \in \mathcal{B}, i \in I, \ I \text{ beliebig} \Big\}. \tag{1.2}$$

7. *Sei $\mathcal{B} \subseteq 2^\Omega$. Dann wird durch* (1.1) *oder* (1.2) *eine Topologie $\mathcal{O}(\mathcal{B})$ definiert, die* von $\mathcal{B}$ erzeugte Topologie.

8. *Sei (Ω, r) ein metrischer Raum und*
$$\mathcal{B} := \{B_\varepsilon(\omega) : \varepsilon > 0, \omega \in \Omega\}. \tag{1.3}$$
Dann ist $\mathcal{O}(\mathcal{B})$ die von r erzeugte Topologie. *Falls speziell $\Omega \subseteq \mathbb{R}^d$ und r der euklidische Abstand ist, heißt die in* (1.1) *oder* (1.2) *definierte Topologie die* euklidische Topologie.

9. *Der Raum $(\Omega, \mathcal{O})$ heißt (vollständig) metrisierbar, wenn es eine (vollständige) Metrik r auf Ω gibt, so dass* (1.1) *gilt. Der Raum $(\Omega, \mathcal{O})$ heißt* polnisch, *falls er separabel und vollständig metrisierbar ist.*

10. *Seien* $(\Omega, \mathcal{O})$ *und* $(\Omega', \mathcal{O}')$ *topologische Räume. Dann heißt eine Abbildung* $f : \Omega \to \Omega'$ stetig, *falls* $f^{-1}(A') \in \mathcal{O}$ *für alle* $A' \in \mathcal{O}'$ *gilt.*

Beispiel 1.2 (Der Raum $\overline{\mathbb{R}}$). Wir werden oftmals Funktionen mit Werten in

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\} \quad \text{oder} \quad \overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$$

betrachten.[2] Um diese Räume als topologische Räume betrachten zu können, setzen wir

$$\varphi : \begin{cases} \overline{\mathbb{R}} & \to [-1, 1], \\ x & \mapsto \begin{cases} \frac{2}{\pi} \arctan(x), & x \in \mathbb{R}, \\ 1, & x = \infty, \\ -1, & x = -\infty \end{cases} \end{cases}$$

und definieren die Metrik

$$r_{\overline{\mathbb{R}}}(x, y) := |\varphi(x) - \varphi(y)|, \qquad x, y \in \overline{\mathbb{R}}.$$

Der von $r_{\overline{\mathbb{R}}}$ definierte topologische Raum $(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ erweitert die euklidische Topologie $(\mathbb{R}, \mathcal{O})$ auf $\mathbb{R}$ insofern, als dass $\{A \cap \mathbb{R} : A \in \overline{\mathcal{O}}\} = \mathcal{O}$. Dies gilt deshalb, weil φ stetig auf $\mathbb{R}$ ist mit stetiger Umkehrfunktion. Weiter gilt, dass $(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ separabel ist und $r_{\overline{\mathbb{R}}}$ ist eine vollständige Metrik.

Auf $\overline{\mathbb{R}}$ kann man wie in der Analysis gewohnt rechnen. Etwa ist $a \cdot \infty = \infty$ für $a > 0$. Allerdings sind die Ausdrücke wie $\infty - \infty$ und ∞/∞ nicht definiert.

Bemerkung 1.3 (Zusammenhang zwischen Metrik und Topologie). Sei $(\Omega, \mathcal{O})$ ein topologischer Raum und $\omega, \omega_1, \omega_2, \ldots \in \Omega$. Wir definieren

$$\omega_n \xrightarrow{n\to\infty} \omega :\Longleftrightarrow \forall O \in \mathcal{O} : \omega \in O \Rightarrow \omega_n \in O \text{ für fast alle } n \in \mathbb{N}. \tag{1.4}$$

Insbesondere ergibt so jede Topologie auf Ω einen Konvergenzbegriff für Folgen in Ω.

Dieser Konvergenzbegriff stimmt mit dem auf metrischen Räumen bekannten Begriff überein: ist nämlich r eine Metrik auf Ω, die $\mathcal{O}$ erzeugt, dann gilt die rechte Seite aus (1.4) genau dann, wenn für alle $\varepsilon > 0$ gilt, dass $r(\omega_n, \omega) < \varepsilon$ für fast alle $n \in \mathbb{N}$.

Mit Hilfe des Konvergenzbegriffes aus (1.4) sehen wir nun folgende bekannte Eigenschaft ein:

Lemma 1.4 (Abschluss bei metrischen Räumen). *Sei* (Ω, r) *ein metrischer Raum und* $\mathcal{O}$ *die von* r *erzeugte Topologie. Für* $F \subseteq \Omega$ *sind äquivalent:*

1. F *ist abgeschlossen.*
2. *Für alle* $\omega_1, \omega_2, \ldots \in F$ *und* $\omega \in \Omega$ *mit* $\omega_n \xrightarrow{n\to\infty} \omega$ *gilt* $\omega \in F$.

Insbesondere gilt: für jedes $A \subseteq \Omega$ *besteht der Abschluss* $\overline{A}$ *genau aus den Häufungspunkten von* A.

[2] Die Schreibweise $\overline{\mathbb{R}}$ suggeriert, dass hier der Abschluss von $\mathbb{R}$ gemeint ist. Dies stimmt nicht, da die hinzugewonnenen Elemente $-\infty, \infty$ nicht in $\mathbb{R}$ liegen, Abschlüsse von Mengen aber immer höchstens die Elemente des Grundraumes enthalten können. Topologisch gesehen ist $\overline{\mathbb{R}}$ die Zwei-Punkte-Kompaktifizierung von $\mathbb{R}$.

Beweis. '1.⇒2.': Angenommen, es gäbe $\omega_1, \omega_2, \ldots \in F$, konvergent gegen $\omega \in F^c$, dann wäre, da $F^c \in \mathcal{O}$ auch $\omega_n \in F^c$ für fast alle n. Dies ist im Widerspruch zur Voraussetzung.
'2.⇒1.': Angenommen, F wäre nicht abgeschlossen, F^c also nicht offen. Dann gibt es $\omega \in F^c$, so dass für alle $\varepsilon > 0$ gilt, dass $B_\varepsilon(\omega) \not\subseteq F^c$. Wähle $\varepsilon_1, \varepsilon_2, \ldots > 0$ mit[3] $\varepsilon_n \downarrow 0$ und $\omega_n \in B_{\varepsilon_n}(\omega) \cap F$. Dann gilt $\omega_1, \omega_2, \ldots \in F$, $\omega_n \xrightarrow{n\to\infty} \omega$, aber $\omega \in F^c$. □

Lemma 1.5 (Abzählbare Basis und separable Räume). *Sei (Ω, r) ein separabler, metrischer Raum, $\mathcal{O}$ die von r erzeugte Topologie, Ω' abzählbar mit $\overline{\Omega'} = \Omega$ und*

$$\widetilde{\mathcal{B}} := \{B_\varepsilon(\omega) : \varepsilon \in \mathbb{Q}_+, \omega \in \Omega'\}.$$

Dann ist $\widetilde{\mathcal{B}}$ abzählbar und $\mathcal{O}(\widetilde{\mathcal{B}}) = \mathcal{O}$.

Beweis. Klar ist, dass $\widetilde{\mathcal{B}}$ abzählbar ist und $\mathcal{O}(\widetilde{\mathcal{B}}) \subseteq \mathcal{O}$. Sei $\mathcal{B}$ wie in (1.3). Dann ist für $B_\varepsilon(\omega) \in \mathcal{B}$

$$B_\varepsilon(\omega) = \bigcup_{\widetilde{\mathcal{B}} \ni B \subseteq B_\varepsilon(\omega)} B,$$

also $\mathcal{B} \subseteq \mathcal{O}(\widetilde{\mathcal{B}})$ und damit $\mathcal{O} = \mathcal{O}(\mathcal{B}) \subseteq \mathcal{O}(\widetilde{\mathcal{B}})$. □

Beispiel 1.6 (Zwei polnische Räume).

1. Sei $\mathcal{O}$ die euklidische Topologie auf $\mathbb{R}^d$. Diese wird nach Definition 1.1.9 durch die euklidische Metrik definiert. Bekannt ist, dass diese vollständig ist. Weiter ist $\mathbb{Q}^d$ abzählbar und jedes $\omega \in \mathbb{R}^d$ ist Häufungspunkt einer Folge in $\mathbb{Q}^d$. Insbesondere ist also $\overline{\mathbb{Q}^d} = \mathbb{R}^d$ nach Lemma 1.4, also ist $\mathbb{R}^d$ separabel. Insgesamt ist also $(\mathbb{R}^d, \mathcal{O})$ polnisch.

2. Sei $K \subseteq \mathbb{R}$ kompakt und $\Omega = \mathcal{C}_\mathbb{R}(K)$ die Menge der stetigen Funktionen $\omega : K \to \mathbb{R}$. Auf Ω sei
$$r(\omega_1, \omega_2) := \sup_{x \in K} |\omega_1(x) - \omega_2(x)|$$
der Supremumsabstand. Bekannt ist wieder, dass r vollständig ist. Außerdem lässt sich jedes $\omega \in \Omega$ nach dem Weierstraß'schen Approximationssatz gleichmäßig durch Polynome approximieren. Sei Ω' die abzählbare Menge der Polynome mit rationalen Koeffizienten. Dann gilt auch, dass $\overline{\Omega'} = \Omega$. Damit ist $(\Omega, \mathcal{O})$ separabel, also polnisch.

1.2 Kompakte Mengen

Topologische Räume können sehr groß sein. Man denke schon an den Raum $\mathbb{R}$, in dem es Folgen gibt, die divergieren. Nun werden *kompakten Menge* als kleinere Teilmengen eines topologischen Raumes betrachtet. In solchen kompakten Mengen gibt es immer konvergente Teilfolgen.

Definition 1.7 (Relativ kompakt, kompakt, relativ folgenkompakt, total beschränkt). *Sei $(\Omega, \mathcal{O})$ ein topologischer Raum und $K \subseteq \Omega$.*

1. *Die Menge K heißt* kompakt*, falls jede offene Überdeckung eine endliche Teilüberdeckung besitzt. Das bedeutet: sind $O_i \in \mathcal{O}, i \in I$ und $K \subseteq \bigcup_{i \in I} O_i$, dann gibt es[4] $J \Subset I$ mit $K \subseteq \bigcup_{i \in J} O_i$.*

[3] Wir schreiben $\varepsilon_n \downarrow 0$ falls $\varepsilon_1 \geq \varepsilon_2 \geq \ldots$ und $\varepsilon_n \xrightarrow{n\to\infty} 0$

[4] Wir schreiben $J \Subset I$, falls $J \subseteq I$ und J endlich ist.

2. *Die Menge K heißt* relativ kompakt*, falls $\overline{K}$ kompakt ist.*

3. *Die Menge K heißt* relativ folgenkompakt*, falls gilt: für jede Folge $\omega_1, \omega_2, \ldots \in K$ gibt es eine konvergente Teilfolge, d.h. $\omega_{k_1}, \omega_{k_2}, \ldots \in K$ und $\omega \in \Omega$ mit $\omega_{k_n} \xrightarrow{n\to\infty} \omega$ wie in* (1.4).

4. *Sei r eine Metrik, die $\mathcal{O}$ erzeugt. Dann heißt $K \subseteq \Omega$* total beschränkt*, wenn es für jedes $\varepsilon > 0$ ein $N \in \mathbb{N}$ und $\omega_1, \ldots, \omega_N \in K$ gibt, so dass $K \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$.*

Lemma 1.8 (Kompakte Mengen sind abgeschlossen). *Sei (Ω, r) ein metrischer Raum und $\mathcal{O}$ die von r erzeugte Topologie. Ist $K \subseteq \Omega$ kompakt, dann ist K auch abgeschlossen.*

Beweis. Wir zeigen, dass K^c offen ist. Sei hierzu $\omega \in K^c$. Für alle $\omega' \in K$ wählen wir $\delta_{\omega'}$ und $\varepsilon_{\omega'}$, so dass $B_{\delta_{\omega'}}(\omega) \cap B_{\varepsilon_{\omega'}}(\omega') = \emptyset$. Dann ist offenbar $\bigcup_{\omega' \in K} B_{\varepsilon_{\omega'}}(\omega') \supseteq K$, also gibt es $J \Subset K$ mit $K \subseteq \bigcup_{\omega' \in J} B_{\varepsilon_{\omega'}}(\omega')$. Setze $\delta := \min_{\omega' \in J} \delta_{\omega'} > 0$. Dann ist $B_\delta(\omega) \cap K \subseteq B_\delta(\omega) \cap \bigcup_{\omega' \in J} B_{\varepsilon_{\omega'}}(\omega') = \emptyset$, also $B_\delta(\omega) \subseteq K^c$. Da $\omega \in K^c$ beliebig war, ist K^c offen, K also abgeschlossen. □

Folgender Satz über kompakte Mengen unterstreicht zum ersten Mal, warum polnischen Räumen eine besondere Bedeutung zukommt.

Proposition 1.9 (Charakterisierung relativ kompakter Mengen). *Sei (Ω, r) ein metrischer Raum, $\mathcal{O}$ die von r erzeugte Topologie und $K \subseteq \Omega$. Betrachte folgende Aussagen:*

1. *K ist relativ kompakt.*

2. *Es gilt: sind $F_i \subseteq \overline{K}$ abgeschlossen, $i \in I$ und $\bigcap_{i\in I} F_i = \emptyset$, dann gibt es $J \Subset I$ mit $\bigcap_{i \in J} F_i = \emptyset$.*

3. *K ist relativ folgenkompakt.*

4. *K ist total beschränkt.*

Dann gilt

$$4. \Longleftarrow 1. \iff 2. \Longrightarrow 3.$$

Außerdem gilt auch 3. $\Longrightarrow$ 2. *falls $(\Omega, \mathcal{O})$ separabel ist und* 4. $\Longrightarrow$ 3. *falls (Ω, r) vollständig ist. Insbesondere sind alle vier Aussagen äquivalent, falls $(\Omega, \mathcal{O})$ polnisch ist.*

Korollar 1.10. *Sei (Ω, r) ein metrischer Raum, $\mathcal{O}$ die von r erzeugte Topologie. Dann sind abgeschlossene Teilmengen kompakter Mengen wieder kompakt.*

Beweis. Sei $K \subseteq \Omega$ kompakt und $A \subseteq K$ abgeschlossen. Eine abgeschlossene Menge ist genau dann kompakt, wenn sie relativ kompakt ist. Aus Proposition 1.9.2 liest man wegen der Relativkompaktheit von K ab, dass für $F_i \in \mathcal{O}^c, i \in I$ mit $F_i \subseteq A \subseteq K$ und $\bigcap_{i \in I} F_i = \emptyset$ ein $J \Subset I$ existiert mit $\bigcap_{i\in J} F_i = \emptyset$. Wieder mit Proposition 1.9.2 folgt daraus, dass A relativ kompakt, also kompakt ist. □

Beweis von Proposition 1.9. '1.⇒4.': Sei $\overline{K}$ kompakt und $\varepsilon > 0$. Offenbar ist $\bigcup_{\omega \in K} B_\varepsilon(\omega) \supseteq \overline{K}$ eine offene Überdeckung. Damit gibt es eine endliche Teilüberdeckung, d.h. es gibt $\omega_1, \ldots, \omega_N$ mit $\overline{K} \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$. Da $\varepsilon > 0$ beliebig war, folgt die Behauptung.

'1.⇒2.': Sei nun $F_i, i \in I$ wie angegeben. Dann ist $\bigcup_{i\in I} F_i^c = \Big(\bigcap_{i\in I} F_i\Big)^c = \Omega \supseteq \overline{K}$. Da $\overline{K}$ kompakt ist, gibt es $J \Subset I$ mit $\overline{K} \subseteq \bigcup_{i\in J} F_i^c$. Damit ist $\bigcap_{i\in J} F_i = \Big(\bigcup_{i\in J} F_i^c\Big) \subseteq \overline{K}^c$. Da aber $F_i \subseteq \overline{K}$ vorausgesetzt war, ist $\bigcap_{i\in J} F_i = \emptyset$.

'2.⇒1.': Sei $O_i \in \mathcal{O}, i \in I$ mit $\overline{K} \subseteq \bigcup_{i\in I} O_i$. Setze $F_i = O_i^c \cap \overline{K}$, dann ist $F_i^c \in \mathcal{O}$ und $\bigcap_{i\in I} F_i = \overline{K} \cap \Big(\bigcup_{i\in I} O_i\Big)^c = \emptyset$. Also gibt es $J \Subset I$ mit $\bigcap_{i\in J} F_i = \emptyset$. Damit ist $\overline{K}^c \cup \bigcup_{i\in J} O_i = \bigcup_{i\in J} F_i^c = \Omega$, also $\bigcup_{i\in J} O_i \supseteq \overline{K}$. Mit anderen Worten ist $\overline{K}$ kompakt.

'2.⇒3.': Sei $\omega_1, \omega_2, \ldots \in K$. Wir setzen $F_n = \overline{\{\omega_n, \omega_{n+1}, \ldots\}} \subseteq \overline{K}$. Angenommen, es gibt keine konvergente Teilfolge von $\omega_1, \omega_2, \ldots$ Dann ist $\bigcap_{n=1}^\infty F_n = \emptyset$. Aus 2. folgt dann, dass es ein $N \in \mathbb{N}$ gibt mit $\emptyset = \bigcap_{n=1}^N F_n = F_N$. Dies ist ein Widerspruch, da F_N nach Konstruktion nicht leer ist; also gibt es eine konvergente Teilfolge.

'3.⇒1.' falls $(\Omega, \mathcal{O})$ separabel ist: Sei Ω' abzählbar mit $\overline{\Omega'} = \Omega$ und $\mathcal{B} := \{B_{1/n}(\omega) : \omega \in \Omega', n \in \mathbb{N}\}$. Damit ist $\mathcal{B}$ eine Basis von $\mathcal{O}$ und auch abzählbar. Wir schreiben $\mathcal{B} = \{B_1, B_2, \ldots\}$.

Angenommen $\overline{K}$ ist nicht kompakt. Das heißt, es gibt $A_i \in \mathcal{O}, i \in I$ mit $\overline{K} \subseteq \bigcup_{i\in I} A_i$ und es gibt keine endliche Teilüberdeckung. Wir setzen für $i \in I$

$$J_i = \{j \in \mathbb{N} : B_j \subseteq A_i\} \subseteq \mathbb{N}$$

sowie $J := \bigcup_{i\in I} J_i \subseteq \mathbb{N}$. Damit ist $A_i = \bigcup_{j\in J_i} B_j$, also

$$K \subseteq \bigcup_{i\in I} A_i = \bigcup_{i\in I} \bigcup_{j\in J_i} B_j = \bigcup_{j\in J} B_j.$$

Klar ist, dass nun $B_j \in \mathcal{O}, j \in J$ eine abzählbare Überdeckung von $\overline{K}$ darstellt. Da es keine endliche Teilüberdeckung für die $A_i, i \in I$ gibt, kann es auch keine endliche Teilüberdeckung für $B_j, j \in J$ geben. (Jedes B_j ist schließlich Teilmenge eines A_i's.) Für $n \in \mathbb{N}$ setzen wir $\omega_n \in \overline{K} \setminus \bigcup_{j\in J, j\leq n} B_j$. Nach Voraussetzung hat die Folge $\omega_1, \omega_2, \ldots \in K$ einen Häufungspunkt $\omega \in \overline{K}$. Da $\overline{K} \subseteq \bigcup_{j\in J} B_j$, gibt es $k \in J \subseteq \mathbb{N}$ mit $\omega \in B_k$. Damit liegen einerseits (da B_k offen ist) unendlich viele der ω_n in B_k, andererseits ist $\omega_i \notin B_k$ für alle $i \geq k$ nach Konstruktion. Dies ist ein Widerspruch, also ist $\overline{K}$ kompakt.

'4.⇐'3. falls (Ω, r) vollständig ist: Sei $\omega_1, \omega_2, \ldots \in K$. Wir konstruieren eine Teilfolge, die Cauchy-Folge ist. Diese konvergiert, da (Ω, r) vollständig ist, und K ist als relativ folgenkompakt erkannt. Wähle eine Folge $\varepsilon_1, \varepsilon_2, \ldots > 0$ mit $\varepsilon_n \downarrow 0$. Da K total beschränkt ist, gibt es endlich viele ε_1-Bälle, die K überdecken. Mindestens einer dieser Bälle muss unendlich viele der ω_n enthalten. Diese haben jeweils höchstens Abstand $2\varepsilon_1$. Wähle ω_{k_1} als einer dieser unendlich vielen Punkte. Da dieser ε_1-Ball durch endlich viele ε_2-Bälle überdeckt wird, gibt es einen dieser ε_2-Bälle, der unendlich viele der ω_n enthält. Diese haben jeweils höchstens Abstand $2\varepsilon_2$. Wähle $\omega_{k_2} \neq \omega_{k_1}$ als einen dieser unendlich vielen Punkten. Durch weiteres Vorgehen erhalten wir eine Folge $\omega_{k_1}, \omega_{k_2}, \ldots \in K$, so dass $r(\omega_{k_n}, \omega_{k_m}) \leq 2\varepsilon_{m\wedge n}$. Mit anderen Worten haben wir wie angekündigt eine Cauchy-Teilfolge in K gefunden. □

Beispiel 1.11 (Kompakte metrische Räume sind polnisch). Sei (Ω, r) ein metrischer Raum und $\mathcal{O}$ die von r erzeugte Topologie. Falls Ω kompakt ist, dann ist $(\Omega, \mathcal{O})$ polnisch.

Wir zeigen, dass r vollständig ist. Sei $\omega_1, \omega_2, \ldots \in \Omega$ eine Cauchy-Folge. Da K relativ folgenkompakt nach Proposition 1.9 ist, gibt es eine konvergente Teilfolge $\omega_{k_1}, \omega_{k_2}, \ldots$ Sei $\omega \in \Omega$ der Grenzwert der konvergenten Teilfolge. Sei $\varepsilon > 0$ und $N \in \mathbb{N}$, so dass $r(\omega_m, \omega_n) < \varepsilon/2$ für $m, n > N$ und $r(\omega_{k_n}, \omega) < \varepsilon/2$ für $k_n > N$. Dann gilt für $m > N$, dass $r(\omega_m, \omega) \leq r(\omega_m, \omega_{k_n}) + r(\omega_{k_n}, \omega) \leq \varepsilon$. Daraus folgt, dass $\omega_n \xrightarrow{n\to\infty} \omega$.

Für die Separabilität von $(\Omega, \mathcal{O})$ sei $\varepsilon_1, \varepsilon_2, \ldots > 0$ mit $\varepsilon_n \downarrow 0$. Da K total beschränkt ist, gibt es für alle $n \in \mathbb{N}$ ein k_n und $\omega_{n1}, \ldots, \omega_{nk_n}$ mit $K \subseteq \bigcup_{k=1}^{k_n} B_{\varepsilon_n}(\omega_{nk_n})$. Sei $\Omega' = \{\omega_{nk} : n \in \mathbb{N}, k = 1, \ldots, k_n\}$. Dann ist Ω' abzählbar und für jedes $\omega \in \Omega$ und jedes $n \in \mathbb{N}$ gibt es ein $k(\omega, n) \in \{1, \ldots, k_n\}$ mit $r(\omega_{k(\omega,n)}, \omega) < \varepsilon_n$. Damit gilt $(\omega_{k(\omega,n)}, \omega) \xrightarrow{n\to\infty} \omega$. Also ist $\overline{\Omega'} = \Omega$.

2 Mengensysteme

Die Wahrscheinlichkeitstheorie formalisiert das umgangssprachlich verwendete Wort *wahrscheinlich*. Dies ist (im weitesten Sinne) eine Eigenschaft eines möglichen Ausgangs eines Experimentes. Grundlegend in der Wahrscheinlichkeitstheorie ist der Begriff des *Ereignisses*, der alles beschreiben soll, was bei dem Experiment passieren kann. Ereignisse werden durch Teilmengen eines abstrakten Grundraumes, der immer Ω genannt wird, dargestellt. Ziel dieses Abschnittes ist es, möglichst vielen Teilmengen von Ω eine Wahrscheinlichkeit zuzuordnen. Dies führt auf den Begriff der σ-Algebra, denn diese enthalten genau die Teilmengen des Grundraumes, denen dann im nächsten Abschnitt Wahrscheinlichkeiten zugeordnet werden. Mit anderen Worten werden Elemente von σ-Algebren Ereignisse im obigen Sinne sein. Die anderen in diesem Abschnitt eingeführten Mengensysteme werden dazu dienen, geeignete σ-Algebren zu definieren.

2.1 Halbringe, Ringe und σ-Algebren

Die in diesem Abschnitt eingeführten Begriffe haben einen einfachen Zusammenhang. Ist nämlich $\mathcal{C} \subseteq 2^\Omega$, so gilt

$$\mathcal{C} \text{ ist } \sigma\text{-Algebra} \quad \Longrightarrow \quad \mathcal{C} \text{ ist Ring} \quad \Longrightarrow \quad \mathcal{C} \text{ ist Halbring.}$$

Beziehungen zwischen den Mengensystemen sind in Tabelle 2.1 festgehalten.

Definition 2.1 (Halbring, Ring, σ-Algebra).

1. *Ein Mengensystem $\mathcal{H}$ heißt* schnittstabil*, falls mit $A, B \in \mathcal{H}$ auch $A \cap B \in \mathcal{H}$ gilt. Es heißt* vereinigungsstabil*, falls mit $A, B \in \mathcal{H}$ auch $A \cup B \in \mathcal{H}$ gilt. Es heißt komplementstabil, wenn mit $A \in \mathcal{H}$ auch $A^c \in \mathcal{H}$ gilt. Es heißt* differenzmengenstabil*, falls mit $A, B \in \mathcal{H}$ auch $B \setminus A \in \mathcal{H}$ gilt.*

2. *Ein nicht-leeres Mengensystem $\mathcal{H}$ ist ein* Halbring *(auf Ω), falls es (i) schnittstabil ist, und (ii) wenn es für $A, B \in \mathcal{H}$ Mengen $C_1, \ldots, C_n \in \mathcal{H}$ gibt mit[5] $B \setminus A = \biguplus_{i=1}^n C_i$.*

3. *Ein Mengensystem $\mathcal{R}$ heißt* Ring *(auf Ω), falls (i) mit $A, B \in \mathcal{R}$ ist auch $A \cup B \in \mathcal{R}$ und (ii) mit $A, B \in \mathcal{R}$ ist auch $B \setminus A \in \mathcal{R}$.*

[5] Wir schreiben $A \uplus B$ für $A \cup B$, falls $A \cap B = \emptyset$.

	$\mathcal{C}$ Halbring	$\mathcal{C}$ Ring	$\mathcal{C}$ σ-Algebra
$\mathcal{C}$ schnittstabil	$\bullet$	$\circ$	$\circ$
$\mathcal{C}$ σ-schnittstabil			$\circ$
$\mathcal{C}$ vereinigungsstabil		$\bullet$	$\circ$
$\mathcal{C}$ σ-vereinigungsstabil			$\bullet$
$\mathcal{C}$ differenzmengenstabil		$\bullet$	$\circ$
$\mathcal{C}$ komplementstabil			$\bullet$
$B \setminus A = \biguplus_{i=1}^n C_i$	$\bullet$	$\circ$	$\circ$
$\Omega \in \mathcal{C}$			$\circ$

Tabelle 2.1: Für $\mathcal{C} \subseteq 2^\Omega$ wird hier die Beziehung zwischen Halbringen, Ringen und σ-Algebren dargestellt. Ein $\bullet$ bedeutet, dass in der Definition des Mengensystems (Spalte) die entsprechende Eigenschaft (Zeile) gefordert ist. Ein $\circ$ bedeutet, dass die entsprechende Eigenschaft aus den definierenden Eigenschaften des Mengensystems folgt.

4. *Ein Mengensystem $\mathcal{F}$ heißt* σ-Algebra *(auf Ω), falls mit $A, A_1, A_2, \ldots \in \mathcal{F}$ auch $A^c, \bigcup_{n=1}^\infty A_n \in \mathcal{F}$ gilt. Dann heißt $(\Omega, \mathcal{F})$ ein* Messraum.

Bemerkung 2.2 (Beziehungen zwischen den und weitere Mengensystemen).

1. *Jeder Ring ist ein Halbring:* Um dies einzusehen, ist nur nachzuprüfen, dass jeder Ring $\mathcal{R}$ schnittstabil ist. Dies folgt aus der Differenzmengenstabilität, da mit $A, B \in \mathcal{R}$ auch $A \cap B = A \setminus (A \setminus B) \in \mathcal{R}$.

2. *Jede σ-Algebra ist schnittstabil und damit ein Ring:* Das ist klar, da $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c$ und $B \setminus A = B \cap A^c$.

3. *Algebra und Verband:* Eine *Algebra* ist ein vereinigungs- und komplementstabiles Mengensystem. Ein schnitt- und vereinigungsstabiles Mengensystem heißt *Verband*. Unsere Darstellung kommt jedoch ohne diese Begriffe aus.

Beispiel 2.3 (Halbringe, σ-Algebren).

1. *Halboffene Intervalle bilden einen Halbring:* Sei $\Omega = \mathbb{R}$. Dann definiert

$$\mathcal{H} := \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$$

einen Halbring. Denn es gilt für $a_1 \leq b_1, a_1' \leq b_1'$, dass[6] $(a_1, b_1] \cap (a_1', b_1'] = (a_1 \vee a_1', b_1 \wedge b_1']$ und $(a_1, b_1] \setminus (a_1', b_1'] = (a_1, a_1' \wedge b_1] \uplus (b_1', b_1]$, wobei $(a, b] = \emptyset$, falls $a \geq b$.

[6] Wie üblich bezeichnet $x \wedge y := \min(x, y)$ und $x \vee y := \max(x, y)$

2. *Einfache Beispiele für σ-Algebren:* Die einfachsten Beispiele für σ-Algebren sind $\{\emptyset, \Omega\}$ und 2^Ω. (Beides sind übrigens auch Topologien.) Für überabzählbares Ω ist ein nicht ganz so triviales Beispiel für eine σ-Algebra das Mengensystem
$$\{A \subseteq \Omega : \ A \text{ abzählbar oder } A^c \text{ abzählbar}\}.$$
(Dieses Mengensystem ist kein Topologie.)
Ein weiteres Beispiel werden wir in Abschnitt 4.1 antreffen: Falls $\mathcal{F}'$ eine σ-Algebra auf Ω' ist und $f : \Omega \to \Omega'$. Dann ist
$$\sigma(f) := \{f^{-1}(A') : A' \in \mathcal{F}'\} \subseteq 2^\Omega \tag{2.1}$$
eine σ-Algebra auf Ω. Ist nämlich $A', A'_1, A'_2, \ldots \in \sigma(f)$, so ist $(f^{-1}(A'))^c = f^{-1}((A')^c) \in \sigma(f)$ und $\bigcup_{n=1}^\infty f^{-1}(A'_n) = f^{-1}\Big(\bigcup_{n=1}^\infty A'_n\Big) \in \sigma(f)$.
In der Praxis am weitaus häufigsten benötigt man Borel'sche σ-Algebren; siehe Definition 2.7.

2.2 Erzeuger und Erweiterungen

Auf der einen Seite ist es Ziel der Maßtheorie, Mengenfunktionen auf σ-Algebren zu definieren. Auf der anderen Seite kann man oftmals nur Halbringe konkret angeben, nicht jedoch σ-Algebren; siehe Beispiel 2.3.1. Allerdings gibt es für jeden Halbring $\mathcal{H}$ einen kleinsten Ring, der $\mathcal{H}$ enthält, $\mathcal{R}(\mathcal{H})$. Genauso gibt es eine kleinste σ-Algebra $\mathcal{F}$, die einen Halbring $\mathcal{H}$ (oder einen Ring $\mathcal{R}$) enthält, $\sigma(\mathcal{H})$. Damit erzeugt jeder Halbring $\mathcal{H}$ eine σ-Algebra.

Bemerkung 2.4 (Erzeugte Mengensysteme). Leicht überzeugt man sich, dass der Schnitt von Ringen (σ-Algebren) wieder ein Ring (σ-Algebra) ist. Insbesondere benötigen wir folgende Begriffe:
Sei $\mathcal{C} \subseteq 2^\Omega$, dann ist
$$\mathcal{R}(\mathcal{C}) := \bigcap \Big\{\mathcal{R} \supseteq \mathcal{C} : \mathcal{R} \text{ Ring}\Big\}$$
der *von $\mathcal{C}$ erzeugte Ring* und
$$\sigma(\mathcal{C}) := \bigcap \Big\{F \supseteq \mathcal{C} : \mathcal{F}\ \sigma\text{-Algebra}\Big\}$$
die *von $\mathcal{C}$ erzeugte σ-Algebra.* Klar ist, dass $\mathcal{R}(\mathcal{R}(\mathcal{H})) = \mathcal{R}(\mathcal{H})$ und $\sigma(\sigma(\mathcal{H})) = \sigma(\mathcal{H})$.

Lemma 2.5 (Von Halbring erzeugter Ring). *Sei $\mathcal{H}$ ein Halbring. Dann ist*
$$\mathcal{R}(\mathcal{H}) = \Big\{ \biguplus_{k=1}^n A_k : A_1, \ldots, A_n \in \mathcal{H} \text{ disjunkt}, n \in \mathbb{N}\Big\}$$
der von $\mathcal{H}$ erzeugte Ring.

Beispiel 2.6 (Von Intervallen erzeugter Ring). Sei $\mathcal{H}$ der von halboffenen Intervallen erzeugte Halbring aus Beispiel 2.3. Dann ist also
$$\mathcal{R}(\mathcal{H}) = \Big\{ \biguplus_{k=1}^n (a_k, b_k] : a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{Q}, \\ a_k < b_k, k = 1, \ldots, n \text{ und } a_k < b_{k+1}, k = 1, \ldots, n-1\Big\}$$
der von $\mathcal{H}$ erzeugte Ring.

Beweis von Lemma 2.5. Klar ist, dass $\mathcal{R}(\mathcal{H})$ schnittstabil ist. Um zu zeigen, dass $\mathcal{R}(\mathcal{H})$ ein Ring ist, zeigen wir zunächst die Differenzmengenstabilität. Sei $A_1, \ldots, A_n \in \mathcal{H}$ und $B_1, \ldots, B_m \in \mathcal{H}$, jeweils disjunkt. Dann gilt

$$\Big(\biguplus_{i=1}^{n} A_i\Big) \setminus \Big(\biguplus_{j=1}^{m} B_j\Big) = \biguplus_{i=1}^{n} \bigcap_{j=1}^{m} A_i \setminus B_j \in \mathcal{R}(\mathcal{H}).$$

Um die Vereinigungsstabilität von $\mathcal{R}(\mathcal{H})$ zu zeigen sei $A, B \in \mathcal{R}(\mathcal{H})$. Dann ist auch $A \cup B = (A \cap B) \uplus (A \setminus B) \uplus (B \setminus A) \in \mathcal{R}(\mathcal{H})$, da Schnitt- und Differenzmengenstabilität schon gezeigt sind.

Es gibt keinen kleineren Ring, der $\mathcal{H}$ enthält. Schließlich müsste dieser vereinigungsstabil sein, und ist damit eine Obermenge von $\mathcal{R}(\mathcal{H})$. □

Definition 2.7 (Borel'sche σ-Algebra). *Sei $(\Omega, \mathcal{O})$ ein topologischer Raum. Dann bezeichnet man mit $\mathcal{B}(\Omega) := \sigma(\mathcal{O})$ die* Borel'sche σ-Algebra *auf Ω. Ist $\Omega \subseteq \mathbb{R}^d$, so bezeichnen wir mit $\mathcal{B}(\Omega)$ die Borel'sche σ-Algebra, die von der euklidischen Topologie auf $\mathbb{R}^d$ erzeugt wird. Ist $\Omega \subseteq \overline{\mathbb{R}}$, so ist $\mathcal{B}(\Omega)$ die Borel'sche σ-Algebra, die von der Topologie aus Beispiel 1.2 erzeugt wird. Mengen in $\mathcal{B}(\overline{\mathbb{R}})$ bezeichnet man auch als* (Borel-)messbare Mengen.

Lemma 2.8 (Abzählbare Basis und Borel'sche σ-Algebra). *Sei $(\Omega, \mathcal{O})$ ein topologischer Raum mit abzählbarer Basis $\mathcal{C} \subseteq \mathcal{O}$. Dann gilt $\sigma(\mathcal{O}) = \sigma(\mathcal{C})$. Insbesondere ist Ω separabel.*

Beweis. Wir müssen nur zeigen, dass $\mathcal{O} \subseteq \sigma(\mathcal{C})$. Dies ist jedoch klar, da $A \in \mathcal{O}$ als abzählbare Vereinigung von Mengen aus $\mathcal{C}$ dargestellt werden kann. Siehe Lemma 1.5. □

Lemma 2.9 (Borel'sche σ-Algebra wird von Intervallen erzeugt). *Sei*

$$\begin{aligned}
\mathcal{C}_1 &= \{[-\infty, b] : b \in \mathbb{Q}\} \text{ oder} \\
\mathcal{C}_2 &= \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \\
\mathcal{C}_3 &= \{(a, b) : a, b \in \mathbb{Q}, a \leq b\} \\
\mathcal{C}_4 &= \{[a, b] : a, b \in \mathbb{Q}, a \leq b\}.
\end{aligned}$$

Dann ist $\sigma(\mathcal{C}_i) = \mathcal{B}(\overline{\mathbb{R}})$, $i = 1, \ldots, 4$.

Beweis. Das Mengensystem $\mathcal{C}_3$ ist eine abzählbare Basis der euklidischen Topologie auf $\overline{\mathbb{R}}$. Für diesen Fall folgt die Aussage aus Lemma 2.8.

Wir zeigen die Aussage nur für $\mathcal{C}_1$ und $\mathcal{C}_2$, die für $\mathcal{C}_4$ folgt analog. Zunächst ist $\mathcal{C}_2 := \{A \setminus B : A, B \in \mathcal{C}_1\} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \subseteq \sigma(\mathcal{C}_2)$ der von $\mathcal{C}_1$ erzeugte Halbring aus Beispiel 2.3. Damit ist $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$ und es genügt zu zeigen, dass $\sigma(\mathcal{C}_2) = \mathcal{B}(\overline{\mathbb{R}})$.

Sei hierzu $\mathcal{O}$ wie in Definition 1.1.8 mit $\Omega = \mathbb{R}$. Wir zeigen (i) aus $A \in \mathcal{O}$ folgt, dass $A \in \sigma(\mathcal{C}_2)$, und (ii) aus $A \in \mathcal{C}_2$ folgt, dass $A \in \sigma(\mathcal{O})$. Daraus folgt dann $\mathcal{O} \subseteq \sigma(\mathcal{C}_2) \subseteq \sigma(\mathcal{O})$, also $\sigma(\mathcal{O}) = \sigma(\mathcal{C}_2)$.
Für (i) sei also $A \in \mathcal{O}$. Wir behaupten

$$A = \bigcup \{(a, b] : (a, b] \subseteq A, a, b \in \mathbb{Q}\}, \tag{2.2}$$

und bemerken, dass die rechte Seite ein Element von $\sigma(\mathcal{C}_2)$ ist. In (2.2) ist '$\supseteq$' klar. Um '$\subseteq$' einzusehen, wählen wir $x \in A$. Dann gibt es nach Definition von $\mathcal{O}$ ein $\varepsilon > 0$, so dass

$B_\varepsilon(x) \subseteq A$. Allerdings gibt es auch $a, b \in \mathbb{Q}$ mit $a \leq b$ und $x \in (a, b] \subseteq B_\varepsilon(x)$. Damit ist '$\subseteq$' gezeigt und (i) folgt.
Für (ii) gehen wir ähnlich vor; sei $A \in \mathcal{C}_2$. Dann ist offenbar

$$A = \bigcap_{n=1}^{\infty} \big(a, b + \tfrac{1}{n}\big).$$

Da $\big(a, b + \frac{1}{n}\big) \in \mathcal{O}$, ist also $A \in \sigma(\mathcal{O})$. □

Beispiel 2.10 (Borel-messbare Mengen). Natürlich sind alle abzählbaren Durchschnitte und Vereinigungen von Intervallen nach Lemma 2.9 in $\mathcal{B}(\overline{\mathbb{R}})$. Sei beispielsweise

$$\begin{aligned}
&A_1 = [0, \tfrac{1}{3}] \cup [\tfrac{2}{3}, 1],\\
&A_2 = [0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{3}{9}] \cup [\tfrac{6}{9}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, 1],\\
&A_3 = [0, \tfrac{1}{27}] \cup [\tfrac{2}{27}, \tfrac{3}{27}] \cup [\tfrac{6}{27}, \tfrac{7}{27}] \cup [\tfrac{8}{27}, \tfrac{9}{27}] \cup [\tfrac{18}{27}, \tfrac{19}{27}] \cup [\tfrac{20}{27}, \tfrac{21}{27}] \cup [\tfrac{24}{27}, \tfrac{25}{27}] \cup [\tfrac{26}{27}, 1],\\
&\dots
\end{aligned}$$

dann bezeichnet $A = \bigcap_{n=1}^{\infty}$ das Cantor'sche Diskontinuum. Diese Menge ist als abzählbarer Schnitt von endlichen Vereinigungen von Intervallen messbar. Wir werden in Beispiel 3.26 ein Beispiel für eine nicht Borel-messbare Menge kennen lernen.

2.3 Dynkin-Systeme

Oftmals muss man in der Maßtheorie zeigen, dass ein bestimmtes Mengensystem $\mathcal{F}$ eine σ-Algebra ist und einen Halbring $\mathcal{H}$ beinhaltet. Die in diesem Abschnitt behandelten Dynkin-Systeme sind dabei sehr hilfreich. Es genügt nämlich wegen Theorem 2.13 zu zeigen, dass $\mathcal{F}$ ein schnittstabiles Dynkin-System mit $\mathcal{H} \subseteq \mathcal{F}$ ist. Dies ist oftmals einfacher, als direkt zu zeigen, dass $\mathcal{F}$ eine σ-Algebra ist.

Definition 2.11 (Dynkin-System). *Ein Mengensystem $\mathcal{D}$ heißt* Dynkin-System *(auf Ω), falls (i) $\Omega \in \mathcal{D}$, (ii) mit $A, B \in \mathcal{D}$ und $A \subseteq B$ ist auch $B \setminus A \in \mathcal{D}$ und (iii) mit $A_1, A_2, \ldots \in \mathcal{D}$ mit $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ ist auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$.*

Beispiel 2.12 (σ-Algebren sind Dynkin-Systeme). *Jede σ-Algebra ist ein Dynkin-System:* Sei $\mathcal{F}$ eine σ-Algebra. Dann ist mit $A, B \in \mathcal{F}$ auch $A^c \in \mathcal{F}$ und damit $\Omega = A \cup A^c \in \mathcal{F}$ sowie $B \setminus A = B \cap A^c \in \mathcal{F}$.

Theorem 2.13 (Schnittstabile Dynkin-Systeme). *Sei $\mathcal{D}$ ein Dynkin-System und $\mathcal{C} \subseteq \mathcal{D}$ schnittstabil. Dann ist $\sigma(\mathcal{C}) \subseteq \mathcal{D}$. Insbesondere ist jedes schnittstabile Dynkin-System eine σ-Algebra.*

Beweis. Genau wie in Lemma 2.4 definiert man das von $\mathcal{C}$ erzeugte Dynkin-System $\lambda(\mathcal{C})$. Da offenbar der Schnitt von Dynkin-Systemen wieder ein Dynkin-System ist, gilt $\lambda(\mathcal{C}) \subseteq \mathcal{D}$. Wir werden zeigen, dass $\lambda(\mathcal{C})$ eine σ-Algebra ist, denn dann ist $\sigma(\mathcal{C}) \subseteq \sigma(\lambda(\mathcal{C})) = \lambda(\mathcal{C}) \subseteq \mathcal{D}$.

Wir zeigen hierzu, dass $\lambda(\mathcal{C})$ schnittstabil ist. Dann ist nämlich für $A_1, A_2, \ldots \in \lambda(\mathcal{C})$

$$\bigcup_{i=1}^{n} A_i = \Big(\bigcap_{i=1}^{n} A_i^c\Big)^c \in \lambda(\mathcal{C})$$

und, da $\lambda(\mathcal{C})$ ein Dynkin-System ist $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty \bigcup_{i=1}^n A_i \in \lambda(\mathcal{C})$.

Wir müssen also zeigen, dass mit $A, B \in \lambda(\mathcal{C})$ auch $A \cap B \in \lambda(\mathcal{C})$ gilt. Falls $A, B \in \mathcal{C}$, so ist dies wegen der Schnittstabilität von $\mathcal{C}$ klar. Für $B \in \mathcal{C}$ setzen wir

$$\mathcal{D}_B := \{A \subseteq \Omega : A \cap B \in \lambda(\mathcal{C})\}.$$

Dann ist $\mathcal{D}_B$ ein Dynkin-System, da (i) $\Omega \in \mathcal{D}_B$, (ii) für $A, C \subseteq \mathcal{D}_B$ ist $A \cap B, C \cap B \in \lambda(\mathcal{C})$ also mit $A \subseteq C$ auch $A \cap B \subseteq C \cap B$, also $(C \setminus A) \cap B = (C \cap B) \setminus (A \cap B) \in \lambda(\mathcal{C})$ und (iii) für $A_1, A_2, \ldots \in \mathcal{D}_B$ ist $A_1 \cap B, A_2 \cap B, \ldots \in \lambda(\mathcal{C})$ und falls $A_1 \subseteq A_2 \subseteq \cdots$ ist $A_1 \cap B \subseteq A_2 \cap B \subseteq \cdots$, woraus $\left(\bigcup_{n=1}^\infty A_n\right) \cap B = \left(\bigcup_{n=1}^\infty A_n \cap B\right) \in \lambda(\mathcal{C})$ folgt.

Da $\mathcal{C} \subseteq \mathcal{D}_B$ und $\mathcal{D}_B$ ein Dynkin-System ist, ist auch $\lambda(\mathcal{C}) \subseteq \mathcal{D}_B$. Dies bedeutet, dass für $A \in \lambda(\mathcal{C})$ und $B \in \mathcal{C}$ auch $A \cap B \in \lambda(\mathcal{C})$. Wir setzen nun für ein $A \in \lambda(\mathcal{C})$

$$\mathcal{B}_A := \{B \subseteq \Omega : A \cap B \in \lambda(\mathcal{C})\}.$$

Wie oben zeigt man, dass $\mathcal{B}_A$ ein Dynkin-System mit $\mathcal{C} \subseteq \mathcal{B}_A$ ist. Damit ist wieder $\lambda(\mathcal{C}) \subseteq \mathcal{B}_A$. Insbesondere ist mit $A, B \in \lambda(\mathcal{C})$ auch $A \cap B \in \lambda(\mathcal{C})$ und die erste Behauptung ist gezeigt. Die zweite Behauptung folgt, wenn man $\mathcal{C} := \mathcal{D}$ setzt. □

2.4 Kompakte Systeme

Nachdem wir im Abschnitt 1.2 kompakte Teilmengen von Ω kennen gelernt haben, stellen wir nun einen wichtigen Zusammenhang zwischen Systemen kompakter Mengen und der Maßtheorie vor. Solche kompakten Systeme spielen eine wichtige Rolle beim Beweis von Theorem 3.9. Hier wird gezeigt, dass aus der Additivität einer Mengenfunktion und einer Approximationseigenschaft bzgl. eines kompakten Systems die σ-Additivität der Mengenfunktion folgt.

Definition 2.14 (Kompaktes System). *Ein schnittstabiles Mengensystem $\mathcal{K}$ heißt* kompaktes System *(auf Ω), falls aus $\bigcap_{n=1}^\infty K_n = \emptyset$ mit $K_1, K_2, \ldots \in \mathcal{K}$ folgt, dass es ein $N \in \mathbb{N}$ gibt mit $\bigcap_{n=1}^N K_n = \emptyset$.*

Beispiel 2.15 (Kompakte Mengen). *Kompakte Mengen bilden ein kompaktes System:* Sei (Ω, r) ein metrischer Raum und $\mathcal{O}$ die von r erzeugte Topologie. Dann ist jedes schnittstabile $\mathcal{K} \subseteq \{K \subseteq \Omega : K \text{ kompakt}\}$ ein kompaktes System. Denn: sei $\bigcap_{n=1}^\infty K_n = \emptyset$. Dann ist sowohl K_1, als auch $L_n := K_1 \cap K_n \subseteq K_1$ für $n = 1, 2, \ldots$ nach Lemma 1.8 abgeschlossen und wegen der Kompaktheit von K_1 gibt es ein N mit $\bigcap_{n=1}^N K_n = \emptyset$ nach Proposition 1.9.

Lemma 2.16 (Erweiterung kompakter Systeme). *Sei $\mathcal{K}$ ein kompaktes System. Dann ist auch*

$$\mathcal{K}_\cup := \left\{ \bigcup_{i=1}^n K_i : K_1, \ldots, K_n \in \mathcal{K}, n \in \mathbb{N} \right\}$$

ein kompaktes System.

Beweis. Klar ist, dass $\mathcal{K}_\cup$ schnittstabil ist. Seien $L_1, L_2, \ldots \in \mathcal{K}_\cup$ mit $\bigcap_{n=1}^N L_n \neq \emptyset$ für alle $N \in \mathbb{N}$. Zu zeigen ist, dass dann auch $\bigcap_{n=1}^\infty L_n \neq \emptyset$. Dazu zeigen wir mit Induktion über N folgende Aussage:

> Es gibt für jedes $N \in \mathbb{N}$ Mengen $K_1, \ldots, K_N \in \mathcal{K}$ mit $K_n \subseteq L_n$, $n = 1, \ldots, N$, so dass für alle $k \in \mathbb{N}_0$ gilt dass $K_1 \cap \cdots \cap K_N \cap L_{N+1} \cap \cdots \cap L_{N+k} \neq \emptyset$.

Sei $N = 1$. Da $L_1 = \bigcup_{j=1}^{m_1} K'_j$ für $K'_1, \dots, k'_{m_1} \in \mathcal{K}$ und $\bigcap_{n=1}^N L_n = \bigcup_{j=1}^{m_1} K'_j \cap \bigcap_{n=1}^N L_n \neq \emptyset$ für jedes $N \in \mathbb{N}$, gibt es ein $j = 1, \dots, m_1$, so dass $K'_j \cap \bigcap_{n=1}^N L_n \neq \emptyset$ für jedes $N \in \mathbb{N}$. Setze $K_1 := K'_j$, und die Behauptung ist für $N = 1$ gezeigt.

Gelte die Behauptung für $N-1$. Wieder ist $L_N = \bigcup_{j=1}^{m_N} K''_j$ für $K''_1, \dots, K''_{m_N} \in \mathcal{K}$. Damit gilt nach der Induktionsvoraussetzung für alle $k \in \mathbb{N}$, dass

$$\begin{aligned} K_1 \cap \dots \cap K_{N-1} \cap \Big(\bigcup_{j=1}^{m_N} K''_j \Big) \cap L_{N+1} \cap \dots \cap L_{N+k} \\ = \bigcup_{j=1}^{m_N} K_1 \cap \dots \cap K_{N-1} \cap K''_j \cap L_{N+1} \cap \dots \cap L_{N+k} \neq \emptyset. \end{aligned}$$

Damit gibt es ein j, so dass $K_1 \cap \dots \cap K_{N-1} \cap K''_j \cap L_{N+1} \cap \dots \cap L_{N+k} \neq \emptyset$ für alle $k \in \mathbb{N}$. Setze $K_N := K''_j$, was den Induktionsbeweis abschließt.

Setzt man $k = 0$ in obiger Behauptung, sieht man, dass es $K_1, K_2, \dots \in \mathcal{K}$ und $K_n \subseteq L_n$, $n \in \mathbb{N}$ gibt mit $\bigcap_{n=1}^N K_n \neq \emptyset$ für alle $N \in \mathbb{N}$. Da $\mathcal{K}$ ein kompaktes System ist, gilt insbesondere

$$\emptyset \neq \bigcup_{n=1}^{\infty} K_n \subseteq \bigcup_{n=1}^{\infty} L_n$$

und die Behauptung ist gezeigt. $\square$

3 Maße

Will man Mengen *messen*, also etwa jeder Menge ihren Inhalt zuordnen, scheinen einige Forderungen natürlich. Dies ist aus der Stochastik bekannt, wo 'messen' bedeutet, dass einer Menge eine Wahrscheinlichkeit zugeordnet wird. Etwa sollte die leere Menge (ein Ereignis, das nie eintritt) *Masse* 0 zugeordnet werden, oder die Massen sollten sich abzählbar additiv verhalten, siehe (3.1). Zentral in der Wahrscheinlichkeitstheorie ist deshalb der Begriff des Wahrscheinlichkeitsmaßes. Wie sich herausstellt, sind Maße auf σ-Algebren zu definieren, damit die Forderung der abzählbaren Additivität erfüllt werden kann. In diesem Abschnitt geben wir die wichtigsten Schritte an, solche Maße zu konstruieren. Als wichtigstes Beispiel eines Maßes dient dabei das Lebesgue-Maß. Dies ist zwar kein Wahrscheinlichkeitsmaß, jedoch funktioniert seine Konstruktion analog zur Konstruktion von Wahrscheinlichkeitsmaßen.

3.1 Mengenfunktionen

Wir werden nun Funktionen $\mu : \mathcal{C} \to \overline{\mathbb{R}}_+$ betrachten, wenn $\mathcal{C}$ ein Halbring, Ring oder eine σ-Algebra ist. Für die Stochastik am wichtigsten ist dabei sicherlich der Begriff des *Wahrscheinlichkeitsmaßes*, das den Spezialfall $\mu(\Omega) = 1$ beschreibt.

Definition 3.1 (Maß und äußeres Maß). *Sei $\mathcal{F} \subseteq 2^\Omega$ und $\mu : \mathcal{F} \to \overline{\mathbb{R}}_+$.*

1. *Die Abbildung μ heißt* endlich additiv, *falls für disjunkte $A_1, \dots, A_n \in \mathcal{F}$ mit $\biguplus_{k=1}^n A_k \in \mathcal{F}$ gilt, dass*

$$\mu\Big(\biguplus_{k=1}^n A_k \Big) = \sum_{k=1}^n \mu(A_k). \tag{3.1}$$

Sie heißt subadditiv*, falls für (beliebige)* $A_1, \dots, A_n \in \mathcal{F}$ *mit* $\bigcup_{k=1}^n A_k \in \mathcal{F}$ *gilt, dass*

$$\mu\Big(\bigcup_{k=1}^n A_k\Big) \leq \sum_{k=1}^n \mu(A_k). \tag{3.2}$$

2. *Eine Abbildung* $\mu : \mathcal{F} \to \mathbb{R}_+$ *heißt* σ-additiv*, falls* (3.1) *auch für* $n = \infty$ *gilt. Sie heißt* σ-sub-additiv*, wenn* (3.2) *auch für* $n = \infty$ *gilt. Sie heißt* monoton*, wenn aus* $A, B \in \mathcal{F}$ *mit* $A \subseteq B$ *folgt, dass* $\mu(A) \leq \mu(B)$.

3. *Falls es eine Folge* $\Omega_1, \Omega_2, \ldots \in \mathcal{F}$ *mit* $\bigcup_{n=1}^\infty \Omega_n = \Omega$ *und* $\mu(\Omega_n) < \infty$ *für alle* $n = 1, 2, \ldots$ *gilt, so heißt* μ σ-endlich.

4. *Sei* $\mathcal{F}$ *eine* σ*-Algebra und* $\mu : \mathcal{F} \to \mathbb{R}_+$*. Ist* μ *auch* σ*-additiv, so heißt* μ *ein* Maß *(auf* $\mathcal{F}$*) und* $(\Omega, \mathcal{F}, \mu)$ *ist ein* Maßraum*. Gilt* $\mu(\Omega) < \infty$*, so heißt* μ endliches Maß *und falls* $\mu(\Omega) = 1$*, so heißt* μ *ein* Wahrscheinlichkeitsmaß *oder eine* Wahrscheinlichkeitsverteilung *oder auch einfach eine* Verteilung*. Außerdem heißt* $(\Omega, \mathcal{F}, \mu)$ *dann ein* Wahrscheinlichkeitsraum*.*

5. *Sei* $(\Omega, \mathcal{O})$ *ein topologischer Raum und* μ *ein Maß auf* $\mathcal{B}(\mathcal{O})$*. Dann heißt die kleinste abgeschlossene Menge* F *(d.h.* $F^c \in \mathcal{O}$*) mit* $\mu(F^c) = 0$ *der* Träger von μ*.*

6. *Eine* σ*-subadditiv, monotone Abbildung* $\mu^* : 2^\Omega \to \mathbb{R}_+$ *heißt* äußeres Maß*, falls* $\mu^*(\emptyset) = 0$*. Eine Menge* $A \subseteq \Omega$ *heißt* μ^*-messbar*, falls*

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^c) \tag{3.3}$$

für alle $E \subseteq \Omega$ *gilt.*

7. *Sei* $\mathcal{F}$ *schnittstabil und* $\mathcal{K} \subseteq \mathcal{F}$ *ein kompaktes System. Dann heißt* μ *von innen* $\mathcal{K}-$*regulär, falls für alle* $A \in \mathcal{K}$

$$\mu(A) = \sup_{\mathcal{K} \ni K \subseteq A} \mu(K).$$

Beispiel 3.2 (Beispiele von Mengenfunktionen). 1. Wir werden uns oft mit Mengenfunktionen auf $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ aus 2.3 beschäftigen. Etwa definiert $\mu((a, b]) = b - a$ eine additive, σ-endliche Mengenfunktion. Wir werden diese Funktion in eindeutiger Weise auf die Borel'sche σ-Algebra $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{H})$ (siehe Lemma 2.9) erweitern und so das Lebesgue-Maß konstruieren, siehe Korollar 3.17.

2. Einfache Beispiele für Mengenfunktionen sind die Dirac-Maße. Ist $\omega' \in \Omega$, so ist

$$\delta_{\omega'} : \begin{cases} 2^\Omega & \to \{0, 1\} \\ A & \mapsto 1_{\{\omega' \in A\}} \end{cases}$$

ein (Wahrscheinlichkeits-)Maß.

3. Für $\mu_i = \delta_{\omega_i}$, $i \in I$ heißt $\sum_{i \in I} \delta_{\omega_i}$ ein Zählmaß.

4. Sind $\mu_i, i \in I$ Maße auf einer σ-Algebra $\mathcal{F}$. Dann ist für $a_i \in \mathbb{R}_+, i \in I$, auch $\sum_{i \in I} a_i \mu_i$ ein Maß. Beispiele hierfür sind aus der Vorlesung *Stochastik* wohlbekannt. Dort war etwa mit $\mathcal{F} = 2^{\mathbb{N}_0}$ und δ_k wie in 2.

$$\mu_{\mathrm{Poi}(\gamma)} := \sum_{k=0}^{\infty} e^{-\gamma} \frac{\gamma^k}{k!} \cdot \delta_k$$

die Poisson-Verteilung auf $2^{\mathbb{N}_0}$ mit Parameter γ,

$$\mu_{\mathrm{geo}(p)} := \sum_{k=1}^{\infty} (1-p)^{k-1} p \cdot \delta_k$$

die geometrische Verteilung mit Erfolgsparameter p und

$$\mu_{B(n,p)} := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot \delta_k$$

die Binomialverteilung $B(n,p)$.

Bemerkung 3.3 (Inhalte und Prämaße). Endlich additive Mengenfunktionen heißen oft *Inhalt*, σ-additive Mengenfunktionen, die nicht auf σ-Algebren definiert sind, heißen oft Prämaße. Die auf einer Borel'schen σ-Algebra definierten, bezüglich der kompakten Mengen von innen regulären Maße heißen Radon-Maße. Wir werden diese Begriffe nicht verwenden.

Lemma 3.4 (Mengenfunktionen auf Halbringen). *Sei $\mathcal{H}$ ein Halbring und $\mu : \mathcal{H} \to \overline{\mathbb{R}}_+$ endlich additiv. Dann ist μ monoton und subadditiv. Außerdem ist μ genau dann σ-additiv wenn es σ-subadditiv ist.*

Beweis. Wir beginnen mit einer Argumentation, die wir häufiger brauchen werden. Für $A_1, A_2, \ldots \in \mathcal{H}$ und $B_n = A_n \setminus \Big(\bigcup_{i=1}^{n-1} A_i \Big)$ und $n \geq 2$ gibt es $D_1, \ldots, D_k \in \mathcal{H}$ disjunkt mit $B_n = \biguplus_{i=1}^{k} D_i$. Die Behauptung ist klar für $n = 2$, da $\mathcal{H}$ ein Halbring ist. Gilt die Behauptung für ein n, so ist $B_{n+1} = A_{n+1} \setminus \Big(\bigcup_{i=1}^{n} A_i \Big) = \Big(A_{n+1} \setminus \Big(\bigcup_{i=2}^{n} A_i \Big) \Big) \setminus A_1$. Nach Voraussetzung gibt es $D_1, \ldots, D_k$ mit $A_{n+1} \setminus \Big(\bigcup_{i=2}^{n} A_i \Big) = \biguplus_{i=1}^{k} D_i$. Damit ist $B_{n+1} = \Big(\biguplus_{i=1}^{k} D_i \Big) \cap A_1^c = \biguplus_{i=1}^{k} (D_i \setminus A_1)$. Für alle $i = 1, \ldots, k$ gibt es Mengen $E_{i1}, \ldots, E_{k_i} \in \mathcal{H}$ mit $D_i \setminus A_1 = \biguplus_{j=1}^{k_i} E_{ij}$. Damit gilt $B_{n+1} = \biguplus_{i=1}^{k} \biguplus_{j=1}^{k_i} E_{ij}$ und die Behauptung ist gezeigt.

Wir zeigen nun, dass μ monoton ist. Sei $A, B \in \mathcal{H}$ mit $A \subseteq B$. Dann gibt es disjunkte $D_1, \ldots, D_k \in \mathcal{H}$ mit $B \setminus A = \biguplus_{i=1}^{k} D_i$, also $B = A \uplus \biguplus_{i=1}^{k} D_i$ und damit wegen der Additivität von μ auch $\mu(B) = \mu(A) + \sum_{i=1}^{k} \mu(D_i) \geq \mu(A)$. Genauso zeigt man, dass aus $E_1, \ldots, E_k \in \mathcal{H}$ disjunkt mit $\biguplus_{i=1}^{k} E_i \subseteq A \in \mathcal{H}$ folgt, dass $\sum_{i=1}^{k} \mu(E_i) \leq \mu(A)$.

Für die Subadditivität seien $A_1, \ldots, A_n \in \mathcal{H}$ mit $\bigcup_{i=1}^{n} A_i \in \mathcal{H}$. Wir setzen $B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2), \ldots$ Nach oben Gesagtem gibt es für jedes $i = 1, \ldots, k$ ein $k_i \in \mathbb{N}$ und Mengen $D_{ij} \in \mathcal{H}$, $j = 1, \ldots, k_i$ mit $B_i = \biguplus_{j=1}^{k_i} D_{ij}$. Damit gilt $\bigcup_{i=1}^{n} A_i = \biguplus_{i=1}^{n} B_i = \biguplus_{i=1}^{n} \biguplus_{j=1}^{k_i} D_{ij}$. Weiter ist $A_i = B_i \uplus \bigcup_{j=1}^{i-1} A_j$, also

$$\mu\Big(\bigcup_{i=1}^{n} A_i \Big) = \sum_{i=1}^{n} \sum_{j=1}^{k_i} \mu(D_{ij}) \leq \sum_{i=1}^{n} \mu(A_i).$$

Wir zeigen nun μ ist σ-additiv $\iff$ μ ist σ-sub-additiv.
'$\Rightarrow$' folgt genauso wie die Subadditivität von μ. Für Für '$\Leftarrow$' sei $A_1, A_2, \ldots \in \mathcal{H}$ disjunkt mit $A = \bigcup_{n=1}^\infty A_n \in \mathcal{H}$. Da μ monoton ist, gilt

$$\sum_{n=1}^\infty \mu(A_n) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n) = \lim_{N\to\infty} \mu\Big(\biguplus_{n=1}^N A_n\Big) \le \mu(A) \le \sum_{n=1}^\infty \mu(A_n)$$

wegen der σ-Subadditivität. □

Lemma 3.5 (Fortsetzung von Mengenfunktionen auf Halbringen). *Sei $\mathcal{H}$ ein Halbring, $\mathcal{R}$ der von $\mathcal{H}$ erzeugte Ring aus Lemma 2.5 und μ eine endlich additive Funktion auf $\mathcal{H}$. Definiere die Mengenfunktion $\widetilde{\mu}$ auf $\mathcal{R}$ durch*

$$\widetilde{\mu}\Big(\biguplus_{i=1}^n A_i\Big) := \sum_{i=1}^n \mu(A_i)$$

für $A_1, \ldots, A_n \in \mathcal{H}$ disjunkt. Dann ist $\widetilde{\mu}$ die einzige additive Fortsetzung von μ auf $\mathcal{R}$, die auf $\mathcal{H}$ mit μ übereinstimmt. Außerdem ist $\widetilde{\mu}$ genau dann σ-additiv, wenn μ σ-additiv ist.

Beweis. Es ist nur zu zeigen, dass $\widetilde{\mu}$ wohldefiniert ist. Sei hierzu $A_1, \ldots, A_m, B_1, \ldots, B_n \in \mathcal{H}$ mit $\biguplus_{i=1}^m A_i = \biguplus_{j=1}^n B_j$. Da

$$A_i = \biguplus_{j=1}^n A_i \cap B_j, \qquad B_j = \biguplus_{i=1}^m A_i \cap B_j,$$

gilt wegen der endlichen Additivität von $\widetilde{\mu}$

$$\sum_{i=1}^m \mu(A_i) = \sum_{i=1}^m \sum_{j=1}^n \mu(A_i \cap B_j) = \sum_{j=1}^n \sum_{i=1}^m \mu(A_i \cap B_j) = \sum_{j=1}^n \mu(B_j).$$ □

Proposition 3.6 (Einschluss-/Ausschlussformel). *Sei μ eine additive Mengenfunktion auf einem Ring $\mathcal{R}$ und I endlich. Dann gilt für $A_i \in \mathcal{R}$, $i \in I$, dass*

$$\mu\Big(\bigcup_{i\in I} A_i\Big) = \sum_{J\subseteq I} (-1)^{|J|+1} \mu\Big(\bigcap_{j\in J} A_j\Big)$$

Für $I = \{1,2\}$ gilt also

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2)$$

und für $I = \{1,2,3\}$

$$\begin{aligned}\mu(A_1 \cup A_2 \cup A_3) = {} & \mu(A_1) + \mu(A_2) + \mu(A_3) \\ & - \mu(A_1 \cap A_2) - \mu(A_1 \cap A_3) - \mu(A_2 \cap A_3) + \mu(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Beweis. Wir verwenden Induktion über $|I|$. Für $|I| = 2$ ist die Behauptung wegen $A_1 \cup A_2 = A_1 \uplus (A_2 \setminus A_1)$ und $(A_2 \setminus A_1) \uplus (A_1 \cap A_2) = A_2$ klar. Gilt sie für $I = \{1, \dots, n\}$, so ist wegen der Additivität von μ

$$\begin{aligned}
\mu\Big(\bigcup_{i=1}^{n+1} A_i\Big) &= \mu\Big(\bigcup_{i=1}^{n} (A_i \cup A_{n+1})\Big) \\
&= \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \mu\Big(A_{n+1} \cup \bigcap_{j \in J} A_j)\Big) \\
&= \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \Big(\mu(A_{n+1}) + \mu\Big(\bigcap_{j \in J} A_j)\Big) - \mu\Big(\bigcap_{j \in J} A_j \cap A_{n+1})\Big)\Big) \\
&= \mu(A_{n+1}) + \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \Big(\mu\Big(\bigcap_{j \in J} A_j)\Big) - \mu\Big(\bigcap_{j \in J} A_j \cap A_{n+1})\Big)\Big) \\
&= \sum_{J \subseteq \{1,\dots,n+1\}} (-1)^{|J|+1} \mu\Big(\bigcap_{j \in J} A_j)\Big).
\end{aligned}$$

□

3.2 σ-Additivität

Die endliche Additivität von Mengenfunktionen ist eine Forderung, die man oft nachprüfen kann. Anders sieht es mit der σ-Additivität aus. Wir werden nun alternative Formulierungen für σ-Additivität kennen lernen.

Proposition 3.7 (Stetigkeit von unten und von oben). *Sei $\mathcal{R}$ ein Ring und $\mu : \mathcal{R} \to \overline{\mathbb{R}}_+$ endlich additiv. Betrachte folgende Eigenschaften:*

1. *μ ist σ-additiv*
2. *μ ist σ-subadditiv*
3. *μ ist stetig von unten, d.h. für $A, A_1, A_2, \dots \in \mathcal{R}$ und $A_1 \subseteq A_2 \subseteq \dots$ mit $A = \bigcup_{n=1}^\infty A_n$ ist $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*
4. *μ ist stetig von oben in $\emptyset$, d.h. für $A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ und $A_1 \supseteq A_2 \supseteq \dots$ mit $\bigcap_{n=1}^\infty A_n = \emptyset$ ist $\lim_{n\to\infty} \mu(A_n) = 0$.*
5. *μ ist stetig von oben, d.h. für $A, A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ und $A_1 \supseteq A_2 \supseteq \dots$ mit $A = \bigcap_{n=1}^\infty A_n$ ist $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*

Dann gilt

$$1. \iff 2. \iff 3. \Longrightarrow 4. \iff 5..$$

Außerdem gilt 4. $\Longrightarrow$ 3., falls $\mu(A) < \infty$ für alle $A \in \mathcal{R}$.

Beweis. 1.⇔2. folgt aus Lemma 3.5, da $\mathcal{R}$ ein Halbring ist.

1.⇒3.: Sei μ also σ-additiv und $A, A_1, A_2, \dots \in \mathcal{R}$ wie in 3. Dann gilt mit $A_0 = \emptyset$

$$\mu(A) = \sum_{n=1}^\infty \mu(A_n \setminus A_{n-1}) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n \setminus A_{n-1}) = \lim_{N\to\infty} \mu(A_N).$$

3.⇒1.: Sei $B_1, B_2, \dots \in \mathcal{R}$ paarweise disjunkt und $B = \biguplus_{n=1}^\infty B_n \in \mathcal{R}$. Dann gilt für $A_N = \biguplus_{n=1}^N B_n$

$$\mu(B) = \lim_{N\to\infty} \mu(A_N) = \sum_{n=1}^\infty \mu(B_n).$$

4.⇒5.: Sei $A, A_1, A_2, \dots \in \mathcal{R}$ wie in 5. vorausgesetzt. Weiter sei $B_n := A_n \setminus A$. Dann erfüllt $B_1, B_2, \dots$ die Voraussetzungen von 4., also gilt $\mu(B_n) \xrightarrow{n\to\infty} 0$, also $\mu(A_n) = \mu(B_n) + \mu(A) \xrightarrow{n\to\infty} \mu(A)$.

5.⇒4.: ist klar.

3.⇒4.: Sei $A_1, A_2, \dots \in \mathcal{R}$ wie in 4. vorausgesetzt. Setze $B_n := A_1 \setminus A_n, n \in \mathbb{N}$. Dann erfüllt $B = A_1, B_1, B_2, \dots \in \mathcal{R}$ die Voraussetzungen von 3., und damit gilt $\mu(A_1) = \lim_{n\to\infty} \mu(B_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n)$, woraus 4. folgt.

4.⇒3. falls $\mu(A) < \infty$ für alle $A \in \mathcal{R}$. Sei $A, A_1, A_2, \dots \in \mathcal{R}$ wie in 3. vorausgesetzt. Setze $B_n := A \setminus A_n \in \mathcal{R}, n \in \mathbb{N}$. Dann gilt $\bigcap_{n=1}^\infty B_n = \emptyset$, also $0 = \lim_{n\to\infty} \mu(B_n) = \mu(A) - \lim_{n\to\infty} \mu(A_n)$, woraus 3. folgt. Hier geht in der letzten Gleichheit die Voraussetzung ein, dass $\mu(A) < \infty$. □

Wir wollen nun von innen bezüglich eines kompakten Systems reguläre Mengenfunktionen näher beleuchten. In vielen Situationen ist die Forderung, dass ein Maß bezüglich eines kompakten Systems von innen regulär ist, erfüllt. Diese Tatsache, die durch das folgende Lemma gezeigt wird, wird für die Theorie schwacher Konvergenz eine wichtige Rolle spielen.

Lemma 3.8. *Ist $(\Omega, \mathcal{O})$ polnisch und μ ein endliches Maß auf $\mathcal{B}(\mathcal{O})$, so existiert zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subseteq \Omega$ mit $\mu(\Omega \setminus K) < \varepsilon$.*

Beweis. Zuerst bemerken wir, dass alle kompakten Mengen nach Lemma 1.8 auch abgeschlossen sind, also sind alle kompakten Mengen in $\mathcal{B}(\mathcal{O})$ und damit ist $\mu(\Omega \setminus K)$ im Lemma wohldefiniert.

Sei $\varepsilon > 0$. Da Ω separabel ist, gibt es für jedes $n = 1, 2, \dots$ Punkte $\omega_1^n, \omega_2^n, \dots \in \Omega$ mit $\Omega = \bigcup_{k=1}^\infty B_{1/n}(\omega_k^n)$. Da μ von oben stetig ist (Proposition 3.7), gilt

$$0 = \mu\Big(\Omega \setminus \bigcup_{k=1}^\infty B_{1/n}(\omega_k^n)\Big) = \lim_{N\to\infty} \mu\Big(\Omega \setminus \bigcup_{k=1}^N B_{1/n}(\omega_k^n)\Big).$$

Damit gibt es ein $N_n \in \mathbb{N}$ mit $\mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\Big) < \varepsilon/2^n$. Nun ist

$$A := \bigcap_{n=1}^\infty \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)$$

nach Definition total beschränkt, also relativ kompakt nach Lemma 1.9. Außerdem gilt für die kompakte Menge $\overline{A}$

$$\mu(\Omega \setminus \overline{A}) \le \mu(\Omega \setminus A) \le \mu\Big(\bigcup_{n=1}^\infty \Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\Big)\Big) \le \sum_{n=1}^\infty \mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\Big) < \varepsilon.$$

Damit ist die Behauptung gezeigt. □

Theorem 3.9 (Von innen reguläre additive Mengenfunktionen sind σ-additiv). *Sei $\mathcal{H}$ ein Halbring und $\mu : \mathcal{H} \to \mathbb{R}_+$ endlich, endlich additiv und für ein kompaktes System $\mathcal{K} \subseteq \mathcal{H}$ von innen regulär. Dann ist μ auch σ-additiv.*

Beweis. Wie in Lemma 3.5 kann man die Mengenfunktion μ auf eindeutige Weise auf den von $\mathcal{H}$ erzeugten Ring $\mathcal{R}(\mathcal{H})$ (siehe Lemma 2.5) erweitern. Weiter ist nach Lemma 2.16 das System $\mathcal{K}_\cup \subseteq \mathcal{R}(\mathcal{H})$, das durch Vereinigungsbildung aus Mengen von $\mathcal{K}$ besteht, ebenfalls kompakt. Wählt man nun $\varepsilon > 0$ und $A = \bigcup_{i=1}^n A_i \in \mathcal{R}(\mathcal{H})$ mit $A_1, \dots, A_n \in \mathcal{H}$, so gibt es kompakte Mengen $K_1, \dots, K_n \in \mathcal{K} \subseteq \mathcal{H}$ mit $\mu(A_i) \le \mu(K_i) + \frac{\varepsilon}{n}$ für $i = 1, \dots, n$. Damit gilt für die Erweiterung von μ auf den Ring $\mathcal{R}(\mathcal{H})$

$$\mu\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n \mu(A_i) \le \Big(\sum_{i=1}^n \mu(K_i)\Big) + \varepsilon = \mu\Big(\bigcup_{i=1}^n K_i\Big) + \varepsilon.$$

Damit ist μ von innen $\mathcal{K}_\cup$-regulär. Also ist o.E. der Halbring $\mathcal{H}$ ein Ring und $\mathcal{K}$ vereinigungsstabil.

Wir zeigen nun, dass μ stetig von oben in $\emptyset$ ist. Dies genügt nach Proposition 3.7 wegen der Endlichkeit von μ auf $\mathcal{H}$. Seien $A_1, A_2, \dots \in \mathcal{H}$ mit $A_1 \supseteq A_2 \supseteq \cdots$ und $\bigcap_{n=1}^\infty A_n = \emptyset$ und $\varepsilon > 0$. Wähle $K_1, K_2, \dots \in \mathcal{K}$ mit $K_n \subseteq A_n, n \in \mathbb{N}$ und

$$\mu(A_n) \le \mu(K_n) + \varepsilon 2^{-n}.$$

Es gilt $\bigcap_{n=1}^\infty K_n \subseteq \bigcap_{n=1}^\infty A_n = \emptyset$. Deswegen gibt es ein $N \in \mathbb{N}$ mit $\bigcap_{n=1}^N K_n = \emptyset$. Damit gilt

$$A_N = A_N \cap \Big(\bigcup_{n=1}^N K_n^c\Big) = \bigcup_{n=1}^N A_N \setminus K_n \subseteq \bigcup_{n=1}^N A_n \setminus K_n.$$

Daraus folgt wegen der Subadditivität und der Monotonie von μ für alle $m \ge N$

$$\mu(A_m) \le \mu(A_N) \le \sum_{n=1}^N \mu(A_n \setminus K_n) \le \varepsilon \sum_{n=1}^N 2^{-n} \le \varepsilon.$$

Damit ist die Behauptung gezeigt, da $\varepsilon > 0$ beliebig war. □

3.3 Eindeutigkeit und Fortsetzung von Maßen

Angenommen, eine additive Mengenfunktion $\mu : \mathcal{H} \to \overline{\mathbb{R}}_+$ ist gegeben. Wir beschäftigen uns damit, wann es möglich ist, diese Mengenfunktion zu einem Maß (d.h. einer σ-additiven Mengenfunktion) auf $\sigma(\mathcal{H})$ auszuweiten. Ziel ist es dabei, Bedingungen aufzustellen, wann das Maß durch μ schon eindeutig gegeben ist. Das Resultat ist in Theorem 3.15 zusammen gefasst. Siehe auch Tabelle 3.1 für eine Übersicht, wie die Resultate vergangener Kapitel damit zusammen hängen.

Proposition 3.10 (Eindeutigkeit von Maßen). *Sei $\mathcal{F}$ eine σ-Algebra und $\mu, \nu : \mathcal{F} \to \mathbb{R}_+$ Maße. Sei $\mathcal{C}$ ein durchschnittstabiles Mengensystem, so dass $\sigma(\mathcal{C}) = \mathcal{F}$ und $\mu|_\mathcal{C}, \nu|_\mathcal{C}$ sind σ-endlich. Dann ist $\mu = \nu$ genau dann, wenn $\mu(A) = \nu(A)$ für alle $A \in \mathcal{C}$ gilt.*

	Lemma 3.4	Theorem 3.9	Theorem 3.15
μ endlich additiv	$\circ$	$\circ$	
μ endlich		$\circ$	
μ σ-endlich			$\circ$
μ definiert auf Halbring	$\circ$	$\circ$	$\circ$
μ σ-additiv	$\circ/\bullet$	$\bullet$	$\circ$
μ σ-subadditiv	$\bullet/\circ$		
μ von innen $\mathcal{K}$-regulär		$\circ$	
μ eindeutig auf $\sigma(\mathcal{H})$ fortsetzbar			$\bullet$

Tabelle 3.1: Um den Carathéodory'schen Fortsetzungssatz anwenden zu können, spielen Lemma 3.4 und Theorem 3.9 eine Rolle. In der Tabelle bedeuten die $\circ$'s jeweils die Voraussetzungen des Satzes und $\bullet$ die Folgerungen. Wie man leicht ablesen kann, gilt der Carathéodory'sche Fortsetzungssatz z.B. dann, wenn μ endlich und von innen $\mathcal{K}$-regulär ist.

Korollar 3.11 (Eindeutigkeit von Wahrscheinlichkeitsmaßen). *Sei $\mathcal{F}$ eine σ-Algebra und $\mu, \nu : \mathcal{F} \to [0,1]$ seien Wahrscheinlichkeitsmaße. Sei $\mathcal{C}$ ein durchschnittsstabiles Mengensystem mit $\sigma(\mathcal{C}) = \mathcal{F}$. Dann ist $\mu = \nu$ genau dann, wenn $\mu(A) = \nu(A)$ für alle $A \in \mathcal{C}$ gilt.*

Beweis. O.E. ist $\Omega \in \mathcal{C}$, da $\mu(\Omega) = \nu(\Omega) = 1$ gilt. Damit sind μ und ν insbesondere σ-endlich und die Aussage folgt aus Proposition 3.10. $\square$

Beweis von Proposition 3.10. Die 'genau dann'-Richtung ist klar. Für die 'wenn'-Richtung setzen wir für ein $C \in \mathcal{C}$ mit $\mu(C) = \nu(C) < \infty$

$$\mathcal{D}_C := \{A \in \mathcal{F} : \mu(A \cap C) = \nu(A \cap C)\} \supseteq \mathcal{C}.$$

Wir zeigen, dass $\mathcal{D}_C$ ein Dynkin System ist. Klar ist, dass $\Omega \in \mathcal{D}_C$. Ist weiter $A, B \in \mathcal{D}$ und $A \subseteq B$, so ist $\mu((B \setminus A) \cap C) = \mu(B \cap C) - \mu(A \cap C) = \nu(B \cap C) - \nu(A \cap C) = \nu((B \setminus A) \cap C)$, also $B \setminus A \in \mathcal{D}_C$. Ist $A_1, A_2, \dots \in \mathcal{D}$ mit $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \in \mathcal{D}_C$ und $A = \bigcup_{n=1}^\infty A_n \in \mathcal{F}$, so ist wegen Proposition 3.7.

$$\mu(A \cap C) = \lim_{n\to\infty} \mu(A_n \cap C) = \lim_{n\to\infty} \nu(A_n \cap C) = \nu(A \cap C),$$

also $A \in \mathcal{D}_C$. Damit ist $\mathcal{D}_C$ für alle $C \in \mathcal{C}$ mit $\mu(C) < \infty$ ein Dynkin-System und damit gilt wegen Theorem 2.13, dass $\mathcal{F} = \sigma(\mathcal{C}) = \mathcal{D}_C$. Sei $\Omega_1, \Omega_2, \dots \in \mathcal{C}$ mit $\Omega_n \uparrow \Omega$ und $\mu(\Omega_n), \nu(\Omega_n) < \infty, n = 1, 2, \dots$ Dann gilt für alle $n = 1, 2, \dots$, dass $\mu(A \cap \Omega_n) = \nu(A \cap \Omega_n)$ für alle $A \in \mathcal{F}$. Daraus folgt dann für $A \in \mathcal{F}$, da μ und ν stetig von unten sind,

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \nu(A \cap \Omega_n) = \nu(A),$$

d.h. $\mu = \nu$. $\square$

Das folgende Theorem erklärt, warum der Begriff der σ-Algebra zentral ist.

Theorem 3.12 (μ^*-messbare Mengen sind eine σ-Algebra). *Sei μ^* ein äußeres Maß auf Ω und $\mathcal{F}^*$ die Menge der μ^*-messbaren Mengen. Dann ist $\mathcal{F}^*$ eine σ-Algebra und $\mu := \mu^*|_{\mathcal{F}^*}$ ein Maß. Außerdem ist $\mathcal{N} := \{N \subseteq \Omega : \mu^*(N) = 0\} \subseteq \mathcal{F}^*$.*

Bemerkung 3.13 (Nullmengen und Eigenschaften fast überall). Mengen N mit $\mu(N) = 0$ heißen auch *(μ-)Nullmengen*. Weiter sagen wir, dass $A \subseteq \Omega$ (μ)-*fast überall* gilt, wenn $A^c \subseteq N$ und N eine μ-Nullmenge ist. Ist μ ein Wahrscheinlichkeitsmaß, sagen wir anstatt 'fast überall' auch *fast sicher*.

Beweis von Theorem 3.12. Wir zeigen zunächst, dass $\mathcal{F}^*$ eine σ-Algebra ist. Klar ist, dass

$$\mu^*(E) = \mu^*(\emptyset) + \mu^*(E) = \mu^*(E \cap \emptyset) + \mu^*(E \cap \Omega),$$

d.h. $\emptyset \in \mathcal{F}^*$. Weiter ist klar, dass aus $A \in \mathcal{F}^*$ auch $A^c \in \mathcal{F}^*$ folgt. Mit $A, B \in \mathcal{F}^*$ ist wegen $(E \cap A \cap B^c) \uplus (E \cap A^c) = E \cap (A \cap B)^c$ und der Subadditivität von μ^*

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*((E \cap A) \cap B) + \mu^*((E \cap A) \cap B^c) + \mu^*(E \cap A^c)\\ &\geq \mu^*(E \cap (A \cap B)) + \mu^*(E \cap (A \cap B)^c) \geq \mu^*(E).\end{aligned}$$

Damit ist $A \cap B \in \mathcal{F}^*$. Seien nun $A_1, A_2, \dots \in \mathcal{F}^*$ disjunkt und $B_n = \biguplus_{k=1}^n A_k$ und $B = \bigcup_{n=1}^\infty B_n = \biguplus_{k=1}^\infty A_k$. Da $\mathcal{F}^*$ durchschnitt- und komplementstabil ist, ist auch $B_1, B_2, \dots \in \mathcal{F}^*$. Weiter zeigen wir, dass $\mu^*(E \cap B_n) = \sum_{k=1}^n \mu^*(E \cap A_k)$ für alle $E \subseteq \Omega$ gilt. Für $n = 1$ ist dies klar, und gilt es für ein n, so folgt, da $B_n \in \mathcal{F}^*$,

$$\begin{aligned}\mu^*(E \cap B_{n+1}) &= \mu^*(E \cap B_{n+1} \cap B_n) + \mu^*(E \cap B_{n+1} \cap B_n^c)\\ &= \mu^*(E \cap B_n) + \mu^*(E \cap A_{n+1}) = \sum_{k=1}^{n+1} \mu^*(E \cap A_k).\end{aligned}$$

Damit gilt wegen der Monotonie von μ^*

$$\mu^*(E \cap B) \leq \sum_{k=1}^\infty \mu^*(E \cap A_k) = \lim_{n\to\infty} \sum_{k=1}^n \mu^*(E \cap A_k) = \lim_{n\to\infty} \mu^*(E \cap B_n) \leq \mu^*(E \cap B),$$

also

$$\mu^*(E \cap B) = \lim_{n\to\infty} \mu^*(E \cap B_n) = \sum_{k=1}^\infty \mu^*(E \cap A_k). \tag{3.4}$$

Als nächstes zeigen wir, dass $B \in \mathcal{F}^*$ ist, woraus folgt, dass $\mathcal{F}^*$ eine σ-Algebra ist. Es gilt nämlich für $E \subseteq \Omega$, wegen (3.4) und der Subadditivität von μ^*

$$\mu^*(E) = \lim_{n\to\infty} \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E),$$

woraus $B \in \mathcal{F}^*$ folgt. Weiter folgt aus (3.4), dass μ^* auf $\mathcal{F}^*$ sogar σ-additiv ist, d.h. dass $\mu = \mu^*|_{\mathcal{F}^*}$ ein Maß ist.

Sei nun $N \subseteq \Omega$ so, dass $\mu^*(N) = 0$ und $E \subseteq \Omega$. Dann ist wegen der Monotonie von μ^* auch $\mu^*(E \cap N) = 0$, also

$$\mu^*(E) \geq \mu^*(E \cap N^c) = \mu^*(E \cap N^c) + \mu^*(E \cap N)$$

und damit $N \in \mathcal{F}^*$. $\square$

Proposition 3.14 (Von endlich additiver Mengenfunktion erzeugtes äußeres Maß). *Sei $\mathcal{H}$ ein Halbring und $\mu : \mathcal{H} \to \mathbb{R}_+$ endlich additiv. Für $A \subseteq \Omega$ sei*

$$\mu^*(A) := \inf_{\mathcal{G} \in \mathcal{U}(A)} \sum_{G \in \mathcal{G}} \mu(G)$$

wobei

$$\mathcal{U}(A) := \big\{\mathcal{G} \subseteq \mathcal{H} \text{ höchstens abzählbar, } A \subseteq \bigcup_{G \in \mathcal{G}} G\big\}$$

die Menge der höchstens abzählbaren Überdeckungen von A ist und $\mu^*(A) = \infty$ falls $\mathcal{U}(A) = \emptyset$. Dann ist μ^* ein äußeres Maß.

Beweis. Die Abbildung μ^* ist monoton, und da $\emptyset \in \mathcal{H}$ ist $\mu^*(\emptyset) = 0$ wegen der endlichen Additivität von μ; siehe Lemma 3.4. (Es ist nämlich $\mu(\emptyset) = \mu(\emptyset) + \mu(\emptyset)$, woraus $\mu(\emptyset) = 0$ folgt.) Um die σ-Subadditivität von μ^* zu prüfen, wählen wir $A_1, A_2, \dots \subseteq \Omega$. Für $n = 1, 2, \dots$ und $\epsilon > 0$ gibt es Mengen $G_{nk} \in \mathcal{H}$, $k \in \mathcal{K}_n$ höchstens abzählbar mit

$$A_n \subseteq \bigcup_{k \in \mathcal{K}_n} G_{nk},$$
$$\mu^*(A_n) \geq \sum_{k \in \mathcal{K}_n} \mu(G_{nk}) - \varepsilon 2^{-n}.$$

Daraus folgt wegen $\bigcup_{n=1}^\infty A_n \subseteq \bigcup_{n=1}^\infty \bigcup_{k \in \mathcal{K}_n} G_{nk}$, der Monotonie von μ^* und der Definition von μ^*

$$\mu^*\Big(\bigcup_{n=1}^\infty A_n\Big) \leq \sum_{n=1}^\infty \sum_{k \in \mathcal{K}_n} \mu(G_{nk}) \leq \varepsilon + \sum_{n=1}^\infty \mu^*(A_n).$$

Mit $\varepsilon \to 0$ folgt die σ-Subadditivität von μ^*, d.h. μ^* ist ein äußeres Maß. □

Theorem 3.15 (Fortsetzung einer $\boldsymbol{\sigma}$-additiven Mengenfunktion). *Sei $\mathcal{H}$ ein Halbring und $\mu : \mathcal{H} \to \mathbb{R}_+$ σ-endlich und σ-additiv. Weiter sei $\widetilde{\mu} = \mu^*|_{\mathcal{F}^*}$ mit μ^* aus Proposition 3.14 und $\mathcal{F}^*$ aus Theorem 3.12. Dann ist $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ und $\widetilde{\mu}|_{\sigma(\mathcal{H})}$ ist das einzige Maß, das auf $\mathcal{H}$ mit μ übereinstimmt.*

Beweis. Zunächst stellen wir fest, dass μ sowohl endlich additiv als auch σ-subadditiv ist nach Lemma 3.4. Nach Proposition 3.14 ist μ^* ein äußeres Maß und nach Theorem 3.12 ist $\mathcal{F}^*$ eine σ-Algebra.

Schritt 1: μ^* stimmt auf $\mathcal{H}$ mit μ überein: Sei $H \in \mathcal{H}$. Wähle $\mathcal{K}$ höchstens abzählbar und $H_k \in \mathcal{H}$, $k \in \mathcal{K}$ mit $H \subseteq \bigcup_{k \in \mathcal{K}} H_k$ und

$$\mu^*(H) \geq \sum_{k \in \mathcal{K}} \mu(H_k) - \varepsilon.$$

Damit gilt wegen $H = \bigcup_{k \in \mathcal{K}} H_k \cap H$ und der σ-Subadditivität von μ

$$\mu^*(H) \leq \mu(H) \leq \sum_{k \in \mathcal{K}} \mu(H_k \cap H) \leq \sum_{k \in \mathcal{K}} \mu(H_k) \leq \mu^*(H) + \varepsilon,$$

wobei wir im zweiten '$\leq$' die endliche Additivität von μ verwendet haben. Mit $\varepsilon \to 0$ folgt, dass $\mu^*(H) = \mu(H)$.

Schritt 2: $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$: Sei $E \subseteq \Omega, H \in \mathcal{H}$ und $\varepsilon > 0$. Wähle $\mathcal{K}$ höchstens abzählbar und $H_k \in \mathcal{H}, k \in \mathcal{K}$ mit $\mu^*(E) \geq \sum_{k\in\mathcal{K}} \mu(H_k) - \varepsilon$. Dann gilt wegen der endlichen Additivität von μ

$$\mu^*(E) + \varepsilon \geq \sum_{k\in\mathcal{K}} \mu(H_k) = \sum_{k\in\mathcal{K}} \mu(H_k \cap H) + \sum_{k\in\mathcal{K}} \mu(H_k \cap H^c) \geq \mu^*(E \cap H) + \mu^*(E \cap H^c).$$

Mit $\varepsilon \to 0$ und der σ-Subadditivität von μ^* folgt $\mu^*(E) = \mu^*(E \cap H) + \mu^*(E \cap H^c)$, d.h. H ist μ^*-messbar und damit $\mathcal{H} \subseteq \mathcal{F}^*$. Da $\mathcal{F}^*$ nach Theorem 3.12 eine σ-Algebra ist, folgt auch $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$.

Schritt 3: Eindeutigkeit: Nach Theorem 3.12 ist $\widetilde{\mu}$ ein Maß, also auch $\widetilde{\mu}|_{\sigma(\mathcal{H})} = \mu^*|_{\sigma(\mathcal{H})}$. Sei $\nu : \sigma(\mathcal{H}) \to \mathbb{R}_+$ ein weiteres Maß, das auf $\mathcal{H}$ mit μ übereinstimmt. Da $\mu = \widetilde{\mu}|_{\mathcal{H}}$ als σ-endlich vorausgesetzt war, ist auch $\nu|_{\mathcal{H}}$ σ-endlich. Mit Proposition 3.10 folgt wegen der Durchschnittstabilität von $\mathcal{H}$, dass $\widetilde{\mu} = \nu$ auf $\sigma(\mathcal{H})$ gilt.

Nun sind alle Behauptungen bewiesen. □

In obigem Theorem wird nur klar, dass $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$. Das nächste Resultat zeigt, wodurch sich Mengen in $\mathcal{F}^*$ von Mengen in $\sigma(\mathcal{H})$ unterscheiden.

Proposition 3.16 (Charakterisierung von $\mathcal{F}^*$ aus Proposition 3.14). *Sei $\mathcal{H}$ ein Halbring, $\mu : \mathcal{H} \to \mathbb{R}_+$ σ-endlich und σ-additiv, μ^* wie in Proposition 3.14 und $\mathcal{F}^*, \mathcal{N}$ wie in Theorem 3.12. Dann ist*

$$\mathcal{F}^* = \{A \setminus N : A \in \sigma(\mathcal{H}), N \in \mathcal{N}\}.$$

Insbesondere ist die rechte Seite eine σ-Algebra.

Beweis. '$\supseteq$': Nach Theorem 3.15 ist $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ und nach Theorem 3.12 ist $\mathcal{N} \subseteq \mathcal{F}^*$. Daraus folgt bereits '$\supseteq$', da $\mathcal{F}^*$ komplementstabil ist.

'$\subseteq$': Sei $B \in \mathcal{F}^*$. Weiter seien $\Omega_1, \Omega_2, \dots \in \mathcal{H}$ mit $\mu(\Omega_n) < \infty, n = 1, 2, \dots$ und $\Omega = \bigcup_{n=1}^\infty \Omega_n$. Sei $\varepsilon_1, \varepsilon_2, \dots > 0$ mit $\varepsilon_i \downarrow 0$. Für $B_n := B \cap \Omega_n$ wählen wir $\mathcal{K}_{ni}$ höchstens abzählbar, $A_{nik} \in \mathcal{H}, n \in \mathbb{N}, k \in \mathcal{K}_{ni}$, $B_n \subseteq \bigcup_{k\in\mathcal{K}_{ni}} A_{nik}$ und

$$\mu^*(B_n) \geq \sum_{k\in\mathcal{K}_{ni}} \mu(A_{nik}) - 2^{-n}\varepsilon_i.$$

Klar ist $A_i := \bigcup_{n=1}^\infty \bigcup_{k\in\mathcal{K}_{ni}} A_{nik} \in \sigma(\mathcal{H})$, $B \subseteq A_i, i \in \mathbb{N}$ und $A_i \setminus B = \bigcup_{n=1}^\infty \bigcup_{k\in\mathcal{K}_{ni}} A_{nik} \setminus B_n$. Damit gilt

$$\mu^*(A_i \setminus B) \leq \sum_{n=1}^\infty 2^{-n}\varepsilon_i = \varepsilon_i.$$

Setze $A = \bigcap_{i=1}^\infty A_i \in \sigma(\mathcal{H})$. Dann ist $B \subseteq A$, $N := A \setminus B \subseteq A_n \setminus B, n \in \mathbb{N}$ und

$$\mu^*(N) = \mu^*(A \setminus B) \leq \limsup_{i\to\infty} \mu^*(A_i \setminus B) \leq \limsup_{i\to\infty} \varepsilon_i = 0.$$

Damit folgt die Behauptung, da $B = A \setminus N$. □

3.4 Maße auf $\mathcal{B}(\mathbb{R})$

Aus der Vorlesung *Stochastik* kennt man bereits Wahrscheinlichkeitsverteilungen mit Dichte. Diese sind Maße auf $\mathcal{B}(\mathbb{R})$. Wir werden hier die in den letzten Kapiteln erarbeitete allgemeine Theorie anwenden, um solche Maße zu charakterisieren.

Proposition 3.17 (Lebesgue-Maß auf $\mathbb{R}$). *Es gibt genau ein Maß λ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so dass*

$$\lambda((a,b]) = b - a \tag{3.5}$$

für $a, b \in \mathbb{Q}$ mit $a \leq b$.

Beweis. Nach Beispiel 2.3 ist $\mathcal{H} = \{(a,b] : a,b \in \mathbb{Q}, a \leq b\}$ ein Halbring. Genauso ist $\widetilde{H} := \{(a,b], [a,b), (a,b), [a,b] : a,b \in \mathbb{Q}, a \leq b\}$ ein Halbring mit $\sigma(\widetilde{\mathcal{H}}) = \mathcal{B}(\mathbb{R})$. Wir definieren $\widetilde{\lambda}$ auf $\widetilde{\mathcal{H}}$ durch

$$\widetilde{\lambda}((a,b]) = \widetilde{\lambda}([a,b)) = \widetilde{\lambda}((a,b)) = \widetilde{\lambda}([a,b]) = b - a.$$

Dann ist $\widetilde{\lambda}$ klar σ-endlich. Es ist $\mathcal{K} = \{[a,b] : a,b \in \mathbb{Q}, a \leq b\} \subseteq \widetilde{\mathcal{H}}$ ein kompaktes System nach Beispiel 2.3. Außerdem ist klar $\widetilde{\lambda}$ von innen $\mathcal{K}$-regulär und damit σ-additiv nach Theorem 3.9. Nach Theorem 3.15 gibt es damit genau eine Fortsetzung von $\widetilde{\lambda}$ auf $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$. □

Proposition 3.18 (Charakterisierung der $\boldsymbol{\sigma}$-endlichen Maße auf $\mathbb{R}$). *Eine Funktion $\mu : \mathcal{B}(\mathbb{R}) \to \overline{\mathbb{R}}_+$ ist genau dann ein σ-endliches Maß, wenn es eine nicht-fallende und rechtsstetige Funktion $G : \mathbb{R} \to \mathbb{R}$ gibt mit*

$$\mu((a,b]) = G(b) - G(a) \tag{3.6}$$

für $a, b \in \mathbb{Q}$ mit $a \leq b$. Ist $\widetilde{G}$ eine weitere Funktion $\widetilde{G} : \mathbb{R} \to \mathbb{R}$ für die (3.6) gilt, so ist $\widetilde{G} = G + c$ für ein $c \in \mathbb{R}$.

Korollar 3.19 (Charakterisierung der Wahrscheinlichkeitsmaße auf $\mathbb{R}$). *Eine Funktion $\mu : \mathcal{B}(\mathbb{R}) \to [0,1]$ ist genau dann ein Wahrscheinlichkeitsmaß, wenn es eine nicht-fallende und rechtsstetige Funktion $F : \mathbb{R} \to [0,1]$ gibt mit $\lim_{b\to\infty} F(b) = 1$ und*

$$\mu((a,b]) = F(b) - F(a) \tag{3.7}$$

für $a, b \in \mathbb{Q}$ mit $a \leq b$. In diesem Fall ist F eindeutig durch μ festgelegt.

Beweis. Die Behauptung folgt direkt aus Proposition 3.18, da der Bildbereich von F nur für eine rechtsstetige, nicht-fallende Funktion genau $[0,1]$ sein kann. □

Beweis von Proposition 3.18. '⇒': Sei also μ ein σ-endliches Maß auf $\mathcal{B}(\mathbb{R})$. Definiere dann $G(0) = 0$, sowie $G(x) := \mu((0,x])$ für $x > 0$ und $G(x) := \mu((x,0])$ für $x < 0$. Dann ist G rechtsstetig, nicht fallend, und (etwa für $0 < a < b$) $\mu((a,b]) = \mu((0,b]) - \mu((0,a]) = G(b) - G(a)$.

'⇐': Der Beweis verläuft ähnlich zu dem Beweis von Korollar 3.17. Sei $\mathcal{H} = \{(a,b] : a,b \in \mathbb{Q}, a \leq b\}$ der Halbring der halboffenen Intervalle mit Enden in rationalen Zahlen. Wir zeigen, dass durch (3.6) eine σ-additive Mengenfunktion μ auf $\mathcal{H}$ definiert wird. Dann sieht man mit Theorem 3.15, dass μ auf eindeutige Weise zu einem σ-endlichen Maß auf $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$ erweitert werden kann. Sei also $a_1, a_2, \ldots$ disjunkt so, dass $\bigcup_{n=1}^{\infty}(a_n, a_{n+1}] = (a,b] \in \mathcal{H}$.

Ohne Einschränkung ist $a_1 \geq a_2 \geq \dots$ Dann ist $b = a_1$ und $a_n \xrightarrow{n\to\infty} a$. Es gilt wegen der Rechtsstetigkeit von G, dass

$$\mu(a,b] = G(b) - G(a) = G(a_1) - \lim_{N\to\infty} G(a_N) = \sum_{n=1}^{\infty} G(a_n) - G(a_{n+1}) = \sum_{n=1}^{\infty} \mu((a_n, a_{n+1}]),$$

und damit ist die σ-Additivität von μ gezeigt.

Sei nun $\widetilde{G}$ eine weitere Funktion, für die (3.6) gilt. Dann gilt für alle $a \in \mathbb{R}$

$$\widetilde{G}(b) = \widetilde{G}(a) + \mu((a,b]) = G(b) + \widetilde{G}(a) - G(a)$$

und die Behauptung folgt mit $c = \widetilde{G}(a) - G(a)$. □

Definition 3.20 (Lebesgue-Maß und Verteilungsfunktion). *1. Das eindeutig definierte Maß λ aus Korollar 3.17 heißt ein-dimensionales Lebesgue-Maß.*

2. Für ein Wahrscheinlichkeitsmaß μ auf $\mathcal{B}(\mathbb{R})$ heißt die Funktion F aus Korollar 3.19 Verteilungsfunktion.

Beispiel 3.21 (Bekannte Verteilungsfunktionen). Sei $f : \mathbb{R} \to \mathbb{R}_+$ stückweise stetig, so dass[7] $\int_{-\infty}^{\infty} f(x)dx = 1$ gilt. Wie aus der Vorlesung *Stochastik* bekannt, heißt eine solche Funktion eine *Dichte*. Solche Dichtefunktionen definieren vermöge

$$F(x) := \int_{-\infty}^{x} f(a)da$$

eine Verteilungsfunktion. Auf der anderen Seite definiert jede dieser Verteilungsfunktionen wegen Korollar 3.19 auf eindeutige Art und Weise ein Wahrscheinlichkeitsmaß. Verteilungen mit Dichten werden wir genauer im Satz von Radon-Nikodým (siehe Abschnitt 5.4) beleuchten.

Wie bereits bekannt, ist

$$F_{U(0,1)}(x) = \int_{-\infty}^{x} 1_{[0,1]}(a)da = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x \leq 1, \\ 1, & x > 1 \end{cases} \tag{3.8}$$

die Verteilungsfunktion der Gleichverteilung auf $[0,1]$. Weiter ist für $x \geq 0$

$$F_{\exp(\lambda)}(x) = \int_{-\infty}^{x} 1_{[0,\infty)}(a)\lambda e^{-\lambda a} da = 1 - e^{-\lambda x} \tag{3.9}$$

die Verteilungsfunktion der Exponentialverteilung mit Parameter λ. Außerdem ist

$$F_{N(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\Big(- \frac{(a-\mu)^2}{2\sigma^2}\Big) da =: \Phi(x) \tag{3.10}$$

die Verteilungsfunktion der Normalverteilung $N(\mu, \sigma^2)$ mit dem Erwartungswert μ und der Varianz σ^2.

[7] Wir setzen hier das Riemann-Integral $\int_a^b f(x)dx$ als bekannt voraus. (Siehe auch Definition 4.21.) Einen weiteren Integralbegriff, das Lebesgue-Integral, werden wir in Kapitel 4 kennen lernen.

3.5 Bildmaße

Sei μ ein Maß auf $\mathcal{F}$. Wenn man nun den Grundraum vermöge einer Funktion $f : \Omega \to \Omega'$ transformiert, kann man auf Ω' ein zu der Transformation korrespondierendes Maß definieren, das sogenannte Bildmaß. Sei etwa $\Omega := [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ und $f : u \mapsto 1-u$. Wir werden dann sehen, dass das Bildmaß von $\mu_{U(0,1)}$ unter f wieder $\mu_{U(0,1)}$ ist. Wir erinnern zunächst an die Situation aus Beispiel 2.3.2 und definieren das Bildmaß.

Definition 3.22 (Bildmaß). *Ist $(\Omega, \mathcal{F}, \mu)$ ein Maßraum, $(\Omega', \mathcal{F}')$ ein Messraum und $f : \Omega \to \Omega'$, so dass $\sigma(f) \subseteq \mathcal{F}$ für $\sigma(f)$ aus* (2.1). *Dann definieren wir eine Mengenfunktion $f_*\mu : \mathcal{F}' \to \mathbb{R}_+$ durch*

$$f_*\mu(A') := \mu(f^{-1}(A')) = \mu(f \in A'), \qquad A' \in \mathcal{F}'.$$

Hier heißt $f_*\mu$ auch Bildmaß *von μ unter f.*

Ist μ ein Wahrscheinlichkeitsmaß, so heißt $f_*\mu$ auch Verteilung *von f.*

Bemerkung 3.23 (Messbare Funktionen). Ist $\sigma(f) \subseteq \mathcal{F}$ wie in obiger Definition, so sagen wir, dass f (nach $\mathcal{F}/\mathcal{F}'$) messbar ist. Dieses Konzept wird uns im nächsten Abschnitt beschäftigen.

Proposition 3.24 (Bildmaß ist ein Maß). *Sei $(\Omega, \mathcal{F}, \mu)$, $(\Omega', \mathcal{F}')$, $f : \Omega \to \Omega'$ und $f_*\mu$ wie in Definition 3.22. Dann ist $f_*\mu$ ein Maß auf $\mathcal{F}'$.*

Beweis. Seien $A'_1, A'_2, \dots \in \mathcal{F}'$ disjunkt, so gilt

$$f_*\mu\Big(\biguplus_{n=1}^\infty A'_n\Big) = \mu\Big(f^{-1}\Big(\biguplus_{n=1}^\infty A'_n\Big)\Big) = \mu\Big(\biguplus_{n=1}^\infty (f^{-1}(A'_n)\Big) = \sum_{n=1}^\infty \mu(f^{-1}(A'_n)) = \sum_{n=1}^\infty f_*\mu(A'_n).$$

Damit ist $f_*\mu$ also σ-additiv und die Behauptung ist gezeigt. □

Beispiel 3.25 (Bekannte Transformationen).

1. Für $\Omega = [0,1]$ ist $\{[0,b) : 0 \le b \le 1\}$ ein schnittstabiles Erzeugendensystem von $\mathcal{B}([0,1])$. Sei $\mu = \mu_{U(0,1)}$ die Gleichverteilung auf $[0,1]$ mit Verteilungsfunktion $F_{U(0,1)}$ aus (3.8) und $f : u \mapsto 1-u$. Dann ist $f_*\mu = \mu$, denn

$$f_*\mu([0,b)) = \mu(f^{-1}([0,b))) = \mu([1-b,1]) = F_{U(0,1)}(1) - F_{U(0,1)}(1-b) = b.$$

   Damit stimmen μ und $f_*\mu$ auf einem schnittstabilen Erzeuger überein und die Aussage folgt mit Proposition 3.10.

2. Sei $\Omega = \mathbb{R}$, $f_y : x \mapsto x + y$ für ein $y \in \mathbb{R}$ und λ das Lebesgue-Maß aus Korollar 3.17. Dann ist $(f_y)_*\lambda = \lambda$, denn

$$(f_y)_*\lambda([a,b]) = \lambda(f_y^{-1}([a,b])) = \lambda([a-y, b-y]) = b-a.$$

   Man sagt auch, dass das Lebesgue-Maß translationsinvariant ist.

3. Sei $\Omega = [0,1], \Omega' = \mathbb{R}_+$, jeweils versehen mit der Borel'schen σ-Algebra. Weiter sei $\mu = \mu_{U(0,1)}$ mit Verteilungsfunktion $F_{U(0,1)}$ aus (3.8) und $f : x \mapsto -\frac{1}{\lambda}\log(x)$ für ein $\lambda > 0$. Dann ist $f_*\mu = \mu_{\exp(\lambda)}$, wobei $\mu_{\exp(\lambda)}$ die Verteilungsfunktion $F_{\exp(\lambda)}$ aus (3.9) hat. Denn es gilt für $x \ge 0$

$$f_*\mu([0,x]) = \mu(f^{-1}([0,x])) = \mu([e^{-\lambda x}, 1]) = 1 - e^{-\lambda x}.$$

Beispiel 3.26 (Beispiel einer nicht Borel-messbaren Menge, Vitali-Menge). Bisher gab es noch kein Beispiel für eine Menge, die nicht in $\mathcal{B}(\mathbb{R})$ liegt. Eine solche werden wir nun konstruieren. Sie ist als *Vitali-Menge* bekannt. Hierzu definieren wir auf $\mathbb{R}$ eine Äquivalenzrelation durch $x \sim y \iff y - x \in \mathbb{Q}$. Bezüglich dieser Äquivalenzrelation zerfällt $\mathbb{R}$ in Äquivalenzklassen der Form $\{x + q : q \in \mathbb{Q}\}$. Wir wählen aus jeder Äquivalenzklasse eine Zahl aus $[0, 1]$ aus und bilden damit die Menge V. (Es sei hier angemerkt, dass dieses Auswählen mittels des Auswahlaxioms der Mengenlehre geschieht und damit keinen trivialen Schritt darstellt.) Weiter sei nun für $q \in \mathbb{Q} \cap [-1, 1]$

$$V_q := \{x + q : x \in V\}.$$

Dann ist $[0, 1] \subseteq \biguplus_{q\in\mathbb{Q}\cap[-1,1]} V_q \subseteq [-1, 2]$.

Angenommen, die Menge V wäre messbar. Dann wären auch die Mengen V_q messbar und, wegen der Translationsinvarianz des Lebesgue-Maßes aus Beispiel 3.25.2 wäre $\lambda(V_q)$ nicht von q abhängig. Sei also $\lambda(V_q) =: a \geq 0$. Außerdem wäre dann wegen der Monotonie des Lebesgue-Maßes

$$1 \leq \sum_{q\in\mathbb{Q}\cap[-1,1]} \lambda(V_q) = \sum_{q\in\mathbb{Q}\cap[-1,1]} a \leq 3.$$

Dies ist jedoch nicht möglich, weder für $a = 0$ noch für $a > 0$. Wegen diesem Widerspruch muss also $V \notin \mathcal{B}(\mathbb{R})$ gelten.

4 Messbare Funktionen und das Integral

In diesem Kapitel sei immer $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Wir können mittlerweile mit dem Maß μ Maßen den Inhalt von Mengen aus $\mathcal{F}$ messen. Ziel der Einführung des Integrals ist es, bei einer solchen Messung die Elemente von Ω verschieden zu gewichten. Diese Gewichtung wird mit einer Funktion $f : \Omega \to \mathbb{R}$ vorgenommen. Solche Funktionen müssen die Minimalforderung der Messbarkeit genügen. Das Resultat dieser Gewichtung führt auf den Begriff des Integrals.

4.1 Messbare Funktionen

Wir kennen bereits den Begriff der (bezüglich der σ-Algebra $\mathcal{F}$) messbaren Menge, der nun erweitert wird. Sei $A \in \mathcal{F}$ also eine messbare Menge, so definiert diese mit $\omega \mapsto 1_A(\omega)$ eine messbare Funktion in dem in Definition 4.3 eingeführte Sinne. Die Linearkombination von solchen Indikatorfunktionen werden wir einfache Funktion nennen. Diesen kommt wegen Theorem 4.8 eine besondere Bedeutung zu, da jede nicht-negative messbare Funktion (im Sinne der punktweisen Konvergenz) von unten durch einfache Funktionen approximiert werden kann.

Bemerkung 4.1 (Notation). Seien Ω, Ω' Mengen, $f : \Omega \to \Omega'$ und I beliebig.

1. Wir schreiben $f(A) := \{f(\omega) : \omega \in A\}$ für $A \subseteq \Omega$ sowie $f^{-1}(A') := \{f^{-1}(\omega') : \omega' \in A'\}$ für $A' \subseteq \Omega'$. Wir bemerken, dass folgende Regeln für $A', A'_i \subseteq \Omega', i \in I$ gelten:

$$f^{-1}((A')^c) = (f^{-1}(A'))^c, \quad f^{-1}\Big(\bigcap_{i\in I} A'_i\Big) = \bigcap_{i\in I} f^{-1}(A'_i), \quad f^{-1}\Big(\bigcup_{i\in I} A'_i\Big) = \bigcup_{i\in I} f^{-1}(A'_i).$$

   Etwas Vorsicht ist jedoch geboten, da für $A, A_i \subseteq \Omega, i \in I$ nur $f(\bigcup_{i\in I} A_n) = \bigcup_{i\in I} f(A_i)$, im allgemeinen jedoch $f(A^c) \neq (f(A))^c$ und $f(\bigcap_{i\in I} A_i) \neq \bigcap_{i\in I} f(A_i)$.

2. Für $\mathcal{C} \subseteq 2^{\Omega'}$ schreiben wir ganz analog

$$f^{-1}(\mathcal{C}) := \{f^{-1}(A') : A' \in \mathcal{C}\}.$$

Lemma 4.2 (Urbilder von $\boldsymbol{\sigma}$-Algebren). *Sei Ω eine Menge und $(\Omega', \mathcal{F}')$ ein Messraum, $f : \Omega \to \Omega'$ und $\mathcal{C}' \subseteq \mathcal{F}'$ mit $\sigma(\mathcal{C}') = \mathcal{F}'$. Dann ist $\sigma(f^{-1}(\mathcal{C}')) = f^{-1}(\sigma(\mathcal{C}'))$. Insbesondere ist $f^{-1}(\mathcal{F}')$ eine σ-Algebra auf Ω.*

Beweis. '$\subseteq$': Nach Bemerkung 4.1 ist klar, dass $f^{-1}(\sigma(\mathcal{C}'))$ eine σ-Algebra ist. Damit ist $\sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(f^{-1}(\sigma(\mathcal{C}'))) = f^{-1}(\sigma(\mathcal{C}'))$.
'$\supseteq$': Wir definieren

$$\widetilde{\mathcal{F}}' := \{A' \in \sigma(\mathcal{C}') : f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))\} \subseteq \sigma(\mathcal{C}').$$

Dann ist, wieder wegen Bemerkung 4.1, $\widetilde{\mathcal{F}}'$ eine σ-Algebra und $\mathcal{C}' \subseteq \widetilde{\mathcal{F}}' \subseteq \sigma(\mathcal{C}')$, also $\widetilde{\mathcal{F}}' = \sigma(\mathcal{C}')$. Für $A' \in \sigma(\mathcal{C}')$ ist also $f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))$, was gleichbedeutend ist mit $f^{-1}(\sigma(\mathcal{C}')) \subseteq \sigma(f^{-1}(\mathcal{C}'))$. □

Definition 4.3 (Messbare Funktionen). *Seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ Messräume und $f : \Omega \to \Omega'$.*

1. *Die Funktion f heißt* $\mathcal{F}/\mathcal{F}'$-messbar, *falls $f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$. Die σ-Algebra $f^{-1}(\mathcal{F}')$ heißt* die von f erzeugte σ-Algebra *auf Ω und wird auch mit $\sigma(f)$ bezeichnet.*

2. *Ist $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \Omega'$ messbar, so heißt X eine* Ω'-wertige Zufallsvariable. *Das Bildmaß $X_*\mathbf{P}$ aus Definition 3.22 heißt dann die* Verteilung von X.

3. *Ist $(\Omega', \mathcal{F}') = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, so heißt f eine* reell(wertig)e Funktion. *Ist f nach $\mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$-messbar, so sagen wir, die Funktion f sei (Borel-)messbar.*

4. *Ist $\Omega' = \overline{\mathbb{R}}$ und $f = 1_A$ für $A \subseteq \Omega$, so heißt f auch* Indikatorfunktion. *Ist $f = \sum_{k=1}^n c_k 1_{A_k}$ für $c_1, \dots, c_n \in \overline{\mathbb{R}}$ paarweise verschieden und $A_1, \dots, A_n \subseteq \Omega$, so heißt f* einfach.

Beispiel 4.4. Sei $(\Omega, \mathcal{F})$ ein Messraum.

1. Die allermeisten Funktionen $f : \Omega \to \overline{\mathbb{R}}$, die man sich vorstellen kann, sind (Borel-)messbar. Etwa ist die Identitätsabbildung $f : \omega \mapsto \omega$ messbar, da $f^{-1}(\mathcal{F}) = \mathcal{F}$.

2. Es ist wichtig zu sehen, dass für viele messbaren Funktionen f gilt, dass $\sigma(f) \subsetneq \mathcal{F}$, siehe etwa das nächste Beispiel.

3. Eine Funktion $f : \Omega \to \{0, 1\}$ ist genau dann messbar, wenn $f^{-1}(\{1\}) \in \mathcal{F}$. Es ist nämlich in diesem Fall $\sigma(f) = \{\emptyset, f^{-1}(\{1\}), (f^{-1}(\{1\}))^c, \Omega\}$.

4. Sei $\mathcal{F} = \mathcal{B}(\mathbb{R})$. Um eine nicht $\mathcal{F}$-messbare Funktion anzugeben, muss man sich genauso anstrengen, wie eine nicht Borel-messbare Menge zu konstruieren. Es ist etwa für die Vitali-Menge V aus Beispiel 3.26 die Funktion 1_V nicht messbar.

Lemma 4.5 (Eigenschaften von Messbarkeit). *Seien $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$ und $(\Omega'', \mathcal{F}'')$ Messräume.*

1. *Ist $\mathcal{C}' \subseteq \mathcal{F}'$ mit $\mathcal{F}' = \sigma(\mathcal{C}')$, so ist $f : \Omega \to \Omega'$ genau dann nach $\mathcal{F}/\mathcal{F}'$-messbar, wenn $f^{-1}(\mathcal{C}') \subseteq \mathcal{F}$.*

2. *Ist $f : \Omega \to \Omega'$ nach $\mathcal{F}/\mathcal{F}'$ und $g : \Omega' \to \Omega''$ nach $\mathcal{F}'/\mathcal{F}''$ messbar. Dann ist $g \circ f : \Omega \to \Omega''$ nach $\mathcal{F}/\mathcal{F}''$-messbar.*

3. *Seien $(\Omega, \mathcal{O})$ und $(\Omega', \mathcal{O}')$ topologische Räume, $f : \Omega \to \Omega'$ stetig sowie $\mathcal{F} = \sigma(\mathcal{O})$ und $\mathcal{F}' = \sigma(\mathcal{O}')$ die Borel'schen σ-Algebren. Dann ist f nach $\mathcal{F}/\mathcal{F}'$-messbar.*

4. *Eine reelle Funktion f (d.h. $f : \Omega \to \mathbb{R}$) ist genau dann messbar, wenn $\{\omega : f(\omega) \leq x\} \in \mathcal{F}$ für alle $x \in \mathbb{Q}$.*

5. *Eine einfache Funktion $f = \sum_{k=1}^n c_k 1_{A_k}$ mit paarweise verschiedenen $c_1, \dots, c_n \in \overline{\mathbb{R}}$ und $A_1, \dots, A_n \subseteq \Omega$ ist genau dann messbar, wenn $A_1, \dots, A_n \in \mathcal{F}$.*

Beweis. 1. Die 'genau dann'-Richtung ist klar. Für die 'wenn'-Richtung verwenden wir Lemma 4.2 und erhalten $f^{-1}(\mathcal{F}') = f^{-1}(\sigma(\mathcal{C}')) = \sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(\mathcal{F}) = \mathcal{F}$. Damit ist f nach $\mathcal{F}/\mathcal{F}'$-messbar.
2. Wir schreiben direkt $(g \circ f)^{-1}(\mathcal{F}'') = f^{-1}(g^{-1}(\mathcal{F}'')) \subseteq f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$, was die Behauptung bereits zeigt.
3. Nach Definition der Borel'schen σ-Algebra ist $\mathcal{O}'$ ein Erzeuger für $\mathcal{B}(\Omega')$. Da f stetig ist (d.h. $f^{-1}(\mathcal{O}') \subseteq \mathcal{O}$) folgt $f^{-1}(\mathcal{O}') \subseteq \mathcal{O} \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega)$. Nach 1. ist f also $\mathcal{B}(\Omega)/\mathcal{B}(\Omega')$-messbar.
4. Wir verwenden 1. Ist nämlich $\Omega' = \mathbb{R}$, so wird nach Lemma 4.2 $\mathcal{B}(\Omega')$ von $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{Q}\}$ erzeugt. Also ist ein reelles f genau dann messbar, wenn $f^{-1}(\mathcal{C}') = \{\{\omega : f(\omega) \leq x\} : x \in \mathbb{R}\} \subseteq \mathcal{F}$.
5. Sei $A := \left(\bigcup_{k=1} A_k\right)^c \in \mathcal{F}$. Dann ist $f^{-1}(\mathcal{B}(\overline{\mathbb{R}})) = \left\{A \cup \bigcup_{j \in J} A_j, \bigcup_{j \in J} A_j : J \subseteq \{1, \dots, n\}\right\}$, woraus die Behauptung folgt. □

Lemma 4.6 (Messbarkeit und Rechenregeln). *Sei $(\Omega, \mathcal{F})$ ein Messraum.*

1. *Sind $f, g : \Omega \to \overline{\mathbb{R}}$ messbar. Dann sind auch fg, sowie $af + bg$ für $a, b \in \mathbb{R}$ und f/g messbar, falls $g(\omega) \neq 0$ für alle $\omega \in \Omega$.*

2. *Sind $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ messbar, so sind auch*
$$\sup_{n \in \mathbb{N}} f_n, \quad \inf_{n \in \mathbb{N}} f_n, \quad \limsup_{n \to \infty} f_n, \quad \liminf_{n \to \infty} f_n$$
messbar. Falls existent, ist auch $\lim_{n \to \infty} f_n$ messbar.

Beweis. 1. Betrachte die Abbildung $\psi : \Omega \to \mathbb{R}^2$, definiert durch $\psi(\omega) = (f(\omega), g(\omega))$. Man überlegt sich leicht, dass ψ nach $\mathcal{F}/\mathcal{B}(\mathbb{R}^2)$-messbar ist. Weiter ist $(x, y) \mapsto ax + by$ und $(x, y) \mapsto xy$ auf $\mathbb{R}$ und $(x, y) \mapsto x/y$ auf $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$ stetig, also messbar nach Lemma 4.5.3. Damit folgen die Behauptungen nach Lemma 4.5.2.
2. Wir zeigen nur die Messbarkeit von $\sup_{n \in \mathbb{N}} f_n$. Die anderen Aussagen folgen dann mittels $\inf_{n \in \mathbb{N}} f_n = -\sup_{n \in \mathbb{N}}(-f_n)$, sowie $\limsup_{n \to \infty} f_n = \inf_{n \in \mathbb{N}} \sup_{k \geq n} f_k$ und $\liminf_{n \to \infty} f_n = \sup_{n \in \mathbb{N}} \inf_{k \geq n} f_k$. Es gilt für $x \in \mathbb{R}$ nach Lemma 4.5.4
$$\left\{\omega : \sup_{n \in \mathbb{N}} f_n(\omega) \leq x\right\} = \bigcap_{n=1}^{\infty} \underbrace{\left\{\omega : f_n(\omega) \leq x\right\}}_{\in \mathcal{F}} \in \mathcal{F}$$
und die Behauptung ist gezeigt. □

Korollar 4.7 (Messbarkeit von Positiv- und Negativteil). *Sei* $(\Omega, \mathcal{F})$ *ein Messraum und* $f : \Omega \to \overline{\mathbb{R}}$. *Dann ist* f *genau dann messbar, wenn* $f^+ := f \vee 0$ *und* $f^- := (-f) \vee 0$ *messbar sind. Dann ist auch* $|f|$ *messbar.*

Beweis. Es gilt $f = f^+ - f^-$ und $|f| = f^+ + f^-$. Damit folgt die Behauptung aus Lemma 4.6.2. □

Theorem 4.8 (Approximation mit messbaren Funktionen). *Sei* $(\Omega, \mathcal{F})$ *ein Messraum und* $f : \Omega \to \overline{\mathbb{R}}$ *messbar. Dann gibt es eine Folge* $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ *von einfachen Funktionen mit*[8] $f_n \uparrow f$.

Beweis. Wir schreiben einfach für[9] $\omega \in \Omega, n \in \mathbb{N}$

$$f_n(\omega) = n \wedge 2^{-n}[2^n f(\omega)],$$

und bemerken, dass $f_n \uparrow f$ per Konstruktion gilt. Außerdem ist $\omega \mapsto [2^n f(\omega)]$ nach Lemma 4.5.4 messbar, wenn f messbar ist. □

4.2 Definition

Die Konstruktion des Integrals einer Funktion f nach einem Maß erfolgt nun in mehreren Schritten. Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Wir bemerken, dass wir für das Integral von $f : \Omega \to \mathbb{R}$ nach μ verschiedene synonyme Schreibweisen verwenden, nämlich

$$\mu[f] = \int f d\mu = \int f(\omega)\mu(d\omega). \tag{4.1}$$

Die Definition des Integrals erfolgt zunächst für einfache Funktionen und anschließend (siehe Theorem 4.8) durch Approximation für allgemeine nicht-negative messbare Funktionen. Das Integral für (nicht notwendig nicht-negative) messbare Funktionen wird dann durch das Integral des Positiv- und Negativteils definiert.

Später werden wir Erwartungswerte von Zufallsvariablen $X : \Omega \to \mathbb{R}$ berechnen. Dann ist $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und wir verwenden die Schreibweise

$$\mathbf{E}[X] := \mathbf{P}[X],$$

wobei $\mathbf{P}[X]$ wie in (4.1) definiert ist.

Definition 4.9 (Integral von einfachen Funktionen). *Sei* $(\Omega, \mathcal{F}, \mu)$ *ein Maßraum und* $f = \sum_{k=1}^m c_k 1_{A_k}$ *eine einfache Funktion mit* $c_1, \dots, c_m \geq 0, A_1, \dots, A_m \in \mathcal{F}$. *Dann heißt*

$$\mu[f] := \int f d\mu := \sum_{k=1}^n c_k \mu(A_m)$$

das Integral von f bezüglich μ.

---

[8] Analog zu '↓' schreiben wir für $x, x_1, x_2, \dots \in \mathbb{R}$, dass $x_n \uparrow x$, falls $x_1 \leq x_2 \leq \dots$ und $x_n \xrightarrow{n\to\infty} x$. Für Funktionen $f, f_1, f_2, \dots : \Omega \to \mathbb{R}$ bedeutet $f_n \uparrow f$, dass $f_n(\omega) \uparrow f(\omega)$ für alle $\omega \in \Omega$.

[9] Hier ist $[x]$ für $x \in \mathbb{R}$ die größte ganze Zahl, die kleiner als x ist, also $[x] := \sup\{n \in \mathbb{Z} : n \leq x\}$.

Bemerkung 4.10 (Wohldefiniertheit des Integrals). Wir müssen sicherstellen, dass obiges Integral wohldefiniert ist. Sei dazu $f = \sum_{l=1}^n d_l 1_{B_l}$ eine weitere Darstellung von f mit $d_1, \dots, d_n \geq 0$ und $B_1, \dots, B_n \in \mathcal{F}$. Dann gilt

$$\sum_{k=1}^m c_k \mu(A_k) = \sum_{k=1}^m \sum_{l=1}^n c_k \mu(A_k \cap B_l) = \sum_{k=1}^m \sum_{l=1}^n d_l \mu(A_k \cap B_l) = \sum_{l=1}^n d_l \mu(B_l),$$

woraus die Wohldefiniertheit folgt.

Lemma 4.11 (Einfache Eigenschaften). *Seien f, g nicht-negative, einfache Funktionen und $\alpha \geq 0$. Dann gilt*[10]

$$\mu[af + bg] = a\mu[f] + b\mu[g], \qquad f \leq g \Rightarrow \mu[f] \leq \mu[g].$$

Beweis. Klar. □

Beispiel 4.12 (Das Integral von Indikatorfunktionen und Riemann-Integral). Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A \in \mathcal{F}$. Dann ist $f = 1_A$ eine einfache Funktion und es gilt

$$\mu[f] = \mu(A)$$

nach Definition 4.9. Es sei hierbei schon bemerkt, dass die Funktion $f = 1_A$ nicht mehr stückweise stetig sein muss. (Sei etwa A das in Beispiel 2.10 betrachtete Cantor'sche Kontinuum.) Deswegen ist es nicht klar, dass die Funktion 1_A integrierbar bezüglich des Riemann-Integrals ist.

Definition 4.13 (Das Integral von messbaren, nicht-negativen Funktionen). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \to \overline{\mathbb{R}}_+$ messbar. Das Integral von f bezüglich μ ist gegeben durch*

$$\mu[f] := \int f d\mu := \int f(\omega)\mu(d\omega) := \sup\{\mu[g] : g \text{ einfach, nicht-negativ}, g \leq f\}. \tag{4.2}$$

Bemerkung 4.14 (Das Integral als Fortsetzung). Nach Lemma 4.11 ist klar, dass die Definition von $\mu[f]$ für einfache, nicht-negative Funktionen aus Definition 4.9 und Definition 4.13 übereinstimmt. Damit ist obige Definition eine Fortsetzung von $\mu[f]$ auf den Raum der nicht-negativen, messbaren Funktionen.

Weiter ist wichtig zu sehen, dass nach Theorem 4.8 jede der in Definition 4.13 auftretenden Funktionen beliebig genau (punktweise) durch einfache Funktionen approximierbar ist. Insbesondere beinhaltet die Menge, deren Supremum in (4.2) gesucht wird, auch einfache Funktionen g, die beliebig dicht an f liegen.

Proposition 4.15 (Eigenschaften des Integrals). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f, g, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}_+$ messbar. Dann gilt:*

1. *Für $f \leq g$ gilt $\mu[f] \leq \mu[g]$.*

2. *Es gilt monotone Konvergenz, d.h.*

$$f_n \uparrow f \qquad \Rightarrow \qquad \mu[f_n] \uparrow \mu[f].$$

[10] Für Funktionen $f, g : \Omega \to \overline{\mathbb{R}}$ schreiben wir $f \leq g$, falls $f(\omega) \leq g(\omega)$ für alle $\omega \in \Omega$ gilt.

3. *Für $a, b \geq 0$ gilt $\mu[af + bg] = a\mu[f] + b\mu[g]$.*

Beweis. 1. ist klar nach der Definition des Integrals. Für 2. ist zunächst, da $f_n \leq f$ für $n = 1, 2, \ldots$

$$\lim_{n\to\infty} \mu[f_n] = \sup_{n\in\mathbb{N}} \mu[f_n] \leq \mu[f].$$

Für '$\geq$' genügt es zu zeigen, dass

$$\mu[g] \leq \sup_{n\in\mathbb{N}} \mu[f_n] \tag{4.3}$$

für alle einfachen Funktionen $g \leq f$. Sei also $g = \sum_{k=1}^m c_k 1_{A_k} \leq f$ für disjunkte Mengen $A_1, \ldots, A_m$ und $c_1, \ldots, c_m > 0$. Für $\varepsilon > 0$ und $n = 1, 2, \ldots$ sei $B_n^\varepsilon := \{f_n \geq (1-\varepsilon)g\}$. Da $f_n \uparrow f$ und $g \leq f$ gilt $\bigcup_{n=1}^\infty B_n^\varepsilon = \Omega$ für alle $\varepsilon > 0$. Also ist

$$\begin{aligned} \mu[f_n] \geq \mu[(1-\varepsilon)g1_{B_n^\varepsilon}] &= \sum_{k=1}^m (1-\varepsilon)c_k\mu(A_k \cap B_n^\varepsilon) \\ &\xrightarrow{n\to\infty} \sum_{k=1}^m (1-\varepsilon)c_k\mu(A_k) = (1-\varepsilon)\mu[g]. \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt (4.3).
Für 3. seien $f_1, g_1, f_2, g_2, \ldots$ einfache Funktionen mit $f_n \uparrow f$ und $g_n \uparrow g$. Dann ist $af_n + bg_n \uparrow af + bg$ und es folgt

$$\mu[af + bg] = \lim_{n\to\infty} \mu[af_n + bg_n] = \lim_{n\to\infty} a\mu[f_n] + b\mu[g_n] = a\mu[f] + b\mu[g]$$

aus 3. wegen Lemma 4.11. □

Wir können nun das Integral für messbare Funktionen definieren. Hierzu sei zunächst bemerkt, dass für $f : \Omega \to \overline{\mathbb{R}}$ messbar gilt, dass $f^+, f^- \leq |f|$. Ist insbesondere $\mu[|f|] < \infty$, so ist auch $\mu[f^+], \mu[f^-] < \infty$.

Definition 4.16 (Integral messbarer Funktionen). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \to \overline{\mathbb{R}}$ messbar. Dann heißt f nach μ integrierbar, falls $\mu[|f|] < \infty$ und wir definieren*

$$\mu[f] := \int f(\omega)\mu(d\omega) := \int f d\mu := \mu[f^+] - \mu[f^-]. \tag{4.4}$$

Außerdem setzen wir

$$\mathcal{L}^1(\mu) := \Big\{ f : \Omega \to \overline{\mathbb{R}} : \mu[|f|] < \infty \Big\}$$

Für $A \in \mathcal{F}$ schreiben wir außerdem

$$\mu[f, A] := \int_A f d\mu := \mu[f1_A].$$

Bemerkung 4.17 (Erweiterung des Integrals und $\mathcal{L}^p$-Räume). 1. Ist höchstens einer der beiden Terme $\mu[f^+]$ oder $\mu[f^-]$ unendlich, so definieren wir das Integral $\mu[f]$ weiterhin mittels (4.4). In anderen Fällen bleibt das Integral undefiniert.

2. Die Funktionenräume $\mathcal{L}^p(\mu) := \Big\{ f : \Omega \to \overline{\mathbb{R}} : \mu[|f|^p] < \infty \Big\}$, $p > 0$, werden in Abschnitt 5 eine besondere Rolle spielen.

4.3 Eigenschaften des Integrals

Wir stellen zunächst einige Eigenschaften des eingeführten Integralbegriffes fest. Diese sind etwa die Monotonie und Linearität. Weiter werden wir sehen, dass sich das Integral einer Funktion nicht ändert, wenn man sie auf einer Nullmenge abändert; siehe Proposition 4.20.

Proposition 4.18 (Einfache Eigenschaften des Integrals). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f, g \in \mathcal{L}^1(\mu)$. Dann gilt:*

1. *Das Integral ist monoton, d.h.*
$$f \leq g \text{ fast überall} \qquad \Longrightarrow \qquad \mu[f] \leq \mu[g].$$
2. *Es gilt die Dreiecksungleichung*
$$|\mu[f]| \leq \mu[|f|].$$
3. *Das Integral ist linear, seien also $a, b \in \mathbb{R}$, dann ist $af + bg \in \mathcal{L}^1(\mu)$ und*
$$\mu[af + bg] = a\mu[f] + b\mu[g].$$

Beweis. Alle Eigenschaften folgen aus Proposition 4.15.1 und 3., sowie aus der Definition des Integrals für messbare Funktionen. □

Proposition 4.19 (Substitutionssatz). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum, $(\Omega', \mathcal{F}')$ ein Messraum, $f : \Omega \to \Omega'$ messbar und $f_*\mu$ das Bildmaß von f aus Definition 3.22. Dann gilt für $g \in \mathcal{L}^1(f_*\mu)$, dass $g \circ f \in \mathcal{L}^1(\mu)$ und*
$$\mu[g \circ f] = f_*\mu[g].$$

Beweis. Es genügt, die Behauptung für einfache, nicht-negative Funktionen g zu zeigen. Der allgemeine Fall folgt dann mittels Approximation durch einfache Funktionen. Sei also $g = \sum_{k=1}^m c_k 1_{A'_k}$ mit $A'_k \in \mathcal{F}'$. Dann ist $g \circ f = \sum_{k=1}^m c_k 1_{f \in A'_k}$ und
$$\mu[g \circ f] = \sum_{k=1}^m c_k \mu(f \in A'_k) = \sum_{k=1}^m c_k f_*\mu(A'_k) = f_*\mu[g].$$
□

Proposition 4.20 (Integrale und Eigenschaften fast überall). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \to \overline{\mathbb{R}}_+$ messbar.*

1. *Es ist $f = 0$ fast überall genau dann, wenn $\mu[f] = 0$.*
2. *Falls $\mu[f] < \infty$, so ist $f < \infty$ fast überall.*

Beweis. 1. Sei $N := \{f > 0\} \in \mathcal{F}$.
'$\Rightarrow$': Es gilt $\mu(N) = 0$. Dann ist $f \leq \infty \cdot 1_N$, also gilt wegen Proposition 4.15.2
$$0 \leq \mu[f] \leq \mu[\infty, N] = \lim_{n\to\infty} \mu[n, N] = 0.$$
Für '$\Leftarrow$' sei $N_n := \{f \geq 1/n\}$ und damit $N_n \uparrow N$ und $nf \geq 1_{N_n}$, also
$$0 = \mu[f] \geq \tfrac{1}{n}\mu(N_n).$$

Damit ist $\mu(N_n) = 0$ und $\mu(N) = \mu(\bigcup_{n=1}^\infty N_n) = 0$.
Für 2. sei $A := \{f = \infty\}$. Wegen $f1_{f\geq n} \geq n1_{f\geq n}$ ist

$$\mu(A) = \mu[1_A] \leq \mu[1_{f\geq n}] \leq \tfrac{1}{n}\mu[f, 1_{f\geq n}] \leq \tfrac{1}{n}\mu[f] \xrightarrow{n\to\infty} 0.$$

Damit ist $\mu(f = \infty) = 0$, also $f < \infty$ fast überall; siehe Bemerkung 3.13. □

Zum Abschluss dieses Abschnittes stellen wir noch die Beziehung des (Lebesgue-)Integrals zum Riemann-Integral dar.

Definition 4.21 (Treppenfunktion und Riemann-Integral). *Eine Treppenfunktion $f : \mathbb{R} \to \mathbb{R}$, für die es eine Darstellung*

$$f(t) = \sum_{j=-\infty}^{\infty} a_j 1_{t_{j-1},t_j)}(t)$$

mit $t_{j-1} \leq t_j, j \in \mathbb{Z}$ gibt, wobei $a_j \in \mathbb{R}, j \in \mathbb{Z}$. Eine messbare Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt eigentlich Riemann-integrierbar*, falls $\lambda[|f|] < \infty$ und es Treppen-Funktionen $f_1^+, f_1^-, f_2^+, f_2^-, ...$ gibt mit $f_n^- \leq f \leq f_n^+$ und $\lambda[f_n^+ - f_n^-] \xrightarrow{n\to\infty} 0$. Das Riemann-Integral von f ist dann durch $\lambda[f]$ definiert. (Insbesondere stimmen dann Riemann- und Lebesgue-Integral überein.*

Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt uneigentlich Riemann-integrierbar*, wenn $f1_K$ für alle kompakten Intervalle $K \subseteq \mathbb{R}$ eigentlich Riemann-integrierbar ist und $\lambda[f1_{[-n,n]}]$ konvergiert. Dieser Grenzwert ist dann das Riemann-Integral von f bezüglich λ.*

Proposition 4.22 (Riemann-Integrierbarkeit). *Die Funktion $f : \mathbb{R}_+ \to \mathbb{R}$ habe eine diskrete Menge von Sprungstellen. Dann ist f genau dann eigentlich Riemann-integrierbar, wenn f Lebesgue-integrierbar ist und es gilt*

$$\lambda[f] = \lim_{n\to\infty} \sum_{k=1}^{\infty} f(s_{n,k})(t_{n,k} - t_{n,k-1}) \tag{4.5}$$

für $0 = t_{n,0} \leq ... \leq t_{n,k_n} = t$ mit $\max_k |t_{n,k} - t_{n,k-1}| \xrightarrow{n\to\infty} 0$ und beliebiges $t_{n,k-1} \leq s_{n,k} \leq t_{n,k}$.

Beweis. Es genügt, die Behauptung für stetiges f zu zeigen. Andernfalls kann man f in die stetigen Teilabschnitte zerlegen. Weiter genügt es, die Behauptung für f mit kompaktem Träger K zu zeigen. Da f auf K gleichmäßig stetig ist, wählt man zunächst $\varepsilon_n \downarrow 0$ und $t_{n,0} \leq ... \leq t_{n,k_n}$ so, dass $K \subseteq [t_{n,0}, t_{n,k_n}]$ und $\max_{t_{n,k-1}\leq s<t_{n,k}} |f(t_{n,k-1}) - f(s)| < \varepsilon_n$. Nun ist es leicht, Treppenfuntionen f_n^+ und f_n^- zu finden, so dass $f_n^- \leq f \leq f_n^+$ und $||f_n^+ - f_n^-|| \leq \varepsilon_n$. Daraus folgt die erste Behauptung. Die Formel (19.4) gilt wegen der gleichmäßigen Approximation der Funktion f. □

Beispiel 4.23 (Unterschiede zwischen Riemann- und Lebesgue-Integral). 1. Wir beginnen mit einer Funktion, die Lebesgue- aber nicht Riemann-integrierbar ist. Sei $f = 1_{[0,1]\cap\mathbb{Q}}$. Dann ist $1_{[0,1]} \leq f^+$ für jede reguläre Treppenfunktion $f^+ \geq f$ und $f^- \leq 0$ für jede reguläre Treppenfunktion $f^- \leq f$. Insbesondere ist f nicht Riemann-integrierbar.

2. Wie aus der Definition des Riemann-Integrals hervorgeht, ist jede eigentlich Riemann-integrierbare Funktion auch Lebesgue-integrierbar. Anders verhält es sich mit uneigentlich Riemann-integrierbaren Funktionen. Sei hierzu f gegeben durch $f(t) = \frac{(-1)^{\lceil t \rceil}+1}{\lceil t \rceil}$. Dann ist
$$\lambda[f1_{[0,2n]}] = \sum_{k=1}^{2n} \frac{(-1)^{k+1}}{k} = \sum_{k=1}^{n} \frac{1}{k} - \frac{1}{k+1} = 1 - \frac{1}{n+1} \xrightarrow{n\to\infty} 1.$$
Damit ist f uneigentlich Riemann-integrierbar. Allerdings gilt
$$\lambda[|f|] = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$
Also ist f nicht Lebesgue-integrierbar.

4.4 Konvergenzsätze

Man kann sich fragen, ob es wirklich so wichtig ist, dass man mehr Funktionen bezüglich des Lebesgue-Integrals als bezüglich des Riemann Integrals integrieren kann. Es kommen schließlich in Anwendungen meistens Riemann-integrierbare Funktionen vor. Es gibt allerdings einen weiteren Vorteil des Lebesgue-Integrals, auf den wir nun eingehen werden. Es gelten nämlich Vertauschungsgesetze von Grenzwerten und Integralen (die hier Konvergenzsätze genannt werden), die relativ wenige Voraussetzungen benötigen. Die Wichtigsten sind der Satz von der monotonen und der von der majorisierten Konvergenz.

Theorem 4.24 (Monotone Konvergenz). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum, $f_1, f_2, \dots \in \mathcal{L}^1(\mu)$ und $f : \Omega \to \overline{\mathbb{R}}$ messbar mit $f_n \uparrow f$ fast überall. Dann gilt*
$$\lim_{n\to\infty} \mu[f_n] = \mu[f],$$
wobei beide Seiten den Wert ∞ annehmen können.

Beweis. Sei $N \in \mathcal{F}$ so, dass $\mu(N) = 0$ und $f_n(\omega) \uparrow f(\omega)$ für $\omega \notin N$. Setze $g_n := (f_n - f_1)1_{N^c} \geq 0$. Damit ist $g_n \uparrow (f - f_1)1_{N^c} =: g$ und mit Proposition 4.18, Proposition 4.20 und Proposition 4.15.2 gilt
$$\mu[f_n] = \mu[f_1] + \mu[g_n] \xrightarrow{n\to\infty} \mu[f_1] + \mu[g] = \mu[f].$$
□

Theorem 4.25 (Lemma von Fatou). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}_+$ messbar. Dann gilt*
$$\liminf_{n\to\infty} \mu[f_n] \geq \mu[\liminf_{n\to\infty} f_n].$$

Beweis. Für alle $k \geq n$ gilt $f_k \geq \inf_{k\geq n} f_k$ und damit
$$\inf_{k\geq n} \mu[f_k] \geq \mu[\inf_{k\geq n} f_k]$$
wegen Proposition 4.15.1. Deshalb gilt mit $n \to \infty$
$$\liminf_{n\to\infty} \mu[f_n] \geq \sup_{n\in\mathbb{N}} \mu[\inf_{k\geq n} f_k] = \mu[\liminf_{n\to\infty} f_n]$$
wegen der monotonen Konvergenz, da $\inf_{k\geq n} f_k \uparrow \liminf_{n\to\infty} f_n$. □

Theorem 4.26 (Satz von der majorisierten Konvergenz). *Sei* $(\Omega, \mathcal{F}, \mu)$ *ein Maßraum und* $f, g, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ *messbar mit* $|f_n| \le g$ *fast überall,* $\lim_{n\to\infty} f_n = f$ *und* $g \in \mathcal{L}^1(\mu)$. *Dann gilt*

$$\lim_{n\to\infty} \mu[f_n] = \mu[f].$$

Beweis. Ohne Einschränkung gilt $|f_n| \le g$ überall. Wir verwenden das Lemma von Fatou, sowie $g - f_n, g + f \ge 0$, also

$$\begin{aligned}\mu[g+f] &\le \liminf_{n\to\infty} \mu[g+f_n] = \mu[g] + \liminf_{n\to\infty} \mu[f_n],\\ \mu[g-f] &\le \liminf_{n\to\infty} \mu[g-f_n] = \mu[g] - \limsup_{n\to\infty} \mu[f_n].\end{aligned}$$

Nach Subtrahieren von $\mu[g]$ ist also

$$\mu[f] \le \liminf_{n\to\infty} \mu[f_n] \le \limsup_{n\to\infty} \mu[f_n] \le \mu[f].$$

□

Beispiel 4.27 (Beispiele zu den Konvergenzsätzen).

1. Im Lemma von Fatou ist nicht vorausgesetzt, dass eines der f_n integrierbar ist. Wir geben nun ein Beispiel dafür, dass im Lemma von Fatou wirklich '<' und nicht '=' gilt. Sei hierzu λ das Lebesgue-Maß und $f_n = 1/n$ (also insbesondere f_n konstant), $n = 1, 2, \dots$. Dann gilt $f_n \downarrow 0$, also
$$\liminf_{n\to\infty} \mu[f_n] = \infty > 0 = \mu[0] = \mu[\liminf_{n\to\infty} f_n].$$
2. Im Satz von der majorisierten Konvergenz kann auf die Voraussetzung, dass $|f_n| \le g$ und $g \in \mathcal{L}^1(\mu)$ nicht verzichtet werden. Sei etwa λ das Lebesgue-Maß auf $[0,1]$ und $f_n = n \cdot 1_{[0,1/n]}$. Dann ist $\sup_{n\in\mathbb{N}} f_n(x) = \sup\{n : x \le 1/n\} = \left[\frac{1}{x}\right]$[11]. Also existiert kein $g \in \mathcal{L}^1(\lambda)$ mit $f_n \le g$. Außerdem ist $\lim_{n\to\infty} f_n = 0$ fast überall (wobei $\{0\}$ die Ausnahmemenge ist) und
$$\lim_{n\to\infty} \mu[f_n] = 1 \neq 0 = \mu[\lim_{n\to\infty} f_n].$$
Anders verhält es sich für $f_n = n \cdot 1_{[0,1/n^2]}$. Hier ist
$$\sup_{n\in\mathbb{N}} f_n(x) = \sup\{n : x \le 1/n^2\} = \left[\frac{1}{\sqrt{x}}\right] \le \frac{1}{\sqrt{x}} =: g(x).$$
Hier ist einerseits $g \in \mathcal{L}^1(\lambda)$, der Satz von der majorisierten Konvergenz ist also anwendbar, andererseits ist $\lim_{n\to\infty} f_n = 0$ fast überall und
$$\lim_{n\to\infty} \mu[f_n] = \lim_{n\to\infty} \tfrac{1}{n} = 0 = \mu[0] = \mu[\lim_{n\to\infty} f_n].$$

5 $\mathcal{L}^p$-Räume

Im ganzen folgenden Abschnitt sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Wir werden uns nun mit den messbaren Funktionen $f : \Omega \to \overline{\mathbb{R}}$ beschäftigen, für die $\mu[|f|^p] < \infty$ gilt. Die resultierenden Funktionenräume $\mathcal{L}^p(\mu)$ werden wir als normierte, vollständige Räume (Proposition 5.7) erkennen, was automatisch zu einem neuen Konvergenzbegriff führt. Weiter wird der Raum $\mathcal{L}^2(\mu)$ eine große Rolle spielen. Da man diesen mit einem Skalarprodukt (nämlich $\langle f, g\rangle := \mu[fg]$) ausstatten kann, stehen hier allgemeine Aussagen zur Verfügung, etwa der Satz von Riesz-Fréchet (Proposition 5.10). Dies werden wir verwenden, um σ-endliche Maße mit Dichte durch den Satz von Radon-Nikodým (Korollar 5.16) zu charakterisieren.

[11] Mit $[x] := \sup\{n \in \mathbb{Z} : n \le x\}$ bezeichnen wir die Abrundungsfunktion.

5.1 Grundlagen

Bereits in Bemerkung 4.17 haben wir die Räume $\mathcal{L}^p(\mu)$ erwähnt. Durch die Definition des Integrals im letzten Abschnitt können wir diese nun näher beleuchten. Vor allem zeigen wir die wichtige Hölder- und die Minkowski-Ungleichung, siehe Proposition 5.2.

Definition 5.1 ($\mathcal{L}^p(\mu)$-Räume). *Sei $0 < p \leq \infty$. Wir setzen*

$$\mathcal{L}^p := \mathcal{L}^p(\mu) := \{f : \Omega \to \overline{\mathbb{R}} \text{ messbar mit } ||f||_p < \infty\}$$

für

$$||f||_p := (\mu[|f|^p])^{1/p}, \qquad 0 < p < \infty \tag{5.1}$$

und

$$||f||_\infty := \inf\{K : \mu(|f| > K) = 0\}.$$

Auf den Räumen $\mathcal{L}^p$, $p \geq 1$ zeigen wir nun eine Dreiecksungleichung, die Minkowski-Ungleichung. Außerdem sei bemerkt, dass die Hölder-Ungleichung im Spezialfall $p = q = 2$ auch Cauchy-Schwartz-Ungleichung heißt.

Proposition 5.2 (Hölder und Minkowski's Ungleichung). *Seien f, g messbar.*

1. *Sei $0 < p, q, r \leq \infty$ so, dass $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Dann gilt*
$$||fg||_r \leq ||f||_p ||g||_q \qquad \text{(Hölder-Ungleichung)}$$

2. *Für $1 \leq p \leq \infty$ ist*
$$||f + g||_p \leq ||f||_p + ||g||_p. \qquad \text{(Minkowski-Ungleichung)}$$

Beweis. Wir starten mit dem Beweis der Hölder-Ungleichung. Im Fall $p = \infty$ oder $q = \infty$ ist die Aussage klar, sei also $p, q < \infty$. Ist entweder $||f||_p = 0$, $||f||_p = \infty$, $||g||_q = 0$ oder $||g||_q = \infty$, ist die Aussage ebenso klar. Sei also o.E. $f, g \geq 0$ und $0 < ||f||_p, ||g||_q < \infty$ und

$$\widetilde{f} := \frac{f}{||f||_p}, \qquad \widetilde{g} = \frac{g}{||g||_q}.$$

Dann ist zu zeigen, dass $||\widetilde{f}\widetilde{g}||_r \leq 1$. Wegen der Konvexität der Exponentialfunktion ist

$$(xy)^r = \exp\left(\tfrac{r}{p} p \log x + \tfrac{r}{q} q \log y\right) \leq \tfrac{r}{p} x^p + \tfrac{r}{q} y^q$$

und damit

$$||\widetilde{f}\widetilde{g}||_r^r = \mu[(\widetilde{f}\widetilde{g})^r] \leq \tfrac{r}{p}\mu[\widetilde{f}^p] + \tfrac{r}{q}\mu[\widetilde{g}^q] = 1$$

und die Behauptung folgt.

Zum Beweis der Minkowski-Ungleichung bemerken wir zunächst, dass in den Fällen $p = 1$ und $p = \infty$ die Behauptung klar ist. Im Falle $1 < p < \infty$ gilt mit $q = p/(p-1)$ und $r = 1/p + 1/q = 1$ mit der Hölder-Ungleichung

$$\begin{aligned}
||f + g||_p^p &\leq \mu[|f| \cdot |f + g|^{p-1}] + \mu[|g| \cdot |f + g|^{p-1}] \\
&\leq ||f||_p \cdot ||(f + g)^{p-1}||_q + ||g||_p \cdot ||(f + g)^{p-1}||_q \\
&= (||f||_p + ||g||_p) \cdot ||f + g||_p^{p-1},
\end{aligned}$$

da $||(f+g)^{p-1}||_q = ||(f+g)^{q(p-1)}||_1^{1/q} = ||(f+g)^p||_1^{(p-1)/p} = ||f+g||_p^{p-1}$. Dividieren durch $||f + g||_p^{p-1}$ bringt das Resultat. □

Proposition 5.3 (Zusammenhang zwischen $\mathcal{L}^r$ und $\mathcal{L}^q$). *Sei μ endlich und $1 \leq r < q \leq \infty$. Dann ist $\mathcal{L}^q(\mu) \subseteq \mathcal{L}^r(\mu)$.*

Beweis. Die Behauptung ist klar für $q = \infty$. Sei also $q < \infty$. Wir verwenden die Hölder-Ungleichung. Es gilt für $f \in \mathcal{L}^q$, da $||1||_p < \infty$ wegen der Endlichkeit von μ,

$$||f||_r = ||1 \cdot f||_r \leq ||1||_p \cdot ||f||_q < \infty \tag{5.2}$$

für $\frac{1}{p} = \frac{1}{r} - \frac{1}{q} > 0$, woraus die Behauptung sofort folgt. □

Bemerkung 5.4 (Gegenbeispiel für σ-endliches μ). Sicherlich gilt Proposition 5.3 nicht, falls μ nicht endlich ist. Sei etwa λ das eindimensionale Lebesgue-Maß und $f : x \mapsto \frac{1}{x} \cdot 1_{x>1}$. Dann ist $f \in \mathcal{L}^2(\lambda)$, aber $f \notin \mathcal{L}^1(\lambda)$.

5.2 $\mathcal{L}^p$-Konvergenz

Wir haben im Satz von der majorisierten Konvergenz (Theorem 4.26) gesehen, dass für eine Folge von Funktionen, die fast überall konvergiert, oft auch deren Integrale konvergieren. Die hier betrachtete $\mathcal{L}^p$-Konvergenz geht jetzt von Konvergenz von Integralen aus. Wir werden sehen, dass der resultierende Konvergenzbegriff zur Folge hat, dass jede Cauchy-Folge konvergiert (Proposition 5.7).

Definition 5.5 (Konvergenz im p-ten Mittel). *Eine Folge $f_1, f_2, \dots$ in $\mathcal{L}^p(\mu)$* konvergiert in $\mathcal{L}^p(\mu)$ *(oder* im p-ten Mittel*) gegen $f \in \mathcal{L}^p(\mu)$, falls*

$$||f_n - f||_p \xrightarrow{n\to\infty} 0.$$

Wir schreiben dann auch $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} f$.

Proposition 5.6 (Konvergenz im p-ten und q-ten Mittel). *Sei $\mu(\Omega) < \infty$, $1 \leq r < q \leq \infty$ und $f, f_1, f_2, \dots \in \mathcal{L}^q$. Falls $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^q} f$, so auch $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^r} f$.*

Beweis. Die Behauptung ist klar für $q = \infty$, sei also $q < \infty$. Aus (5.2) folgt sofort dass $||f - g||_r \leq ||f - g||_q$, woraus die Behauptung bereits folgt. □

Proposition 5.7 (Vollständigkeit von $\mathcal{L}^p$). *Sei $p \geq 1$ und $f_1, f_2, \dots$ eine Cauchy-Folge in $\mathcal{L}^p$. Dann gibt es ein $f \in \mathcal{L}^p$ mit $||f_n - f||_p \xrightarrow{n\to\infty} 0$.*

Beweis. Sei $\varepsilon_1, \varepsilon_2, \dots$ summierbar, also z.B. $\varepsilon_n := 2^{-n}$. Da $f_1, f_2, \dots$ eine Cauchy-Folge ist, gibt es für jedes k einen Index n_k mit $||f_m - f_n||_p \leq \varepsilon_k$ für alle $m, n \geq n_k$. Insbesondere gilt

$$\sum_{k=1}^{\infty} ||f_{n_{k+1}} - f_{n_k}||_p \leq \sum_{k=1}^{\infty} \varepsilon_k < \infty.$$

Nach monotoner Konvergenz und der Minkowski'schen Ungleichung gilt damit

$$\Big|\Big| \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \Big|\Big|_p \leq \sum_{k=1}^{\infty} ||f_{n_{k+1}} - f_{n_k}||_p < \infty.$$

Damit ist insbesondere $\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty$ fast überall, also $f_{n_1}(\omega), f_{n_2}(\omega), \dots$ für fast alle ω eine Cauchy-Folge in $\mathbb{R}$. Damit gibt es eine messbare Abbildung f mit $f_{n_k} \xrightarrow{k\to\infty} f$ fast überall. Nach Fatou's Lemma gilt

$$||f_n - f||_p \leq \liminf_{k\to\infty} ||f_{n_k} - f_n||_p \leq \sup_{m\geq n} ||f_m - f_n||_p \xrightarrow{n\to\infty} 0,$$

d.h. $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} f$. □

5.3 Der Raum $\mathcal{L}^2$

Für jedes $1 \leq p \leq \infty$ gilt, dass $||af||_p = |a| \cdot ||f||_p$ für $a \in \mathbb{R}$ gilt. Zusammen mit der Minkowski'schen Ungleichung bedeutet das, dass $\mathcal{L}^p(\mu)$ ein reeller Vektorraum ist. Es ist entscheidend zu bemerken, dass die Abbildung $f \mapsto ||f||_p$ eine Pseudo-Norm ist. Sie kann keine volle Norm[12] sein, weil $||f||_p = 0$ nach Proposition 4.20 nur impliziert, dass $\mu(f \neq 0) = 0$, aber nicht, dass $f = 0$. Wir werden deswegen im Folgenden Funktionen f und g miteinander identifizieren, wenn $f = g$ μ-fast überall gilt. Nach dem eben gesagtem ist dann $(\mathcal{L}^p, ||\cdot||_p)$ ein normierter Raum. Da $||\cdot||_p$ nach Proposition 5.7 vollständig ist, ist $(\mathcal{L}^p, ||\cdot||_p)$ sogar ein Banachraum für jedes $1 \leq p \leq \infty$. Als nächstes betrachten wir den Spezialfall $p = 2$. Wir definieren eine Abbildung $\mathcal{L}^2 \times \mathcal{L}^2 \to \mathbb{R}$ durch

$$\langle f, g \rangle := \mu[fg].$$

Dann ist $\langle \cdot, \cdot \rangle$ offensichtlich linear, symmetrisch und positiv semi-definit, also ein Skalarprodukt[13]. Wir schreiben konsequenterweise $f \perp g$ genau dann, wenn $\mu[fg] = 0$. Da $||f|| := ||f||_2 = \langle f, f \rangle^{1/2}$, ist $(\mathcal{L}^2, \langle \cdot, \cdot \rangle)$ also ein Hilbertraum.

Lemma 5.8 (Parallelogrammidentität). *Seien $f, g \in \mathcal{L}^2$. Dann gilt*

$$||f + g||^2 + ||f - g||^2 = 2||f||^2 + 2||g||^2.$$

Beweis. Aus der Definition von $||\cdot||$ und der Symmetrie und Bilinearität von $\langle \cdot, \cdot \rangle$ folgt

$$||f + g||^2 + ||f - g||^2 = \langle f + g, f + g \rangle + \langle f - g, f - g \rangle = 2\langle f, f \rangle + 2\langle g, g \rangle = 2||f||^2 + 2||g||^2.$$

□

Proposition 5.9 (Zerlegung von $f \in \mathcal{L}^2$). *Sei M ein abgeschlossener, linearer Teilraum von $\mathcal{L}^2$. Dann hat jede Funktion $f \in \mathcal{L}^2$ eine fast sicher eindeutige Zerlegung $f = g + h$ mit $g \in M, h \perp M$.*

Beweis. Für $f \in \mathcal{L}^2$ definieren wir

$$d_f := \inf_{g \in M} \{||f - g||\}.$$

Wähle $g_1, g_2, \ldots$ mit $||f - g_n|| \xrightarrow{n \to \infty} d_f$. Nach der Parallelogrammidentität gilt

$$4d_f^2 + ||g_m - g_n||^2 \leq ||2f - g_m - g_n||^2 + ||g_m - g_n||^2 = 2||f - g_m||^2 + 2||f - g_n||^2 \xrightarrow{m,n \to \infty} 4d_f^2.$$

Also ist $||g_m - g_n||^2 \xrightarrow{m,n \to \infty} 0$, d.h. $g_1, g_2, \ldots$ ist eine Cauchy-Folge. Nach Proposition 5.7 gibt es damit ein $g \in \mathcal{L}^2$ mit $||g_n - g||_p \xrightarrow{n \to \infty} 0$, und wegen der Abgeschlossenheit von M

[12] Sei V ein reeller Vektorraum. Eine Abbildung $||\cdot|| : V \to \mathbb{R}$ heißt *Norm*, wenn (i) $||x|| = 0$ gilt genau dann, wenn $x = 0$, (ii) $||a \cdot x|| = |a| \cdot ||x||$ für alle $x \in \mathbb{R}$ und (iii) $||x + y|| \leq ||x|| + ||y||$. Dann heißt das Paar $(V, ||\cdot||)$ ein normierter Raum.

[13] Sei V ein reeller Vektorraum. Dann heißt eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ ein *Skalarprodukt*, falls (i) $\langle x, \alpha y + z \rangle = \alpha \langle x, y \rangle + \langle x, z \rangle$ für alle $x, y, z \in V$ und $\alpha \in \mathbb{R}$ (Linearität), (ii) $\langle x, y \rangle = \langle y, x \rangle$ (Symmetrie) und (iii) $\langle x, x \rangle > 0$ für jedes $x \in V \setminus \{0\}$ (Positive Definitheit). Durch ein Skalarprodukt wird die Norm $||x|| := \langle x, x \rangle^{1/2}$ auf V definiert. Ist $(V, ||\cdot||)$ vollständig, so heißt $(V, \langle \cdot, \cdot \rangle)$ ein *Hilbertraum*.

auch $g \in M$. Da $||h|| = d_f$ für $h := f - g$, folgt für alle $t > 0, l \in M$, wegen der Definition von d_f,

$$d_f^2 \le ||h + tl||^2 = d_f^2 + 2t\langle h, l\rangle + t^2||l||^2.$$

Da dies für alle t gilt, folgt $\langle h, l\rangle = 0$, also $h \perp M$.

Zum Beweis der Eindeutigkeit sei $g' + h'$ eine weitere Zerlegung von f. Dann ist aufgrund der Linearität von M einerseits $g - g' \in M$, andererseits ist fast sicher $g - g' = h - h' \perp M$, also $g - g' \perp g - g'$. Dies bedeutet $||g - g'|| = \langle g - g', g - g'\rangle = 0$, also $g = g'$ fast überall. □

Proposition 5.10 (Riesz-Fréchet). *Eine Abbildung $F : \mathcal{L}^2 \to \mathbb{R}$ ist genau dann stetig und linear wenn es ein $h \in \mathcal{L}^2$ gibt, so dass für alle $f \in \mathcal{L}^2$*

$$F(f) = \langle f, h\rangle.$$

Dann ist $h \in \mathcal{L}^2$ fast sicher eindeutig bestimmt.

Beweis. '$\Leftarrow$': Die Linearität von $f \mapsto \langle f, h\rangle$ folgt aus der Bilinearität von $\langle\cdot,\cdot\rangle$. Die Stetigkeit folgt aus der Cauchy-Schwartz Ungleichung mittels

$$|\langle |f - f'|, h\rangle| \le ||f - f'|| \cdot ||h||.$$

'$\Rightarrow$': Falls $F \equiv 0$ wähle $h = 0$. Falls $F \not\equiv 0$, ist $M = F^{-1}\{0\}$ ein (wegen der Stetigkeit von F) abgeschlossener und (wegen der Linearität von F) linearer Unterraum von $\mathcal{L}^2$. Wähle $f' \in \mathcal{L}^2 \setminus M$ mit der (nach Proposition 5.9 fast sicher eindeutigen) orthogonalen Zerlegung $f' = g' + h'$ mit $g' \in M$ und $h' \perp M$. Da $f' \notin M$, ist $h' \not\equiv 0$ und $F(h') = F(f') - F(g') = F(f') \neq 0$. Wir setzen $h'' = \frac{h'}{F(h')}$, so dass $h'' \perp M$ und $F(h'') = 1$ und es gilt für alle $f \in \mathcal{L}^2$

$$F(f - F(f)h'') = F(f) - F(f)F(h'') = 0.$$

d.h. $f - F(f)h'' \in M$, also insbesondere $\langle F(f)h'', h''\rangle = \langle f, h''\rangle$ und

$$F(f) = \tfrac{1}{||h''||^2} \cdot \langle F(f)h'', h''\rangle = \tfrac{1}{||h''||^2} \cdot \langle f, h''\rangle = \langle f, \tfrac{h''}{||h''||^2}\rangle.$$

Nun folgt die Behauptung mit $h = \frac{h''}{||h''||^2}$.

Zur Eindeutigkeit sei $\langle f, h_1 - h_2\rangle = 0$ für alle $f \in \mathcal{L}^2$; insbesondere ist mit $f = h_1 - h_2$

$$||h_1 - h_2||^2 = \langle h_1 - h_2, h_1 - h_2\rangle = 0,$$

also $h_1 = h_2$ μ-fast sicher. □

Bemerkung 5.11 (Allgemeingültigkeit der letzten Aussagen). Lemma 5.8, sowie die Propositionen 5.9 und 5.10 gelten ebenso, falls man $\mathcal{L}^2$ durch einen anderen Hilbert-Raum ersetzt.

5.4 Satz von Radon-Nikodým

Aus der Vorlesung *Stochastik* sind bereits Verteilungen mit Dichte bekannt. Dieses Konzept wird nun aufgegriffen und in den Kontext von Integralen eingebettet. Sei hierzu ν ein weiteres Maß auf $\mathcal{F}$. Ziel ist es, Bedingungen anzugeben, wann das Maß ν durch eine Dichte darstellbar ist. Die Antwort findet sich im Satz von Radon-Nikodým (Korollar 5.16). Er ist ein Spezialfall des Lebesgue'schen Zerlegungssatzes, Theorem 5.15. Dieser zeigt, dass für je zwei σ-endliche Maße μ, ν das Maß ν (additiv) in zwei Anteile zerlegt werden kann: einen absolutstetigen bzgl. μ und einen zu μ singulären. Der absolutstetige Anteil hat dabei eine Dichte bzgl. μ. Zunächst müssen wir alle Begriffe erklären.

Definition 5.12 (Absolutstetige Maße). 1. *Wir sagen, ν besitzt eine* Dichte f *bzgl.* μ, *falls für alle* $A \in \mathcal{F}$

$$\nu(A) = \mu[f; A].$$

gilt. Wir schreiben dann $f = \frac{d\nu}{d\mu}$ *und* $\nu = f \cdot \mu$.

2. *Das Maß ν heißt* absolutstetig bzgl. μ, *falls alle μ-Nullmengen auch ν-Nullmengen sind. Wir schreiben dann $\nu \ll \mu$. Ist sowohl $\nu \ll \mu$ als auch $\mu \ll \nu$, so heißen μ und ν* äquivalent.

3. *Die Maße μ und ν heißen* singulär, *falls es ein $A \in \mathcal{F}$ gibt mit $\mu(A) = 0$ und $\nu(A^c) = 0$. Wir schreiben dann $\mu \perp \nu$.*

Lemma 5.13 (Kettenregel und Eindeutigkeit). *Sei μ ein Maß auf $\mathcal{F}$.*

1. *Sei ν ein σ-endliches Maß. Sind g_1 und g_2 Dichten von ν bzgl. μ, so ist $g_1 = g_2$ μ-fast überall.*

2. *Sei $f : \Omega \to \mathbb{R}_+$ und $g : \Omega \to \mathbb{R}$ messbar. Dann gilt*

$$(f \cdot \mu)[g] = \mu[fg],$$

falls eine der beiden Seiten existiert.

Beweis. 1. Sei $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ so, dass $\Omega_n \uparrow \Omega$ und $\nu(\Omega_n) < \infty$. Setze $A_n := \Omega_n \cap \{g_1 > g_2\}$. Da sowohl g_1 als auch g_2 Dichten von ν bzgl μ sind, folgt

$$\mu[g_1 - g_2; A_n] = 0.$$

Da auf A_n nur $g_1 > g_2$ möglich ist, ist $g_1 = g_2$ $1_{A_n}\mu$-fast überall. Außerdem ist

$$\mu\{g_1 > g_2\} = \mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = 0.$$

Analog folgt $\mu\{g_1 < g_2\} = 0$ und damit $g_1 = g_2$ μ-fast überall.

2. gilt per Definition für $g = 1_A$ mit $A \in \mathcal{F}$. Damit erweitert man die Aussage schrittweise für einfach Funktionen, positive messbare Funktionen und schlussendlich auf den allgemeinen Fall. □

Beispiel 5.14 (Bekannte Dichten). 1. Aus der Vorlesung *Stochastik* sind bereits einige Dichtefunktionen bekannt. Sei etwa für $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$

$$f_{N(\mu,\sigma^2)}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)$$

und λ das eindimensionale Lebesgue-Maß. Dann heißt das Wahrscheinlichkeitsmaß $f_{N(\mu,\sigma^2)} \cdot \lambda$ auch *Normalverteilung mit Erwartungswert μ und Varianz σ^2*.

Für $\gamma \geq 0$ und

$$f_{\exp(\gamma)}(x) := 1_{x\geq 0} \cdot \gamma e^{-\gamma x}$$

heißt $f_{\exp(\gamma)} \cdot \lambda$ auch *Exponentialverteilung zum Parameter γ*. Man kann nun beispielsweise mit Lemma 5.13 berechnen, dass

$$f_{\exp(\gamma)} \cdot \lambda[\mathrm{id}] = \int_0^\infty \gamma e^{-\gamma x} x dx = -e^{-\gamma x} x\Big|_0^\infty + \int_0^\infty e^{-\gamma x} dx = \frac{1}{\gamma}.$$

Den Wert $1/\gamma$ haben wir bereits als Erwartungswert der Exponentialverteilung zum Parameter γ interpretiert.

2. Es gibt natürlich nicht nur Dichten bezüglich des Lebesgue-Maßes. Sei beispielsweise

$$\mu = \sum_{n=0}^{\infty} \delta_n$$

das Zählmaß auf $\mathbb{N}_0$ (siehe Beispiel 3.2) und $f : \mathbb{N}_0 \to \mathbb{R}_+$, gegeben für ein $\gamma \geq 0$ durch

$$f(k) = e^{-\gamma} \frac{\gamma^k}{k!}.$$

Dann ist $f \cdot \mu$ die Poisson-Verteilung zum Parameter γ auf $2^{\mathbb{N}_0}$ nach Beispiel 3.2.

Theorem 5.15 (Lebesgue'scher Zerlegungssatz). *Seien μ, ν σ-endliche Maße auf $(\Omega, \mathcal{F})$. Dann lässt sich ν eindeutig schreiben als*

$$\nu = \nu_a + \nu_s \qquad \textit{mit} \qquad \nu_a \ll \mu, \nu_s \perp \mu.$$

Das Maß ν_a hat eine Dichte bzgl. μ, die μ-fast überall endlich ist.

Beweis. Durch eine Ausschöpfung $\Omega_1, \Omega_2, \dots \subseteq \Omega$ mit $\Omega_n \uparrow \Omega$ und $\nu(\Omega_n), \mu(\Omega_n) < \infty$ können wir uns auf den Fall endlicher Maße zurückziehen. Mit Proposition 5.6 ist die lineare Abbildung

$$\begin{cases} \mathcal{L}^2(\mu + \nu) & \to \mathbb{R}) \\ f & \mapsto \nu[f] \end{cases}$$

stetig. Nach Proposition 5.10 gibt es also ein $h \in \mathcal{L}^2(\mu + \nu)$ mit

$$\nu[f] = (\mu + \nu)[fh], \tag{5.3}$$

also

$$\nu[f(1-h)] = \mu[fh] \tag{5.4}$$

für jedes $f \in \mathcal{L}^2(\mu + \nu)$. Wählt man $f = 1_{\{h<0\}}$ in (5.3) folgt

$$0 \leq \nu\{h < 0\} = (\mu + \nu)[h; h < 0] \leq 0,$$

also $h \geq 0$ $(\mu + \nu)$-fast überall. Analog kann man mittels $f = 1_{\{h>1\}}$ aus (5.4) folgern, dass

$$0 \leq \mu[h; \{h > 1\}] = \nu[1 - h; \{h > 1\} \leq 0,$$

also $h \leq 1$ $(\mu + \nu)$-fast sicher. Sei nun $f \geq 0$ messbar und $f_1, f_2, \dots \in \mathcal{L}^2(\mu + \nu)$ mit $f_n \uparrow f$. Mit dem Satz von der monotonen Konvergenz folgt

$$\nu[f(1-h)] = \lim_{n\to\infty} \nu[f_n(1-h)] = \lim_{n\to\infty} \mu[f_N h] = \mu[fh],$$

d.h. (5.4) gilt für alle messbaren $f \geq 0$.

Sei nun $E := h^{-1}\{1\}$. Aus (5.4) folgt mit $f = 1_E$, dass

$$\mu(E) = \mu[h; E] = \nu[1-h; E] = 0.$$

Wir definieren für $A \in \mathcal{F}$ zwei Maße ν_a und ν_s durch

$$\nu_a(A) = \nu(A \setminus E), \qquad \nu_s(A) = \nu(A \cap E),$$

so dass $\nu = \nu_a + \nu_s$ und $\nu_s \perp \mu$. Um zu zeigen, dass $\nu_a \ll \mu$ wähle $A \in \mathcal{F}$ mit $\mu(A) = 0$. Damit ist nach (5.4)

$$\nu[1-h; A \setminus E] = \mu[h; A \setminus E] = 0.$$

Da $h < 1$ auf $A \setminus E$ ist, gilt damit $\nu_a(A) = \nu(A \setminus E) = 0$, also $\nu_a \ll \mu$.

Um die Dichte von ν_a bzgl. μ zu bestimmen, setzen wir $g := \frac{h}{1-h} 1_{\Omega \setminus E}$ und verwenden (5.4), so dass

$$\mu[g; A] = \mu\Big[\frac{h}{1-h}; A \setminus E\Big] = \nu(A \setminus E) = \nu_a(A).$$

Also ist $g = \frac{d\nu_a}{d\mu}$.

Um die Eindeutigkeit der Zerlegung zu zeigen, sei $\nu = \nu_a + \nu_s = \widetilde{\nu}_a + \widetilde{\nu}_s$ für $\nu_a, \widetilde{\nu}_a \ll \mu$, $\nu_s, \widetilde{\nu}_s \perp \mu$. Wähle $A, \widetilde{A} \in \mathcal{A}$ mit $\nu_s(A) = \mu(A^c) = \widetilde{\nu}_s(\widetilde{A}) = \mu(\widetilde{A}^c) = 0$. Dann gilt

$$\nu_s(A \cap \widetilde{A}) = \widetilde{\nu}_s(A \cap \widetilde{A}) = \nu_a(A^c \cup \widetilde{A}^c) = \widetilde{\nu}_a(A^c \cup \widetilde{A}^c) = 0$$

und deshalb

$$\begin{aligned} \nu_a &= 1_{A \cap \widetilde{A}} \cdot \nu_a = 1_{A \cap \widetilde{A}} \cdot \nu = 1_{A \cap \widetilde{A}} \cdot \widetilde{\nu}_a = \widetilde{\nu}_a, \\ \nu_s &= \nu - \nu_a = \nu - \widetilde{\nu}_a = \widetilde{\nu}_s. \end{aligned}$$

□

Korollar 5.16 (Satz von Radon-Nikodým). *Seien μ und ν σ-endliche Maße. Dann hat ν genau dann eine Dichte bzgl. μ, wenn $\nu \ll \mu$.*

Beweis. '$\Rightarrow$': klar.
'$\Leftarrow$': Nach Theorem 5.15 gibt es eine eindeutige Zerlegung $\nu = \nu_a + \nu_s$ mit $\nu_a \ll \mu, \nu_s \perp \mu$. Da $\nu \ll \mu$, muss $\nu_s = 0$ gelten und damit $\nu = \nu_a$. Insbesondere existiert die Dichte von ν bzgl. μ. □

Beispiel 5.17. Im Zerlegungssatz von Lebesgue 5.15 und im Satz von Radon-Nikodým 5.16 darf man die Voraussetzung, dass μ und ν σ-endlich sind nicht weglassen, wie folgendes Beispiel zeigt:

Sei $(\Omega, \mathcal{F})$ ein Messraum mit überabzählbarem Ω und

$$\mathcal{F} := \{A : A \text{ oder } A^c \text{ abzählbar}\}.$$

Seien μ und ν unendliche Maße auf $(\Omega, \mathcal{F})$, gegeben durch

$$\nu(A) := \begin{cases} 0, & A \text{ abzählbar}, \\ \infty, & \text{sonst}, \end{cases} \qquad \mu(A) := \begin{cases} |A|, & A \text{ endlich}, \\ \infty, & \text{sonst}. \end{cases}$$

Dann ist offenbar $\nu \ll \mu$. Angenommen, es gäbe eine $\mathcal{F}$-messbare Dichte von ν bzgl. μ, so wäre für alle $\omega \in \Omega$

$$0 = \nu\{\omega\} = \mu[f; \{\omega\}] = f(\omega)\mu(\{\omega\}) = f(\omega).$$

Damit wäre $f = 0$ und $\nu = 0$ im Widerspruch zur Definition von ν.

6 Produkträume

Sei $(\Omega_i)_{i\in I}$ eine Familie von Mengen. Dann heißt

$$\Omega := \bigtimes_{i\in I} \Omega_i := \{(\omega_i)_{i\in I} : \omega_i \in \Omega_i\}$$

Produktraum der $(\Omega_i)_{i\in I}$. Weiter definieren wir für $H \subseteq J \subseteq I$ die Projektionen

$$\pi_H^J : \bigtimes_{i\in J} \Omega_i \to \bigtimes_{i\in H} \Omega_i.$$

sowie $\pi_H := \pi_H^I$ und $\pi_i := \pi_{\{i\}}$, $i \in I$. In diesem Kapitel werden wir alle bisher eingeführten Konzepte auf solche Produkträume anwenden. Für die Anwendung der Maßtheorie auf stochastische Prozesse besonders wichtig wird Satz über projektive Limiten von Wahrscheinlichkeitsmaßen, Theorem 6.24.

6.1 Topologie

Wir beginnen mit der Definition der Topologie auf Produkträumen. Diese ist gerade so gemacht, dass Projektionen stetige Funktionen sind.

Definition 6.1 (Produktraum und Produkt-Topologie). *Ist $(\Omega_i, \mathcal{O}_i)_{i\in I}$ eine Familie von topologischen Räumen, dann heißt die von*

$$\mathcal{B} := \{\bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i : J \Subset I, A_i \in \mathcal{O}_i\}$$

erzeugte Topologie (siehe Definition 1.1.7) $\mathcal{O}$ die Produkttopologie *auf Ω.*

Bemerkung 6.2 (Stetigkeit der Koordinatenabbildungen). Es sei bemerkt, dass bezüglich der Produkttopologie $\mathcal{O}$ alle Projektionen $\pi_i, i \in I$ stetig sind. Es ist nämlich

$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{I\ni j\neq i} \Omega_j \in \mathcal{O}$$

für $A_i \in \mathcal{O}_i$. Damit ist die Projektion stetig (siehe Definition 1.1.10).

Proposition 6.3 (Abzählbare Produkte polnischer Räume sind polnisch). *Sei $(\Omega_i, \mathcal{O}_i)_{i\in\mathbb{N}}$ eine Familie polnischer Räume. Dann ist der Produktraum $(\Omega, \mathcal{O})$ aus Definition 6.1 polnisch.*

Beweis. Sei $\Omega_i' \subseteq \Omega_i$ abzählbar mit $\overline{\Omega_i'} = \Omega_i$ und r_i eine vollständige Metrik, die $\mathcal{O}_i$ erzeugt, $i \in \mathbb{N}$. Weiter sei $(\omega_i')_{i\in\mathbb{N}} \in \bigtimes_{i\in\mathbb{N}} \Omega_i'$.

Zunächst bemerken wir, dass für $\omega^1 = (\omega_i^1)_{i\in\mathbb{N}}, \omega^2 = (\omega_i^2)_{i\in\mathbb{N}} \in \Omega$ durch

$$r(\omega^1, \omega^2) := \sum_{i=1}^{\infty} 2^{-i}\big(r_i(\omega_i^1, \omega_i^2) \wedge 1\big)$$

eine vollständige Metrik auf Ω definiert wird, die $\mathcal{O}$ erzeugt. Außerdem ist damit die Menge

$$\Big\{(\omega_i)_{i\in\mathbb{N}} \in \bigtimes_{i\in\mathbb{N}} \Omega_i' : \omega_i \neq \omega_i' \text{ für endlich viele } i \in \mathbb{N}\Big\}$$

abzählbar und dicht in Ω. Also ist $(\Omega, \mathcal{O})$ polnisch. $\square$

6.2 Mengensysteme

Analog zur Topologie ist die Produkt-σ-Algebra gerade so, dass Projektionen messbare Funktionen sind.

Definition 6.4 (Produkt-σ-Algebra). *Ist $(\Omega_i, \mathcal{F}_i)_{i\in I}$ eine Familie von Messräumen, dann heißt die σ-Algebra*

$$\bigotimes_{i\in I} \mathcal{F}_i := \sigma\Big(\{\mathop{\times}_{i\in J} A_i \times \mathop{\times}_{i\in I\setminus J} \Omega_i : J \Subset I, A_i \in \mathcal{F}_i\}\Big) \tag{6.1}$$

die Produkt-σ-Algebra *auf Ω. Ist $(\Omega_i, \mathcal{F}_i) = (\Omega, \mathcal{F}), i \in I$, so setzen wir $\mathcal{F}^I := \bigotimes_{i\in I} \mathcal{F}$.*

Bemerkung 6.5 (Messbarkeit der Koordinatenabbildung). Analog zur Produkttopologie gilt, dass die Projektionen π_i messbar bzgl. $\bigotimes_{i\in I} \mathcal{F}_i$ sind. Es ist nämlich für $A_i \in \mathcal{F}_i$

$$\pi_i^{-1}(A_i) = A_i \times \mathop{\times}_{I\ni j\neq i} \Omega_j \in \bigotimes_{i\in I} \mathcal{F}_i.$$

Außerdem gilt

$$\bigotimes_{i\in I} \mathcal{F}_i = \sigma\Big(\big\{A_i \times \mathop{\times}_{j\neq i} \Omega_j : A_i \in \mathcal{F}_i, i \in I\big\}\Big). \tag{6.2}$$

Lemma 6.6 (Produkt-σ-Algebra bei abzählbaren Produkten). *Sei I abzählbar und $(\Omega_i, \mathcal{O}_i)_{i\in I}$ eine Familie polnischer Räume sowie $(\Omega, \mathcal{O})$ der Produktraum, versehen mit der Produkttopologie aus Definition 6.1. Dann ist $\mathcal{B}(\Omega) = \bigotimes_{i\in I} \mathcal{B}(\Omega_i)$. Insbesondere ist $\mathcal{B}(\mathbb{R}^d) = \bigotimes_{i=1}^d \mathcal{B}(\mathbb{R})$.*

Beweis. Da alle $(\Omega_i, \mathcal{O}_i)$, $i \in I$, separabel sind, gibt es nach Lemma 1.5 für jedes $i \in I$ eine abzählbare Basis $\mathcal{C}_i$ von $\mathcal{O}_i$. Damit ist

$$\mathcal{C} := \{\mathop{\times}_{i\in J} A_i \times \mathop{\times}_{i\in I\setminus J} \Omega_i : J \Subset I, A_i \in \mathcal{O}_i\}$$

eine abzählbare Basis von $(\Omega, \mathcal{O})$. Nach Lemma 2.8 gilt $\sigma(\mathcal{C}) = \mathcal{B}(\Omega)$. Außerdem ist offenbar $\mathcal{C} \subseteq \bigotimes_{i\in I} \mathcal{B}(\Omega_i)$ und damit $\mathcal{B}(\Omega) \subseteq \bigotimes_{i\in I} \mathcal{B}(\Omega_i)$. Andersherum liefert (6.2) und, dass für $A_i \in \mathcal{F}_i$

$$A_i \times \mathop{\times}_{j\neq i} \Omega_j \in \sigma\Big(\big\{A_i \times \mathop{\times}_{j\neq i} \Omega_j : A_i \in \mathcal{O}_i\big\}\Big) = \pi_i^{-1}(\mathcal{F}_i) \subseteq \mathcal{B}(\Omega),$$

woraus die Behauptung folgt. □

Lemma 6.7 (Produkte von Erzeugern/Halbringen sind Erzeuger/Halbringe). *Seien $(\Omega_i, \mathcal{F}_i)$ Messräume und $\Omega = \mathop{\times}_{i\in I} \Omega_i$.*

1. *Sei I endlich und $\mathcal{H}_i$ ein Halbring mit $\sigma(\mathcal{H}_i) = \mathcal{F}_i$. Dann ist*

$$\mathcal{H} := \{\mathop{\times}_{i\in I} A_i : A_i \in \mathcal{H}_i, i \in I\} \tag{6.3}$$

ein Halbring mit $\sigma(\mathcal{H}) = \bigotimes_{i\in I} \mathcal{F}_i$.

2. Sei I beliebig und $\mathcal{H}_i$ ein schnittstabiler Erzeuger von $\mathcal{F}_i$, $i \in I$. Dann ist

$$\mathcal{H} := \{\bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i : J \Subset I, A_i \in \mathcal{H}_i, i \in J\}$$

ein schnittstabiler Erzeuger von $\bigotimes_{i\in I} \mathcal{F}_i$.

Beweis. Für 1. ist o.E. $I = \{1, \ldots, d\}$. Es ist zunächst klar, dass $\mathcal{H}$ schnittstabil ist. Die Eigenschaft (ii) für Halbringe zeigt man durch Induktion über d. Die Behauptung ist klar für $d = 1$, da $\mathcal{H}_1$ ein Halbring ist. Gilt sie für $d - 1$, so gilt

$$\begin{aligned}
&(A_1 \times \cdots \times A_d) \setminus (B_1 \times \cdots \times B_d) \\
&\quad = \big(A_1 \times \cdots \times A_{d-1} \times (A_d \setminus B_d)\big) \uplus \big((A_1 \times \cdots \times A_{d-1}) \setminus (B_1 \times \cdots \times B_{d-1})\big) \times (A_d \cap A'_d)
\end{aligned}$$

Der erste Term der letzten Zeile ist als disjunkte Vereinigung von Mengen aus $\mathcal{H}$ darstellbar, da $\mathcal{H}_d$ ein Halbring ist. Der zweite Term ist als disjunkte Vereinigung darstellbar, da nach Induktionsvoraussetzung $(A_1 \times \cdots \times A_{d-1}) \setminus (B_1 \times \cdots \times B_{d-1})$ als disjunkte Vereinigung aus Mengen der Form $A_1 \times \cdots \times A_{d-1}$ mit $A_i \in \mathcal{H}_i, i = 1, \ldots, d-1$ darstellbar ist.

Für 2. ist wieder klar, dass $\mathcal{H}$ schnittstabil ist. Aus (6.1) folgt sofort, dass $\mathcal{H} \subseteq \bigotimes_{i\in I} \mathcal{F}_i$, also $\sigma(\mathcal{H}) \subseteq \bigotimes_{i\in I} \mathcal{F}_i$. Andersherum ist klar, dass für $A_i \in \mathcal{F}_i$

$$A_i \times \bigtimes_{j\neq i} \Omega_j \in \sigma\Big(\big\{A_i \times \bigtimes_{j\neq i} \Omega_j : A_i \in \mathcal{H}_i\big\}\Big) \subseteq \sigma(\mathcal{H}),$$

woraus $\bigotimes_{i\in I} \mathcal{F}_i \subseteq \sigma(\mathcal{H})$ wegen (6.2) und damit die Behauptung folgt. □

Korollar 6.8 (Borel'sche $\boldsymbol{\sigma}$-Algebra auf $\mathbb{R}^d$ wird von Quadern erzeugt). *Sei $\Omega = \mathbb{R}^d$. Für $\underline{a} = (a_1, \ldots, a_d), \underline{b} = (b_1, \ldots, b_d) \in \mathbb{R}^d$ setzen wir $\underline{a} \leq \underline{b}$ genau dann, wenn $a_i \leq b_i, i = 1, \ldots, d$, sowie mit*

$$(\underline{a}, \underline{b}] := (a_1, b_1] \times \cdots \times (a_d, b_d]$$

den halboffenen Quader. Dann definiert

$$\mathcal{H} := \{(\underline{a}, \underline{b}] : \underline{a}, \underline{b} \in \mathbb{Q}, \underline{a} \leq \underline{b}\}$$

einen Halbring mit $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R}^d)$.

Beweis. Nach Beispiel 2.3.1 und Lemma 6.7.1 ist $\mathcal{H}$ ein Halbring, der $\bigotimes_{i=1}^d \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$ erzeugt; siehe Lemma 6.6. □

6.3 Maße und Integrale

Mehrfachintegrale sind schon aus der Analysis bekannt. Wir definieren nun zunächst Maße auf Produkträumen und im gleichen Atemzug auch die dazu gehörigen (Mehrfach-)Integrale. Mit dem Satz von Fubini (Theorem 6.13) können dann Integrale nach Maßen auf Produkträumen als Mehrfachintegrale interpretiert und ausgewertet werden. Hierzu ist es notwendig, dass die in den Mehrfachintegralen auftauchenden Integranden messbar sind. Dies wird in Lemma 6.11 sichergestellt. Um Maße auf Produkträumen in genügender Allgemeinheit definieren zu können, benötigen wir zunächst den Begriff des Übergangskernes.

Definition 6.9 (Übergangskern). *Seien* $(\Omega_i, \mathcal{F}_i), i = 1, 2$ *Messräume. Eine Abbildung* $\kappa : \Omega_1 \times \mathcal{F}_2 \to \mathbb{R}_+$ *heißt ein* Übergangskern von $(\Omega_1, \mathcal{F}_1)$ nach $(\Omega_2, \mathcal{F}_2)$, *wenn (i) für alle* $\omega_1 \in \Omega_1$ *ist* $\kappa(\omega_1, .)$ *ein Maß auf* $\mathcal{F}_2$ *und (ii) für alle* $A_2 \in \mathcal{F}_2$ *ist* $\kappa(., A_2)$ *nach* $\mathcal{F}_1$*-messbar.*

Ein Übergangskern heißt σ*-endlich, wenn es eine Folge* $\Omega_{21}, \Omega_{22}, \dots \in \mathcal{F}_2$ *gibt mit* $\Omega_{2n} \uparrow \Omega_2$ *und* $\sup_{\omega_1} \kappa(\omega_1, \Omega_{2n}) < \infty$ *für alle* $n = 1, 2, \dots$ *Er heißt* stochastischer Kern *oder* Markov'scher Kern, *falls für alle* $\omega_1 \in \Omega_1$ *gilt, dass* $\kappa(\omega_1, \Omega_2) = 1$.

Beispiel 6.10 (Markov-Kette). Sei $\Omega = \{\omega_1, \dots, \omega_n\}$ endlich und $P = (p_{ij})_{1 \le i,j \le n}$ mit $p_{ij} \in [0, 1]$ und $\sum_{j=1}^n p_{ij} = 1$. Dann definiert für $A \subset \Omega$

$$\kappa(\omega_i, .) := \sum_{j=1}^n p_{ij} \cdot \delta_{\omega_j}$$

einen Markov-Kern von $(\Omega, 2^\Omega)$ nach $(\Omega, 2^\Omega)$. Hier ist P als stochastische Matrix die Übergangsmatrix einer homogenen, Ω-wertigen Markov-Kette.

Lemma 6.11 (Messbarkeit integrierbarer Schnitte). *Seien* $(\Omega_i, \mathcal{F}_i), i = 1, 2$ *Messräume,* κ *ein* σ*-endlicher Übergangskern von* $(\Omega_1, \mathcal{F}_1)$ *nach* $(\Omega_2, \mathcal{F}_2)$ *und* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}_+$ *nach* $\mathcal{F}_1 \otimes \mathcal{F}_2$ *messbar. Dann ist*

$$\omega_1 \mapsto \kappa(\omega_1, .)[f] := \int \kappa(\omega_1, d\omega_2) f(\omega_1, \omega_2)$$

nach $\mathcal{F}_1$*-messbar.*

Beweis. Wir nehmen an, dass $\kappa(\omega_1, \Omega_2) < \infty$ für alle $\omega_1 \in \Omega_1$ gilt. (Der allgemeine Fall erfolgt dann mittels einer Folge $\Omega_{11}, \Omega_{12}, \dots \in \mathcal{F}_1$ mit $\Omega_{1n} \uparrow \Omega_1$.) Sei

$$\mathcal{D} := \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 : \omega_1 \mapsto \kappa(\omega_1, .)[1_A] \text{ ist } \mathcal{F}_1\text{-messbar}\}.$$

Dann prüft man leicht nach, dass $\mathcal{D}$ ein schnittstabiles Dynkin-System ist. Weiter ist sicherlich $\mathcal{H} \subseteq \mathcal{D}$, wobei $\mathcal{H}$ wie in (6.3) definiert ist. Damit ist nach Theorem 2.13 $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{H}) \subseteq \mathcal{D} \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$. Also ist $\omega_1 \mapsto \kappa(\omega_1, .)[1_A]$ für alle $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ nach $\mathcal{F}_1$-messbar. Diese Aussage lässt sich sofort erweitern, indem man anstatt 1_A eine Treppenfunktionen einsetzt. Durch monotone Konvergenz folgt dann auch, dass $\omega_1 \mapsto \kappa(\omega_1, .)[f]$ für alle messbaren, nichtnegativen Funktionen nach $\mathcal{F}_1$-messbar ist. □

Theorem 6.12 (Satz von Ionescu-Tulcea). *Seien* $(\Omega_i, \mathcal{F}_i), i = 0, \dots, n$ *Messräume,* μ *ein* σ*-endliches Maß auf* $\mathcal{F}_0$ *und* κ_i *ein* σ*-endlicher Übergangskern von* $\left(\times_{j=0}^{i-1} \Omega_j, \bigotimes_{j=0}^{i-1} \mathcal{F}_j \right)$ *nach* $(\Omega_i, \mathcal{F}_i)$, $i = 1, \dots, n$. *Dann gibt es genau ein* σ*-endliches Maß* $\mu \bigotimes_{i=1}^n \kappa_i$ *auf* $\left(\times_{i=0}^n \Omega_i, \bigotimes_{i=0}^n \mathcal{F}_i \right)$ *mit*

$$\left(\mu \bigotimes_{i=1}^n \kappa_i\right)(A_0 \times \dots \times A_n) = \int_{A_0} \mu(d\omega_0) \left(\int_{A_1} \kappa_1(\omega_0, d\omega_1) \cdots \left(\int_{A_n} \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) \right) \cdots \right). \tag{6.4}$$

Beweis. Wir zeigen das Theorem nur für $n = 1$, der allgemeine Fall erfolgt dann durch Induktion.

Der Beweis ist eine Anwendung von Theorem 3.15. Zunächst stellen wir fest, dass nach Lemma 6.7 das in (6.3) definierte Mengensystem $\mathcal{H}$ ein Halbring auf $\times_{i=1}^n \Omega_i$ ist. Wir zeigen zunächst, dass die angegebene Mengenfunktion σ-endlich auf $\mathcal{H}$ ist. Es gibt nämlich $\Omega_{i1}, \Omega_{i2}, \in \mathcal{F}_i$ mit $\Omega_{in} \uparrow \Omega_i, i = 0, 1$ mit $\mu(\Omega_{0n}) < \infty, \kappa_1(\omega_0, \Omega_{1n}) < \infty, n = 1, 2, \ldots, \omega_0 \in \Omega_0$ und $\sup_{\omega_0 \in \Omega_0} \kappa_1(\omega_0, \Omega_{1n}) =: C_n < \infty$. Damit ist $\mu \otimes \kappa_1(\Omega_{0n} \times \Omega_{1n}) \leq C_n \cdot \mu(\Omega_{0n}) < \infty$ und $\Omega_{0n} \times \Omega_{1n} \uparrow \Omega_0 \times \Omega_1$. Damit ist also $\mu \otimes \kappa_1$ auch σ-endlich. Definiert man $\widetilde{\mu}$ auf $\mathcal{H}$ mittels (6.4), so ist dies also eine σ-endliche Mengenfunktion.

Wir zeigen nun, dass $\widetilde{\mu}$ σ-subadditiv und endlich additiv auf $\mathcal{H}$ ist. Für $A_1, \ldots, A_n \in \mathcal{H}$ und $A = \bigcup_{n=1}^\infty A_n \in \mathcal{H}$ ist wegen der σ-Subadditivität von $\kappa_1(\omega_0, .)$ für alle $\omega_0 \in \Omega_0$

$$\begin{aligned} \widetilde{\mu}(A) &= \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_A(\omega_0, \omega_1) \\ &\leq \sum_{n=1}^\infty \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_{A_n}(\omega_0, \omega_1) = \sum_{n=1}^\infty \widetilde{\mu}(A_n). \end{aligned}$$

Analog zeigt man die endliche Additivität. Nach Lemma 3.4 ist $\widetilde{\mu}$ damit σ-additiv. Aus Theorem 3.15 folgt nun, dass es genau eine Erweiterung von $\widetilde{\mu}$ auf $\sigma(\mathcal{H}) = \bigotimes_{i=1}^n \sigma(\mathcal{H}_i)$ gibt, welche die im Theorem angegebene ist. □

Wir beschäftigen uns nun mit dem in Theorem 6.12 definierten Maß.

Theorem 6.13 (Satz von Fubini). *Seien $(\Omega_i, \mathcal{F}_i)$, μ, κ_i und $\mu \bigotimes_{i=1}^n \kappa_i$ wie in Theorem 6.12. Weiter sei $f : \times_{i=0}^n \Omega_i \to \mathbb{R}_+$ messbar bezüglich $\bigotimes_{i=0}^n \mathcal{F}_i$. Dann gilt*

$$\int f d\big(\mu \bigotimes_{i=0}^n \kappa_i\big) = \int \mu(d\omega_0) \Big(\int \kappa_1(\omega_1, d\omega_2) \cdots \Big(\int \kappa_n(\omega_0, \ldots, \omega_{n-1}, d\omega_n) f(\omega_0, \ldots, \omega_n) \Big) \cdots \Big). \tag{6.5}$$

Diese Gleichheit gilt auch, falls $f : \times_{i=0}^n \Omega_i \to \mathbb{R}$ messbar ist mit $\int |f| d\big(\mu \bigotimes_{i=0}^n \kappa_i\big) < \infty$.

Beweis. Betrachte die Mengenfunktion $\widetilde{\mu}$ auf $\bigotimes_{i=0}^n \mathcal{F}_i$, gegeben durch

$$\widetilde{\mu} : A \mapsto \int \mu(d\omega_0) \Big(\int \kappa_1(\omega_1, d\omega_2) \cdots \Big(\int \kappa_n(\omega_0, \ldots, \omega_{n-1}, d\omega_n) 1_A(\omega_0, \ldots, \omega_n) \Big) \cdots \Big).$$

Man sieht, dass $\widetilde{\mu}$ auf $\mathcal{H}$ aus (6.3) mit $\mu \bigotimes_{i=1}^n \kappa_i$ übereinstimmt. Da $\mathcal{H}$ schnittstabil ist, folgt die Gleichheit (6.5) für Indikatorfunktionen wegen Proposition 3.10. Mittels Linearität des Integrals erweitert man (6.5) zunächst auf einfache Funktionen und dann mit Monotonie auf beliebige, nicht-negative, messbare Funktionen. Man beachte hierbei, dass alle vorkommenden Integranden nach Lemma 6.11 messbar sind. □

Korollar 6.14 (Produktmaße). *Sei $\Omega = \times_{i=1}^n \Omega_i$ und $\mathcal{H}_i \subseteq 2^{\Omega_i}$ ein Halbring, $i = 1, \ldots, n$, sowie $\mu_i : \mathcal{H}_i \to \mathbb{R}_+$ σ-endlich und, σ-additiv, $i = 1, \ldots, n$. Dann gibt es genau ein Maß $\mu_1 \otimes \cdots \otimes \mu_n$ auf $\bigotimes_{i=1}^n \sigma(\mathcal{H}_i)$ mit*

$$\mu_1 \otimes \cdots \otimes \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n). \tag{6.6}$$

Für eine Funktion $f : \Omega \to \mathbb{R}_+$ und jede Permutation π auf $\{1, \ldots, n\}$ gilt dann

$$\int f d\mu_1 \otimes \cdots \otimes \mu_n = \int \Big(\cdots \Big(\int f(\omega_1, \ldots, \omega_n) \mu_{\pi(1)}(d\omega_{\pi_1}) \Big) \cdots \Big) \mu_{\pi(n)}(d\omega_{\pi(n)}).$$

Diese Formel gilt auch für $f : \Omega \to \mathbb{R}$, falls $\int |f| d\mu_1 \otimes \cdots \otimes \mu_n < \infty$.

Beweis. Das Korollar folgt direkt aus Theorem 6.12 und Theorem 6.13, wenn man $\kappa_i(\omega_0, \dots, \omega_{i-1}, .) = \mu_i(.)$ für alle $\omega_0, \dots, \omega_{i-1}$ setzt. □

Definition 6.15 (Endliches Produktmaß). *Betrachte dieselbe Situation wie in Korollar 6.14. Dann heißt das eindeutige Maß $\mu_1 \otimes \dots \otimes \mu_n$ aus Korollar 6.14 das* Produktmaß der *$\mu_1, \dots, \mu_n$. Wir schreiben auch*

$$\bigotimes_{i=1}^{n} \mu_i := \mu_1 \otimes \dots \otimes \mu_n.$$

Gilt $(\Omega_i, \mathcal{H}_i, \mu_i) = (\Omega_0, \mathcal{H}_i, \mu_0)$, $i = 1, \dots, n$, sind also alle Räume gleich, so bezeichnen wir es auch mit

$$\mu_0^{\otimes n} := \mu_1 \otimes \dots \otimes \mu_n.$$

Beispiel 6.16 (Mehrdimensionales Lebesgue-Maß).

1. Sei λ das ein-dimensionale Lebesgue-Maß auf $\mathcal{B}(\mathbb{R})$ aus Proposition 3.17. Dann heißt $\lambda^{\otimes d}$ das d-dimensionale Lebesgue-Maß.

2. Sei $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch

$$f(x, y) = \frac{xy}{(x^2 + y^2)^2}.$$

Dann ist für jedes $x \in \mathbb{R}$

$$\int \lambda(dy) f(x, y) = 0,$$

da $f(x, .) \in \mathcal{L}^1(\lambda)$ und $f(x, y) = -f(x, -y)$. Also gilt insbesondere

$$\int \lambda(dx)\Big(\int \lambda(dy) f(x, y)\Big) = \int \lambda(dy)\Big(\int \lambda(dx) f(x, y)\Big) = 0,$$

allerdings ist $|f|$ nicht nach $\lambda^{\otimes 2}$ integrierbar, weil f in $(0, 0)$ eine nicht integrierbare Polstelle besitzt. Wie dieses Beispiel zeigt, muss man mit Mehrfachintegralen aufpassen. Insbesondere folgt aus der Gleichheit und Endlichkeit der Mehrfachintegrale nicht, dass der Integrand integrierbar ist.

6.4 Faltung von Maßen

Wir betrachten nun eine einfache Verknüpfung von Produktmaßen und Bildmaßen. Zur Faltung von Maßen μ, ν auf $\mathcal{B}(\mathbb{R})$ betrachten wir zunächst das Produktmaß $\mu \otimes \nu$. Das Bildmaß unter Summenbildung ist dann die Faltung aus μ, ν. Diese Faltung werden wir später als die Verteilung von $X + Y$ identifizieren, wenn X, Y *unabhängige* Zufallsvariablen mit Verteilung μ und ν sind. Manchmal, etwa bei Poisson-Verteilungen und bei Normalverteilungen, ist die Faltung wieder eine Poisson- bzw. Normalverteilung.

Definition 6.17 (Faltung von Maßen). *Seien $\mu_1, \dots, \mu_n$ σ-endliche Maße auf $\mathcal{B}(\mathbb{R})$ und $\mu_1 \otimes \dots \otimes \mu_n$ deren Produktmaß. Weiter sei $S(x_1, \dots, x_n) := x_1 + \dots + x_n$. Dann heißt das Bildmaß $S_*(\mu_1 \otimes \dots \otimes \mu_n)$ die* Faltung *der Maße $\mu_1, \dots, \mu_n$ und wird mit $\mu_1 * \dots * \mu_n$ oder $*_{i=1}^n \mu_i$ bezeichnet.*

Beispiel 6.18 (Faltung von Poisson- und geometrischen Verteilungen). 1. Für $\gamma_1, \gamma_2 \geq 0$ seien $\mu_{\text{Poi}(\gamma_1)}$ und $\mu_{\text{Poi}(\gamma_2)}$ zwei Poisson-Verteilungen aus Beispiel 3.2. Wir berechnen die Faltung der beiden Verteilungen durch

$$\begin{aligned}\mu_{\text{Poi}(\gamma_1)} * \mu_{\text{Poi}(\gamma_2)} &= \sum_{m,n} 1_{m+n=k} e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^n}{m!n!} \cdot \delta_k \\ &= \sum_{m=0}^{k} e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^{k-m}}{m!(k-m)!} \cdot \delta_k \\ &= e^{-(\gamma_1+\gamma_2)} \frac{(\gamma_1+\gamma_2)^k}{k!} \cdot \delta_k \sum_{m=0}^{k} \binom{k}{m} \frac{\gamma_1^m \gamma_2^{k-m}}{(\gamma_1+\gamma_2)^k} \\ &= \mu_{\text{Poi}(\gamma_1+\gamma_2)}.\end{aligned}$$

2. Die geometrische Verteilung zum Parameter $p \in [0,1]$ ist bereits aus Beispiel 3.2 bekannt. Die Faltung zweier Maße $\mu_{\text{geom}(p)}$ ist gegeben durch

$$\begin{aligned}\mu_{\text{geom}(p)} * \mu_{\text{geom}(p)} &= \sum_{m=2}^{k} (1-p)^{m-1} p (1-p)^{k-m-1} p \cdot \delta_k \\ &= (k-1)(1-p)^{k-2} p^2 \cdot \delta_k.\end{aligned}$$

Dies ist eine negative Binomialverteilung zu den Parametern p und 2.

Lemma 6.19 (Faltung von Verteilungen mit Dichten). *Sei λ ein Maß auf $\mathcal{B}(\mathbb{R})$, $\mu = f_\mu \cdot \lambda$ und $\nu = f_\nu \cdot \lambda$ für messbare Dichten $f_\mu, f_\nu : \mathbb{R} \to \mathbb{R}_+$. Dann gilt $\mu * \nu = f_{\mu*\nu} \cdot \lambda$ mit*

$$f_{\mu*\nu}(t) = \int f_\mu(s) f_\nu(t-s) \lambda(ds).$$

Beweis. Der Beweis ist eine einfache Anwendung des Satzes von Fubini, Theorem 6.13. □

Beispiel 6.20 (Faltung von Normalverteilungen). Seien $f_{N(\mu_1,\sigma_1^2)}$ und $f_{N(\mu_2,\sigma_2^2)}$ die Dichtefunktionen zweier Normalverteilungen mit Erwartungswert μ_1, μ_2 und Varianz σ_1^2 und σ_2^2. Sei weiter $\mu := \mu_1 + \mu_2$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Dann ist die Dichte der Faltung gegeben durch

$$\begin{aligned}x \mapsto & \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \int \exp\Big(- \frac{(y-\mu_1)^2}{2\sigma_1^2} - \frac{(x-y-\mu_2)^2}{2\sigma_2^2}\Big) dy \\ \overset{y \to (y-\mu_1)\sigma/(\sigma_1\sigma_2)}{=} & \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{\sigma_2^2 y^2}{2\sigma^2} - \frac{\Big((x-\mu) - y\frac{\sigma_1\sigma_2}{\sigma}\Big)^2}{2\sigma_2^2}\Big) dy \\ = & \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{\sigma_2^2 y^2 + \Big((x-\mu)\frac{\sigma}{\sigma_2} - \sigma_1 y\Big)^2}{2\sigma^2}\Big) dy \\ = & \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{(\sigma y - \frac{\sigma_1}{\sigma_2}(x-\mu))^2}{2\sigma^2} - \frac{(x-\mu)^2\Big(\frac{\sigma^2}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma_2^2}\Big)}{2\sigma^2}\Big) dy \\ = & \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big),\end{aligned}$$

also ist die Faltung wieder eine Normalverteilung. Diese hat nun Erwartungswert μ und Varianz σ^2.

6.5 Projektive Familien von Wahrscheinlichkeitsmaßen

Bisher haben wir σ-endliche Maße auf endlichen Produkträumen definiert. Dies ist für die zu behandelnde Wahrscheinlichkeitstheorie nicht ausreichend. Um das einzusehen, sei an den unendlichen Münzwurf erinnert, der schon in der Vorlesung *Stochastik* betrachtet wurde. Hier würde man sagen, dass $\Omega = \{\text{Kopf}, \text{Zahl}\}^{\mathbb{N}}$ und das dazu gehörige Wahrscheinlichkeitsmaß das Produktmaß $\mathbf{P}^{\otimes\infty}$ von $\mathbf{P} = \frac{1}{2}\delta_{\text{Kopf}} + \frac{1}{2}\delta_{\text{Zahl}}$ ist. Dies ist jedoch ein unendliches (aber immerhin doch abzählbares) Produktmaß, dessen Existenz wir noch nicht gezeigt haben. Oftmals ist es auch notwendig, dass wir Maße auf überabzählbaren Produkten betrachten. Ein Großteil der Vorlesung *Stochastische Prozesse* wird solche enthalten. Wir geben hier nun die allgemeine Konstruktion von Wahrscheinlichkeitsmaßen auf Produktmaßen, die auf Kolmogorov zurück geht. Es sei hier erwähnt, dass in dem resultierenden Satz von Kolmogorov (Theorem 6.24) die Voraussetzung getroffen wird, dass Ω polnisch ist.

Definition 6.21 (Projektiver Limes). *1. Sei $(\Omega, \mathcal{F})$ ein Messraum, I eine beliebige Indexmenge und $(\Omega^I, \mathcal{F}^I)$ wie in Definition 6.4. Eine Familie von Wahrscheinlichkeitsmaßen $(\mathbf{P}_J)_{J \Subset I}$ heißt projektive Familie auf $\mathcal{F}$, falls $\mathbf{P}_J$ für alle $J \Subset I$ ein Wahrscheinlichkeitsmaß auf $\mathcal{F}^J$ ist und*

$$\mathbf{P}_H = (\pi^J_H)_* \mathbf{P}_J$$

für alle $H \subseteq J \Subset I$.

2. Existiert für eine projektive Familie $(\mathbf{P}_J)_{J \Subset I}$ von Wahrscheinlichkeitsmaßen auf $\mathcal{F}$ ein Wahrscheinlichkeitsmaß $\mathbf{P}_I$ auf $\mathcal{F}^I$ mit $\mathbf{P}_J = (\pi_J)_* \mathbf{P}_I$ für alle $J \Subset I$, so heißt $\mathbf{P}_I$ projektiver Limes der projektiven Familie. Wir schreiben dann

$$\mathbf{P}_I = \varprojlim_{J \Subset I} \mathbf{P}_J.$$

Beispiel 6.22 (Projektive Limiten und stochastische Prozesse). In mindestens zwei Situationen spielen projektive Familien eine große Rolle.

1. Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und I eine unendliche Indexmenge. In Definition 6.15 haben wir für jedes $J \Subset I$ das Produktmaß $\mathbf{P}^{\otimes J}$ auf $\mathcal{F}^J$ definiert. Die Familie $(\mathbf{P}^{\otimes J})_{J \Subset I}$ ist projektiv. Ist nämlich $H \subseteq J \Subset I$, so ist für $A_i \in \mathcal{F}, i \in H$,

$$\begin{aligned}
(\pi^J_H)_* \mathbf{P}^{\otimes J}\Big(\bigtimes_{i \in H} A_i \Big) &= \mathbf{P}^{\otimes J}\Big((\pi^J_H)^{-1}\Big(\bigtimes_{i \in H} A_i \Big)\Big) \\
&= \mathbf{P}^{\otimes J}\Big(\bigtimes_{i \in H} A_i \times \bigtimes_{i \in J \setminus H} \Omega \Big) \\
&= \prod_{i \in H} \mathbf{P}(A_i) \cdot \prod_{i \in J \setminus H} \mathbf{P}(\Omega) \\
&= \prod_{i \in H} \mathbf{P}(A_i) \\
&= \mathbf{P}^{\otimes H}\Big(\bigtimes_{i \in H} A_i \Big).
\end{aligned}$$

Allerdings haben wir noch nicht gezeigt, wann es den projektiven Limes von $(\mathbf{P}^{\otimes J})_{J \Subset I}$ gibt. Diesen würden wir dann das unendliche Produktmaß $\mathbf{P}^{\otimes I}$ nennen.

2. Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge, $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ ein Messraum und $X_i : \Omega \to \widetilde{\Omega}, i \in I$ Zufallsvariable. Wir werden die Familie $\mathcal{X} := (X_i)_{i\in I}$ einen stochastischen Prozess nennen. Also ist $\mathcal{X} : \Omega \to \widetilde{\Omega}^I$ mit $\mathcal{X}(\omega) = (X_i(\omega))_{i\in I}$. Man kann sich nun fragen, ob die Verteilung von $\mathcal{X}$ (d.h. das Bildmaß $\mathcal{X}_*\mathbf{P}$) als Verteilung auf $\widetilde{\mathcal{F}}^I$ existiert.

   Hierzu sei bemerkt, dass $\widetilde{\mathbf{P}}_J := ((X_j)_{j\in J})_*\mathbf{P}, J \Subset I$ eine projektive Familie ist. Ist nämlich $H \subset J \Subset I$ und $\widetilde{A}_i \in \widetilde{\mathcal{F}}, i \in H$, dann ist

$$\begin{aligned}(\pi^J_H)_*\widetilde{\mathbf{P}}_J\big(\bigtimes_{j\in H} \widetilde{A}_j\big) &= \widetilde{\mathbf{P}}_J\big((\pi^J_H)^{-1} \bigtimes_{j\in H} \widetilde{A}_j\big)\\ &= \widetilde{\mathbf{P}}_J\big(\bigtimes_{j\in H} \widetilde{A}_j \times \bigtimes_{j\in J\setminus H} \widetilde{\Omega}\big)\\ &= \mathbf{P}\big(X_j \in \widetilde{A}_j, j \in H \text{ und } X_j \in \widetilde{\Omega}, j \in J \setminus H\big)\\ &= \mathbf{P}\big(X_j \in \widetilde{A}_j, j \in H\big)\\ &= \widetilde{\mathbf{P}}_H\big(\bigtimes_{j\in H} \widetilde{A}_j\big).\end{aligned}$$

   Wie Theorem 6.24 zeigt, gibt es die Verteilung $\mathcal{X}_*\mathbf{P}$ (was dann der projektive Limes der $(\widetilde{\mathbf{P}}_J)_{J\Subset I}$ ist) zumindest dann, wenn $\widetilde{\mathcal{F}}$ die Borel'sche σ-Algebra eines polnischen Raumes ist.

Bemerkung 6.23 (Eindeutigkeit des projektiven Limes). Zu jeder projektiven Familie $(\mathbf{P}_J)_{J\Subset I}$ gibt es höchstens einen projektiven Limes. Denn: seien $\mathbf{P}_I$ und $\widetilde{\mathbf{P}}_I$ zwei projektive Limiten, so ist für $A := \bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i \in \mathcal{H}$ mit $\mathcal{H}$ aus Lemma 6.7 und $J \Subset I$

$$\mathbf{P}_I(A) = \mathbf{P}_J\Big(\bigtimes_{i\in J} A_i\Big) = \widetilde{\mathbf{P}}_J\Big(\bigtimes_{i\in J} A_i\Big) = \widetilde{\mathbf{P}}_I(A).$$

Damit stimmen $\mathbf{P}_I$ und $\widetilde{\mathbf{P}}_I$ auf dem schnittstabilen Erzeuger überein und nach Proposition 3.10 gilt $\mathbf{P}_I = \widetilde{\mathbf{P}}_I$. Inhalt des nächsten Theorems ist, dass es bei polnischen Räumen genau einen projektiven Limes gibt.

Theorem 6.24 (Existenz von Prozessen, Kolmogorov). *Sei $(\Omega, \mathcal{O})$ polnisch, $\mathcal{F} = \mathcal{B}(\mathcal{O})$ und $(\mathbf{P}_J)_{J\Subset I}$ eine projektive Familie von Wahrscheinlichkeitsmaßen auf $\mathcal{F}$. Dann gibt es den projektiven Limes $\varprojlim_{J\Subset I} \mathbf{P}_J$.*

Beweis. Sei $\mathcal{H}$ wie in Lemma 6.7 und μ eine endlich additive Mengenfunktion auf $\mathcal{H}$, definiert durch die projektive Familie mittels

$$\mu\Big(\bigtimes_{j\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega\Big) := \mathbf{P}_J\Big(\bigtimes_{j\in J} A_j\Big).$$

Nach Lemma 6.7 ist $\mathcal{H}$ ein Halbring und μ ein wohldefinierter Inhalt auf $\mathcal{H}$ ist. Weiter ist

$$\mathcal{K} := \{\bigtimes_{j\in J} K_j \times \bigtimes_{i\in I\setminus J} \Omega : J \Subset I, K_j \text{ kompakt}\} \subseteq \mathcal{H}$$

ein kompaktes System.

Wir zeigen nun, dass μ von innen $\mathcal{K}$-regulär ist. Sei $\varepsilon > 0$, $\times_{i\in J} A_i \times \times_{i\in I\setminus J} \Omega \in \mathcal{H}$ für $J \Subset I$ und $A_i \in \mathcal{F}, i \in J$. Da $\mathbf{P}_j$ für $j \in I$ ein Maß ist, gibt es nach Lemma 3.8 kompakte Mengen $K_j \in \mathcal{F}$ mit $K_j \subseteq A_j$ und $\mathbf{P}_j(A_j \setminus K_j) \le \varepsilon$. Damit ist

$$\begin{aligned}
\mu\Big(\Big(\bigtimes_{i\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega\Big)\setminus\Big(\bigtimes_{i\in J} K_i \times \bigtimes_{i\in I\setminus J} \Omega\Big)\Big) &= \mu\Big(\Big((\bigtimes_{i\in J} A_i) \setminus (\bigtimes_{i\in J} K_i)\Big) \times \bigtimes_{i\in I\setminus J} \Omega\Big) \\
&= \mathbf{P}_J\Big((\bigtimes_{j\in J} A_j) \setminus (\bigtimes_{j\in J} K_j)\Big) \\
&\le \mathbf{P}_J\Big(\bigcup_{j\in J}(A_j \setminus K_j) \times \bigtimes_{i\neq j} \Omega\Big) \\
&\le \sum_{j\in J} \mathbf{P}_J\Big((A_j \setminus K_j) \times \bigtimes_{i\neq j} \Omega\Big) \\
&= \sum_{j\in J} \mathbf{P}_j(A_j \setminus K_j) \\
&\le |J|\varepsilon.
\end{aligned}$$

Da J endlich und $\varepsilon > 0$ beliebig war, ist also μ von innen $\mathcal{K}$-regulär. Nach Theorem 3.9 ist μ σ-additiv. Weiter ist $\mu(\Omega^I) = 1$, also lässt μ sich nach Theorem 3.15 in eindeutiger Weise auf ein Maß $\mathbf{P}$ auf $\sigma(\mathcal{H}) = \mathcal{F}^I$ fortsetzen. Dies muss der projektive Limes von $(\mathbf{P}_J)_{J\Subset I}$ sein. $\square$

Teil II
Grundlagen der Wahrscheinlichkeitstheorie

Wir kommen nun zum eigentlichen Teil der Vorlesung, der Stochastik. Hierzu sei im Folgenden immer $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Das Integral bezüglich $\mathbf{P}$ bezeichnen wir mit $\mathbf{E}[\cdot] := \mathbf{P}[\cdot]$. Weiter kürzen wir $\mathcal{L}^p := \mathcal{L}^p(\mathbf{P})$ ab, falls das nicht zu Verwechslungen führt.

Ziel dieses Abschnittes ist es, die wichtigsten probabilistischen Aussagen bereit zu stellen. Grundlegend hierfür ist sicherlich der Begriff der Zufallsvariable, den wir in Kapitel 7 beleuchten wollen. Oft werden wir den Fall von E-wertigen Zufallsvariablen betrachten, wobei E ein polnischer Raum ist. Die wichtigsten Sätze der Stochastik sind das starke Gesetz der großen Zahlen (Theorem 9.21) und der zentrale Grenzwertsatz (Theorem 11.8). Diese beiden Sätze sind Grenzwertaussagen für Zufallsvariable, wobei wichtig ist, dass die Art der Konvergenz in beiden Sätzen grundverschieden ist. Während das starke Gesetz der großen Zahlen eine fast sichere Konvergenz beschreibt, ist der zentrale Grenzwertsatz eine Aussage über Konvergenz in Verteilung (d.h. über die schwache Konvergenz der Verteilungen der Zufallsvariablen). Konsequenterweise wird es unter anderem darum gehen, Zusammenhänge zwischen verschiedenen Konvergenzarten einzusehen, siehe Kapitel 8 und 10.

7 Zufallsvariable

Meist werden wir reellwertige Zufallsvariable $X : \Omega \to \mathbb{R}$ (d.h. Borel-messbare Funktionen) betrachten. Wir wiederholen nun alles, was wir bereits über Zufallsvariablen wissen und im Folgenden direkt benötigen werden.

7.1 Wiederholung

Viele Begriff fielen bereits im Abschnitt *Maßtheorie*. Die Wichtigsten wiederholen wir kurz. Außerdem steht nun der Zusammenhang zwischen der Maßtheorie und der Vorlesung *Stochastik* im Vordergrund.

Bemerkung 7.1 (Zufallsvariable und deren Verteilung). Sei $(\Omega', \mathcal{F}')$ ein Maßraum.

1. Jede $\mathcal{F}/\mathcal{F}'$-messbare Funktion X heißt (Ω'-wertige) *Zufallsvariable*. Ist $(\Omega', \mathcal{F}') = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, so heißt sie *reellwertig*. Die σ-Algebra $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{F}')\}$ ist *die von X erzeugte σ-Algebra*.

2. Das Wahrscheinlichkeitsmaß $X_*\mathbf{P}$ auf $\mathcal{F}'$ heißt *Verteilung von X*. Ist weiter Y eine Zufallsvariable und $X_*\mathbf{P} = Y_*\mathbf{P}$ (d.h. $\mathbf{P}(X \in A') = \mathbf{P}(Y \in A')$ für alle $A' \in \mathcal{F}'$), so heißen X und Y identisch verteilt und wir schreiben $X \stackrel{d}{=} Y$. Diese Schreibweise ist jedoch mit Vorsicht zu genießen, da man die Gleichheit $X \stackrel{d}{=} Y$ nicht durch Äquivalenzumformungen zu anderen Aussagen erweitern kann. (Etwa gilt $X - Y \stackrel{d}{=} 0$ im Allgemeinen nicht, wenn X und Y identisch verteilt sind.)

3. Für eine Familie $(X_i)_{i\in I}$ von Zufallsvariablen heißt $((X_i)_{i\in I})_*\mathbf{P}$ die *gemeinsame Verteilung von $(X_i)_{i\in I}$*. (Dies ist das Bildmaß unter der Abbildung $(X_i)_{i\in I} : \omega \mapsto (X_i(\omega))_{i\in I}$).

4. Wir werden folgende Redewendung verwenden: *Sei X eine nach $N(\mu, \sigma^2)$-verteilte Zufallsvariable...* Damit ist gemeint, dass $X : \Omega \mapsto \mathbb{R}$ eine messbare Abbildung ist und $X_*\mathbf{P} = \mu_{N(\mu,\sigma^2)}$, siehe Beispiel 3.21. In dieser Situation schreiben wir auch $X \sim N(\mu, \sigma^2)$. Hier bedeutet '$\sim$' *ist so verteilt wie*.

5. Sei μ ein weiteres Maß auf $\mathcal{F}$ und $f : \Omega \to \mathbb{R}$ mit $f \geq 0$ fast überall und $\mu[f] = 1$. Dann hat X genau dann die Dichte f bezüglich μ, wenn $X_*\mathbf{P} = f \cdot \mu$ (siehe Definition 5.12). Dann gilt also für $A \in \mathcal{F}$
$$\mathbf{P}(X \in A) = \mu[f, A].$$
In diesem Fall gilt für $g : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, dass (siehe Lemma 5.13)
$$\mathbf{E}[g(X)] = (X_*\mathbf{P})[g] = (f \cdot \mu)[g] = \mu[fg],$$
falls die rechte Seite existiert.

6. Die Monotonie und Linearität des Integrals bedeutet nun für Zufallsvariable $X, Y \in \mathcal{L}^1$ und $a, b \in \mathbb{R}$:
$$\begin{aligned} &X \leq Y \text{ fast sicher } \Longrightarrow \mathbf{E}[X] \leq \mathbf{E}[Y], \\ &\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]. \end{aligned}$$
Außerdem gilt nach Proposition 4.20
$$\mathbf{E}[X] < \infty \Longrightarrow \mathbf{P}(X < \infty) = 1.$$

Obwohl die σ-Algebra $\mathcal{F}$ gegeben ist, wird im weiteren Verlauf der Vorlesung, insbesondere bei der Einführung der bedingten Erwartung in Kapitel 12, die von X erzeugte σ-Algebra eine besondere Rolle spielen. Einfach gesagt ist eine reellwertige Zufallsvariable Y genau dann $\sigma(X)$-messbar, wenn $Y = \varphi(X)$ für eine Borel-messbare Abbildung φ. Anders ausgedrückt heißt das, dass man den Wert von $Y(\omega)$ kennt, falls man $X(\omega)$ kennt, obwohl man nicht weiß, welchen Wert ω angenommen hat.

Lemma 7.2 (Messbarkeit bezüglich $\boldsymbol{\sigma(X)}$). *Sei $(\Omega', \mathcal{F}')$ ein Messraum und X ein Zufallsvariable mit Werten in Ω' und $Z : \Omega \to \overline{\mathbb{R}}$. Genau dann ist Z $\sigma(X)$-messbar, wenn es eine $\mathcal{F}'/\mathcal{B}(\overline{\mathbb{R}})$-messbare Abbildung $\varphi : \Omega' \to \overline{\mathbb{R}}$ gibt mit $\varphi \circ X = Z$.*

Beweis. '$\Longleftarrow$': klar
'$\Longrightarrow$': Es genügt, den Fall $Z \geq 0$ zu betrachten; ansonsten teilt man $Z = Z^+ - Z^-$ auf. Sei zunächst $Z = 1_A$ für $A \in \sigma(X)$. Dann gibt es ein $A' \in \mathcal{F}'$ mit $X^{-1}(A') = A$, d.h. $Z = 1_{X^{-1}(A')} = 1_{A'} \circ X$, d.h. $\varphi = 1_{A'}$ erfüllt die Aussage. Durch Linearität ist die Aussage auch für einfache Funktionen, d.h. endliche Linearkombinationen von Indikatorfunktionen erfüllt. Im allgemeinen Fall gibt es einfache Funktionen $Z_1, Z_2, \dots \geq 0$ mit $Z_n \uparrow Z$. Hierzu gibt es $\mathcal{F}'$-messbare Funktionen φ_n mit $Z_n = \varphi_n \circ X$. Dann ist $\varphi = \sup_n \varphi_n$ wieder $\mathcal{F}'$-messbar und, da $Z \geq 0$ ist,
$$\varphi \circ X = (\sup_n \varphi_n) \circ X = \sup_n(\varphi_n \circ X) = \sup_n Z_n = Z. \qquad \square$$

Wir wiederholen nun kurz die Konvergenzsätze für Integrale im Kontext von Zufallsvariablen.

Proposition 7.3 (Integral-Konvergenzsätze). *Seien $X, X_1, X_2, \dots$ reellwertige Zufallsvariablen.*

1. Lemma von Fatou, Theorem 4.25: *Es gilt*
$$\liminf_{n\to\infty} \mathbf{E}[X_n] \geq \mathbf{E}[\liminf_{n\to\infty} X_n].$$
2. Satz von der monotonen Konvergenz, Theorem 4.24: *Ist $X_1, X_2, \dots \in \mathcal{L}^1$ und gilt $X_n \uparrow X$ fast sicher, dann ist*
$$\lim_{n\to\infty} \mathbf{E}[X_n] = \mathbf{E}[X],$$
wobei beide Seiten den Wert ∞ annehmen können.
3. Satz von der majorisierten Konvergenz, Theorem 4.26: *Sei $X_n \xrightarrow{n\to\infty} X$ fast sicher und Y eine weitere reellwertige Zufallsvariable mit $|X_1|, |X_2|, \dots \leq Y$ fast sicher und $\mathbf{E}[Y] < \infty$. Dann gilt*
$$\mathbf{E}[X_n] \xrightarrow{n\to\infty} \mathbf{E}[X].$$

Wir sammeln nun bereits bekannte Ungleichungen. Sie helfen oft, Wahrscheinlichkeiten oder Erwartungswerte abzuschätzen. Die meisten Ungleichungen sind schon aus der Vorlesung *Stochastik* bekannt.

Proposition 7.4 (Markov- und Chebyshev-Ungleichung).

1. *Sei X eine Zufallsvariable mit Werten in $\overline{\mathbb{R}}_+$ und $x \in \mathbb{R}_+$. Dann gilt die Markov-Ungleichung*
$$\mathbf{P}(X \geq x) \leq \frac{\mathbf{E}[X]}{x}.$$
2. *Ist X eine reellwertige Zufallsvariable und $p, x \in \mathbb{R}_+$, dann gilt die Chebyshev-Ungleichung*
$$\mathbf{P}(|X| \geq x) \leq \frac{\mathbf{E}[|X|^p]}{x^p}.$$

Beweis. 1. Da X nicht-negativ ist, gilt $x \cdot 1_{X\geq x} \leq X$. Also ist auch
$$x \cdot \mathbf{P}(X \geq x) = \mathbf{E}[x \cdot 1_{X\geq x}] \leq \mathbf{E}[X],$$
woraus die Ungleichung folgt. Die Ungleichung in 2. folgt aus 1. durch
$$\mathbf{P}(|X| \geq x) = \mathbf{P}(|X|^p \geq x^p) \leq \frac{\mathbf{E}[|X|^p]}{x^p}.$$
□

Proposition 7.5 (Minkowski und Hölder-Ungleichung). *Seien X, Y reellwertige Zufallsvariablen.*

1. *Ist $0 < p, q, r \leq \infty$ so, dass $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Dann gilt*
$$\mathbf{E}[|XY|^r]^{1/r} \leq \mathbf{E}[|X|^p]^{1/p} \cdot \mathbf{E}[|Y|^q]^{1/q} \qquad \textit{(Hölder-Ungleichung)}$$
Speziell für $p = q = 2$ ergibt sich
$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^2]^{1/2} \cdot \mathbf{E}[|Y|^2]^{1/2}. \qquad \textit{(Cauchy-Schwartz-Ungleichung)}$$

2. *Für* $1 \le p \le \infty$ *ist*

$$\begin{aligned} \mathbf{E}[|X+Y|^p]^{1/p} &\le \mathbf{E}[|X|^p]^{1/p} + \mathbf{E}[|Y|^p]^{1/p}, \quad 1 \le p \le \infty \qquad \textit{(Minkowski-Ungleichung)} \\ \mathbf{E}[|X+Y|^p] &\le \mathbf{E}[|X|^p] + \mathbf{E}[|Y|^p], \qquad 0 < p < 1 \end{aligned} \tag{7.1}$$

Beweis. Siehe Proposition 5.2 für die Hölder-Ungleichung und die Minkowski-Ungleichung für $1 \le p \le \infty$. Für $0 < p < 1$ ist $x \mapsto x^p$ konkav, also $(x+y)^p = (\frac{2x+2y}{2})^p \le \frac{1}{2}(2x)^p + \frac{1}{2}(2x)^p \le x^p + y^p$ für reelle x, y, woraus die Minkowski-Ungleichung auch im Fall $0 < p < 1$ folgt. $\square$

Proposition 7.6 (Jensen'sche Ungleichung). *Sei I ein offenes Intervall und $X \in \mathcal{L}^1$ mit Werten in I und $\varphi : I \to \mathbb{R}$ konvex.*[14] *Dann gilt*

$$\mathbf{E}[\varphi(X)] \ge \varphi(\mathbf{E}[X]).$$

Beweis. Da φ konvex ist, ist φ stetig und

$$t \mapsto \frac{\varphi(tx + (1-t)y) - \varphi(y)}{t(x-y)}$$

für $y \le x$ monoton fallend. Insbesondere existiert für $y \in I$

$$\lambda(y) := \lim_{x \downarrow y} \frac{\varphi(x) - \varphi(y)}{x - y} = \lim_{t \downarrow 0} \frac{\varphi(tx + (1-t)y) - \varphi(y)}{t(x-y)} \tag{7.2}$$

und es gilt

$$\frac{\varphi(x) - \varphi(y)}{x-y} \ge \lambda(y) \quad \Longrightarrow \quad \varphi(y) + \lambda(y)(x-y) \le \varphi(x) \tag{7.3}$$

für alle $x \in I$. (Für $y > x$ argumentiert man analog wie oben.)

Nun also zum Beweis der Jensen'schen Ungleichung. Da I ein Intervall ist, ist $\mathbf{E}[X] \in I$. Nach (7.3) ist für $x \in I$ mit $y = \mathbf{E}[X]$

$$\varphi(x) \ge \varphi(\mathbf{E}[X]) + \lambda(\mathbf{E}[X])(x - \mathbf{E}[X])$$

und damit

$$\mathbf{E}[\varphi(X)] \ge \varphi(\mathbf{E}[X]) + \lambda(\mathbf{E}[X])\mathbf{E}[X - \mathbf{E}[X]] = \varphi\big(\mathbf{E}[X]\big). \qquad \square$$

Mit der Jensen'schen Ungleichung kann man etwa zeigen, dass $\mathcal{L}^q \subseteq \mathcal{L}^p$ für $p \le q$. Alternativ liest man diese Eigenschaft aus Proposition 5.3 ab.

Lemma 7.7 (p-fach und q-fach integrierbare Zufallsvariable). *Sei $q > 0$ und $X \in \mathcal{L}^q$ eine reelwertige Zufallsvariable. Dann ist für $p \le q$*

$$\mathbf{E}[|X|^p] \le \mathbf{E}[|X^q|]^{p/q}.$$

Insbesondere ist $\mathcal{L}^q \subseteq \mathcal{L}^p$.

Beweis. Die Abbildung $y \mapsto y^{p/q}$ ist konkav auf $\mathbb{R}_+$, also gilt mir der Jensen'schen Ungleichung

$$\mathbf{E}[|X|^p] = \mathbf{E}[(|X|^q)^{p/q}] \le \mathbf{E}[|X|^q]^{p/q}. \qquad \square$$

[14] Eine Abbildung $\varphi : I \to \mathbb{R}$ heißt konvex, falls $\varphi(tx + (1-t)y) \le t\varphi(x) + (1-t)\varphi(y)$ für alle $0 \le t \le 1$ und $x, y \in I$.

7.2 Momente

Aus der Vorlesung *Stochastik* sind bereits Begriffe wie Erwartungswert, Varianz und Covarianz bekannt. Diese wiederholen wir nun. Es gelten alle schon bekannten Rechenregeln. Der einzige Unterschied ist, dass nun $\mathbf{E}[\cdot]$ das Integral bezüglich eines Wahrscheinlichkeitsmaßes ist.

Definition 7.8 (Momente). *Seien X, Y reellwertige Zufallsvariable. Dann heißt, falls existent,* $\mathbf{E}[X]$ *der* Erwartungswert *der Zufallsvariable X. Außerdem ist, falls existent,*

$$\mathbf{V}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2]$$

die Varianz *von X und*

$$\mathbf{COV}[X, Y] := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

die Covarianz *von X und Y. Ist $\mathbf{COV}[X, Y] = 0$, so heißen X und Y unkorreliert. Weiter heißt $\mathbf{E}[X^p]$ für $p > 0$ das p-te Moment von X und $\mathbf{E}[(X - \mathbf{E}[X])^p]$ das zentrierte p-te Moment von X.*

Wir wiederholen hier nur ein paar Eigenschaften.

Proposition 7.9 (Eigenschaften der zweiten Momente). *Seien $X, Y \in \mathcal{L}^2$ reellwertige Zufallsvariable. Dann ist $\mathbf{V}[X], \mathbf{V}[Y], \mathbf{COV}[X, Y] < \infty$ und es gilt*

$$\begin{aligned} \mathbf{V}[X] &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2, \\ \mathbf{COV}[X, Y] &= \mathbf{E}[XY] - \mathbf{E}[X] \cdot \mathbf{E}[Y]. \end{aligned}$$

Außerdem gilt die Cauchy-Schwartz-Ungleichung

$$\mathbf{COV}[X, Y]^2 \leq \mathbf{V}[X] \cdot \mathbf{V}[Y].$$

Sind $X_1, \ldots, X_n \in \mathcal{L}^2$, so gilt die Gleichung von Bienamyé

$$\mathbf{V}\Big[\sum_{k=1}^n X_k\Big] = \sum_{k=1}^n \mathbf{V}[X_k] + 2 \sum_{1 \leq k < l \leq n} \mathbf{COV}[X_k, X_l].$$

Beweis. Da $\mathbf{V}[X] = \mathbf{COV}[X, X]$ genügt es für die erste Aussage, die zweite Gleichung zu zeigen. Diese folgt aus der Linearität des Erwartungswertes mittels

$$\begin{aligned} \mathbf{COV}[X, Y] &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[XY] - \mathbf{E}[\mathbf{E}[X]Y] - \mathbf{E}[X\mathbf{E}[Y]] + \mathbf{E}[X]\mathbf{E}[Y] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]. \end{aligned}$$

Die Cauchy-Schwartz-Ungleichung folgt durch Anwenden von Proposition 7.5 aus die Zufallsvariablen $X - \mathbf{E}[X]$ und $Y - \mathbf{E}[Y]$. Insbesondere ist $\mathbf{COV}[X, Y] < \infty$. Für die Gleichung von Bienamyé sei o.E. $\mathbf{E}[X_k] = 0$, $k = 1, \ldots, n$ (ansonsten geht man zu den Zufallsvariablen $X_k - \mathbf{E}[X_k]$ über). Dann ist

$$\begin{aligned} \mathbf{V}\Big[\sum_{k=1}^n X_k\Big] &= \mathbf{E}\Big[\Big(\sum_{k=1}^n X_k\Big)^2\Big] = \sum_{k=1}^n \sum_{l=1}^n \mathbf{E}[X_k X_l] = \sum_{k=1}^n \mathbf{E}[X_k^2] + 2 \sum_{1 \leq k < l \leq n} \mathbf{E}[X_k X_l] \\ &= \sum_{k=1}^n \mathbf{V}[X_k] + 2 \sum_{1 \leq k < l \leq n} \mathbf{COV}[X_k X_l]. \end{aligned}$$

□

Proposition 7.10 (Momente nicht-negativer Zufallsvariable). *Sei X eine Zufallsvariable mit Werten in $\mathbb{R}_+$. Dann gilt*

$$\mathbf{E}[X^p] = p\int_0^\infty \mathbf{P}(X > t)t^{p-1}dt = p\int_0^\infty \mathbf{P}(X \geq t)t^{p-1}dt.$$

Beweis. Wir verwenden den Satz von Fubini,

$$\mathbf{E}[X^p] = p\mathbf{E}\Big[\int_0^X t^{p-1}dt\Big] = p\int_0^\infty \mathbf{E}\Big[1_{X>t}t^{p-1}\Big]dt = p\int_0^\infty \mathbf{P}(X > t)t^{p-1}dt.$$

Der Beweis der zweiten Gleichung ist analog. □

7.3 Charakteristische Funktionen und Laplace-Transformierte

Wir führen nun Erwartungswerte bestimmter Funktionen von Zufallsvariablen ein. Die daraus resultierenden Funktionen sind die charakteristische Funktion (der Verteilung reellwertiger Zufallsvariablen) und die Laplace-Transformierte (der Verteilung nicht-negativer Zufallsvariablen). Die Nützlichkeit dieser beiden Funktionen ist darauf zurückzuführen, dass man mit ihrer Hilfe leicht die Momente der Verteilungen der Zufallsvariablen berechnen kann (siehe Proposition 7.14). Außerdem werden wir später in Proposition 10.25 zeigen, dass diese Funktionen verteilungsbestimmend sind.

Definition 7.11 (Charakteristische Funktion und Laplace-Transformierte).

1. *Die* charakteristische Funktion *einer Zufallsvariablen X mit Werten in $\mathbb{R}^d$ ist gegeben durch*
$$\psi_X := \psi_{X_*\mathbf{P}} := \begin{cases} \mathbb{R}^d & \to \mathbb{C}, \\ t & \mapsto \mathbf{E}[e^{itX}] := \mathbf{E}[\cos(tX)] + i\mathbf{E}[\sin(tX)], \end{cases}$$
wobei $tx := \langle t, x\rangle$ das Skalarprodukt in $\mathbb{R}^d$ ist.

2. *Die* Laplace-Transformierte *von X ist gegeben durch*
$$\mathscr{L}_X := \mathscr{L}_{X_*\mathbf{P}} := \begin{cases} \mathbb{R}^d & \to \mathbb{R}, \\ t & \mapsto \mathbf{E}[e^{-tX}], \end{cases}$$
gegeben das Integral auf der rechten Seite existiert. Diese wird meistens für Wahrscheinlichkeitsmaße auf $\mathbb{R}_+^d$ betrachtet.

In diesem einführenden Teil über charakteristische Funktionen und Laplace-Transformierte werden wir nur einige wichtige Eigenschaften herleiten.

Proposition 7.12 (Eigenschaften von charakteristischen Funktionen). *Seien X, Y Zufallsvariable mit Werten in $\mathbb{R}^d$ und charakteristischen Funktionen ψ_X, ψ_Y. Dann gilt*

1. *$|\psi_X(t)| \leq 1$ für jedes $t \in \mathbb{R}^d$ und $\psi_X(0) = 1$.*
2. *ψ_X ist gleichmäßig stetig.*
3. *$\psi_{aX+b}(t) = \psi_X(at)e^{ibt}$ für alle $a \in \mathbb{R}, b \in \mathbb{R}^d$.*

Beweis. 1. ist klar. Für die gleichmäßige Stetigkeit sei erwähnt, dass

$$
\begin{aligned}
|e^{ihx} - 1| &= \sqrt{|\cos(hx) + i\sin(hx) - 1|^2} = \sqrt{(\cos(hx) - 1)^2 + \sin(hx)^2} \\
&= \sqrt{2(1 - \cos(hx))} = 2|\sin(hx/2)| \leq |hx| \wedge 2.
\end{aligned}
$$

Damit gilt 2. wegen

$$
\begin{aligned}
\sup_{t\in\mathbb{R}^d} |\psi_X(t+h) - \psi_X(t)| &= \sup_{t\in\mathbb{R}^d} |\mathbf{E}[e^{i(t+h)X} - e^{itX}]| = \sup_{t\in\mathbb{R}^d} |\mathbf{E}[e^{itX}(e^{ihX} - 1)]| \\
&\leq \mathbf{E}[|e^{ihX} - 1|] \leq \mathbf{E}[|hX| \wedge 2] \xrightarrow{h\to 0} 0.
\end{aligned}
$$

Für 3. berechnen wir

$$\mathbf{E}[e^{it(aX+b)}] = e^{itb}\mathbf{E}[e^{i(at)X}] = e^{itb}\psi_X(at).$$

□

Beispiel 7.13 (Beispiele für charakteristische Funktionen). 1. Die charakteristische Funktion einer nach $B(n,p)$ verteilten Zufallsvariable X ist gegeben durch

$$\psi_{B(n,p)}(t) = (1 - p + pe^{it})^n.$$

Nach Definition gilt nämlich

$$\mathbf{E}[e^{itX}] = \sum_{k=1}^{n} \binom{n}{k} p^k(1-p)^{n-k}e^{itk} = (1 - p + pe^{it})^n.$$

2. Die charakteristische Funktion einer nach Poi(γ)-verteilten Zufallsvariable ist gegeben durch

$$\psi_{\mathrm{Poi}(\gamma)} = e^{\gamma(e^{it}-1)},$$

denn

$$\psi_{\mathrm{Poi}(\gamma)} = e^{-\gamma}\sum_{n=1}^{\infty} \frac{\gamma^n e^{itn}}{n!} = e^{\gamma(e^{it}-1)}.$$

3. Die charakteristische Funktion einer nach $N(\mu,\sigma^2)$-verteilten Zufallsvariable X ist gegeben durch

$$\psi_{N(\mu,\sigma^2)}(t) = e^{it\mu}e^{-\sigma^2 t^2/2}.$$

Nach Proposition 7.12.2 genügt es, diese Behauptung für $\mu = 0, \sigma^2 = 1$ nachzurechnen. Für diesen Fall gilt mittels partieller Integration

$$\frac{d}{dt}\psi_{N(0,1)}(t) = \frac{i}{\sqrt{2\pi}}\int xe^{-x^2/2}e^{itx}dx = \frac{i}{\sqrt{2\pi}}\int e^{-x^2/2}ite^{itx}dx = -t\psi_{N(0,1)}(t).$$

Diese Differentialgleichung mit $\psi_{N(0,1)}(0) = 1$ hat die eindeutige Lösung $\psi_{N(0,1)}(t) = e^{-t^2/2}$.

4. Die Laplace-Transformierte einer nach $\exp(\gamma)$-verteilten Zufallsvariablen X ist gegeben durch

$$\mathscr{L}_{\exp(\gamma)}(t) = \frac{\gamma}{\gamma + t}.$$

Es ist nämlich

$$\mathbf{E}[e^{-tX}] = \int_0^\infty \gamma e^{-\gamma x}e^{-tx}dx = \frac{\gamma}{\gamma + t}.$$

Oftmals sind charakteristische Funktionen und Laplace-Transformierte ein einfaches Hilfsmittel, um Momente von Zufallsvariablen zu berechnen.

Proposition 7.14 (Charakteristische Funktion und Momente). *Sei X eine Zufallsvariable mit Werten in $\mathbb{R}$.*

1. *Ist $X \in \mathcal{L}^p$, so ist ψ_X p-mal stetig differenzierbar und für $k = 0, \dots, p$ gilt*
$$\psi_X^{(k)}(t) = \mathbf{E}[(iX)^k e^{itX}].$$
Insbesondere ist $\psi_X^{(k)}(0) = i^k \mathbf{E}[X^k]$.
2. *Ist speziell $X \in \mathcal{L}^2$, so ist*
$$\psi_X(t) = 1 + it\mathbf{E}[X] - \frac{t^2}{2}\mathbf{E}[X^2] + \varepsilon(t)t^2$$
mit $\varepsilon(t) \xrightarrow{t\to 0} 0$.

Beweis. 1. Mit $|X|^p$ ist auch $|X|^p \vee 1$ integrierbar. Damit haben alle $|X|^k$ eine integrierbare Majorante und die rechte Seite existiert. Da die Aussage offensichtlich für $k = 0$ gilt, nehmen wir an, dass sie für ein $k < n$ gilt. Dann ist
$$\Big|\frac{d^{k+1}}{dt^{k+1}} e^{itX}\Big| = \lim_{h\to 0}\Big|\frac{(iX)^k e^{i(t+h)X} - (iX)^k e^{itX}}{h}\Big| \le \Big|X^k \frac{e^{ihX} - 1}{h}\Big| \le \big|X^{k+1}\big|.$$
Wegen majorisierter Konvergenz darf man Ableitung und Integral vertauschen und es folgt
$$\psi_X^{(k+1)}(t) = \mathbf{E}\Big[\frac{d}{dt}(iX)^k e^{itX}\Big] = \mathbf{E}[(iX)^{k+1} e^{itX}].$$
Die Stetigkeit der Ableitung folgt ebenso mit majorisierter Konvergenz.
2. Für die Abschätzung benötigen wir die Taylorentwicklung von ψ_X mit Restglied. Es gilt
$$e^{itX} = 1 + itX - \frac{t^2X^2}{2}(\cos(\theta_1 tX) + i\sin(\theta_2 tX))$$
mit Zufallszahlen θ_1, θ_2, so dass $|\theta_1|, |\theta_2| \le 1$. Deshalb bekommen wir
$$\psi_X(t) = 1 + it\mathbf{E}[X] - \frac{t^2}{2}\mathbf{E}[X^2] + \varepsilon(t)t^2$$
mit $2\varepsilon(t) = \mathbf{E}[X^2(1 - \cos(\theta_1 tX) + i\sin(\theta_2 tX))] \xrightarrow{t\to 0} 0$ aus majorisierter Konvergenz. □

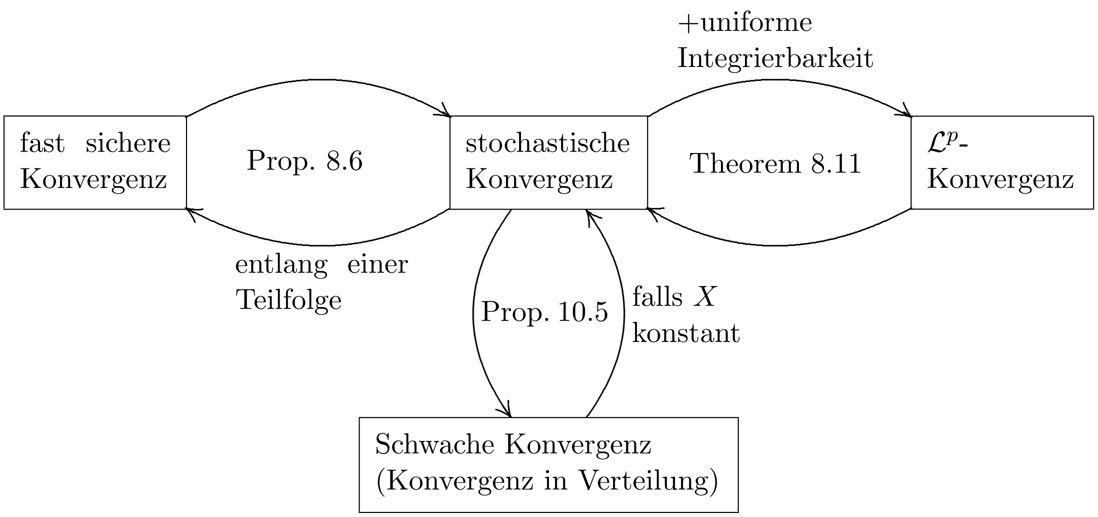
Beispiel 7.15 (Momente der Exponential- und Normalverteilung). 1. Sei X eine nach $\exp(\gamma)$-verteile Zufallsvariable. Wir haben bereits die Laplace-Transformierte von X, $\mathscr{L}_{\exp(\gamma)}(t) = \gamma/(\gamma + t)$, berechnet. Daraus ergeben sich leicht alle Momente von X, nämlich
$$\mathbf{E}[X^n] = (-1)^n \frac{d^n}{dt^n}\mathbf{E}[e^{-tX}]\Big|_{t=0} = (-1)^n \frac{d^n}{dt^n}\frac{\gamma}{\gamma + t}\Big|_{t=0} = \frac{n!\gamma}{(\gamma + t)^{n+1}}\Big|_{t=0} = \frac{n!}{\gamma^n}.$$

2. Für eine nach $N(\mu, \sigma^2)$-verteilte Zufallsvariable X kennen wir bereits die charakteristische Funktion $\psi_{N(\mu,\sigma^2)}(t) = e^{it\mu - \sigma^2 t^2/2}$. Für kleine t entwickeln wir dies mit
$$\psi_{N(\mu,\sigma^2)}(t) = 1 + it\mu - \sigma^2 t^2/2 - \mu^2 t^2/2 + \varepsilon(t)t^2$$
mit $\varepsilon(t) \xrightarrow{t\to 0} 0$. Daraus liest man mittels Proposition 7.14.2 ab, dass
$$\mathbf{E}[X] = \mu, \qquad \mathbf{V}[X] = \mathbf{E}[X^2] - \mu^2 = \sigma^2.$$

8 Fast sichere, stochastische und $\mathcal{L}^p$-Konvergenz

Aus der Analysis ist bereits bekannt, dass es für Funktionenfolgen verschiedene Konvergenzarten gibt, etwa die gleichmäßige und die punktweise Konvergenz. Wir wollen nun die wichtigsten Konvergenzarten besprechen, bezüglich derer Zufallsvariable konvergieren können.

Neben der fast sicheren Konvergenz werden wir die stochastische Konvergenz und die $\mathcal{L}^p$-Konvergenz (siehe auch Abschnitt 5) kennenlernen. Im Abschnitt 10 werden wir außerdem die Konvergenz in Verteilung (was dasselbe ist wie die schwache Konvergenz der Verteilungen der Zufallsvariablen) kennen lernen. Folgendes Schaubild fasst alle Konvergenzarten zusammen:



8.1 Definition und Beispiele

Wir beginnen mit einigen Definitionen.

Definition 8.1 (Fast sichere und stochastische Konvergenz). *Seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit Werten in einem metrischen Raum (E, r).*

1. *Ist*
$$\mathbf{P}(\lim_{n\to\infty} r(X_n, X) = 0) = 1,$$
sagen wir, dass die Folge $X_1, X_2, \ldots$ fast sicher gegen X konvergiert und schreiben $X_n \xrightarrow{n\to\infty}_{fs} X$.

2. *Ist*
$$\lim_{n\to\infty} \mathbf{P}(r(X_n, X) > \varepsilon) = 0,$$
für alle $\varepsilon > 0$, sagen wir, dass die Folge $X_1, X_2, \ldots$ in Wahrscheinlichkeit (oder stochastisch) gegen X konvergiert und schreiben $X_n \xrightarrow{n\to\infty}_p X$.

3. *Sind die Zufallsvariablen reellwertig und ist für ein $p > 0$*
$$\lim_{n\to\infty} \mathbf{E}[|X_n - X|^p] = 0,$$
sagen wir, dass die Folge $X_1, X_2, \ldots$ in $\mathcal{L}^p$ (oder im p-ten Mittel) gegen X konvergiert und schreiben auch $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$.

Bemerkung 8.2 (Eigenschaften der $\mathcal{L}^p$-Konvergenz). Wir wissen aus Abschnitt 5 schon einiges über die $\mathcal{L}^p$-Konvergenz. Ist etwa $X, X_1, X_2, \dots$ so, dass $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^q} X$ und $p < q$, dann gilt auch $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$ nach Proposition 5.6. Außerdem sind die Räume $\mathcal{L}^p$ vollständig nach Proposition 5.7. Gibt es also für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$, so dass für alle $m, n \geq n$

$$\mathbf{E}[|X_n - X_m|^p] < \varepsilon,$$

so gibt es eine Zufallsvariable $X \in \mathcal{L}^p$ mit $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$.

Beispiel 8.3 (Gegenbeispiele). Betrachten wir uns das Schaubild am Anfang des Kapitels, stellen wir fest, dass aus der fast sicheren zwar die stochastische Konvergenz folgt, jedoch nicht umgekehrt. Außerdem folgt zwar aus der $\mathcal{L}^1$-Konvergenz die stochastische, jedoch folgt nicht einmal aus der fast sicheren Konvergenz die $\mathcal{L}^1$-Konvergenz. Wir geben zunächst zwei Beispiele für diese beiden Fälle.

1. Aus der stochastischen Konvergenz folgt nicht die fast sichere: Sei U eine auf $[0, 1]$ uniform verteilte Zufallsvariable. Weiter setzen wir

$$\begin{aligned}
&A_1 = [0, \tfrac{1}{2}], \quad && A_2 = [\tfrac{1}{2}, 1], && && \\
&A_3 = [0, \tfrac{1}{4}], && A_4 = [\tfrac{1}{4}, \tfrac{2}{4}], \quad && A_5 = [\tfrac{2}{4}, \tfrac{3}{4}], \quad && A_6 = [\tfrac{3}{4}, 1], \\
&\dots
\end{aligned}$$

und $X_n := 1_{U \in A_n}$. Dann gilt klar für $0 < \varepsilon < 1$

$$\lim_{n\to\infty} \mathbf{P}(|X_n| > \varepsilon) = \lim_{n\to\infty} \mathbf{P}(U \in A_n) = 0,$$

d.h. $X_n \xrightarrow{n\to\infty}_p 0$, jedoch gibt es für jedes $n \in \mathbb{N}$ eine $m > n$ mit $X_m = 1$. Deshalb konvergiert die Folge $X_1, X_2, \dots$ nicht fast sicher gegen 0.

2. Aus der fast sicheren folgt nicht die $\mathcal{L}^1$-Konvergenz: Sei wieder U eine auf $[0, 1]$ uniform verteilte Zufallsvariable. Weiter ist $B_n = [0, \frac{1}{n}]$ und $Y_n = n \cdot 1_{U \in B_n}$. Dann ist $Y_n \xrightarrow{n\to\infty}_{fs} Y = \infty \cdot 1_{U=0}$, also $Y = 0$ fast sicher. Andererseits ist

$$\mathbf{E}[Y_n] = n \cdot \mathbf{P}(U \in A_n) = 1,$$

also konvergiert $Y_1, Y_2, \dots$ nicht in $\mathcal{L}^1$ gegen 0.

Lemma 8.4 (Stochastischer Limes eindeutig). *Sei $X, Y, X_1, X_2, \dots$ Zufallsvariable mit Werte in einem metrischen Raum (E, r) und $X_n \xrightarrow{n\to\infty}_p X$ sowie $X_n \xrightarrow{n\to\infty}_p Y$. Dann ist $X = Y$ fast sicher.*

Beweis. Es gilt für alle $\varepsilon > 0$

$$\mathbf{P}(r(X, Y) > 2\varepsilon) \leq \mathbf{P}(r(X_n, X) > \varepsilon \text{ oder } r(X_n, Y) > \varepsilon) \xrightarrow{n\to\infty} 0.$$

Also folgt

$$\mathbf{P}(X \neq Y) = \mathbf{P}\Big(\bigcup_{k=1}^{\infty} \Big\{r(X, Y) > 1/k\Big\}\Big) \leq \sum_{k=1}^{\infty} \mathbf{P}(r(X, Y) > 1/k) = 0,$$

woraus die Aussage folgt. $\square$

8.2 Fast sichere und stochastische Konvergenz

Wir zeigen nun ein Resultat, das fast sichere und Konvergenz in Wahrscheinlichkeit in Beziehung setzt.

Lemma 8.5 (Charakterisierung von stochastischer Konvergenz). *Seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit Werten in einem metrischen Raum (E, r). Dann gilt*

$$X_n \xrightarrow{n\to\infty}_p X \qquad \iff \qquad \mathbf{E}[r(X_n, X) \wedge 1] \xrightarrow{n\to\infty} 0. \tag{8.1}$$

Beweis. Falls $X_n \xrightarrow{n\to\infty}_p X$, so gilt für alle $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{E}[r(X_n, X) \wedge 1] \le \lim_{n\to\infty} \big(\varepsilon + \mathbf{P}(r(X_n, X) > \varepsilon)\big) = \varepsilon,$$

womit die rechte Seite gezeigt ist. Gilt hingegen die rechte Seite, folgt mit der Chebyshev-Ungleichung für $0 < \varepsilon \le 1$, dass

$$\mathbf{P}(r(X_n, X) > \varepsilon) \le \frac{\mathbf{E}[r(X_n, X) \wedge 1]}{\varepsilon} \xrightarrow{n\to\infty} 0. \qquad \square$$

Proposition 8.6 (Konvergenz in Wahrscheinlichkeit und fast sichere Konvergenz). *Seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit Werten in einem metrischen Raum (Ω', r). Dann sind äquivalent:*

1. $X_n \xrightarrow{n\to\infty}_p X$
2. *Für jede Folge $(n_k)_{k=1,2,\ldots}$ gibt es eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\ldots}$ mit $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty}_{fs} X$.*

Insbesondere gilt

$$X_n \xrightarrow{n\to\infty}_{fs} X \quad \Longrightarrow \quad X_n \xrightarrow{n\to\infty}_p X.$$

Beweis. 1. ⇒ 2.: Wegen (8.1) können wir für jede Folge $(n_k)_{k=1,2,\ldots}$ eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\ldots}$ wählen, so dass

$$\mathbf{E}\Big[\sum_{\ell=1}^\infty (r(X_{n_{k_\ell}}, X) \wedge 1)\Big] = \sum_{\ell=1}^\infty \mathbf{E}[r(X_{n_{k_\ell}}, X) \wedge 1] < \infty.$$

Das erste Gleichheitszeichen gilt dabei wegen monotoner Konvergenz. Damit ist

$$1 = \mathbf{P}\Big(\sum_{\ell=1}^\infty (r(X_{n_{k_\ell}}, X) \wedge 1) < \infty\Big) \le \mathbf{P}\big(\limsup_{\ell\to\infty} r(X_{n_{k_\ell}}, X) = 0\big) \le 1,$$

d.h. $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty}_{fs} X$.
2. ⇒ 1.: Nehmen wir an, dass 1. nicht gilt. Wegen (8.1) gibt es ein $\varepsilon > 0$ und eine Teilfolge $(n_k)_{k=1,2,\ldots}$, so dass $\lim_{k\to\infty} \mathbf{E}[r(X_{n_k}, X) \wedge 1] > \varepsilon$. Angenommen, es gäbe nun eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\ldots}$, so dass $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty} X$ fast sicher. Dann wäre auch

$$\lim_{\ell\to\infty} \mathbf{E}[r(X_{n_{k_\ell}}, X) \wedge 1] = \mathbf{E}\big[\lim_{\ell\to\infty} r(X_{n_{k_\ell}}, X) \wedge 1\big] = 0$$

wegen majorisierter Konvergenz, also ein Widerspruch. Also haben wir eine Folge $(n_k)_{k=1,2,\ldots}$ gefunden, für die es keine weitere Teilfolge $(n_{k_\ell})_{\ell=1,2,\ldots}$ gibt mit $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty}_{fs} X$, also haben wir gezeigt, dass 2. nicht gilt. $\square$

8.3 Stochastische und $\mathcal{L}^p$-Konvergenz

In Beispiel 8.3 hatten wir bereits gesehen, dass die fast sichere Konvergenz (und damit auch die stochastische Konvergenz) nicht die $\mathcal{L}^1$-Konvergenz impliziert. Das erstaunt nicht, da ja der Satz von der majorisierten Konvergenz besagt, dass eine Folge $X_1, X_2, \dots$, die fast sicher gegen X konvergiert und eine integrierbare Majorante besitzt auch in $\mathcal{L}^1$ gegen X konvergiert. Würde die fast sichere Konvergenz die $\mathcal{L}^1$-Konvergenz implizieren, bräuchte man die Forderung einer integrierbaren Majorante nicht zu machen. Wir wollen im Folgenden die Bedingung der integrierbaren Majorante für die $\mathcal{L}^1$-Konvergenz abschwächen. Siehe Theorem 8.11 und Korollar 8.12. Der Begriff der gleichgradigen Integrierbarkeit ist hierfür zentral, siehe Definition 8.7.

Definition 8.7 (Gleichgradige Integrierbarkeit). *Eine Familie $(X_i)_{i\in I}$ heißt* gleichgradig integrierbar*, falls*

$$\inf_K \sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] = 0$$

Beispiel 8.8 (Gleichgradige Integrierbarkeit).

1. Sei $Y \in \mathcal{L}^1$ und $(X_i)_{i\in I}$ mit $\sup_i |X_i| < |Y|$. Dann ist $(X_i)_{i\in I}$ gleichgradig integrierbar. Denn
$$\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] \le \mathbf{E}[|Y|; |Y| > K] \xrightarrow{K\to\infty} 0$$
nach majorisierter Konvergenz. Insbesondere ist jedes $Y \in \mathcal{L}^1$ gleichgradig integrierbar.

2. Jede endliche Familie $(X_i)_{i=1,\dots,n}$ mit $X_i \in \mathcal{L}^1, i = 1, \dots, n$ ist gleichgradig integrierbar, denn $\sup_{1\le i\le n} |X_i| \in \mathcal{L}^1$ und
$$\inf_K \sup_{1\le i\le n} \mathbf{E}[|X_i|; |X_i| > K] \le \inf_K \mathbf{E}[\sup_{1\le i\le n} |X_i|; \sup_{1\le i\le n} |X_i| > K] = 0$$
wegen majorisierter Konvergenz.

3. Betrachten wir das Beispiel 8.3.2. Hier ist für $n > K$
$$\mathbf{E}[|Y_n|; |Y_n| > K] = \mathbf{E}[Y_n] = 1.$$
Insbesondere ist $(Y_n)_{n=1,2,\dots}$ nicht gleichgradig integrierbar.

4. Sei $p > 1$. Dann ist $(X_i)_{i\in I}$ mit $X_i \in \mathcal{L}^p, i \in I$ gleichgradig integrierbar, falls $\sup_{i\in I} ||X_i||_p < \infty$. Denn es gilt mit der Markov-Ungleichung (Proposition 7.4)
$$\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] \le \sup_{i\in I} \frac{\mathbf{E}[|X_i|^p]}{K^{p-1}} \xrightarrow{K\to\infty} 0.$$

Lemma 8.9 (Charakterisierung von gleichgradiger Integrierbarkeit). *Sei $(X_i)_{i\in I}$ eine Familie von Zufallsvariablen. Dann sind äquivalent:*

1. *$(X_i)_{i\in I}$ ist gleichgradig integrierbar*
2. *Es gilt*
$$\sup_{i\in I} \mathbf{E}[|X_i|] < \infty \quad \textit{und} \quad \lim_{\varepsilon\to 0} \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] = 0.$$

3. *Es gilt*
$$\lim_{K\to\infty} \sup_{i\in I} \mathbf{E}[(|X_i| - K)^+] = 0.$$

4. *Es gibt eine Funktion* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *so, dass* $\frac{f(x)}{x} \xrightarrow{x\to\infty} \infty$ *und* $\sup_{i\in I} \mathbf{E}[f(|X_i|)] < \infty$.

Gilt eine der vier Aussagen, so kann die Funktion f in 4. monoton wachsend und konvex gewählt werden.

Beweis. '1. ⇒ 2.': Sei $\delta > 0$ gegeben und $K = K_\delta$ so, dass $\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] \le \delta$. Dann ist für $A \in \mathcal{F}$
$$\mathbf{E}[|X_i|; A] = \mathbf{E}[|X_i|; A \cap \{|X_i| > K\}] + \mathbf{E}[|X_i|; A \cap \{|X_i| \le K\}] \le \delta + K \cdot \mathbf{P}(A).$$

Insbesondere ist
$$\sup_{i\in I} \mathbf{E}[|X_i|] = \sup_{i\in I} \mathbf{E}[|X_i|; \Omega] \le \delta + K < \infty$$
und
$$\sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] \le \delta + K\varepsilon \xrightarrow{\varepsilon\to 0} \delta.$$
Da $\delta > 0$ beliebig war, muss
$$\lim_{\varepsilon\to 0} \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] = 0$$
gelten.
'2. ⇒ 3.': Zunächst bemerken wir, dass $(|X_i| - K)^+ \le |X_i| 1_{|X_i|\ge K}$. Sei $\varepsilon > 0$. Wähle $K = K_\varepsilon$ groß genug, so dass nach der Markov-Ungleichung
$$\sup_{i\in I} \mathbf{P}(|X_i| > K) \le \sup_{i\in I} \frac{\mathbf{E}[|X_i|]}{K} < \varepsilon$$
ist. Damit folgt 3. aus
$$\begin{aligned}\lim_{K\to\infty} \sup_{i\in I} \mathbf{E}[(|X_i| - K)^+] &= \lim_{\varepsilon\to 0} \sup_{i\in I} \mathbf{E}[(|X_i| - K_\varepsilon)^+] \le \lim_{\varepsilon\to 0} \sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K_\varepsilon] \\ &\le \lim_{\varepsilon\to 0} \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] = 0.\end{aligned}$$

'3. ⇒ 4.': Es gibt eine Folge $K_1, K_2, \ldots$ mit $K_n \uparrow \infty$ und $\sup_{i\in I} \mathbf{E}[(|X_i| - K_n)^+] \le 2^{-n}$. Wir setzen
$$f(x) := \sum_{n=1}^{\infty} (x - K_n)^+.$$

Dann ist f monoton wachsend und als Summe konvexer Funktionen wieder konvex. Außerdem ist für $x \ge 2K_n$
$$\frac{f(x)}{x} \ge \sum_{k=1}^{n} \Big(1 - \frac{K_k}{x}\Big) \ge \frac{n}{2},$$

also $\frac{f(x)}{x} \xrightarrow{x\to\infty} \infty$. Wegen monotoner Konvergenz gilt außerdem

$$\mathbf{E}[f(|X_i|)] = \sum_{n=1}^{\infty} \mathbf{E}[(|X_i| - K_n)^+] \le \sum_{n=1}^{\infty} 2^{-n} = 1.$$

'4. ⇒ 1.': Setze $a_K := \inf_{x\ge K} \frac{f(x)}{x}$, so dass $a_K \xrightarrow{K\to\infty} \infty$. Also ist

$$\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| \ge K] \le \frac{1}{a_K} \sup_{i\in I} \mathbf{E}[f(|X_i|); |X_i| \ge K] \le \frac{1}{a_K} \sup_{i\in I} \mathbf{E}[f(|X_i|)] \xrightarrow{K\to\infty} 0. \qquad \square$$

Beispiel 8.10 (Differenz und gleichgradige Integrierbarkeit). Für $X \in \mathcal{L}^1$ ist $(X_i)_{i\in I}$ genau dann gleichgradig integrierbar, wenn $(X_i - X)_{i\in I}$ gleichgradig integrierbar ist.

Um dies zu sehen, sei $(X_i)_{i\in I}$ gleichgradig integrierbar. Nach Beispiel 8.8.2 ist X gleichgradig integrierbar. Außerdem ist

$$\sup_{i\in I} \mathbf{E}[|X_i - X|] \le \mathbf{E}[|X|] + \sup_{i\in I} \mathbf{E}[|X_i|] < \infty$$

und es gilt

$$\sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i - X|; A] \le \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] + \sup_{A:\mathbf{P}(A)<\varepsilon} \mathbf{E}[|X|; A] \xrightarrow{\varepsilon\to 0} 0,$$

d.h. nach Lemma 8.9 ist $(X_i - X)_{i\in I}$ gleichgradig integrierbar. Die Umkehrung folgt analog.

Theorem 8.11 (Stochastische und Konvergenz im p-ten Mittel). *Sei $X_1, X_2, \ldots$ eine Folge in $\mathcal{L}^p$ mit $0 < p < \infty$. Folgende Aussagen sind äquivalent:*

1. *Es gibt eine messbare Funktion $X \in \mathcal{L}^p$ mit $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$.*
2. *Die Familie $(|X_i|^p)_{i=1,2,\ldots}$ ist gleichgradig integrierbar und es gibt eine messbare Funktion X mit $X_n \xrightarrow{n\to\infty} X$ stochastisch.*

Gilt 1. oder 2. dann stimmen die Limiten überein.

Beweis. 1.⇒ 2.: Zunächst ist wegen der Chebyshev'schen Ungleichung für jedes $\varepsilon > 0$

$$\mathbf{P}(|X_n - X| > \varepsilon) \le \frac{\mathbf{E}[|X_n - X|^p]}{\varepsilon^p} = \frac{||X_n - X||_p^p}{\varepsilon^p} \xrightarrow{n\to\infty} 0,$$

d.h. die stochastische Konvergenz gilt. Für den Beweis der gleichgradigen Integrierbarkeit verwenden wir Lemma 8.9. Sei $\varepsilon > 0$ und $N = N_\varepsilon$ so, dass $||X_n - X||_p < \varepsilon$ für $n \ge N$. Dann ist mit der Minkowski-Ungleichung (7.1)

$$\begin{aligned}\sup_{n\in\mathbb{N}} (\mathbf{E}[|X_n|^p])^{1\wedge 1/p} &= \sup_{n\in\mathbb{N}} ||X_n||_p^{p\wedge 1} \le \sup_{n<N} ||X_n||_p^{p\wedge 1} + \sup_{n\ge N} ||X_n - X||_p^{p\wedge 1} + ||X||_p^{p\wedge 1} \\ &\le \sup_{n<N} (\mathbf{E}[|X_n|^p])^{1\wedge 1/p} + \varepsilon^{p\wedge 1} + (\mathbf{E}[|X|^p])^{1\wedge 1/p} < \infty,\end{aligned}$$

und für $A \in \mathcal{F}$, wieder mit der Minkowski Ungleichung

$$\begin{aligned}\sup_{n\in\mathbb{N}}(\mathbf{E}[|X_n|^p; A])^{1\wedge 1/p} &= \sup_{n\in\mathbb{N}} ||X_n 1_A||_p^{p\wedge 1}\\ &\leq \sup_{n<N} ||X_n 1_A||_p^{p\wedge 1} + \sup_{n\geq N} ||(X_n - X)1_A||_p^{p\wedge 1} + ||X 1_A||_p^{p\wedge 1}\\ &\leq \sup_{n<N} (\mathbf{E}[|X_n|^p; A])^{1\wedge 1/p} + \varepsilon^{p\wedge 1} + (\mathbf{E}[|X|^p; A])^{1\wedge 1/p}.\end{aligned}$$

Da N endlich ist, folgt

$$\lim_{\delta\to 0} \sup_{A:\mathbf{P}(A)<\delta} \sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|^p; A] \leq \varepsilon^p.$$

Weil $\varepsilon > 0$ beliebig war, folgt die Behauptung.
2.$\Rightarrow$ 1.: Da $X_n \xrightarrow{n\to\infty}_p X$, gibt es nach Proposition 8.6 eine Teilfolge $n_1, n_2, \dots$ mit $X_{n_k} \xrightarrow{k\to\infty} X$ fast sicher. Mit dem Lemma von Fatou ist

$$\mathbf{E}[|X|^p] = \mathbf{E}[\liminf_{k\to\infty} |X_{n_k}|^p] \leq \sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|^p] < \infty$$

wegen Lemma 8.9. Insbesondere ist $X \in \mathcal{L}^p$. Genau wie in Beispiel 8.10 ist auch $\{|X_n - X|^p : n \in \mathbb{N}\}$ gleichgradig integrierbar. Für jedes $\delta > 0$ gilt wegen der stochastischen Konvergenz

$$\mathbf{P}(|X_n - X| > \delta) \xrightarrow{n\to\infty} 0.$$

Aus Lemma 8.9 folgt nun mit majorisierter Konvergenz

$$\lim_{n\to\infty} \mathbf{E}[|X_n - X|^p] = \lim_{n\to\infty} \mathbf{E}[|X_n - X|^p; |X_n - X| > \delta] + \mathbf{E}[|X_n - X|^p; |X_n - X| \leq \delta] \leq \delta^p.$$

Da $\delta > 0$ beliebig war, folgt $X_n \xrightarrow{n\to\infty} X$ in $\mathcal{L}^p$. $\square$

Korollar 8.12 (Erwartungswert-Konvergenz und gleichgradige Integrierbarkeit). *Sei $0 < p < \infty$ und $X_1, X_2, \dots \in \mathcal{L}^p$ und X messbar mit $X_n \xrightarrow{n\to\infty}_p X$. Folgende Aussagen sind äquivalent:*

1. $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$,
2. $||X_n||_p \xrightarrow{n\to\infty} ||X||_p$,
3. *Die Familie $(|X_n|^p)_{n=1,2,\dots}$ ist gleichgradig integrierbar.*

Beweis. Die Äquivalenz 1. $\Leftrightarrow$ 3. ist klar aus Theorem 8.11.
1. $\Rightarrow$ 2.: folgt aus der Minkowski'schen Ungleichung mit

$$\big|||X_n||_p^{p\wedge 1} - ||X||_p^{p\wedge 1}\big| \leq ||X_n - X||_p^{p\wedge 1} \xrightarrow{n\to\infty} 0.$$

2. $\Rightarrow$ 3.: Für festes K ist

$$\mathbf{E}[|X_n|^p; |X_n| > K] \leq \mathbf{E}[|X_n|^p - (|X_n| \wedge (K - |X_n|)^+)^p] \xrightarrow{n\to\infty} \mathbf{E}[|X|^p - (|X| \wedge (K - |X|)^+)^p].$$

Die Konvergenz folgt hierbei, da $\mathbf{E}[|X_n|^p] \xrightarrow{n\to\infty} \mathbf{E}[|X|^p]$, und $(|X_n| \wedge (K - |X_n|)^+)^p \xrightarrow{n\to\infty}_{\mathcal{L}^1} |X| \wedge (K - X)^+)^p$, da die Konvergenz nach Proposition 8.6 stochastisch gilt und $((|X_n| \wedge (K - |X_n|)^+)^p)_{n=1,2,\dots}$ beschränkt, insbesondere gleichgradig integrierbar ist. Da $\mathbf{E}[|X|^p - (|X| \wedge (K - |X|)^+)^p] \xrightarrow{K\to\infty} 0$ nach majorisierter Konvergenz, ist $(|X_n|^p)_{n=1,2,\dots}$ gleichgradig integrierbar. $\square$

9 Unabhängigkeit und das starke Gesetz

Mit unserem Wissen über Wahrscheinlichkeitsmaße und σ-Algebren beleuchten wir nun den Begriff der stochastischen Unabhängigkeit. Insbesondere werden wir in diesem Kapitel das starke Gesetz der großen Zahlen beweisen, siehe Theorem 9.21. Auf dem Weg dahin beweisen wir das Borel-Cantelli Lemma (Theorem 9.8) und das Kolmogorov'sche 0-1-Gesetz (Theorem 9.15).

9.1 Definition und einfache Eigenschaften

Bereits in der Vorlesung *Stochastik* wurden unabhängige Zufallsvariablen betrachtet. Die intuitive Vorstellung von Unabhängigkeit ist oft richtig, manchmal jedoch mit Vorsicht zu genießen.

Definition 9.1 (Unabhängigkeit). 1. *Eine Familie von Mengen* $(A_i)_{i\in I}$ *mit* $A_i \in \mathcal{F}$ *heißt* unabhängig, *falls*

$$\mathbf{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J}\mathbf{P}(A_j) \tag{9.1}$$

für alle $J \Subset I$.

2. *Eine Familie* $(\mathcal{C}_i)_{i\in I}$ *von Mengensystemen* $\mathcal{C}_i \subseteq \mathcal{F}$ *heißt* unabhängig, *falls* (9.1) *für alle* $J \Subset I$ *und* $A_j \in \mathcal{C}_j, j \in J$ *gilt.*

3. *Eine Familie von Zufallsvariablen* $(X_i)_{i\in I}$ *heißt unabhängig, falls* $(\sigma(X_i))_{i\in I}$ *unabhängig ist.*

Wir beschäftigen uns zuerst mit der Frage, ob es Wahrscheinlichkeitsräume gibt, auf denen es beliebig viele unabhängige Zufallsvariablen gibt. Hierbei kommt uns das Wissen über Produktmaße zu Gute.

Proposition 9.2 (Unabhängigkeit und Produktmaße). *Eine Familie* $(X_i)_{i\in I}$ *von Zufallsvariablen ist genau dann unabhängig, falls für jedes* $J \Subset I$

$$((X_i)_{i\in J})_*\mathbf{P} = \bigotimes_{i\in J}(X_i)_*\mathbf{P},$$

die gemeinsame Verteilung jeder endlichen Teilfamilie also gleich der Produktverteilung der einzelnen Verteilungen ist.

Beweis. Nach Definition ist die Familie $(X_i)_{i\in I}$ genau dann unabhängig, falls für jedes $J \Subset I$ und $A_i \in \mathcal{F}, i \in J$,

$$\mathbf{P}(X_i \in A_i, i \in J) = \prod_{i\in J}\mathbf{P}(X_i \in A_i).$$

Die Behauptung folgt nun daraus, dass $\mathbf{P}(X_i \in A_i) = (X_i)_*\mathbf{P}(A_i)$ (siehe Definition 3.22) und $\mathbf{P}(X_i \in A_i, i \in J) = ((X_i)_{i\in J})_*\mathbf{P}\big(\times_{i\in J} A_i\big)$ (siehe Korollar 6.14). □

Korollar 9.3 (Existenz von überabzählbar vielen unabhängigen Zufallsvariablen). *Sei* E *ein polnischer Raum,* I *eine beliebige Indexmenge. Seien* $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$ *Wahrscheinlichkeitsräume und* X_i E*-wertige Zufallsvariable,* $i \in I$*. Dann gibt es einen Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbf{P})$ *und eine Familie* $(Y_i)_{i\in I}$ E*-wertiger, unabhängiger Zufallsvariable mit* $Y_i \stackrel{d}{=} X_i$.

Beweis. Es sei bemerkt, dass $(((X_i)_{i\in J})_* \bigotimes_{i\in J} \mathbf{P}_i)_{J\Subset I}$ eine projektive Familie von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{B}(E))$ ist. Wegen Theorem 6.24 gibt es also den projektiven Limes $\mathbf{P}_I$. Dieser ist ein Wahrscheinlichkeitsmaß auf $(E^I, (\mathcal{B}(E))^I)$. Außerdem ist mit $\pi_i : E^I \to E$, der i-ten Projektion, $(\pi_i)_*\mathbf{P}_I = (X_i)_*\mathbf{P}_i$, d.h. $\pi_i \stackrel{d}{=} X_i$. □

Lemma 9.4 (Funktionen unabhängiger Zufallsvariablen). *Seien $(\Omega_i', \mathcal{F}_i'), (\Omega_i'', \mathcal{F}_i'')$, $i \in I$, Messräume. Sei $(X_i)_{i\in I}$ eine Familie unabhängiger Zufallsvariablen, $X_i : \Omega \to \Omega_i'$, und $\varphi_i : \Omega_i' \to \Omega_i''$ messbar, $i \in I$. Dann ist auch die Familie $(\varphi_i(X_i))_{i\in I}$ unabhängig.*

Beweis. Nach Lemma 7.2 ist die Zufallsvariable $\varphi_i(X_i)$ nach $\sigma(X_i)$ messbar, $i \in I$, d.h. $\sigma(\varphi_i(X_i)) \subseteq \sigma(X_i)$. Da $(\sigma(X_i))_{i\in I}$ nach Voraussetzung unabhängig sind, folgt die Behauptung aus der Definition der Unabhängigkeit. □

Proposition 9.5 (Unabhängigkeit und Unkorreliertheit). *Seien $X, Y \in \mathcal{L}^1$ unabhängige, reellwertige Zufallsvariablen. Dann ist $XY \in \mathcal{L}^1$ und es gilt*

$$\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Beweis. Zunächst bemerken wir: gilt die Behauptung für die Paare (X_i, Y_j), $i, j = 1, \ldots, n$, so auch für $\sum_{i=1}^n X_i$ und $\sum_{j=1}^n Y_j$. Wegen der Linearität des Erwartungswertes ist nämlich

$$\mathbf{E}\Big[\sum_{i=1}^n X_i \cdot \sum_{j=1}^n Y_j\Big] = \sum_{i=1}^n\sum_{j=1}^n \mathbf{E}[X_iY_j] = \sum_{i=1}^n\sum_{j=1}^n \mathbf{E}[X_i]\mathbf{E}[Y_j] = \mathbf{E}\Big[\sum_{i=1}^n X_i\Big] \cdot \mathbf{E}\Big[\sum_{j=1}^n Y_j\Big].$$

Die Behauptung ist klar, wenn X und Y Indikatorfunktionen sind. Wegen oben gesagtem gilt sie damit auch für einfache Funktionen, und damit mit monotoner Konvergenz auch für nichtnegative messbare Funktionen. Der allgemeine Fall folgt mit der Zerlegung $X = X^+ - X^-$ und $Y = Y^+ - Y^-$. □

Beispiel 9.6 (Unkorrelierte, nicht unabhängige Zufallsvariablen). Sei U eine auf $[0, 1]$ uniform verteilte Zufallsvariable, $X = \cos(2\pi U)$ und $Y = \sin(2\pi U)$. Dann ist $\mathbf{E}[X] = \mathbf{E}[Y] = 0$ und

$$\mathbf{E}[XY] = \int_0^1 \cos(2\pi u)\sin(2\pi u)du = \tfrac{1}{2}\int_0^1 \sin(4\pi u)du = 0$$

und damit sind X, Y unkorreliert. Allerdings ist $\{|X| < \varepsilon, |Y| < \varepsilon\} = \emptyset$ für $\varepsilon > 0$ klein genug und damit ist $\mathbf{P}(X^{-1}(-\varepsilon, \varepsilon), Y^{-1}(-\varepsilon, \varepsilon)) = 0 < \mathbf{P}(X^{-1}(-\varepsilon, \varepsilon)) \cdot \mathbf{P}(Y^{-1}(-\varepsilon, \varepsilon))$. Damit sind X und Y nicht unabhängig.

Hat man einen Wahrscheinlichkeitsraum und (abzählbar) viele Ereignisse, kann man sich fragen, wie viele dieser Ereignisse wohl eintreten. Das Borel-Cantelli Lemma gibt ein scharfes Kriterium dafür, wann nur endlich viele Ereignisse eintreten.

Definition 9.7 (limsup von Mengen). *Für $A_1, A_2, \dots \in \mathcal{F}$ ist*

$$\limsup_{n\to\infty} A_n := \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m$$

das Ereignis 'unendlich viele der A_n treten ein'.

Theorem 9.8 (Borel-Cantelli Lemma). *1. Sei $A_1, A_2, ... \in \mathcal{F}$. Dann gilt*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty \Longrightarrow \mathbf{P}(\limsup_{n\to\infty} A_n) = 0.$$

2. Sind $A_1, A_2, \ldots$ unabhängig, so gilt auch

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty \Longrightarrow \mathbf{P}(\limsup_{n\to\infty} A_n) = 1.$$

Beweis. Wir beginnen mit 1. Wegen der Stetigkeit von $\mathbf{P}$ von oben (siehe Proposition 3.7) gilt

$$\mathbf{P}(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}\Big(\bigcup_{m\geq n} A_m\Big) \leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbf{P}(A_m) = 0$$

nach Voraussetzung. Für 2. verwenden wir, dass $\log(1-x) \leq -x$ für $x \in [0,1]$. Damit gilt nämlich, wegen der Stetigkeit von $\mathbf{P}$ von unten und der Unabhängigkeit von $(A_n)_{n=1,2,...}$

$$\begin{aligned}
\mathbf{P}((\limsup_{n\to\infty} A_n)^c) &= \mathbf{P}\Big(\bigcup_{n=1}^{\infty} \bigcap_{m\geq n} A_m^c\Big) \\
&= \lim_{n\to\infty} \mathbf{P}\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) \\
&= \lim_{n\to\infty} \prod_{m=n}^{\infty} (1 - \mathbf{P}(A_m)) \\
&= \lim_{n\to\infty} \exp\Big(\sum_{m=n}^{\infty} \log(1 - \mathbf{P}(A_m))\Big) \\
&\leq \lim_{n\to\infty} \exp\Big(-\sum_{m=n}^{\infty} \mathbf{P}(A_m)\Big) \\
&= 0
\end{aligned}$$

und die Behauptung folgt. □

Beispiel 9.9 (Unendlicher Münzwurf und geometrische Verteilungen).

1. Wir betrachten einen unendlichen Münzwurf. Das bedeutet, dass wir einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und unabhängige Zufallsvariablen $X_1, X_2, \ldots$ mit Werte in $\{\text{Kopf}, \text{Zahl}\}$ haben. Der Münzwurf sei fair, d.h. $\mathbf{P}(X_n = \text{Kopf}) = 1/2$. Wir betrachten die Ereignisse $A_n = \{X_n = \text{Kopf}\}$. Da

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \sum_{n=1}^{\infty} \tfrac{1}{2} = \infty$$

   und die Familie $(A_n)_{n\in\mathbb{N}}$ unabhängig ist, folgt aus dem Borel-Cantelli Lemma, dass fast sicher unendlich oft *Kopf* kommt.

2. Wir betrachten dieselbe Situation wie in 1., jedoch die Ereignisse $B_n := \{X_1 = \text{Kopf}\}$. Klar ist, dass die Familie $(B_n)_{n\in\mathbb{N}}$ nicht unabhängig ist. (Z.B. ist ja $\mathbf{P}(B_1 \cap B_2) = \mathbf{P}(B_1) = 1/2 \neq \frac{1}{4} = \mathbf{P}(B_1) \cdot \mathbf{P}(B_2)$.) Genau wie in 1. ist $\sum_{n=1}^{\infty} \mathbf{P}(B_n) = \infty$. Klar ist auch, dass $\mathbf{P}(\limsup_{n\to\infty} B_n) = \frac{1}{2}$. Daraus folgt, dass im Borel-Cantelli Lemma auf die Bedingung der Unabhängigkeit in 2. nicht verzichtet werden kann.

3. Seien $X_1, X_2, \dots$ zum Erfolgsparameter p geometrisch verteilte Zufallsvariablen. Wir betrachten die Ereignisse $A_n := \{X_n \geq n\}$ und fragen uns, ob unendlich viele dieser Ereignisse eintreten können. Da
$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \sum_{n=1}^{\infty} \mathbf{P}(X_n \geq n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \tfrac{1}{p} < \infty.$$
Deshalb treten fast sicher nur endlich viele der Ereignisse $\{X_n \geq n\}$ ein.

9.2 Das Kolmogorov'sche 0-1-Gesetz

Bereits das Borel-Cantelli Lemma ist eine Aussage darüber, wann ein von unendlich vielen Ereignissen abhängiges Ereignis fast sicher eintritt. Diese Situation werden wir nun weiter beleuchten.

Proposition 9.10 (Unabhängigkeit erzeugter σ-Algebren). *Sei $(\mathcal{C}_i)_{i\in I}$ eine Familie unabhängiger, schnittstabiler Mengensysteme. Dann ist auch $(\sigma(\mathcal{C}_i))_{i\in I}$ eine unabhängige Familie.*

Beweis. Sei $J = \{i_1, \dots, i_n\} \Subset I$ und o.E. $|J| > 1$. Dann gilt (9.1) für beliebige $A_{i_1}, \dots, A_{i_n}$ mit $A_{i_k} \in \mathcal{C}_{i_k}, k = 1, \dots, n$. Wir halten $A_{i_2}, \dots, A_{i_n}$ fest und definieren
$$\mathcal{D} := \{A_{i_1} \in \mathcal{F} : (9.1) \text{ gilt}\}.$$
Wir werden nun zeigen, dass $\mathcal{D}$ ein Dynkin-System ist. Ist nämlich $A \subseteq B \in \mathcal{D}$, so ist auch $B \setminus A \in \mathcal{D}$, weil
$$\begin{aligned}
\mathbf{P}\Big((B \setminus A) \cap \bigcap_{k=2}^{n} A_{i_k}\Big) &= \mathbf{P}\Big(B \cap \bigcap_{k=2}^{n} A_{i_k}\Big) - \mathbf{P}\Big(A \cap \bigcap_{k=2}^{n} A_{i_k}\Big) \\
&= (\mathbf{P}(B) - \mathbf{P}(A)) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}) \\
&= \mathbf{P}(B \setminus A) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}).
\end{aligned}$$
Ist außerdem $A_1, A_2, \dots \in \mathcal{D}$ mit $A_1 \subseteq A_2 \subseteq A_2 \dots$, so ist wegen der Stetigkeit von $\mathbf{P}$ von unten
$$\begin{aligned}
\mathbf{P}\Big(\Big(\bigcup_{j=1}^{\infty} A_j\Big) \cap \bigcap_{k=2}^{n} A_{i_k}\Big) &= \sup_{j\in\mathbb{N}} \mathbf{P}\Big(A_j \cap \bigcap_{k=2}^{n} A_{i_k}\Big) \\
&= \sup_{j\in\mathbb{N}} \mathbf{P}(A_j) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}) \\
&= \mathbf{P}\Big(\bigcup_{j=1}^{\infty} A_j\Big) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}).
\end{aligned}$$

Da $\mathcal{C}_{i_1}$ schnittstabil ist und $\mathcal{C}_{i_1} \subseteq \mathcal{D}$, ist $\sigma(\mathcal{C}_{i_1}) \subseteq \mathcal{D}$ nach Theorem 2.13. Insbesondere gilt (9.1) für $A_{i_1} \in \sigma(\mathcal{C}_{i_1}), A_{i_2} \in \mathcal{C}_{i_2}, \dots, A_{i_n} \in \mathcal{C}_{i_n}$. Iteriert man obiges Verfahren für $k = 2, \dots, n$, erhält man die Aussage. □

Korollar 9.11 (Unabhängigkeit von Indikatorfunktionen). *Eine Familie von Mengen $(A_i)_{i \in I}$ ist genau dann unabhängig, wenn die Familie der Zufallsvariablen $(1_{A_i})_{i \in I}$ unabhängig ist. Insbesondere gilt*

$$\mathbf{P}\Big(\bigcap_{j \in J} B_j\Big) = \prod_{j \in J} \mathbf{P}(B_j)$$

für $J \Subset I$, $B_j \in \{A_j, A_j^c\}, j \in J$.

Beweis. Für $i \in I$ sei $\mathcal{C}_i = \{A_i\}$. Dann ist $\sigma(1_{A_i}) = \{\emptyset, A_i, A_i^c, \Omega\} = \sigma(\mathcal{C}_i)$. Da $\mathcal{C}_i$ trivialerweise schnittstabil ist, folgt die Aussage aus Proposition 9.10. □

Korollar 9.12 (Gruppierung). *Sei $(\mathcal{F}_i)_{i \in I}$ eine Familie unabhängiger σ-Algebren. Weiter sei $\mathcal{I}$ eine Partition von I, d.h. $\mathcal{I} = \{I_k, k \in K\}$ mit $\biguplus_{k \in K} I_k = I$, die I_k sind also disjunkt und deren Vereinigung ist I. Dann ist auch $(\sigma(\mathcal{F}_i : i \in I_k))_{k \in K}$ ein unabhängiges System.*

Beweis. Das Mengensystem $\mathcal{C}_k := \big\{ \bigcap_{i \in J_k} A_i : J_k \Subset I_k, A_i \in \mathcal{F}_i \big\}$ ist schnittstabil und $\sigma(\mathcal{C}_k) = \sigma(\mathcal{F}_i : i \in I_k)$, $k \in K$. Da nach Voraussetzung die Familie $(\mathcal{C}_k)_{k \in K}$ unabhängig ist, folgt die Behauptung aus Proposition 9.10. □

Wir kommen nun zur Hauptaussage dieses Abschnittes, dem Kolmogorov'schen 0-1-Gesetz. Hierzu führen wir eine bestimmte σ-Algebra ein, die terminale σ-Algebra.

Definition 9.13 (Terminale und triviale $\boldsymbol{\sigma}$-Algebren).

1. *Sei $\mathcal{F}_1, \mathcal{F}_2, ... \subseteq \mathcal{F}$ eine Folge von σ-Algebren. Dann ist*

$$\mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots) = \bigcap_{n \geq 1} \sigma\Big(\bigcup_{m > n} \mathcal{F}_m\Big)$$

die σ-Algebra der terminalen Ereignisse *von $\mathcal{F}_1, \mathcal{F}_2, \dots$*

2. *Eine σ-Algebra $\widetilde{\mathcal{F}} \subseteq \mathcal{F}$ heißt $\mathbf{P}$-trivial, falls $\mathbf{P}(A) \in \{0,1\}$ für alle $A \in \widetilde{\mathcal{F}}$.*

Lemma 9.14 (Triviale σ-Algebren).

1. *Eine σ-Algebra $\widetilde{\mathcal{F}}$ ist genau dann $\mathbf{P}$-trivial, wenn $\widetilde{\mathcal{F}}$ von sich selbst unabhängig ist.*

2. *Sei $\widetilde{\mathcal{F}}$ eine $\mathbf{P}$-triviale σ-Algebra und X eine $\widetilde{\mathcal{F}}$-messbare Zufallsvariable mit Werten in einem separablen metrischen Raum E. Dann ist X fast sicher konstant.*

Beweis. 1. Sei $\widetilde{\mathcal{F}}$ zunächst $\mathbf{P}$-trivial und $A, B \in \widetilde{\mathcal{F}}$. Dann gilt $\mathbf{P}(A \cap B) = \mathbf{P}(A) \wedge \mathbf{P}(B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$, also ist $\widetilde{\mathcal{F}}$ von sich selbst unabhängig. Ist andererseits $\widetilde{\mathcal{F}}$ von sich selbst unabhängig und $A \in \widetilde{\mathcal{F}}$, dann ist $\mathbf{P}(A) = \mathbf{P}(A \cap A) = (\mathbf{P}(A))^2$, also $\mathbf{P}(A) \in \{0,1\}$.

2. Für $n \in \mathbb{N}$ sei $(B_{nj})_{j=1,2,\dots}$ eine abzählbare Überdeckung von E mit Bällen vom Radius $1/n$. Da $\widetilde{\mathcal{F}}$ eine $\mathbf{P}$-triviale σ-Algebra ist, gilt $\mathbf{P}(X \in B_{nj}) \in \{0,1\}$ für alle n, j. Für $n \in \mathbb{N}$ sei $J_n := \{j \in \mathbb{N} : \mathbf{P}(X \in B_{nj}) = 1\} \neq \emptyset$. Damit ist wegen der Stetigkeit von oben $\mathbf{P}\Big(X \in \bigcap_{n=1}^\infty \bigcap_{j \in J_n} B_{nj}\Big) = 1$. Da $\bigcap_{n=1}^\infty \bigcap_{j \in J_n} B_{nj}$ höchstens ein Element hat, folgt die Behauptung. □

Unter Unabhängigkeit ist die σ-Algebra der terminalen Ereignisse besonders einfach.

Theorem 9.15 (Kolmogorov'sches 0-1-Gesetz). *Sei $\mathcal{F}_1, \mathcal{F}_2, \dots \subseteq \mathcal{F}$ eine Folge unabhängiger σ-Algebren. Dann ist $\mathcal{T} := \mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots)$ $\mathbf{P}$-trivial.*

Beweis. Sei $\mathcal{T}_n := \sigma\Big(\bigcup_{m>n} \mathcal{F}_m\Big)$, $n = 1, 2, \dots$. Nach Korollar 9.12 sind $(\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{T}_n)$ unabhängig, $n = 1, 2, \dots$ Damit sind auch $(\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{T})$ unabhängig, $n = 1, 2, \dots$ und damit auch $(\mathcal{T}, \mathcal{F}_1, \mathcal{F}_2, \dots)$. Wieder mit Korollar 9.12 folgt, dass $(\mathcal{T}_0, \mathcal{T})$ unabhängig sind und, da $\mathcal{T} \subseteq \mathcal{T}_0$ auch, dass $\mathcal{T}$ von sich selbst unabhängig ist. Deswegen folgt die Behauptung aus Lemma 9.14. □

9.3 Summen unabhängiger Zufallsvariable

Viele wichtige Sätze der Wahrscheinlichkeitstheorie beschäftigen sich mit unabhängigen Zufallsvariablen. In dieser Vorlesung sind dies vor allem das starke Gesetz der großen Zahlen (Theorem 9.21) und der zentrale Grenzwertsatz (Theorem 11.8). Bereits hier geben wir wichtige Hilfsmittel zur Analyse von Summen unabhängiger Zufallsvariablen an. Das erste ist der Zusammenhang mit der Faltung von Wahrscheinlichkeitsmaßen (siehe Abschnitt 6.4).

Proposition 9.16 (Faltung ist Verteilung der Summe). *Seien $X_1, \dots, X_n$ unabhängige, reellwertige Zufallsvariablen. Dann gilt*

$$(X_1 + \dots + X_n)_*\mathbf{P} = (X_1)_*\mathbf{P} * \dots * (X_n)_*\mathbf{P}.$$

Weiter gilt für die charakteristischen Funktionen

$$\psi_{X_1+\dots+X_n} = \psi_{X_1} \cdots \psi_{X_n}$$

und, falls $X_1, \dots, X_n$ Werte in $\mathbb{R}_+$ annehmen,

$$\mathscr{L}_{X_1+\dots+X_n} = \mathscr{L}_{X_1} \cdots \mathscr{L}_{X_n}.$$

Beweis. Zunächst ist nach Proposition 9.2 $((X_1, \dots, X_n))_*\mathbf{P} = (X_1)_*\mathbf{P} \otimes \dots \otimes (X_n)_*\mathbf{P}$. Damit folgt die erste Behauptung bereits aus der Definition 6.17 der Faltung von Maßen. Die weiteren Behauptungen folgen aus Proposition 9.5, da etwa

$$\begin{aligned}\psi_{X_1+\dots+X_n}(t) &= \mathbf{E}[e^{it(X_1+\dots+X_n)}] = \mathbf{E}\big[e^{itX_1} \cdots e^{itX_n}\big] \\ &= \mathbf{E}[e^{itX_1}] \cdots \mathbf{E}[e^{itX_n}] = \psi_{X_1}(t) \cdots \psi_{X_n}(t).\end{aligned}$$

□

Das Kolmogorov'sche 0-1-Gesetz stellt recht einfach eine Aussage zur Verfügung, wann Summen unabhängiger Zufallsvariable fast sicher konvergieren.

Proposition 9.17 (Konvergenz von Summen unabhängiger Zufallsvariablen). *Seien $X_1, X_2, \dots$ unabhängige Zufallsvariablen und $S_n := X_1 + \dots + X_n$.*

1. *Es gilt*

$$\mathbf{P}(\omega : S_n(\omega) \text{ konvergiert für } n \to \infty) \in \{0, 1\}$$

2. *Weiter ist*
$$\mathbf{P}(\omega : S_n(\omega)/n \text{ konvergiert für } n \to \infty) \in \{0, 1\}.$$
Falls $\mathbf{P}(S_n/n \text{ konvergiert}) = 1$, *ist der Grenzwert fast sicher konstant.*

Beweis. Setze $\mathcal{F}_i := \sigma(X_i)$, $i = 1, 2, \dots$ Damit ist die Familie $(\mathcal{F}_i)_{i=1,2,\dots}$ unabhängig. Die Menge $\{\omega : S_n(\omega) \text{ konvergiert für } n \to \infty\}$ ist messbar bezüglich $\mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots)$ und damit folgt die erste Aussage aus Theorem 9.15. Genauso folgt, dass $\mathbf{P}(S_n/n \text{ konvergiert}) \in \{0, 1\}$. Sei $S = \lim_{n\to\infty} S_n(n)/n$. Damit gilt für alle $m = 1, 2, \dots$
$$S = \lim_{n\to\infty} \frac{X_1 + \dots + X_n}{n} = \lim_{n\to\infty} \frac{X_m + \dots + X_n}{n},$$
also ist S nach $\sigma\Big(\bigcup_{k\geq m} \mathcal{F}_k\Big)$ messbar. Damit ist S auch $\mathcal{T}$-messbar und damit fast sicher konstant nach Theorem 9.15 und Lemma 9.14. □

Proposition 9.18 (Maximal-Ungleichung von Kolmogorov). *Seien* $X_1, X_2, \dots \in \mathcal{L}^2$ *unabhängige Zufallsvariablen. Dann gilt für* $K > 0$
$$\mathbf{P}\Big(\sup_{n\in\mathbb{N}} \Big|\sum_{k=1}^n X_k - \mathbf{E}[X_k]\Big| > K\Big) \leq \frac{\sum_{n=1}^\infty \mathbf{V}(X_n)}{K^2}.$$

Beweis. O.E. sei $\mathbf{E}[X_k] = 0, k = 1, 2, \dots$. Weiter setzen wir $S_n = X_1 + \dots + X_n$ und $T := \inf\{n : |S_n| > K\}$. Dann gilt $\mathbf{P}(\sup_n |S_n| > K) = \mathbf{P}(T < \infty)$. Wegen Korollar 9.12 sind $S_k \cdot 1_{T=k}$ und $S_n - S_k$ unabhängig für $k \leq n$. Deshalb gilt
$$\begin{aligned}
\sum_{k=1}^n \mathbf{E}[X_k^2] = \mathbf{E}[S_n^2] &\geq \sum_{k=1}^n \mathbf{E}[S_n^2, T = k] \\
&= \sum_{k=1}^n \mathbf{E}[S_k^2 + (S_n - S_k + 2S_k)(S_n - S_k), T = k] \\
&\geq \sum_{k=1}^n \mathbf{E}[S_k^2, T = k] + 2\mathbf{E}[S_k(S_n - S_k), T = k] \\
&= \sum_{k=1}^n \mathbf{E}[S_k^2, T = k] \geq K^2 \mathbf{P}(T \leq n)
\end{aligned}$$
Nun folgt die Behauptung mit $n \to \infty$. □

Theorem 9.19 (Konvergenzkriterium für Reihen). *Seien* $X_1, X_2, \dots \in \mathcal{L}^2$ *unabhängige Zufallsvariablen mit* $\sum_{n=1}^\infty \mathbf{V}[X_n] < \infty$. *Dann konvergiert* $\sum_{k=1}^n X_k - \mathbf{E}[X_k]$ *fast sicher.*

Beweis. Wieder sei o.E. $\mathbf{E}[X_k] = 0, k = 1, 2, \dots$ und wir schreiben $S_n = X_1 + \dots + X_n$. Für $\varepsilon > 0$ gilt nach Proposition 9.18
$$\lim_{k\to\infty} \mathbf{P}(\sup_{n\geq k} |S_n - S_k| > \varepsilon) \leq \lim_{k\to\infty} \frac{\sum_{n=k+1}^\infty \mathbf{E}[X_n^2]}{\varepsilon^2} = 0.$$
Deswegen konvergiert $\sup_{n\geq k} |S_n - S_k| \xrightarrow{k\to\infty}_p 0$. Nach Proposition 8.6 gibt es also eine Teilfolge $k_1, k_2, \dots$ mit $\sup_{n\geq k_i} |S_n - S_{k_i}| \xrightarrow{i\to\infty}_{fs} 0$. Da aber $(\sup_{n\geq k} |S_n - S_k|)_{k=1,2,\dots}$ fallend ist, gilt $\sup_{n\geq k} |S_n - S_k| \xrightarrow{k\to\infty}_{fs} 0$. Das bedeutet aber, dass $(S_n)_{n=1,2,\dots}$ konvergiert. □

9.4 Das starke Gesetz der großen Zahlen

Bereits in der Vorlesung *Stochastik* haben wir das schwache Gesetz der großen Zahlen bewiesen: sind $X_1, X_2, \dots \in \mathcal{L}^2$ identisch verteilt und unkorreliert, dann ist für $\varepsilon > 0$

$$\mathbf{P}\Big(\frac{1}{n}\Big|\sum_{k=1}^n (X_k - \mathbf{E}[X_k])\Big| > \varepsilon\Big) \leq \frac{1}{\varepsilon^2}\mathbf{V}\Big[\frac{1}{n}\sum_{k=1}^n X_k\Big] = \frac{1}{\varepsilon^2 n^2}\sum_{k=1}^n \mathbf{V}[X_k] = \frac{\mathbf{V}[X_1]}{\varepsilon^2 n} \xrightarrow{n\to\infty} 0.$$

Diese Aussage wollen wir nun in zwei Richtungen verschärfen. Wir wollen einerseits die stochastische Konvergenz durch die fast sichere Konvergenz ersetzen, und außerdem nur die Existenz erster Momente (nicht jedoch die Existenz zweiter Momente) fordern. Zunächst jedoch definieren wir, was wir genau meinen, wenn wir sagen, dass eine Folge von Zufallsvariablen einem Gesetz großer Zahlen folgt.

Definition 9.20 (Gesetz der großen Zahlen). *Sei $X_1, X_2, \dots \in \mathcal{L}^1$ eine Folge reellwertiger Zufallsvariablen. Wir sagen, dass die Folge dem schwachen Gesetz der großen Zahlen folgt, falls*

$$\frac{1}{n}\sum_{k=1}^n (X_k - \mathbf{E}[X_k]) \xrightarrow{n\to\infty}_p 0.$$

Die Folge genügt dem starken Gesetz der großen Zahlen, falls

$$\frac{1}{n}\sum_{k=1}^n (X_k - \mathbf{E}[X_k]) \xrightarrow{n\to\infty}_{fs} 0.$$

Theorem 9.21 (Starkes Gesetz für unabhängige Zufallsvariablen). *Eine Folge $X_1, X_2, \dots \in \mathcal{L}^1$ unabhängiger und identisch verteilter Zufallsvariablen genügt dem starken Gesetz der großen Zahlen, d.h. es gilt*

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{n\to\infty}_{fs} \mathbf{E}[X_1].$$

Bemerkung 9.22 (Schwaches Gesetz). Da aus der fast sicheren Konvergenz die stochastische Konvergenz folgt (siehe Proposition 8.6), genügt die Folge $X_1, X_2, \dots$ aus dem Theorem auch dem schwachen Gesetz der großen Zahlen. Weiter genügt auch die Folge $X_1^+, X_2^+, \dots$ dem starken Gesetz und $\mathbf{E}[\frac{1}{n}(X_1^+ + \cdots X_n^+)] = \mathbf{E}[X_1^+]$. Damit ist die Folge $(\frac{1}{n}(X_1^+ + \cdots + X_n^+))_{n=1,2,\dots}$ nach Korollar 8.12 gleichgradig integrierbar. Genauso ist die Folge der Partialsummen der Negativteile gleichgradig integrierbar. Es folgt aus Theorem 8.11, dass $\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{n\to\infty}_{\mathcal{L}^1} \mathbf{E}[X_1]$.

Bemerkung 9.23 (Endliche vierte und zweite Momente). Die Schwierigkeit im Beweis des starken Gesetzes ist, dass nur verwendet werden darf, dass $X_1 \in \mathcal{L}^1$. Wesentlich einfacher wird der Beweis, wenn man $X_1 \in \mathcal{L}^4$ bzw. $X_1 \in \mathcal{L}^2$ voraussetzt. Diese beiden Beweise geben wir zunächst an. Es sei $S_n := X_1 + \cdots + X_n$.

1. Der Fall $X_1 \in \mathcal{L}^4$: Hier kommt man ganz ohne weitere Hilfsmittel aus:

   Aus der Linearität des Erwartungswertes ist klar, dass $\mathbf{E}[S_n/n] = \mathbf{E}[X_1]$. O.E. sei $\mathbf{E}[X_1] = 0$, ansonsten geht man zu den Zufallsvariablen $X_1 - \mathbf{E}[X_1], X_2 - \mathbf{E}[X_2], \dots \in \mathcal{L}^4$

über. Zunächst berechnen wir mit Hilfe der Unabhängigkeit von $(X_k)_{k=1,2,\dots}$

$$\mathbf{E}[S_n^4] = \sum_{k=1}^n \mathbf{E}[X_k^4] + 3 \sum_{\substack{k,l=1\\k\neq l}}^{\infty} \mathbf{E}[X_k^2 X_l^2] \leq (n+6n^2)\mathbf{E}[X_1^4]$$

wegen der Cauchy-Schwartz-Ungleichung. Daraus folgt

$$\mathbf{E}\Big[\sum_{n=1}^\infty \Big(\frac{S_n}{n}\Big)^4\Big] \leq \sum_{n=1}^\infty \frac{n+6n^2}{n^4}\mathbf{E}[X_1^4] < \infty.$$

Deswegen gilt $\sum_{n=1}^\infty \Big(\frac{S_n}{n}\Big)^4 < \infty$ fast sicher, insbesondere also $\frac{S_n}{n} \xrightarrow{n\to\infty}_{fs} 0$.

2. Der Fall $X_1 \in \mathcal{L}^2$: Hier ist das Konvergenzkriterium für Reihen, Theorem 9.19 von entscheidender Hilfe. Außerdem benötigen wir noch folgendes Resultat:

Lemma 9.24 (Kronecker Lemma). *Seien $x_1, x_2, \dots \in \mathbb{R}$, $y_1, y_2, \dots \in \mathbb{R}$ monoton mit $y_n \uparrow \infty$ und $\sum_{n=1}^\infty x_n/y_n < \infty$. Dann gilt $\sum_{k=1}^n x_k/y_n \xrightarrow{n\to\infty} 0$.*

Beweis. Sei $z_0 = 0, z_n := \sum_{k=1}^n x_k/y_k$. Dann gilt $z_n \xrightarrow{n\to\infty} z_\infty < \infty$ und $x_k = y_k(z_k - z_{k-1})$. Wir schreiben mit $y_0 = 0$

$$\begin{aligned}\frac{\sum_{k=1}^n x_k}{y_n} &= \frac{1}{y_n}\sum_{k=1}^n y_k(z_k - z_{k-1}) = z_n + \frac{1}{y_n}\Big(\sum_{k=0}^{n-1} y_k z_k - \sum_{k=1}^n y_k z_{k-1}\Big)\\ &= z_n - \frac{1}{y_n}\Big(\sum_{k=1}^n y_k z_{k-1} - y_{k-1}z_{k-1}\Big)\\ &\xrightarrow{n\to\infty} z_\infty - z_\infty \cdot \lim_{n\to\infty}\frac{1}{y_n}\sum_{k=1}^n y_k - y_{k-1} = 0.\end{aligned}$$

□

Zurück zum Beweis des starken Gesetzes im Fall $X_1 \in \mathcal{L}^2$. O.E. sei wieder $\mathbf{E}[X_1] = 0$. Betrachte die Folge $X_1/1, X_2/2, \dots$ Wegen $\sum_{n=1}^\infty \mathbf{V}[X_n/n] = \mathbf{V}[X_1]\sum_{n=1}^\infty 1/n^2$ gilt nach Theorem 9.19, dass $\sum_{k=1}^n X_k/k$ fast sicher konvergiert. Mit Lemma 9.24 folgt, dass $S_n/n \xrightarrow{n\to\infty}_{fs} 0$.

Beweis von Theorem 9.21 falls $X_1 \in \mathcal{L}^1$. Es genügt, den Fall von nicht-negativen Zufallsvariablen zu betrachten. Im allgemeinen Fall sei bemerkt, dass $X_1^+, X_2^+, \dots \in \mathcal{L}^1$ und $X_1^-, X_2^-, \dots \in \mathcal{L}^1$ die Voraussetzungen des Satzes erfüllen, und aus $(X_1^+ + \dots + X_n^+)/n \xrightarrow{n\to\infty}_{fs} \mathbf{E}[X_1^+]$ und $(X_1^- + \dots + X_n^-)/n \xrightarrow{n\to\infty}_{fs} \mathbf{E}[X_1^-]$ die Aussage wegen Linearität des Erwartungswertes folgt.

Für $S_n = X_1 + \dots + X_n$ werden wir zeigen, dass

$$\mathbf{E}[\limsup_{n\to\infty} S_n/n] \leq \mathbf{E}[X_1]. \tag{9.2}$$

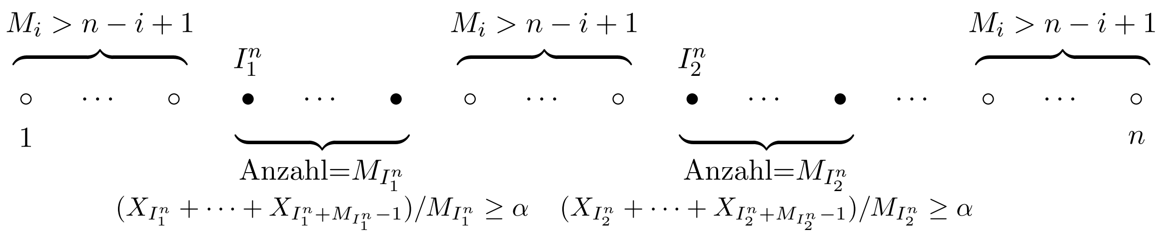


Abbildung 9.1: Illustration von M_i, I_j^n, eingeführt unterhalb von (9.3). Die Größe L_n ist die Anzahl der zusammenhängenden Bereiche von $\bullet$'s.

Gilt dies, so folgt erstens

$$\begin{aligned}
\mathbf{E}[\liminf_{n\to\infty} S_n/n] &\geq \mathbf{E}[\liminf_{n\to\infty} (X_1 \wedge k + \cdots + X_n \wedge k)/n] \\
&= k - \mathbf{E}[\limsup_{n\to\infty} ((k-X_1)^+ + \cdots + (k-X_n)^+)/n] \\
&\geq \mathbf{E}[k - (k-X_1)^+] \xrightarrow{k\to\infty} \mathbf{E}[X_1]
\end{aligned}$$

Zweitens ist dann $\mathbf{E}[\limsup_{n\to\infty} S_n/n - \liminf_{n\to\infty} S_n/n] = 0$, also $\limsup_{n\to\infty} S_n/n = \liminf_{n\to\infty} S_n/n = 0$ fast sicher, da sowohl $\liminf_{n\to\infty} S_n/n$ als auch $\limsup_{n\to\infty} S_n/n$ terminale Funktionen, und damit nach Theorem 9.15 und Lemma 9.14 fast sicher konstant sind. Außerdem ist damit

$$\liminf_{n\to\infty} S_n/n = \mathbf{E}[\liminf_{n\to\infty} S_n/n] \geq \mathbf{E}[X_1] \geq \mathbf{E}[\limsup_{n\to\infty} S_n/n] = \limsup_{n\to\infty} S_n/n,$$

woraus die Behauptung folgt.

Es bleibt also (9.2) zu zeigen. O.E. sei $\mathbf{E}[X_1] > 0$, ansonsten ist $X_k = 0$ fast sicher, $k = 1, 2, \ldots$ und die Aussage ist trivial. Hierzu werden wir

$$0 < \alpha < \mathbf{E}[\limsup_{n\to\infty} S_n/n] \qquad \Longrightarrow \qquad \alpha \leq \mathbf{E}[X_1] \tag{9.3}$$

beweisen. Nach Voraussetzung ist für $i = 0, 1, 2, \ldots$

$$\alpha < \mathbf{E}[\limsup_{n\to\infty} S_n/n] = \limsup_{n\to\infty} S_n/n = \limsup_{n\to\infty} (X_{i+1} + \cdots X_{i+n})/n.$$

Damit ist

$$M_i := \inf\{n \in \mathbb{N} : (X_i + \cdots + X_{i+n-1})/n \geq \alpha\}$$

fast sicher endlich, $i = 1, 2, \ldots$ Die M_i's sind identisch verteilt. Wir definieren für $n = 1, 2, \ldots$ rekursiv (siehe auch Abbildung 9.1) $I_1^n = 0$ sowie für $j = 0, 1, 2, \ldots$ (mit $M_0 := 0$)

$$I_{j+1}^n := \inf\{i \in \mathbb{N} : i \geq I_j^n + M_{I_j}^n, M_i \leq n - i + 1\}$$

mit $\inf \emptyset = \infty$ und $L_n := \sup\{n \in \mathbb{N}_0 : I_j^n < \infty\}$. Damit gilt für $1 \leq j \leq L_n$, dass

$I_j^n + M_{I_j^n} \leq n$, also $(X_{I_j^n} + \cdots + X_{I_j^n + M_{I_j^n} - 1})/M_{I_j^n} \geq \alpha$. Dies verwenden wir nun mittels

$$\begin{aligned}
\mathbf{E}[X_1] &= \mathbf{E}[(X_1 + \cdots + X_n)/n] \\
&\geq \tfrac{1}{n}\mathbf{E}\Big[\sum_{j=1}^{L_n} M_{I_j^n} \cdot (X_{I_j^n} + \cdots + X_{I_j^n + M_{I_j^n} - 1})/M_{I_j^n}\Big] \\
&\geq \tfrac{\alpha}{n}\mathbf{E}\Big[\sum_{j=1}^{L_n} M_{I_j^n}\Big] = \alpha - \tfrac{\alpha}{n}\mathbf{E}\Big[n - \sum_{j=1}^{L_n} M_{I_j^n}\Big] \\
&\geq \alpha - \tfrac{\alpha}{n}\mathbf{E}\Big[\sum_{i=1}^{n} 1_{M_i > n-i+1}\Big] \\
&= \alpha\Big(1 - \tfrac{1}{n}\sum_{i=1}^{n}\mathbf{P}(M_i > i)\Big) \xrightarrow{n\to\infty} \alpha,
\end{aligned}$$

da $\big(\frac{1}{n}\sum_{i=1}^{n}\mathbf{P}(M_i > i)\big)_{n=1,2,\ldots}$ als Cesàro-Limes von $(\mathbf{P}(M_i > i))_{i=1,2,\ldots}$ wegen der Identität der Verteilungen der M_i's gegen 0 konvergiert. Damit ist (9.3) gezeigt und die Behauptung ist bewiesen. $\Box$

Wir geben nun eine einfach Anwendung des starken Gesetzes an. Gerade in der Statistik kommt es oft vor, dass man eine große Anzahl unabhängiger, identisch verteilter, reellwertiger Zufallsgrößen betrachtet. Der Satz von Glivenko-Cantelli (Theorem 9.26) besagt, dass die empirische Verteilung der Zufallsvariablen fast sicher gegen die zu Grunde liegende Verteilung konvergiert.

Definition 9.25 (Empirische Verteilung). *Seien $X_1, X_2, \ldots$ Zufallsvariable. Für $n = 1, 2, \ldots$ heißt die (zufällige) Wahrscheinlichkeitsverteilung*

$$\widehat{\mu}_n := \frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$$

die empirische Verteilung *von $X_1, \ldots, X_n$. Sind die Zufallsvariablen reellwertig, ist außerdem*

$$\widehat{F}_n(x) := \frac{1}{n}\sum_{k=1}^{n} 1_{X_k \leq x}$$

die empirische Verteilungsfunktion *von $X_1, \ldots, X_n$.*

Theorem 9.26 (Satz von Glivenko-Cantelli). *Seien $X_1, X_2, \ldots$ unabhängige, reellwertige Zufallsvariablen mit identischer Verteilung mit Verteilungsfunktion F. Dann gilt*

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \xrightarrow{n\to\infty}_{fs} 0.$$

Beweis. Für $x \in \mathbb{R}$ und $n = 1, 2, \ldots$ sei $Y_n(x) := 1_{X_n \leq x}$ und $Z_n(x) := 1_{X_n < x}$. Nach Theorem 9.21 gilt für jedes $x \in \mathbb{R}$

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} Y_k(x) \xrightarrow{n\to\infty}_{fs} \mathbf{E}[Y_1(x)] = \mathbf{P}(X_1 \leq x) = F(x),$$

$$\widehat{F}_n(x-) = \frac{1}{n}\sum_{k=1}^{n} Z_k(x) \xrightarrow{n\to\infty}_{fs} \mathbf{E}[Z_1(x)] = \mathbf{P}(X_1 < x) = F(x-).$$

Wir müssen zeigen, dass diese Konvergenzen auch uniform für alle $x \in \mathbb{R}$ gelten. Für $N = 1, 2, \dots$ und $j = 0, \dots, N$ setzen wir

$$x_j^N := \inf\{x \in \mathbb{R} : F(x) \geq j/N\}$$

und

$$R_n^N := \max_{j=1,\dots,N-1} \big(|\widehat{F}_n(x_j^N) - F(x_j^N)| + |\widehat{F}_n(x_j^N-) - F(x_j^N-)|\big).$$

Für $N = 1, 2, \dots$ gilt also $R_n^N \xrightarrow{n\to\infty}_{fs} 0$. Außerdem ist für $x \in (x_{j-1}^N, x_j^N)$

$$\begin{aligned}
\widehat{F}_n(x) &\leq \widehat{F}_n(x_j^N) \leq \widehat{F}_n(x_j^N) + R_n^N \leq F(x) + R_n^N + \tfrac{1}{N},\\
\widehat{F}_n(x) &\geq \widehat{F}_n(x_{j-1}^N) \geq F(x_{j-1}^N) - R_n^N \geq F(x) - R_n^N - \tfrac{1}{N},
\end{aligned}$$

also, für jedes $N = 1, 2, \dots$

$$\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \leq \tfrac{1}{N} + R_n^N \xrightarrow{n\to\infty}_{fs} \tfrac{1}{N}.$$

Da die linke Seite nicht von N abhängt, folgt die Behauptung mit $N \to \infty$. □

10 Schwache Konvergenz

Für Maß- und Messräume haben wir bisher oft verwendet, dass die σ-Algebra die Borel'sche ist, d.h. die σ-Algebra, die von einer Topologie erzeugt wird. In diesem Abschnitt werden wir oft voraussetzen, dass der topologische Raum polnisch ist, also separabel und metrisierbar durch eine vollständige Metrik. Um uns Schreibarbeit zu ersparen, setzen wir nun gleich voraus, dass (E, r) ein metrischer Raum ist und manchmal werden wir voraussetzen, dass er vollständig und separabel ist.

Für eine messbare Abbildung $f : E \to \mathbb{R}$ und ein Maß μ auf $\mathcal{B}(E)$ (der Borel'schen σ-Algebra von E) werden wir in diesem und im nächsten Kapitel durchgehend $\mu[f] := \int f d\mu$ schreiben.

10.1 Definition und einfache Eigenschaften

Bisher haben wir uns mit verschiedenen Konvergenzarten von Zufallsvariablen beschäftigt. Die Konvergenz in Verteilung von Zufallsvariablen ist dasselbe wie die schwache Konvergenz der Verteilungen der Zufallsvariablen. Zur Motivation hinter den folgenden Definitionen sei an ein Faktum erinnert: in einem metrischen Raum (E, r) ist $x_n \xrightarrow{n\to\infty} x$ genau dann wenn $f(x_n) \xrightarrow{n\to\infty} f(x)$ für alle stetigen Funktionen auf E (d.h. $f \in \mathcal{C}(E, \mathbb{R})$).

Definition 10.1 (Schwache Konvergenz und Konvergenz in Verteilung).

1. *Wir bezeichnen mit $\mathcal{P}(E)$ die Menge der Wahrscheinlichkeitsmaße auf $\mathcal{B}(E)$ und mit $\mathcal{P}_{\leq 1}(E)$ die Menge der endlichen Maße μ auf $\mathcal{B}(E)$ mit $\mu(E) \leq 1$. Weiter ist $\mathcal{C}_b(E)$ die Menge der reellwertigen, beschränkten, stetigen Funktionen auf E und $\mathcal{C}_c(E) \subseteq \mathcal{C}_b(E)$ die Menge der reellwertigen, beschränkten stetigen Funktionen auf E mit kompaktem Träger.*

2. *Eine Folge* $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$ *konvergiert* schwach *gegen* $\mathbf{P} \in \mathcal{P}(E)$*, falls*
$$\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f] \tag{10.1}$$
für alle $f \in \mathcal{C}_b(E)$*. Wir schreiben dann*
$$\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}.$$

3. *Ist* $\mu_1, \mu_2, \dots \in \mathcal{P}_{\leq 1}$ *und* μ *ein Maß auf* E*. Gilt* (10.1) *nur für alle* $f \in \mathcal{C}_c(E)$*, so sagen wir,* μ_n *konvergiert* vage *gegen* μ*. Wir schreiben dann*
$$\mu_n \xRightarrow{n\to\infty}_v \mu.$$

4. *Seien* $X, X_1, X_2, \dots$ *Zufallsvariablen auf Wahrscheinlichkeitsräumen* $(\Omega, \mathcal{A}, \mathbf{P})$, $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1)$, $(\Omega_2, \mathcal{A}_2, \mathbf{P}_2), \dots$ *mit Werten in* E*. Dann konvergiert* $X_1, X_2, \dots$ in Verteilung *gegen* X*, falls* $(X_n)_*\mathbf{P}_n \xRightarrow{n\to\infty} X_*\mathbf{P}$*. Wir schreiben dann*
$$X_n \xRightarrow{n\to\infty} X.$$

Bemerkung 10.2. 1. Man beachte, dass für Zufallsvariablen $X, X_1, X_2, \dots$ mit Werten in E genau dann $X_n \xRightarrow{n\to\infty} X$ gilt, wenn
$$\mathbf{P}[f(X_n)] \xrightarrow{n\to\infty} \mathbf{P}[f(X)]$$
für alle $f \in \mathcal{C}_b(E)$ gilt. Viele der folgenden Resultate lassen sich deswegen auf zwei Arten und Weisen formulieren: entweder mittels Wahrscheinlichkeitsverteilungen, oder mittels Zufallsvariablen. Der Zusammenhang ist dabei immer, dass die Aussage über die Wahrscheinlichkeitsverteilungen ebenfalls eine Aussage über die Verteilungen der Zufallsvariablen ist.

2. Der schwache Limes von Wahrscheinlichkeitsmaßen muss wieder ein Wahrscheinlichkeitsmaß sein, da $1 \in \mathcal{C}_b(E)$. Der vage Limes von Wahrscheinlichkeitsmaßen muss jedoch nicht unbedingt wieder ein Wahrscheinlichkeitsmaß sein, da $1 \notin \mathcal{C}_c(E)$, falls E nicht kompakt ist; siehe auch Beispiel 10.3.1. Immerhin sind vage Grenzwerte in $\mathcal{P}_{\leq 1}(E)$, wie Lemma 10.12.

3. Wir kennen bereits die fast sicher Konvergenz, die stochastische und die Konvergenz in $\mathcal{L}^p$ von Zufallsvariablen $X_1, X_2, \dots$ gegen X. Entscheidender Unterschied zur Konvergenz in Verteilung ist, dass letztere nicht voraussetzt, dass die Zufallsvariablen auf demselben Wahrscheinlichkeitsraum definiert sind.

4. Nach Definition 10.1 ist die Topologie der schwachen Konvergenz auf $\mathcal{P}(E)$ die schwächste (d.h. die kleinste) Topologie, für die $\mathbf{P} \mapsto \mathbf{P}[f]$ für alle $f \in \mathcal{C}_b(E)$ stetig ist.

Beispiel 10.3. 1. Sei $x, x_1, x_2, \dots \in \mathbb{R}$ mit $x_n \xrightarrow{n\to\infty} x$ sowie $\mathbf{P} = \delta_x, \mathbf{P}_1 = \delta_{x_1}, \mathbf{P}_2 = \delta_{x_2}, \dots$ Dann gilt $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$, da
$$\mathbf{P}_n[f] = f(x_n) \xrightarrow{n\to\infty} f(x) = \mathbf{P}[f]$$

für alle $f \in \mathcal{C}_b(\mathbb{R})$.

Falls die Folge $x_1, x_2, \ldots$ divergiert, etwa $x_n = n$, so gilt $\mathbf{P}_n \xrightarrow{n\to\infty}_v 0$ (das ist das 0-Maß auf $\mathcal{B}(\mathbb{R})$), da

$$\mathbf{P}_n[f] = f(x_n) \xrightarrow{n\to\infty} 0 = 0[f]$$

für alle $f \in \mathcal{C}_c(\mathbb{R})$ gilt. Allerdings gilt die schwache Konvergenz nicht, da $\mathbf{P}_n[1] = 1 \neq 0 = 0[1]$.

2. Seien $X, X_1, X_2, \ldots$ identisch verteilt. Dann ist $X_n \xRightarrow{n\to\infty} X$, jedoch gilt die Konvergenz im allgemeinen weder fast sicher, noch stochastisch oder in $\mathcal{L}^p$.

3. Wie wir noch sehen werden, ist der zentrale Grenzwertsatz, den wir in Kapitel 11 kennenlernen werden (Theorem 11.8), ein Resultat über Konvergenz in Verteilung. In seiner einfachsten Form, dem Satz von deMoivre-Laplace (siehe auch Bemerkung 10.8 und Beispiel 10.34), besagt dieser: sei $p \in (0,1)$, $X_n \sim \mathrm{B}(n,p), n = 1, 2, \ldots$ und $X \sim N(0,1)$. Dann gilt
$$\frac{X_n - np}{\sqrt{np(1-p)}} \xRightarrow{n\to\infty} X.$$

4. Ebenso ist die Poisson-Approximation der Binomialverteilung eine Aussage über Konvergenz in Verteilung (siehe auch Theorem 11.5): sei $X_n \sim \mathrm{B}(n, p_n), n = 1, 2, \ldots$ mit $n \cdot p_n \xrightarrow{n\to\infty} \lambda$ und $X \sim \mathrm{Poi}(\lambda)$. Dann gilt
$$X_n \xRightarrow{n\to\infty} X.$$

Lemma 10.4 (Eindeutigkeit des schwachen Limes). *Seien $\mu, \nu, \mu_1, \mu_2, \cdots \in \mathcal{P}(E)$ mit $\mu_n \xRightarrow{n\to\infty} \mu$ und $\mu_n \xRightarrow{n\to\infty} \nu$. Dann ist $\mu = \nu$.*

Beweis. Nach Proposition 3.10 genügt es zu zeigen, dass $\mu(A) = \nu(A)$ für alle abgeschlossenen $A \subseteq E$. Die Menge aller abgeschlossenen Mengen ist nämlich ein schnittstabiler Erzeuger von $\mathcal{B}(E)$. Sei also $A \subseteq E$ abgeschlossen. Wir setzen

$$r(x, A) := \inf_{y \in A} r(x, y)$$

sowie

$$f_m(x) \mapsto (1 - m \cdot r(x, A))^+.$$

für $m = 1, 2, \ldots$. Dann ist $f_m \xrightarrow{m\to\infty} 1_A$, da A abgeschlossen ist. Weiter gilt mit majorisierter Konvergenz

$$\mu(A) = \lim_{m\to\infty} \mu[f_m] = \lim_{m\to\infty} \lim_{n\to\infty} \mu_n[f_m] = \lim_{m\to\infty} \nu[f_m] = \nu(A)$$

und die Behauptung folgt. □

Wir erinnern an die Eingangsgrafik von Kapitel 8. Eine Folge von Zufallsvariablen kann fast sicher, in Wahrscheinlichkeit, in $\mathcal{L}^p$ oder in Verteilung konvergieren. Die Konvergenz in Verteilung ist der schwächste dieser Begriffe in folgendem Sinne.

Proposition 10.5 (Konvergenz in Wahrscheinlichkeit und in Verteilung). *Seien $X, X_1, X_2, \ldots$ Zufallsvariable mit Werten in E. Falls $X_n \xrightarrow{n\to\infty}_p X$, so auch $X_n \xRightarrow{n\to\infty} X$. Ist X konstant, so gilt auch die Umkehrung.*

Beweis. Sei $X_n \xrightarrow{n\to\infty}_p X$. Angenommen, es gäbe ein $f \in \mathcal{C}_b(E)$ so, dass $\lim_{n\to\infty} \mathbf{P}[f(X_n)] \neq \mathbf{P}[f(X)]$. Dann gibt es eine Teilfolge $(n_k)_{k=1,2,\dots}$ und ein $\varepsilon > 0$ mit

$$\lim_{k\to\infty} |\mathbf{P}[f(X_{n_k})] - \mathbf{P}[f(X)]| > \varepsilon. \tag{10.2}$$

Wegen $X_n \xrightarrow{n\to\infty}_p X$ und Proposition 8.6 gibt es eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$, so dass $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty} X$ fast sicher. Wegen majorisierter Konvergenz wäre damit auch

$$\lim_{\ell\to\infty} \mathbf{P}[f(X_{n_{k_\ell}})] = \mathbf{P}[f(X)]$$

im Widerspruch zu (10.2).

Für die Umkehrung sei $X = s \in E$. Es ist $x \mapsto r(x,s) \wedge 1$ eine beschränkte, stetige Funktion und damit

$$\mathbf{P}[r(X_n,s) \wedge 1] \xrightarrow{n\to\infty} \mathbf{P}[r(X,s)\wedge 1] = 0.$$

Also gilt $X_n \xrightarrow{n\to\infty}_p X$ wegen (8.1). □

Theorem 10.6 (Portmanteau Theorem). *Seien $X, X_1, X_2, \dots$ Zufallsvariablen mit Werten in E. Folgende Bedingungen sind äquivalent:*

(i) $X_n \xRightarrow{n\to\infty} X$

(ii) $\mathbf{P}[f(X_n)] \xrightarrow{n\to\infty} \mathbf{P}[f(X)]$ *für alle beschränkten, Lipschitz-stetigen Funktionen f.*

(iii) $\liminf_{n\to\infty} \mathbf{P}(X_n \in G) \geq \mathbf{P}(X \in G)$ *für alle offenen $G \subseteq E$.*

(iv) $\limsup_{n\to\infty} \mathbf{P}(X_n \in F) \leq \mathbf{P}(X \in F)$ *für alle abgeschlossenen $F \subseteq E$.*

(v) $\lim_{n\to\infty} \mathbf{P}(X_n \in B) = \mathbf{P}(X \in B)$ *für alle $B \in \mathcal{B}(E)$ mit*[15] $\mathbf{P}(X \in \partial B) = 0$.

Beweis. *(i)* ⇒ *(ii)*: klar

(ii) ⇒ *(iv)* Sei $F \subseteq E$ abgeschlossen und $f_1, f_2, \dots$ Lipschitz-stetig, so dass $f_k \downarrow 1_F$. (Beispielsweise wählt man $\varepsilon_k \downarrow 0$ und $f_k(x) = (1 - \frac{1}{\varepsilon_k} r(x,F))^+$, wobei $r(x,F) := \inf_{y\in F} r(x,y)$.) Damit gilt

$$\limsup_{n\to\infty} \mathbf{P}(X_n \in F) \leq \inf_{k=1,2,\dots} \limsup_{n\to\infty} \mathbf{P}[f_k(X_n)] = \inf_{k=1,2,\dots} \mathbf{P}[f_k(X)] = \mathbf{P}(X \in F).$$

(iii) $\iff$ *(iv)* Das ist klar. Man muss in der Richtung *(iii)* ⇒ *(iv)* nur $F := E \setminus G$ und in der Richtung *(iv)* ⇒ *(iii)* analog $G := E \setminus F$ setzen.

(iii) ⇒ *(i)* Sei $f \geq 0$ stetig. Wegen Proposition 7.10 und Fatou's Lemma ist damit

$$\begin{aligned}\mathbf{P}[f(X)] &= \int_0^\infty \mathbf{P}(f(X) > t)dt \leq \int_0^\infty \liminf_{n\to\infty} \mathbf{P}(f(X_n) > t)dt \\ &\leq \liminf_{n\to\infty} \int_0^\infty \mathbf{P}(f(X_n) > t)dt = \liminf_{n\to\infty} \mathbf{P}[f(X_n)].\end{aligned}$$

[15] Für den Abschluss $\overline{B}$ und das Innere B° bezeichne hier $\partial B := \overline{B} \setminus B^\circ$ den Rand von B.

Für $-c < f < c$ gilt damit, da $-f + c \geq 0$ ist,

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}[f(X_n)] &= c - \liminf_{n\to\infty} \mathbf{P}[-f(X_n) + c] \leq c - \mathbf{P}[-f(X) + c] = \mathbf{P}[f(X)] \\ &\leq \liminf_{n\to\infty} \mathbf{P}[f(X_n)],\end{aligned}$$

also $\mathbf{P}[f(X_n)] \xrightarrow{n\to\infty} \mathbf{P}[f(X)]$.
$(iii), (iv) \Rightarrow (v)$ Für $B \in \mathcal{B}(E)$ ist

$$\mathbf{P}(X \in B^\circ) \leq \liminf_{n\to\infty} \mathbf{P}(X_n \in B^\circ) \leq \limsup_{n\to\infty} \mathbf{P}(X_n \in \overline{B}) \leq \mathbf{P}(X \in \overline{B}).$$

Gegeben $\mathbf{P}(X \in \partial B) = \mathbf{P}(X \in \overline{B}) - \mathbf{P}(X \in B^\circ) = 0$, ist damit $\mathbf{P}(X_n \in B) \xrightarrow{n\to\infty} \mathbf{P}(X \in B)$.
$(v) \Rightarrow (iv)$ Angenommen, (v) trifft zu und $F \subseteq E$ ist abgeschlossen. Wir schreiben $F^\varepsilon := \{x \in E : r(x, F) \leq \varepsilon\}$ für $\varepsilon > 0$. Die Mengen $\partial F^\varepsilon \subseteq \{x : r(x, F) = \varepsilon\}$ sind disjunkt, also gilt

$$\mathbf{P}(X \in \partial F^\varepsilon) = 0 \tag{10.3}$$

für Lebesgue-fast jedes ε. Sei $\varepsilon_1, \varepsilon_2, \ldots$ eine Folge mit $\varepsilon_k \downarrow 0$, so dass (10.3) für alle $\varepsilon_1, \varepsilon_2, \ldots$ gilt. Damit ist

$$\limsup_{n\to\infty} \mathbf{P}(X_n \in F) \leq \inf_{k=1,2,\ldots} \limsup_{n\to\infty} \mathbf{P}(X_n \in F^{\varepsilon_k}) = \inf_{k=1,2,\ldots} \mathbf{P}(X \in F^{\varepsilon_k}) = \mathbf{P}(X \in F).$$

□

Korollar 10.7 (Konvergenz von Verteilungsfunktionen). *Seien $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ mit Verteilungsfunktionen $F, F_1, F_2, \ldots$ Dann gilt $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ genau dann wenn $F_n(x) \xrightarrow{n\to\infty} F(x)$ für alle Stetigkeitsstellen x von F.*

Beweis. '$\Rightarrow$': Ist x eine Stetigkeitsstelle von F, so ist $\mathbf{P}(\partial(-\infty; x]) = \mathbf{P}(\{x\}) = 0$. Damit gilt nach Theorem 10.6 (Richtung $(i) \Rightarrow (v)$), dass

$$F_n(x) = \mathbf{P}_n((-\infty; x]) \xrightarrow{n\to\infty} \mathbf{P}((-\infty; x]) = F(x).$$

'$\Leftarrow$': Nach Theorem 10.6 (Richtung $(ii) \Rightarrow (i)$) genügt es zu zeigen, dass $\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f]$ für alle beschränkten, Lipschitzfunktionen f. O.E. nehmen wir an, dass $|f| \leq 1$ und f Lipschitz-Konstante 1 hat. Für $\varepsilon > 0$ wähle $N \in \mathbb{N}$ und Stetigkeitspunkte $y_0 < \cdots < y_N$ von F so, dass $F(y_0) < \varepsilon$, $F(y_N) > 1 - \varepsilon$ und $y_i - y_{i-1} < \varepsilon$ für $i = 1, \ldots, N$. Dann ist $F_n(y_i) \xrightarrow{n\to\infty} F(y_i)$ und

$$f \leq 1_{(-\infty, y_0]} + 1_{(y_N, \infty)} + \sum_{i=1}^{N-1} (f(y_i) + \varepsilon) 1_{(y_i, y_{i+1}]},$$

also

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}_n[f] &\leq \limsup_{n\to\infty} F_n(y_0) + 1 - F_n(y_N) + \sum_{i=1}^{N} (f(y_i) + \varepsilon)(F_n(y_i) - F_n(y_{i-1}) \\ &\leq 3\varepsilon + \sum_{i=1}^{N} f(y_i)(F(y_i) - F(y_{i-1})) \leq 4\varepsilon + \mathbf{P}[f]\end{aligned}$$

Mit $\varepsilon \to 0$ und durch Ersetzen von f mit $1 - f$ folgt $\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f]$. □

Bemerkung 10.8 (Der Satz von deMoivre-Laplace). In Beispiel 10.3 hatten wir behauptet, dass der Satz von deMoivre-Laplace eine Aussage über schwache Konvergenz macht. Man zeigt gewöhnlich, dass für $B(n,p)$-verteilte Zufallsgrößen X_n gilt, dass

$$\mathbf{P}\Big(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x\Big) \xrightarrow{n\to\infty} \Phi(x),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist. Wie Korollar 10.7 zeigt, bedeutet das genau die Konvergenz in Verteilung gegen eine Standardnormalverteilung.

Korollar 10.9 (Satz von Slutzky). *Seien $X, X_1, X_2, ..., Y_1, Y_2, ...$ Zufallsvariable mit Werten in E. Gilt $X_n \xRightarrow{n\to\infty} X$ und $r(X_n, Y_n) \xrightarrow{n\to\infty}_p 0$, so gilt auch $Y_n \xRightarrow{n\to\infty} X$.*

Beweis. Siehe Übung. □

Theorem 10.10 (Continuous mapping theorem). *Sei E separabel, (E', r') ein weiterer metrischer Raum und $\varphi : E \to E'$ messbar und $U_\varphi \subseteq E$ die Menge der Unstetigkeitsstellen von φ.*

1. *Sind $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$ und $\mathbf{P}(U_\varphi) = 0$ sowie $\mathbf{P}_i \xRightarrow{n\to\infty} \mathbf{P}$, dann ist auch $\varphi_*\mathbf{P}_n \xRightarrow{n\to\infty} \varphi_*\mathbf{P}$.*
2. *Sind $X, X_1, X_2, \dots$ Zufallsvariable mit Werten in E und $\mathbf{P}(X \in U_\varphi) = 0$ und $X_n \xRightarrow{n\to\infty} X$, dann ist auch $\varphi(X_n) \xRightarrow{n\to\infty} \varphi(X)$.*

Beweis. Zunächst bemerken wir, dass 2. eine Anwendung von 1. ist, falls man $\mathbf{P}_n = (X_n)_*\mathbf{P}$ setzt. Die Menge U_φ ist Borel-messbar, da

$$U_\varphi^{\delta,\varepsilon} := \{x \in E : \exists y, z \in B_\delta(x), r'(\varphi(y), \varphi(z)) > \varepsilon\}$$

Borel-messbar ist (hier geht die Separabilität von E ein) und

$$U_\varphi = \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} U_\varphi^{1/k,1/n}.$$

Sei $G \subseteq E'$ offen und $x \in \varphi^{-1}(G) \cap U_\varphi^c$. Da φ in x also stetig ist, gibt es ein $\delta > 0$ mit $\varphi(y) \in G$ (d.h. $y \in \varphi^{-1}(G)$) für alle y mit $r(x,y) < \delta$. Also ist $\varphi^{-1}(G) \cap U_\varphi^c \subseteq (\varphi^{-1}(G))^\circ$. Daraus folgt mit Theorem 10.6 (Richtung $(i) \Rightarrow (iii)$)

$$\begin{aligned}\varphi_*\mathbf{P}(G) &= \mathbf{P}(\varphi^{-1}(G)) = \mathbf{P}(\varphi^{-1}(G) \cap U_\varphi^c) \leq \mathbf{P}((\varphi^{-1}(G))^\circ) \\ &\leq \liminf_{n\to\infty} \mathbf{P}_n((\varphi^{-1}(G))^\circ) \leq \liminf_{n\to\infty} \mathbf{P}_n(\varphi^{-1}(G)) = \liminf_{n\to\infty} \varphi_*\mathbf{P}_n(G).\end{aligned}$$

Wieder wegen Theorem 10.6 (Richtung $(iii) \Rightarrow (i)$) folgt also $\varphi_*\mathbf{P}_n \xRightarrow{n\to\infty} \varphi_*\mathbf{P}$. □

Sieht man von der vagen Konvergenz einmal ab, so ist die Konvergenz in Verteilung der schwächste Konvergenzbegriff. Jedoch besteht ein Zusammenhang mit der fast sicheren Konvergenz, wie folgendes Theorem zeigt.

Theorem 10.11 (Schwache und fast sichere Konvergenz, Skorohod). *Seien $X, X_1, X_2, \dots$ Zufallsvariablen mit Werten in einem vollständigen und separablen Raum (E, r). Dann gilt $X_n \xRightarrow{n\to\infty} X$ genau dann, wenn es einen Wahrscheinlichkeitsraum gibt, auf dem Zufallsgrößen $Y, Y_1, Y_2, \dots$ definiert sind mit $Y_n \xrightarrow{n\to\infty}_{fs} Y$ und $Y \stackrel{d}{=} X, Y_1 \stackrel{d}{=} X_1, Y_2 \stackrel{d}{=} X_2, \dots$*

Beweis. '$\Leftarrow$': klar, da aus der fast sicheren die schwache Konvergenz nach Proposition 10.5 folgt.
'$\Rightarrow$': Als Wahrscheinlichkeitsraum erweitern wir den Raum, auf dem X definiert ist, setzen also insbesondere $Y = X$. Sei zunächst $E = \{1, \dots, m\}$ endlich. Sei U uniform auf $[0,1]$ verteilt und unabhängig von Y, sowie $W_1, W_2, \dots$ unabhängig mit

$$\mathbf{P}(W_n = k) = \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l=1}^m \mathbf{P}(X = l) \wedge \mathbf{P}(X_n = l)}.$$

Wir setzen $Y_n = k$ falls entweder

$$X = k \text{ und } U \leq \frac{\mathbf{P}(X_n = k)}{\mathbf{P}(X = k)}$$

oder

$$X = l \text{ und } U > \frac{\mathbf{P}(X_n = l)}{\mathbf{P}(X = l)} \text{ und } W_n = k.$$

Dann ist

$$\begin{aligned}
\mathbf{P}(Y_n = k) &= \mathbf{P}(X = k) \cdot \frac{\mathbf{P}(X_n = k)}{\mathbf{P}(X = k)} \wedge 1 \\
&\quad + \sum_{l=1}^m \mathbf{P}(X = l) \cdot \Big(1 - \frac{\mathbf{P}(X_n = l)}{\mathbf{P}(X = l)}\Big)^+ \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l'=1}^m \mathbf{P}(X = l') \wedge \mathbf{P}(X_n = l')} \\
&= \mathbf{P}(X_n = k) \wedge \mathbf{P}(X = k) \\
&\quad + \sum_{l=1}^m (\mathbf{P}(X = l) - \mathbf{P}(X_n = l) \wedge \mathbf{P}(X = l)) \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l'=1}^m \mathbf{P}(X = l') \wedge \mathbf{P}(X_n = l')} \\
&= \mathbf{P}(X_n = k).
\end{aligned}$$

Damit ist $Y_n \stackrel{d}{=} X_n$. Da nach Voraussetzung $\mathbf{P}(X_n = k) \xrightarrow{n\to\infty} \mathbf{P}(X = k)$, folgt die fast sichere Konvergenz.

Für allgemeines E wählen wir für jedes p eine Partition von E in Mengen $B_1, B_2, \dots$ in E mit $\mathbf{P}(Y \in \partial B_k) = 0$ und Durchmesser höchstens 2^{-p}. Wähle m groß genug, damit $\mathbf{P}(Y \notin B_0) < 2^{-p}$ mit $B_0 := E \setminus \bigcup_{k \leq m} B_k$. Für $k = 1, 2, \dots,$ definiere Zufallsgrößen $\widetilde{Z}, \widetilde{Z}_1, \widetilde{Z}_2, \dots$ so, dass $\widetilde{Z} = k$ genau dann, wenn $Y \in B_k$ und $\widetilde{Z}_n = k$ wenn $Y_n \in B_k$. Dann gilt $\widetilde{Z}_n \xRightarrow{n\to\infty} \widetilde{Z}$. Da $\widetilde{Z}, \widetilde{Z}_1, \dots$ nur Werte in einer endlichen Menge annehmen, können wir damit Zufallsgrößen $Z, Z_1, Z_2, \dots$ definieren mit $Z_n \xrightarrow{n\to\infty}_{fs} Z$. Weiter seien $W_{n,k}$ Zufallsgrößen mit Verteilung $\mathbf{P}[X_n \in .|X_n \in B_k]$ und $\widetilde{Y}_{n,p} = \sum_k W_{n,k} 1_{Z_n = k}$, so dass $\widetilde{Y}_{n,p} \stackrel{d}{=} X_n$ für alle n. Klar ist nun

$$\Big\{r(\widetilde{Y}_{n,p}, Y) > 2^{-p}\Big\} \subseteq \{Z_n \neq Z\} \cup \{Y \in B_0\}.$$

Da $Z_n \xrightarrow{n\to\infty}_{fs} Z$ und $\mathbf{P}\{Y \in B_0\} < 2^{-p}$, gibt es für jedes p Zahlen $n_1 < n_2 < \dots$ mit

$$\mathbf{P}\Big(\bigcup_{n \geq n_p} \big\{r(\widetilde{Y}_{n,p}, Y) > 2^{-p}\big\}\Big) < 2^{-p}$$

für alle p. Mit dem Borel-Cantelli-Lemma bekommen wir

$$\sup_{n \geq n_p} r(\widetilde{Y}_{n,p}, Y) \leq 2^{-p}$$

für fast alle p. Wir definieren also $Y_n := \widetilde{Y}_{n,p}$ für $n_p \leq n < n_{p+1}$ und beachten, dass $X_n \stackrel{d}{=} Y_n \xrightarrow{n\to\infty}_{fs} Y$. $\square$

10.2 Der Satz von Prohorov

In diesem Abschnitt beleuchten wir zunächst den Begriff der vagen Konvergenz. Dabei werden wir uns auf den Raum $E = \mathbb{R}$ beschränken. (Die meisten der hier gezeigten Aussagen gelten auch noch in lokal-kompakten Räumen.) Klar ist schon, dass aus der schwachen Konvergenz von Verteilungen die vage folgt, und dass die schwache Konvergenz äquivalent mit der Konvergenz der Verteilungsfunktionen ist (Korollar 10.7). Hauptresultat ist hier der Satz von Helly (Theorem 10.13), der angibt, dass jede Folge von Wahrscheinlichkeitsmaßen ein vage konvergente Teilfolge hat.

Anschließend untersuchen wir die Frage, inwieweit eine Folge von Wahrscheinlichkeitsmaßen auch schwach konvergente Teilfolgen besitzt. Dies führt uns auf den Begriff der Straffheit von Wahrscheinlichkeitsmaßen und den Satz von Prohorov (Theorem 10.19).
Wie wir bereits in Bemerkung 10.2.1 gesehen haben, kann es sein, dass das Grenzmaß bei vager Konvergenz kein Wahrscheinlichkeitsmaß ist. Allerdings zeigt nachfolgendes Resultat, dass das Grenzmaß zumindest Gesamtmasse höchstens 1 besitzt.

Lemma 10.12 (Massenverlust bei vager Konvergenz). *Sei $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ und $\mu \in \mathcal{P}(\mathbb{R})$ mit $\mathbf{P}_n \xRightarrow{n\to\infty}_v \mu$, so gilt $\mu(\mathbb{R}) \leq 1$.*

Beweis. Sei $f_1, f_2, \dots \in \mathcal{C}_c(\mathbb{R})$ mit $f_k \uparrow 1$. Dann gilt mit monotoner Konvergenz

$$\mu(\mathbb{R}) = \sup_{k\in\mathbb{N}} \mu[f_k] = \sup_{k\in\mathbb{N}} \limsup_{n\to\infty} \mathbf{P}_n[f_k] \leq 1.$$

$\square$

Theorem 10.13 (Satz von Helly). *Sei $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$. Dann gibt es eine Teilfolge $(n_k)_{k=1,2,\dots}$ und ein $\mu \in \mathcal{P}_{\leq 1}(\mathbb{R})$ mit $\mathbf{P}_{n_k} \xRightarrow{k\to\infty}_v \mu$.*

Beweis. Seien $F_1, F_2, \dots$ die Verteilungsfunktionen von $\mathbf{P}_1, \mathbf{P}_2, \dots$ Weiter sei $(x_1, x_2, \dots)$ eine Abzählung von $\mathbb{Q}$. Da $[0,1]$ kompakt ist, gibt es zu jeder Folge $(F_n(x_i))_{n=1,2,\dots}$ eine konvergente Teilfolge. Mittels eines Diagonal-Argumentes gibt es eine Folge $(n_k)_{k=1,2,\dots}$, so dass $(F_{n_k}(x_i))_{k=1,2,\dots}$ für alle i gegen einen Limes $G(x_i)$ auf $\mathbb{Q}$ konvergiert. Wir definieren

$$F(x) := \inf\{G(r) : r \in \mathbb{Q}, r > x\}.$$

Da alle F_n und damit G nicht-negative Zuwächse haben, gilt dasselbe für F. Aus der Definition von F und der Monotonie von G folgt weiter, dass F rechtsstetig ist. Nach Proposition 3.18 gibt es ein Maß μ auf $\mathbb{R}$ mit $\mu((x,y]) = F(y) - F(x)$ für alle $x, y \in \mathbb{R}, x \leq y$. Es bleibt zu zeigen, dass $\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mu[f]$ für alle $f \in \mathcal{C}_c(\mathbb{R})$. O.E. können wir annehmen, dass $f \geq 0$ ist.

Es ist $F_n(x) \xrightarrow{n\to\infty} F(x)$ an allen Stetigkeitsstellen x von F nach Konstruktion. Es gibt eine abzählbare Menge $D \subseteq \mathbb{R}$, so dass F auf D^c stetig ist. Damit ist $\mathbf{P}_n(U) \xrightarrow{n\to\infty} \mu(U)$ für alle endlichen Vereinigungen U von Intervallen mit Ecken in D^c. Sei nun $B \subseteq \mathbb{R}$ offen und beschränkt. Seien $U_1, U_2, \dots$ und $V_1, V_2, \dots$ Folgen von endlichen Vereinigungen von offenen Intervallen mit Ecken in C, so dass $U_k \uparrow B$, $V_k \downarrow \overline{B}$. Dann gilt

$$\begin{aligned}\mu(B) &= \lim_{k\to\infty} \mu(U_k) = \lim_{k\to\infty} \liminf_{n\to\infty} \mathbf{P}_n(U_k) \leq \liminf_{n\to\infty} \mathbf{P}_n(B) \\ &\leq \limsup_{n\to\infty} \mathbf{P}_n(B) \leq \lim_{k\to\infty} \limsup_{n\to\infty} \mathbf{P}_n(V_k) = \lim_{k\to\infty} \mu(V_k) = \mu(\overline{B}).\end{aligned}$$

Da $\mu(f = t) > 0$ für höchstens abzählbar viele t, und da $\mathbf{P}_n(f > t) \le 1_{t \ge ||f||}$, folgt mit majorisierter Konvergenz

$$\begin{aligned}\mu[f] &= \int_0^\infty \mu(f > t)dt \le \int_0^\infty \liminf_{n\to\infty} \mathbf{P}_n(f > t)dt = \liminf_{n\to\infty} \int_0^\infty \mathbf{P}_n(f > t)dt = \liminf_{n\to\infty} \mathbf{E}_n[f] \\ &\le \limsup_{n\to\infty} \mathbf{P}_n[f] = \limsup_{n\to\infty} \int_0^\infty \mathbf{P}_n(f > t)dt = \int_0^\infty \limsup_{n\to\infty} \mathbf{P}_n(f > t)dt \le \int_0^\infty \mu(f \ge t) \\ &= \mu[f]. \end{aligned}$$ □

Wir kehren nun zurück zum Fall eines allgemeinen metrischen Raumes (E, r). Um die Existenz von Häufungspunkten im Sinne der schwachen Konvergenz zu zeigen, muss sichergestellt sein, dass Grenzmaße wieder Wahrscheinlichkeitsmaße sind. Insbesondere darf also beim Grenzübergang keine Masse verloren gehen wie bei der vagen Konvergenz (siehe Lemma 10.12). Hierbei ist das Konzept der *Straffheit* zentral.

Definition 10.14 (Straffheit). *Sei $\mathcal{K}$ das System aller kompakten Mengen in E. Eine Familie $(\mathbf{P}_i)_{i\in I}$ in $\mathcal{P}(E)$ ist* straff*, falls*

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} \mathbf{P}_i(K) = 1.$$

Eine Familie $(X_i)_{i\in I}$ von E-wertigen Zufallsvariablen ist straff*, falls $((X_i)_*\mathbf{P})_{i\in I}$ straff ist, d.h.*

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} \mathbf{P}(X_i \in K) = 1.$$

Bemerkung 10.15 (Äquivalente Formulierungen).

1. Die Definition der Straffheit einer Familie $(\mathbf{P}_i)_{i\in I}$ in $\mathcal{P}(E)$ ist äquivalent zu folgender Bedingung: für alle $\varepsilon > 0$ gibt es $K \subseteq E$ kompakt mit $\inf_{i\in I} \mathbf{P}_i(K) \ge 1 - \varepsilon$.

2. Falls $E = \mathbb{R}^d$, ist eine Familie $(\mathbf{P}_i)_{i\in I}$ genau dann straff, wenn

$$\sup_{r>0} \inf_{i\in I} \mathbf{P}_i(B_r(0)) = 1,$$

wobei $B_r(0)$ die Kugel um 0 mit Radius r ist.

3. In Lemma 3.8 haben wir gezeigt, dass $\mathbf{P} \in \mathcal{P}(E)$ straff ist, falls (E, r) vollständig und separabel ist. Daraus folgt auch, dass jede endliche Familie von Wahrscheinlichkeitsmaßen auf der Borel'schen σ-Algebra eines polnischen Raumes straff ist.

4. Weiter ist eine abzählbare Familie $(\mathbf{P}_i)_{i=1,2,\dots}$ von Wahrscheinlichkeitsmaßen auf einem polnischen Raum (E, r) genau dann straff, falls

$$\sup_{K\in\mathcal{K}} \liminf_{i=1,2,\dots} \mathbf{P}_i(K) = 1.$$

Beweis. '⇒': klar, da $\liminf_{i=1,2,\dots} \mathbf{P}_i(K) \ge \inf_{i=1,2,\dots} \mathbf{P}_i(K) = 1$.

'⇐': Sei $\varepsilon > 0$ und K so, dass $\liminf_{i=1,2,\dots} \mathbf{P}_i(K) \ge 1 - \varepsilon/2$. Wähle N so, dass $\inf_{i=N+1,N+2,\dots} \mathbf{P}_i(K) \ge 1 - \varepsilon$ und $K_1, \dots, K_N$ kompakt so, dass $\mathbf{P}_i(K_i) \ge 1 - \varepsilon$ für $i = 1, \dots, N$. Da $\widetilde{K} = K \cup K_1 \cup \dots \cup K_N$ kompakt ist und $\inf_{i=1,2,\dots} \mathbf{P}_i(\widetilde{K}) \ge 1 - \varepsilon$ folgt die Straffheit von $(\mathbf{P}_i)_{i=1,2,\dots}$. □

Beispiel 10.16 (Straffe Mengen von Wahrscheinlichkeitsmaßen). 1. Ist E kompakt, so ist jede Familie von Wahrscheinlichkeitsmaßen auf $\mathcal{B}(E)$ straff.

2. Eine Familie $(X_i)_{i\in I}$ von reellwertigen Zufallsvariablen mit

$$\sup_{i\in I} \mathbf{P}[|X_i|] < \infty,$$

ist straff. Denn es gilt

$$\inf_{r>0} \sup_{i\in I} \mathbf{P}(|X_i| \geq r) \leq \inf_{r>0} \sup_{i\in I} \frac{\mathbf{P}[|X_i|]}{r} = 0.$$

3. Die Familie $(\delta_n)_{n=1,2,\ldots}$, wobei δ_n das Diracmaß auf n ist, ist nicht straff.

Lemma 10.17 (Vage Konvergenz und Straffheit). *Seien* $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ *und* $\mu \in \mathcal{P}_{\leq 1}(\mathbb{R})$ *mit*

$$\mathbf{P}_n \xRightarrow{n\to\infty}_v \mu.$$

Dann ist

$$\mu(\mathbb{R}) = 1 \qquad \Longleftrightarrow \qquad (\mathbf{P}_n)_{n=1,2,\ldots} \text{ ist straff.}$$

In diesem Fall gilt $\mathbf{P}_n \xRightarrow{n\to\infty} \mu$.

Beweis. Für $r > 0$ wähle ein $g_r \in \mathcal{C}_c(\mathbb{R})$, $1_{B_r(0)} \leq g_r \leq 1_{B_{r+1}(0)}$. Dann ist $(\mathbf{P}_n)_{n=1,2,\ldots}$ genau dann straff, wenn

$$\sup_{r>0} \liminf_{n\to\infty} \mathbf{P}_n[g_r] = 1.$$

'$\Rightarrow$': Es gilt, da μ von unten stetig ist

$$1 = \sup_{r>0} \mu(B_r(0)) \leq \sup_{r>0} \mu[g_r] = \sup_{r>0} \liminf_{n\to\infty} \mathbf{P}_n[g_r] \leq 1.$$

'$\Leftarrow$': Sei $(\mathbf{P}_n)_{n=1,2,..}$ straff. Dann folgt mit Lemma 10.12

$$1 \geq \mu(\mathbb{R}) = \sup_{r>0} \mu(B_r(0)) = \sup_{r>0} \mu[g_r] = \sup_{r>0} \liminf_{n\to\infty} \mathbf{P}_n[g_r] = 1.$$

Es bleibt, die schwache Konvergenz zu zeigen. Angenommen, $(\mathbf{P}_n)_{n=1,2,\ldots}$ ist straff und $f \in \mathcal{C}_b(\mathbb{R})$. Dann gilt

$$\begin{aligned}\limsup_{n\to\infty} |\mathbf{P}_n[f] - \mu[f]| &\leq \inf_{r>0} \limsup_{n\to\infty} \big(|\mathbf{P}_n[f - fg_r]| + |\mathbf{P}_n[fg_r]| - \mu[fg_r]| + |\mu[f - fg_r]|\big) \\ &\leq ||f|| \inf_{r>0} \limsup_{n\to\infty} \mathbf{P}_n(B_r(0)^c) + \inf_{r>0} \mu[B_r(0)^c] = 0,\end{aligned}$$

woraus $\mathbf{P}_n \xRightarrow{n\to\infty} \mu$ folgt. □

Korollar 10.18 (Schwache Konvergenz und Straffheit). *Seien* $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$. *Falls* $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$, *so ist* $(\mathbf{P}_n)_{n\in\mathbb{N}}$ *straff.*

Beweis. Da aus der schwachen Konvergenz von $\mathbf{P}_1, \mathbf{P}_2, \dots$ gegen $\mathbf{P}$, die vage Konvergenz folgt, gelten für $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots$ die Voraussetzungen von Lemma 10.17 und $\mathbf{P}(\mathbb{R}) = 1$. Also ist $(\mathbf{P}_n)_{n\in\mathbb{N}}$ straff. □

Um die schwache Konvergenz von Wahrscheinlichkeitsmaßen festzustellen, ist Theorem 10.6 hilfreich. Wir wenden uns nun der Frage zu, ob eine Folge von Wahrscheinlichkeitsmaßen überhaupt einen Häufungspunkt haben kann. Das bedeutet, dass es eine Teilfolge gibt, die schwach gegen ein Wahrscheinlichkeitsmaß konvergiert.

Theorem 10.19 (Satz von Prohorov). *Sei (E, r) vollständig und separabel und $(\mathbf{P}_i)_{i\in I}$ eine Familie in $\mathcal{P}(E)$. Dann sind äquivalent:*

1. *Die Familie $(\mathbf{P}_i)_{i\in I}$ ist relativ kompakt bezüglich der Topologie der schwachen Konvergenz, d.h. jede Folge in $(\mathbf{P}_i)_{i\in I}$ hat eine schwach konvergente Teilfolge.*
2. *Zu jedem $\varepsilon > 0$ gibt es ein $N \in \mathbb{N}$ und $x_1, \dots, x_N \in E$, so dass*
$$\inf_{i\in I} \mathbf{P}_i\Big(\bigcup_{k=1}^N B_\varepsilon(x_k)\Big) \geq 1 - \varepsilon.$$
3. *Die Familie $(\mathbf{P}_i)_{i\in I}$ ist straff.*

Beweis. Sei $x_1, x_2, \dots$ eine dichte Teilfolge in E.
1. $\Rightarrow$ 2.: Angenommen, 2. ist nicht wahr. Dann gibt es ein $\varepsilon > 0$ und für jedes $N = 1, 2, \dots$ ein $\mathbf{P}_{i_N}$ mit $\mathbf{P}_{i_N}\Big(\bigcup_{k=1}^N B_\varepsilon(x_k)\Big) \leq 1 - \varepsilon$. Wegen der Relativkompaktheit gäbe es dann eine schwach gegen ein $\mathbf{P} \in \mathcal{P}(E)$ konvergente Teilfolge $(\mathbf{P}_{i_M})_{M=1,2,\dots}$. Damit gilt wegen Theorem 10.6 ($(i) \Rightarrow (iii)$), dass

$$1 = \mathbf{P}(E) = \sup_{N\in\mathbb{N}} \mathbf{P}\Big(\bigcup_{k=1}^N B_\varepsilon(x_i)\Big) \leq \sup_{N\in\mathbb{N}} \liminf_{M\to\infty} \mathbf{P}_{i_M}\Big(\bigcup_{k=1}^N B_\varepsilon(x_i)\Big) \leq 1 - \varepsilon,$$

also ein Widerspruch.
2. $\Rightarrow$ 3.: Sei $\varepsilon > 0$. Für $j = 1, 2, \dots$ wählen wir $x_{j1}, \dots, x_{jN_j}$, so dass

$$\inf_{i\in I} \mathbf{P}_i\Big(\bigcup_{k=1}^{N_j} B_{\varepsilon 2^{-j}}(x_{jk})\Big) > 1 - \varepsilon 2^{-j}.$$

Weiter setzen wir

$$K := \bigcap_{j=1}^\infty \bigcup_{k=1}^{N_j} B_{\varepsilon 2^{-j}}(x_{jk}).$$

Dann ist $K \subseteq E$ nach Konstruktion total beschränkt, nach Proposition 1.9 also relativ kompakt, $\overline{K}$ also kompakt. Außerdem gilt

$$\sup_{i\in I} \mathbf{P}_i(\overline{K}^c) \leq \sup_{i\in I} \sum_{j=1}^\infty \mathbf{P}_i\Big(\bigcap_{k=1}^{N_j} (B_{\varepsilon 2^{-j}}(x_{jk}))^c\Big) \leq \varepsilon.$$

Damit ist die Familie $(\mathbf{P}_i)_{i\in I}$ straff.
3. $\Rightarrow$ 1.: Sei $\mathbf{P}_1, \mathbf{P}_2, \dots$ eine Folge in der Familie $(\mathbf{P}_i)_{i\in I}$. Ziel ist es, eine konvergente Teilfolge zu bestimmen. Hierzu wählen wir kompakte Mengen $K_1 \subseteq K_2 \subseteq \cdots \subseteq E$ aus mit $\inf_{n=1,2,\dots} \mathbf{P}_n(K_j) \geq 1 - 1/j$. Weiter wählen wir das System kompakter Mengen

$$\mathcal{K} := \Big\{\bigcup_{k=1}^N K_{j_k} \cap \overline{B_{\varepsilon_k}(x_k)} : N, j_k \in \mathbb{N}, \varepsilon_k \in \mathbb{Q}^+\Big\}.$$

Damit ist $\mathcal{K}$ ein Halbring und auch ein schnittstabiler Erzeuger von $\sigma(E)$. Da $\mathcal{K}$ abzählbar ist, können wir eine Teilfolge $\mathbf{P}_{n_1}, \mathbf{P}_{n_2}, \ldots$ von $\mathbf{P}_1, \mathbf{P}_2, \ldots$ finden, so dass $\mathbf{P}_{n_k}(A)$ für alle $A \in \mathcal{K}$ konvergiert. Wir setzen für $A \in \mathcal{K}$

$$\mu(A) = \lim_{k\to\infty} \mathbf{P}_{n_k}(A).$$

Damit ist nach Theorem 3.9 und Theorem 3.15 der Inhalt μ auf $\mathcal{K}$ eindeutig zu einem Maß auf $\sigma(\mathcal{K}) = \mathcal{B}(E)$ fortsetzbar. Das so definierte μ ist ein Wahrscheinlichkeitsmaß, weil ja

$$1 \geq \mu(E) = \sup_{j=1,2,\ldots} \mu(K_j) = \sup_{j=1,2,\ldots} \lim_{k\to\infty} \mathbf{P}_{n_k}(K_j) \geq \sup_{j=1,2,\ldots} 1 - 1/j = 1.$$

Außerdem gilt für jede offene Menge A

$$\mu(A) = \sup_{A\supseteq H\in\mathcal{K}} \mu(H) = \sup_{A\supseteq H\in\mathcal{K}} \lim_{k\to\infty} \mathbf{P}_{n_k}(H) \leq \liminf_{k\to\infty} \mathbf{P}_{n_k}(A).$$

Nun folgt $\mathbf{P}_{n_k} \xRightarrow{k\to\infty} \mu$ aus Theorem 10.6 ($(iii) \Rightarrow (i)$). $\square$

10.3 Separierende Funktionenklassen

Nun werden wir das Konzept der separierenden Funktionenklasse vorstellen. Dabei wird klar werden, welche Nützlichkeit charakteristische Funktionen und Laplace-Transformierte von Verteilungen (siehe Definition 7.11) besitzen. Diese basieren nämlich auf zwei bestimmten Funktionenklassen, die separierend sind.

Definition 10.20 (Punktetrennende und separierende Funktionenklassen).

1. *Eine Funktionenklasse $\mathcal{M} \subseteq \mathcal{C}(E)$ heißt punktetrennend in E, falls es für alle $x, y \in E$ mit $x \neq y$ ein $f \in \mathcal{M}$ gibt mit $f(x) \neq f(y)$.*
2. *Eine Funktionenklasse $\mathcal{M} \subseteq \mathcal{C}(E)$ heißt separierend in $\mathcal{P}(E)$, falls aus $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$ und*
$$\mathbf{P}[f] = \mathbf{Q}[f] \text{ für alle } f \in \mathcal{M}$$
folgt, dass $\mathbf{P} = \mathbf{Q}$.

Beispiel 10.21. 1. Die Funktionenklasse $\mathcal{M} := \mathcal{C}_b(E)$ ist sowohl punktetrennend als auch separierend. Ist nämlich $x \neq y$, so ist $z \mapsto r(x, z) \wedge 1$ eine beschränkte, stetige Funktion, die x und y trennt. Ist außerdem $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$ und $\mathbf{P} \neq \mathbf{Q}$, so gibt es einen offenen Ball A mit $\mathbf{P}(A) \neq \mathbf{Q}(A)$. Sei $f_1, f_2, \ldots$ eine Folge in $\mathcal{C}_b(E)$ mit $f_n \uparrow 1_A$. Wäre $\mathbf{P}[f_n] = \mathbf{Q}[f_n]$ für alle $n = 1, 2, \ldots$, so wäre auch
$$\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}[f_n] = \lim_{n\to\infty} \mathbf{Q}[f_n] = \mathbf{Q}(A)$$
im Widerspruch zur Voraussetzung.

2. Die Funktionenklasse $\{x \mapsto cx : c \in \mathbb{R}\}$ aller linearen Funktionen ist zwar punktetrennend, aber nicht separierend.

Das nächste Resultat benötigt das Stone-Weierstrass-Theorem, das wir zunächst wiederholen.

Definition 10.22 (Algebra). *Ein Mengensystem $\mathcal{M} \subseteq \mathcal{C}(E)$ heißt Algebra, falls $1 \in \mathcal{M}$, sowie mit f, g und $\alpha, \beta \in \mathbb{R}$ auch $\alpha f + \beta g$, sowie fg in $\mathcal{M}$ enthalten sind.*

Theorem 10.23 (Stone-Weierstrass). *Sei (E, r) kompakt und $\mathcal{M} \subseteq C_b(E)$ eine punktetrennende Algebra. Dann ist $\mathcal{M}$ dicht in $\mathcal{C}_b(E)$ bzgl. der Supremumsnorm.*

Beweis. Siehe Analysis. □

Theorem 10.24 (Punktetrennende und separierende Algebren).
Sei (E, r) vollständig und separabel. Ist $\mathcal{M} \subseteq \mathcal{C}_b(E)$ punktetrennend und so, dass mit $f, g \in \mathcal{M}$ auch $fg \in \mathcal{M}$. Dann ist $\mathcal{M}$ separierend.

Beweis. Sei $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$. Ohne Einschränkung ist $1 \in \mathcal{M}$, da $\mathbf{P}[1] = \mathbf{Q}[1]$ immer gilt. Also ist $\mathcal{M}$ o.E. eine Algebra. Sei $\varepsilon > 0$ und K kompakt, so dass $\mathbf{P}(K) > 1 - \varepsilon$, $\mathbf{Q}(K) > 1 - \varepsilon$. Für $g \in \mathcal{C}_b(E)$ gibt es nach dem Stone-Weierstrass Theorem 10.23 eine Folge $(g_n)_{n=1,2,\ldots}$ in $\mathcal{M}$ mit

$$\sup_{x \in K} |g_n(x) - g(x)| \xrightarrow{n \to \infty} 0. \tag{10.4}$$

Nun,

$$\begin{aligned}
\big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}]\big| \leq{} & \big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{P}[ge^{-\varepsilon g^2}; K]\big| \\
& + \big|\mathbf{P}[ge^{-\varepsilon g^2}; K] - \mathbf{P}[g_n e^{-\varepsilon g_n^2}; K]\big| \\
& + \big|\mathbf{P}[g_n e^{-\varepsilon g_n^2}; K] - \mathbf{P}[g_n e^{-\varepsilon g_n^2}]\big| \\
& + \big|\mathbf{P}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[g_n e^{-\varepsilon g_n^2}]\big| \\
& + \big|\mathbf{Q}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[g_n e^{-\varepsilon g_n^2}; K]\big| \\
& + \big|\mathbf{Q}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}; K]\big| \\
& + \big|\mathbf{Q}[ge^{-\varepsilon g^2}; K] - \mathbf{Q}[ge^{-\varepsilon g^2}]\big|
\end{aligned}$$

Den ersten Term beschränken wir durch

$$\big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{P}[ge^{-\varepsilon g^2}; K]\big| \leq \frac{C}{\sqrt{\varepsilon}} \mathbf{P}(K^c) \leq C\sqrt{\varepsilon}$$

mit $C = \sup_{x \geq 0} xe^{-x^2}$; analog den dritten, fünften und letzten. Der zweite und vorletzte Term konvergieren für $n \to \infty$ gegen 0 wegen (10.4). Da $\mathcal{M}$ eine Algebra ist, kann $g_n e^{-\varepsilon g_n^2}$ durch Funktionen in $\mathcal{M}$ approximiert werden, womit auch der vierte Term für $n \to \infty$ gegen 0 konvergiert. Damit gilt

$$\big|\mathbf{P}[g] - \mathbf{Q}[g]\big| = \lim_{\varepsilon \to 0} \big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}]\big| \leq 4C \lim_{\varepsilon \to 0} \sqrt{\varepsilon} = 0.$$

Da g beliebig war und $\mathcal{C}_b(E)$ separierend ist, folgt $\mathbf{P} = \mathbf{Q}$. □

Wir kommen nun auf die charakteristische Funktion und die Laplace-Transformierte zurück. Wie bereits erwähnt ist die Nützlichkeit der charakteristischen Funktion und der Laplace-Transformierten darauf zurückzuführen, dass sie verteilungsbestimmend sind.

Proposition 10.25 (Charakteristische Funktion verteilungsbestimmend). *Ein Wahrscheinlichkeitsmaß $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ ($\mathbf{P} \in \mathcal{P}(\mathbb{R}_+^d)$) wird eindeutig durch die charakteristische Funktion $\psi_{\mathbf{P}}$ (die Laplace-Transformierte $\mathscr{L}_{\mathbf{P}}$) bestimmt.*

Beweis. Wir zeigen die Aussage nur für charakteristische Funktionen, die für Laplace-Transformierte beweist man analog. Wir stellen fest, dass die Menge $\mathcal{M} := \{x \mapsto e^{itx}; t \in \mathbb{R}^d\}$ in $\mathbb{R}^d$ Punkte trennt. Da $\mathcal{M} \subseteq \mathcal{C}_b(\mathbb{R}^d)$ und abgeschlossen unter Produktbildung ist, ist sie nach Theorem 10.24 auch separierend. Dies zeigt die Aussage. □

Korollar 10.26 (Unabhängigkeit und charakteristische Funktion).

1. *Eine Familie $(X_j)_{j\in I}$ von reellwertigen Zufallsvariablen ist genau dann unabhängig, falls für alle $J \Subset I$*
$$\mathbf{E}\Big[\prod_{j\in J} e^{it_jX_j}\Big] = \prod_{j\in J} \mathbf{E}[e^{it_jX_j}] \tag{10.5}$$
für alle $(t_j)_{j\in J} \in \mathbb{R}^J$ gilt.

2. *Eine Familie $(X_j)_{j\in I}$ von Zufallsvariablen mit Werten in $\mathbb{R}_+$ ist genau dann unabhängig, falls für alle $J \Subset I$*
$$\mathbf{E}\Big[\prod_{j\in I} e^{-t_jX_j}\Big] = \prod_{j\in J} \mathbf{E}[e^{-t_jX_j}]$$
für alle $(t_j)_{j\in J} \in \mathbb{R}^J$ gilt.

Beweis. Wir zeigen nur die erste Aussage, die zweite folgt analog. Ist $(X_j)_{j\in I}$ unabhängig, so sind nach Lemma 9.4 auch die Zufallsvariablen $(e^{it_jX_j})_{j\in I}$ für alle $(t_j)_{j\in J} \in \mathbb{R}^J$ unabhängig. Damit folgt (10.5) aus Proposition 9.5. Umgekehrt gelte (10.5). Einerseits stellt die linke Seite von (10.5) die charakteristische Funktion der Verteilung $((X_j)_{j\in J})_*\mathbf{P}$ dar. Andererseits ist die rechte Seite von (10.5) die charakteristische Funktion von $\bigotimes_{j\in J}(X_j)_*\mathbf{P}$. Da die charakteristische Funktion nach Proposition 10.25 die gemeinsame Verteilung von $(X_j)_{j\in J}$ eindeutig bestimmt, gilt also $((X_j)_{j\in J})_*\mathbf{P} = \bigotimes_{j\in J}(X_j)_*\mathbf{P}$. Damit folgt die Unabhängigkeit von $(X_j)_{j\in I}$ aus Proposition 9.2. □

10.4 Der Satz von Lévy

Wir wollen nun den Zusammenhang zwischen schwacher Konvergenz und der Konvergenz der charakteristischen Funktionen der zu Grunde liegenden Verteilungen beleuchten. Seien hierzu $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}^d)$. Wie man aus Proposition 10.27 sieht, folgt die schwache Konvergenz $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ aus der punktweisen Konvergenz der charakteristischen Funktionen, $\psi_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \psi_{\mathbf{P}}(t)$, $t \in \mathbb{R}^d$, gegeben $(\mathbf{P}_n)_{n\in\mathbb{N}}$ ist straff. Entscheidend ist nun, dass man die Straffheit der Familie $(\mathbf{P}_n)_{n\in\mathbb{N}}$ ebenfalls aus den charakteristischen Funktionen ablesen kann, wie wir in Proposition 10.32 zeigen werden. Dies führt zu der Aussage des Lévy'schen Stetigkeitssatzes (Theorem 10.33), der angibt, wann der punktweise Limes von charakteristischen Funktionen wieder charakteristische Funktion eines Wahrscheinlichkeitsmaßes ist.

Proposition 10.27 (Separierende Funktionenklasse und schwache Konvergenz). *Sei (E, r) vollständig und separabel und $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$. Dann sind äquivalent:*

1. $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$.

2. $(\mathbf{P}_n)_{n=1,2,\dots}$ *ist straff und es gibt eine separierende Familie* $\mathcal{M} \subseteq \mathcal{C}_b(E)$ *mit*

$$\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f] \text{ für alle } f \in \mathcal{M}.$$

Beweis. 1. $\Rightarrow$ 2. Nach Korollar 10.18 ist $(\mathbf{P}_n)_{n=1,2,\dots}$ straff. Der zweite Teil von 2. gilt wegen der Definition der schwachen Konvergenz.
2. $\Rightarrow$ 1. Angenommen, $(\mathbf{P}_n)_{n=1,2,\dots}$ ist straff und $\mathbf{P}_1, \mathbf{P}_2, \dots$ konvergiert nicht schwach gegen $\mathbf{P}$. Dann gibt es $\varepsilon > 0$, ein $f \in \mathcal{C}_b(E)$ und eine Teilfolge $(n_k)_{k=1,2,\dots}$, so dass

$$|\mathbf{P}_{n_k}[f] - \mathbf{P}[f]| > \varepsilon \text{ für alle } k. \tag{10.6}$$

Nach Theorem 10.19 gibt es eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$ und ein $\mathbf{Q} \in \mathcal{P}(E)$, so dass $\mathbf{P}_{n_{k_\ell}} \xRightarrow{\ell\to\infty} \mathbf{Q}$. Wegen (10.6) ist

$$|\mathbf{P}[f] - \mathbf{Q}[f]| \geq |\liminf_{\ell\to\infty} \big(\mathbf{P}[f] - \mathbf{P}_{n_{k_\ell}}[f]|\big) + \liminf_{\ell\to\infty}(\mathbf{P}_{n_{k_\ell}}[f] - \mathbf{Q}[f])| > \varepsilon,$$

insbesondere also $\mathbf{P} \neq \mathbf{Q}$. Andererseits haben wir für alle $g \in \mathcal{M}$

$$\mathbf{P}[g] = \lim_{\ell\to\infty} \mathbf{P}_{n_{k_\ell}}[g] = \mathbf{Q}[g].$$

Da $\mathcal{M}$ separierend ist, ist dies ein Widerspruch und 1. ist gezeigt. $\square$

Sei $\mathbf{P} \in \mathcal{P}(\mathbb{R})$ und $\psi_{\mathbf{P}}$ deren charakteristische Funktion. Wir zeigen zunächst eine Abschätzung, die wichtig ist, um Straffheit und die $\psi_{\mathbf{P}}$ in Verbindung zu bringen.

Lemma 10.28 (Straffheit und die charakteristische Funktion). *Sei* $\mathbf{P} \in \mathcal{P}(\mathbb{R})$. *Dann gilt für alle* $r > 0$

$$\mathbf{P}((-\infty; -r] \cup [r; \infty)) \leq \frac{r}{2} \int_{-2/r}^{2/r} (1 - \psi_{\mathbf{P}}(t))dt, \tag{10.7}$$

Beweis. Es ist $\sin(x)/x \leq 1$ für $x \leq 2$ und $\sin x \leq x/2$ für $x \geq 2$. Sei X eine Zufallsvariable mit Verteilung $\mathbf{P}$. Also gilt für jedes $c > 0$ nach Fubini

$$\begin{aligned}
\int_{-c}^{c} (1 - \psi_{\mathbf{P}}(t))dt &= \mathbf{P}\Big[\int_{-c}^{c} (1 - e^{itX})dt\Big] = \mathbf{P}\Big[2c - \frac{1}{iX} e^{itX}\Big|_{t=-c}^{c}\Big] \\
&= 2c\mathbf{P}\Big[1 - \frac{\sin(cX)}{cX}\Big] \\
&\geq 2c\mathbf{P}\Big[1 - \frac{\sin(cX)}{cX}; |cX| \geq 2\Big] \\
&\geq c \cdot \mathbf{P}(|cX| \geq 2) = c\mathbf{P}((-\infty; -\tfrac{2}{c}] \cup [\tfrac{2}{c}; \infty)),
\end{aligned}$$

und die Behauptung folgt mit $c = 2/r$. $\square$

Definition 10.29 (Gleichgradige Stetigkeit). *Wir wiederholen eine Definition aus der Analysis. Eine Menge* $\mathcal{M} \subseteq \mathcal{C}(\mathbb{R}^d)$ *heißt in* $x \in \mathbb{R}^d$ *gleichgradig stetig, falls*

$$\sup_{f\in\mathcal{M}} |f(y) - f(x)| \xrightarrow{y\to x} 0.$$

Bemerkung 10.30 (Äquivalente Bedingung für Folgen). Falls $\mathcal{M} = \{f_1, f_2, \dots\}$, so ist die Bedingung

$$\limsup_{n\to\infty} |f_n(y) - f_n(x)| \xrightarrow{y\to x} 0$$

äquivalent.

Lemma 10.31 (Gleichgradige Stetigkeit und Konvergenz). *Seien $f_1, f_2, \dots \in \mathcal{C}(\mathbb{R}^d)$, so dass $f_n \xrightarrow{n\to\infty} f$ punktweise für eine Funktion $f : \mathbb{R}^d \to \mathbb{R}$ gilt. Genau dann ist f in 0 stetig, wenn $(f_n)_{n=1,2,\dots}$ in 0 gleichgradig stetig ist.*

Beweis. Ist $(f_n)_{n=1,2,\dots}$ gleichgradig stetig in 0, so folgt

$$|f(t) - f(0)| = |\lim_{n\to\infty}(f_n(t) - f_n(0))| \le \limsup_{n\to\infty} |f_n(t) - f_n(0)| \xrightarrow{t\to 0} 0.$$

Ist andersherum f stetig in 0, so gilt

$$\limsup_{n\to\infty} |f_n(t) - f_n(0)| \le \limsup_{n\to\infty} |f_n(t) - f(t)| + |f(t) - f(0)| + |f(0) - f_n(0) = |f(t) - f(0)| \xrightarrow{t\to 0} 0.$$

□

Proposition 10.32 (Straffheit und gleichgradige Stetigkeit). *Sei $(\mathbf{P}_i)_{i\in I}$ eine Familie in $\mathcal{P}(\mathbb{R}^d)$. Ist $(\psi_{\mathbf{P}_i})_{i\in I}$ gleichgradig stetig in 0, so ist $(\mathbf{P}_i)_{i\in I}$ straff.*

Beweis. Es genügt zu zeigen, dass $((\pi_k)_*\mathbf{P}_i)_{i\in I}$ für alle Projektionen $\pi_1, \dots, \pi_d$ straff ist. Es gilt offenbar $\psi_{(\pi_k)_*\mathbf{P}_i}(t) = \psi_{\mathbf{P}_i}(te_k)$, falls e_k der k-te Einheitsvektor ist. Damit genügt es, die Behauptung im Fall $d = 1$ zu zeigen. Da $\psi_{\mathbf{P}_i}(0) = 1$ für alle $i \in I$ gilt, folgern wir aus der gleichgradigen Stetigkeit, dass

$$\sup_{i\in I} |1 - \psi_{\mathbf{P}_i}(t)| \xrightarrow{t\to 0} 0,$$

also, siehe Bemerkung 10.15,

$$\begin{aligned}
\sup_{r>0} \inf_{i\in I} \mathbf{P}_i([-r;r]) &\ge 1 - \inf_{r>0} \sup_{i\in I} \frac{r}{2} \int_{-2/r}^{2/r} (1 - \psi_{\mathbf{P}_i}(t))dt \\
&\ge 1 - \inf_{r>0} \frac{r}{2} \int_{-2/r}^{2/r} \sup_{i\in I} |1 - \psi_{\mathbf{P}_i}(t)|dt \\
&\ge 1 - 2 \inf_{r>0} \sup_{t\in[0;2/r]} \sup_{i\in I} |1 - \psi_{\mathbf{P}_i}(t)| = 1.
\end{aligned}$$

Damit ist die Behauptung gezeigt. □

Theorem 10.33 (Lévy'scher Stetigkeitssatz). *Seien $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}^d)$ und $\psi : \mathbb{R}^d \to \mathbb{C}$, so dass $\psi_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \psi(t)$ für alle $t \in \mathbb{R}^d$. Falls ψ stetig in 0 ist, so ist $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ für ein $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ mit $\psi_{\mathbf{P}} = \psi$.*

Beweis. Da $\psi_{\mathbf{P}_n}$ punktweise gegen die in 0 stetige Funktion ψ konvergieren, folgt mit Lemma 10.31, dass $(\psi_{\mathbf{P}_n})_{n=1,2,\dots}$ in 0 gleichgradig stetig ist. Mit Proposition 10.32 folgt, dass $(\mathbf{P}_n)_{n=1,2,\dots}$ straff ist. Sei $(n_k)_{k=1,2,\dots}$ eine Teilfolge und $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$, so dass $\mathbf{P}_{n_k} \xRightarrow{k\to\infty} \mathbf{P}$. Da $x \mapsto e^{itx}$ eine stetige, beschränkte Funktion ist, folgt $\psi_{\mathbf{P}_{n_k}}(t) \xrightarrow{k\to\infty} \psi_{\mathbf{P}}(t)$ für alle $t \in \mathbb{R}^d$. Andererseits ist nach Voraussetzung auch $\psi_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \psi(t)$, woraus $\psi_{\mathbf{P}} = \psi$ folgt. Damit ist ψ als charakteristische Funktion von $\mathbf{P}$ identifiziert, und da diese $\mathbf{P}$ eindeutig bestimmt, gilt damit $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$. □

Beispiel 10.34 (Satz von deMoivre-Laplace). Seien $S_n \sim B(n,p)$. Der Satz von deMoivre-Laplace besagt, dass

$$S_n^* := \frac{S_n - np}{\sqrt{np(1-p)}} \xRightarrow{n\to\infty} N(0,1). \tag{10.8}$$

Wir wollen dies nun nochmal mit Hilfe von charakteristischen Funktionen zeigen, also $\psi_{S_n^*} \xrightarrow{n\to\infty} \psi_{N(0,1)}$ punktweise. Dazu verwenden wir Proposition 7.12.3 und schreiben mit $q := 1-p$ und $C_1, C_2, \dots \in \mathbb{C}$ mit $\limsup_{n\to\infty} |C_n| < \infty$

$$\begin{aligned}
\psi_{S_n^*}(t) &= \exp\Big(-it\sqrt{\frac{np}{q}}\Big)\cdot\psi_{B(n,p)}\Big(\frac{t}{\sqrt{npq}}\Big)\\
&= \exp\Big(-it\sqrt{\frac{np}{q}}\Big)\Big(q + p\exp\Big(\frac{it}{\sqrt{npq}}\Big)\Big)^n\\
&= \Big(q\exp\Big(-it\sqrt{\frac{p}{nq}}\Big) + p\exp\Big(it\sqrt{\frac{q}{np}}\Big)\Big)^n\\
&= \Big(1 - qit\sqrt{\frac{p}{nq}} - q\frac{t^2}{2}\frac{p}{nq} + pit\sqrt{\frac{q}{np}} - p\frac{t^2}{2}\frac{q}{np} + \frac{C_n}{n^{3/2}}\Big)^n\\
&= \Big(1 - \frac{t^2}{2}\frac{1}{n} + \frac{C_n}{n^{3/2}}\Big)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}} = \psi_{N(0,1)}(t).
\end{aligned}$$

Aus Theorem 10.33 folgt nun (10.8).

Der Lévy-sche Stetigkeitssatz lässt sich auch mit Laplace-Transformierten formulieren. Wir geben den Satz ohne Beweis an:

Theorem 10.35 (Lévy'scher Stetigkeitssatz für Laplace-Transformierte). *Seien $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}_+^d)$ und $\mathscr{L} : \mathbb{R}^d \to [0,1]$, so dass $\mathscr{L}_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \mathscr{L}(t)$ für alle $t \in \mathbb{R}^d$. Falls $\mathscr{L}$ stetig in 0 ist, so ist $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ für ein $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ mit $\mathscr{L}_{\mathbf{P}} = \mathscr{L}$.*

Beispiel 10.36 (Konvergenz der geometrischen zur Exponentialverteilung). Sei etwa $X_n \sim \mu_{\text{geo}(p_n)}$ verteilt und $n\cdot p_n \xrightarrow{n\to\infty} \lambda$. Dann ist

$$\begin{aligned}
\mathscr{L}_{X_n/n}(t) = \mathbf{P}[e^{-tX_n/n}] &= \sum_{k=1}^\infty (1-p_n)^{k-1} p_n e^{-tk/n}\\
&= p_n e^{-t/n}\frac{1}{1-(1-p_n)e^{-t/n}}\\
&= \frac{\lambda}{n(1-(1-\lambda/n)(1-t/n))} + o(1/n)\\
&\xrightarrow{n\to\infty} \frac{\lambda}{\lambda + t}.
\end{aligned}$$

Also gilt $\frac{X_n}{n} \xRightarrow{n\to\infty} Y$, wobei $Y \sim \mu_{\exp(\lambda)}$, da

$$\mathscr{L}_{\exp(\lambda)}(t) = \int_0^\infty \lambda e^{-\lambda a} e^{-ta} da = \frac{\lambda}{\lambda + t}.$$

11 Grenzwertsätze in Verteilung

Wir werden nun unsere Kenntnisse über schwache Konvergenz und charakteristische Funktionen in speziellen Situationen einsetzen. In Abschnitt 11.1 geht es um Aussagen, wann die Summe von Zufallsvariablen gegen eine Poissonverteilte Zufallsvariable konvergiert. In Abschnitt 11.2 werden wir den zentralen Grenzwertsatz von Lindeberg-Feller kennenlernen, der eine Charakterisierung für die schwache Konvergenz gegen eine Normalverteilung darstellt. In Abschnitt 11.3 geht es schließlich um Erweiterungen für den Fall von mehrdimensionalen Zufallsvariablen.

11.1 Poisson-Konvergenz

Bereits bekannt ist die Aussage, dass $B(n, p_n)$ für $n \cdot p_n \xrightarrow{n\to\infty} \lambda$ für große n schwach gegen die Poi(λ)-Verteilung konvergiert, siehe Beispiel 11.1. In diesem Abschnitt verallgemeinern wir diese Aussage, siehe Theorem 11.5.

Beispiel 11.1 (Poisson-Approximation der Binomialverteilung). Sei $p_1, p_2, \dots \in [0,1]$ so, dass $n \cdot p_n \xrightarrow{n\to\infty} \lambda$. Dann ist bereits aus der Vorlesung Stochastik bekannt, dass

$$B(n, p_n)(\{k\}) \xrightarrow{n\to\infty} Poi(\lambda)(\{k\}).$$

Anders ausgedrückt ist das eine Aussage über schwache Konvergenz:

$$B(n, p_n) \xRightarrow{n\to\infty} \mathrm{Poi}(\lambda). \tag{11.1}$$

Der Satz von Lévy gibt eine weitere Möglichkeit an, dieses Resultat zu beweisen. Wir erinnern an die charakteristischen Funktionen der Binomial- und Poisson-Verteilung aus Beispiel 7.13. Wir schreiben direkt

$$\begin{aligned}\psi_{B(n,p_n)}(t) &= \Big(1 - p_n\big(1 - e^{it}\big)\Big)^n \\ &= \Big(1 - \frac{n \cdot p_n}{n}\big(1 - e^{it}\big)\Big)^n \\ &\xrightarrow{n\to\infty} \exp\big(-\lambda(1 - e^{it})\big) = \psi_{\mathrm{Poi}(\lambda)}(t).\end{aligned}$$

Insbesondere konvergieren die charakteristischen Funktionen der Binomialverteilungen punktweise gegen eine in 0 stetige Funktion, nämlich die charakteristische Funktion der Poisson-Verteilung. Aus Theorem 10.33 folgt (11.1).

Im folgenden werden wir sehen, dass die schwache Konvergenz gegen eine Poisson-Verteilung noch allgemeiner gilt. Hierzu werden wir Erzeugendenfunktionen verwenden.

Bemerkung 11.2 (Erzeugendenfunktion). Betrachte eine Zufallsvariable X mit Werten in $\mathbb{Z}_+$ und definiere die *Erzeugendenfunktion*

$$z \mapsto \varphi_X(z) := \mathbf{P}[z^X] = \sum_{k=0}^{\infty} z^k \mathbf{P}[X = k].$$

Wir bemerken, dass diese für $z \in [0,1]$ mit der Laplace-Transformierten von X eng verwandt ist, weil ja (mit $z = e^{-t}$)

$$\mathscr{L}_X(t) = \mathbf{P}[e^{-tX}] = \mathbf{P}[z^X] = \varphi_X(z).$$

Insbesondere übertragen sich zwei Eigenschaften von Laplace-Transformierten zu Erzeugendenfunktionen.

1. *Erzeugendenfunktionen verteilungsbestimmend, siehe Proposition 10.25:* Die Verteilung von X ist eindeutig bestimmt durch $z \mapsto \varphi_X(z)$ für $z \in [0,1]$.
2. *Schwache Konvergenz äquivalent zur Konvergenz der Erzeugendenfunktionen, siehe Theorem 10.33:* Sei $X_1, X_2, \dots$ eine Folge von Zufallsvariablen mit Werten in $\mathbb{Z}_+$, so dass $\varphi_{X_n}(z) \xrightarrow{n\to\infty} \varphi(z)$ für $z \in [0,1]$ für eine Funktion φ, die in 1 stetig von unten ist. Dann ist $X_n \xRightarrow{n\to\infty} X$ für eine Zufallsvariable X mit Erzeugendenfunktion φ.

Manchmal sind Erzeugendenfunktionen praktische Werkzeuge. Durch ihre Definition sind sie Potenzreihen mit Konvergenzradius $r \geq 1$. Man weiß, dass im Inneren des Konvergenzradius Ableitung und Summe vertauschen. Ist also etwa $r > 1$, schreiben wir

$$\varphi_X'(1) = \sum_{k=0}^{\infty} k z^{k-1} \mathbf{P}(X = k)\Big|_{z=1} = \sum_{k=0}^{\infty} k \mathbf{P}(X = k) = \mathbf{P}[X].$$

Analoge Rechnungen für höhere Ableitungen sind ebenfalls möglich.

Definition 11.3 (Asymptotische Vernachlässigbarkeit). *Eine triagonale Familie von Zufallsvariablen* $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ *mit* $m_1, m_2, \dots \in \mathbb{N}$ *ist* asymptotisch vernachlässigbar *falls für* $n = 1, 2, \dots$ *die Zufallsvariablen* $X_{n1}, \dots, X_{n,m_n}$ *unabhängig sind und*

$$\sup_{j=1,\dots,m_n} \mathbf{P}(|X_{nj}| > \varepsilon) \xrightarrow{n\to\infty} 0 \tag{11.2}$$

für alle $\varepsilon > 0$. *Falls* $X_{ij} \geq 0$ *für alle* i, j, *so ist auch* $m_n = \infty$ *zugelassen.*

Bemerkung 11.4 (Äquivalente Formulierung).
1. Für eine triagonale Familie von Zufallsvariablen $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ gilt (11.2) genau dann, wenn
$$\sup_{j=1,\dots,m_n} \mathbf{E}[|X_{nj}| \wedge 1] \xrightarrow{n\to\infty} 0.$$
2. Sei $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ eine triagonale von $\mathbb{Z}_+$-wertigen Zufallsvariablen. Dann ist (11.2) genau dann, wenn
$$\inf_{z\in[0,1]} \inf_{j=1,\dots,m_n} \varphi_{X_{nj}}(z) = \inf_{j=1,\dots,m_n} \varphi_{X_{nj}}(0) = \inf_{j=1,\dots,m_n} \mathbf{P}(|X_{nj}| = 0) \xrightarrow{n\to\infty} 1. \tag{11.3}$$

Theorem 11.5 (Poisson Konvergenz). *Sei* $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ *eine Familie asymptotisch vernachlässigbarer Zufallsvariablen mit Werten in* $\mathbb{Z}_+$ *und* $X \sim Poi(\lambda)$. *Dann gilt*

$$\sum_{j=1}^{m_n} X_{nj} \xRightarrow{n\to\infty} X$$

genau dann, wenn gilt:

1. $\displaystyle\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 1) \xrightarrow{n\to\infty} 0$
2. $\displaystyle\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} = 1) \xrightarrow{n\to\infty} \lambda.$

Wir bereiten den Beweis mit einem Lemma vor.

Lemma 11.6. *Sei $(\lambda_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ eine triagonale Familie asymptotisch vernachlässigbarer, nicht-negativer Konstanten und $\lambda \in [0;\infty]$. Dann gilt*

$$\prod_{j=1}^{m_n}(1-\lambda_{nj}) \xrightarrow{n\to\infty} e^{-\lambda} \qquad \iff \qquad \sum_{j=1}^{m_n}\lambda_{nj} \xrightarrow{n\to\infty} \lambda.$$

Beweis. Zunächst sei bemerkt, dass $\log(1-x) = -x + \varepsilon(x)$ für $x > 0$ mit $\varepsilon(x)/x \xrightarrow{x\to 0} 0$. Da $\sup_{j=1,\dots,m_n}\lambda_{nj} < 1$ für große n, ist die linke Aussage äquivalent zu

$$-\lambda = \lim_{n\to\infty}\sum_{j=1}^{m_n}\log(1-\lambda_{nj}) = -\lim_{n\to\infty}\sum_{j=1}^{m_n}\lambda_{nj}\big(1-\tfrac{\varepsilon(\lambda_{nj})}{\lambda_{nj}}\big) = -\lim_{n\to\infty}\sum_{j=1}^{m_n}\lambda_{nj},$$

da

$$\sup_{j=1,\dots,m_n}\tfrac{\varepsilon(\lambda_{nj})}{\lambda_{nj}} \xrightarrow{n\to\infty} 0.$$

Daraus folgt die Aussage. □

Beweis von Theorem 11.5. Wir bezeichnen mit $\varphi_{n,j}$ die Erzeugendenfunktion von $X_{n,j}$. Nach Bemerkung 11.2.2 ist die Konvergenzaussage der schwachen Konvergenz im Theorem äquivalent zur punktweisen Konvergenz von $\prod_{j=1}^{m_n}\varphi_{nj}(z) \xrightarrow{n\to\infty} e^{-\lambda(1-z)}$, da

$$\varphi_X(z) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}z^k = e^{-\lambda(1-z)}.$$

Wegen Lemma 11.6 gilt dies genau dann, wenn

$$A_n(z) := \sum_{j=1}^{m_n}(1-\varphi_{nj}(z)) \xrightarrow{n\to\infty} \lambda(1-z), \tag{11.4}$$

da die Familie $(1-\varphi_{nj}(z))_{n=1,2,\dots,j=1,\dots,m_n}$ für jedes $z\in[0,1]$ nach (11.3) asymptotisch vernachlässigbar ist. Wir zerlegen $A(z) = A_n^1(z) + A_n^2(z)$ mit

$$\begin{aligned}
A_n^1(z) &= \sum_{k=1}^{\infty}(1-z)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}=k) = (1-z)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}>0),\\
A_n^2(z) &= \sum_{k=2}^{\infty}(z-z^k)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}=k).
\end{aligned}$$

Zunächst sei festgestellt, dass $z(1-z) \le z - z^k \le z$ für alle $k = 2,3,\dots$ Damit ist

$$z(1-z)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}>1) \le A_n^2(z) \le z\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}>1). \tag{11.5}$$

Wenden wir uns nun dem Beweis der Behauptung zu.

'$\Rightarrow$': Es gelte also nun (11.4). Für $z = 0$ bedeutet das, da $\varphi_{nj}(0) = \mathbf{P}(X_{nj} = 0)$, dass

$$\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 0) = \sum_{j=1}^{m_n} (1 - \varphi_{nj}(0)) \xrightarrow{n\to\infty} \lambda,$$

also $A_n^1(z) \xrightarrow{n\to\infty} \lambda(1-z)$ für $z \in [0,1]$. Dann muss aber auch $A_n^2(z) \xrightarrow{n\to\infty} 0$ für $z \in [0,1]$ gelten. Wegen (11.5) bedeutet dies, dass 1. gilt. Die Aussage 2. folgt daraus durch Subtraktion.
'$\Leftarrow$': Es gelte also 1. und 2. Klar ist, dass $A_n^2(z) \xrightarrow{n\to\infty} 0$ wegen (11.5). Dann gilt aber auch $A_n^1(z) \xrightarrow{n\to\infty} (1-z)\lambda$ wegen der Voraussetzung, d.h. (11.4) ist gezeigt. □

Beispiel 11.7 (Konvergenz geometrischer Verteilungen gegen Poisson). Sei $X_{nj}, j = 1,\dots,n, n = 1,2,\dots$ geometrisch verteilt mit Parameter p_n (d.h. $\mathbf{P}(X_{nj} = k) = (1-p_n)^{k-1}p_n$, siehe Beispiel 3.2.4) und $Y_{nj} = X_{nj} - 1$. (Somit gibt Y_{nj} die Anzahl der *Misserfolge* vor dem ersten *Erfolg* an.) Wir setzen $Y_n := \sum_{j=1}^n Y_{nj}$, was so verteilt ist wie die Anzahl der Misserfolge vor dem n-ten Erfolg. Falls $Y \sim \text{Poi}(\lambda)$ und $(1-p_n)\cdot n \xrightarrow{n\to\infty} \lambda$, so ist $Y_n \xRightarrow{n\to\infty} Y$. Denn es gilt

$$\begin{aligned}\sum_{j=1}^{n} \mathbf{P}(Y_{nj} = 1) &= n(1-p_n)p_n \xrightarrow{n\to\infty} \lambda,\\ \sum_{j=1}^{n} \mathbf{P}(Y_{nj} > 1) &= n(1-p_n)^2 \xrightarrow{n\to\infty} 0\end{aligned}$$

und Theorem 11.5 liefert das Ergebnis.

11.2 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz, Theorem 11.8, verallgemeinert den Satz von deMoivre Laplace. Die Verallgemeinerung besteht darin, dass beliebige Summen unabhängiger (nicht notwendig identisch verteilter) Zufallsgrößen schwach gegen eine normalverteilte Zufallsgröße konvergieren, falls sie die *Lindeberg-Bedingung* (siehe 2. in Theorem 11.8) erfüllen.

Theorem 11.8 (Zentraler Grenzwertsatz von Lindeberg-Feller). *Sei $(X_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ eine Familie von Zufallsvariablen, so dass für $n = 1,2,\dots$ die Zufallsvariablen $X_{n1},\dots,X_{nm_n}$ unabhängig sind. Sei außerdem*

$$\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}] \xrightarrow{n\to\infty} \mu, \qquad \sum_{j=1}^{m_n} \mathbf{V}[X_{nj}] \xrightarrow{n\to\infty} \sigma^2$$

und $X \sim N(\mu,\sigma^2)$. Dann sind folgende Aussagen äquivalent:

1. $\displaystyle\sum_{j=1}^{m_n} X_{nj} \xRightarrow{n\to\infty} X$ *und* $\displaystyle\sup_{j=1,\dots,m_n} \mathbf{V}[X_{nj}] \xrightarrow{n\to\infty} 0,$

2. $\displaystyle\sum_{j=1}^{m_n} \mathbf{E}[(X_{nj} - \mathbf{E}[X_{nj}])^2; |X_{nj} - \mathbf{E}[X_{nj}]| > \varepsilon] \xrightarrow{n\to\infty} 0$ *für alle* $\varepsilon > 0$.

Bevor wir den zentralen Grenzwertsatz beweisen, verweisen wir auf den Spezialfall von identisch verteilten Zufallsvariablen, der bereits in der Vorlesung Stochastik behandelt wurde.

Korollar 11.9 (Zentraler Grenzwertsatz für identische verteilte Zufallsvariable). *Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit $\mathbf{E}[X_1] = \mu, \mathbf{V}[X_1] = \sigma^2 > 0$. Sei $S_n := \sum_{k=1}^n X_k$ und $X \sim N(0,1)$. Dann gilt*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xRightarrow{n\to\infty} X.$$

Beweis. Sei $m_n = n$ und $X_{nj} = \frac{X_n - \mu}{\sqrt{n\sigma^2}}$. Dann erfüllt die Familie $(X_{nj})_{n=1,2,\ldots,j=1,\ldots,n}$ die Voraussetzungen von Theorem 11.8 mit $\mu = 0, \sigma^2 = 1$. Außerdem gilt

$$\sum_{j=1}^n \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] = \frac{1}{\sigma^2}\mathbf{E}[(X_1 - \mu)^2; |X_1 - \mu| > \varepsilon\sqrt{n\sigma^2}] \xrightarrow{n\to\infty} 0$$

wegen majorisierter Konvergenz. □

Oftmals ist die Lindeberg-Bedingung nicht einfach nachzuprüfen. Einfacher ist oft die stärkere Lyapunoff-Bedingung.

Bemerkung 11.10 (Lyapunoff-Bedingung). Die Familie $(X_{nj})_{n=1,2,\ldots,j=1,\ldots,m_n}$ aus Theorem 11.8 genügt der Lyapunoff-Bedingung, falls für ein $\delta > 0$

$$\sum_{j=1}^{m_n} \mathbf{E}\big[|X_{nj} - \mathbf{E}[X_{nj}]|^{2+\delta}\big] \xrightarrow{n\to\infty} 0.$$

Unter den Voraussetzungen von Theorem 11.8 impliziert die Lyapunoff-Bedingung die Lindeberg-Bedingung. Um dies zu sehen, sei ohne Einschränkung $\mathbf{E}[X_{nj}] = 0$. Es gilt für alle $\varepsilon > 0$

$$x^2 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta} 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta}.$$

Gilt nun die Lyapunoff-Bedingung, so folgt die Lindeberg-Bedingung aus

$$\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \leq \frac{1}{\varepsilon^\delta} \sum_{j=1}^{m_n} \mathbf{E}[|X_{nj}|^{2+\delta}] \xrightarrow{n\to\infty} 0.$$

Der Beweis von Theorem 11.8 basiert auf der geschickten Verwendung der charakteristischen Funktionen der Zufallsvariable X_{nj} und Taylor-Approximationen. Wir bereiten den Beweis des Theorems mit zwei Lemmata vor.

Lemma 11.11 (Abschätzung). *Für komplexe Zahlen $z_1, \ldots, z_n, z'_1, \ldots, z'_n$ mit $|z_i| \leq 1, |z'_i| \leq 1$ für $i = 1, \ldots, n$ gilt*

$$\Big|\prod_{k=1}^n z_k - \prod_{k=1}^n z'_k\Big| \leq \sum_{k=1}^n |z_k - z'_k|. \tag{11.6}$$

Beweis. Für $n = 1$ ist die Gleichung offensichtlich richtig. Gilt (11.6) für ein n, so ist

$$\Big|\prod_{k=1}^{n+1} z_k - \prod_{k=1}^{n+1} z_k'\Big| \le \Big|z_{n+1}\Big(\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} z_k'\Big)\Big| + \Big|(z_{n+1} - z_{n+1}')\prod_{k=1}^{n} z_k'\Big|$$
$$\le \sum_{k=1}^{n} |z_k - z_k'| + |z_{n+1} - z_{n+1}'|.$$

Daraus folgt die Behauptung. □

Lemma 11.12 (Taylor-Approximation der Exponentialfunktion). *Sei $t \in \mathbb{C}$ und $n \in \mathbb{Z}_+$. Dann gilt*

$$\Big|e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!}\Big| \le \frac{2|t|^n}{n!} \wedge \frac{|t|^{n+1}}{(n+1)!}. \tag{11.7}$$

Beweis. Bezeichne $h_n(t)$ die Differenz auf der linken Seite. Für $n = 0$ folgt (11.7) aus

$$|h_0(t)| = \Big|\int_0^t e^{is}ds\Big| \le \int_0^t |e^{is}|ds = |t|$$

und

$$|h_0(t)| \le |e^{it}| + 1 = 2.$$

Allgemein gilt für $t \in \mathbb{R}, n \in \mathbb{N}$

$$\Big|\int_0^t h_n(s)ds\Big| = \Big| - i(e^{it} - 1) + i\sum_{k=0}^{n} \frac{(it)^{k+1}}{(k+1)!}\Big| = \Big|ie^{it} - i\sum_{k=0}^{n+1} \frac{(it)^k}{k!}\Big| = |h_{n+1}(t)|,$$

woraus (11.7) mittels Induktion folgt. □

Bemerkung 11.13 (Notation). Im folgenden Beweis werden wir für Funktionen a und b genau dann $a \lesssim b$ schreiben, falls es eine Konstante C gibt mit $a \le Cb$.

Beweis von Theorem 11.8. O.E. sei $\mathbf{E}[X_{nj}] = \mu = 0$ und $\sigma^2 = 1$; ansonsten ersetzen wir X_{nj} durch $\frac{X_{nj} - \mathbf{E}[X_{nj}]}{\sqrt{\sigma^2}}$. Sei $\sigma_{nj}^2 := \mathbf{V}[X_{nj}]$ sowie $\sigma_n^2 := \sum_{j=1}^{m_n} \sigma_{nj}^2 \xrightarrow{n\to\infty} 1$. Bezeichne außerdem ψ_{nj} die charakteristische Funktion von X_{nj}.
2. ⇒ 1. Da für jedes $\varepsilon > 0$

$$\sup_{j=1,\dots,m_n} \sigma_{nj}^2 \le \varepsilon^2 + \sup_{j=1,\dots,m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \le \varepsilon^2 + \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \xrightarrow{n\to\infty} \varepsilon^2, \tag{11.8}$$

ist der zweite Teil von 1. bereits gezeigt.

Seien $(Z_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ unabhängige Zufallsvariablen mit $Z_{nj} \sim N(0, \sigma_{nj}^2)$. Damit ist $Z_n := \sum_{j=1}^{m_n} Z_{nj} \sim N(0, \sigma_n^2)$. Insbesondere gilt also $Z_n \xRightarrow{n\to\infty} X$, was man etwa direkt aus der Form der charakteristischen Funktionen der Normalverteilung, Beispiel 7.13.3 abliest. Sei $\widetilde{\psi}_{nj}$ die charakteristische Funktion von Z_{nj}. Dann genügt es zu zeigen, siehe Theorem 10.33, dass

$$\prod_{j=1}^{n_j} \psi_{nj}(t) - \prod_{j=1}^{m_n} \widetilde{\psi}_{nj}(t) \xrightarrow{n\to\infty} 0 \tag{11.9}$$

für alle t. Mittels Lemma 11.11 und Lemma 11.12 schreiben wir

$$\begin{aligned}\Big|\prod_{j=1}^{m_n}\psi_{nj}(t)-\prod_{j=1}^{m_n}\widetilde{\psi}_{nj}(t)\Big| &\le \sum_{j=1}^{m_n}|\psi_{nj}(t)-\widetilde{\psi}_{nj}(t)|\\ &\le \sum_{j=1}^{m_n}|\psi_{nj}(t)-1+\tfrac{1}{2}t^2\sigma_{nj}^2|+\sum_{j=1}^{m_n}|\widetilde{\psi}_{nj}(t)-1+\tfrac{1}{2}t^2\sigma_{nj}^2|\\ &\lesssim 2\sum_{j=1}^{m_n}\mathbf{E}[X_{nj}^2(1\wedge|X_{nj}|)+\sum_{j=1}^{m_n}|e^{-\frac{1}{2}\sigma_{nj}^2t^2}-1+\tfrac{1}{2}t^2\sigma_{nj}^2|.\end{aligned}$$

Weiter ist

$$\sum_{j=1}^{m_n}\mathbf{E}[X_{nj}^2(1\wedge|X_{nj}|)]\le\varepsilon\sum_{j=1}^{m_n}\sigma_{nj}^2+\sum_{j=1}^{m_n}\mathbf{E}[X_{nj}^2;|X_{nj}|>\varepsilon]\xrightarrow{n\to\infty}\varepsilon$$

und

$$\sum_{j=1}^{m_n}|e^{-\frac{1}{2}\sigma_{nj}^2t^2}-1+\tfrac{1}{2}t^2\sigma_{nj}^2|\lesssim\sum_{j=1}^{m_n}\sigma_{nj}^4\le\sigma_n^2\sup_{j=1,\dots,m_n}\sigma_{nj}^2\xrightarrow{n\to\infty}0$$

wegen (11.8). Damit ist (11.9) bereits bewiesen.
1. $\Rightarrow$ 2. Nach dem zweiten Teil von 1. ist für jedes $\varepsilon>0$ mit der Chebyshev-Ungleichung

$$\sup_{j=1,\dots,m_n}\mathbf{P}[|X_{nj}|>\varepsilon]\le\sup_{j=1,\dots,m_n}\frac{\sigma_{nj}^2}{\varepsilon^2}\xrightarrow{n\to\infty}0. \tag{11.10}$$

Mit Lemma 11.12 gilt

$$\sup_{j=1,\dots,m_n}|\psi_{nj}(t)-1|\le\sup_{j=1,\dots,m_n}\mathbf{E}[2\wedge|t\cdot X_{nj}|]\le 2\sup_{j=1,\dots,m_n}\mathbf{P}[|X_{nj}|>\varepsilon]+\varepsilon|t|\xrightarrow{n\to\infty}\varepsilon|t|.$$

Insbesondere ist $\sum_{j=1}^{m_n}\log\psi_{nj}(t)$ für jedes t definiert, falls n groß genug ist. Aus der Gültigkeit von 1. folgt

$$\sum_{j=1}^{m_n}\log\psi_{nj}(t)\xrightarrow{n\to\infty}-\frac{t^2}{2}. \tag{11.11}$$

Außerdem gilt wegen $\psi'_{nj}(0)=i\mathbf{E}[X_{nj}]=0,\psi''_{nj}(0)=-\mathbf{V}[X_{nj}]=-\sigma_{nj}^2$ mit Hilfe einer Taylorentwicklung von ψ_{nj} um 0

$$|\psi_{nj}(t)-1|\lesssim\sigma_{nj}^2|t|^2$$

und

$$\begin{aligned}\Big|\sum_{j=1}^{m_n}\log\psi_{nj}(t)-\sum_{j=1}^{m_n}(\psi_{nj}(t)-1)\Big| &\lesssim\sum_{j=1}^{m_n}|\psi_{nj}(t)-1|^2\\ &\lesssim\sum_{j=1}^{m_n}(\sigma_{nj}^2)^2|t|^4\lesssim|t|^4\sup_{j=1,\dots,m_n}\sigma_{nj}^2\xrightarrow{n\to\infty}0.\end{aligned} \tag{11.12}$$

Da aus der Konvergenz einer imaginären Reihe die Konvergenz ihres Realteils folgt, folgern wir aus (11.11) und (11.12) wegen $\mathrm{Re}(\psi_{nj}(t)) = \mathbf{E}[\cos(tX_{nj})]$

$$\sum_{j=1}^{m_n} \mathbf{E}[\cos(tX_{nj}) - 1] \xrightarrow{n\to\infty} -\frac{t^2}{2}$$

Für $\varepsilon > 0$ ist nun wegen $0 \leq 1 - \cos(\theta) \leq \frac{\theta^2}{2}$

$$\begin{aligned} 0 \leq \limsup_{n\to\infty} \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] &= \limsup_{n\to\infty}\ 1 - \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| \leq \varepsilon] \\ &\leq \limsup_{n\to\infty}\ 1 - \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{E}[1 - \cos(tX_{nj}); |X_{nj}| \leq \varepsilon] \\ &= \limsup_{n\to\infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{E}[1 - \cos(tX_{nj}); |X_{nj}| > \varepsilon] \qquad (11.13) \\ &\leq \limsup_{n\to\infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{P}[|X_{nj}| > \varepsilon] \\ &\leq \frac{2}{\varepsilon^2 t^2} \limsup_{n\to\infty} \sum_{j=1}^{m_n} \sigma_{nj} = \frac{2}{\varepsilon^2 t^2}. \end{aligned}$$

Da $t, \varepsilon > 0$ willkürlich waren, ist 2. gezeigt, wenn man in der letzten Ungleichungskette $t\infty$ betrachtet. □

11.3 Mehrdimensionale Grenzwertsätze

Bisher haben wir schwache Grenzwertsätze nur für den Fall $\mathbb{R}$–wertiger Zufallsvariablen betrachtet. Wir verallgemeinern dies nun zu $\mathbb{R}^d$–wertigen Zufallsgrößen. Insbesondere geben wir eine Variante des mehrdimensionalen zentralen Grenzwertsatzes an.

Definition 11.14 (Mehrdimensionale Normalverteilung). *Seien $\mu \in \mathbb{R}^d$ und $C \in \mathbb{R}^{d\times d}$ eine strikt positiv definite symmetrische Matrix.*[16,17] *Die* d-dimensionale Normalverteilung *mit Erwartungswert μ und Covarianzmatrix C ist das Wahrscheinlichkeitsmaß $N_{\mu,C}$ auf $\mathbb{R}^d$ mit Dichte*

$$f_{\mu,C}(x) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\Big(-\tfrac{1}{2}(x-\mu)C^{-1}(x-\mu)^\top \Big).$$

Proposition 11.15 (Eigenschaften der mehrdimensionalen Normalverteilung). *Seien $\mu \in \mathbb{R}^d$, $C = AA^\top \in \mathbb{R}^{d\times d}$ eine strikt positiv definite symmetrische Matrix und I die d-dimensionale Einheitsmatrix. Es sind äquivalent:*

1. $X \sim N_{\mu,C}$
2. $tX^\top \sim N_{t\mu^\top, tCt^\top}$ *für jedes $t \in \mathbb{R}^d$*

[16] Wir bezeichnen hier Zeilenvektoren mit x und Spaltenvektoren mit $x^\top$.

[17] Strikt positiv definit bedeutet $xCx^\top > 0$ für alle $x \in \mathbb{R}^d$. Aus der linearen Algebra ist bekannt, dass es für eine strikt positiv definite Matrix C immer eine invertierbare Matrix A gibt mit $C = AA^\top$.

3. $\psi_X(t) = e^{it\mu^\top} e^{-\frac{1}{2}tCt^\top}$ *für jedes* $t \in \mathbb{R}^d$.

In jedem dieser Fälle gilt

4. $X \stackrel{d}{=} AY + \mu$ *für* $Y \sim N_{0,I}$

5. $\mathbf{E}[X_i] = \mu_i$ *für* $i = 1, \ldots, d$

6. $\mathbf{COV}[X_i, X_j] = C_{ij}$ *für* $i, j = 1, \ldots, d$

Beweis. Sei zunächst $X \sim N_{\mu,C}$. Wir zeigen zunächst 4.-6. Die Eigenschaft 4. ist eine Anwendung des Transformationssatzes. Für $B \in \mathcal{B}(\mathbb{R}^d)$ und $T : y \mapsto Ay^\top + \mu^\top$ ist

$$\begin{aligned}
N_{0,I}(T^{-1}(B)) &= \frac{1}{\sqrt{(2\pi)^d}} \int_{T^{-1}(B)} e^{-\frac{1}{2}yy^\top} dy \\
&\stackrel{y=A^{-1}(x-\mu)}{=} \frac{1}{\sqrt{(2\pi)^d}} \frac{1}{\det A} \int_B \exp\big(-\tfrac{1}{2}(x-\mu)(A^\top)^{-1}A^{-1}(x-\mu)^\top\big)dx \\
&= \frac{1}{\sqrt{(2\pi)^d \det C}} \int_B \exp\big(-\tfrac{1}{2}(x-\mu)C^{-1}(x-\mu)^\top\big)dx \\
&= N_{\mu,C}(B).
\end{aligned}$$

5. folgt aus 4. mit

$$\mathbf{E}[X_i] = \mathbf{E}[\pi_i(AY + \mu)] = \pi_i\mu = \mu_i,$$

wobei π_i die Projektion auf die i-te Koordinate ist.
6. folgt ebenso aus 4. mit

$$\begin{aligned}
\mathbf{COV}[X_i, X_j] &= \mathbf{E}[(\pi_i AY^\top)(\pi_j AY^\top)] = \mathbf{E}[(A_{i\cdot}Y^\top)(A_{j\cdot}Y^\top)] = \mathbf{E}[A_{i\cdot}Y^\top Y A_{j\cdot}^\top] \\
&= A_{i\cdot}A_{j\cdot}^\top = (AA^\top)_{ij} = C_{ij}.
\end{aligned}$$

Wir kommen nun zur Äquivalenz von 1.-3.: '1. $\Rightarrow$ 2.': Da $X \stackrel{d}{=} AY^\top + \mu^\top$ wie in 4., ist $tX^\top = tAY^\top + t\mu^\top$ als Linearkombination von (eindimensionalen) Normalverteilungen wieder normalverteilt. Der Erwartungswert ist offenbar $t\mu^\top$ und die Varianz

$$\mathbf{V}[tX^\top] = \mathbf{E}[(tAY^\top)^2] = \mathbf{E}[tAY^\top YA^\top t^\top] = tAA^\top t^\top = tCt^\top.$$

'2. $\Rightarrow$ 3.': Da $tX^\top \sim N_{t\mu^\top, tCt^\top}$, folgt die Aussage aus Beispiel 7.13.3.
'3. $\Rightarrow$ 1.': Dies folgt aus Proposition 10.25. $\square$

Bemerkung 11.16 (Spezialfälle).

1. Falls C in Definition 11.14 zwar positiv, aber nicht strikt positiv definit ist (d.h. es gibt $x \in \mathbb{R}^d$, $x \neq 0$ mit $xCx = 0$), kann man $N_{\mu,C}$ nicht durch Angabe der Dichte wie in obiger Definition bestimmen. In diesem Fall definiert man $N_{\mu,C}$ durch Angabe der charakteristischen Funktion, d.h. $N_{\mu,C}$ ist die eindeutig bestimmte Verteilung auf $\mathbb{R}^d$ mit $\psi_{N_{\mu,C}}(t) = e^{it\mu}e^{-\frac{1}{2}tCt^\top}$.

2. Ist $Y \sim N_{0,I}$ und A eine orthogonale Matrix, so ist auch $X := AY \sim N_{0,I}$. Dies folgt aus Proposition 11.15, wenn man $I = AA^\top$ schreibt und 4. benutzt.

Proposition 11.17 (Cramér-Wold Device). *Sind $X, X_1, X_2, \dots$ Zufallsvariablen mit Werten in $\mathbb{R}^d$. Dann gilt $X_n \xRightarrow{n\to\infty} X$ genau dann, falls $tX_n \xRightarrow{n\to\infty} tX$ für alle $t \in \mathbb{R}^d$ (wobei $(t,x) \mapsto tx$ das Skalarprodukt im $\mathbb{R}^d$ ist).*

Beweis. '$\Rightarrow$': Sei $t \in \mathbb{R}^d$ und $f \in \mathcal{C}_b(\mathbb{R})$. Dann ist $f(t\cdot) \in \mathcal{C}_b(\mathbb{R}^d)$. Damit gilt $\mathbf{E}[f(tX_n)] \xrightarrow{n\to\infty} \mathbf{E}[f(tX)]$, d.h. $tX_n \xRightarrow{n\to\infty} tX$.
'$\Leftarrow$': Sei π_i die Projektion auf die i-te Koordinate. Da $(\pi_i X_n)_{n=1,2,\dots}$ nach Korollar 10.18 straff für alle i ist, sieht man, dass $(X_n)_{n=1,2,\dots}$ straff ist. Da $\{x \mapsto e^{itx} : t \in \mathbb{R}^d\}$ eine separierende Funktionenklasse ist, folgt die Behauptung aus $\mathbf{E}[e^{itX_n}] \xrightarrow{n\to\infty} \mathbf{E}[e^{itX}]$ für alle $t \in \mathbb{R}^d$ und Proposition 10.27. □

Theorem 11.18 (Mehrdimensionaler zentraler Grenzwertsatz). *Seien $X_1, X_2, \dots$ unabhängige, identische verteilte Zufallsvariablen mit Werten in $\mathbb{R}^d$ mit $\mathbf{E}[X_n] = \mu \in \mathbb{R}^d$ und $\mathbf{COV}[X_{n,i}, X_{n,j}] = C_{ij}$ für $i,j = 1,\dots,d$ und $S_n = \sum_{i=1}^n X_i$. ist $X \sim N_{0,C}$, so gilt*

$$\frac{S_n - n\mu}{\sqrt{n}} \xRightarrow{n\to\infty} X.$$

Beweis. Wir wenden den eindimensionalen zentralen Grenzwertsatz, Korollar 11.9, auf die unabhängigen, identisch verteilten Zufallsvariablen $tX_1, tX_2, \dots$ an. Dieser liefert

$$t\frac{S_n - n\mu}{\sqrt{n}} \xRightarrow{n\to\infty} tX.$$

Da t beliebig war, folgt die Aussage aus Proposition 11.17. □

12 Die bedingte Erwartung

Sei $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Wir schreiben $\mathcal{L}^1 := \mathcal{L}^1(\mathbf{P})$ für die Menge aller reellen Zufallsvariablen, deren Erwartungswert existiert. In diesem Kapitel verwenden wir wieder die Notation $\mathbb{E}[\cdot]$ für das Integral bezüglich des Wahrscheinlichkeitsmaßes $\mathbf{P}$, sowie $\mathcal{L}^p := \mathcal{L}^p(\mathbf{P})$.

12.1 Motivation

Definiere wie in der elementaren Stochastik für $A, G \in \mathcal{A}$ und $\mathbf{P}(G) > 0$

$$\mathbf{P}(A|G) := \frac{\mathbf{P}(A \cap G)}{\mathbf{P}(G)}$$

und analog die *bedingte Erwartung*

$$\mathbf{E}[X|G] := \frac{\mathbf{E}[X;G]}{\mathbf{P}(G)}.$$

Dann gilt $\mathbf{P}(A|G) = \mathbf{E}[1_A|G]$. Dieser Zusammenhang bedeutet, dass man bedingte Erwartungen dazu verwenden kann, bedingte Wahrscheinlichkeiten zu berechnen. Insbesondere ist der Begriff der bedingten Erwartung allgemeiner als der Begriff der bedingten Wahrscheinlichkeit.

Wir werden in diesem Kapitel die bedingte Erwartung $\mathbf{E}[X|\mathcal{G}]$ für eine Zufallsvariable X und eine σ-Algebra $\mathcal{G} \subseteq \mathcal{F}$ kennen lernen. Hierbei wird $\mathbf{E}[X|\mathcal{G}]$ eine $\mathcal{G}$-messbare Zufallsvariable sein. Als einfaches Beispiel sei $\{G_1, G_2, \dots\} \subseteq \mathcal{F}$ eine Partition von Ω mit $\mathbf{P}(G_i) > 0$ für $i = 1, 2, \dots$ und $\mathcal{G}$ die erzeugte σ-Algebra. Dann setzen wir für $X \in \mathcal{L}^1$

$$\mathbf{E}[X|\mathcal{G}](\omega) := \sum_{i=1}^{\infty} \mathbf{E}[X|G_i] \cdot 1_{G_i}(\omega). \tag{12.1}$$

Es gilt also: für $\omega \in G_i$ ist die Zufallsvariable $\mathbf{E}[X|\mathcal{G}]$ gegeben durch $\mathbf{E}[X|\mathcal{G}](\omega) = \mathbf{E}[X|G_i] = \mathbf{E}[X; G_i]/\mathbf{P}(G_i)$. Insbesondere ist sie auf G_i konstant, $i = 1, 2, \dots$ Mit anderen Worten ist $\mathbf{E}[X|\mathcal{G}]$ messbar bezüglich $\mathcal{G}$. Weiter gilt etwa für $J \subseteq \mathbb{N}$ und $A = \bigcup_{j \in J} G_j \in \mathcal{G}$

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] &= \mathbf{E}\Big[\sum_{i=1}^{\infty} \mathbf{E}[X|G_i] 1_{G_i} 1_A\Big] \\
&= \sum_{j \in J} \mathbf{E}\big[\mathbf{E}[X|G_j] 1_{G_j}\big] \\
&= \sum_{j \in J} \mathbf{E}[X|G_j] \cdot \mathbf{P}(G_j) \\
&= \mathbf{E}[X; A].
\end{aligned} \tag{12.2}$$

Insbesondere gilt mit $J = \mathbb{N}$ also $\mathbf{E}\big[\mathbf{E}[X|\mathcal{F}]\big] = \mathbf{E}[X]$. Die Definition der bedingten Erwartung (12.1) lässt sich mit Hilfe der Eigenschaft (12.2) auf beliebige σ-Algebren $\mathcal{G} \subseteq \mathcal{F}$ verallgemeinern.

Beispiel 12.1 (Binomialverteilung mit zufälliger Erfolgswahrscheinlichkeit). Sei X gleichverteilt auf $[0, 1]$, d.h. die Verteilung von X hat Dichte $1_{[0;1]}$. Gegeben $X = x$ sei $Y_1, \dots, Y_n$ eine Folge von Bernoulli-verteilten Zufallsvariablen mit Erfolgswahrscheinlichkeit x. Also ist $Y = Y_1 + \dots + Y_n$ binomialverteilt mit n und x, d.h. Y zählt die Anzahl der Erfolge in n unabhängigen Experimenten mit Erfolgswahrscheinlichkeit x. Intuitiv ist klar, dass das

$$\mathbf{P}(Y = k|X) = \binom{n}{k} X^k (1 - X)^{n-k}$$

bedeuten sollte. Dies ist allerdings bisher nicht definiert, da $\mathbf{P}(X = x) = 0$ gilt. Bemerkenswert ist jedoch, dass die rechte Seite eine $\sigma(X)$-messbare Zufallsvariable ist.

12.2 Definition und Eigenschaften

Wir definieren nun formal die bedingte Erwartung $\mathbf{E}[X|\mathcal{G}]$ für $\mathcal{G} \subseteq \mathcal{F}$. Wie oben schon erwähnt, ist dies eine $\mathcal{G}$-messbare Zufallsvariable, deren Erwartungen wie in (12.2) mit denen von X übereinstimmen.

Theorem 12.2 (Existenz und Eigenschaften der bedingten Erwartung). *Sei $\mathcal{G} \subseteq \mathcal{F}$ eine σ-Algebra. Dann gibt es einen fast sicher eindeutigen, linearen Operator $\mathbf{E}[.|\mathcal{G}] : \mathcal{L}^1 \to \mathcal{L}^1$, so dass $\mathbf{E}[X|\mathcal{G}]$ für alle $X \in \mathcal{L}^1$ eine $\mathcal{G}$-messbare Zufallsvariable ist mit*

1. $\mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]; A\big] = \mathbf{E}[X; A]$ *für alle* $A \in \mathcal{G}$.

Weiter gilt

2. $\mathbf{E}[X|\mathcal{G}] \geq 0$, *falls* $X \geq 0$.

3. $\mathbf{E}\big[|\mathbf{E}[X|\mathcal{G}]|\big] \leq \mathbf{E}[|X|]$.

4. *Falls* $0 \leq X_n \uparrow X$ *für* $n \to \infty$, *so ist auch* $\mathbf{E}[X_n|\mathcal{G}] \uparrow \mathbf{E}[X|\mathcal{G}]$ *in* $\mathcal{L}^1$, *falls alle Erwartungen existieren.*

5. *Falls* X *eine* $\mathcal{G}$*-messbare Funktion ist, so gilt* $\mathbf{E}[XY|\mathcal{G}] = X\mathbf{E}[Y|\mathcal{G}]$, *falls alle Erwartungen existieren.*

6. $\mathbf{E}\big[X\mathbf{E}[Y|\mathcal{G}]\big] = \mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]Y\big] = \mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]\mathbf{E}[Y|\mathcal{G}]\big]$, *falls alle Erwartungen existieren.*

7. *Ist* $\mathcal{H} \subseteq \mathcal{G}$, *so ist* $\mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]|\mathcal{H}\big] = \mathbf{E}[X|\mathcal{H}]$.

8. *Ist* X *unabhängig von* $\mathcal{G}$, *so ist* $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.

Beweis. 1. im Fall $X \in \mathcal{L}^2$: Sei M der abgeschlossene lineare Teilraum von $\mathcal{L}^2$, der aus allen Funktionen besteht, die bis auf eine Nullmenge mit einer $\mathcal{G}$-messbaren Funktion übereinstimmen. Nach Proposition 5.9 gibt es nun fast sicher eindeutige Funktionen $Y \in M, Z \perp M$ mit $X = Y + Z$. Wir definieren $\mathbf{E}[X|\mathcal{G}] := Y$. Damit gilt $X - \mathbf{E}[X|\mathcal{G}] \perp M$, also $\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]; A] = 0$ für $A \in \mathcal{G}$, woraus 1. für $X \in \mathcal{L}^2$ folgt.
3. im Fall $X \in \mathcal{L}^2$: Wähle $A := \{\mathbf{E}[X|\mathcal{G}] \geq 0\}$. Nach 1. gilt dann

$$\mathbf{E}[|\mathbf{E}[X|\mathcal{G}]|] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] - \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A^c] = \mathbf{E}[X; A] - \mathbf{E}[X; A^c] \leq \mathbf{E}[|X|].$$

1. im Fall $X \in \mathcal{L}^1$: Ist $X \in \mathcal{L}^1 \supset \mathcal{L}^2$, so wähle $X_1, X_2, \dots \in \mathcal{L}^2$ mit $||X_n - X||_1 \xrightarrow{n\to\infty} 0$ (etwa so, dass $|X_n| := |X| \wedge n$), und definiere $\mathbf{E}[X|\mathcal{G}] := \lim_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}]$. Dieser Grenzwert existiert in $\mathcal{L}^1$, da wegen 3.

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X_m|\mathcal{G}]|] = \mathbf{E}[|\mathbf{E}[X_n - X_m|\mathcal{G}]|] \leq \mathbf{E}[|X_n - X_m|] \xrightarrow{n,m\to\infty} 0$$

die Folge $(\mathbf{E}[X_n|\mathcal{G}])_{n=1,2,\dots}$ eine Cauchy-Folge ist und $\mathcal{L}^1$ vollständig ist. Außerdem gilt damit $||\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]||_1 \xrightarrow{n\to\infty} 0$. Weiter ist für $A \in \mathcal{G}$

$$\begin{aligned}|\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]; A]| &\leq \mathbf{E}[|X1_A - X_n1_A|] \\ &\quad + |\mathbf{E}[X_n - \mathbf{E}[X_n|\mathcal{G}]; A]]| \\ &\quad + \mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}]1_A - \mathbf{E}[X|\mathcal{G}]1_A|] \\ &\quad \xrightarrow{n\to\infty} 0\end{aligned}$$

wegen majorisierter Konvergenz und 1. folgt im Fall $X \in \mathcal{L}^1$.
3. im Fall $X \in \mathcal{L}^1$. Auch hier sieht man durch ein Approximationsargument, falls $X_1, X_2, \dots \in \mathcal{L}^2$ mit $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^1} X$

$$\mathbf{E}[|\mathbf{E}[X|\mathcal{G}]|] = \lim_{n\to\infty} \mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}]|] \leq \lim_{n\to\infty} \mathbf{E}[|X_n|] = \mathbf{E}[|X|],$$

da wegen der umgekehrten Dreiecksungleichung etwa

$$\mathbf{E}[||\mathbf{E}[X_n|\mathcal{G}]| - |\mathbf{E}[X|\mathcal{G}]||] \leq \mathbf{E}[|\mathbf{E}[X|\mathcal{G}] - \mathbf{E}[X_n|\mathcal{G}]|] \xrightarrow{n\to\infty} 0.$$

2. Setze $A = \{\mathbf{E}[X|\mathcal{G}] \leq 0\}$ und damit

$$0 \geq \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = 0,$$

also wegen $\mathbf{E}[X|\mathcal{G}]1_A \le 0$ auch $\mathbf{E}[X|\mathcal{G}]1_A = 0$ fast sicher.
4. Wegen monotoner Konvergenz ist $||X_n - X||_1 \xrightarrow{n\to\infty} 0$, also nach 3.

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]|] = \mathbf{E}[|\mathbf{E}[X_n - X|\mathcal{G}]|] \le \mathbf{E}[|X_n - X|] \xrightarrow{n\to\infty} 0.$$

6. im Fall $X, Y \in \mathcal{L}^2$. Nach der Definition der bedingten Erwartung ist $\mathbf{E}[X|\mathcal{G}], \mathbf{E}[Y|\mathcal{G}] \in M$, wenn M der lineare Teilraum von $\mathcal{L}^2$ besteht, der Funktionen beinhaltet, die bis auf eine Nullmenge mit einer $\mathcal{G}$-messbaren Funktion übereinstimmen. Außerdem ist $X - \mathbf{E}[X|\mathcal{G}] \perp M$. Damit ist

$$\mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}])\mathbf{E}[Y|\mathcal{G}]] = 0.$$

6. im Fall $X, Y \in \mathcal{L}^1$. Wähle $X_1, Y_1, X_2, Y_2, \dots \in \mathcal{L}^2$ mit $X_n \uparrow X, Y_n \uparrow Y$. Wegen 4. und majorisierter Konvergenz gilt dann, falls alle Erwartungen existieren,

$$\mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}])\mathbf{E}[Y|\mathcal{G}]] = \lim_{n\to\infty} \mathbf{E}[(X_n - \mathbf{E}[X_n|\mathcal{G}])\mathbf{E}[Y_n|\mathcal{G}]] = 0.$$

5. Wegen 1. ist $\mathbf{E}[X|\mathcal{G}]1_A = X1_A$ für $A \in \mathcal{G}$ fast sicher. Damit ist auch

$$\mathbf{E}[XY; A] = \mathbf{E}[X\mathbf{E}[Y|\mathcal{G}]; A]$$

nach 6. Hieraus folgt nach 1. bereits $\mathbf{E}[XY|\mathcal{G}] = X\mathbf{E}[Y|\mathcal{G}]$.
7. Da $\mathcal{H} \subseteq \mathcal{G}$, ist für $A \in \mathcal{H}$

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = \mathbf{E}[\mathbf{E}[X|\mathcal{H}]; A]$$

nach 1. Hier aus folgt aber $\mathbf{E}[\mathbf{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbf{E}[X|\mathcal{H}]$.
8. Sicherlich ist $\mathbf{E}[X]$ messbar bezüglich $\mathcal{G}$. Für $A \in \mathcal{G}$ ist außerdem

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = \mathbf{E}[X]\mathbf{E}[1_A] = \mathbf{E}\big[\mathbf{E}[X]; A\big]$$

und damit $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$. □

Bemerkung 12.3 (Interpretation und alternativer Beweis).

1. Sei $X \in \mathcal{L}^2$. Wie der Beweis von 1. in Theorem 12.2 zeigt, ist $X - \mathbf{E}[X|\mathcal{G}]$ senkrecht auf dem linearen Teilraum aller $\mathcal{G}$-messbaren Funktionen. Insbesondere ist $\mathbf{E}[X|\mathcal{G}]$ diejenige $\mathcal{G}$-messbare Zufallsvariable, die (im Sinne der $\mathcal{L}^2$-Norm) der Zufallsvariable X am nächsten kommt. Deswegen kann man sagen, dass $\mathbf{E}[X|\mathcal{G}]$ die beste Schätzung von X ist, wenn Informationen aus der σ-Algebra $\mathcal{G}$ zur Verfügung stehen.

2. Die fast sicher eindeutige Existenz der bedingten Erwartung mit der Eigenschaft 1. in Theorem 12.2 kann man anders als oben mit Hilfe des Satzes von Radon-Nikodým (Korollar 5.16) beweisen:

   Sei zunächst $X \ge 0$. Setze $\widetilde{\mathbf{P}} := \mathbf{P}|_{\mathcal{G}}$, die Einschränkung von $\mathbf{P}$ auf $\mathcal{G}$, und $\mu(.) := \widetilde{\mathbf{E}}[X; .]$ ein endliches Maß. Dann gilt offenbar $\mu \ll \widetilde{\mathbf{P}}$. Der Satz von Radon-Nikodým stellt sicher, dass μ eine Dichte bzgl. $\widetilde{\mathbf{P}}$ hat, d.h. es eine $\mathcal{G}$-messbare Zufallsvariable Z gibt mit

   $$\mathbf{E}[X; A] = \widetilde{\mathbf{E}}[X; A] = \mu(A) = \widetilde{\mathbf{E}}[Z; A] = \mathbf{E}[Z; A]$$

   für alle $A \in \mathcal{G}$. Damit erfüllt Z die Eigenschaften von 1. aus Theorem 12.2. Der allgemeinen Fall (d.h. X kann auch negative Werte annehmen) folgt dann mit der Zerlegung $X = X^+ - X^-$.

Zum Beweis der (fast sicheren) Eindeutigkeit der bedingten Erwartung sei Z' eine weitere, $\mathcal{G}$-messbare Zufallsvariable mit $\mathbf{E}[Z';A] = \mathbf{E}[X;A]$ für alle $A \in \mathcal{G}$. Dann ist $B := \{Z' - \mathbf{E}[X|\mathcal{G}] > 0\} \in \mathcal{G}$ und $\mathbf{E}[\mathbf{E}[X|\mathcal{G}] - Z'; B] = \mathbf{E}[X - X; B] = 0$ und ebenso $\mathbf{E}[\mathbf{E}[X|\mathcal{G}] - Z'; B^c] = 0$. Das heißt also $Z' = \mathbf{E}[X|\mathcal{G}]$ fast sicher.

Proposition 12.4 (Jensen'sche Ungleichung für bedingte Erwartungen). *Sei I ein offenes Intervall, $\mathcal{G} \subseteq \mathcal{A}$ und $X \in \mathcal{L}^1$ mit Werten in I und $\varphi : I \to \mathbb{R}$ konvex. Dann gilt*

$$\mathbf{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbf{E}[X|\mathcal{G}]).$$

Beweis. Der Beweis verläuft analog zu dem der Jensen'schen Ungleichung im unbedingten Fall, Proposition 7.6: Da I offen ist, liegt $\mathbf{E}[X|\mathcal{G}] \in I$ fast sicher. Wir erinnern an die Definition von λ in (7.2). Weiter ist, wie in (7.3) für $x \in I$

$$\varphi(x) \geq \varphi(\mathbf{E}[X|\mathcal{G}]) + \lambda(\mathbf{E}[X|\mathcal{G}])(x - \mathbf{E}[X|\mathcal{G}])$$

und damit

$$\begin{aligned}\mathbf{E}[\varphi(X)|\mathcal{G}] &\geq \mathbf{E}[\varphi(\mathbf{E}[X|\mathcal{G}])|\mathcal{G}] + \mathbf{E}[\lambda(\mathbf{E}[X|\mathcal{G}]) \cdot (X - \mathbf{E}[X|\mathcal{G}])|\mathcal{G}] \\ &= \varphi\big(\mathbf{E}[X|\mathcal{G}]\big).\end{aligned}$$ □

Lemma 12.5 (Gleichgradige Integrierbarkeit und bedingte Erwartung). *Sei $X \in \mathcal{L}^1$. Dann ist die Familie $(\mathbf{E}[X|\mathcal{G}])_{\mathcal{G} \subseteq \mathcal{A}}$ gleichgradig integrierbar.*

Beweis. Da $\{X\}$ gleichgradig integrierbar ist, gibt es nach Lemma 8.9 eine monoton wachsende konvexe Funktion $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ mit $\frac{\varphi(x)}{x} \xrightarrow{x\to\infty} \infty$ und $\mathbf{E}[\varphi(|X|)] < \infty$. Mit Theorem 12.2.3 folgt

$$\sup_{\mathcal{F} \subseteq \mathcal{A}} \mathbf{E}[\varphi(|\mathbf{E}[X|\mathcal{F}]|)] \leq \mathbf{E}[\varphi(|X|)] < \infty.$$

Damit ist $\{\mathbf{E}[X|\mathcal{F}] : \mathcal{F} \subseteq \mathcal{A}\ \sigma\text{-Algebra}\}$ gleichgradig integrierbar, wieder nach Lemma 8.9. □

Theorem 12.6 (Majorisierte und monotone Konvergenz für bedingte Erwartungen). *Sei $\mathcal{G} \subseteq \mathcal{F}$ und $X_1, X_2, \dots \in \mathcal{L}^1$. Es gelte eine der Bedingungen:*

1. *Sei $X \in \mathcal{L}^1$, so dass $X_n \uparrow X$ fast sicher.*
2. *Ist $Y \in \mathcal{L}^1$, so dass $|X_n| \leq |Y|$ für alle n, und $X_n \xrightarrow{n\to\infty} X$ fast sicher.*

Dann gilt

$$\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{n\to\infty} \mathbf{E}[X|\mathcal{G}]$$

fast sicher und in $\mathcal{L}^1$.

Beweis. Für die $\mathcal{L}^1$-Konvergenz hat man in beiden Fällen mit Theorem 12.2.3

$$\begin{aligned}\mathbf{E}\big[\big|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]\big|\big] &= \mathbf{E}\big[\big|\mathbf{E}[X_n - X|\mathcal{G}]\big|\big] \\ &\leq \mathbf{E}[|X_n - X|] \xrightarrow{n\to\infty} 0.\end{aligned}$$

Die fast sichere Konvergenz teilen wir in die beiden Fälle auf: im Fall 1. ist nach Theorem 12.2.2 klar, dass $\mathbf{E}[X_n|\mathcal{G}]$ monoton wächst. Außerdem ist für $A \in \mathcal{F}$ mit dem Satz der monotonen Konvergenz

$$\mathbf{E}\big[\sup_n \mathbf{E}[X_n|\mathcal{G}]; A\big] = \sup_n \mathbf{E}\big[\mathbf{E}[X_n|\mathcal{G}]; A\big] = \sup_n \mathbf{E}[X_n; A] = \mathbf{E}[\sup_n X_n; A] = \mathbf{E}[X; A].$$

Damit ist aber gezeigt, dass fast sicher $\sup_n \mathbf{E}[X_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$ gilt.

Im Fall 2. setzen wir

$$\begin{aligned} Y_n &:= \sup_{k\geq n} X_k \downarrow \limsup_n X_n = X \text{ fast sicher,} \\ Z_n &:= \inf_{k\geq n} X_k \uparrow \liminf_n X_n = X \text{ fast sicher.} \end{aligned}$$

Damit ist $-Y \leq Z_n \leq X_n \leq Y_n \leq Y$, also insbesondere $Y_1, Z_1, Y_2, Z_2, \dots \in \mathcal{L}^1$, also ist nach 1.

$$\mathbf{E}[X|\mathcal{G}] = \lim_{n\to\infty} \mathbf{E}[Z_n|\mathcal{G}] \leq \lim_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}] \leq \lim_{n\to\infty} \mathbf{E}[Y_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$$

fast sicher. Insbesondere ist also $\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{n\to\infty} \mathbf{E}[X|\mathcal{G}]$ fast sicher. □

12.3 Der Fall $\mathcal{G} = \sigma(X)$

Im Falle $\mathcal{G} = \sigma(X)$ bedeutet $\mathbf{E}[Y|X] := \mathbf{E}[Y|\sigma(X)]$ die Erwartung von Y, gegeben, dass die Zufallsvariable X festgelegt ist. Dies ist eine Funktion von X, wie Proposition 12.7 zeigt.

Proposition 12.7 (Bedingung auf eine Zufallsvariable). *Sei $(\Omega', \mathcal{F}')$ ein Messraum, X eine Ω'-wertige Zufallsvariable mit Werten in Ω' und $Y \in \mathcal{L}^1$. Dann existiert eine $\mathcal{F}'/\mathcal{B}(\mathbb{R})$-messbare Abbildung $\varphi : \Omega' \to \mathbb{R}$ mit $\mathbf{E}[Y|X] = \varphi(X)$.*

Beweis. Klar nach Lemma 7.2 □

Beispiel 12.8 (Zufällige Erfolgswahrscheinlichkeit). Betrachten wir die in Beispiel 12.1 gestellte Frage nach der Existenz der bedingten Wahrscheinlichkeit $\mathbf{P}(Y = k|X)$, wobei X uniform auf $[0, 1]$ ist und X unabhängig binomial verteilt mit n und X. Wir zeigen nun, dass

$$\mathbf{P}(Y = k|X) = \binom{n}{k} X^k (1 - X)^{n-k}. \tag{12.3}$$

Sei $A = \{X \in I\}$ für $I \in \mathcal{B}([0, 1])$, d.h. A ist eine $\sigma(X)$-messbare Menge. Dann gilt

$$\mathbf{E}[1_{Y=k}; A] = \mathbf{P}(Y = k, X \in I) = \int_I \binom{n}{k} x^k (1 - x)^{n-k} dx = \mathbf{E}\Big[\binom{n}{k} X^k (1 - X)^{n-k}; A\Big]$$

Dies bedeutet aber, dass (12.3) stimmt.

Beispiel 12.9 (Summen unabhängiger identisch verteilter Zufallsvariable). Seien $X_1, X_2, \dots$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen, $\mu = \mathbf{E}[X_1]$ und $S_n := X_1 + \dots + X_n$. Dann ist

$$\begin{aligned} \mathbf{E}[S_n|X_1] &= \mathbf{E}[X_1|X_1] + \mathbf{E}[X_2 + \dots + X_n|X_1] = X_1 + (n - 1)\mu, \\ \mathbf{E}[X_1|S_n] &= \tfrac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i|S_n] = \tfrac{1}{n}\mathbf{E}[S_n|S_n] = \tfrac{1}{n} S_n. \end{aligned}$$

In der zweiten Rechnung ist also beispielsweise für $X = S_n$ und $Y = X_1$ die Funktion φ aus Proposition 12.7 gegeben durch $\varphi(x) = \frac{1}{n} x$.

Beispiel 12.10 (Buffon's Nadelproblem). Auf einer Ebene liegen Geraden im horizontalen Abstand 1. Es werden Nadeln der Länge 1 auf die Ebene geworfen; siehe Abbildung 12.1. Betrachten wir eine Nadel. Wir setzen

$$Z := \begin{cases} 1, & \text{falls die Nadel eine Gerade schneidet} \\ 0, & \text{sonst} \end{cases}.$$

Dabei ist der Mittelpunkt der Nadel X von der linken Geraden entfernt und die Verlängerung

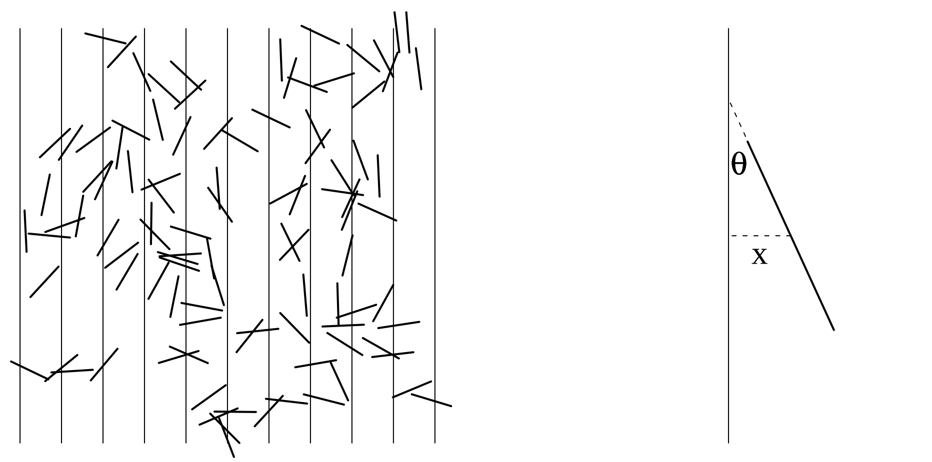


Abbildung 12.1: Skizze zu Buffon's Nadelproblem.

der Nadel geht einen spitzen Winkel Θ mit der linken Gerade ein. Damit ist X uniform auf $[0;1]$, Θ uniform auf $[0;\frac{\pi}{2}]$ unabhängig und es gilt

$$\mathbf{P}(Z=1|\Theta) = \mathbf{P}(X \le \tfrac{1}{2}\sin(\Theta) \text{ oder } X \ge 1 - \tfrac{1}{2}\sin(\Theta)|\Theta) = \sin(\Theta).$$

Damit ist

$$\mathbf{P}(Z=1) = \mathbf{E}[\mathbf{P}(Z=1|\Theta)] = \mathbf{E}[\sin(\Theta)] = \frac{2}{\pi}\int_0^{\pi/2}(\sin(\theta)d\theta = \frac{2}{\pi}.$$

Dies kann man so interpretieren: will man durch Simulation (d.h. also durch ein Monte-Carlo Verfahren) den numerischen Wert von π herausfinden, kann man Buffon's Nadeln simulieren. Da jede einzelne Nadel die Wahrscheinlichkeit $\frac{2}{\pi}$ hat, eine vertikale Linien zu treffen, ist etwa

$$\pi \approx \frac{2}{\text{Anteil der Nadeln, die eine Vertikale treffen}}$$

nach dem Gesetz der großen Zahlen.

Beispiel 12.11 (Suchen in Listen). Gegeben seien n Namen von Personen, die aus r verschiedenen Städten kommen. Jede Person kommt (unabhängig von jeder anderen) mit Wahrscheinlichkeit p_j aus Stadt j, $j = 1, \dots, r$. Die Namen (zusammen mit anderen persönlichen Daten) werden in r verschiedene (ungeordnete) Listen eingetragen. Will man nun eine (zufällige, nach den Wahrscheinlichkeiten $p_1, \dots, p_r$ verteilte) Person in der Liste suchen, bestimmt man zunächst die Stadt j, aus der die Person kommt und sucht anschließend in Liste

j nach dem Personennamen. Bis man feststellt, dass der Name nicht auf der Liste auftaucht, muss man die zu findende Person mit Namen auf der Liste vergleichen. Die Frage ist nun: Wir oft hat man im Mittel ohne Erfolg den Namen der zu findenden Person mit Namen auf der Liste vergleichen, bis man endgültig weiß, dass die Person nicht auf der Liste steht?

Wir definieren zunächst ein paar Zufallsvariablen:
J : Nummer der Stadt, aus der die zu suchende Person kommt
L : Anzahl der unerfolgreichen Vergleiche, bis man den Namen der zu findenden Person findet
Z_j : Anzahl der Personen aus Stadt j
sowie $Z = (Z_1, \dots, Z_r)$. Um $\mathbf{E}[L]$ zu ermitteln, bestimmen wir zunächst

$$\mathbf{P}(L = a|J, Z) = 1_{Z_J = a}$$

und damit

$$\mathbf{P}(L = a|Z) = \sum_{j=1}^{r} p_j 1_{Z_j = a}.$$

Daraus folgern wir

$$\mathbf{E}[L|Z] = \sum_{a=1}^{\infty} \sum_{j=1}^{r} a \cdot p_j \cdot 1_{Z_j = a} = \sum_{j=1}^{r} p_j Z_J$$

und deshalb

$$\mathbf{E}[L] = \mathbf{E}[\mathbf{E}[L|Z]] = \sum_{j=1}^{r} p_j \mathbf{E}[Z_j] = n \cdot \sum_{j=1}^{r} p_j^2.$$

12.4 Bedingte Unabhängigkeit

In Abschnitt 9 haben wir bereits die Unabhängigkeit von σ-Algebren (oder von Zufallsvariablen) kennen gelernt. Bedingte Erwartungen und Unabhängigkeit sind eng verwandt, wie das erste Lemma zeigt.

Lemma 12.12 (Bedingte Wahrscheinlichkeit und Unabhängigkeit). *Die σ-Algebren $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ sind genau dann unabhängig, wenn $\mathbf{P}(G|\mathcal{H}) = \mathbf{P}(G)$ für alle $G \in \mathcal{G}$.*

Beweis. '$\Rightarrow$': Wenn $\mathcal{G}$ und $\mathcal{H}$ unabhängig sind, so ist für $G \in \mathcal{G}, H \in \mathcal{H}$

$$\mathbf{E}[\mathbf{P}(G), H] = \mathbf{P}(G \cap H) = \mathbf{E}[\mathbf{P}(G|\mathcal{H}), H].$$

Damit ist $\mathbf{P}(G|\mathcal{H}) = \mathbf{P}(G)$ nach der Definition der bedingten Erwartung.
'$\Leftarrow$': Gilt also $\mathbf{P}(G|\mathcal{H}) = \mathbf{P}(G)$, so folgt für $H \in \mathcal{H}$

$$\mathbf{P}(G \cap H) = \mathbf{E}[1_G, H] = \mathbf{E}[\mathbf{P}(G|\mathcal{H}), H] = \mathbf{E}[\mathbf{P}(G), H] = \mathbf{P}(G) \cdot \mathbf{P}(H).$$ □

Oft benötigt man das Konzept der Unabhängigkeit auch noch in einer bedingten Form. Dafür geben wir zunächst ein wichtiges Beispiel an.

Beispiel 12.13 (Markov-Ketten). Sei E eine abzählbare Menge. Eine Markov-Kette $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ ist eine Familie von E-wertigen Zufallsvariablen, so dass für alle $A \subseteq E$

$$\mathbf{P}(X_{t+1} \in A|X_0, \dots, X_t) = \mathbf{P}(X_{t+1} \in A|X_t). \tag{12.4}$$

Dies bedeutet: wenn man die Verteilung von X_{t+1} wissen will, und dabei schon die Informationen der Zufallsvariable X_t zur Verfügung hat, bringt die Information über die Zufallsvariablen $X_0, \dots, X_{t-1}$ keine zusätzliche Information. Man sagt auch:

Gegeben X_t ist X_{t+1} unabhängig von $X_0, \ldots, X_{t-1}$.

Oder in Termen von σ-Algebren:

Gegeben $\sigma(X_t)$ ist $\sigma(X_{t+1})$ unabhängig von $\sigma(X_0, \ldots, X_{t-1})$.

Etwas umgangssprachlich sagt man auch: gegeben die Gegenwart (das ist der Zustand zur Zeit t, X_t) ist die Zukunft (d.h. X_{t+1}) unabhängig von der Vergangenheit (das sind die Zustände $X_0, \ldots, X_{t-1}$).

Ein einfaches Beispiel für eine Markov-Kette ist die ein-dimensionale Irrfahrt: seien $Y_1, Y_2, \ldots$ unabhängig und identisch verteilt, so dass $\mathbf{P}(Y_1 = 1) = p$ und $\mathbf{P}(Y_1 = -1) = q$ für ein $p \in [0,1]$. Weiter sei $X_0 = 0$ und $X_t = Y_1 + \cdots + Y_t$. Dann ist $(X_t)_{t \geq 0}$ eine Markov-Kette, denn

$$\mathbf{P}(X_{t+1} = k | X_0, \ldots, X_t) = \begin{cases} p, & k = X_t + 1, \\ q, & k = X_t - 1. \end{cases}$$

Insbesondere definiert die rechte Seite eine X_t-messbare Zufallsvariable und ist damit gleich $\mathbf{P}(X_{t+1} = k | X_t)$.

Definition 12.14 (Bedingte Unabhängigkeit). *Sei $\mathcal{G} \subseteq \mathcal{F}$. Eine Familie $(\mathcal{C}_i)_{i \in I}$ von Mengensystemen mit $\mathcal{C}_i \subseteq \mathcal{F}$ heißt* unabhängig gegeben $\mathcal{G}$, *falls*

$$\mathbf{P}\Big(\bigcap_{j \in J} A_j | \mathcal{G}\Big) = \prod_{j \in J} \mathbf{P}(A_j | \mathcal{G}) \tag{12.5}$$

für alle $J \Subset I$ und $A_j \in \mathcal{C}_j, j \in J$, gilt.

Analog definiert man die bedingte Unabhängigkeit für Zufallsvariablen. Sei Y eine Zufallsvariable. Eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen ist unabhängig gegeben $\mathcal{G}$ (bzw. Y) falls $(\sigma(X_i))_{i \in I}$ unabhängig gegeben $\mathcal{G}$ (bzw. $\sigma(Y)$) ist.

Beispiel 12.15 (Einfache Fälle). Seien $\mathcal{G} \subseteq \mathcal{F}$ eine σ-Algebra und $(\mathcal{C}_i)_{i \in I}$ eine Familie von Mengensystemen.

1. Ist $\mathcal{G} = \mathcal{F}$, so ist $(\mathcal{C}_i)_{i \in I}$ immer unabhängig gegeben $\mathcal{G}$.

2. Ist $\mathcal{G} = \{\emptyset, \Omega\}$, so ist $(\mathcal{C}_i)_{i \in I}$ genau dann unabhängig gegeben $\mathcal{G}$, wenn $(\mathcal{C}_i)_{i \in I}$ unabhängig sind.

Beispiel 12.16 (Binomialverteilung mit zufälliger Erfolgswahrscheinlichkeit). Wir betrachten nochmal den Münzwurf mit zufälliger Erfolgswahrscheinlichkeit aus Beispiel 12.1 und 12.8. Hier war X uniform auf $[0,1]$ verteilt und, gegeben X sind $Y_1, \ldots, Y_n$ Bernoulli-verteilt. Nun sollte ja gelten, dass $(Y_1, \ldots, Y_n)$ unabhängig gegeben X sind. Genau wie in Beispiel 12.8 berechnen wir für $A = \{X \in I\}$ für ein $I \in \mathcal{B}([0,1])$ und $y_1, \ldots, y_n \in \{0,1\}$ und $k := y_1 + \cdots + y_n$

$$\begin{aligned} \mathbf{E}[1_{Y_1 = y_1, \ldots, Y_n = y_n}, A] &= \mathbf{P}(Y_1 = y_1, \ldots, Y_n = y_n, X \in I) \\ &= \int_I x^{y_1 + \cdots + y_n} (1-x)^{n - y_1 - \cdots - y_n} dx = \mathbf{E}\big[X^k (1-X)^{n-k}, A\big], \end{aligned}$$

also

$$\mathbf{P}(Y_1 = y_1, \ldots, Y_n = y_n | X) = X^k (1-X)^{n-k}.$$

Analog zeigt man für $i = 1, \dots, n$

$$\mathbf{P}(Y_i = y_1|X) = X^{y_i}(1-X)^{1-y_i}.$$

Daraus folgt

$$\mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n|X) = \prod_{i=1}^n \mathbf{P}(Y_i = y_1|X),$$

also sind $(Y_1, \dots, Y_n)$ unabhängig gegeben X.

Lemma 12.12 gibt es auch in folgender Version, in der die Unabhängigkeit durch die bedingte Unabhängigkeit ausgetauscht ist.

Proposition 12.17 (Bedingte Wahrscheinlichkeit und bedingte Unabhängigkeit). *Sei $\mathcal{K} \subseteq \mathcal{F}$ eine σ-Algebra. Die σ-Algebren $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ sind genau dann unabhängig gegeben $\mathcal{K}$, wenn $\mathbf{P}(G|\sigma(\mathcal{H}, \mathcal{K})) = \mathbf{P}(G|\mathcal{K})$ für alle $G \in \mathcal{G}$.*

Beweis. '$\Rightarrow$': Wenn $\mathcal{G}$ und $\mathcal{H}$ unabhängig gegeben $\mathcal{K}$ sind, so ist für $G \in \mathcal{G}, H \in \mathcal{H}, K \in \mathcal{K}$

$$\mathbf{E}[\mathbf{P}(G|\mathcal{K}), H \cap K] = \mathbf{E}[\mathbf{P}(G|\mathcal{K})\mathbf{P}(H|\mathcal{K}), K] = \mathbf{E}[\mathbf{P}(G \cap H|\mathcal{K}), K] = \mathbf{P}(G \cap H \cap K).$$

Nun kann man zeigen, dass das Mengensystem

$$\mathcal{D} := \{A \in \sigma(\mathcal{H}, \mathcal{K}) : \mathbf{E}[\mathbf{P}(G|\mathcal{K}), A] = \mathbf{P}(G \cap A)\}$$

ein schnittstabiles Dynkin-System ist mit $\mathcal{D} \supseteq \mathcal{H}, \mathcal{K}$. Nun folgt mit Theorem 2.13, dass $\mathcal{D} = \sigma(\mathcal{H}, \mathcal{K})$, woraus $\mathbf{P}(G|\sigma(\mathcal{H}, \mathcal{K})) = \mathcal{P}(G|\mathcal{K})$ folgt.
'$\Leftarrow$': Gilt also $\mathbf{P}(G|\sigma(\mathcal{H}, \mathcal{K})) = \mathbf{P}(G|\mathcal{K})$, so folgt für $H \in \mathcal{H}$

$$\mathbf{P}(G \cap H|\mathcal{K}) = \mathbf{E}[\mathbf{P}(G|\sigma(\mathcal{H}, \mathcal{K})), H|\mathcal{K}] = \mathbf{E}[\mathbf{P}(G|\mathcal{K}), H|\mathcal{K}] = \mathbf{P}(G|\mathcal{K}) \cdot \mathbf{P}(H|\mathcal{K}).$$ □

Beispiel 12.18 (Markov-Ketten). Betrachten wir nochmal die Markov-Kette $(X_t)_{t=0,1,2,\dots}$ aus Beispiel 12.13. Für festes t setzen wir $\mathcal{G} = \sigma(X_{t+1}), \mathcal{H} = \sigma(X_0, \dots, X_{t-1}), \mathcal{K} = \sigma(X_t)$. Die Markov-Eigenschaft (12.4) sagt nun für $G \in \mathcal{G}, H \in \mathcal{H}, K \in \mathcal{K}$, dass $\mathbf{P}(G|\sigma(\mathcal{H}, \mathcal{K})) = \mathbf{P}(G|\mathcal{K})$. Nach Proposition 12.17 bedeutet dies, dass X_{t+1} und $(X_0, \dots, X_{t-1})$ unabhängig gegeben X_t sind.

12.5 Reguläre Version der bedingten Verteilung

Wir haben im Abschnitt 12.1 gesehen, wie die bedingte Wahrscheinlichkeit $\mathbf{P}(A|\mathcal{G}) := \mathbf{E}[1_A|\mathcal{G}]$ für eine σ-Algebra $\mathcal{G} \subseteq \mathcal{F}$ definiert ist. Dies bedeutet jedoch noch *nicht*, dass wir ein Wahrscheinlichkeitsmaß $A \mapsto \mathbf{P}(A|\mathcal{G})$ definiert haben; siehe hierzu die nächste Bemerkung. In den meisten Fällen kann man jedoch ein solches (zufälliges, $\mathcal{G}$-messbares) Maß definieren, die (oder besser: eine) reguläre Version der bedingten Verteilung.

Bemerkung 12.19 (Bedingte Wahrscheinlichkeiten und bedingte Verteilungen). Sei $\mathcal{G} \subseteq \mathcal{F}$ eine σ-Algebra und $A_1, A_2, \dots \in \mathcal{F}$ mit $A_i \cap A_j = \emptyset$. Dann ist für $B \in \mathcal{G}$

$$\begin{aligned}\mathbf{E}\Big[\mathbf{P}\Big(\bigcup_{n=1}^{\infty} A_n|\mathcal{G}\Big); B\Big] &= \mathbf{E}\big[\mathbf{E}[1_{\bigcup_{n=1}^{\infty} A_n}|\mathcal{G}]; B\big] = \mathbf{E}[1_{\bigcup_{n=1}^{\infty} A_n}; B] \\ &= \mathbf{E}\Big[\sum_{n=1}^{\infty} 1_{A_n}; B\Big] = \sum_{n=1}^{\infty} \mathbf{E}[1_{A_n}; B] \\ &= \sum_{n=1}^{\infty} \mathbf{E}\big[\mathbf{P}(A_n|\mathcal{G}); B\big] = \mathbf{E}\Big[\sum_{n=1}^{\infty} \mathbf{P}(A_n|\mathcal{G}); B\Big]\end{aligned}$$

und damit

$$\mathbf{P}\Big(\bigcup_{n=1}^{\infty} A_n|\mathcal{G}\Big) = \sum_{n=1}^{\infty} \mathbf{P}(A_n|\mathcal{G}) \tag{12.6}$$

$\mathbf{P}$-fast überall. Das bedeutet, dass es eine von $A_1, A_2, \dots$ abhängige Nullmenge gibt, so dass (12.6) für alle ω außerhalb dieser Nullmenge gilt. Da es aber überabzählbar viele Folgen $A_1, A_2, \dots \in \mathcal{F}$ gibt, muss es damit nicht notwendigerweise eine Nullmenge N geben, so dass (12.6) für jede Wahl von $A_1, A_2, \dots \in \mathcal{F}$ und $\omega \notin N$ gilt. Falls es jedoch ein solches N gibt, werden wir sagen, dass eine *reguläre Version der bedingten Verteilung* von $\mathbf{P}$ gegeben $\mathcal{G}$ existiert. Bedingungen hierfür werden wir in Theorem 12.22 kennenlernen.

Wir erinnern an den Begriff des stochastischen Kernes; siehe Definition 6.9.

Definition 12.20 (Reguläre Version der bedingten Verteilung). *Sei $(\Omega', \mathcal{F}')$ ein Messraum, Y eine Ω'-wertige messbare Zufallsvariable und $\mathcal{G} \subseteq \mathcal{F}$. Ein stochastischer Kern $\kappa_{Y,\mathcal{G}}$ von $(\Omega, \mathcal{G})$ nach $(\Omega', \mathcal{F}')$ heißt* reguläre Version der bedingten Verteilung *von Y, gegeben $\mathcal{G}$, falls*

$$\kappa_{Y,\mathcal{G}}(\omega, B) = \mathbf{P}(Y \in B|\mathcal{G})(\omega)$$

für $\mathbf{P}$-fast alle ω und jedes $B \in \mathcal{F}'$.

Bemerkung 12.21 (Auf eine Zufallsvariable bedingte Verteilung).

1. Für den stochastischen Kern aus Definition 12.20 reicht es, die Eigenschaft (ii) aus Definition 6.9 nur für einen schnittstabilen Erzeuger $\mathcal{C}$ von $\mathcal{F}$ zu fordern. Es ist nämlich stets

$$\mathcal{D} := \{A' \in \mathcal{F}' : \omega \mapsto \kappa(\omega, A') \text{ ist } \mathcal{A}\text{-messbar}\}$$

ein Dynkin-System. Damit ist nach Theorem 2.13 auch $\mathcal{D} = \sigma(\mathcal{C})$.

2. Sei $\mathcal{G} = \sigma(X)$ für eine Zufallsvariable X in Definition 12.20.2. Ist dann $\kappa_{Y,\sigma(X)}$ eine reguläre Version der bedingten Erwartung von Y gegeben $\sigma(X)$, so ist $\omega \mapsto \kappa_{Y,\sigma(X)}(\omega, A')$ $\sigma(X)$-messbar für alle $A' \in \mathcal{A}'$. Damit gibt es nach Proposition 12.7 eine nach $\sigma(X)/\mathcal{B}([0;1])$-messbare Abbildung $\varphi_{A'} : \Omega \to [0;1]$ mit $\varphi_{A'} \circ X = \kappa_{Y,\sigma(X)}(.,A')$. Wir setzen dann

$$\kappa_{Y,X}(x, A') := \varphi_{A'}(x)$$

und sagen $\kappa_{Y,X}$ ist die reguläre Version der bedingten Verteilung von Y gegeben X.

Theorem 12.22 (Existenz der regulären Version der bedingten Verteilung). *Sei (E, r) ein vollständiger und separabler metrischer Raum, ausgestattet mit der Borel'schen σ-Algebra, $\mathcal{G} \subseteq \mathcal{F}$ eine σ-Algebra und Y eine (nach $\mathcal{F}$ messbare) Zufallsvariable mit Werten in E. Dann existiert eine reguläre Version der bedingten Verteilung von Y gegeben $\mathcal{G}$.*

Bevor wir das Theorem beweisen können, benötigen wir eine Eigenschaft (Proposition 12.24) über vollständige, separable metrische Räume.

Definition 12.23 (Borel'scher Raum).
1. *Zwei Messräume $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ heißen* isomorph*, falls es eine bijektive, nach $\mathcal{F}/\mathcal{F}'$-messbare Abbildung $\varphi : \Omega \to \Omega'$ gibt, so dass φ^{-1} nach $\mathcal{F}'/\mathcal{F}$-messbar ist.*
2. *Ein Messraum $(\Omega, \mathcal{F})$ heißt* Borel'scher Raum*, falls es eine Borel'sche Menge $A \in \mathcal{B}(\mathbb{R})$ gibt, so dass $(\Omega, \mathcal{F})$ und $(A, \mathcal{B}(A))$ isomorph sind.*

Proposition 12.24 (Polnische und Borel'sche Räume). *Jeder vollständige und separable metrische Raum (E, r), ausgestattet mit der Borel'schen σ-Algebra, ist ein Borel'scher Raum.*

Beweis. Siehe etwa Dudley, Real analysis and probability, Theorem 13.1.1. □

Beweis von Theorem 12.22. Wir beweisen das Theorem unter der schwächeren Voraussetzung, dass E, ausgestattet mit der Borel'schen σ-Algebra, ein Borel'scher Raum ist. O.E. können wir also annehmen, dass $E \in \mathcal{B}(\mathbb{R})$ ist. Die Strategie unseres Beweises besteht darin, eine Verteilungsfunktion der bedingten Verteilung zu finden, indem diese erst für rationale Werte festgelegt wird, bevor sie auf alle reellen Zahlen fortgesetzt wird.

Für $r \in \mathbb{Q}$ sei F_r eine Version von $\mathbf{P}(Y \leq r|\mathcal{G})$ (d.h. $F_r = \mathbf{P}(Y \leq r|\mathcal{G})$ fast sicher. Sei $A \in \mathcal{F}$ so, dass für $\omega \in A$ die Abbildung $r \mapsto F_r(\omega)$ nicht-fallend ist mit Grenzwerten 1 und 0 bei $\pm\infty$. Da A durch abzählbar viele Bedingungen gegeben ist, die alle fast sicher erfüllt sind, folgt $\mathbf{P}(A) = 1$. Definiere nun für $x \in \mathbb{R}$

$$F_x(\omega) := 1_A(\omega) \cdot \inf_{r>x} F_r(\omega) + 1_{A^c}(\omega) \cdot 1_{x \geq 0}.$$

Damit ist $x \mapsto F_x(\omega)$ für alle ω eine Verteilungsfunktion. Definiere

$$\kappa(\omega, .) := \text{Maß, das durch } x \mapsto F_x(\omega) \text{ definiert ist.}$$

Für $r \in \mathbb{Q}$ und $B = (-\infty; r]$ ist

$$\omega \mapsto \kappa(\omega, B) = 1_A(\omega) \cdot \mathbf{P}(Y \leq r|\mathcal{G})(\omega) + 1_{A^c}(\omega) \cdot 1_{r \geq 0} \tag{12.7}$$

(nach $\mathcal{F}$) messbar. Da $\{(-\infty; r] : r \in \mathbb{Q}\}$ ein schnittstabiler Erzeuger von $\mathcal{B}(\mathbb{R})$ ist, ist nach Bemerkung 12.21 die Abbildung $\omega \mapsto \kappa(\omega, B)$ für alle $B \in \mathcal{F}$ messbar. Also ist κ ein stochastischer Kern.

Es bleibt zu zeigen, dass κ eine reguläre Version der bedingten Verteilung ist. Da (12.7) auf einem schnittstabilen Erzeuger von $\mathcal{E}$ gilt, gilt für $\omega \in A$

$$\kappa(\omega, B) = \mathbf{P}(Y \in B|\mathcal{G})(\omega).$$

Mit anderen Worten, κ ist eine reguläre Version der bedingten Verteilung. □

13 Ausblick: Finanzmathematik

Als Anwendung der Wahrscheinlichkeitstheorie etabliert sich seit ein paar Jahrzehnten die Finanzmathematik. Aus diesem Bereich werden wir in diesem abschließenden Abschnitt das Binomialmodell der Optionspreisbewertung kennen lernen.

Bemerkung 13.1 (Finanzderivate). Auf Finanzmärkten gibt es verschiedene Arten von Wertpapieren. Die einfachsten hiervon sind etwa festverzinsliche Wertpapiere oder Aktien. Derivate sind Wertpapiere, die direkt an den Preis eines anderen, zu Grunde liegenden Wertpapieres, gekoppelt sind. Betrachten wir als Beispiel eine *europäische Call-Option*. Das zugrunde liegende Wertpapier ist eine bestimmte Aktie. Der Besitzer der europäischen Option hat das Recht, aber nicht die Pflicht, die Aktie zu einer Zeit τ zu einem Preis K zu erwerben. Die Frage, die hier beantwortet werden soll ist:

> Was ist der faire Preis einer europäischen Call-Option mit Parametern τ und K?

Klar ist, dass der faire Preis vom heutigen Aktienkurs sowie von der vermutlichen Wertentwicklung der Aktie abhängt. Die Wahrscheinlichkeitstheorie stellt für den Preis der Aktie einen geeigneten Rahmen zur Verfügung, da man den Aktienpreis der Zukunft als Zufallsvariable betrachten kann.

Die Anzahl verschiedener Finanzderivate wächst ständig. Bei Put-Optionen besteht das Recht des Besitzers nicht darin, die Aktie zu kaufen, sondern sie zu verkaufen. Bei einem Future besteht nicht die Recht auf Erwerb/Verkauf der Aktie, sondern die Pflicht. Amerikanische Optionen erlauben dem Besitzer, schon vor der Zeit τ das Recht auf Erwerb/Verkauf der Aktie auszuüben. Bei all diesen Fragen stellt sich die Frage nach dem fairen Preis.

Bemerkung 13.2 (Finanzmarktmodell). Neben der Aktie und dem Finanzderivat wird hier angenommen, dass es auch ein festverzinsliches Wertpapier gibt. Für 1 Geldeinheit (GE) in Zeit $t-1$ bekommt man $1+r$ GEen zur Zeit $1+r$, d.h. der Zinssatz ist r. Es wird stets angenommen, dass jederzeit zu diesem Zinssatz Geld geliehen oder angelegt werden kann.

Geht man von einem Händler in einem solchen Finanzmarkt aus, so hat er jederzeit die Wahl zwischen folgenden Möglichkeiten:

- Geld zum Zinssatz von r leihen,
- Geld zum Zinssatz r anlegen,
- in die Aktie investieren,
- Aktien verkaufen.

Zur letzten Möglichkeit sei bemerkt, dass auch Leerverkäufe getätigt werden können. Das heißt, die Aktie wird zum heutigen Preis verkauft, obwohl man sie nicht besitzt, und wird erst zu einem späteren Zeitpunkt geliefert.

Folgende Definition stellt die Grundlage für ein einfaches Aktienpreismodell dar. Natürlich macht das Modell viele Vereinfachungen, aber das haben mathematische Modelle so an sich.

Definition 13.3 (Binomialmodell des Aktienpreisprozesses). *Seien $M_1, M_2, \ldots$ unabhängig und identisch verteilt mit $\mathbf{P}(M_1 = u) = p, \mathbf{P}(M_1 = d) = q$ mit $0 < d < u$ und $p + q = 1$. Für $S_0 \in \mathbb{R}_+$ definieren wir den* Aktienpreisprozess *$S_0, S_1, \ldots$ durch*

$$S_t = M_t \cdots M_1 \cdot S_0.$$

Weiter sei $0 < r < 1$ der Zinssatz eines festverzinslichen Wertpapieres.

Definition 13.4 (Portfolio und Handelsstrategien).

1. *Ein* Portfolio *besteht aus einer Anzahl von Aktien und einer Anzahl von festverzinslichen Wertpapieren.*
2. *Eine Handelsstrategie ist gegeben durch den Wert X_0 eines Anfangsportfolios und Funktionen $\Delta_0 = \Delta_0(S_0), \Delta_1 = \Delta_1(S_0, S_1), \ldots$, die die Anzahl der Aktien im Portfolio zu den Zeitpunkten $t = 0, 1, 2, \ldots$ beschreiben. Wir fordern, dass Δ_t messbar bezüglich $S_0, \ldots, S_t$ ist, $t = 1, 2, \ldots$*

   Das bedeutet: Zum Zeitpunkt $t = 0$ bezeichnet $\Delta_0 = \Delta_0(S_0)$ die Anzahl der Aktien des Startportfolios, außerdem ist $X_0 - \Delta_0 S_0$ die Anzahl des festverzinslichen Wertpapiers im Portfolio. Der Wert des Portfolios zur Zeit $t = 1$ ist dann

   $$X_1 = \Delta_0 S_1 + (X_0 - \Delta_0 S_0)(1 + r). \tag{13.1}$$

   Auf Grundlage des Aktienpreises S_1 wird festgelegt, wie viel des Wertes X_1 in Aktien investiert werden sollen (das sind dann $\Delta_1(S_0, S_1)$). Die Anzahl an festverzinslichen Wertpapieren ist dann $X_1 - \Delta_1 S_1$.

Bemerkung 13.5 (Annahmen). Im Binomialmodell werden einige Annahmen getroffen, die mehr oder weniger realistisch sind:

1. Sowohl die Aktie als auch das festverzinsliche Wertpapier sind beliebig teilbar. (Da es sich üblicherweise um eine Großzahl an gehandelter Papiere handelt, ist diese Annahme oft gerechtfertigt.)
2. Der Zinssatz des festverzinslichen Wertpapieres hängt nicht davon ab, ob der Händler sich Geld leiht oder Geld in das Wertpapier investiert. (Dies gilt für Privatanleger sicher nicht, für Banken annähernd schon.)
3. Der Kauf- und Verkaufspreis der Aktie sind gleich. (Dies ist in der Praxis nicht erfüllt, und kann zu großen Abweichungen des Modells führen.)
4. Gegeben den Aktienpreis S_t zur Zeit t, kann der Aktienpreis S_{t+1} zur Zeit $t + 1$ nur zwei Werte annehmen. (Dies ist in der Praxis sicher verletzt, macht die Sache aber mathematisch handhabbarer.)

Bemerkung 13.6 (Arbitrage). Grundlegend bei der Theorie der Finanzmärkte ist die Annahme, dass man nie Geld risikolos gewinnen kann. Die Begründung ist, dass es in echten Märkten zwar manchmal solche Situationen gibt, diese aber sofort durch geeigneten Handel wieder verschwinden. Für das Binomialmodell heißt das: Arbitragefreiheit herrscht genau dann, wenn für alle Werte von Δ_0 gilt, dass

$$\mathbf{P}(X_1 > 0 | X_0 = 0) < 1.$$

Lemma 13.7 (Keine Arbitrage im Binomialmodell). *Das Binomialmodell ist genau dann arbitrage-frei, wenn $d < 1 + r < u$.*

Beweis. Wir berechnen direkt

$$\mathbf{P}(X_1 > 0|X_0 = 0) = \mathbf{P}(\Delta_0 S_1/S_0 > (1+r)\Delta_0) = \mathbf{P}(\Delta_0 M_1 > (1+r)\Delta_0).$$

Da Δ_0 beliebig ist, ist die Arbitragefreiheit äquivalent zu

$$\mathbf{P}(M_1 > 1+r) < 1,$$
$$\mathbf{P}(M_1 < 1+r) < 1.$$

Und dies bedeutet wiederum $d < 1 + r < u$. □

Wir kommen nun zum nicht-trivialen Teil dieses Abschnittes. Ziel ist es, den fairen Preis der europäischen Call-Option für das arbitragefreie Binomial-Modell zu berechnen.

Definition 13.8 (Europäisches Finanzderivat). *Ein (europäisches) Finanzderivat wird bestimmt durch die Ausübungszeit τ und eine Funktion φ, so dass der Wert des Finanzderivats zur Zeit τ gerade $\varphi(S_\tau)$ beträgt. Im Fall der europäischen Call-Option ist $\varphi(S_\tau) = (S_\tau - K)^+$, denn: (i) Ist $S_\tau < K$, so ist die Option zum Zeitpunkt τ wertlos, weil man ja die Aktie billiger als zu K GEen erwerben kann. (ii) Ist $S_\tau > K$, so kann man die Option ausüben, die Aktie also zu K GEen erwerben und gleich wieder für S_τ GEen verkaufen. Der Erlös ist gerade $S_\tau - K$.*

Die Strategie, um den fairen Preis der Call-Option zu ermitteln, veranschaulichen wir an einem Beispiel im Fall $\tau = 1$.

Beispiel 13.9 (Hedging). Sei $u = 2, d = \frac{1}{2}, r = \frac{1}{4}$, sowie $S_0 = 4$ und $K = 5$. Das bedeutet, dass die europäische Option zur Zeit 1 entweder 3 GE wert ist, falls $M_1 = u$ (Wahrscheinlichkeit p), oder 0 GEen wert ist, falls $M_1 = l$ (Wahrscheinlichkeit q), also

$$\varphi(X_1) = \begin{cases} 3, & \text{falls } M_1 = u, \\ 0, & \text{falls } M_1 = d. \end{cases}$$

Um den fairen Preis der Option zu ermitteln, versuchen wir, mit Hilfe eines Portfolios aus Aktien und festverzinslichem Wertpapier dieselbe Auszahlungsfunktion herzustellen. Wir beginnen mit einem Portfolio aus 0.5 Aktien und -0.8 Einheiten des festverzinslichen Wertpapiers. Dieses Portfolio hat einen Wert von $X_0 = 0.5 \cdot 4 - 0.8 = 1.2$. Zur Zeit 1 hat dieses Portfolio einen Wert von

$$X_1 = 0.5 \cdot M_1 \cdot 4 - 0.8 \cdot (1+r) = \begin{cases} 3, & \text{falls } M_1 = u, \\ 0, & \text{falls } M_1 = d. \end{cases}$$

Dies ist dieselbe Auszahlungsfunktion wie die der europäischen Option. Da der Wert des Portfolios 1.2 GEen war, muss der faire Preis der europäischen Option also 1.2 GEen sein. Wäre der Preis etwa höher, könnte man eine solche Option verkaufen, und gleichzeitig die oben beschriebene Handelsstrategie wählen, wobei der Differenzbetrag noch in das festverzinsliche Wertpapier angelegt wird. Damit ist $X_0 = 0$, aber $X_1 > 0$ fast sicher, es würde sich also eine Arbitrage-Möglichkeit ergeben.

Man sagt, dass die angegebene Handelsstrategie die Option repliziert. Außerdem wird das Replizieren auch als Hedging bezeichnet.

Proposition 13.10 (Fairer Preis der europäischen Call-Option für $\tau = 1$). *Der faire Preis einer europäischen Call-Option, die zur Zeit* 1 *einen Wert von* $V_1 = \varphi(S_1) = (S_1 - K)^+$ *besitzt, ist im Binomialmodell*

$$V_0 = \frac{1}{1+r}\big(\widetilde{p} \cdot \varphi(uS_0) + \widetilde{q} \cdot \varphi(dS_0)\big),$$

wobei

$$\widetilde{p} = \frac{1+r-d}{u-d}, \qquad \widetilde{q} = \frac{u-1-r}{u-d}.$$

Ist außerdem $X_0 = V_0$ *mit*

$$\Delta_0 = \frac{\varphi(uS_0) - \varphi(dS_0)}{uS_0 - dS_0},$$

so ist $X_1 = \varphi(S_1)$ *fast sicher.*

Bemerkung 13.11 (Interpretation). Die Formel für den Optionspreis interpretiert man am besten so: Sei $\widetilde{\mathbf{P}}$ ein Wahrscheinlichkeitsmaß mit

$$\widetilde{\mathbf{P}}(M_1 = u) = \widetilde{p}, \qquad \widetilde{\mathbf{P}}(M_1 = d) = \widetilde{q}.$$

Dann gilt

$$(1+r)V_0 = \widetilde{\mathbf{E}}[\varphi(S_1)|S_0],$$

wobei $\widetilde{\mathbf{E}}$ der Erwartungswert bezüglich $\widetilde{\mathbf{P}}$ ist.

Beweis von Proposition 13.10. Wir gehen genau wie im Beispiel 13.9 vor. Wir nehmen ein Portfolio an, das aus Δ_0 Aktien besteht, und einen Wert von X_0 hat. Damit sind $X_0 - \Delta_0 S_0$ in das festverzinsliche Wertpapier investiert. Der Wert des Portfolios zur Zeit 1 ist nun

$$X_1 = \Delta_0 S_1 + (X_0 - \Delta_0 S_0)(1+r)$$

Damit $X_1 = \varphi(S_1)$ (fast sicher) gelten soll, muss also

$$\begin{aligned}\varphi(uS_0) &= \Delta_0 uS_0 + (X_0 - \Delta_0 S_0)(1+r),\\ \varphi(dS_0) &= \Delta_0 dS_0 + (X_0 - \Delta_0 S_0)(1+r).\end{aligned}$$

Löst man diese zwei Gleichungen nach X_0, Δ_0 auf, so erhält man

$$\begin{aligned}\Delta_0 &= \frac{\varphi(uS_0) - \varphi(dS_0)}{uS_0 - dS_0},\\ X_0 &= \frac{1}{1+r}\big(\widetilde{p} \cdot \varphi(uS_0) + \widetilde{q} \cdot \varphi(dS_0)\big).\end{aligned}$$ □

Iteriert man die letzte Proposition, bekommt man den fairen Preis der europäischen Call-Option für beliebiges τ.

Theorem 13.12 (Fairer Preis der europäischen Call-Option). *Der faire Preis zur Zeit* t *einer europäischen Call-Option, die zur Zeit* τ *einen Wert von* $V_\tau = \varphi(S_\tau) = (S_\tau - K)^+$ *besitzt, ist im Binomialmodell*

$$V_t = V_t(S_t) = \frac{1}{(1+r)^{\tau-t}} \sum_{s=0}^{\tau-t} \binom{\tau-t}{s} \widetilde{p}^s \widetilde{q}^{\tau-t-s} \varphi(u^s d^{\tau-t-s} S_t), \tag{13.2}$$

wobei

$$\widetilde{p} = \frac{1+r-d}{u-d}, \qquad \widetilde{q} = \frac{u-1-r}{u-d}.$$

Ist außerdem $X_t = V_t$

$$\Delta_s = \frac{V_{s+1}(uS_s) - V_{s+1}(dS_s)}{uS_s - dS_s},$$

so ist $X_s = V_s$ fast sicher für $s = t, \dots, \tau - 1$.

Bemerkung 13.13 (Preisberechnung mit bedingten Erwartungen). Der Preis der Option berechnet sich am einfachsten mittels bedingter Erwartungen. Wir haben ja schon in Bemerkung 13.11 gesehen, dass $V_0 = \widetilde{\mathbf{E}}[\varphi(S_1)|S_0]/(1+r)$ gilt. Damit ist wohl auch für den Fall $\tau > 1$

$$\begin{aligned} V_t &= \widetilde{\mathbf{E}}[\varphi(S_\tau)|S_t]/(1+r)^{\tau-t} \\ &= \widetilde{\mathbf{E}}[\widetilde{\mathbf{E}}[\cdots \widetilde{\mathbf{E}}[\varphi(S_\tau)|S_{\tau-1}]|S_{\tau-2}]\cdots|S_t]/(1+r)^{\tau-t} \\ &= \widetilde{\mathbf{E}}[\widetilde{\mathbf{E}}\cdots \widetilde{\mathbf{E}}[\widetilde{p}\varphi(uS_{\tau-1}) + \widetilde{q}\varphi(dS_{\tau-1})|S_{\tau-2}]\cdots|S_t]/(1+r)^{\tau-t-1} \\ &= \cdots = \sum_{s=0}^{\tau-t} \binom{\tau-t}{s} \widetilde{p}^s \widetilde{q}^{\tau-t-s} \varphi(u^s d^{\tau-t-s} S_t). \end{aligned}$$

Beweis von Theorem 13.12. Am einfachsten zeigt man das Theorem durch Induktion über $t = \tau-1, \tau-2, \dots, 0$. Für $t = \tau - 1$ ist die Aussage gerade die von Proposition 13.10. Gilt also das Theorem schon für ein t. Gegeben S_{t-1}, weiß man also schon, dass $V_t(uS_{t-1})$ und $V_t(dS_t)$ die fairen Preise der Option zur Zeit t sind, je nachdem ob $M_t = u$ oder $M_t = d$. Damit befindet man sich wieder in der Situation von Proposition 13.10, also

$$V_{t-1} = \frac{1}{1+r}\big(\widetilde{p} \cdot V_t(uS_{t-1}) + \widetilde{q} \cdot V_t(dS_{t-1})\big).$$

Diese Rekursion wird aber gerade von (13.2) gelöst. Genauso sieht man die Form für Δ_s ein. □

Teil III
Stochastische Prozesse

In der modernen Stochastik spielen stochastische Prozesse eine zentrale Rolle. Anwendungen finden sie sowohl in der Finanzmathematik, als auch der Biologie oder Physik. Stochastische Prozesse werden immer dann verwendet, wenn sich eine Größe – also zum Beispiel ein Aktienkurs, die Frequenz eines Allels in einer Population oder die Position eines kleinen Teilchens – zufällig in der Zeit verändert.

Hier soll wichtiges Rüstzeug bereit gestellt werden, um mit stochastischen Prozessen umzugehen. Weiter werden wichtige Beispiele behandelt, etwa der Poisson-Prozess oder die Brown'sche Bewegung. Letztere spielt außerdem bei der Konstruktion von stochastischen Integralen eine entscheidende Rolle.

14 Einführung

Stochastische Prozesse sind nichts weiteres als Familien von Zufallsvariablen. Wichtig ist es einzusehen, dass diese Familie mit der Zeit indiziert ist. Im Laufe der Zeit werden also immer mehr Zufallsvariablen realisiert.

Im Folgenden sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, (E, r) ein vollständiger und separabler metrischer Raum mit Borel'scher σ-Algebra $\mathcal{B}(E)$ und I eine geordnete Teilmenge von $\overline{\mathbb{R}}$, die wir auch Indexmenge nennen. Wir werden immer die beiden Fälle $I \subseteq \overline{\mathbb{Z}}$ oder $I \subseteq \overline{\mathbb{R}}$ überabzählbar unterscheiden. Man beachte bereits hier, dass eine überabzählbare Indexmenge, etwa $I = \mathbb{R}$, neue Fragen aufwirft, da bekanntermaßen Wahrscheinlichkeitsmaße nur mit abzählbaren vielen Ausnamemengen umgehen können.

14.1 Definition und Existenz

Zunächst kümmern wir uns um die elementare Frage danach, was ein stochastische Prozess überhaupt ist, und wie man ihn in eindeutiger Art und Weise definieren kann.

Definition 14.1 (Stochastischer Prozess). *1. Sei $\mathcal{X} = (X_t)_{t\in I}$ so, dass $X_t : \Omega \to E$ $\mathcal{F}/\mathcal{B}(E)$-messbar ist. Dann heißt $\mathcal{X}$ ein E-wertiger* (stochastischer) Prozess. *Für $\omega \in \Omega$ heißt die Abbildung $I \mapsto E$, gegeben durch $X(\omega) : t \mapsto X_t(\omega)$ ein* Pfad *von X.*

2. *Ist in 1. der Wahrscheinlichkeitsraum $\Omega = E^I$ und $X_t = \pi_t$ die Projektion, so heißt $\mathcal{X}$* kanonischer Prozess.

3. *Sei $0 < p < \infty$. Ein reellwertiger Prozess $\mathcal{X} = (X_t)_{t\in I}$ heißt p-fach integrierbar, falls $\mathbf{E}[|X_t|^p] < \infty$ für alle $t \in I$. Er heißt L^p-beschränkt, falls $\sup_{t\in I} \mathbf{E}[|X_t|^p] < \infty$.*

In den Abschnitten 14.2 und 14.3 werden wir zwei Beispiele stochastischer Prozesse kennen lernen. Insbesondere wird der Poisson-Prozess (siehe Abschnitt 14.2) der erste Prozess mit überabzählbarer Indexmenge $I = [0, \infty)$ werden.

Beispiel 14.2 (Summen unabhängiger Zufallsvariable und Markov-Ketten). Aus der Vorlesung *Wahrscheinlichkeitstheorie* sind schon einige stochastische Prozesse bekannt, auch wenn sie da nicht so genannt wurden.

1. Seien etwa $X_1, X_2, ...$ reellwertige, unabhängige, identisch verteilte Zufallsvariable. Dann ist $\mathcal{S} = (S_t)_{t=0,1,2,...}$ mit $S_0 = 0$ und
$$S_t = \sum_{i=1}^{t} X_i$$
für $t = 1, 2, ...$ ein reellwertiger, stochastischer Prozess mit Indexmenge $I = \{0, 1, 2, ...\}$. Ist insbesondere $\mathbf{P}(X_i = \pm 1) = 1/2$, so heißt $\mathcal{S}$ eine eindimensionale, einfache Irrfahrt; siehe Abbildung 14.1.

2. Sei $\kappa(.,.)$ ein stochastischer Kern (siehe Definition 6.9) von $(E, \mathcal{B}(E))$ auf $(E, \mathcal{B}(E))$. Weiter sei X_0 eine E-wertige Zufallsvariable und gebeben X_t habe X_{t+1} die Verteilung $\kappa(X_t, .)$, $t = 0, 1, 2, ...$ Dann heißt $(X_t)_{t=0,1,...}$ eine E-wertige *Markov-Kette*.

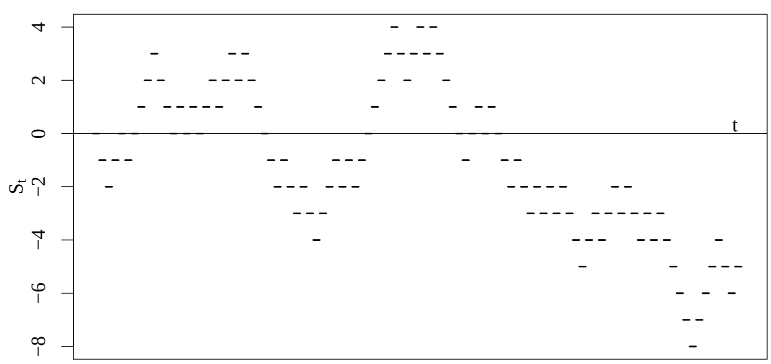


Abbildung 14.1: Ein Pfad einer eindimensionalen Irrfahrt.

Bemerkung 14.3 (Wiederholung: Existenz von stochastischen Prozessen).

1. Zur Wiederholung: die Produkt-σ-Algebra auf dem Raum E^I ist definiert als die kleinste σ-Algebra, bzgl. der alle Projektionen $\pi_t, t \in I$ messbar sind. Insbesondere ist für einen E-wertigen stochastischen Prozess $\mathcal{X} = (X_t)_{t\in I}$ die Abbildung $\omega \mapsto X(\omega)$ nach $\mathcal{F}/\mathcal{B}(E)^I$-messbar. Weiter ist eine projektive Familie auf $\mathcal{F}$ eine Familie von Verteilungen $(\mathbf{P}_J)_{J \Subset I}$ mit $\mathbf{P}_H = (\pi_H^J)_* \mathbf{P}_J$ für $H \subseteq J$, wobei π_H^J die Projektion von E^J auf E^H ist.

2. Oftmals kann man die endlich-dimensionalen Verteilungen eines stochastischen Prozesses $\mathcal{X} = (X_t)_{t\in I}$ definieren, also die gemeinsame Verteilung von $(X_{t_1}, ..., X_{t_n})$ für beliebige $t_1, ..., t_n \in I$. Beispielsweise werden wir in Abschnitten 14.2 und 14.3 den Poisson-Prozess und die Brown'sche Bewegung durch die Angabe der gemeinsamen Verteilung von $(X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}})$ definieren. Damit sind auch die endlich-dimensionalen Verteilungen festgelegt. Um sicher zu stellen, dass es zu diesen endlich-dimensionalen Verteilungen einen stochastischen Prozess gibt, benötigt man den Satz von Kolmgorov; siehe Theorem 6.24. Es ist zu beachten, dass endlich dimensionale Verteilungen von stochastischen Prozessen immer projektiv sind; siehe auch Beispiel 6.22.2.

Definition 14.4 (Gleichheit von stochastischen Prozessen). *Seien* $\mathcal{X} = (X_t)_{t \in I}$ *und* $\mathcal{Y} = (Y_t)_{t \in I}$ *zwei E-wertige stochastische Prozesse.*

1. *Gilt* $\mathcal{X} \stackrel{d}{=} \mathcal{Y}$, *so heißt* $\mathcal{Y}$ *eine* Version *von* $\mathcal{X}$ *(und umgekehrt).*
2. *Sind* $\mathcal{X}$ *und* $\mathcal{Y}$ *auf demselben Wahrscheinlichkeitsraum definiert und gilt* $\mathbf{P}(X_t = Y_t) = 1$ *für alle* $t \in I$, *so heißt* $\mathcal{X}$ *eine* Modifikation *von* $\mathcal{Y}$ *(und umgekehrt).*
3. *Sind* $\mathcal{X}$ *und* $\mathcal{Y}$ *auf demselben Wahrscheinlichkeitsraum definiert und gilt* $\mathbf{P}(X_t = Y_t \text{ für alle } t \in I) = 1$, *so heißen* $\mathcal{X}$ *und* $\mathcal{Y}$ ununterscheidbar.

Die Pfade $t \mapsto X_t(\omega)$ eines stochastischen Prozesses können bestimmte Eigenschaften besitzen. Beispielsweise können es stetige Funktionen $I \to E$ sein. Neben Prozessen mit stetigen Pfaden werden wir Prozesse mit rechtsstetigen Pfaden und linksseitigen Grenzwerten benötigen.

Definition 14.5 (Rechtsstetige Funktionen, linksseitige Grenzwerte). *Eine Funktion* $f : I \to E$ *heißt in* $t \in I$ *rechtsstetig mit linksseitigen Grenzwert*[18] *falls*

$$f(t) = \lim_{s \downarrow t} f(s) \text{ und } \lim_{s \uparrow t} f(s) \text{ existiert.}$$

Sie heißt rechtsstetig mit linksseitigen Grenzwerten, falls diese Eigenschaft für alle $t \in I$ *gilt. Die Menge der rechtsstetigen Funktionen mit linksseitigen Grenzwerten bezeichnen wir mit* $\mathcal{D}_E(I)$.

Proposition 14.6 (Versionen, Modifikationen, ununterscheidbare Prozesse). *Seien* $\mathcal{X} = (X_t)_{t \in I}$ *und* $\mathcal{Y} = (Y_t)_{t \in I}$ *stochastische Prozesse mit Werten in E.*

1. *Der Prozess* $\mathcal{Y}$ *ist genau dann eine Version von* $\mathcal{X}$ *(und umgekehrt), wenn beide Prozesse dieselben endlich-dimensionalen Verteilungen besitzen, d.h.* $(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (Y_{t_1}, ..., Y_{t_n})$ *für jede Wahl von* $n \in \mathbb{N}$ *und* $t_1, ..., t_n \in I$.
2. *Sind* $\mathcal{X}$ *und* $\mathcal{Y}$ *ununterscheidbar, so ist* $\mathcal{X}$ *eine Modifikation von* $\mathcal{Y}$ *(oder umgekehrt). Ist* $\mathcal{X}$ *eine Modifikation von* $\mathcal{Y}$, *so ist* $\mathcal{X}$ *eine Version von* $\mathcal{Y}$.
3. *Ist* I *höchstens abzählbar und ist* $\mathcal{X}$ *eine Modifikation von* $\mathcal{Y}$ *(oder umgekehrt), dann sind* $\mathcal{X}$ *und* $\mathcal{Y}$ *ununterscheidbar.*
4. *Ist* $I = [0, \infty)$ *und haben* $\mathcal{X}$ *und* $\mathcal{Y}$ *fast sicher rechtsstetige Pfade und ist* $\mathcal{X}$ *eine Modifikation von* $\mathcal{Y}$, *so sind* $\mathcal{X}$ *und* $\mathcal{Y}$ *ununterscheidbar.*

Beweis. 1. '$\Rightarrow$': klar. '$\Leftarrow$': Seien $(\Omega, \mathcal{F}, \mathbf{P})$ und $(\Omega', \mathcal{F}', \mathbf{P}')$ die Wahrscheinlichkeitsräume, auf denen $\mathcal{X}$ und $\mathcal{Y}$ definiert sind. Wir betrachten den durchschnittstabilen Erzeuger

$$\mathcal{C} := \{\pi_J^{-1}(A) : A \in \mathcal{B}(E)^{|J|}; J \Subset I\} \subseteq \mathcal{B}(E)^I$$

von $\mathcal{B}(E)^I$. Weiter ist für $J \Subset I$, $A \in \mathcal{B}(E)^{|J|}$,

$$\mathbf{P}((X_t)_{t \in J} \in A) = \mathbf{P}'_I((Y_t)_{t \in J} \in A),$$

[18] Solche Funktionen bezeichnet man auch als rcll (right-continuous with left limits) oder càdlàg (continue à droite, limite à gauche).

d.h. $\mathcal{X}_*\mathbf{P}$ und $\mathcal{Y}_*\mathbf{P}'$ stimmen auf $\mathcal{C}$ überein. Nach Theorem 3.10 ist damit $\mathcal{X}_*\mathbf{P} = \mathcal{Y}_*\mathbf{P}'$. Also ist $\mathcal{Y}$ eine Version von $\mathcal{X}$.
2. Sei $t \in I$. Sind $\mathcal{X}$ und $\mathcal{Y}$ ununterscheidbar, dann gilt $\mathbf{P}(X_t \neq Y_t) \leq \mathbf{P}(X_s \neq Y_s \text{ für ein } s \in I) = 0$. Sind $\mathcal{X}$ und $\mathcal{Y}$ Modifikationen und $t_1, ..., t_n \in I$, so ist

$$\mathbf{P}(X_{t_1} = Y_{t_1}, ..., X_{t_n} = Y_{t_n}) = 1.$$

Insbesondere haben $\mathcal{X}$ und $\mathcal{Y}$ dieselben endlich-dimensionalen Verteilungen. Nach 1. ist also $\mathcal{Y}$ eine Version von $\mathcal{X}$.
3. Die Aussage ist klar wegen der σ-Additivität von Wahrscheinlichkeitsmaßen,

$$\mathbf{P}(X_t \neq Y_t \text{ für ein } t \in I) = \sum_{t \in I} \mathbf{P}(X_t \neq Y_t) = 0.$$

4. Sei R eine Menge mit $\mathbf{P}(R) = 1$, so dass $\mathcal{X}$ und $\mathcal{Y}$ auf R rechtsstetige Pfade besitzen, sowie $N_t := \{X_t \neq Y_t\}$. Weiter sei $I' = I \cap \mathbb{Q}$. Dann ist $\mathbf{P}(\bigcup_{t \in I'} N_t) = 0$ sowie

$$\mathbf{P}\Big(\bigcup_{t \in I} N_t\Big) \leq \mathbf{P}\Big(R \cap \bigcup_{t \in I} \bigcup_{r \geq t, r \in I'} N_r\Big) = \mathbf{P}\Big(R \cap \bigcup_{r \in I'} N_r\Big) = 0.$$

□

Bemerkung 14.7 (Versionen mit verschiedenen Pfadeigenschaften). Sei $\mathcal{X} = (X_t)_{t \in I}$ ein E-wertiger stochastischer Prozess und $I = [0, \infty)$. Jeder Pfad $t \mapsto X_t(\omega)$ ist also eine Abbildung $I \to E$. Man unterscheidet stochastische Prozesse nach ihren Pfadeigenschaften. Ist zum Beispiel $t \mapsto X_t(\omega)$ für fast alle ω eine stetige Funktion, so sagt man, $\mathcal{X}$ habe (fast sicher) stetige Pfade. Es ist wichtig einzusehen, dass die Eigenschaft des Prozesses, stetige Pfade zu haben, nicht an dessen Verteilung ablesbar ist:

Sei $\mathcal{Y} = (Y_t)_{t \in I}$ mit $Y_t = 0$, sowie $T \sim \exp(1)$ und $\mathcal{X} = (X_t)_{t \in I}$ gegeben durch

$$X_t = \begin{cases} 1, & t = T, \\ 0, & \text{sonst.} \end{cases}$$

Dann ist $\mathbf{P}(X_t = Y_t) = \mathbf{P}(T \neq t) = 1$ für jedes $t \in I$. Also ist $\mathcal{X}$ eine Modifikation von $\mathcal{Y}$, insbesondere stimmen nach der letzten Proposition die Verteilungen von $\mathcal{X}$ und $\mathcal{Y}$ überein. Allerdings hat nur $\mathcal{Y}$ stetige Pfade, aber jeder Pfad von $\mathcal{X}$ ist unstetig. Insbesondere sind $\mathcal{X}$ und $\mathcal{Y}$ nicht ununterscheidbar.

Theorem 14.8 (Stetige Modifikationen; Kolmogorov, Chentsov). *Sei $\mathcal{X} = (X_t)_{t \in I}$ ein E-wertiger stochastischer Prozess mit $I = \mathbb{R}$ oder $I = [0, \infty)$. Für jedes $\tau > 0$ gebe es Zahlen $\alpha, \beta, C > 0$ mit*

$$\mathbf{E}[r(X_s, X_t)^\alpha] \leq C|t - s|^{1+\beta}$$

für alle $0 \leq s, t \leq \tau$. Dann existiert eine Modifikation $\widetilde{\mathcal{X}} = (\widetilde{X}_t)_{t \in I}$ von $\mathcal{X}$ mit stetigen Pfaden. Die Pfade sind sogar fast sicher lokal Hölder-stetig von jeder Ordnung $\gamma \in (0, \beta/\alpha)$.[19]

[19]Zur Erinnerung: eine Funktion $f : I \to E$ heißt lokal Hölder-stetig von der Ordnung γ, wenn es zu jedem $\tau > 0$ ein C gibt mit $r(f(s), f(t)) \leq C|t - s|^\gamma$ für alle $0 \leq s, t \leq \tau$.

Beweis. Es genügt, die Aussage für $I = [0,1]$ zu zeigen. Der allgemeine Fall folgt durch Aufteilung von I in abzählbar viele Intervalle der Länge 1. Wir betrachten die Menge der Zeitpunkte

$$D_n := \{0,1,...,2^n\} \cdot 2^{-n}$$

für $n = 0,1,...$, $D = \bigcup_{n=0}^\infty D_n$ und die Zufallsvariable

$$\xi_n := \max\{r(X_s,X_t) : s,t \in D_n, |t-s| = 2^{-n}\}.$$

Sei $0 < \gamma < \beta/\alpha$. Dann gilt für ein $C > 0$

$$\begin{aligned}\mathbf{E}\Big[\sum_{n=0}^\infty (2^{\gamma n}\xi_n)^\alpha\Big] &= \sum_{n=0}^\infty 2^{\alpha\gamma n}\mathbf{E}[\xi_n^\alpha] \leq \sum_{n=0}^\infty 2^{\alpha\gamma n} \sum_{s,t\in D_n, |t-s|=2^{-n}} \mathbf{E}[r(X_s,X_t)^\alpha] \\ &\leq C\sum_{n=0}^\infty 2^{\alpha\gamma n}2^n 2^{-n(1+\beta)} = C\sum_{n=0}^\infty 2^{(\alpha\gamma-\beta)n} < \infty.\end{aligned} \tag{14.1}$$

Deshalb gibt es eine Zufallsvariable C' mit $\xi_n \leq C' 2^{-\gamma n}$ für alle $n = 0,1,...$ Sei nun $m \in \{0,1,...\}$ und $r \in [2^{-m-1}, 2^{-m}] \cap D$. Dann gilt

$$\begin{aligned}\sup\big\{r(X_s,X_t) : s,t \in D, |s-t| \leq r\big\} &= \sup_{n\geq m}\big\{r(X_s,X_t) : s,t \in D_n, |s-t| \leq r\big\} \\ &\leq 2\sum_{n\geq m}\xi_n \leq 2C'\sum_{n\geq m} 2^{-\gamma n} \leq C'' 2^{-\gamma(m-1)} \leq C'' r^\gamma.\end{aligned} \tag{14.2}$$

für eine Zufallsvariable C''. Daraus folgt, dass fast jeder Pfad auf D Hölder-stetig zum Parameter γ ist. Damit lässt sich $\mathcal{X}$ Hölder-stetig auf I fortsetzen. Diese Fortsetzung nennen wir $\mathcal{Y} = (Y_t)_{t\in I}$. Um zu zeigen, dass $\mathcal{Y}$ eine Modifikation von $\mathcal{X}$ ist, betrachten wir ein $t \in I$ und eine Folge $t_1, t_2, ... \in D$ mit $t_n \to t$ mit $n \to \infty$. Wegen der Voraussetzung folgt $\mathbf{P}(r(X_{t_n}, X_t) > \varepsilon) \leq \mathbf{E}[r(X_{t_n}, X_t)^\alpha]/\varepsilon^\alpha \xrightarrow{n\to\infty} 0$ für jedes $\varepsilon > 0$, also $X_{t_n} \xrightarrow{n\to\infty}_p X_t$. Außerdem gilt wegen der Stetigkeit von $\mathcal{Y}$, dass $Y_{t_n} \xrightarrow{n\to\infty}_{fs} Y_t$. Insbesondere gilt also $\mathbf{P}(X_t = Y_t) = 1$. Damit ist der Beweis vollständig. □

14.2 Beispiel 1: Der Poisson-Prozess

Wir betrachten zum ersten Mal einen interessanten stochastischen Prozess mit Indexmenge $I = [0,\infty)$. Ein Pfad des Poisson-Prozesses ist in Abbildung 14.2 dargestellt.

Bemerkung 14.9 (Modellierung durch einen Poisson-Prozess). Unter Beobachtung stehen die Anzahl von Clicks eines Geigerzählers, Anrufe in einer Telefonzentrale, Mutationsereignissen entlang von Ahnenlinien... Solche Zähl-Vorgänge wollen wir mit Hilfe eines stochastischen Prozesses $\mathcal{X} = (X_t)_{t\in I}$ mit $I = [0,\infty)$ modellieren. Dabei sei X_t die Anzahl der Clicks/Anrufe/Mutationen bis zur Zeit t. An einen solchen Prozess stellt man sinnvollerweise ein paar Annahmen:

1. *Unabhängige Zuwächse:* Ist $0 = t_0 < t_1 < ... < t_n$, so ist $(X_{t_i} - X_{t_{i-1}} : i = 1,...,n)$ eine unabhängige Familie.

2. *Identisch verteilte Zuwächse:* Ist $0 < t_1 < t_2$, so ist $X_{t_2} - X_{t_1} \stackrel{d}{=} X_{t_2-t_1} - X_0$.

3. *Keine Doppelpunkte:* Es ist $\limsup_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbf{P}(X_\varepsilon - X_0 > 1) = 0$

Definition 14.10 (Poisson-Prozess). *Ein reellwertiger stochastischer Prozess* $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ *mit* $X_0 = 0$ *heißt Poisson-Prozess mit Intensität* λ*, wenn folgendes gilt:*

1. *Für* $0 = t_0 < ... < t_n$ *ist die Familie* $(X_{t_i} - X_{t_{i-1}} : i = 1, ..., n)$ *unabhängig.*
2. *Für* $0 \leq t_1 < t_2$ *ist* $X_{t_2} - X_{t_1} \sim$ *Poi*$(\lambda(t_2 - t_1))$.

Proposition 14.11 (Existenz von Poisson-Prozessen). *Sei* $\lambda \geq 0$*. Dann gibt es genau eine Verteilung* $\mathbf{P}_I$ *auf* $(\mathcal{B}(\mathbb{R}))^I$*, so dass der bzgl.* $\mathbf{P}_I$ *kanonische Prozess ein Poisson-Prozess mit Intensität* λ *ist.*

Beweis. Zunächst zur Eindeutigkeit: Die endlich-dimensionalen Randverteilungen von $\mathbf{P}_I$ sind durch 1. und 2. aus Definition 14.10 eindeutig festgelegt. Deswegen folgt die Eindeutigkeit aus Proposition 14.6.1.

Zur Existenz definieren wir den Poisson-Prozess als projektiven Limes. Für $J = \{t_1, ..., t_n\} \Subset I$ mit $0 = t_0 < t_1 < ... < t_n$ setzen wir für $x_0 = 0$

$$S^n : (x_1 - x_0, ..., x_n - x_{n-1}) \mapsto (x_1, ..., x_n).$$

Weiter ist

$$\mathbf{P}_J := S^n_* \bigotimes_{i=1}^n \mathrm{Poi}(\lambda(t_i - t_{i-1})). \tag{14.3}$$

Mit anderen Worten: Sind $Y_{t_i - t_{i-1}}$ für $i = 1, ..., n$ unabhängig Poisson-verteilt mit Parameter $\lambda(t_i - t_{i-1})$, dann ist $S^n(Y_{(t_1-t_0)}, ..., Y_{t_n-t_{n-1})}) \sim \mathbf{P}_J$.

Wir zeigen nun, dass die Familie $(\mathbf{P}_J : J \Subset I)$ projektiv ist: sei $J = \{t_1, ..., t_n\}$ wie oben und $H = J \setminus \{t_i\}$ für ein i. Dann ist

$$\mathrm{Poi}(\lambda(t_{i+1} - t_i)) * \mathrm{Poi}(\lambda(t_i - t_{i-1})) = \mathrm{Poi}(\lambda(t_{i+1} - t_{i-1}))$$

und damit

$$(\pi^J_H)_*\mathbf{P}_J = (\pi^J_H \circ S^n)_* \bigotimes_{j=1}^n \mathrm{Poi}(\lambda(t_j - t_{j-1})) = \mathbf{P}_H.$$

Nach Theorem 6.24 gibt es den projektiven limes $\mathbf{P}_I$. Betrachten wir den bzgl. $\mathbf{P}_I$ kanonischen Prozess $\mathcal{X} = (X_t)_{t\in I}$, so hat dieser die endlich-dimensionalen Verteilungen $(\mathbf{P}_J : J \Subset I)$. Insbesondere sind wegen (14.3) Zuwächse unabhängig und Poisson-verteilt. Damit erfüllt $\mathcal{X}$ die Bedingungen 1. und 2. aus Definition 14.10. □

Proposition 14.12 (Charakterisierung von Poisson-Prozessen). *Ein nicht-fallender Prozess* $\mathcal{X} = (X_t)_{t\in I}$ *mit* $X_0 = 0$ *und Werten in* $\mathbb{Z}_+$ *ist genau dann ein Poissonprozess mit Intensität* λ*, falls* $\lambda := \mathbf{E}[X_1 - X_0] < \infty$ *und 1.-3. aus Bemerkung 14.9 erfüllt sind.*

Beweis. '⇒': 1. und 2. aus Bemerkung 14.9 sind klar erfüllt. Für 3. berechnen wir direkt

$$\tfrac{1}{\varepsilon}\mathbf{P}(X_\varepsilon > 1) = \frac{1 - e^{-\lambda\varepsilon}(1 + \lambda\varepsilon)}{\varepsilon} \leq \frac{1 - (1 - \lambda\varepsilon)(1 + \lambda\varepsilon)}{\varepsilon} \xrightarrow{\varepsilon\to 0} 0.$$

'$\Leftarrow$': Es ist 1. aus Definition 14.10 erfüllt. Bleibt zu zeigen, dass $X_t \sim \text{Poi}(\lambda t)$ ist. Sei hierzu für $n \in \mathbb{N}, k = 1, ..., n,$

$$Z_k^n := (X_{tk/n} - X_{t(k-1)/n}) \wedge 1, \qquad X_t^n = \sum_{k=1}^n Z_k^n.$$

Das heißt, Z_k^n gibt an, ob im Intervall $(t(k-1)/n; tk/n]$ mindestens ein Sprung stattgefunden hat. Dann gilt, da X_t^n monoton in n ist,

$$\begin{aligned}
\mathbf{P}(\lim_{n\to\infty} X_t^n \neq X_t) &= \lim_{n\to\infty} \mathbf{P}(X_t^n \neq X_t) \\
&= \lim_{n\to\infty} \mathbf{P}(X_{tk/n} - X_{t(k-1)/n} > 1 \text{ für ein } k) \\
&\leq \lim_{n\to\infty} \sum_{k=1}^n \mathbf{P}(X_{tk/n} - X_{t(k-1)/n} > 1) \\
&= \lim_{n\to\infty} n\mathbf{P}(X_{t/n} > 1) \xrightarrow{n\to\infty} 0
\end{aligned}$$

nach 3. Weiter ist X_t^n binomialverteilt mit n und Erfolgswahrscheinlichkeit $p_n := \mathbf{P}(X_{t/n} > 0)$. Wegen der Linearität der Abbildung $t \mapsto \mathbf{E}[X_t]$ und, da $X_t^n \uparrow X_t$, folgt mit dem Satz über die monotone Konvergenz,

$$\lambda t = \mathbf{E}[X_t] = \lim_{n\to\infty} \mathbf{E}[X_t^n] = \lim_{n\to\infty} np_n.$$

Durch eine Poisson-Approximation (siehe Beispiel 11.1) gilt nun

$$\begin{aligned}
\mathbf{P}(X_t = k) &= \lim_{n\to\infty} \mathbf{P}(X_t^n = k) - \mathbf{P}(X_t^n = k; X_t \neq X_t^n) + \mathbf{P}(X_t = k; X_t \neq X_t^n) \\
&= \lim_{n\to\infty} \mathbf{P}(X_t^n = k) = \text{Poi}(\lambda t)(k),
\end{aligned}$$

d.h. $X_t \sim \text{Poi}(\lambda t)$ und die Behauptung folgt. □

Proposition 14.13 (Konstruktion mittels Exponentialverteilungen). *Seien $S_1, S_2, ...$ unabhängig, exponentialverteilt mit Parameter λ. Weiter sei $\mathcal{X} = (X_t)_{t\in I}$ gegeben durch*

$$X_t := \max\{i : S_1 + ... + S_i < t\}$$

mit $\max \emptyset = 0$. *Dann ist $\mathcal{X}$ ein Poisson-Prozess mit Intensität λ.*

Beweis. Wir müssen zeigen, dass für $0 = t_0 < ... < t_n$, $k_1, ..., k_n \in \mathbb{N}$

$$\mathbf{P}(X_{t_1} - X_{t_0} = k_1, ..., X_{t_n} - X_{t_{n-1}} = k_n) = \prod_{j=1}^n \text{Poi}(\lambda(t_j - t_{j-1}))(k_j)$$

gilt. Dies werden wir nur für den Fall $n = 2$ ausrechnen, der allgemeine Fall folgt analog. Sei in der folgenden Rechnung $0 \leq s < t$ und $U_1, U_2, ...$ uniform verteilte Zufallsvariablen auf

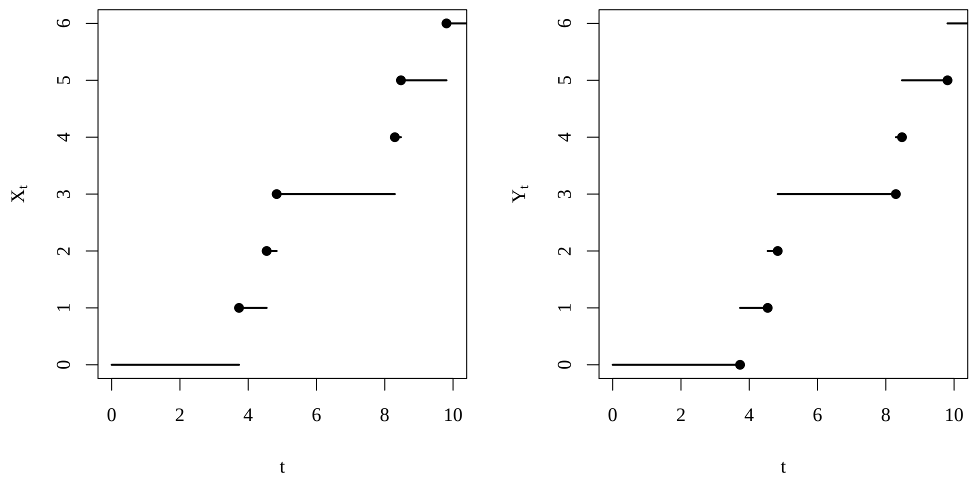


Abbildung 14.2: Der rechtsstetige Poisson-Prozess $\mathcal{X}$ und der linksstetige Poisson-Prozess $\mathcal{Y}$.

$[0, t]$. Wir berechnen

$$\begin{aligned}
&\mathbf{P}(X_s - X_0 = k, X_t - X_s = \ell) \\
&\quad = \int_0^s \int_{s_1}^s \cdots \int_{s_{k-1}}^s \int_s^t \int_{s_{k+1}}^t \cdots \int_{s_{k+\ell-1}}^t \int_t^\infty \lambda^{k+\ell+1} e^{-\lambda s_1} e^{-\lambda(s_2 - s_1)} \cdots \\
&\qquad\qquad \cdots e^{-\lambda(s_{k+\ell+1} - s_{k+\ell})} ds_{k+\ell+1} ... ds_1 \\
&\quad = \lambda^{k+\ell} \int_0^s \int_{s_1}^s \cdots \int_{s_{k-1}}^s \int_s^t \int_{s_{k+1}}^t \cdots \int_{s_{k+\ell-1}}^t \Big(\int_t^\infty \lambda e^{-\lambda s_{k+\ell+1}} ds_{k+\ell+1} \Big) ds_{k+\ell} \cdots ds_1 \\
&\quad = e^{-\lambda t} \lambda^{k+\ell} t^{k+\ell} \mathbf{P}[U_1 < ... < U_k < s < U_{k+1} < ... < U_{k+\ell}] \\
&\quad = e^{-\lambda t} \lambda^\ell t^\ell \binom{k+\ell}{k} \Big(\frac{s}{t}\Big)^k \Big(\frac{t-s}{t}\Big)^\ell \frac{1}{(k+\ell)!} \\
&\quad = e^{-\lambda s} \frac{(\lambda s)^k}{k!} \cdot e^{-\lambda(t-s)} \frac{(\lambda(t-s))^\ell}{\ell!},
\end{aligned}$$

woraus die Behauptung folgt. □

Beispiel 14.14 (Links- und rechtsstetiger Poisson-Prozess). Sei, ähnlich wie in Proposition 14.13, der stochastische Prozess $\mathcal{Y} = (Y_t)_{t \in I}$ gegeben durch

$$Y_t := \max\{i : S_1 + ... + S_i \leq t\}.$$

Pfade der Prozesse $\mathcal{X}$ aus Proposition 14.13 und $\mathcal{Y}$ sind in Abbildung 14.2 zu sehen. Die beiden Prozesse unterschieden sich dadurch, dass $\mathcal{X}$ rechtsstetig und $\mathcal{Y}$ linksstetig ist. Allerdings sind beide Prozesse Poisson-Prozesse mit Intensität λ, wie man sich leicht überzeugt. Es gilt nämlich $\mathbf{P}(X_t = Y_t) = 1$ für alle $t \in [0, \infty)$ und damit ist $\mathcal{Y}$ eine Version von $\mathcal{X}$ nach Proposition 14.6. Wie man an diesem Beispiel nochmals sieht, können zwei Prozesse mit derselben Verteilung völlig unterschiedliche Pfade besitzen.

14.3 Beispiel 2: Die Brown'sche Bewegung

Die Brown'sche Bewegung ist benannt nach dem Botaniker Robert Brown, der in einem Mikroskop beobachtete, wie sich Pollen unter thermischen Fluktuationen scheinbar erratisch bewegt. Wir geben für diesen gerade in der stochastischen Analysis besonders wichtigen Prozess eine mathematische Definition. Hier wird die Normalverteilung eine ausgezeichnete Rolle spielen. Einen Pfad einer ein-dimensionalen Brown'schen Bewegung findet man in Abbildung 14.3.

Dieser Abschnitt dient nur der Einführung der Brown'schen Bewegung. Weiteres über die Brown'sche Bewegung werden wir später lernen.

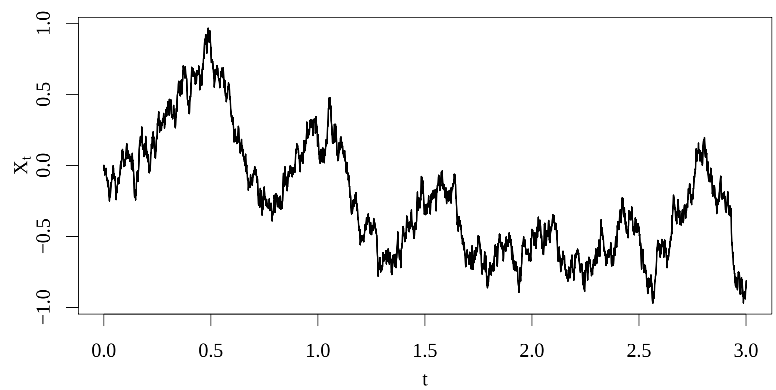


Abbildung 14.3: Ein Pfad einer Brown'schen Bewegung..

Definition 14.15 (Brown'sche Bewegung und Gauss'sche Prozesse).
Sei $\mathcal{X} = (X_t)_{t\in I}$ ein stochastischer Prozess mit Werten in $\mathbb{R}$.

1. *Der Prozess $\mathcal{X}$ heißt* Gauss'sch, *wenn $c_1 X_{t_1} + \cdots + c_n X_{t_n}$ für jede Wahl von $c_1, ..., c_n \in \mathbb{R}$ und $t_1, ..., t_n \in I$ normalverteilt ist. Für einen Gauss'schen Prozess bezeichnet $(s,t) \mapsto$ $\mathbf{COV}(X_s, X_t)$ dessen Covarianzstruktur.*

2. *Ist $I = [0,\infty)$, so heißt $\mathcal{X}$ eine* Brown'sche Bewegung *mit Start in x, falls der Prozess stetige Pfade hat und wenn für jede Wahl von $0 = t_0 < t_1 < \cdots < t_n$ gilt, dass $X_{t_0} = x$ und $X_{t_i} - X_{t_{i-1}}$ unabhängig davon nach $N(0, t_i - t_{i-1})$ verteilt sind, $i = 1, ..., n$. Gilt $x = 0$, so heißt $\mathcal{X}$ auch* standardisiert *oder auch* Wiener Prozess.

3. *Seien $\mathcal{X}^1 = (X_t^1)_{t\in[0,\infty)}, ..., \mathcal{X}^d = (X_t^d)_{t\in[0,\infty)}$ Brown'sche Bewegungen. Dann heißt der $\mathbb{R}^d$-wertige Prozess $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ mit $X_t = (X_t^1, ..., X_t^d)$ eine d-*dimensionale Brown'sche Bewegung.

Bemerkung 14.16 (Stetigkeit der Brown'schen Bewegung). Nach Theorem 6.24 ist klar, dass es einen Prozess gibt, dessen Zuwächse normalverteilt sind wie in Definition 14.15.2 angegeben. Es handelt sich bei der angegebenen Verteilungen $(X_{t_1}, ..., X_{t_n})_{n\in\mathbb{N}, t_1,...,t_n \in I}$ nämlich um eine projektive Familie. Ist etwa $X_{t_i} - X_{t_{i-1}} \sim N(0, t_i - t_{i-1})$ und $X_{t_{i-1}} - X_{t_{i-2}} \sim N(0, t_{i-1} - t_{i-2})$, so ist $X_{t_i} - X_{t_{i-2}} = X_{t_i} - X_{t_{i-1}} + X_{t_{i-1}} - X_{t_{i-2}} \sim N(0, t_i - t_{i-2})$ wegen Beispiel 6.20. Weniger klar ist jedoch, ob es auch einen Prozess mit solchen Zuwächsten gibt, der stetige Pfade hat. Um dies zu überprüfen, verwenden wir das Kriterium aus Theorem 14.8.

Proposition 14.17 (Existenz der Brown'schen Bewegung). *Sei $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ ein reellwertiger stochastischer Prozess, so dass für jede Wahl von $0 = t_0 < t_1 < ... < t_n$ gilt, dass $X_{t_0} = x$ und $X_{t_i} - X_{t_{i-1}}$ unabhängig nach $N(0, t_i - t_{i-1})$ verteilt sind, $i = 1, ..., n$. Dann existiert eine Modifikation $\mathcal{Y}$ von $\mathcal{X}$ mit stetigen Pfaden. Mit anderen Worten ist $\mathcal{Y}$ eine Brown'sche Bewegung. Der Prozess $\mathcal{Y}$ ist sogar lokal Hölder-stetig zu jedem Parameter $\gamma < 1/2$. Weiterhin ist die Covarianzstruktur der Brown'schen Bewegung $\mathcal{Y}$ gegeben durch $\mathbf{COV}(X_s, X_t) = s \wedge t$.*

Beweis. ObdA sei $x = 0$. Die Existenz und Eindeutigkeit eines Prozesses mit unabhängigen normalverteilten Zuwächsen folgt wie im Beweis von Proposition 14.11. Da $X_s \sim N(0, s)$, ist $X_s \stackrel{d}{=} s^{1/2} X_1$, wie man zum Beispiel aus Beispiel 7.13.3 abliest. Für $a > 2$ gilt weiterhin

$$\mathbf{E}[|X_t - X_s|^a] = \mathbf{E}[|X_{t-s}|^a] = \mathbf{E}[((t-s)^{1/2}|X_1|)^a] = (t-s)^{a/2}\mathbf{E}[|X_1|^a].$$

Nach Theorem 14.8 gibt es also eine Modifikation von $\mathcal{X}$ mit stetigen Pfaden. Diese sind nach obiger Rechnung Hölder-stetig zu jedem Parameter $\gamma \in (0, ((a/2)-1)/a) = (0, (a-2)/(2a))$. Da a beliebig war, folgt die Hölder-Stetigkeit für jedes $\gamma \in (0, 1/2)$.

Um die Covarianzstruktur von $\mathcal{X}$ zu bestimmen, berechnen wir für $0 \leq s \leq t$

$$\mathbf{COV}(X_s, X_t) = \mathbf{COV}(X_s, X_s) + \mathbf{COV}(X_s, X_t - X_s) = \mathbf{V}[X_s] = s.$$

Eine analoge Rechnung für $t < s$ liefert das Ergebnis $\mathbf{COV}(X_s, X_t) = s \wedge t$. □

Lemma 14.18 (Charakterisierung von Gauss'schen Prozessen).
Seien $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ und $\mathcal{Y} = (Y_t)_{t\in[0,\infty)}$ Gauss'sche Prozesse mit $\mathbf{E}[X_t] = \mathbf{E}[Y_t]$ und $\mathbf{COV}(X_s, X_t) = \mathbf{COV}(Y_s, Y_t)$. Dann ist $\mathcal{Y}$ eine Version von $\mathcal{X}$ (und umgekehrt).

Beweis. Sei $n \in \mathbb{N}$ und $c_1, ..., c_n \in \mathbb{R}$ beliebig. Dann sind für jede Wahl von $t_1, ..., t_n \in I$ sowohl $Z_X := c_1 X_{t_1} + \cdots + c_n X_{t_n}$ als auch $Z_Y := c_1 Y_{t_1} + \cdots c_n Y_{t_n}$ normalverteilt. Außerdem gilt nach Voraussetzung

$$\mathbf{E}[Z_X] = c_1\mathbf{E}[X_{t_1}] + \cdots + c_n\mathbf{E}[X_{t_n}] = c_1\mathbf{E}[Y_{t_1}] + \cdots + c_n\mathbf{E}[Y_{t_n}] = \mathbf{E}[Z_Y]$$

sowie

$$\mathbf{V}(Z_X) = \sum_{i,j=1}^n c_i c_j \mathbf{COV}(X_{t_i}, X_{t_j}) = \sum_{i,j=1}^n c_i c_j \mathbf{COV}(Y_{t_i}, Y_{t_j}) = \mathbf{V}(Z_Y).$$

Damit gilt $Z_X \stackrel{d}{=} Z_Y$. Da $c_1, ..., c_n$ beliebig waren, folgt aus Proposition 11.17, dass $(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (Y_{t_1}, ..., Y_{t_n})$. Mit Proposition 14.6.1 folgt die Aussage. □

Theorem 14.19 (Brown'sche Skalierung). *Sei $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ eine Brown'sche Bewegung. Dann sind die Prozesse $(X_{c^2t}/c)_{t\in[0,\infty)}$ für jedes $c > 0$ und $(tX_{1/t})_{t\in[0,\infty)}$ ebenfalls Brown'sche Bewegungen.*

Beweis. Klar ist, dass sowohl $(X_{c^2t}/c)_{t\in[0,\infty)}$ als auch $(tX_{1/t})_{t\in[0,\infty)}$ Gauss'sche Prozesse sind. Weiter ist

$$\mathbf{E}[X_{c^2t}/c] = 0,$$
$$\mathbf{E}[tX_{1/t}] = 0,$$

und für $s, t \geq 0$

$$\mathbf{COV}[X_{c^2 s}/c, X_{c^2 t}/c] = \frac{1}{c^2}(c^2 s \wedge c^2 t) = s \wedge t,$$
$$\mathbf{COV}[sX_{1/s}, tX_{1/t}] = st\Big(\frac{1}{s} \wedge \frac{1}{t}\Big) = s \wedge t.$$

Nun folgt die Behauptung mit Lemma 14.18. □

14.4 Filtrationen und Stoppzeiten

Bei einem stochastischen Prozess werden im Laufe der Zeit immer mehr der zu grunde liegenden Zufallsvariablen realisiert. Das bedeutet, dass im Laufe der Zeit immer mehr Informationen über den Pfad des Prozesses sichtbar werden. Nun ist Information gleichbedeutend mit der Messbarkeit bezüglich einer σ-Algebra, wie aus der Vorlesung *Wahrscheinlichkeitstheorie* bekannt. Da die Information im Laufe der Zeit wächst, sind mit einem stochastischen Prozess eine aufsteigende Familie von σ-Algebren verbunden, die wir im Folgenden eine Filtration nennen wollen.

Definition 14.20 (Filtrationen, Stoppzeiten). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein E-weriger stochastischer Prozess, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$.*

1. *Eine Familie $(\mathcal{F}_t)_{t\in I}$ von σ-Algebren mit $\mathcal{F}_t \subseteq \mathcal{F}, t \in I$, heißt* Filtration, *falls $\mathcal{F}_s \subseteq \mathcal{F}_t$ für alle $s \leq t$.*
2. *Die Filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ mit $\mathcal{F}_t = \sigma(X_s : s \leq t)$ heißt die von $\mathcal{X}$ erzeugte Filtration.*
3. *Der stochastische Prozess $(X_t)_{t\in I}$ heißt an eine Filtration $(\mathcal{F}_t)_{t\in I}$* adaptiert, *falls X_t für alle $t \in I$ eine $\mathcal{F}_t/\mathcal{B}(E)$-messbare Zufallsvariable ist,*

Sei nun $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ eine Filtration.

5. *Eine zufällige Zeit ist eine Zufallsvariable mit Werten in $\bar{I}$ (dem Abschluss von I). Eine zufällige Zeit T heißt $((\mathcal{F}_t)_{t\in I}$-)*Stoppzeit *falls $\{T \leq t\} \in \mathcal{F}_t$ für alle $t \in I$. Falls $I = [0, \infty)$, so heißt eine zufällige Zeit T $((\mathcal{F}_t)_{t\in I}$-)*Optionszeit *falls $\{T < t\} \in \mathcal{F}_t$ für alle $t \in I$. (Im Fall $I = \{0, 1, 2, ...\}$ benötigen wir diesen Begriff nicht.)*
6. *Jede Stoppzeit T definiert die σ-Algebra*
$$\mathcal{F}_T := \{A \in \mathcal{A} : A \cap \{T \leq t\} \in \mathcal{F}_t, t \in I\}$$
der T-Vergangenheit.
7. *Für eine zufällige Zeit T ist X_T definiert durch $\omega \mapsto X_{T(\omega)}(\omega)$. Weiter ist $\mathcal{X}^T := (X_{T\wedge t})_{t\in I}$ der bei T gestoppte Prozess.*

Bemerkung 14.21 (Interpretation der Definition von Stoppzeiten). Sei $\mathcal{X} = (X_t)_{t\in I}$ ein stochastischer Prozess und $(\mathcal{F}_t)_{t\in I}$ die kanonische Filtration. Man kann $\mathcal{F}_t$ als die Information auffassen, die zum Zeitpunkt t durch Kenntnis von $(X_s)_{0\leq s\leq t}$ verfügbar ist. Ist nun T eine Stoppzeit, so gilt $\{T \leq t\} \in \mathcal{F}_t$. Also lässt sich das Eintreten des Ereignisses $\{T \leq t\}$ durch Kenntnis von $(X_s)_{s\leq t}$ entscheiden. Mit anderen Worten kann durch Kenntnis des stochastischen Prozesses bis zur Zeit t entschieden werden, ob die Stoppzeit T spätestens jetzt gerade abgelaufen ist. Ist T eine Optionszeit, so kann durch Kenntnis des stochastischen Prozesses bis zur Zeit t entschieden werden, ob die Stoppzeit T schon in der Vergangenheit von t abgelaufen ist.

Beispiel 14.22 (Treffzeiten im Poisson-Prozess). Seien $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ und $\mathcal{Y} = (Y_t)_{t\in[0,\infty)}$ der rechts- und linksstetige Poisson-Prozess aus Beispiel 14.14, sowie $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$ und $(\mathcal{F}_t^{\mathcal{Y}})_{t\in[0,\infty)}$ die von $\mathcal{X}$ und $\mathcal{Y}$ erzeugten Filtrationen. Weiter sei

$$T_1 := \inf\{t \geq 0 : X_t = 1\} = \inf\{t \geq 0 : Y_t = 1\}$$

die *Treffzeit* der 1. (Das letzte Gleichheitszeichen gilt, da die Prozesse $\mathcal{X}$ und $\mathcal{Y}$ zum selben Zeitpunkt von 0 nach 1 springen.) Dann gilt:

- T_1 ist sowohl $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$-Stoppzeit, als auch $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$-Optionszeit.

Denn: Ist $T_1 = t$ der Sprungzeitpunkt von 0 nach 1, so gilt $X_t = 1$, also $\{T_1 \leq t\} = \{X_t \geq 1\} \in \sigma((X_s)_{s\leq t}) = \mathcal{F}_t^{\mathcal{X}}$ und $\{T_1 < t\} = \{X_{t-} \geq 1\} \in \sigma((X_s)_{s<t}) \subseteq \mathcal{F}_t^{\mathcal{X}}$.

- T_1 ist zwar $(\mathcal{F}_t^{\mathcal{Y}})_{t\in[0,\infty)}$-Optionszeit, jedoch keine $(\mathcal{F}_t^{\mathcal{Y}})_{t\in[0,\infty)}$-Stoppzeit.

Denn: Ist $T_1 = t$ der Sprungzeitpunkt von 0 nach 1, so gilt $Y_t = 0$, jedoch $Y_{t+} = 1$, also $\{T_1 \leq t\} = \{X_{t+} \geq 1\} \in \sigma((Y_s)_{s\leq t+h})$ für jedes $h > 0$, jedoch nicht $\{T_1 \leq t\} \in \mathcal{F}_t^{\mathcal{Y}}$. Allerdings gilt weiterhin $\{T_1 < t\} = \{Y_t \geq 1\} \in \sigma((Y_s)_{s\leq t}) \subseteq \mathcal{F}_t^{\mathcal{Y}}$.

Lemma 14.23 (Einfache Eigenschaften von Stoppzeiten). *Sei $(\mathcal{F}_t)_{t\in I}$ eine Filtration.*

1. *Jede Zeit $T = s \in I$ ist eine Stoppzeit*
2. *Für Stoppzeiten S, T sind auch $S \wedge T, S \vee T$ Stoppzeiten.*
3. *Für Stoppzeiten $S, T \geq 0$ ist $S + T$ eine Stoppzeit.*
4. *Jede Stoppzeit T ist $\mathcal{F}_T$ messbar.*
5. *Für Stoppzeiten S, T mit $S \leq T$ ist $\mathcal{F}_S \subseteq \mathcal{F}_T$.*

Beweis. 1. Für $t \in I$ ist $\{s \leq t\} \in \{\emptyset, \Omega\} \subseteq \mathcal{F}_t$, d.h. $T = s$ ist eine Stoppzeit.
2. Für $t \in I$ ist $\{S \wedge T \leq t\} = \{S \leq t\} \cup \{T \leq t\} \in \mathcal{F}_t$ und $\{S \vee T \leq t\} = \{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t$.
3. Sei $t \in I$. Es sind $S \wedge t$ und $T \wedge t$ Stoppzeiten, d.h. für $s \leq t$ ist $\{S \wedge t \leq s\} \in \mathcal{F}_s \subseteq \mathcal{F}_t$. Für $s > t$ ist $\{S \wedge t \leq s\} = \Omega \in \mathcal{F}_t$, d.h. $S \wedge t$ ist $\mathcal{F}_t$-messbar. Analog folgt, dass $T \wedge t$ $\mathcal{F}_t$-messbar ist. Weiter ist $1_{\{S>t\}}, 1_{\{T>t\}}$ $\mathcal{F}_t$-messbar. Setzen wir $S' = S \wedge t + 1_{\{S>t\}}, T' = T \wedge t + 1_{\{T>t\}}$, so ist $S' + T'$ $\mathcal{F}_t$-messbar und es gilt $\{S + T \leq t\} = \{S' + T' \leq t\} \in \mathcal{F}_t$.
4. Da T eine Stoppzeit ist, ist $\{T \leq t\} \in \mathcal{F}_t$. Nach Definition von $\mathcal{F}_T$ bedeutet das $\{T \leq t\} \in \mathcal{F}_T$. Da $\mathcal{H} := \{(-\infty; t] : t \in \mathbb{R}\}$ ein Erzeuger von $\mathcal{B}(\mathbb{R})$ sind, folgt die Behauptung.
5. Sei $A \in \mathcal{F}_S$ und $t \in I$. Wegen $B := A \cap \{S \leq t\} \in \mathcal{F}_t$ ist

$$A \cap \{T \leq t\} = B \cap \{T \leq t\} \in \mathcal{F}_t,$$

d.h. $A \in \mathcal{F}_T$. □

Definition 14.24 (Rechtsstetige und vollständige Filtration).

1. *Sei $(\mathcal{F}_t)_{t\in[0,\infty)}$ eine Filtration. Wir definieren $(\mathcal{F}_t^+)_{t\in[0,\infty)}$ durch $\mathcal{F}_t^+ := \bigcap_{s>t} \mathcal{F}_s$. Weiter heißt $(\mathcal{F}_t)_{t\in[0,\infty)}$* rechtsstetig, *falls $\mathcal{F}_t^+ = \mathcal{F}_t$.*
2. *Sei $\mathcal{N} = \{A :$ es gibt ein $N \supseteq A$ mit $N \in \mathcal{F}$ und $\mathbf{P}(N) = 0\}$. Dann heißt die Filtration $(\mathcal{F}_t)_{t\in I}$* vollständig, *falls $\mathcal{N} \subseteq \mathcal{F}_t$ für jedes $t \in I$.*

Lemma 14.25 (Übliche Vervollständigung einer Filtration). *Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, $(\mathcal{F}_t)_{t\in[0,\infty)}$ eine Filtration und $\mathcal{N}$ wie in Definition 14.24. Dann gibt es eine kleinste rechtsstetige und vollständige Filtration $(\mathcal{G}_t)_{t\in[0,\infty)}$ mit $\mathcal{F}_t \subseteq \mathcal{G}_t, t \in [0,\infty)$. Diese ist durch*

$$\mathcal{G}_t = \sigma(\mathcal{F}_t^+, \mathcal{N})$$

gegeben. Außerdem gilt $\sigma(\mathcal{F}_t^+, \mathcal{N}) = \sigma(\mathcal{F}_t, \mathcal{N})^+$.

Beweis. Zunächst zeigen wir die letzte Gleichung. Klar ist, dass

$$\sigma(\mathcal{F}_t^+, \mathcal{N}) \subseteq \sigma(\sigma(\mathcal{F}_t, \mathcal{N})^+, \mathcal{N}) = \sigma(\mathcal{F}_t, \mathcal{N})^+.$$

Sei andersherum $A \in \sigma(\mathcal{F}_t, \mathcal{N})^+$. Dann ist $A \in \sigma(\mathcal{F}_{t+h}, \mathcal{N})$ für alle $h > 0$. Also gibt es ein $A_h \in \mathcal{F}_{t+h}$ mit $\mathbf{P}((A \setminus A_h) \cup (A_h \setminus A)) = 0$. Wähle nun $h_1, h_2, ...$ mit $h_n \downarrow 0$ und

$$A' = \{A_{h_n} \text{ unendlich oft}\}.$$

Dann gilt offenbar $A' \in \mathcal{F}_t^+$ sowie $\mathbf{P}((A \setminus A') \cup (A' \setminus A)) = 0$, also $A \in \sigma(\mathcal{F}_t^+, \mathcal{N})$. Daraus folgt $\sigma(\mathcal{F}_t, \mathcal{N})^+ \subseteq \sigma(\mathcal{F}_t^+, \mathcal{N})$.

Zum Beweis der Minimalität von $(\mathcal{G}_t)_{t\in[0,\infty)}$ sei $(\mathcal{H}_t)_{t\in[0,\infty)}$ eine weitere rechtsstetige vollständige Erweiterung von $(\mathcal{F}_t)_{t\in[0,\infty)}$. Dann gilt

$$\mathcal{G}_t = \sigma(\mathcal{F}_t^+, \mathcal{N}) \subseteq \sigma(\mathcal{H}_t, \mathcal{N}) = \mathcal{H}_t$$

für alle $t \in [0, \infty)$. □

Lemma 14.26 (Options- und Stoppzeiten). *Sei $(\mathcal{F}_t)_{t\in[0,\infty)}$ eine Filtration. Eine zufällige Zeit T ist genau dann eine $(\mathcal{F}_t)_{t\in[0,\infty)}$-Optionszeit, wenn T eine $(\mathcal{F}_t^+)_{t\in[0,\infty)}$-Stoppzeit ist. Dann gilt*

$$\mathcal{F}_T^+ = \{A \in F : A \cap \{T < t\} \in \mathcal{F}_t, t > 0\}.$$

Ist insbesondere $(\mathcal{F}_t)_{t\in[0,\infty)}$ rechsstetig, so ist jede zufällige Zeit genau dann eine $(\mathcal{F}_t)_{t\in[0,\infty)}$-Stoppzeit, wenn sie ein $(\mathcal{F}_t)_{t\in[0,\infty)}$-Optionszeit ist.

Beweis. Zunächst ist

$$\{T \le t\} = \bigcap_{\mathbb{Q}\ni s>t} \{T < s\}, \qquad \{T < t\} = \bigcup_{\mathbb{Q}\ni s<t} \{T \le s\}.$$

Ist nun T eine $(\mathcal{F}_t^+)_{t\in[0,\infty)}$-Stoppzeit und $A \cap \{T \le t\} \in \mathcal{F}_t^+$. Dann gilt

$$A \cap \{T < t\} = \bigcup_{\mathbb{Q}\ni s<t} (A \cap \{T \le s\}) \in \mathcal{F}_t.$$

Ist andererseits $A \cap \{T < t\} \in \mathcal{F}_t$, dann gilt

$$A \cap \{T \le t\} = \bigcap_{h>0} \bigcap_{t<s<t+h} (A \cap \{T < s\}) \in \bigcap_{h>0} \mathcal{F}_{t+h} = \mathcal{F}_t^+.$$

Setzt man $A = \Omega$ in den letzten beiden Gleichungen, so folgt die erste Behauptung. Für allgemeines A folgt auch die zweite. □

Lemma 14.27 (Suprema und Infima von Stoppzeiten). *Seien $T_1, T_2, ...$ zufällige Zeiten und $(\mathcal{F}_t)_{t\in I}$ eine Filtation. Dann gilt:*

1. *Sind $T_1, T_2, ...$ Stoppzeiten, dann ist auch $T := \sup_n T_n$ eine Stoppzeit.*
2. *Ist $I = \{0, 1, 2, ...\}$ und sind $T_1, T_2, ...$ Stoppzeiten, dann ist auch $T := \inf_n T_n$ eine Stoppzeit.*
3. *Ist $I = [0, \infty)$ und sind $T_1, T_2, ...$ Optionszeiten, dann ist auch $T := \inf_n T_n$ eine Optionszeit. Außerdem gilt $\mathcal{F}_T^+ = \bigcap_n \mathcal{F}_{T_n}^+$.*

Beweis. 1. Es gilt $\{T \le t\} = \bigcap_n \{T_n \le t\} \in \mathcal{F}_t$, woraus die Behauptung folgt.
2. Es gilt $\{T \le t\} = \bigcup_n \{T_n \le t\} \in \mathcal{F}_t$, woraus wiederum die Behauptung folgt.
3. Hier ist $\{T < t\} = \bigcup_n \{T_n < t\} \in \mathcal{F}_t$. Da $T \le T_n$, gilt $\mathcal{F}_T^+ \subseteq \bigcap_n \mathcal{F}_{T_n}^+$ nach Lemma 14.23.5. Ist andererseits $A \in \bigcap_n \mathcal{F}_{T_n}^+$, so ist

$$A \cap \{T < t\} = A \cap \bigcup_n \{T_n < t\} = \bigcup_n (A \cap \{T_n < t\}) \in \mathcal{F}_t.$$

Damit ist $A \in \mathcal{F}_T^+$. □

Proposition 14.28 (Approximation durch abzählbare Stoppzeiten). *Ist $I = [0, \infty)$, so kann jede Optionszeit T durch eine Folge von Stoppzeiten $T_1, T_2, ...$ approximiert werden, so dass T_n nur Werte in einer diskreten (insbesondere abzählbaren) Menge annimmt und $T_n \downarrow T$.*

Beweis. Wir definieren $T_n = 2^{-n}[2^n T + 1]$. Dann ist $T_1, T_2, ...$ eine gegen T fallende Folge, wobei T_n nur die Werte $\{1, 2, ...\} \cdot 2^{-n}$ annimmt, $n = 1, 2, ...$ Weiter ist $\{T_n \le k2^{-n}\} = \{T < k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$, also ist T_n eine Stoppzeit, $n = 1, 2, ...$ □

Definition 14.29 (Treffzeit). *Sei $B \in \mathcal{B}(E)$. Dann ist die* Treffzeit von B *definiert als*

$$T_B := \inf\{t : X_t \in B\}.$$

Um herauszufinden, ob die Treffzeit T_B eine Stopp- (oder Options-)zeit ist, ist folgendes Ergebnis wichtig.

Proposition 14.30 (Treffzeiten als Options- und Stoppzeiten). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein E-wertiger Prozess, der an eine Filtration $(\mathcal{F}_t)_{t\in I}$ adaptiert ist. Dann gilt für $B \in \mathcal{B}(E)$:*

1. *Falls $I = \{0, 1, 2, ...\}$, so ist die Zeit T_B eine Stoppzeit.*
2. *Falls $I = [0, \infty)$, B offen ist und $\mathcal{X}$ rechtsstetige Pfade hat, so ist T_B eine Optionszeit.*
3. *Falls $I = [0, \infty)$, B abgeschlossen ist und $\mathcal{X}$ stetige Pfade hat, so ist T_B eine Stoppzeit.*

Beweis. 1. Hier gilt

$$\{T_B \le t\} = \bigcup_{s \le t} \{X_s \in B\} \in \mathcal{F}_t.$$

Für 2. schreiben wir

$$\{T_B < t\} = \bigcup_{\mathbb{Q} \ni s < t} \{X_s \in B\} \in \mathcal{F}_t.$$

Bei 3. ist mit $B_n := \{x \in E : r(x, B) < 1/n\}$

$$\{T_B \le t\} = \bigcap_n \{T_{B_n} \le t\} = \bigcap_n (\{T_{B_n} < t\} \cup \{X_t \in \overline{B_n}\}) \in \mathcal{F}_t.$$

Damit sind alle Behauptungen gezeigt. □

14.5 Progressive Messbarkeit

Per Definition ist für einen stochastischen Prozess $\mathcal{X} = (X_t)_{t\in I}$ jede der Größen X_t messbar, $t \in I$. Jedoch ist (noch) unklar, wann genau für eine zufällige Zeit T die Größe $X_T : \omega \mapsto X_{T(\omega)}(\omega)$ messbar und damit eine Zufallsvariable ist. Hierzu benötigen wir einen stärkeren Messbarkeitsbegriff des Prozesses $\mathcal{X}$.

Definition 14.31 (Progressive Messbarkeit). *Sei $(\mathcal{F}_t)_{t\in I}$ eine Filtration und $\mathcal{X} = (X_t)_{t\in I}$ ein daran adaptierter stochastischer Prozess. Dann heißt $\mathcal{X}$* progressiv messbar *bezüglich $(\mathcal{F}_t)_{t\in I}$, falls für alle $t \in I$ die Abbildung*

$$\begin{cases} I \cap [0,t] \times \Omega & \to E \\ (s,\omega) & \mapsto X_s(\omega) \end{cases}$$

messbar ist bezüglich $I \cap \mathcal{B}([0,t]) \otimes \mathcal{F}_s/\mathcal{B}(E)$.

Lemma 14.32 (Rechtsstetige Pfade und progressive Messbarkeit). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein an die Filtration $(\mathcal{F}_t)_{t\in I}$ adaptierter stochastischer Prozess. Ist entweder I abzählbar, oder hat $\mathcal{X}$ rechtsstetige Pfade, so ist $\mathcal{X}$ progressiv messbar bezüglich $(\mathcal{F}_t)_{t\in I}$.*

Beweis. Sei $t \in I$. Wir betrachten die Abbildung

$$Y : \begin{cases} I \cap [0,t] \times \Omega & \to E \\ (s,\omega) & \mapsto X_s(\omega). \end{cases}$$

Sei zunächst I abzählbar und $B \in \mathcal{B}(E)$. Dann gilt

$$Y^{-1}(B) = \bigcup_{s\in I, s\le t} \{s\} \times X_s^{-1}(B) \in \mathcal{B}(I \cap [0,t]) \otimes \mathcal{F}_t.$$

Als nächstes sei I überabzählbar und $\mathcal{X}$ habe rechtsstetige Pfade. Betrachte die Prozesse $\mathcal{X}^n = (X_s^n)_{t\in I\cap[0,t]}, n = 1, 2, ...$ mit $X_s^n := X_{(2^{-n}\lceil 2^n s\rceil)\wedge t}$ und die entsprechenden Abbildungen Y_n. Wegen der Rechtsstetigkeit der Pfade gilt dann $Y_n \xrightarrow{n\to\infty}_{fs} Y$. Außerdem ist

$$\begin{aligned} Y_n^{-1}(B) = & \bigcup_{k:(k+1)2^{-n}\le t} [k2^{-n}, (k+1)2^{-n}) \times X_{(k+1)2^{-n}}^{-1}(B) \cup [2^{-n}\lfloor 2^n t\rfloor, t] \times X_t^{-1}(B) \\ & \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t. \end{aligned}$$

□

Proposition 14.33 (Messbarkeit von X_T). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein an die Filtration $(\mathcal{F}_t)_{t\in I}$ adaptierter, progressiv messbarer, stochastischer Prozess und T eine $(\mathcal{F}_t)_{t\in I}$-Stoppzeit. Dann ist*

$$X_T : \begin{cases} \{T < \infty\} & \to E \\ \omega & \mapsto X_{T(\omega)}(\omega) \end{cases}$$

messbar bezüglich $\{T < \infty\} \cap \mathcal{F}_T/\mathcal{B}(E)$.

Beweis. Wir müssen zeigen, dass $\{X_T \in B, T \leq t\} \in \mathcal{F}_t$ für $B \in \mathcal{B}(E)$ gilt, $t \in I$. Per Definition von $\mathcal{F}_T$ gilt dann nämlich, dass $\{X_T \in B\} \in \mathcal{F}_T$, woraus die Behauptung folgt. Da aber $\{X_T \in B, T \leq t\} = \{X_{T\wedge t} \in B, T \leq t\}$, genügt es zu zeigen, dass $X_{T\wedge t}$ messbar ist bezüglich $\mathcal{F}_t$, $t \in I$. Wir können also oBdA annehmen, dass $T \leq t$ gilt. Wir schreiben $X_T = Y_t \circ \psi$, wobei $\psi(\omega) := (T(\omega), \omega)$ messbar ist bezüglich $\mathcal{F}_t/(I \cap \mathcal{B}([0,t]) \otimes \mathcal{F}_t)$ und $Y_t(s,\omega) = X_s(\omega)$ nach Voraussetzung $I \cap \mathcal{B}([0,t]) \otimes \mathcal{F}_t/\mathcal{B}(E)$-messbar ist. Die Behauptung folgt nun mit Lemma 4.5.2. □

15 Martingale

Wir kommen nun zu einer wichtigen Klasse stochastischer Prozesse, den Martingalen. Oftmals werden sie als ein *faires Spiel* bezeichnet. Einfach gesagt ist ein Martingal ein reellwertiger stochastischer Prozess, dessen Inkremente im Mittel verschwinden. Im ganzen Abschnitt sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $I \subseteq \mathbb{R}$ eine Indexmenge und $(\mathcal{F}_t)_{t\in I}$ eine Filtration. Adaptiertheit eines stochastischen Prozesses wird immer bezüglich $(\mathcal{F}_t)_{t\in I}$ gemeint sein, ebenso werden wir nur $(\mathcal{F}_t)_{t\in I}$-Stoppzeiten betrachten.

15.1 Definition und Eigenschaften

Für eine $\mathcal{F}$-messbare Zufallsvariable X kann man einen stochastischen Prozess definieren, nämlich $\mathcal{X} = (X_t)_{t\in I}$ mit

$$X_t := \mathbf{E}[X|\mathcal{F}_t]. \tag{15.1}$$

Natürlich ist dann, wegen Theorem 12.2.7

$$\mathbf{E}[X_t|\mathcal{F}_s] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}_t]|\mathcal{F}_s] = \mathbf{E}[X|\mathcal{F}_s] = X_s.$$

Stochastische Prozesse $\mathcal{X}$ mit der Eigenschaft werden wir Martingale nennen. In Abschnitt 15.5 wird es dann (unter anderem) darum gehen, wann es zu einem Martingal $(X_t)_{t\in I}$ eine Zufallsvariable X gibt, so dass (15.1) gilt.

Definition 15.1 ((Sub-, Super-)Martingal). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein adaptierter, reellwertiger stochastischer Prozess mit $\mathbf{E}[|X_t|] < \infty$, $t \in I$. Dann heißt $\mathcal{X}$*

$$\begin{aligned} \textit{Martingal, falls } \mathbf{E}[X_t|\mathcal{F}_s] &= X_s \textit{ für } s,t \in I,\ s < t, \\ \textit{Sub-Martingal, falls } \mathbf{E}[X_t|\mathcal{F}_s] &\geq X_s \textit{ für } s,t \in I,\ s < t, \\ \textit{Super-Martingal, falls } \mathbf{E}[X_t|\mathcal{F}_s] &\leq X_s \textit{ für } s,t \in I,\ s < t. \end{aligned}$$

Genauer sagen wir dass $\mathcal{X}$ ein $(\mathcal{F}_t)_{t\in I}$-(Sub, Super)-Martingal ist.

Bemerkung 15.2 (Martingaleigenschaft bei diskreter Indexmenge). Ist I diskret, etwa $I = \{0, 1, 2, ...\}$, so ist ein reellwertiger stochastischer Prozess $\mathcal{X} = (X_t)_{t\in I}$ genau dann ein Martingal, wenn $\mathbf{E}[|X_t|] < \infty$, $t \in I$ und $\mathbf{E}[X_t|\mathcal{F}_{t-1}] = X_{t-1}$ für alle $t = 1, 2, ...$ gilt. Es gilt dann nämlich für $s, t \in I, s \leq t$,

$$\mathbf{E}[X_t|\mathcal{F}_s] = \mathbf{E}[\cdots \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-2}] \cdots |\mathcal{F}_s] = X_s$$

nach Theorem 12.2.7. Analoges gilt für Sub- und Super-Martingale.

Beispiel 15.3 (Summen und Produkte integrierbarer Zufallsvariablen).

1. Sei $X_1, X_2, ...$ eine Folge unabhängiger, integrierbarer Zufallsvariablen mit $\mathbf{E}[X_i] = 0, i = 1, 2, ...$ und $\mathcal{F}_t := \sigma(X_1, ..., X_t)$. Weiter sei $S_0 := 0$ und für $t = 1, 2, ...$

$$S_t := \sum_{i=1}^{t} X_i.$$

Dann ist

$$\mathbf{E}[S_t|\mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1} + X_t|\mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t|\mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t] = S_{t-1},$$

d.h. $(S_t)_{t=0,1,2,...}$ ist ein Martingal.

Falls $\mathbf{E}[X_i] \geq 0$ für alle $i = 1, 2, ...$, so ist $(S_t)_{t\geq 0}$ ein Sub-Martingal.

2. Sei $I = \{-1, -2, ...\}$ und $X_1, X_2, ...$ seien integrierbare, unabhängige, identisch verteilte Zufallsvariablen. Weiter setzen wir für $t \in I$

$$S_t := \frac{1}{|t|} \sum_{i=1}^{|t|} X_i$$

und $\mathcal{F}_t := \sigma(..., S_{t-1}, S_t)$. Dann ist für $t \in I$

$$\begin{aligned}
\mathbf{E}[S_t|\mathcal{F}_{t-1}] &= \mathbf{E}\Big[\frac{1}{|t|} \sum_{i=1}^{|t|} X_i \Big| S_{t-1}, S_{t-2}, ...\Big] \\
&= \frac{1}{|t|} \sum_{i=1}^{|t|} \mathbf{E}\Big[X_i \Big| \sum_{i=1}^{|t|+1} X_i\Big] \\
&= \mathbf{E}\Big[X_1 \Big| \sum_{i=1}^{|t|+1} X_i\Big] \\
&= \frac{1}{|t-1|} \sum_{i=1}^{|t-1|} X_i \\
&= S_{t-1}
\end{aligned}$$

nach Beispiel 12.9. Speziell ist

$$\begin{aligned}
\mathbf{E}[X_1|\mathcal{F}_t] &= \mathbf{E}\Big[X_1 \Big| \sum_{i=1}^{|t|} X_i\Big] \\
&= \frac{1}{|t|} \sum_{i=1}^{|t|} X_i \\
&= S_t.
\end{aligned}$$

3. Sei $I = \{0, 1, 2, ...\}$ und $X_1, X_2, ...$ eine Folge unabhängiger, integrierbarer Zufallsvariablen mit $\mathbf{E}[X_i] = 1, i = 1, 2, ...$ und $\mathcal{F}_t := \sigma(X_1, ..., X_t)$. Weiter ist $S_0 := 1$ und für $t = 1, 2, ...$
$$S_t := \prod_{i=1}^t X_i.$$
Dann ist, falls $S_1, S_2, ...$ integrierbar sind,
$$\mathbf{E}[S_t|\mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1}X_t|\mathcal{F}_{t-1}] = S_{t-1} \cdot \mathbf{E}[X_t|\mathcal{F}_{t-1}] = S_{t-1} \cdot \mathbf{E}[X_t] = S_{t-1},$$
d.h. $(S_t)_{t\in I}$ ist ein Martingal.
Falls $\mathbf{E}[X_i] \geq 1$ für alle $i = 1, 2, ...$, so ist $(S_t)_{t\in I}$ ein Sub-Martingal.

Beispiel 15.4 (Vom Poisson-Prozess abgeleitete Martingale). Sei $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t\in I}$ ein Poisson-Prozess mit Intensität λ und $\mathcal{F}_t = \sigma(X_s : s \leq t)$. Dann sind
$$(X_t - \lambda t)_{t\in I} \qquad \text{und} \qquad \Big(X_t^2 - \lambda \int_0^t (2X_r + 1)dr\Big)_{t\in I}$$
Martingale. Es gilt nämlich für $0 \leq s \leq t$
$$\mathbf{E}[X_t - \lambda t|\mathcal{F}_s] = \mathbf{E}[X_s + X_t - X_s - \lambda t|\mathcal{F}_s] = X_s + \lambda(t-s) - \lambda t = X_s - \lambda s,$$
$$\begin{aligned}\mathbf{E}\Big[X_t^2 - X_s^2 - \lambda \int_s^t (2X_r + 1)dr|\mathcal{F}_s\Big]& \\ &= \mathbf{E}\Big[(X_t - X_s)^2 + 2(X_t - X_s)X_s - \lambda((2X_s + 1)(t-s) + 2\int_s^t (X_r - X_s)dr)|\mathcal{F}_s\Big] \\ &= \lambda(t-s) + \lambda^2(t-s)^2 + 2\lambda(t-s)X_s - \lambda((2X_s + 1)(t-s) - \lambda^2(t-s)^2 = 0.\end{aligned}$$

Beispiel 15.5 (Von der Brown'schen Bewegung abgeleitete Martingale). Sei $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t\in I}$ eine Brown'sche Bewegung, $\mathcal{F}_t = \sigma(X_s : s \leq t)$ und $\mu \in \mathbb{R}$.

1. Es sind
$$(\mu X_t)_{t\in I}, \qquad (\mu X_t^2 - \mu t)_{t\in I} \qquad \text{und} \qquad \big(\exp(\mu X_t - \mu^2 t/2)\big)_{t\in I} \tag{15.2}$$
Martingale. Es gilt nämlich für $0 \leq s \leq t$
$$\begin{aligned}\mathbf{E}[\mu X_t|\mathcal{F}_s] &= \mathbf{E}[\mu X_s + \mu(X_t - X_s)|\mathcal{F}_s] = \mu X_s, \\ \mathbf{E}\Big[\mu X_t^2 - \mu t|\mathcal{F}_s] &= \mu\mathbf{E}[(X_t - X_s)^2 + 2(X_t - X_s)X_s + X_s^2 - t|\mathcal{F}_s] \\ &= \mu(t-s) + \mu X_s^2 - \mu t = \mu X_s^2 - \mu s, \\ \mathbf{E}\Big[\exp(\mu X_t - \mu^2 t/2)|\mathcal{F}_s] &= \exp(\mu X_s - \mu^2 t/2) \cdot \mathbf{E}[\exp(\mu(X_t - X_s))] \\ &= \exp(\mu X_s - \mu^2 t/2 + \mu^2(t-s)/2) = \exp(\mu X_s - \mu^2 s/2)\end{aligned}$$
nach Beispiel 7.13.3.
Da der Prozess $\big(\exp(\mu X_t - \mu^2 t/2)\big)_{t\in I}$ ein nicht-negatives Martingal mit $\mathbf{E}[\exp(\mu X_t - \mu t/2)] = 1$ ist, stellt es eine Dichte dar. Es ist nämlich für $\tau > 0$
$$\mathbf{Q}_\tau : \begin{cases} \mathcal{B}(\mathbb{R})^{[0,\tau]} & \to [0,1] \\ A & \mapsto \mathbf{E}[\exp(\mu X_\tau - \mu^2\tau/2), A] \end{cases}$$

ein weiteres Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R})^{[0,\tau]}$, das zu einem Wahrscheinlichkeitsmaß $\mathbf{Q}$ auf $\mathcal{B}(\mathbb{R})^I$ mittels

$$\mathbf{Q}|_{\mathcal{F}_\tau} = \mathbf{Q}_\tau \tag{15.3}$$

fortgesetzt werden kann.

2. Für $\mu \in \mathbb{R}$ heißt der Prozess $(X_t + \mu t)_{t\in[0,\infty)}$ *Brown'sche Bewegung mit Drift* μ. Dieser ist genau dann ein Martingal, wenn $\mu = 0$. Für $\mu > 0$ ist es ein Sub-Martingal und für $\mu < 0$ ein Super-Martingal.

Es besteht ein enger Zusammenhang zwischen der Brown'schen Bewegung mit Drift und dem Martingal $\big(\exp(\mu X_t - \mu^2 t/2)\big)_{t\in I}$ aus (15.2).

Proposition 15.6 (Brown'sche Bewegung mit Drift und Maßwechsel). *Sei* $I = [0,\infty)$ *und* $\mathcal{X} = (X_t)_{t\in I}$ *eine auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbf{P})$ *definierte Brown'sche Bewegung. Weiter sei* $\mathcal{Y} = (Y_t)_{t\in I}$ *mit* $Y_t = X_t + \mu t$ *für ein* $\mu \in \mathbb{R}$ *sowie* $\mathbf{Q}$ *aus* (15.3). *Dann gilt*

$$\mathcal{X}_*\mathbf{Q} = \mathcal{Y}_*\mathbf{P} \qquad \textit{sowie} \qquad \mathcal{Y}_*\mathbf{Q} = \mathcal{X}_*\mathbf{P},$$

d.h. die Verteilung von $\mathcal{X}$ *unter dem Maß* $\mathbf{Q}$ *ist die einer Brown'schen Bewegung mit Drift* μ *und* $\mathcal{Y}$ *ist unter* $\mathbf{Q}$ *ein Martingal.*

Beweis. Zunächst sei f stetig und beschränkt, sowie $0 \le s \le t$. Dann gilt

$$\begin{aligned}
\mathbf{E}_{\mathbf{Q}}[f(X_t)|\mathcal{F}_s] &= \mathbf{E}_{\mathbf{P}}[f(X_t)e^{\mu X_t - \mu^2 t/2}|\mathcal{F}_s] \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 t/2} \int f(X_s + y) e^{\mu y} e^{-y^2/(2(t-s))} dy \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 t/2 + \mu^2 (t-s)/2} \int f(X_s + y) e^{-(y-\mu(t-s))^2/(2(t-s))} dy \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 s/2} \int f(X_s + y + \mu(t-s)) e^{-y^2/(2(t-s))} dy \\
&= \mathbf{E}_{\mathbf{P}}[f(X_t + \mu(t-s))|\mathcal{F}_s] \cdot e^{\mu X_s - \mu^2 s/2}.
\end{aligned}$$

Sei nun $0 \le t_1 \le \dots \le t_n$ sowie $f_1, ..., f_n$ stetig und beschränkt. Dann gilt

$$\begin{aligned}
\mathbf{E}_{\mathbf{Q}}[f_1(X_{t_1}) \cdots f_n(X_{t_n})] &= \mathbf{E}_{\mathbf{P}}[f_1(X_{t_1}) \cdots f_{n-1}(X_{t_{n-1}}) \mathbf{E}_{\mathbf{P}}[f_n(X_{t_n}) e^{\mu X_{t_n} - \mu^2 t_n/2}|\mathcal{F}_{t_{n-1}}]] \\
&= \mathbf{E}_{\mathbf{P}}[f_1(X_{t_1}) \cdots f_{n-2}(X_{t_{n-2}}) \cdot \\
&\qquad \mathbf{E}_{\mathbf{P}}[f_{n-1}(X_{t_{n-1}}) \mathbf{E}_{\mathbf{P}}[f_n(X_{t_n} + \mu(t_n - t_{n-1})|\mathcal{F}_{t_{n-1}}] e^{\mu X_{t_{n-1}} - \mu^2 t_{n-1}/2}]|\mathcal{F}_{t_{n-2}}]] \\
&= \dots = \mathbf{E}_{\mathbf{P}}[f_1(X_{t_1} + \mu t_1) \cdots f_n(X_{t_n} + \mu t_n)] = \mathbf{E}_{\mathbf{P}}[f_1(Y_{t_1}) \cdots f_n(Y_{t_n})].
\end{aligned}$$

Da $f_1, ..., f_n$ beliebig waren, sind also die endlich-dimensionalen Verteilungen von $\mathcal{X}_*\mathbf{Q}$ und $\mathcal{Y}_*\mathbf{P}$ identisch. Die Aussage folgt nun aus Proposition 14.6.1. $\square$

Beispiel 15.7 (Verzweigungsprozesse in diskreter Zeit). Wir betrachten ein einfaches Modell für eine sich zufällige entwickelnde Population. Seien $X_i^{(t)}$) unabhängige, $\{0, 1, 2, ...\}$-wertige Zufallsvariable und $\mu = \mathbf{E}[X_i^{(t)}]$. Hier steht $X_i^{(t)}$ für die Anzahl der Nachkommen des

$i$ten Individuums der Generation t mit $i, t = 0, 1, ..., t = 1, 2, ...$ Startend mit $Z_0 = k$ setzen wir

$$Z_{t+1} = \sum_{i=1}^{Z_t} X_i^{(t)},$$

also ist $\mathcal{Z} = (Z_t)_{t=0,1,2,...}$ der stochastische Prozess der Gesamtzahl an Individuen. Die Verteilung von $X_i^{(t)}$ heißt auch die *Nachkommensverteilung*.

Der Prozess $\mathcal{Z}$ ist genau dann ein (nicht-negatives) Martingal (bezüglich der von $\mathcal{Z}$ erzeugten Filtration), wenn $\mathbf{E}[X_i^{(t)}] = 1$, jedes Individuum im Mittel also einen Nachkommen hat. Es gilt nämlich für $t = 1, 2, ...$

$$\mathbf{E}[Z_{t+1} - Z_t | \mathcal{F}_t] = \mathbf{E}\Big[\sum_{i=1}^{Z_t} X_i^{(t)} - Z_t | \mathcal{F}_t\Big] = (\mu - 1) Z_t.$$

Im Fall $\mu > 1$ ist $\mathcal{Z}$ ein Sub-Martingal, und im Fall $\mu < 1$ ein Super-Martingal. Außerdem heißt[20]

$$\begin{aligned} &\mathcal{Z} \text{ ein kritischer Verzweigungsprozess, falls } \mu = 1, \\ &\mathcal{Z} \text{ ein super-kritischer Verzweigungsprozess, falls } \mu > 1, \\ &\mathcal{Z} \text{ ein sub-kritischer Verzweigungsprozess, falls } \mu < 1. \end{aligned}$$

Allgemein ist $(Z_t/\mu^t)_{t=0,1,2,...}$ ein (nicht-negatives) Martingal, weil genau wie in der letzten Rechnung

$$\mathbf{E}[Z_{t+1} - \mu Z_t | \mathcal{F}_t] = \mu Z_t - \mu Z_t = 0.$$

Bemerkenswert ist außerdem, dass $\mathbf{E}[Z_{t+1}|\mathcal{F}_t] = \mu Z_t$ gilt, woraus man rekursiv folgern kann, dass

$$\mathbf{E}[Z_t] = \mu^t.$$

□

Wir schließen diesen Abschnitt mit einer einfachen Aussage, wie man aus bekannten (Sub)-Martingalen weitere Sub-Martingale erhält.

Proposition 15.8 (Konvexe Funktionen von Martingalen sind Sub-Martingale). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein stochastischer Prozess und $\varphi : \mathbb{R} \to \mathbb{R}$ konvex. Falls $\varphi(X) = (\varphi(X_t))_{t\in I}$ integrierbar ist und eine der beiden Bedingungen*

1. *$\mathcal{X}$ ist ein Martingal*
2. *$\mathcal{X}$ ist ein Sub-Martingal und φ ist nicht-fallend*

erfüllt ist, so ist $\varphi(\mathcal{X}) = (\varphi(X_t))_{t\in I}$ ein Sub-Martingal.

Beweis. Ist $\mathcal{X}$ ein Martingal so ist $\varphi(X_s) = \varphi(\mathbf{E}[X_t|\mathcal{F}_s])$. Ist $\mathcal{X}$ ein Sub-Martingal und φ nicht-fallend, gilt $\varphi(X_s) \leq \varphi(\mathbf{E}[X_t|\mathcal{F}_s])$. In beiden Fällen gilt damit für $s \leq t$ fast sicher wegen der Jensen'schen Ungleichung für bedingte Erwartungen, Proposition 12.4, dass

$$\varphi(X_s) \leq \varphi(\mathbf{E}[X_t|\mathcal{F}_s]) \leq \mathbf{E}[\varphi(X_t)|\mathcal{F}_s],$$

d.h. $\varphi(\mathcal{X})$ ist ein Sub-Martingal. □

[20] Es mag irritierend erscheinnen, dass ein *superkritscher Verzweigungsprozess* ein Sub-Martingal und ein *subkritischer Verzweigungsprozess* ein Super-Martingal ist.

15.2 Eigenschaften von Martingalen in diskreter Zeit

In diesem Abschnitt sei immer $I = \{0, 1, 2, ...\}$ (wobei sich alle Ergebnisse auch auf eine diskrete Indexmenge $I = \{t_0, t_1, ...\}$ mit $t_0 < t_1 < ...$ übertragen lassen). Alle hier eingeführten Konzepte haben ein Analogon für Prozesse in stetiger Zeit. Allerdings sind die entsprechenden Aussagen dann deutlich aufwändiger zu formulieren und zu beweisen. Einige dieser analogen Aussagen werden erst in der Vorlesung *Stochastische Integration und Finanzmathematik* formuliert.

Definition 15.9 (Previsibler Prozess). *Ein stochastischen Prozess $\mathcal{X}$ heißt $(\mathcal{F}_t)_{t\in I}$-*previsibel*, falls $X_0 = 0$ und X_t $\mathcal{F}_{t-1}$-messbar ist, $t = 1, 2, ...$*

Proposition 15.10 (Doob-Zerlegung). *Sei $I = \{0, 1, 2, ...\}$. Jeder adaptierte Prozess $\mathcal{X} = (X_t)_{t\in I}$ hat eine fast sicher eindeutige Zerlegung $\mathcal{X} = \mathcal{M} + \mathcal{A}$, wobei $\mathcal{M}$ ein Martingal und $\mathcal{A}$ previsibel ist. Insbesondere ist $\mathcal{X}$ genau dann ein Sub-Martingal falls $\mathcal{A}$ fast sicher nicht fällt.*

Beweis. Definiere den previsiblen Prozess $\mathcal{A} = (A_t)_{t\in I}$ durch

$$A_t = \sum_{s=1}^{t} \mathbf{E}[X_s - X_{s-1}|\mathcal{F}_{s-1}]. \tag{15.4}$$

Dann ist $\mathcal{M} = \mathcal{X} - \mathcal{A}$ ein Martingal, denn

$$\mathbf{E}[M_t - M_{t-1}|\mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] - (A_t - A_{t-1}) = 0.$$

Kommen wir zur Eindeutigkeit der Darstellung. Falls $\mathcal{X} = \mathcal{M} + \mathcal{A}$ für ein Martingal $\mathcal{M}$ und einen previsiblen Prozess $\mathcal{A}$, so ist $A_t - A_{t-1} = \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}]$ für alle $t = 1, 2, ...$, d.h. (15.4) gilt fast sicher. □

Definition 15.11 (Quadratische Variation, wachsender Prozess). *Sei $I = \{0, 1, 2, ...\}$ und $\mathcal{X} = (X_t)_{t\in I}$ ein quadratisch integrierbares Martingal. Der fast sicher eindeutig bestimmte, previsible Prozess $(\langle\mathcal{X}\rangle_t)_{t\in I}$, für den $(X_t^2 - \langle\mathcal{X}\rangle_t)_{t\in I}$ ein Martingal ist, heißt der* quadratische Variationsprozess *(oder auch der* wachsende Prozess*) von $\mathcal{X}$.*

Proposition 15.12 (Wachsender Prozess und Varianz). *Sei $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t\in I}$ ein Martingal mit quadratischem Variationsprozess $\langle\mathcal{X}\rangle = (\langle\mathcal{X}\rangle_t)_{t\in I}$. Dann ist*

$$\langle\mathcal{X}\rangle_t = \sum_{s=1}^{t} \mathbf{E}[X_s^2 - X_{s-1}^2|\mathcal{F}_{s-1}] = \sum_{s=1}^{t} \mathbf{E}[(X_s - X_{s-1})^2|\mathcal{F}_{s-1}]$$

und

$$\mathbf{E}[\langle X\rangle_t] = \mathbf{V}[X_t - X_0].$$

Beweis. Wie im Beweis von Proposition 15.10 kann man den Prozess $\langle\mathcal{X}\rangle$ mittels (15.4) schreiben. Daraus folgt sofort das erste Gleichheitszeichen. Das zweite folgt, da $\mathbf{E}[X_s X_{s-1}|\mathcal{F}_{s-1}] = X_{s-1}^2$. Weiter ist

$$\mathbf{E}[\langle\mathcal{X}\rangle_t] = \sum_{s=1}^{t} \mathbf{E}[X_s^2 - X_{s-1}^2] = \mathbf{E}[X_t^2 - X_0^2] = \mathbf{E}[(X_t - X_0)^2] = \mathbf{V}[X_t - X_0].$$

□

Beispiel 15.13 (Wachsende Prozesse). 1. Sei $\mathcal{S} = (S_t)_{\in I}$ mit $S_t = \sum_{s=1}^t X_s$ wie in Beispiel 15.3.1 mit quadratische integrierbaren Zufallsvariablen $X_1, X_2, \ldots$. Dann ist nach Poposition 15.12

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[X_s^2].$$

Insbesondere ist der quadratische Variationsprozess von $\mathcal{S}$ deterministisch.

2. Sei $\mathcal{S} = (S_t)_{\in I}$ mit $S_t = \prod_{s=1}^t X_s$ wie in Beispiel 15.3.3 mit quadratisch integrierbaren Zufallsvariablen $X_1, X_2, \ldots$. Dann ist

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[(S_s - S_{s-1})^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{E}[(X_s - 1)^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{V}[X_s].$$

Insbesondere ist in diesem Beispiel der Prozess $\langle \mathcal{S} \rangle$ echt stochastisch.

3. Sei $I = [0, \infty)$ und $(X_t)_{t \in I}$ eine Brown'sche Bewegung. Auch in stetiger Zeit ist der wachsende Prozess $(\langle \mathcal{X} \rangle_t)_{t \in I}$ so definiert, dass $(X_t^2 - \langle \mathcal{X} \rangle_t)_{t \in I}$ ein Martingal ist. Nach Beispiel 15.5 ist also $\langle \mathcal{X} \rangle_t = t$ ein Kandidat für den wachsenden Prozess der Brown'schen Bewegung. Allerdings ist in stetiger Zeit schwieriger zu sagen, was das Äquivalent eines *previsiblen* Prozesses sein soll.

Definition 15.14 (Diskretes stochastisches Integral). *Sei $I = \{0, 1, 2, \ldots\}$ und $\mathcal{H} = (H_t)_{t \in I}, \mathcal{X} = (X_t)_{t \in I}$ stochastische Prozesse mit Werten in $\mathbb{R}$. Ist $\mathcal{X}$ adaptiert und $\mathcal{H}$ previsibel, so definieren wir das stochastische Integral $\mathcal{H} \cdot \mathcal{X} = ((\mathcal{H} \cdot \mathcal{X})_t)_{t \in I}$ durch*

$$(\mathcal{H} \cdot \mathcal{X})_t = \sum_{s=1}^t H_s (X_s - X_{s-1})$$

für alle $t \in I$. Ist $\mathcal{X}$ ein Martingal, so nennen wir $\mathcal{H} \cdot \mathcal{X}$ eine Martingaltransformierte von $\mathcal{X}$.

Proposition 15.15 (Stabilität der stochastischen Integrale). *Sei $I = \{0, 1, 2, \ldots\}$ und $\mathcal{X} = (X_t)_{t \in I}$ ein adaptierter, reellwertiger Prozess mit $\mathbf{E}[|X_0|] < \infty$.*

1. *$\mathcal{X}$ ist genau dann ein Martingal, wenn für jeden previsiblen Prozess $\mathcal{H} = (H_t)_{t \in I}$ das stochastische Integral $\mathcal{H} \cdot \mathcal{X}$ ein Martingal ist.*

2. *$\mathcal{X}$ ist genau dann ein Sub-Martingal (Super-Martingal), wenn für jeden previsiblen, nicht-negativen Prozess $\mathcal{H} = (H_t)_{t \in I}$ das stochastische Integral $\mathcal{H} \cdot \mathcal{X}$ ein Sub-Martingal (Super-Martingal) ist.*

Beweis. 1. '$\Rightarrow$': Man schreibt sofort

$$\begin{aligned}
\mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_{t+1} - (\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_t] &= \mathbf{E}[H_{t+1}(X_{t+1} - X_t) | \mathcal{F}_t] \\
&= H_{t+1} \mathbf{E}[X_{t+1} - X_t | \mathcal{F}_t] \\
&= 0.
\end{aligned}$$

'$\Leftarrow$': Sei $t \in I$ und $H_s := 1_{\{s=t\}}$. Dann ist $\mathcal{H} = (H_s)_{s \in I}$ deterministisch, insbesondere previsibel. Da $(\mathcal{H} \cdot \mathcal{X})_{t-1} = 0$ gilt, folgt

$$0 = \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}] = \mathbf{E}[X_t | \mathcal{F}_{t-1}] - X_{t-1}$$

Daraus folgt die Behauptung.
2. folgt analog. □

Beispiel 15.16 (Quadratische Variation für stochastische Integrale). Sei $I = \{0,1,2,...\}$, $\mathcal{X} = (X_t)_{t\in I}$ ein Martingal und $\mathcal{H} = (H_t)_{t\in I}$ previsibel. Dann ist wegen Proposition 15.12

$$\begin{aligned}\langle \mathcal{H}\cdot\mathcal{X}\rangle_t &= \sum_{s=1}^t \mathbf{E}[((\mathcal{H}\cdot\mathcal{X})_s - (\mathcal{H}\cdot\mathcal{X})_{s-1})^2|\mathcal{F}_{s-1}] = \sum_{s=1}^t \mathbf{E}[H_s^2(X_s - X_{s-1})^2|\mathcal{F}_{s-1}]\\ &= \sum_{s=1}^t H_s^2\cdot\mathbf{E}[(X_s - X_{s-1})^2|\mathcal{F}_{s-1}].\end{aligned}$$

Insebsondere gilt also

$$\mathbf{V}[(\mathcal{H}\cdot\mathcal{X})_t] = \sum_{s=1}^t \mathbf{E}[H_s^2\cdot(X_s - X_{s-1})^2].$$

Beispiel 15.17 (Auszahlung bei Spielsystemen). Martingaltransformierte kann man auch als Auszahlungen von Spielsystemen interpretieren. Gegeben, eine zufällige Größe entwickelt sich gemäß des adaptierten Prozesses $\mathcal{X} = (X_t)_{t=0,1,2,...}$. Wettet man vor Zeit t mit einem Einsatz H_t (basierend auf den Erfahrungen, die aus $X_0, ..., X_{t-1}$ gewonnen wurden) auf die Änderung der zufälligen Größe $X_t - X_{t-1}$, so ist $(\mathcal{H}\cdot\mathcal{X})_t$ der bis zur Zeit t realisierte Gewinn. Gegeben der zugrunde liegende Prozess $\mathcal{X}$ ist ein Martingal, zeigt Proposition 15.15, dass der erzielte Gewinn $\mathcal{H}\cdot\mathcal{X}$ für jede Strategie $\mathcal{H}$ ein Martingal ist. Insbesondere ist der erwartete Gewinn 0.

Als Beispiel betrachten wie das Petersburger Paradoxon: eine faire Münze wird unendlich oft geworfen. In jeder Runde setzt ein Spieler einen Einsatz in beliebiger Höhe. Kommt `Kopf`, verliert er ihn, kommt `Zahl`, so wird der Einsatz verdoppelt wieder ausbezahlt. Das Paradox besteht aus folgender Strategie: startend mit einem Einsatz von 1 beim ersten Münzwurf, verdoppelt der Spieler bei jedem Misserfolg seinen Einsatz. Kommt der erste Erfolg im t-ten Wurf, so beträgt sein bisheriger Einsatz $\sum_{i=1}^t 2^{i-1} = 2^t - 1$. Da der letzte Einsatz 2^{t-1}war, bekommt der Spieler also 2^t zurück, hat also sicher einen Gewinn von 1 gemacht, obwohl das Spiel fair war.

Um dieses Spiel mittels Martingalen zu analysieren, sei $X_1, X_2, ...$ eine unabhängig, identisch verteilte Folge mit $\mathbf{P}(X_1 = -1) = \mathbf{P}(X_1 = 1) = \frac{1}{2}$, und $S_0 = 0, S_t = \sum_{i=1}^t X_i$. Dann ist $\mathcal{S} = (S_t)_{t=0,1,2,...}$ ein Martingal. Weiter sei H_t der Einsatz im t-ten Spiel. Also ist

$$(\mathcal{H}\cdot\mathcal{S})_t = \sum_{i=1}^t H_i(S_i - S_{i-1}) = \sum_{i=1}^t H_i X_i$$

der Gewinn nach dem t-ten Spiel. Da mit $\mathcal{S}$ auch $\mathcal{H}\cdot\mathcal{S}$ ein Martingal ist, gilt

$$\lim_{t\to\infty}\mathbf{E}[(\mathcal{H}\cdot\mathcal{S})_t] = \mathbf{E}[(\mathcal{H}\cdot\mathcal{S})_1] = \mathbf{E}[X_1] = 0,$$

d.h. der mittlere Gewinn nach langer Zeit ist 0, unabhängig von der Strategie $\mathcal{H}$. Oben haben wir den Einsatz

$$H_t := 2^{t-1}1_{\{S_{t-1}=-(t-1)\}} \tag{15.5}$$

betrachtet und gezeigt, dass für den Gewinn $(\mathcal{H}\cdot\mathcal{S})_t \xrightarrow{t\to\infty}_{fs} 1$ gilt.

Wie bewerten wir nun die Strategie (15.5)? Sei T die zufällige Zeit des Gewinns, d.h. T ist geometrisch verteilt mit Parameter $\frac{1}{2}$. Insbesondere ist T fast sicher endlich. Dann ist

$$\mathbf{E}\Big[\sum_{t=1}^{\infty} H_t\Big] = \sum_{k=1}^{\infty} \frac{1}{2^k}(2^k - 1) = \infty,$$

d.h. für die obige Strategie benötigt man unter Umständen sehr viel Kapital.

Proposition 15.18 (Optional Stopping). *Sei $I = \{0, 1, 2, ...\}$ und $\mathcal{X} = (X_t)_{t\in I}$ ein (Sub, Super)-Martingal und T eine Stoppzeit. Dann ist $\mathcal{X}^T = (X_{T\wedge t})_{t\in I}$ ein (Sub, Super)-Martingal.*

Beweis. Wir zeigen die Aussage nur für den Fall, dass $\mathcal{X}$ ein Sub-Martingal ist. Die anderen Aussagen ergeben sich aus Symmetriegründen. Für ein Sub-Martingal $\mathcal{X}$ und $\{T > t-1\} \in \mathcal{F}_t$ ist

$$\begin{aligned}\mathbf{E}[X_{T\wedge t} - X_{T\wedge(t-1)}|\mathcal{F}_{t-1}] &= \mathbf{E}[(X_t - X_{t-1})1_{\{T>t-1\}}|\mathcal{F}_{t-1}] \\ &= 1_{\{T>t-1\}}\mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] \geq 0,\end{aligned}$$

d.h. $\mathcal{X}^T$ ist ein Sub-Martingal. □

Lemma 15.19 (Bedingen auf $\mathcal{F}_T$). *Sei $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t\in I}$ ein Martingal und T eine durch t beschränkte Stoppzeit. Dann gilt $X_T = \mathbf{E}[X_t|\mathcal{F}_T]$.*

Beweis. Nach der Definition der bedingten Erwartung und da X_T $\mathcal{F}_T$-messbar ist (siehe Proposition 14.33), müssen wir zeigen, dass $\mathbf{E}[X_t; A] = \mathbf{E}[X_T; A]$ für $A \in \mathcal{F}_T$ gilt. Es ist $\{T = s\} \cap A \in \mathcal{F}_s$ für $s \in I$, also

$$\begin{aligned}\mathbf{E}[X_T; A] &= \sum_{s=1}^{t} \mathbf{E}[X_s; \{T = s\} \cap A] \\ &= \sum_{s=1}^{t} \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_s]; \{T = s\} \cap A] \\ &= \sum_{s=1}^{t} \mathbf{E}[X_t; \{T = s\} \cap A] \\ &= \mathbf{E}[X_t; A].\end{aligned}$$

□

Lemma 15.20 (Gleichgradige Integrierbarkeit und Stoppzeiten).
Sei $I = \{0, 1, 2, ...\}$. Ein Martingal $\mathcal{X} = (X_t)_{t\in I}$ ist genau dann gleichgradig integrierbar, wenn die Familie $\{X_T : T$ fast sicher endliche Stoppzeit$\}$ gleichgradig integrierbar ist.

Beweis. '$\Leftarrow$': klar.
'$\Rightarrow$': Nach Lemma 8.9 gibt es eine konvexe Funktion $f : \mathbb{R}_+ \to \mathbb{R}_+$ mit $\frac{f(x)}{x} \xrightarrow{x\to\infty} \infty$ und $\sup_{t\in I} \mathbf{E}[f(|X_t|)] =: L < \infty$. Sei T eine fast sicher endliche Stoppzeit, dann ist nach

Lemma 15.19 (angewendet auf die fast sicher endliche Stoppzeit $T \wedge t$) $\mathbf{E}[X_t|\mathcal{F}_{T\wedge t}] = X_{T\wedge t}$. Da $\{T \leq t\} \in \mathcal{F}_{T\wedge t}$ folgt mit der Jensen'schen Ungleichung

$$\begin{aligned}\mathbf{E}[f(|X_T|), \{T \leq t\}] &= \mathbf{E}[f(|X_{T\wedge t}|), \{T \leq t\}] \\ &= \mathbf{E}[f(|\mathbf{E}[X_t|\mathcal{F}_{T\wedge t}]|), \{T \leq t\}] \\ &\leq \mathbf{E}[\mathbf{E}[f(|X_t|)|\mathcal{F}_{T\wedge t}], \{T \leq t\}] \\ &= \mathbf{E}[f(|X_t|), \{T \leq t\}] \leq L.\end{aligned}$$

Damit ist $\mathbf{E}[f(|X_T|)] \leq L$, d.h. die Behauptung folgt mit Lemma 8.9. □

In Beispiel 15.17 war $\mathcal{H} \cdot \mathcal{S}$ ein Martingal, T eine Stoppzeit und $\mathbf{E}[(\mathcal{H} \cdot \mathcal{S})_t] = 0 \neq 1 = (\mathcal{H} \cdot \mathcal{S})_T$. Falls T beschränkt gewesen wäre, wäre diese Ungleichung nicht möglich gewesen, wie wir nun zeigen.

Theorem 15.21 (Optional Sampling Theorem). *Sei $I = \{0, 1, 2, ...\}$, $S \leq T$ fast sicher endliche Stoppzeiten und $\mathcal{X} = (X_t)_{t\in I}$ ein Sub-Martingal. Ist entweder T beschränkt oder $\mathcal{X}$ gleichgradig integrierbar, so ist X_T integrierbar und $X_S \leq \mathbf{E}[X_T|\mathcal{F}_S]$.*

Beweis. Wir führen den Beweis zunächst im Fall einer beschränkten Stoppzeit T. Sei also $T \leq t$ für ein $t \in I$. Wir verwenden die Doob-Zerlegung $\mathcal{X} = \mathcal{M} + \mathcal{A}$ von $\mathcal{X}$ in das Martingal $\mathcal{M}$ und den monoton nicht-fallenden Prozess $\mathcal{A}$. Dann ist mit Lemma 15.19 und $\mathcal{F}_S \subseteq \mathcal{F}_T$ nach Theorem 12.2.7

$$\begin{aligned}X_S = M_S + A_S &= \mathbf{E}[M_t + A_S|\mathcal{F}_S] \\ &\leq \mathbf{E}[M_t + A_T|\mathcal{F}_S] \\ &= \mathbf{E}[\mathbf{E}[M_t|\mathcal{F}_T] + A_T|\mathcal{F}_S] \\ &= \mathbf{E}[M_T + A_T|\mathcal{F}_S] \\ &= \mathbf{E}[X_T|\mathcal{F}_S].\end{aligned}$$

Sei nun T unbeschränkt und $\mathcal{X}$ gleichgradig integrierbar. Sei $\mathcal{X} = \mathcal{M} + \mathcal{A}$ die Doob-Zerlegung von $\mathcal{X}$ in das Martingal $\mathcal{M}$ und den nicht-fallenden previsiblen Prozess $\mathcal{A} \geq 0$ mit $A_0 = 0$. Da

$$\mathbf{E}[|A_t|] = \mathbf{E}[A_t] = \mathbf{E}[X_t - X_0] \leq \mathbf{E}[|X_0|] + \sup_{s\in I} \mathbf{E}[|X_s|]$$

gilt $A_t \uparrow A_\infty$ für ein $A_\infty \geq 0$ mit $\mathbf{E}[A_\infty] < \infty$. Mit Lemma 8.9 kann man folgern, dass auch $\mathcal{M}$ gleichgradig integrierbar ist. Wir wenden nun das Optional Sampling Theorem auf die beschränkten Stoppzeiten $S \wedge t$, $T \wedge t$ und $\mathcal{M}$ an. Für $A \in \mathcal{F}_S$ ist $\{S \leq t\} \cap A \in \mathcal{F}_{S\wedge t}$, also

$$\mathbf{E}[M_{T\wedge t}, \{S \leq t\} \cap A] = \mathbf{E}[\mathbf{E}[M_{T\wedge t}|\mathcal{F}_{S\wedge t}], \{S \leq t\} \cap A] = \mathbf{E}[M_{S\wedge t}, \{S \leq t\} \cap A].$$

Da nach Lemma 15.20 die Menge $\{M_{S\wedge t}, M_{T\wedge t} : t \in I\}$ gleichgradig integrierbar ist, gilt mit Theorem 8.11

$$\mathbf{E}[M_T, A] = \lim_{t\to\infty} \mathbf{E}[M_{T\wedge t}, \{S \leq t\} \cap A] = \lim_{t\to\infty} \mathbf{E}[M_{S\wedge t}, \{S \leq t\} \cap A] = \mathbf{E}[M_S, A],$$

d.h. $\mathbf{E}[M_T|\mathcal{F}_S] = M_S$. Ferner folgt

$$\mathbf{E}[X_T|\mathcal{F}_S] = \mathbf{E}[M_T|\mathcal{F}_S] + A_S + \mathbf{E}[A_T - A_S|\mathcal{F}_S] \geq M_S + A_S = X_S.$$

□

Das Optional Sampling Theorem bietet eine einfache Möglichkeit der Charakterisierung von Martingalen.

Lemma 15.22 (Charakterisierung von Martingalen). *Sei* $I = \{0, 1, 2, ...\}$, *und* $\mathcal{X} = (X_t)_{t\in I}$ *ein adaptierter, stochastischer Prozess. Dann ist* $\mathcal{X}$ *genau dann ein Martingal, falls* $\mathbf{E}[X_S] = \mathbf{E}[X_T]$ *für Stoppzeiten* S, T *gilt, die nur zwei Werte annehmen.*

Beweis. Siehe Übung.

□

Beispiel 15.23 (Wald'sche Identitäten, Ruin-Problem).

1. Seien $X_1, X_2, ... \in \mathcal{L}^1$ unabhängig mit $\mu := \mathbf{E}[X_1] = \mathbf{E}[X_2] = ...$, und $S_t := \sum_{s=1}^t X_s$. Weiter sei T eine fast sicher beschränkte Stoppzeit. Dann gilt die erste Wald'sche Identität
$$\mathbf{E}[S_T] = \mathbf{E}[T]\mu.$$
Denn: der Prozess $\mathcal{M} = (M_t)_{t=0,1,2,...}$ mit $M_0 = 0$, $M_t = S_t - t\mu$ für $t = 1, 2, ...$ ist ein Martingal, und nach dem Optional Sampling Theorem
$$0 = \mathbf{E}[M_T] = \mathbf{E}[S_T] - \mathbf{E}[T]\mu.$$
Ist außerdem $X_1, X_2, ... \in L^2$ mit $\sigma^2 = \mathbf{V}[X_1] = \mathbf{V}[X_2] = ...$ und T unabhängig von $X_1, X_2, ...$, so gilt die zweite Wald'sche Identität
$$\mathbf{V}[S_T] = \mathbf{E}[T]\sigma^2 + \mathbf{V}[T]\mu^2.$$
Denn: $(M_t^2 - \langle M\rangle_t)_{t=0,1,2,...}$ ist ein Martingal, und $\langle M\rangle_t = t\sigma^2$ nach Beispiel 15.13, also
$$0 = \mathbf{E}[M_T^2 - \langle M\rangle_T] = \mathbf{E}[M_T^2] - \mathbf{E}[T]\sigma^2.$$
Außerdem ist wegen der Unabhängigkeit von T und $X_1, X_2, ...$
$$\mathbf{COV}[S_T, T] = \mathbf{E}[\mathbf{E}[X_1 + \cdots + X_T|T]T] - \mu\mathbf{E}[T]^2 = \mu\mathbf{V}[T],$$
also
$$\mathbf{E}[M_T^2] = \mathbf{V}[S_T - T\mu] = \mathbf{V}[S_T] + \mu^2\mathbf{V}[T] - 2\mu\mathbf{COV}[S_T, T] = \mathbf{V}[S_T] - \mu^2\mathbf{V}[T].$$
In allen beiden Wald'schen Identitäten kann man die Voraussetzung, dass T beschränkt ist, abschwächen.

2. Seien $k \in \mathbb{N}$ und $X_1, X_2, ...$ unabhängige und identisch verteilte Zufallsvariablen mit $\mathbf{P}(X_1 = 1) = 1 - \mathbf{P}(X_1 = -1) = p := 1 - q$. Für $N \in \mathbb{N}$ mit $0 < k < N$ sei $S_0 = k$ und $S_t = S_0 + \sum_{i=1}^t X_i$. Weiter sei $T := \inf\{t : S_t \in \{0, N\}\}$ und $p_k := \mathbf{P}(S_T = 0)$. Das bedeutet: man spielt ein Spiel, startend mit Kapital k, bis man entweder ruiniert ist oder Kapital N besitzt. In jedem Schritt gewinnt man mit Wahrscheinlichkeit p eine Kapitaleinheit und verliert mit Wahrscheinlichkeit $q = 1 - p$ eine Kapitaleinheit. Dann ist die Wahrscheinlichkeit, ruiniert zu werden (0 Kapitaleinheiten zu besitzen), gegeben durch p_k.

Im Fall $p = \frac{1}{2}$ ist $(S_t)_{t=0,1,2,...}$ ein Martingal, und damit nach dem Optional Sampling Theorem

$$k = \mathbf{E}[S_T] = N(1 - \mathbf{P}(S_T = 0)),$$

also

$$\mathbf{P}(S_T = 0) = \frac{N-k}{N}.$$

Eine ähnliche Rechnung erlaubt die Bestimmung von p_k für den Fall $p \neq \frac{1}{2}$.

Weiter berechnen wir nun mittels des Optional Sampling Theorems für $p \neq \frac{1}{2}$

$$p_k := \mathbf{P}(S_T = 0) = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}. \tag{15.6}$$

Denn: es gilt

$$\mathbf{E}\Big[\Big(\frac{q}{p}\Big)^{X_1}\Big] = \frac{q}{p}p + \frac{p}{q}q = 1$$

und damit ist $\mathcal{Y} = (Y_t)_{t=0,1,2,...}$, definiert durch $Y_t := \left(\frac{q}{p}\right)^{S_t}$ nach Beispiel 15.3.3 ein Martingal. Da T fast sicher endlich ist, ist $Y_{T\wedge t}$ wegen Proposition 15.18 ein Martingal, das durch 1 und $\left(\frac{q}{p}\right)^N$ beschränkt ist. Wegen Theorem 15.21 ist

$$\Big(\frac{q}{p}\Big)^k = \mathbf{E}[Y_0] = \mathbf{E}[Y_T] = p_k + (1-p_k)\Big(\frac{q}{p}\Big)^N,$$

woraus (15.6) folgt.

3. Betrachten wir einen fairen Münzwurf. Wie lange dauert es, bis zum ersten Mal das Muster `ZKZK` auftritt? (`K` und `Z` stehen hier für Kopf und Zahl.)

   Um dies zu berechnen, betrachten wir das folgende Spiel: vor dem ersten Münzwurf wettet eine Spielerin einen Euro auf `Z`. Falls sie verliert, hört sie auf, falls sie gewinnt, setzt sie vor dem nächsten Wurf zwei Euro auf `K`. Verliert sie im zweiten Wurf, hört sie auf, im Fall eines Gewinns wettet sie vier Euro auf `Z`. Verliert sie im dritten Wurf, hört sie auf, gewinnt sie, wettet sie acht Euro auf `K`. Falls sie also beim vierten Wurf gewinnt, hat sie insgesamt 15 Euro gewonen. In allen anderen Fällen verliert sie einen Euro.

   Nehmen wir nun an, dass vor jedem Münzwurf eine neue Spielerin nach obiger Strategie spielt. Das Spiel wird beendet, wenn das erste Mal eine Spielerin 15 Euro gewinnt.

   Sei X_t der Gesamtgewinn aller Spielerinnen bis zur Zeit t und T die Zeit, zu der das Spiel gestoppt wird, weil zum ersten Mal das Muster `ZKZK` aufgetreten ist. Sicher ist

$$|X_t| \leq 15 \cdot t, \qquad \mathbf{P}[T > 4t] \leq \frac{15}{16}^t.$$

   Damit hat $(X_{t\wedge T} : t = 1, 2, ...)$ eine integrierbare Majorante, ist also nach Beispiel 8.8.2 gleichgradig integrierbar. Damit können wir das Optional Stopping-Theorem anwenden, d.h. $(X_{T\wedge t})_{t=1,2,...}$ ist ein Martingal.

   Sicher ist

$$X_T = 15 - 1 + 3 - 1 - (T - 4)$$

da die ersten $T-4$ Spielerinnen, sowie Spielerinnen $T-3$ und $T-1$ einen Verlust von einem Euro hinnehmen mussten. Spielerin $T-2$ hat zur Zeit T einen Gewinn von drei Euro und Spielerin $T-4$ hat 15 Euro gewonnen. Also

$$0 = \mathbf{E}[X_T] = \mathbf{E}[15 - 1 + 3 - 1 - (T-4)] = -\mathbf{E}[T] - 20,$$

also $\mathbf{E}[T] = 20$. Interessant ist, dass erwartet werden kann, dass beispielsweise das Muster `ZZKK` schon nach 16 Münzwürfen auftritt.

15.3 Martingalkonvergenzsätze mit abzählbarer Zeitmenge

Wieder ist $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, I abzählbar (hier ist auch erlaubt, dass I dicht in $[0,\infty)$ ist) und $(\mathcal{F}_t)_{t\in I}$ eine Filtration. Wir kennen Konvergenzsätze, etwa das starke Gesetz der großen Zahlen. Martingale konvergieren unter relativ schwachen Voraussetzungen.

Wir beginnen in Proposition 15.25 mit den Doob'schen Ungleichungen. Diese machen Aussagen über die Verteilung von $\sup_{s\le t} X_s$, falls $\mathcal{X} = (X_t)_{t\in I}$ ein (Sub, Super)-Martingal ist.

Lemma 15.24 (Maximal-Ungleichung). *Ist I höchstens abzählbar und $\mathcal{X} = (X_t)_{t\in I}$ ein Sub-Martingal, so ist für $\lambda > 0$*

$$\lambda \mathbf{P}[\sup_{s\le t} X_s \ge \lambda] \le \mathbf{E}[X_t, \sup_{s\le t} X_s \ge \lambda] \le \mathbf{E}[|X_t|, \sup_{s\le t} X_s \ge \lambda].$$

Beweis. Die zweite Ungleichung ist trivial. Für die erste bemerken wir, dass es wegen monotoner Konvergenz (durch Wahl von immer feineren Indexmengen in $[0,t]$) genügt, den diskreten Fall, also etwa $I = \{0,1,2,...\}$, zu betrachten. Wir erinnern an die Definition der Stoppzeit T_B aus Definition 14.29 und setzen

$$T = t \wedge T_{[\lambda;\infty)}.$$

Nach dem Optional Sampling Theorem 15.21 ist

$$\begin{aligned}\mathbf{E}[X_t] \ge \mathbf{E}[X_T] &= \mathbf{E}[X_T; \sup_{s\le t} X_s \ge \lambda] + \mathbf{E}[X_T; \sup_{s\le t} X_s < \lambda] \\ &\ge \lambda \mathbf{P}[\sup_{s\le t} X_s \ge \lambda] + \mathbf{E}[X_t; \sup_{s\le t} X_s < \lambda].\end{aligned}$$

Subtrahieren des letzten Terms ergibt die Ungleichung. □

Proposition 15.25 (Doob'sche L^p-Ungleichung). *Sei I höchstens abzählbar und $\mathcal{X} = (X_t)_{t\in I}$ ein Martingal oder ein positives Sub-Martingal.*

1. *Für $p \ge 1$ und $\lambda > 0$ ist*
$$\lambda^p \mathbf{P}[\sup_{s\le t} |X_s| \ge \lambda] \le \mathbf{E}[|X_t|^p].$$

2. *Für $p > 1$ ist*
$$\mathbf{E}[|X_t|^p] \le \mathbf{E}[\sup_{s\le t} |X_s|^p] \le \Big(\frac{p}{p-1}\Big)^p \mathbf{E}[|X_t|^p].$$

Beweis. Wieder genügt es – wegen monotoner Konvergenz – den Fall $I = \{0, 1, 2, ...\}$ zu betrachten.
1. Nach Proposition 15.8 ist $(|X_t|^p)_{t\in I}$ ein Sub-Martingal und die Behauptung folgt nach Lemma 15.24.
2. Die erste Ungleichung ist klar. Für die zweite Ungleichung beachte, dass nach Lemma 15.24 gilt, dass

$$\lambda \mathbf{P}\{\sup_{s\leq t} |X_s| \geq \lambda\} \leq \mathbf{E}[|X_s|; \sup_{s\leq t} |X_s| \geq \lambda].$$

Also ist für $K > 0$

$$\begin{aligned}
\mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p] &= \mathbf{E}\Big[\int_0^{\sup_{s\leq t}|X_s|\wedge K} p\lambda^{p-1} d\lambda\Big] \\
&= \mathbf{E}\Big[\int_0^K p\lambda^{p-1} 1_{\{\lambda < \sup_{s\leq t}|X_s|\}} d\lambda\Big] \\
&= \int_0^K p\lambda^{p-1}\mathbf{P}(\sup_{s\leq t}|X_s| \geq \lambda) d\lambda \\
&\leq \int_0^K p\lambda^{p-2}\mathbf{E}[|X_t|, \sup_{s\leq t}|X_s| \geq \lambda] d\lambda \\
&= p\mathbf{E}\Big[|X_t| \int_0^{\sup_{s\leq t}|X_s|\wedge K} \lambda^{p-2} d\lambda\Big] \\
&= \frac{p}{p-1}\mathbf{E}[|X_t|(\sup_{s\leq t}|X_s| \wedge K)^{p-1}] \\
&\leq \frac{p}{p-1}\mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p]^{(p-1)/p} \cdot \mathbf{E}[|X_t|^p]^{1/p},
\end{aligned}$$

wobei wir im letzten Schritt die Hölder-Ungleichung verwendet haben. Potenziert man beide Seiten mit p und teilt anschließend durch $\mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p]^{p-1}$, folgt

$$\mathbf{E}[\sup_{s\leq t}(|X_s|)^p] = \lim_{K\to\infty} \mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p] \leq \Big(\frac{p}{p-1}\Big)^p \mathbf{E}[|X_t|^p].$$

□

Für die Martingalkonvergenzsätze ist das Aufkreuzungslemma 15.27 zentral. Bild 15.1 verdeutlicht die Definition einer Aufkreuzung.

Definition 15.26. *Sei I höchstens abzählbar und $\mathcal{X} = (X_t)_{t\in I}$ ein reellwertiger stochastischer Prozess. Für $a < b$ ist eine Aufkreuzung ein Stück Pfad $(X_r)_{s\leq r\leq s'}$ mit $X_s \leq a$ und $X_{s'} \geq b$. Um die Anzahl solcher Aufkreuzungen zu zählen, führen wir Stoppzeiten $0 =: T_0 < S_1 < T_1 < S_2 < T_2 < ...$ durch*

$$\begin{aligned}
S_k &:= \inf\{t \geq T_{k-1} : X_t \leq a\}, \\
T_k &:= \inf\{t \geq S_k : X_t \geq b\}
\end{aligned}$$

mit $\inf \emptyset = \infty$ ein. Die k-te Aufkreuzung zwischen a und b ist hier zwischen S_k und T_k. Weiter ist

$$U_{a,b}^t := \sup\{k : T_k \leq t\}$$

die Anzahl der Aufkreuzungen zwischen a und b bis Zeit t.

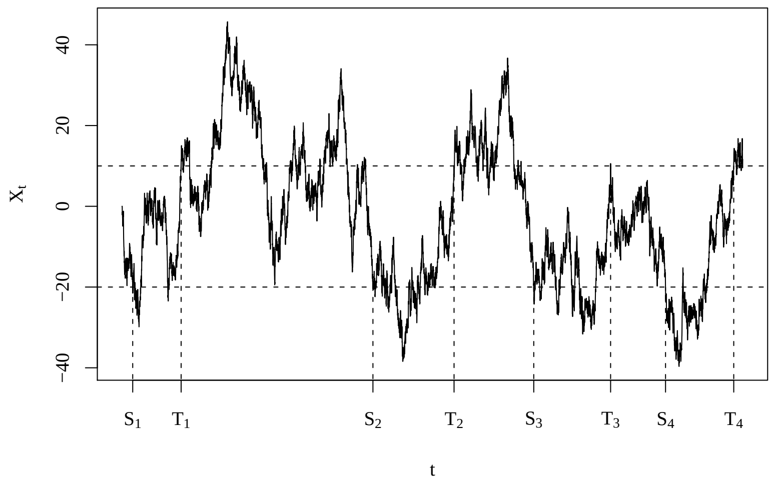


Abbildung 15.1

Eine Illustration der Stoppzeiten $S_1, T_1, S_2, T_2, \ldots$ aus Definition 15.26

Lemma 15.27 (Aufkreuzungslemma). *Sei I höchstens abzählbar und $\mathcal{X} = (X_t)_{t\in I}$ ein Sub-Martingal. Dann ist*

$$\mathbf{E}[U^t_{a,b}] \leq \frac{\mathbf{E}[(X_t - a)^+]}{b-a}.$$

Beweis. Wieder können wir – wegen monotoner Konvergenz – annehmen, dass $I = \{0, 1, 2, \ldots\}$. Da nach Proposition 15.8 mit $\mathcal{X}$ auch $((X_t - a)^+)_{t\in I}$ ein Sub-Martingal ist und die Aufkreuzungen zwischen a und b von $\mathcal{X}$ dieselben sind wie die Aufkreuzungen von $((X_t - a)^+)_{t\in I}$ zwischen 0 und $b - a$, können wir Œ annehmen, dass $\mathcal{X} \geq 0$ und $a = 0$ gilt. Wir definieren den Prozess $\mathcal{H} = (H_t)_{t\in I}$ durch

$$H_t := \sum_{k\geq 1} 1_{\{S_k < t \leq T_k\}},$$

d.h. $H_t = 1$ genau dann, wenn t in einer Aufkreuzung liegt. Da

$$\{H_t = 1\} = \bigcup_{k\geq 1} \{S_k \leq t-1\} \cap \{T_k > t-1\},$$

ist H previsibel.

Gegeben $T_k < \infty$ ist offenbar $X_{T_k} - X_{S_k} \geq b$. Weiter ist in diesem Fall

$$(\mathcal{H}\cdot\mathcal{X})_{T_k} = \sum_{i=1}^{k} \sum_{s=S_i+1}^{T_i} (X_s - X_{s-1}) = \sum_{i=1}^{k} (X_{T_i} - X_{S_i}) \geq kb.$$

Für $t \in \{T_k, \ldots, S_{k+1}\}$ ist $(\mathcal{H}\cdot\mathcal{X})_t = (\mathcal{H}\cdot\mathcal{X})_{T_k}$ und für $t \in \{S_k + 1, \ldots, T_k\}$ ist $(\mathcal{H}\cdot\mathcal{X})_t \geq (\mathcal{H}\cdot\mathcal{X})_{S_k} = (\mathcal{H}\cdot\mathcal{X})_{T_{k-1}}$. Deswegen ist $(\mathcal{H}\cdot\mathcal{X})_t \geq bU^t_{0,b}$. Aus Proposition 15.15 folgt, dass $((1-\mathcal{H})\cdot\mathcal{X})$ ein Sub-Martingal ist, insbesondere $\mathbf{E}[((1-\mathcal{H})\cdot\mathcal{X})_t] \geq 0$. Mit $X_t - X_0 = (1\cdot\mathcal{X})_t = (\mathcal{H}\cdot\mathcal{X})_t + ((1-\mathcal{H})\cdot\mathcal{X})_t$ gilt

$$\mathbf{E}[X_t] \geq \mathbf{E}[X_t - X_0] \geq \mathbf{E}[(\mathcal{H}\cdot\mathcal{X})_t] \geq b\mathbf{E}[U^t_{0,b}].$$

$\square$

Theorem 15.28 (Martingalkonvergenzsatz für Sub-Martingale). *Sei $I \subseteq [0,\infty)$ abzählbar, $\sup I = u \in (0,\infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I}\mathcal{F}_t)$ und $\mathcal{X} = (X_t)_{t\in I}$ ein Sub-Martingal mit $\sup_{t\in I}\mathbf{E}[X_t^+] < \infty$. Dann gibt es eine Nullmenge N, so dass $\mathcal{X}$ außerhalb von N entlang jeder auf- oder absteigenden Folge in I konvergiert.*

Ist insbesondere $I = \{0,1,2,...\}$, $\mathcal{X}$ ein Sub-Martingal mit $\sup_{t\in I}\mathbf{E}[X_t^+] < \infty$, dann existiert eine $\mathcal{F}_\infty$-messbare, integrierbare Zufallsvariable X_∞ und $X_t \xrightarrow{t\to\infty}_{fs} X_\infty$.

Beweis. Wegen Lemma 15.27 ist $\mathbf{P}(U_{a,b}^t < \infty) = 1$ für alle a,b,t. Deshalb ist

$$N := \bigcup_{\substack{a<b\\ a,b\in\mathbb{Q}}} \{\sup_{t\in I} U_{a,b}^t = \infty\}$$

eine Nullmenge. Angenommen es existiert eine auf- oder absteigende Folge $t_1, t_2, ... \in I$, so dass $\mathbf{P}(\liminf_{n\to\infty} X_{t_n} < \limsup_{n\to\infty} X_{t_n}) > 0$. Für $a,b\in\mathbb{Q}$ sei

$$B(a,b) := \{\liminf_{n\to\infty} X_{t_n} < a < b < \limsup_{n\to\infty} X_{t_n}\}.$$

Da $\{\liminf_{n\to\infty} X_{t_n} < \limsup_{n\to\infty} X_{t_n}\} = \bigcup_{a,b\in\mathbb{Q}} B(a,b)$, existieren $a,b\in\mathbb{Q}$ mit $\mathbf{P}(B(a,b)) > 0$. Allerdings gilt $\sup_t U_{a,b}^t = \infty$ auf $B(a,b)$ im Widerspruch dazu, dass N eine Nullmenge ist. Also folgt die fast sichere Konvergenz entlang jeder auf- oder absteigenden Folge.

Sei nun $I = \{0,1,2,...\}$. Da alle X_t $\mathcal{F}_\infty$-messbar sind, ist auch X_∞ $\mathcal{F}_\infty$-messbar. Es bleibt zu zeigen, dass X_∞ integrierbar ist. Nach Fatou's Lemma ist

$$\mathbf{E}[X_\infty^+] \le \sup_{t\in I}\mathbf{E}[X_t^+] < \infty.$$

Außerdem ist, da $\mathcal{X}$ ein Sub-Martingal ist, wieder mit Fatou's Lemma

$$\mathbf{E}[X_\infty^-] \le \liminf_{t\to\infty}\mathbf{E}[X_t^-] = \liminf_{t\to\infty}\big(\mathbf{E}[X_t^+] - \mathbf{E}[X_t]\big) \le \sup_{t\in I}\mathbf{E}[X_t^+] - \mathbf{E}[X_0] < \infty.$$

□

Korollar 15.29 (Martingalkonvergenzsatz für positive Super-Martingale). *Sei $I \subseteq [0,\infty)$ höchstens abzählbar, $\sup I = u \in (0,\infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I}\mathcal{F}_t)$ und $\mathcal{X} = (X_t)_{t\in I}$ ein nicht-negatives Super-Martingal. Dann existiert eine $\mathcal{F}_u$-messbare, integrierbare Zufallsvariable X_u mit $\mathbf{E}[X_u] \le \mathbf{E}[X_0]$ und $X_t \xrightarrow{t\to u}_{fs} X_u$.*

Beweis. Theorem 15.28, angewandt auf das Sub-Martingal $-\mathcal{X}$ liefert den fast sicheren Limes. Mit dem Lemma von Fatou ist außerdem

$$\mathbf{E}[X_u] \le \liminf_{t\to u}\mathbf{E}[X_t] \le \mathbf{E}[X_0].$$

□

Beispiel 15.30 (Konvergenz von Verzweigungsprozessen). Betrachten wir einen kritischen oder sub-kritischen Verzweigungsprozess $\mathcal{Z} = (Z_t)_{t=0,1,2,...}$ aus Beispiel 15.7 (wobei die Nachkommensverteilung nicht degeneriert ist, also nicht $X_i^{(t)} = 1$ fast sicher gilt). Diese sind nicht-negative Super-Martingale, also müssen diese nach Korollar 15.29 fast sicher gegen eine

Zufallsvariable Z_∞ konvergieren. Es muss hierbei $\mathbf{P}(Z_\infty > 0) = 0$ gelten, da sonst die fast sichere Konvergenz verletzt ist. (Eine Population mit einer positiven Anzahl von Individuen hat eine positive Wahrscheinlichkeit, in einer Generation ihre Größe zu verändern.) Also ist

$$Z_t \xrightarrow{t\to\infty} Z_\infty := 0$$

fast sicher.

Im Fall des kritischen Verzweigungsprozesses ist es wichtig einzusehen, dass $(Z_t)_{t=0,1,2,\dots,\infty}$ kein Martingal ist, weil $\mathbf{E}[Z_\infty|\mathcal{F}_t] = \mathbf{E}[0|\mathcal{F}_t] \neq Z_t$ mit positiver Wahrscheinlichkeit gilt.

Ist $\mathcal{Z}$ superkritisch, so ist $(Z_t/\mu^t)_{t=0,1,2,\dots}$ ein nicht-negatives Martingal, das ebenfalls nach obigem Korollar fast sicher konvergiert.

Theorem 15.31 (Konvergenzsatz für gleichgradig integrierbare Martingale). *Sei I abzählbar mit $\sup I = u \in (0,\infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I}\mathcal{F}_t)$ und $\mathcal{X} = (X_t)_{t\in I}$ ein (Super, Sub)-Martingal. Dann sind folgende Aussagen äquivalent:*

1. *$\mathcal{X}$ ist gleichgradig integrierbar.*
2. *Es existiert eine $\mathcal{F}_u$-messbare Zufallsvariable X_u, so dass $(X_t)_{t\in I\cup u}$ ein (Super, Sub)-Martingal ist.*
3. *Es existiert eine $\mathcal{F}_u$-messbare Zufallsvariable X_u mit $X_t \xrightarrow{t\to u}_{fs,L^1} X_u$.*

Beweis. 2. $\Rightarrow$ 1. folgt direkt aus Lemma 12.5. 1. $\Rightarrow$ 3. Wegen Lemma 8.9 ist $\sup_{t\in I}\mathbf{E}[|X_t|] < \infty$. Die fast sichere Konvergenz folgt damit aus Theorem 15.28 und die L^1-Konvergenz damit aus Theorem 8.11.

3. $\Rightarrow$ 2.: Den Beweis, dass $(X_t)_{t\in I\cup\{u\}}$ ein (Super, Sub)-Martingal ist, führen wir exemplarisch für Sub-Martingale, d.h. $\mathbf{E}[\mathbf{E}[X_u|\mathcal{F}_s];A] \geq \mathbf{E}[X_s;A]$ für $A\in\mathcal{F}_s$ und $s\in I$. Wegen der L^1-Konvergenz nach Theorem 12.2.3 ist $\mathbf{E}[|\mathbf{E}[X_t|\mathcal{F}_s] - \mathbf{E}[X_u|\mathcal{F}_s]|] \xrightarrow{t\to u} 0$ und damit gilt für $A\in\mathcal{F}_s$

$$\mathbf{E}[\mathbf{E}[X_u|\mathcal{F}_s];A] = \lim_{t\to\infty}\mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_s];A] \geq \mathbf{E}[X_s;A],$$

d.h. $\mathbf{E}[X_u|\mathcal{F}_s] \geq X_s$ fast sicher. □

Theorem 15.32 (Martingalkonvergenzsatz für L^p-beschränkte Martingale). *Sei I abzählbar mit $\sup I = u \in [0,\infty)$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I}\mathcal{F}_t)$, $p > 1$ und $\mathcal{X} = (X_t)_{t\in I}$ ein L^p-beschränktes Martingal. Dann gibt es eine $\mathcal{F}_u$-messbare Zufallsvariable X_u mit $\mathbf{E}[|X_u|^p] < \infty$, $X_t \xrightarrow{t\uparrow u}_{fs,L^1} X_u$. Außerdem ist $(|X_t|^p)_{t\in I}$ gleichgradig integrierbar.*

Beweis. Wegen Lemma 8.9 ist $\mathcal{X}$ gleichgradig integrierbar. Nach Theorem 15.31 gibt es damit den Limes X_u mit $X_t \xrightarrow{t\uparrow u}_{fs,L^1} X_u$. Nach der Doob'schen Ungleichung aus Proposition 15.25 ist für $t\in I$

$$\mathbf{E}[\sup_{t\in I}|X_t|^p] = \lim_{t\uparrow u}\mathbf{E}[\sup_{s\leq t}|X_s|^p] \leq \lim_{t\uparrow u}\Big(\frac{p}{p-1}\Big)^p\mathbf{E}[|X_t|^p] < \infty.$$

Damit ist $(|X_t|^p)_{t\in I}$ gleichgradig integrierbar nach Beispiel 8.8.3. Nach dem Lemma von Fatou und Lemma 8.9 ist $\mathbf{E}[|X_u|^p] \leq \sup_{t\in I}\mathbf{E}[|X_t|^p] < \infty$ und Theorem 8.11 liefert die Konvergenz in L^p. □

Beispiel 15.33 (Verzweigungsprozess). Sei $\mathcal{Z}$ ein Verzweigungsprozess wie in Beispiel 15.7 und Beispiel 15.30 mit $Z_0 = k$. Die quadratische Variation von $\mathcal{Y} = (Y_t)_{t=0,1,2,...}$, gegeben $Y_t = Z_t/\mu^t$ ist nach Proposition 15.12 gegeben als

$$\begin{aligned}\langle \mathcal{Y}\rangle_t &= \sum_{s=1}^t \frac{1}{\mu^{2s}}\mathbf{E}\Big[\Big(\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} - \mu Z_{s-1}\Big)^2 | \mathcal{F}_{s-1}\Big] \\ &= \sum_{s=1}^t \frac{1}{\mu^{2s}}\mathbf{V}\Big[\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} | Z_{s-1}\Big] \\ &= \sum_{s=1}^t \frac{1}{\mu^{2s}} Z_{s-1}\cdot \mathbf{V}[X_1^{(1)}].\end{aligned}$$

Hat also insbesondere die Nachkommensverteilung ein zweites Moment, gilt also $\mathbf{V}[X_1^{(1)}] =: \sigma^2 < \infty$, folgt

$$\mathbf{V}[Y_t] = \sum_{s=1}^t \frac{1}{\mu^{2s}}\mathbf{E}[Z_s]\cdot\sigma^2 = k\sigma^2\sum_{s=1}^t \frac{1}{\mu^s}.$$

Ist $\mu \leq 1$, ist $\mathcal{Y}$ also nicht L^2-beschränkt, aber für $\mu > 1$ ist $\sup_{t=0,1,2,...}\mathbf{V}[Y_t] < \infty$. Damit gibt es also eine $\mathcal{F}_\infty$-messbare, quadratisch integrierbare Zufallsvariable Y_∞, so dass $(Y_t)_{t=0,1,2,...,\infty}$ ein Martingal ist.

Beispiel 15.34 (Produkt von Zufallsvariablen). Seien $I = \{1,2,...\}$, $X_1, X_2, ...$ nichtnegative, unabhängige, integrierbare Zufallsvariable mit $\mathbf{E}[X_t] = 1, t \in I$ und $S_t := \prod_{s=1}^t X_s$ nach Beispiel 15.3.2 ein Martingal. Nach Korollar 15.29 gibt es damit ein S_∞, so dass $S_t \xrightarrow{t\to\infty}_{fs} S_\infty$. Definiere

$$a_t := \mathbf{E}[\sqrt{X_t}].$$

Wir zeigen nun:

$$\{S_t : t\in I\} \text{ gleichgradig integrierbar} \qquad \Longleftrightarrow \qquad \prod_{t=1}^\infty a_t > 0.$$

Insbesondere gilt dann auch $S_t \xrightarrow{t\to\infty}_{L^1} S_\infty$. Im Beweis setzen wir für $t = 1,2,..$

$$W_t := \prod_{s=1}^t \frac{\sqrt{X_s}}{a_s}.$$

Damit ist $(W_t)_{t=1,2,...}$ ein Martingal. Auch hier folgt, dass es ein W_∞ gibt mit $W_t \xrightarrow{t\to\infty}_{fs} W_\infty$.

'$\Leftarrow$': Wegen der Jensen'schen Ungleichung ist $a_t^2 = (\mathbf{E}[\sqrt{X_t}])^2 \leq \mathbf{E}[X_t] = 1$, also $a_t \leq 1$. Es gilt

$$\sup_{t\in I}\mathbf{E}[W_t^2] = \sup_{t\in I}\mathbf{E}\Big[\prod_{s=1}^t \frac{X_s}{a_s^2}\Big] = \sup_{t\in I}\prod_{s=1}^t \frac{\mathbf{E}[X_s]}{a_s^2} \leq \frac{1}{\Big(\prod_{s=1}^\infty a_s\Big)^2} < \infty.$$

Damit ist $(W_t)_{t\in I}$ ein L^2-beschränktes Martingal, Nach Theorem 15.32 ist $\{W_t^2 : t\in I\}$ gleichgradig integrierbar. Daraus folgt auch die gleichgradige Integrierbarkeit von $\{S_t : t \in I\}$.

'⇒': Nehmen wir an, dass $\prod_{s=1}^{\infty} = 0$. Da W_t einen fast sicheren, endlichen Limes hat, muss $S_t = \prod_{s=1}^t X_s \xrightarrow{t\to\infty}_{fs} 0$ gelten. Falls $\{S_t : t \in I\}$ gleichgradig integrierbar wäre, wäre $0 = \mathbf{E}[S_\infty] = \lim_{t\to\infty} \mathbf{E}[S_t] = 1$, also ein Widerspruch.

Theorem 15.35 (Konvergenz von bedingten Erwartungswerten).

1. *Sei $I \subseteq [0,\infty)$ abzählbar mit $\sup I = u \in (0,\infty]$, $(\mathcal{F}_t)_{t\in I}$ eine Filtration und $\mathcal{F}_u = \sigma(\bigcup_{t\in I}\mathcal{F}_t)$. Dann gilt für $X \in \mathcal{L}^1$, dass*
$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t\uparrow u}_{fs,L^1} \mathbf{E}[X|\mathcal{F}_u].$$

2. *Sei $I \subseteq (-\infty,\infty)$ abzählbar mit $\inf I = u \in [-\infty,\infty)$, $(\mathcal{F}_t)_{t\in I}$ eine Filtration und $\mathcal{F}_u = \bigcap_{t\in I}\mathcal{F}_t$. Dann gilt für $X \in \mathcal{L}^1$, dass*
$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t\downarrow u}_{fs,L^1} \mathbf{E}[X|\mathcal{F}_u].$$

Beweis. Wir zeigen nur 1., da der Beweis von 2. analog verläuft. Mit $\mathbf{E}[|\mathbf{E}[X|\mathcal{F}_t]|] \leq \mathbf{E}[|X|] < \infty$ konvergiert nach Theorem 15.28 das Martingal $(\mathbf{E}[X|\mathcal{F}_t])_{t\in I}$ fast sicher. Die L^1-Konvergenz folgt mit Theorem 15.31 und Lemma 12.5. Der Grenzwert X_u kann hierbei $\mathcal{F}_u$-messbar gewählt werden. Wir werden nun zeigen, dass $X_u = \mathbf{E}[X|\mathcal{F}_u]$, woraus die Behauptung folgt.

Klar ist, dass $\mathbf{E}[\mathbf{E}[X|\mathcal{F}_t], A] = \mathbf{E}[X, A]$ für alle $A \in \mathcal{F}_s$ und $s \leq t$ gilt. Mit $t \uparrow u$ is damit $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ für alle $A \in \mathcal{F}_s$ und mit $s \uparrow u$ gilt auch $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ für alle $A \in \mathcal{F}_u$. Da X_u nach $\mathcal{F}_u$-messbar ist, gilt damit $X_u = \mathbf{E}[X|\mathcal{F}_u]$. □

Wir kommen nun zu Rückwärtsmartingalen, das sind Martingale mit nach unten unbeschränkter Indexmenge $I \subseteq (-\infty, 0]$. Diese konvergieren unter sehr schwachen Voraussetzungen.

Theorem 15.36 (Martingalkonvergenzsatz für Rückwärtsmartingale). *Sei $I \subseteq (\infty, 0]$ diskret, $\inf I = u \in (-\infty, 0]$, $\mathcal{F}_u = \bigcap_{t\in I}\mathcal{F}_t$ und $\mathcal{X} = (X_t)_{t\in I}$ ein Sub-Martingal. Dann sind äquivalent*

1. *Es gibt eine $\mathcal{F}_u$-messbare, integrierbare Zufallsvariable X_u mit $X_t \xrightarrow{t\downarrow u}_{fs,\ L^1} X_u$*
2. $\inf_{t\in I}\mathbf{E}[X_t] > -\infty$.

Dann ist auch $(X_t)_{t\in I\cup\{u\}}$ ein Sub-Martingal. Insbesondere konvergiert jedes Rückwärtsmartingal fast sicher und in L^1.

Beweis. Ohne Einschränkung sei $I = \{..., -2, -1, 0\}$ und $u = -\infty$.
'1. ⇒ 2.': Aus der Konvergenz im Mittel folgt
$$\inf_{t\in I}\mathbf{E}[X_t] = \lim_{t\to-\infty}\mathbf{E}[X_t] = \mathbf{E}[X_{-\infty}] > -\infty.$$

'2. ⇒ 1.': Die fast sichere Konvergenz folgt wie im Beweis von Theorem 15.28, wobei die Bedingung $\sup_{t\in I}\mathbf{E}[X_t^+] < \infty$ wegen $I \subseteq (-\infty, 0]$ durch $\inf_{t\in I}\mathbf{E}[X_t^-] < \infty$ ersetzt werden muss. Weiter definieren wir für $t = ..., -2, .1, 0$
$$Y_t := \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] \geq 0.$$

Dann gilt

$$\mathbf{E}\Big[\sum_{t=0}^{-\infty} Y_t\Big] = \mathbf{E}[X_0] - \inf_{t\in I} \mathbf{E}[X_t] < \infty.$$

Damit ist $\sum_{t=0}^{-\infty} Y_t < \infty$ fast sicher, und wir definieren

$$A_t = \sum_{s\leq t} Y_s, \qquad M_t = X_t - A_t$$

Nun ist $(A_t)_{t\in I}$ wegen $\mathbf{E}[A_0] < \infty$ gleichgradig integrierbar, und $(M_t)_{t\in I}$ ist wegen Lemma 12.5 gleichgradig integrierbar. Damit ist $\mathcal{X}$ gleichgradig integrierbar, und die L^1-Konvergenz folgt. Der Beweis, dass $(X_t)_{t\in I\cup\{-\infty\}}$ ein Sub-Martingal ist, verläuft analog zum Beweis in 15.31. □

Beispiel 15.37 (Das starke Gesetz der großen Zahlen). Seien $X_1, X_2, ... \in L^1$ unabhängig identisch verteilt. Für $t \in \{..., -2, -1\}$ setzen wir wie im Beispiel 15.3.2

$$S_t := \frac{1}{|t|} \sum_{s=1}^{|t|} X_s$$

und $\mathcal{F}_t = \sigma(..., S_{t-1}, S_t) = \sigma(S_t, X_{t+1}, X_{t+2}, ...)$. Dann ist $(S_t)_{t\in I}$ ein Rückwärtsmartingal mit $S_t = \mathbf{E}[X_1|\mathcal{F}_t]$. Nach Theorem 15.36 konvergiert S_t fast sicher und in L^1 gegen eine Zufallsgröße $S_{-\infty}$. Diese ist messbar bzgl. $\mathcal{F}_{-\infty}$, jedoch auch bzgl. $\mathcal{T}(X_1, X_2, ...)$, der terminalen σ-Algebra der Familie $\{X_1, X_2, ...\}$. Da diese σ-Algebra nach dem Kolmogoroff'schen 0-1-Gesetz trivial ist, ist $S_{-\infty}$ fast sicher konstant. Da $(S_t)_{t\in I\cup\{-\infty\}}$ ein Martingal ist, folgt also

$$\frac{1}{|t|} \sum_{s=1}^{|t|} X_s = S_t \xrightarrow{t\to-\infty}_{fs, L^1} S_{-\infty} = \mathbf{E}[S_{-\infty}] = \mathbf{E}[S_{-1}] = \mathbf{E}[X_1].$$

Die fast sichere Konvergenz ist jedoch genau die Aussage des Gesetzes der großen Zahlen.

Wir kommen nun noch zu einer Anwendung der Martingalkonvergenzsätze, einer Verbesserung des Lemmas von Borel-Cantelli, Theorem 9.8. Hierzu benötigen Wir ein Lemma.

Lemma 15.38 (Konvergenz und wachsender Prozess). *Sei $\mathcal{M} = (M_t)_{t=0,1,2,...}$ ein L^2-integrierbares Martingal, wobei $|M_t - M_{t-1}| \leq K$ für ein K und alle $t = 1, 2, ...$ gilt. Dann gibt es eine Nullmenge N, so dass*

$$\{\langle\mathcal{M}\rangle_\infty < \infty\} \subseteq \{\lim_{t\to\infty} M_t \text{ existiert}\} \cup N,$$
$$\{\langle\mathcal{M}\rangle_\infty = \infty\} \subseteq \{\lim_{t\to\infty} M_t/\langle\mathcal{M}\rangle_t = 0\} \cup N.$$

Beweis. Wir beginnen mit der ersten Aussage. Zunächst ist für jedes $k = 1, 2, 3, ...$ die zufällige Zeit

$$T_k := \inf\{t : \langle\mathcal{M}\rangle_t > k\}$$

eine Stoppzeit. Daraus folgt bereits

$$\{\langle\mathcal{M}\rangle_\infty < \infty\} = \bigcup_{k=1}^{\infty} \{T_k = \infty\}. \tag{15.7}$$

Weiter ist damit der gestoppte Prozess $(\langle\mathcal{M}\rangle_{t\wedge T_k})_{t=0,1,2,...}$ prävisibel, denn für $A \in \mathcal{B}(\mathbb{R})$ gilt

$$\{\langle\mathcal{M}\rangle_{t\wedge T_k} \in A\} = \big(\{T_k > t-1\} \cap \{\langle\mathcal{M}\rangle_t \in B\}\big) \cup \bigcup_{s=0}^{t-1}\{T_k = s, \langle\mathcal{M}\rangle_s \in A\} \in \mathcal{F}_{t-1}.$$

Betrachten wir nun die Martingale $(\mathcal{M}^{T_k})^2 - \langle\mathcal{M}\rangle^{T_k} = (\mathcal{M}^2 - \langle\mathcal{M}\rangle)^{T_k}$ für $k = 1, 2, ...$ Es ist $\langle\mathcal{M}^{T_k}\rangle = \langle\mathcal{M}\rangle^{T_k}$ und $\langle\mathcal{M}\rangle^{T_k}$ ist durch $k + K^2$ beschränkt. Also ist $\mathcal{M}^{T_k}$ in L^2 beschränkt und konvergiert damit fast sicher. Auf einer Menge $\{T_k = \infty\}$ konvergiert jedoch $\mathcal{M}^{T_k}$ genau dann, wenn $\mathcal{M}$ konvergiert. Zusammen mit (15.7) folgt die Aussage.

Für die zweite Aussage betrachten wir das Martingal $\mathcal{X} := (1 + \langle\mathcal{M}\rangle)^{-1} \cdot \mathcal{M}$. Da $(1 + \langle\mathcal{M}\rangle)^{-1}$ beschränkt ist und $\mathcal{M}$ ein L^2-integrierbares Martingal ist, ist $\mathcal{X}$ ein L^2-integrierbares Martingal. Außerdem ist nach Beispiel 15.16

$$\begin{aligned}
\langle\mathcal{X}\rangle_t &= \Big(\frac{1}{(1+\langle\mathcal{M}\rangle)^2} \cdot \langle\mathcal{M}\rangle\Big)_t = \sum_{s=1}^t \frac{1}{(1+\langle\mathcal{M}\rangle_s)^2}(\langle\mathcal{M}\rangle_s - \langle\mathcal{M}\rangle_{s-1}) \\
&\le \sum_{s=1}^t \frac{1}{(1+\langle\mathcal{M}\rangle_s)(1+\langle\mathcal{M}\rangle_{s-1})}(\langle\mathcal{M}\rangle_s - \langle\mathcal{M}\rangle_{s-1}) = \sum_{s=1}^t \frac{1}{1+\langle\mathcal{M}\rangle_{s-1}} - \frac{1}{1+\langle\mathcal{M}\rangle_s} \\
&= 1 - \frac{1}{1+\langle\mathcal{M}\rangle_t}.
\end{aligned}$$

Damit konvergiert nach 1. das Martingal $\mathcal{X}$, d.h. insbesondere

$$\sum_{s=1}^\infty \frac{M_s - M_{s-1}}{1+\langle\mathcal{M}\rangle_s} < \infty.$$

Nun liefert das Kronecker-Lemma 9.24, dass

$$\frac{\sum_{s=1}^t M_s - M_{s-1}}{\langle\mathcal{M}\rangle_t} \xrightarrow{t\to\infty} 0$$

auf $\{\langle\mathcal{M}\rangle_\infty = \infty\}$. □

Theorem 15.39 (Erweiterung des Borel-Cantelli Lemmas). *Sei $A_t \in \mathcal{F}_t$, $t = 0, 1, 2, ...$ sowie*

$$X_s := \mathbf{P}(A_s|\mathcal{F}_{s-1}).$$

1. *Auf $\sum_{t=1}^\infty X_t < \infty$ treten nur endlich viele der A_t ein, d.h.*

$$\Big\{\sum_{t=1}^\infty X_t < \infty\Big\} \subseteq \Big\{\sum_{t=1}^\infty 1_{A_t} < \infty\Big\}.$$

2. *Auf $\sum_{t=1}^\infty X_t = \infty$ gilt $\sum_{t=1}^\infty 1_{A_t}/\sum_{t=1}^\infty X_t = 1$, also*

$$\Big\{\sum_{t=1}^\infty X_t = \infty\Big\} \subseteq \Big\{\sum_{t=1}^\infty 1_{A_t}/\sum_{t=1}^\infty X_t = 1\Big\} \subseteq \Big\{\sum_{t=1}^\infty 1_{A_t} = \infty\Big\}.$$

Bemerkung 15.40 (Erweiterung). Das Borel-Cantelli-Lemma aus Theorem 9.8 kann nun einfach hergeleitet werden. Ist nämlich

$$\mathbf{E}\Big[\sum_{t=1}^{\infty} X_t\Big] = \sum_{t=1}^{\infty} \mathbf{P}(A_t) < \infty,$$

so gilt $\sum_{t=1}^{\infty} X_t < \infty$ fast sicher. Die Aussage liefert nun, dass höchstens endlich viele der A_n eintreten. Falls weiter $A_1, A_2, ...$ unabhängig sind, so setzen wir $\mathcal{F}_t = \sigma(A_1, ..., A_t)$ und damit $X_s = \mathbf{E}[1_{A_s}|\mathcal{F}_{s-1}] = \mathbf{P}(A_s)$. Gilt nun $\sum_{t=1}^{\infty} \mathbf{P}(A_t) = \infty$, treten also unendlich viele der A_n ein.

Beweis. Wir betrachten das Martingal $\mathcal{M}$ mit

$$M_t = \sum_{s=1}^{t} 1_{A_s} - X_s.$$

Es gilt

$$\langle \mathcal{M} \rangle_t = \sum_{s=1}^{t} \mathbf{E}[1_{A_s}^2 - X_s^2|\mathcal{F}_{s-1}] = \sum_{s=1}^{t} X_s(1 - X_s) \leq \sum_{s=1}^{t} X_s.$$

Ist nun $\sum_{t=1}^{\infty} X_t < \infty$, so konvergiert $\mathcal{M}$ nach Lemma 15.38.1. Also gilt auch $\sum_{t=1}^{\infty} 1_{A_t} < \infty$.
Ist nun $\sum_{t=1}^{\infty} X_t = \infty$ und $\langle \mathcal{M} \rangle_\infty < \infty$, so konvergiert $\mathcal{M}$ und die Behauptung ist klar.
Ist nun $\sum_{t=1}^{\infty} X_t = \infty$ und $\langle \mathcal{M} \rangle_\infty = \infty$, so gilt $M_t/\langle \mathcal{M} \rangle_t \xrightarrow{t\to\infty} 0$ nach Lemma 15.38.2. Daraus folgt dann

$$\Big|\frac{\sum_{s=1}^{t} 1_{A_s}}{\sum_{s=1}^{t} X_s} - 1\Big| = \Big|\frac{M_t}{\sum_{s=1}^{t} X_s}\Big| \leq \Big|\frac{M_t}{\langle \mathcal{M} \rangle_t}\Big| \xrightarrow{t\to\infty} 0.$$

□

15.4 Der zentrale Grenzwertsatz für Martingale

Der zentrale Grenzwertsatz aus Abschnitt 11.2 gibt die Konvergenz einer Summe von *unabhängigen* Zufallsvariablen – geeignet transformiert – gegen eine Normalverteilung. Nun behandeln wir den Fall einer Folge von Martingalen $\mathcal{M}^1 = (M_t^1)_{t=0,1,2,...}, \mathcal{M}^2 = (M_t^2)_{t=0,1,2,...}, ...$, jeweils gestartet in 0, die wir mittels $X_t^n := M_t^n - M_{t-1}^n, t = 1, 2, ...$ wieder als Summe schreiben können, da ja nun $M_t^n = X_1^n + \cdots + X_t^n$ gilt. Man beachte nun, dass die Familie $X_1^n, X_2^n, ...$ nicht unabhängig sein müssen. Dennoch können wir – unter geeigneten Voraussetzungen die Konvergenz in Verteilung gegen eine normalverteilte Zufallsvariable zeigen.

Theorem 15.41 (Zentraler Grenzwertsatz für Martingale). *Sei $I^n = \{0, 1, 2, ..., t_n\}$ und $\mathcal{M}^n = (M_t^n)_{t\in I^n}$ ein Martingal mit $M_0^n = 0$ bezüglich einer Filtration $\mathcal{F}^n = (\mathcal{F}_t^n)_{t\in\mathcal{I}^n}$, $n = 1, 2, ...$ Für $X_t^n := M_t^n - M_t^{n-1}$ (mit $t = 1, ..., t_n$) gelte*

$$\mathbf{E}[\max_{1\leq s\leq t_n} |X_s^n|] \xrightarrow{n\to\infty} 0, \tag{15.8}$$

$$\sum_{s=1}^{t_n} (X_s^n)^2 \xrightarrow{n\to\infty}_p \sigma^2 > 0. \tag{15.9}$$

Dann ist $M_{t_n}^n \xRightarrow{n\to\infty} X$ mit $X \sim N(0, \sigma^2)$.

Wir benötigen im Beweis des Theorems zwei Lemmas.

Lemma 15.42 (Konvergenz von Produkten von Zufallsvariablen). *Seien $U_1, U_2, ..., T_1, T_2, ...$ Zufallsvariablen, die den folgenden Bedingungen genügen:*

1. $U_n \xrightarrow{n\to\infty}_p u$,
2. $(T_n)_{n=1,2,...}$ *und* $(T_n U_n)_{n=1,2,...}$ *sind uniform integrierbar,*
3. $\mathbf{E}[T_n] \xrightarrow{n\to\infty} 1$,

Dann gilt $\mathbf{E}[T_n U_n] \xrightarrow{n\to\infty} u$.

Beweis. Siehe Übung. □

Lemma 15.43 (Abschätzung der Exponentialfunktion).
1. *Es gibt ein $C > 0$ und eine Funktion r mit $|r(x)| \leq C|x^3|$, so dass*
$$\exp(ix) = (1 + ix)\exp(-x^2/2 + r(x))$$
für alle $x \in \mathbb{R}$ gilt.
2. *Es gilt $|1 + ix| \leq e^{x^2/2}$ für alle $x \in \mathbb{R}$.*

Beweis. Siehe Übung. 1. Es genügt, die Behauptung für kleine $|x|$ zu zeigen, da sie für große $|x|$ trivial ist. Mit Hilfe von Lemma 11.12 schreiben wir

$$\begin{aligned}
&\Big|\exp(ix) - (1 + ix)\exp(-x^2/2)\Big| \\
&\qquad = \Big|\exp(ix) - 1 - ix + x^2/2 - (1 + ix)(\exp(-x^2/2) - 1 + x^2/2) + ix^3/3\Big| \\
&\qquad \leq \Big|\exp(ix) - 1 - ix + x^2/2\Big| + |1 + ix| \cdot \Big|\exp(-x^2/2) - 1 + x^2/2\Big| + \big|x^3/3\big| \\
&\qquad \leq \tfrac{|x^3|}{6} + |1 + ix| \cdot \Big(\tfrac{|x^2|}{2} \wedge \tfrac{|x^4|}{8}\Big) + \tfrac{|x^3|}{3} \leq |x^3|
\end{aligned}$$

für alle x. Daraus folgt die Behauptung für kleine $|x|$, und damit ist 1. bewiesen. Für 2. genügt es, $|1 + ix|^2 = 1 + x^2 \leq e^{x^2}$ zu schreiben und die Wurzel zu ziehen. □

Beweis von Theorem 15.41. Zunächst definieren wir

$$Z_s^n := X_s^n 1_{\sum_{r=1}^{s-1}(X_r^n)^2 \leq 2\sigma^2}$$

sowie $N_t^n := \sum_{s=1}^t Z_s^n$. Dann ist $(N_t^n)_{t=1,2,...}$ ein $(\mathcal{F}_t^n)_{t\in I^n}$-Martingal, denn

$$\mathbf{E}[N_t^n - N_{t-1}^n|\mathcal{F}_{t-1}^n] = \mathbf{E}[Z_t^n|\mathcal{F}_{t-1}^n] = 1_{\sum_{r=1}^{s-1}(X_r^n)^2 \leq 2\sigma^2} \cdot \mathbf{E}[X_t^n|\mathcal{F}_{t-1}^n] = 0,$$

da $M_t^n = X_1^n + \cdots + X_t^n$. Nun ist

$$\begin{aligned}
\mathbf{P}(\max_{t=1,...,t_n} |M_t^n - N_t^n| > 0) &= \mathbf{P}(M_t^n \neq N_t^n \text{ für ein } t \in I^n) \\
&= \mathbf{P}(X_t^n \neq Z_t^n \text{ für ein } t \in I^n) \\
&= \mathbf{P}\Big(\sum_{s=1}^{t_n}(X_s^n)^2 > 2\sigma^2\Big) \xrightarrow{n\to\infty} 0,
\end{aligned} \tag{15.10}$$

wobei die Konvergenz aus (15.9) folgt. Nun gilt also $M^n_{t_n} - N^n_{t_n} \xrightarrow{n\to\infty}_p 0$, also genügt es nach dem Satz von Slutzky, Korollar 10.9, $N^n_{t_n} \overset{n\to\infty}{\Longrightarrow} X \sim N(0,\sigma^2)$ zu zeigen. Hierfür werden wir für beliebiges $\lambda \in \mathbb{R}$

$$\mathbf{E}[e^{i\lambda N^n_{t_n}}] \xrightarrow{n\to\infty} e^{-i\lambda^2\sigma^2/2}$$

zeigen. Mit der Funktion r aus Lemma 15.43 gilt nun

$$\mathbf{E}[e^{i\lambda N^n_{t_n}}] = \prod_{s=1}^{t_n}(1+i\lambda Z^n_s)\cdot \exp\Big(-\tfrac{\lambda^2}{2}\sum_{s=1}^{t_n}(Z^n_s)^2 + \sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big).$$

Wir setzen nun

$$T_n := \prod_{s=1}^{t_n}(1+i\lambda Z^n_s), \qquad U_n := \exp\Big(-\tfrac{\lambda^2}{2}\sum_{s=1}^{t_n}(Z^n_s)^2 + \sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big)$$

und zeigen, dass für diese Zufallsvariablen die Voraussetzungen von Lemma 15.42 gelten (mit $u = e^{-\lambda^2\sigma^2/2}$). Für 1. ist zunächst wegen (15.10)

$$\lim_{n\to\infty}\sum_{s=1}^{t_n}(Z^n_s)^2 = \lim_{n\to\infty}\sum_{s=1}^{t_n}(X^n_s)^2 = \sigma^2.$$

Weiter ist mit C aus Lemma 15.43

$$\begin{aligned}\Big|\sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big| &\le C\cdot|\lambda^3|\cdot\sum_{s=1}^{t_n}|Z^n_s|^3 \le C\cdot|\lambda^3|\cdot\sum_{s=1}^{t_n}|X^n_s|^3\\ &\le C\cdot|\lambda^3|\cdot\max_{1\le s\le t_n}|X^n_s|\cdot\sum_{s=1}^{t_n}|X^n_s|^2 \xrightarrow{n\to\infty} 0,\end{aligned}$$

wobei die Konvergenz aus (15.8) und (15.9) folgt.
Für 2. gilt zunächst $|T_nU_n| = |e^{i\lambda N^n_{t_n}}| = 1$, woraus bereits die uniforme Integrierbarkeit von $(T_nU_n)_{n\in I^n}$ folgt. Für die uniforme Integrierbarkeit von $(T_n)_{n\in I^n}$ definieren wir

$$J_n := \inf\Big\{s \le t_n : \sum_{r=1}^{s}(X^n_r)^2 > 2\sigma^2\Big\}\wedge t_n$$

und schreiben

$$\begin{aligned}|T_n| &= \prod_{s=1}^{J_n-1}|1+i\lambda Z^n_s|\cdot|1+i\lambda Z^n_{J_n}| \le \exp\Big(\tfrac{\lambda^2}{2}\sum_{s=1}^{J_n-1}(X^n_s)^2\Big)(1+\lambda|X^n_{J_n}|)\\ &\le \exp(\lambda^2\sigma^2)\cdot(1+|\lambda|\cdot\max_{1\le s\le t_n}|X^n_s|).\end{aligned}$$

Da $\max_{1\le s\le t_n}|X^n_s| \xrightarrow{n\to\infty}_{L^1} 0$, ist insbesondere die Familie $(\max_{1\le s\le t_n}|X^n_s|)_{n=1,2,...}$ uniform integrierbar, woraus die uniforme Integrierbarkeit von $(T_n)_{n=1,2,...}$ folgt.

Wir kommen nun zu 3., indem wir $\mathbf{E}[T_n] = 1$ zeigen. Wegen $\mathbf{E}[Z^n_s|\mathcal{F}^n_{s-1}] = 0$ für alle $s = 1,...,t_n$ ist nämlich

$$\begin{aligned}\mathbf{E}[T_n] &= \mathbf{E}\Big[\prod_{s=1}^{t_n}(1+i\lambda Z^n_s)\Big]\\ &= \mathbf{E}\Big[(1+i\lambda Z^n_1)\cdot\mathbf{E}[(1+i\lambda Z^n_2)\cdots\mathbf{E}[1+\lambda Z^n_{t_n}|\mathcal{F}^n_{t_n-1}]\cdots|\mathcal{F}^n_1]\Big] = 1.\end{aligned}$$

Nun folgt die behauptung direkt mit Lemma 15.43. □

Beispiel 15.44. 1. Seien $X_1, X_2, ...$ unabhängige, identisch verteilte, reellwertige Zufallsvariable mit $\mathbf{E}[X_1] = 0$ und endlicher Varianz $\mathbf{V}[X_1] = \sigma^2$. Bekannterweise ist dann $\mathcal{M}^n = (M_t^n)_{t=0,1,2,...}$ mit

$$M_t^n = \tfrac{1}{\sqrt{n}} \sum_{s=1}^t X_t$$

ein Martingal und

$$M_n^n \xRightarrow{n\to\infty} X \sim N(0, \sigma^2).$$

Dies lässt sich auch mittels Theorem 15.41 zeigen: zunächst stellen wir fest, dass $\int_0^\infty t\mathbf{P}(|X_1| > t)dt < \infty$ wegen des endlichen zweiten Moments ist. Damit ist $\mathbf{P}(|X_1| > t) = o(1/t^2)$ für $t \to \infty$, lässt sich also schreiben als $\mathbf{P}(|X_1| > t) = a(t)/t^2$ mit $a(t) \xrightarrow{t\to\infty} 0$. Daraus folgt

$$\begin{aligned}\mathbf{E}[\max_{1\le s\le n} |X_s|/\sqrt{n}] &= \int_0^\infty \mathbf{P}(\max_{1\le s\le n} |X_s| > t\sqrt{n})dt = \int_0^\infty 1 - (1 - \mathbf{P}(|X_1| > t\sqrt{n}))^n dt \\ &= \int_0^\infty 1 - \Big(1 - \frac{a(t\sqrt{n})}{t^2 n}\Big)^n dt \xrightarrow{n\to\infty} 0\end{aligned}$$

wegen majorisierter Konvergenz. Weiter gilt mit dem Gesetz großer Zahlen, dass

$$\frac{1}{n}\sum_{s=1}^n X_s^2 \xrightarrow{n\to\infty}_{fs} \sigma^2.$$

Also sind die Voraussetzungen von Theorem 15.41 erfüllt.

2. Wir bringen noch ein Beispiel einer Folge von Martingalen, die auf Summen von abhängigen Zufallsvariablen führen. Hierfür erinnern wir an das stochastische Integral aus Definition 15.14. Seien $Y_1, Y_2, ...$ unabhänige, identisch verteilte, beschränkte Zufallsvariable mit $\mathbf{E}[Y_1] = 0$ und $\mathbf{V}[Y_1] = 1$ sowie $\mathcal{H} = (H_t)_{t=0,1,2,...}$ und $\mathcal{M}^n = (M_t^n)_{t=0,1,2,...}$ gegeben als

$$H_s = \frac{1}{s-1}(Y_1^2 + \cdots + Y_{s-1}^2), \qquad M_t^n = \frac{1}{\sqrt{n}}\sum_{s=1}^t Y_s.$$

Dann ist

$$(\mathcal{H}\cdot\mathcal{M}^n)_t = \frac{1}{\sqrt{n}}\sum_{s=1}^t Y_s \frac{1}{s-1}\sum_{r=1}^{s-1} Y_r^2$$

ein Martingal mit

$$X_t^n := (\mathcal{H}\cdot\mathcal{M}^n)_t - (\mathcal{H}\cdot\mathcal{M}^n)_{t-1} = \frac{1}{\sqrt{n}} Y_t \frac{1}{t-1}\sum_{r=1}^{t-1} Y_r^2.$$

(Man beachte, dass $(X_1^n, X_2^n, ...)$ keine unabhängige Familie ist.) Nun gilt (15.8) wegen der Beschränktheit von $Y_1, Y_2, ...$ Weiter berechnen wir

$$\sum_{s=1}^n (X_s^n)^2 = \frac{1}{n}\sum_{s=1}^n Y_s^2\Big(\frac{1}{s-1}\sum_{r=1}^{s-1} Y_r^2\Big)^2 \xrightarrow{n\to\infty} 1,$$

woraus nun $(\mathcal{H}\cdot\mathcal{M}^n)_n \xRightarrow{n\to\infty} X \sim N(0,1)$ folgt.

15.5 Eigenschaften von Martingalen in stetiger Zeit

Wir werden nun Ergebnisse von Martingalen mit abzählbarer Indexmenge auf den Fall einer überabzählbaren Indexmenge, $I = [0, \infty)$, übertragen. Zentral ist hierbei Theorem 15.45, in dem wir sehen werden, dass es zu sehr vielen Sub-Martingalen eine rechtsstetige Modifikation gibt.

Theorem 15.45 (Regularisierung von Martingalen in stetiger Zeit). *Sei $I = [0, \infty)$ und $\mathcal{X} = (X_t)_{t \in I}$ ein Sub-Martingal. Weiter ist $\mathcal{Y} = (Y_t)_{t \in I \cap \mathbb{Q}}$ mit $Y_t = X_t$ für $t \in I \cap \mathbb{Q}$. Dann gilt mit $(\mathcal{G}_t)_{t \in I}$ aus Lemma 14.25.*

1. *Es gibt eine Nullmenge N, so dass $Y_t^+ := \lim_{s \downarrow t} Y_t$ für alle $t \in I$ außerhalb N existiert. Der Prozess $\mathcal{Z} = (Z_t)_{t \in I}$ mit $Z_t = 1_{N^c} Y_t^+$ ist ein $(\mathcal{G}_t)_{t \in I}$-Sub-Martingal.*

2. *Falls $(\mathcal{F}_t)_{t \in I}$ rechtsstetig ist, dann hat $\mathcal{X}$ genau dann eine Modifikation mit Pfaden in $\mathcal{D}_{\mathbb{R}}([0, \infty))$, wenn $t \mapsto \mathbf{E}[X_t]$ rechtsstetig ist.*

Beweis. Da $(|\mathcal{Y}_t|)_{t \in I \cap \mathbb{Q}}$ ein Sub-Martingal ist, ist $\sup_{t \leq \tau} \mathbf{E}[|Y_t|] < \infty$ für $\tau < \infty$. Also gibt es nach Theorem 15.28 für jedes $t \in I$ die Grenzwerte $Y_{t\pm}, t \in I$ außerhalb einer Nullmenge N. Damit ist $(Z_t)_{t \in I}$ mit $Z_t = 1_{N^c} Y_t^+$ rechtsstetig mit linksseitigen Grenzwerten. Außerdem ist Z_t messbar bezüglich $\sigma(\mathcal{F}_t, \mathcal{N})^+, t \in I$.

Wir zeigen nun, dass $(Z_t)_{t \in I}$ ein Sub-Martingal ist. Seien $s < t$ und $s_n \downarrow s$, sowie $t_n \downarrow t$ (und $s_n \leq t$, $n = 1, 2, ...$). Dann gilt offenbar $Y_{s_m} \leq \mathbf{E}[Y_{t_n} | \mathcal{F}_{s_m}]$ für alle m, n. Damit gilt $Z_s \leq \mathbf{E}[Y_{t_n} | \mathcal{F}_{s+}]$ nach Theorem 15.35. Da $\sup_n \mathbf{E}[Y_{t_n}] < \infty$, ist das Sub-Martingal $(Y_{t_n})_{n=1,2,...}$ nach Theorem 15.36 gleichgrading integrierbar mit $Y_{t_n} \xrightarrow{n \to \infty}_{fs, L^1} Z_t$, und damit $\mathbf{E}[Y_{t_n} | \mathcal{F}_{s+}] \xrightarrow{n \to \infty}_{fs, L^1} \mathbf{E}[Z_t | \mathcal{F}_{s+}]$. Daraus folgt $Z_s \leq \mathbf{E}[Z_t | \mathcal{F}_{s+}] = \mathbf{E}[Z_t | \mathcal{G}_s]$.

2. Mit derselben Notation ist für $t \in I$ und $t_n \downarrow t$ mit $t_1, t_2, ... \in \mathbb{Q}$,

$$\mathbf{E}[X_{t_n}] = \mathbf{E}[Y_{t_n}], \qquad X_t \leq \mathbf{E}[Y_{t_n} | \mathcal{F}_t].$$

Wegen $t_n \downarrow t$ ist $\lim_{s \downarrow t} \mathbf{E}[X_s] = \mathbf{E}[Z_t]$. Weiter ist wegen der Rechtsstetigkeit von $(\mathcal{F}_t)_{t \in I}$ und Theorem 15.36 $X_t \leq \mathbf{E}[Z_t | \mathcal{F}_t] = Z_t$. Falls $\mathcal{X}$ eine rechtsstetige Modifikation besitzt, dann ist $Z_t = X_t$ fast sicher, und damit $\lim_{s \downarrow t} \mathbf{E}[X_s] = \mathbf{E}[X_t]$, also $t \mapsto \mathbf{E}[X_t]$ rechtsstetig. Ist andererseits $t \mapsto \mathbf{E}[X_t]$ rechtsstetig, so folgt $\mathbf{E}[|Z_t - X_t|] = 0$, und damit $Z_t = X_t$ fast sicher. Damit ist $(Z_t)_{t \in I}$ eine rechtsstetige Modifikation von $\mathcal{X}$. □

Bemerkung 15.46 (Übliche Bedingungen). Sei $I = [0, \infty)$. Im Folgenden werden wir immer annehmen, dass die Filtration $(\mathcal{F}_t)_{t \geq 0}$ rechtsstetig und vollständig ist. Weiter zeigt Theorem 15.45, dass es unter diesen Annahmen zu jedem Sub-Martingal $\mathcal{X}$ eine Modifikation mit Pfaden in $\mathcal{D}_{\mathbb{R}}([0, \infty))$ gibt, falls $t \mapsto \mathbf{E}[X_t]$ rechtsstetig ist. Dies wollen wir ebenfalls annehmen, und von jedem Sub-Martingal immer diese Modifikation mit Pfaden in $\mathcal{D}_{\mathbb{R}}([0, \infty))$ nehmen. All dies werden wir zusammen fassen und sagen, dass wir unter den *üblichen Bedingungen* arbeiten.

Theorem 15.47 (Martingalkonvergenzsätze für kontinuierliches I). *Sei $I \subseteq [0, \infty)$ ein Intervall. Unter den üblichen Bedingungen gelten die Aussagen von Lemma 15.24, Proposition 15.25, Lemma 15.27, Theorem 15.28, Korollar 15.29, Theorem 15.31, Theorem 15.32, Theorem 15.35 und Theorem 15.36 entsprechend.*

Beweis. Man beachte, dass alle Aussagen bereits im Falle abzählbarer Indexmenge, also z.B. $I \cap \mathbb{Q}$, gezeigt wurden. Alle Aussagen folgen im kontinuierlichen Fall, weil unter den üblichen Voraussetzungen der Prozess $\mathcal{X} = (X_t)_{t\in I}$, sowie alle seine Grenzwerte, eindeutig aus $(X_t)_{t\in I\cap\mathbb{Q}}$ und dessen Grenzwerten konstruiert werden kann. □

Alle Martingalkonvergenzsätze sind nun auch für den Fall kontinuierlicher Indexmenge gezeigt. Es folgen noch die Aussagen des Optional Sampling (Theorem 15.21) und Optional Stopping Theorems (Proposition 15.18) im kontinuierlichen Fall.

Theorem 15.48 (Optional Sampling Theorem im kontinuierlichen Fall). *Sei $I \subseteq [0,\infty)$ ein Intervall, $S \leq T$ fast sicher endliche Stoppzeiten und $\mathcal{X} = (X_t)_{t\in I}$ ein Sub-Martingal. Ist entweder T beschränkt oder $\mathcal{X}$ gleichgradig integrierbar, so ist X_T integrierbar und $X_S \geq \mathbf{E}[X_T|\mathcal{F}_S]$. Außerdem gilt Lemma 15.22 auch für $I = [0,\infty)$.*

Beweis. Ohne Einschränkung ist $I = [0,\infty)$. Sei $S_n := 2^{-n}[2^n S+1]$ und $T_n := 2^{-n}[2^n T+1]$, so dass $S_n \downarrow S$ und $T_n \downarrow T$ wie in Proposition 14.28. Mit Theorem 15.21 folgt $X_{S_m} \leq \mathbf{E}[X_{T_n}|\mathcal{F}_{S_m}]$ für alle $m \geq n$. Mit $m \to \infty$ und Theorem 15.35.2 folgt

$$X_S \leq \mathbf{E}[X_{T_n}|\mathcal{F}_S] \tag{15.11}$$

Ist T fast sicher beschränkt, so ist $..., X_{T_2}, X_{T_1}$ ein Sub-Martingal mit $\inf_n \mathbf{E}[X_{T_n}] > -\infty$. Also handelt es sich nach Theorem 15.36 um ein uniform integrierbares, fast sicher und in L^1 gegen X_T integrierbares Sub-Martingal. Nun folgt die Aussage aus (15.11) mit $m \to \infty$.

Ist $\mathcal{X}$ uniform integrierbar, dann konvergiert nach Theorem 15.31 (bzw. Theorem 15.47) $X_t \xrightarrow{t\to\infty}_{fs,L^1} X_\infty$ mit integrierbarem X_∞. und es gilt $X_s \leq \mathbf{E}[X_\infty|\mathcal{F}_s]$.

Wie oben ist zunächst $X_S \leq \mathbf{E}[X_{T_n}|\mathcal{F}_S]$, und das Sub-Martingal $..., X_{T_2}, X_{T_1}$ konvergiert fast sicher und in L^1 gegen X_T. Also gilt die Aussage wieder wegen (15.11).

Der Beweis von Lemma 15.22 gilt unverändert. □

Korollar 15.49 (Optional Stopping im kontinuierlichen Fall). *Sei $I \subseteq [0,\infty)$ ein Intervall und $\mathcal{X} = (X_t)_{t\in I}$ ein (Sub, Super)-Martingal und T eine fast sicher endliche Stoppzeit. Dann ist $\mathcal{X}^T = (X_{T\wedge t})_{t\in I}$ ein (Sub, Super)-Martingal.*

Beweis. Das Korollar folgt mit dem Optional Sampling Theorem, da $T \wedge s \leq T \wedge t$, also $X_{T\wedge s} \leq \mathbf{E}[X_{T\wedge t}|\mathcal{F}_{T\wedge s}] \leq \mathbf{E}[X_{T\wedge t}|\mathcal{F}_s]$. □

16 Markov-Prozesse

Die einfachsten stochastischen Prozesse $\mathcal{X} = (X_t)_{t\in I}$ sind die, bei denen $\mathcal{X}$ eine unabhängige Familie ist. Wir kommen nun zur zweit-einfachsten Abhängigkeits-Struktur, die bei stochastischen Prozessen auftritt. Unter einem Markov-Prozess $\mathcal{X}$ verstehen wir einen Prozess, bei dem zur Zeit t der zukünftige Verlauf nur von X_t abhängt, jedoch nicht von $(X_s)_{s<t}$. Mit anderen Worten: $(X_s)_{s>t}$ und $(X_s)_{s<t}$ sind unabhängig gegeben X_t.

Viele der bereits eingeführten stochastische Prozesse sind Markov-Prozesse und werden in diesem Abschnitt als Beispiele dienen. Im ganzen Abschnitt sei (E, r) ein vollständiger und separabler metrischer Raum.

16.1 Definition und Beispiele

In diesem Abschnitt werden wir den Begriff der bedingten Unabhängigkeit aus Abschnitt 12.4 benötigen; siehe auch Beispiel 12.13. Schließlich sind Markov-Prozesse solche, bei denen die Zukunft – gegeben die Gegenwart – nicht von der Vergangenheit abhängt. Nach der Einführung von Markov-Prozessen und einigen Beispielen werden wir in Theorem 16.5 feststellen, wann Gauss'sche Prozesse Markov sind. Ein zentraler Begriff werden Markov-Kerne $\mu_{s,t}^{\mathcal{X}}$ darstellen, die gerade die Übergangswahrscheinlichkeiten zwischen zwei Zeitpunkten s und t beschreiben. Formal äquivalent führen wir Operatoren $T_{s,t}^{\mathcal{X}}$ ein, die angeben, wie sich Erwartungswerte von Funktionen $f(X_t)$ im Laufe der Zeit ändern.

Definition 16.1 (Markov-Prozess). *Sei $(\mathcal{F}_t)_{t\in I}$ eine Filtration und $\mathcal{X} = (X_t)_{t\in I}$ ein adaptierter stochastischer Prozess.*

1. *Der Prozess $\mathcal{X}$ heißt* Markov-Prozess, *falls $\mathcal{F}_s$ unabhängig von X_t ist gegeben X_s, $s \le t$. Das heißt, es gilt für $A \in \mathcal{B}(E)$ (siehe Proposition 12.17)*
$$\mathbf{P}(X_t \in A|\mathcal{F}_s) = \mathbf{P}(X_t \in A|X_s) \tag{16.1}$$
oder äquivalent dazu
$$\mathbf{E}(f(X_t)|\mathcal{F}_s) = \mathbf{E}(f(X_t)|X_s)$$
für alle messbaren und beschränkten $f : E \to \mathbb{R}$.
2. *Die Markov-Kerne (oder Übergangskerne) $\mu_{s,t}^{\mathcal{X}}$ (von E nach E) von $\mathcal{X}$ sind durch*
$$\mu_{s,t}^{\mathcal{X}}(X_s, B) = \mathbf{P}(X_t \in B|X_s) = \mathbf{P}(X_t \in B|\mathcal{F}_s)$$
gegeben.
3. *Sei $\mathcal{B}(E)$ (nicht nur die Borel'sche σ-Algebra auf E, sondern auch) die Menge der beschränkten, messbaren Funktionen $f : E \to \mathbb{R}$. Dann definieren wir für $s \le t$ den Übergangsoperator*
$$T_{s,t}^{\mathcal{X}} : \begin{cases} \mathcal{B}(E) & \to \mathcal{B}(E) \\ f & \mapsto x \mapsto \mathbf{E}[f(X_t)|X_s = x] = \int \mu_{s,t}^{\mathcal{X}}(x, dy) f(y). \end{cases}$$
4. *Für Markov-Kerne μ, ν von E nach E setzen wir außerdem einen Markov-Kern von E nach E^2 durch*
$$(\mu \otimes \nu)(x, A \times B) = \int \mu(x, dy)\nu(y, dz)1_{y\in A, z\in B}$$
und einen Markov-Kern von E nach E durch
$$(\mu\nu)(x, A) = (\mu \otimes \nu)(x, E \times A).$$

Bemerkung 16.2 (Interpretationen). 1. Genau wie bei Martingalen wird die Markov-Eigenschaft bezüglich einer Filtration $(\mathcal{F}_t)_{t\in I}$ formuliert. Im folgenden werden wir jedoch immer $\mathcal{F}_t = \sigma((X_s)_{s\le t})$ setzen, $t \in I$.

2. Wir wollen die Übergangskerne $(\mu^{\mathcal{X}}_{s,t})_{s\leq t}$ als reguläre Versionen der bedingten Erwartung von X_t gegeben X_s interpretieren. Dies ist möglich, da E Polnisch ist und nach Theorem 12.22 dann die reguläre Version der bedingten Verteilung existiert.

3. Den Übergangsoperator $T^{\mathcal{X}}_{s,t}$ interpretiert man am besten so: Gegeben sei eine Funktion f und $f(X_s)$ sei bekannt. Dann ist $(T^{\mathcal{X}}_{s,t}f)(X_s)$ die Erwartung von $f(X_t)$ bei Start in X_s. Diese hängt natürlich vom Wert X_s ab, also ist $T^{\mathcal{X}}_{s,t}f$ eine Funktion in X_s.

4. Zur Interpretation der Markov-Kerne $\mu^{\mathcal{X}}_{s,t}\otimes\mu^{\mathcal{X}}_{t,u}$ und $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}$ für $s\leq t\leq u$ sei folgendes bemerkt: Es ist $\mu^{\mathcal{X}}_{s,t}\otimes\mu^{\mathcal{X}}_{t,u}(x, A\times B)$ die Wahrscheinlichkeit, gegeben $X_s = x$, dass sowohl $X_t\in A$ und $X_u\in B$ ist. Außerdem wird unter $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}$ der Zustand zur Zeit t ausintegriert, d.h. $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}(x,B)$ ist die Wahrscheinlichkeit, gegeben $X_s = x$, dass $X_u\in B$. (Diese muss natürlich im Falle eines Markov-Prozesses gleich $\mu^{\mathcal{X}}_{s,u}(x,B)$ sein; siehe auch die Chapman-Kolmogorov Gleichungen in Korollar 16.16.)

Beispiel 16.3 (Markov-Ketten). (Siehe auch Beispiel 6.10.) Markov-Prozesse $\mathcal{X} = (X_t)_{t\in I}$ mit höchstens abzählbarem Zustandsraum E heißen *Markov-Ketten*. Ist außerdem $I = \{0,1,2,...\}$, so ist der Übergangskern $\mu^{\mathcal{X}}_{t,t+1}$ durch eine Matrix $P_{t,t+1} = (p_{t,t+1}(x,y))_{x,y\in E}$ gegeben, so dass
$$p_{t,t+1}(x,y) = \mathbf{P}(X_{t+1} = y|X_t = x)$$
und
$$\mu^{\mathcal{X}}_{t,t+1}(x,A) = \sum_{y\in A} p_{t,t+1}(x,y).$$
Weiter ist hier
$$(\mu^{\mathcal{X}}_{t,t+1}\otimes\mu^{\mathcal{X}}_{t+1,t+2})(x, A\times B) = \sum_{y\in A, z\in B} p_{t,t+1}(x,y)p_{t+1,t+2}(y,z)$$
und
$$(\mu^{\mathcal{X}}_{t,t+1}\mu^{\mathcal{X}}_{t+1,t+2})(x, A) = \sum_{y\in E, z\in A} p_{t,t+1}(x,y)p_{t+1,t+2}(y,z).$$
Für den Übergangsoperator $(T^{\mathcal{X}}_{s,t})_{s\leq t}$ lässt sich $f: E\to\mathbb{R}$ beschränkt als Vektor schreiben, nämlich als $f = (f(x))_{x\in E}$ und damit ist
$$(T^{\mathcal{X}}_{t,t+1}f)(x) = \sum_{y\in E}\mu^{\mathcal{X}}_{t,t+1}(x,dy)f(y) = \sum_{y\in E} p_{t,t+1}(x,y)f(y),$$
also entspricht die Anwendung von $T^{\mathcal{X}}_{t,t+1}$ auf f einer Multiplikation der Matrix $p_{t,t+1}$ mit dem Vektor f.

Beispiel 16.4 (Summen und Produkte unabhängiger Zufallsvariablen etc.).

1. Seien $X_1, X_2, ...$ reellwertig, fast sicher endlich und unabhängig. Dann sind $\mathcal{S} = (S_t)_{t=0,1,2,...}$ mit $S_t = \sum_{s=1}^t X_s$ and auch $\mathcal{S} = (S_t)_{t=0,1,2,...}$ mit $S_t = \prod_{s=1}^t X_s$ Markov-Prozesse. Es gilt nämlich beispielsweise für $A\in\mathcal{B}(\mathbb{R})$
$$\begin{aligned}\mathbf{P}(S_{t+1}\in A|\mathcal{F}_t) &= \int \mathbf{P}(S_t\in A-x, X_{t+1}\in dx|\mathcal{F}_t)\\ &= \int 1_{S_t\in A-x}\mathbf{P}(X_{t+1}\in dx) = \mathbf{P}(S_{t+1}\in A|S_t).\end{aligned}$$

In diesem Fall ist

$$\mu^{\mathcal{S}}_{t,t+1}(x, A) = \mathbf{P}(X_{t+1} \in A - x)$$

und

$$(T^{\mathcal{S}}_{t,t+1} f)(x) = \mathbf{E}[f(x + X_{t+1})].$$

2. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Poisson-Prozess mit Intensität λ. Dann sind $(X_t)_{t\geq 0}$ sowie $(X_{f(t)})_{t\geq 0}$ für jede wachsende Funktion f Markov-Prozesse, genau wie $(X_t - \lambda t)_{t\geq 0}$. Allerdings ist $(X_t^2 - \lambda \int_0^t (2X_r + 1)dr)_{t\geq 0}$ kein Markov-Prozess; siehe auch Beispiel 15.4. (Für den letzten Prozess sei bemerkt: angenommen $X_t^2 - \lambda \int_0^t (2X_r + 1)dr = x$, fällt der Prozess linear mit Steigung $\lambda(2X_t + 1)$ ab. Allerdings ist diese Steigung keine Funktion von x.)

   Betrachten wir den Poisson-Prozess $\mathcal{X}$. Hier sind die Markov-Kerne für $x \in \{0, 1, 2, ...\}$ gegeben als

$$\mu^{\mathcal{X}}_{s,t}(x, A) = \sum_{k\in A\cap\{x,x+1,...\}} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^{k-x}}{(k-x)!},$$

   und der Übergangsoperator für $f : \{0, 1, 2, ...\} \to \mathbb{R}$ beschränkt ist

$$(T^{\mathcal{X}}_{s,t} f)(x) = \sum_{k=0}^{\infty} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!} f(x + k) = \mathbf{E}[f(x + P)],$$

   wobei P eine Poisson-verteilte Zufallsvariable ist mit Parameter $\lambda(t - s)$.

3. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung. Dann sind sowohl $(\mu X_t)_{t\geq 0}$, als auch $(\mu X_t^2 - \mu t)_{t\geq 0}$ als auch $(\exp(\mu X_t - \mu^2 t/2))_{t\geq 0}$ für $\mu \in \mathbb{R}$ Markov-Prozesse (sowie Martingale nach Beispiel 15.5). Beispielsweise ist

$$\begin{aligned}\mathbf{P}[X_u^2 - u \leq x|\mathcal{F}_t] &= \mathbf{P}[(X_u - X_t)^2 + 2(X_u - X_t)X_t + X_t^2 \leq u + x|\mathcal{F}_t] \\ &= \mathbf{P}[(X_u - X_t)^2 + 2(X_u - X_t)X_t + X_t^2 \leq u + x|X_t] = \mathbf{P}[X_u^2 - u \leq x|X_t].\end{aligned}$$

   Betrachten wir die Brown'sche Bewegung $\mathcal{X}$. Ihre Markov-Kerne ist gegeben durch

$$\mu^{\mathcal{X}}_{s,t}(x, A) = \frac{1}{\sqrt{2\pi(t-s)}} \int_A \exp\Big(-\frac{(y-x)^2}{2(t-s)}\Big) dy$$

   und der Übergangsoperator für $f \in \mathcal{B}(\mathbb{R})$

$$(T^{\mathcal{X}}_{s,t} f)(x) = \frac{1}{\sqrt{2\pi(t-s)}} \int \exp\Big(-\frac{y^2}{2(t-s)}\Big) f(x + y)dy = \mathbf{E}[f(x + \sqrt{t-s}Z)],$$

   wobei Z eine $N(0, 1)$-verteilte Zufallsvariable ist.

Theorem 16.5 (Gauss'sche Markov-Prozesse). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Gauss'scher Prozess. Genau dann ist $\mathcal{X}$ Markov, falls*

$$\mathbf{COV}(X_s, X_u) \cdot \mathbf{V}(X_t) = \mathbf{COV}(X_s, X_t) \cdot \mathbf{COV}(X_t, X_u) \tag{16.2}$$

für alle $s \leq t \leq u$.

Beweis. Durch Subtraktion der Erwartungswerte können wir ohne Beschränkung der Allgemeinheit annehmen, dass $\mathbf{E}[X_t] = 0$ für alle $t \geq 0$ gilt. Wir bemerken, dass (falls $\mathbf{V}(X_t) > 0$) mit

$$X'_u = X_u - \frac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)} X_t$$

gilt, dass $\mathbf{COV}(X'_u, X_t) = 0$. Also sind X'_u und X_t unabhängig (und die gemeinsame Verteilung eine Normalverteilung). Im Falle $\mathbf{V}(X_t) = 0$ setzen wir $X'_u = X_u$ woraus dasselbe folgt.

Sei zunächst $\mathcal{X}$ Markov und $s \leq t$. Dann ist X_s von X_u unabhängig gegeben X_t, also ist auch X_s von X'_u unabhängig gegeben X_t. Da auch X_t und X'_u unabhängig sind, folgt

$$\begin{aligned}
\mathbf{P}(X_s \in A, X'_u \in B) &= \mathbf{E}[\mathbf{P}(X_s \in A|X_t) \cdot \mathbf{P}(X'_u \in B|X_t)] \\
&= \mathbf{E}[\mathbf{P}(X_s \in A|X_t) \cdot \mathbf{P}(X'_u \in B)] = \mathbf{P}(X_s \in A) \cdot \mathbf{P}(X'_u \in B)
\end{aligned}$$

und damit sind X_s und X'_u unabhängig. Damit gilt

$$0 = \mathbf{COV}(X_s, X'_u) = \mathbf{COV}(X_s, X_u) - \frac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)} \mathbf{COV}(X_s, X_t)$$

und (16.2) folgt.

Andersherum erfülle $\mathcal{X}$ die Gleichung (16.2). Dann ist (mit derselben Rechnung wie eben) X_s unabhängig von X'_u für alle $s \leq t$. Damit ist X'_u unabhängig von $\mathcal{F}_t = \sigma((X_s)_{s \leq t})$ und es folgt

$$\begin{aligned}
\mathbf{P}(X_u \in A|\mathcal{F}_t) &= \int \mathbf{P}\Big(X'_u \in dx, \tfrac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)} X_t \in A - x|\mathcal{F}_t\Big) \\
&= \int \mathbf{P}\Big(X'_u \in dx, \tfrac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)} X_t \in A - x|X_t\Big) \\
&= \mathbf{P}(X_u \in A|X_t).
\end{aligned}$$

□

Beispiel 16.6 (Beispiele Gauss'scher Markov-Prozesse).

1. Wir haben schon gezeigt, dass eine Brown'sche Bewegung $\mathcal{X}$ ein Markov-Prozess ist. Zur Sicherheit sei hierzu noch bemerkt, dass in diesem Fall für $s \leq t \leq u$

$$\mathbf{COV}(X_s, X_u) \cdot \mathbf{V}(X_t) = s \cdot t = \mathbf{COV}(X_s, X_t) \cdot \mathbf{COV}(X_t, X_u).$$

2. Eine fraktionale Brown'sche Bewegung mit Hurst-Parameter h ist ein Gauss'scher Prozess $\mathcal{X} = (X_t)_{t \geq 0}$ mit $\mathbf{E}[X_t] = 0$, $t \geq 0$ und

$$\mathbf{COV}(X_s, X_t) = \frac{1}{2}(t^{2h} + s^{2h} - (t - s)^{2h}).$$

   Wie man leicht nachrechnet, ist dies nur für $h = \frac{1}{2}$ ein Markov-Prozess. Dann ist $\mathcal{X}$ die Brown'sche Bewegung.

3. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung und $\mathcal{Y} = (Y_t)_{t\in[0,1]}$ gegeben als $Y_t = X_t - tX_1$. Dann heißt $\mathcal{Y}$ Brown'sche Brücke; siehe auch Abbildung 16.1. Es ist $\mathbf{E}[Y_t] = 0, t \geq 0$ und $s \leq t$
$$\mathbf{COV}(Y_s, Y_t) = \mathbf{COV}(X_s - sX_1, X_t - tX_1) = s - 2st + st = s(1-t).$$
Damit gilt für $s \leq t \leq u$
$$\mathbf{COV}(Y_s, Y_u) \cdot \mathbf{V}(Y_t) = s(1-u)t(1-t) = \mathbf{COV}(Y_s, Y_t) \cdot \mathbf{COV}(Y_t, Y_u),$$
also ist die Brown'sche Brücke ein Markov-Prozess.

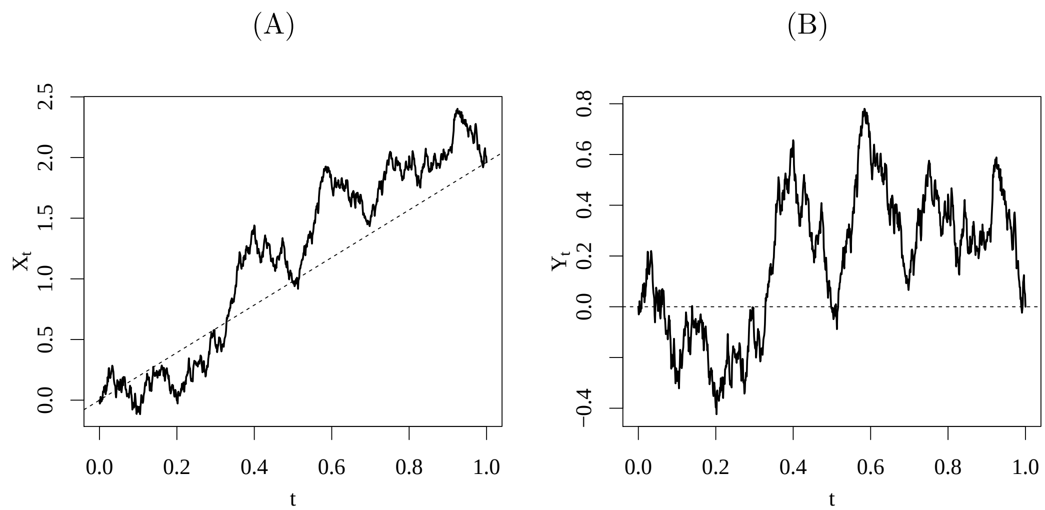


Abbildung 16.1: (A) Der Pfad einer Brown'schen Bewegung $\mathcal{X} = (X_t)_{t\in[0,1]}$. (B) Der entsprechende Pfad der Brown'schen Brücke $\mathcal{Y} = (Y_t)_{t\in[0,1]}$ mit $Y_t = X_t - tX_1$.

Die verbale Beschreibung von Markov-Prozessen besagt, dass der zukünftige Verlauf des Prozesses unabhängig von der Vergangenheit ist, gegeben die Gegenwart. Jedoch wird in Definition (16.1) nur gefordert, dass einzelne Zeitpunkte der Zukunft unabhängig von der Vergangenheit sind, gegeben die Gegenwart. Dass dies in der Tat mit der verbalen Beschreibung übereinstimmt, wird nun gezeigt.

Lemma 16.7 (Erweiterte Markov-Eigenschaft). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Markov-Prozess. Dann ist $(X_u)_{u\geq t}$ unabhängig von $\mathcal{F}_t$ gegeben X_t*

Beweis. Seien $t = t_0 < t_1 < ... < t_n \in I$ und $A_0, ..., A_n \in E$. Dann gilt
$$\begin{aligned}
\mathbf{P}(X_{t_0} \in A_0,..., X_{t_n} \in A_n|\mathcal{F}_t) &= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_{n-1}}\in A_{n-1}} \cdot \mathbf{E}[1_{X_{t_n}\in A_n}|\mathcal{F}_{t_{n-1}}]|\mathcal{F}_t] \\
&= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_{n-1}}\in A_{n-1}} \cdot \mathbf{E}[1_{X_{t_n}\in A_n}|X_{t_{n-1}}]|\mathcal{F}_t] \\
&= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_{n-2}}\in A_{n-2}} \cdot \underbrace{\mathbf{E}[1_{X_{t_{n-1}}\in A_{n-1}}\mathbf{E}[1_{X_{t_n}\in A_n}|X_{t_{n-1}}]|\mathcal{F}_{t_{n-2}}]}_{\substack{=\mathbf{E}[1_{X_{t_{n-1}}\in A_{n-1}}\mathbf{E}[1_{X_{t_n}\in A_n}|X_{t_{n-1}},X_{t_{n-2}}]|X_{t_{n-2}}] \\ =\mathbf{E}[1_{X_{t_{n-1}}\in A_{n-1}}\cdot 1_{X_{t_n}\in A_n}|X_{t_{n-2}}]}} |\mathcal{F}_t] \\
&= \cdots = \mathbf{E}[1_{X_{t_0}\in A_0}\mathbf{E}[1_{X_1\in A_1}, ..., 1_{X_{t_n}\in A_n}|X_{t_0}]|\mathcal{F}_t] \\
&= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_n}\in A_n}|X_t] = \mathbf{P}[X_{t_0} \in A_0, ..., X_{t_n} \in A_n|X_t],
\end{aligned}$$

wobei wir Proposition 12.17 verwendet haben. Damit ist gezeigt, dass $(X_{t_0}, ..., X_{t_n})$ unabhängig von $\mathcal{F}_t$ ist gegeben X_t, also die Unabhängigkeit auf Zylindermengen $\{X_{t_0} \in A_0, ..., X_{t_n} \in A_n\}$. Dies erweitert man mittels eines Argumentes mit einem Dynkin-System zu allen Mengen in $\sigma((X_u)_{u\geq t})$. □

Ein besonderer Fall ist der eines Markov-Prozesses, der räumlich homogen ist. Dieser verhält sich immer gleich, unabhängig wie sein momentaner Wert ist. Solche Prozesse haben wir auch schon kennen gelernt, etwa die Brown'sche Bewegung und der Poisson-Prozess. Äquivalent dazu ist es, dass der Prozess unabhängige Inkremente hat, wie Lemma 16.9 zeigt.

Definition 16.8 (Räumlich homogener Markov-Prozess).
Sei E eine Abelsche Gruppe.

1. *Ein Markov-Kern von E nach E heißt* homogen*, falls $\mu(x, B) = \mu(0, B - x)$ für alle $x \in E$ und $B \in \mathcal{B}(E)$ gilt. (Hierbei ist $B - x = \{y - x : y \in B\}$.)*

2. *Ein Markov-Prozess $\mathcal{X}$ heißt* räumlich homogen*, falls die Markov-Kerne $\mu^{\mathcal{X}}_{s,t}$ homogen sind, $s \leq t$.*

3. *Ein Markov-Prozess $\mathcal{X} = (X_t)_{t\geq 0}$ hat unabhängige Inkremente, falls $X_t - X_s$ unabhängig von $\mathcal{F}_s$ ist, $s \leq t$.*

Lemma 16.9 (Homogeneität und unabhängige Inkremente). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Markov-Prozess mit Zustandsraum E, wobei E eine Abelsche Gruppe ist. Der Prozess $\mathcal{X}$ hat genau dann unabhängige Inkremente, wenn $\mathcal{X}$ räumlich homogen ist. In diesem Fall ist die Vervollständigung der Filtration $(\mathcal{F}_t)_{t\geq 0}$ mit $\mathcal{F}_t = \sigma((X_s)_{s\leq t})$ rechtsstetig.*

Beweis. Sei zunächst $\mathcal{X}$ ein räumlich homogener Markov-Prozess, also $\mu^{\mathcal{X}}_{s,t}(x, B) = \mu^{\mathcal{X}}_{s,t}(0, B - x)$ für alle $x \in E$ und $B \in \mathcal{B}(E)$. Dann gilt

$$\mathbf{P}(X_t - X_s \in B|\mathcal{F}_s) = \mathbf{P}(X_t \in X_s + B|\mathcal{F}_s) = \mu_{s,t}(X_s, X_s + B) = \mu^{\mathcal{X}}_{s,t}(0, B).$$

Damit ist $X_t - X_s$ nach Lemma 12.12 unabhängig von $\mathcal{F}_s$, also hat $\mathcal{X}$ unabhängige Inkremente.

Andersherum habe $\mathcal{X}$ unabhängige Inkremente. Dann ist $(X_t - X_s)_{t\geq s}$ ebenfalls ein Markov-Prozess mit denselben Markov-Kernen und es gilt

$$\mu^{\mathcal{X}}_{s,t}(X_s, B) = \mathbf{P}(X_t \in B|\mathcal{F}_s) = \mathbf{P}(X_t - X_s \in B - X_s|\mathcal{F}_s) = \mu^{\mathcal{X}}_{s,t}(0, B - X_s).$$

Wir kommen nun zum zweiten Teil der Aussage, der Rechtsstetigkeit der von $\mathcal{X}$ erzeugten Filtration. Sei $t \in I$ und $u_1, u_2, ... \in I$ mit $u_n \downarrow t$. Ohne Einschränkung nehmen wir an, dass $\mathcal{F}_t$ vollständig ist. Wir müssen zeigen, dass $\mathcal{F}_t^+ = \bigcap_n \mathcal{F}_{u_n} = \mathcal{F}_t$. Zunächst ist $(\mathcal{F}_t, \mathcal{G}_1, \mathcal{G}_2, ...)$ mit $\mathcal{G}_n = \sigma(X_{u_{n-1}} - X_{u_n})$ eine unabhängige Familie. Es ist $\mathcal{F}_t^+$ unabhängig von $(\mathcal{G}_1, ..., \mathcal{G}_n)$ für jedes n. Sei $A \in \mathcal{F}_t^+$. Dann ist nach Proposition 12.17

$$\mathbf{P}(A|\mathcal{F}_t) = \mathbf{P}(A|\mathcal{F}_t, \mathcal{G}_1, ..., \mathcal{G}_n) \xrightarrow{n\to\infty} 1_A$$

fast sicher wegen Theorem 15.35 und weil 1_A messbar ist bezüglich $\sigma(\mathcal{F}_t, \mathcal{G}_1, .\mathcal{G}_2, ...)$. Insbesondere folgt, da $\mathcal{F}_t$ vollständig ist, $\mathcal{F}_t^+ \subseteq \mathcal{F}_t \subseteq \mathcal{F}_t^+$. □

16.2 Starke Markov-Prozesse

Bei Martingalen haben wir das Vorgehen kennen gelernt, dass eine Eigenschaft, die für feste Zeiten gilt (z.B. $X_s = \mathbf{E}[X_t|\mathcal{F}_s]$) auf Stoppzeiten übertragen wird. (Dies führte etwa zum Optional Sampling Theorem, also $X_S = \mathbf{E}[X_T|\mathcal{F}_S]$ für fast sicher beschränkte Stoppzeiten $S \leq T$.)

Die Markov-Eigenschaft ist zunächst wieder eine Eigenschaft für feste Zeitpunkte, die man z.B. schreiben kann als

$$\mathbf{P}(X_{s+t} \in A|\mathcal{F}_s) = \mu^{\mathcal{X}}_{s,s+t}(X_s, A).$$

Das Ersetzen der festen Zeit s in der letzten Gleichung durch eine Stoppzeit S führt zu starken Markov-Prozessen. Die meisten hier behandelten Prozesse gehören zu dieser Klasse, jedoch bildet Beispiel 16.14 eine Ausnahme.

Definition 16.10 (Starker Markov-Prozess). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Markov-Prozess mit erzeugter Filtration $(\mathcal{F}_t)_{t\in I}$ und progressiv messbar. Weiter sei S eine $(\mathcal{F}_t)_{t\in I}$-Stoppzeit. Dann hat $\mathcal{X}$ die* starke Markov-Eigenschaft bei S*, falls*

$$\mathbf{P}(X_{S+t} \in A|\mathcal{F}_S) = \mu^{\mathcal{X}}_{S,S+t}(X_S, A)$$

für $A \in \mathcal{B}(E)$ oder äquivalent dazu

$$\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] = (T^{\mathcal{X}}_{S,S+t}f)(X_S)$$

für $f \in \mathcal{B}(E)$ gilt. Weiter heißt $\mathcal{X}$ starker Markov-Prozess*, falls $\mathcal{X}$ die starke Markov-Eigenschaft bei allen fast sicher endlichen Stoppzeiten hat.*

Proposition 16.11 (Stark Markov bei diskreten Stoppzeiten). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Markov-Prozess mit erzeugter Filtration $(\mathcal{F}_t)_{t\in I}$ und progressiv messbar. Weiter sei S eine fast sicher endliche $(\mathcal{F}_t)_{t\in I}$-Stoppzeit, die nur diskrete (also insbesondere nur abzählbar viele) Werte annimmt. Dann hat $\mathcal{X}$ die starke Markov-Eigenschaft bei S.*

Ist insbesondere I diskret, so hat jeder Markov-Prozess $\mathcal{X}$ auch die starke Markov-Eigenschaft.

Beweis. Sei $\{s_1, s_2, ...\}$ der Wertebereich von S und $f \in \mathcal{B}(E)$ sowie $A \in \mathcal{F}_S$. Dann ist (da der Wertebereich von S diskret ist) $A \cap \{S = s_i\} \in \mathcal{F}_{s_i}$ und

$$\begin{aligned}
\mathbf{E}[f(X_{S+t}), A] &= \sum_i \mathbf{E}[f(X_{S+t}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}[f(X_{s_i+t}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}\big[\mathbf{E}[f(X_{s_i+t})|X_{s_i}], A \cap \{S = s_i\}\big] \\
&= \sum_i \mathbf{E}[(T_{s_i,s_i+t}f)(X_{s_i}), A \cap \{S = s_i\}\big] \\
&= \sum_i \mathbf{E}[(T_{S,S+t}f)(X_S), A \cap \{S = s_i\}\big] \\
&= \mathbf{E}[(T_{S,S+t}f)(X_S), A\big].
\end{aligned}$$

Da $(T_{S,S+t}f)(X_S)$ nach $\mathcal{F}_S$-messbar ist, folgt die Behauptung. □

Theorem 16.12 (Stark Markov bei stetigem Übergangsoperator). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Markov-Prozess mit erzeugter Filtration $(\mathcal{F}_t)_{t\in I}$ mit rechtsstetigen Pfaden. Ist $T^{\mathcal{X}}_{s,t}f$ stetig für $f \in \mathcal{C}_b(E)$ und $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ stetig für alle $f \in \mathcal{C}_b(E)$ (bezüglich der Supremumsnorm auf $\mathcal{C}_b(E)$), dann ist $\mathcal{X}$ ein starker Markov-Prozess.*

Beweis. Zunächst ist nach Lemma 14.32 der Prozess $\mathcal{X}$ progressiv messbar. Sei S eine fast sicher endliche Stoppzeit, die wir nach Proposition 14.28 durch Stoppzeiten $S_1, S_2, ...$ mit $S_n \downarrow S$ approximieren, so dass S_n nur diskrete Werte annimmt, $n = 1, 2, ...$ Dann gilt wegen der Rechtsstetigkeit der Pfade von $\mathcal{X}$, dass $X_{S_n} \xrightarrow{n\to\infty} X_S$ fast sicher und für $f \in \mathcal{C}_b(E)$ ist

$$\begin{aligned}
\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] &= \lim_{n\to\infty} \mathbf{E}[\mathbf{E}[f(X_{S_n+t})|\mathcal{F}_{S_n}]|\mathcal{F}_S] \\
&= \lim_{n\to\infty} \mathbf{E}[(T^{\mathcal{X}}_{S_n,S_n+t}f)(X_{S_n})|\mathcal{F}_S] \\
&= \mathbf{E}[(T^{\mathcal{X}}_{S,S+t}f)(X_S)|\mathcal{F}_S] = (T^{\mathcal{X}}_{S,S+t}f)(X_S),
\end{aligned}$$

wobei die Stetigkeitsvoraussetzungen im dritten Gleichheitszeichen eingegangen sind. □

Beispiel 16.13 (Poisson-Prozess und Brown'sche Bewegung sind stark Markov).

1. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Poisson-Prozess mit Intensität $\lambda \geq 0$. Dann ist $\mathcal{X}$ stark Markov, denn:
   Nach Beispiel 16.4.2 ist $(T^{\mathcal{X}}_{s,t}f)(x) = \mathbf{E}[f(x+P)]$, wobei $P \sim \text{Poi}(\lambda(t-s))$. Damit ist $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ konstant. Weiter ist $x \mapsto (T^{\mathcal{X}}_{s,s+t}f)(x)$ messbar und wegen der diskreten Topologie auf $\{0,1,2,...\}$ auch stetig. Die starke Markov-Eigenschaft folgt damit aus Theorem 16.12.

2. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung. Dann ist $\mathcal{X}$ stark Markov, denn:
   Nach Beispiel 16.4.3 ist $(T^{\mathcal{X}}_{s,t}f)(x) = \mathbf{E}[f(x+\sqrt{t-s}Z)]$, wobei $Z \sim N(0,1)$. Damit ist $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ konstant und $x \mapsto (T^{\mathcal{X}}_{s,s+t}f)(x)$ stetig. Wieder folgt die starke Markov-Eigenschaft aus Theorem 16.12.

Es ist gar nicht so einfach, nicht-starke Markov-Prozesse anzugeben. Hier jedoch ein Beispiel.

Beispiel 16.14 (Ein nicht-starker Markov-Prozess). Sei $T \sim \exp(1)$-verteilt. Weiter definieren wir den stochastischen Prozess $\mathcal{X} = (X_t)_{t\geq 0}$ mit

$$X_t = (t-T)^+$$

und Vervollständigung der kanonischen Filtration $(\mathcal{F}_t)_{t\geq 0}$. Dann ist für $f \in \mathcal{B}(\mathbb{R})$

$$\mathbf{E}[f(X_{s+t})|\mathcal{F}_s] = \begin{cases} \mathbf{E}[f((t-T)^+)], & \text{falls } X_s = 0, \\ f(x+t), & \text{falls } X_s > 0. \end{cases}$$

Insbesondere hängt die rechte Seite nur von X_s ab und damit ist $\mathcal{X}$ ein Markov-Prozess mit Übergangsoperator

$$(T^{\mathcal{X}}_{s,s+t})f(x) = 1_{x=0}\mathbf{E}[f((t-T)^+)] + 1_{x>0}f(x+t).$$

Betrachte nun die zufällige Zeit $S = \inf\{t : X_t > 0\}$ (also $S = T$). Nach Proposition 14.30.2 ist T eine Optionszeit und damit, da $\{T = t\}$ eine Nullmenge und $\mathcal{F}_t$ vollständig ist, $\{T \leq t\} = \{T < t\} \cup \{T = t\} \in \mathcal{F}_t$. Damit ist T eine $(\mathcal{F}_t)_{t\geq 0}$ eine Stoppzeit. Nun gilt

$$\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] = f(t),$$

da S nach $\mathcal{F}_S$-messbar ist und $X_{S+t} = t$ fast sicher gilt. Andererseits ist $X_S = 0$ und damit

$$(T^{\mathcal{X}}_{S,S+t}f)(X_S) = (T^{\mathcal{X}}_{S,S+t}f)(0) = \mathbf{E}[f((t-T)^+)].$$

Da die rechten Seiten der letzten beiden Gleichungen für viele $f \in \mathcal{B}(E)$ nicht übereinstimmen, ist $\mathcal{X}$ kein starker Markov-Prozess.

16.3 Verteilung von Markov-Prozessen

Für einen Markov-Prozess $\mathcal{X}$ stellen die Markov-Kerne $\mu^{\mathcal{X}}_{s,t}$ sowie die Übergangsoperatoren $T^{\mathcal{X}}_{s,t}$ ein wichtiges Werkzeug dar. Wir werden in Theorem 16.17 lernen, dass eine Konsistenz-Bedingung (die Chapman-Kolmogorov-Gleichungen, siehe Korollar 16.16) nicht nur notwendig sondern auch hinreichend für eine Familie von Markov-Kernen ist, um Markov-Kerne für einen Markov-Prozess zu sein.

Lemma 16.15 (Endlich-dimensionale Verteilungen). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Markov-Prozess mit $X_t \sim \nu_t^{\mathcal{X}}$ für Verteilungen $\nu_t^{\mathcal{X}}$ auf E und Markov-Kernen $(\mu_{s,t}^{\mathcal{X}})_{s\leq t}$. Dann gilt für $t_0 < \cdots < t_n$*

$$(X_{t_0}, ...X_{t_n}) \sim \nu_{t_0}^{\mathcal{X}} \otimes \mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_{n-1},t_n}^{\mathcal{X}}$$

und

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = (\mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_{n-1},t_n}^{\mathcal{X}})(X_{t_0}, \cdot)$$

Beweis. Der Beweis der ersten Formel erfolgt mittels Induktion. Für $n = 0$ ist die Aussage klar. Gilt sie für ein n, so gilt für $f \in \mathcal{C}_b(E^{n+2})$

$$\begin{aligned}\mathbf{E}[f(X_{t_0},..., X_{t_{n+1}})] &= \mathbf{E}[\mathbf{E}[f(X_{t_0}, ..., X_{t_{n+1}})|\mathcal{F}_{t_n}]] \\ &= \mathbf{E}\Big[\int f(X_{t_0}, ..., X_{t_n}, x_{n+1})\mu_{t_n,t_{n+1}}^{\mathcal{X}}(X_{t_n}, dx_{n+1})\Big] \\ &= \int \nu_{t_0}^{\mathcal{X}} \otimes \mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n,t_{n+1}}^{\mathcal{X}}(dx_0, ..., dx_{n+1})f(x_0, ..., x_{n+1})\end{aligned}$$

also gilt die erste Formel für $n+1$. Für die zweite Formel bemerken wir, dass die rechte Seite X_{t_0}-messbar ist. Außerdem gilt mit Lemma 16.7

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = \mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|X_{t_0})$$

sowie für $A \in \mathcal{B}(E)$ und $B \in \mathcal{B}(E^n)$ mit der ersten Formel

$$\begin{aligned}&\mathbf{E}[1_{(X_{t_1},...,X_{t_n})\in B}, X_{t_0} \in A] = \mathbf{P}((X_{t_0}, ..., X_{t_n}) \in A \times B) \\ &\quad = \int_A \nu_{t_0}^{\mathcal{X}}(dx)(\mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n,t_{n+1}}^{\mathcal{X}}(x, B) = \mathbf{E}[(\mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n,t_{n+1}}^{\mathcal{X}})(X_{t_0}, B), X_{t_0} \in A],\end{aligned}$$

woraus die Behauptung folgt. □

Korollar 16.16 (Chapman-Kolmogorov Gleichungen). *Sei $\mathcal{X}$ ein Markov-Prozess mit $X_t \sim \nu_t^{\mathcal{X}}$ für Verteilungen $\nu_t^{\mathcal{X}}$ auf E, Markov-Kernen $(\mu_{s,t}^{\mathcal{X}})_{s\leq t}$ und Übergangsoperatoren $(T_{s,t}^{\mathcal{X}})_{s\leq t}$. Dann gilt für $s \leq t \leq u$*

$$\mu_{s,t}^{\mathcal{X}}\mu_{t,u}^{\mathcal{X}} = \mu_{s,u}^{\mathcal{X}}, \tag{16.3}$$

und für $f \in \mathcal{B}(E)$

$$(T_{s,t}^{\mathcal{X}}(T_{t,u}^{\mathcal{X}}f))(X_s) = (T_{s,u}^{\mathcal{X}}f)(X_s) \tag{16.4}$$

$\nu_s^{\mathcal{X}}$-fast sicher.

Beweis. Nach Proposition 16.15 gilt für $\nu_s^{\mathcal{X}}$-fast alle X_s für $A \in \mathcal{B}(E)$

$$\begin{aligned}\mu_{s,u}^{\mathcal{X}}(X_s, A) &= \mathbf{P}(X_u \in A|\mathcal{F}_s) = \mathbf{P}((X_t, X_u) \in E \times A|\mathcal{F}_s) \\ &= (\mu_{s,t}^{\mathcal{X}} \otimes \mu_{t,u}^{\mathcal{X}})(X_s, E \times A) = (\mu_{s,t}^{\mathcal{X}}\mu_{t,u}^{\mathcal{X}})(X_s, A)\end{aligned}$$

sowie für $f \in \mathcal{B}(E)$

$$\begin{aligned}(T_{s,u}^{\mathcal{X}}f)(X_s) &= \mathbf{E}[f(X_u)|\mathcal{F}_s] = \mathbf{E}\big[\mathbf{E}[f(X_u)|\mathcal{F}_t]|\mathcal{F}_s\big] \\ &= \mathbf{E}[(T_{t,u}^{\mathcal{X}}f)(X_t)|\mathcal{F}_s] = (T_{s,t}^{\mathcal{X}}(T_{t,u}^{\mathcal{X}}f))(X_s).\end{aligned}$$

□

Klar ist, dass es zu jedem Markov-Prozess die Markov-Kerne $(\mu_{s,t}^{\mathcal{X}})_{s\leq t}$ gibt. Andersherum zeigen wir nun, dass es zu jeder Familie von Markov-Kernen $(\mu_{s,t})_{s\leq t}$, die den Chapman-Kolmogorov Gleichungen genügt, einen Markov-Prozess gibt.

Theorem 16.17 (Existenz von Markov-Prozessen).
Sei I eine Indexmenge mit $\min I = 0$, ν_0 ein Wahrscheinlichkeitsmaß auf E. Dann gilt:

1. *Ist $(\mu_{s,t})_{s\leq t}$ eine Familie von Markov-Kernen mit $\mu_{s,t}\mu_{t,u} = \mu_{s,u}$ für alle $s \leq t \leq u$. Dann gibt es einen Markov-Prozess mit Startverteilung ν_0 und Übergangskernen $(\mu_{s,t})_{s\leq t}$.*
2. *Ist $(T_{s,t})_{s\leq t}$ eine Familie von Übergangsoperatoren mit $T_{s,t}T_{t,u} = T_{s,u}$ für alle $s \leq t \leq u$. Dann gibt es einen Markov-Prozess mit Startverteilung ν_0 und Übergangsoperatoren $(T_{s,t})_{s\leq t}$.*

Beweis. Gegeben $(\mu_{s,t})_{s\leq t}$ rechnet man leicht nach, dass

$$(T_{s,t}f)(x) := \int \mu_{s,t}(x, dy)f(y)$$

mit $f \in \mathcal{B}(E)$ eine Familie von Übergangsoperatoren $(T_{s,t})_{s\leq t}$ definiert, der genau dann (16.4) erfüllt wenn $(\mu_{s,t})_{s\leq t}$ die Bedingungen (16.3) erfüllt. Ist andersherum $(T_{s,t})_{s\leq t}$ gegeben, so definiert

$$\mu_{s,t}(x, A) = (T_{s,t}1_A)(x)$$

eine Familie von Markov-Kernen, die genau dann (16.3) erfüllt wenn $(T_{s,t})_{s\leq t}$ die Bedingung (16.4) erfüllt. Deshalb genügt es, 1. zu zeigen. Hierfür definieren wir zunächst die Maße für $t_1 < ... < t_n$ mit $\{t_1, ..., t_n\} \Subset I$

$$\nu_{t_1,...,t_n} = \nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \otimes \mu_{t_{n-1},t_n}.$$

Um zu zeigen, dass $(\nu_{t_1,...,t_n})_{\{t_1,...,t_n\}\Subset I}$ eine projektive Familie ist sei $J = \{t_1, ..., t_n\}$ und $H = \{t_1, ..., t_{k-1}, t_{k+1}, ..., t_n\}$. Dann ist für $B = B_1 \times \cdots \times B_{k-1} \times B_{k+1} \times \cdots \times B_n \in \mathcal{B}(E^H)$

$$\begin{aligned}
(\pi^J_H)_*\nu_J(B) &= \nu_J((\pi^J_H)^{-1}(B)) \\
&= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{n-1},t_n})(B_1 \times \cdots \times B_{k-1} \times E \times B_{k+1} \times \cdots \times B_n) \\
&= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{k-1},t_k}\mu_{t_k,t_{k+1}} \otimes \mu_{t_{k+1},t_{k+2}} \otimes \cdots \mu_{t_{n-1},t_n})(B) \\
&= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{k-1},t_{k+1}} \otimes \mu_{t_{k+1},t_{k+2}} \otimes \cdots \mu_{t_{n-1},t_n})(B) \\
&= \nu_H(B).
\end{aligned}$$

Nach Theorem 6.24 gibt es einen Prozess $\mathcal{X} = (X_t)_{t\in I}$ mit den endlich-dimensionalen Verteilungen $(\nu_J)_{J\Subset I}$ und Startverteilung ν_0. Es bleibt zu zeigen, dass $\mathcal{X}$ ein Markov-Prozess ist. Hierfür sei $A \in \mathcal{B}(E^J)$ für ein $J \Subset I$ und $\max J = s \leq t$ sowie $B \in \mathcal{B}(E)$. Dann gilt

$$\mathbf{P}((X_r)_{r\in J} \in A, X_t \in B) = \nu_{J\cup\{t\}}(A \times B) = \mathbf{E}[\mu_{s,t}(X_s, B), (X_r)_{r\in J} \in A].$$

Ist $(\mathcal{F}_t)_{t\in I}$ die von $\mathcal{X}$ erzeugte Filtration, so gilt also für $A \in \mathcal{F}_s$

$$\mathbf{P}(X_t \in B, A) = \mathbf{E}[\mu_{s,t}(X_s, B), A].$$

Aus der Definition der bedingten Erwartung lesen wir ab, dass $\mathbf{P}(X_s \in B|\mathcal{F}_s) = \mu_{s,t}(X_s, B) = \mathbf{P}(X_s \in B|X_s)$. Daraus folgt die Behauptung. □

Korollar 16.18 (Verteilung von Markov-Prozessen). *Sei ν und $(\mu_{s,t})_{s\leq t}$ wie in Theorem 16.17. Dann gibt es eine Wahrscheinlichkeitsverteilung $\mathbf{P}_\nu$ auf $\mathcal{B}(E)^I$, so dass $\mathbf{P}_\nu$ die Verteilung des Markov-Prozesses mit Übergangskernen $(\mu_{s,t})_{s\leq t}$ und Anfangsverteilung ν ist. Weiter definiert $x \mapsto \mathbf{P}_x := \mathbf{P}_{\delta_x}$ einen Übergangskern von E nach $\mathcal{B}(E)^I$ und es gilt*

$$\mathbf{P}_\nu = \int \nu(dx)\mathbf{P}_x.$$

Beweis. Man rechnet leicht nach, dass $\mathbf{P}_\nu(A) = \int \nu(dx)\mathbf{P}_x(A)$ für Zylindermengen A gilt. Wie üblich erweiter man diese Aussage auf alle $A \in \mathcal{B}(E)^I$. □

16.4 Halbgruppen und Generatoren

Eine besondere Rolle spielen zeitlich homogene Markov-Prozesse. Bei diesen hängt $\mu_{s,t}^{\mathcal{X}}$ nur von der Zeitdifferenz $t-s$ ab.

Definition 16.19 (Zeitlich homogener Markov-Prozess und seine Halbgruppen). *Sei I abgeschlossen unter Addition. Ein Markov-Prozess $\mathcal{X}$ heißt* zeitlich homogen, *falls es eine Familie von Markov-Kernen $(\mu_t)_{t\in I}$ gibt mit $\mu_{s,t}^{\mathcal{X}} = \mu_{t-s}$. Dann schreiben wir auch $\mu_t^{\mathcal{X}} = \mu_t$ und bezeichnen $(\mu_t^{\mathcal{X}})_{t\in I}$ als* Übergangshalbgruppe[21].

Dies ist (natürlich) genau dann der Fall, wenn es eine Familie von Übergangsoperatoren $(T_t)_{t\in I}$ gibt mit $T_{s,t}^{\mathcal{X}} = T_{t-s}$. In diesem Fall schreiben wir $T_t^{\mathcal{X}} = T_t$ und bezeichnen $(T_t^{\mathcal{X}})_{t\in I}$ als Operatorhalbgruppe.

Bemerkung 16.20 (Übertragung auf zeitlich homogene Markov-Prozesse). Sei $\mathcal{X}$ ein zeitlich homogener Markov-Prozess mit Übergangs- und Operator-Halbgruppe $(\mu_t^{\mathcal{X}})_{t\in I}$ und $(T_t^{\mathcal{X}})_{t\in I}$. Dann gilt nach Ergebnissen aus Abschnitt 16.3

$$(X_{t_0}, ...X_{t_n}) \sim \nu_{t_0}^{\mathcal{X}} \otimes \mu_{t_1-t_0}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n-t_{n-1}}^{\mathcal{X}}$$

und

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = (\mu_{t_1-t_0}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n-t_{n-1}}^{\mathcal{X}})(X_{t_0}, \cdot).$$

Außerdem werden die Chapman-Kolmogorov Gleichungen zu

$$\begin{aligned} \mu_s^{\mathcal{X}}\mu_t^{\mathcal{X}} &= \mu_{s+t}^{\mathcal{X}}, \\ T_s^{\mathcal{X}}T_t^{\mathcal{X}} &= T_{s+t}^{\mathcal{X}} \end{aligned}$$

für alle $s, t \in I$. Die starke Markov-Eigenschaft ist in diesem Falle

$$\begin{aligned} \mathbf{P}[X_{S+t} \in A|\mathcal{F}_S] &= \mu_t(X_S, A), \\ \mathbf{E}[f(X_{S+t})|\mathcal{F}_S] &= (T_t f)(X_S) \end{aligned}$$

für alle fast sicher endlichen Stoppzeiten S, $A \in \mathcal{B}(E)$ bzw. $f \in \mathcal{B}(E)$.

[21] Eine Halbgruppe ist ein Paar $(I, *)$, wobei $*$ eine zweistellige, assoziative Verknüpfung auf I ist.

Bemerkung 16.21 (Halbgruppen-Eigenschaft). Sei $(\mu_t^{\mathcal{X}})_{t\in I}$ die Übergangshalbgruppe und $(T_t^{\mathcal{X}})_{t\in I}$ die Operatorhalbgruppe eines zeitlich homogenen Markov-Prozesses $\mathcal{X}$. Dann gilt wegen der Chapman-Kolmogorov Gleichungen

$$\begin{aligned} \mu_s^{\mathcal{X}}\mu_t^{\mathcal{X}} &= \mu_{s+t}^{\mathcal{X}}, \\ T_s^{\mathcal{X}}T_t^{\mathcal{X}} &= T_{s+t}^{\mathcal{X}} \end{aligned}$$

für alle $s, t \in I$. Aus diesem Grund spricht man von (kommutativen) Übergangs- und Operator-*Halbgruppen.*

Bestimmte Eigenschaften von Operatorhalbgruppen erleichtern oftmals Beweise. Dies führt nun auf den Begriff der Feller-Halbgruppe. Um uns Schreibarbeit zu sparen, verwenden wir die Verteilungen $\mathbf{P}_x$ aus Korollar 16.18 und bezeichnen den Erwartungswert bezüglich dieser Verteilung mit $\mathbf{E}_x$.

Definition 16.22 (Feller-Halbgruppe, Feller-Prozess). *Sei $I = \mathbb{R}_+$.*

1. *Sei $(T_t)_{t\in I}$ eine Familie von Operatoren mit $T_t : \mathbf{B}(E) \to \mathcal{B}(E)$. Diese heißt eine* Operatorhalbgruppe*, falls $T_t(T_s f) = T_{t+s} f$ für alle $f \in \mathcal{B}(E)$. Eine solche Halbgruppe heißt*
   (a) positiv*, falls $T_t f \geq 0$ falls $f \geq 0$ für alle $t \in I$,*
   (b) Kontraktion*, falls $0 \leq T_t f \leq 1$ für $0 \leq f \leq 1$ für a*
   (c) konservativ*, falls $T_t 1 = 1$ für alle $t \in I$,*
   (d) stark stetig*, falls $||T_t f - f||_\infty \xrightarrow{t\to 0} 0$ für alle $f \in \mathcal{C}_b(E)$.*
   (e) Feller-Halbgruppe*, falls $T_t f(x) \xrightarrow{t\to 0} f(x)$ für $x \in E$ und $f \in \mathcal{C}_b(E)$ und $T_t f \in \mathcal{C}_b(E)$ für alle $f \in \mathcal{C}_b(E)$ und $t \in I$ gilt.*
2. *Ein zeitlich homogener Markov-Prozess $\mathcal{X} = (X_t)_{t\in\mathcal{I}}$ heißt* Feller-Prozess*, falls seine Operator-Halbgruppe $(T_t^{\mathcal{X}})_{t\in I}$ eine Feller-Halbgruppe ist.*

Bemerkung 16.23 (Probabilistische Eigentschaften von Feller-Prozessen). Sei $I = \mathbb{R}_+$ und $(T_t^{\mathcal{X}})_{t\in I}$ die Operator-Halbgruppe eines Markov-Prozesses $\mathcal{X} = (X_t)_{t\in I}$.

1. Die Halbgruppe $(T_t^{\mathcal{X}})_{t\in I}$ ist konservativ und eine positive Kontraktion.
   Denn: Natürlich ist $T_t^{\mathcal{X}} 1(x) = \mathbf{E}_x[1] = 1$, was die Konservativität von $(T_t^{\mathcal{X}})_{t\in I}$ zeigt. Ganz ähnlich schreibt man für $f \in \mathcal{B}(E)$ mit $0 \leq f \leq 1$
   $$T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(x)] \leq \mathbf{E}_x[1] = 1$$
   und damit ist $(T_t^{\mathcal{X}})_{t\in I}$ eine Kontraktion.
2. Sei $X_0 = x$. Dann ist $T_t^{\mathcal{X}} f(x) \xrightarrow{t\to 0} f(x)$ für alle $f \in \mathcal{C}_b(E)$ genau dann, wenn $X_t \xrightarrow{t\to 0}_p x$.
   Denn: '⇒': Es folgt mit $g(y) := r(x, y) \wedge 1$, dass $\mathbf{E}_x[r(x, Y_t) \wedge 1] = T_t^{\mathcal{X}} g(x) \xrightarrow{t\to 0} g(x) = 0$, was die behauptete Konvergenz zeigt. '⇐': Es gilt $X_t \overset{t\to 0}{\Longrightarrow} x$ und damit nach Definition der schwachen Konvergenz für $f \in \mathcal{C}_b(E)$ insbesondere $T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(X_t)] \xrightarrow{t\to\infty} \mathbf{E}_x[f(x)] = f(x)$.

Lemma 16.24 (Poisson-Prozess und Brown'sche Bewegung sind Feller). *Sowohl der Poisson-Prozess (mit Rate $\lambda \geq 0$) als auch die Brown'sche Bewegung sind Feller-Prozesse.*

Beweis. Sei $\mathcal{X}^x = (X_t^x)_{t\geq 0}$ ein Poisson-Prozess und $\mathcal{Y}^y = (Y_t^y)_{t\geq 0}$ eine Brown'sche Bewegung, jeweils gestartet in $x \in \mathbb{R}$ und $y \in \mathbb{R}$. Es gilt $\mathcal{X}^x \stackrel{d}{=} x + \mathcal{X}^0$ sowie $\mathcal{Y}^y \stackrel{d}{=} y + \mathcal{Y}^0$. Dann gilt $X_t^x \sim N(x,t)$ und $Y_t^y \sim y + \text{Poi}(t\lambda)$. Insbesondere ist offenbar $X_t \xrightarrow{t\to 0}_p x, Y_t \xrightarrow{t\to 0}_p$. Deshalb gilt $T_t^{\mathcal{X}} f(x) \xrightarrow{t\to 0} f(x)$ und $T_t^{\mathcal{Y}} f(y) \xrightarrow{t\to 0} f(y)$ für $f \in \mathcal{C}_b(\mathbb{R})$ nach Bemerkung 16.23.2. Weiter gilt

$$T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(X_t)] = \mathbf{E}_0[f(x + X_t)] \xrightarrow{x\to x'} \mathbf{E}_0[f(x' + X_t)] = T_t^{\mathcal{X}} f(x')$$

und analog für den Prozess $\mathcal{Y}$. Daraus folgen alle Behauptungen. □

Für konkrete Markov-Prozesse sind Halbgruppen meist schwer angebbar. (Siehe jedoch die Ausnahmen des Poisson-Prozesses und der Brown'schen Bewegung aus Beispiel 16.4.) Einfacher fällt zumeist zu definieren, was in infinitesimal kurzer Zeit passiert. Dies wird durch den Generator der Operatorhalbgruppe beschrieben.

Definition 16.25 (Generator). *Sei $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t\in I}$ ein zeitlich homogener Markov-Prozess mit Operatorhalbgruppe $(T_t^{\mathcal{X}})_{t\in I}$. Dann ist der* Generator *von $\mathcal{X}$ (oder von dessen Operatorhalbgruppe) definiert als*

$$(G^{\mathcal{X}} f)(x) = \lim_{t\to 0} \frac{\mathbf{E}_x[f(X_t) - f(x)]}{t} = \lim_{t\to 0} \frac{1}{t}((T_t^{\mathcal{X}} f)(x) - f(x)),$$

für alle f für die der Grenzwert existiert. Die Menge der Funktionen f, für die $(G^{\mathcal{X}} f)(x)$ für alle $x \in E$ existiert, ist der Domain von $G^{\mathcal{X}}$ *und wird mit $\mathcal{D}(G^{\mathcal{X}})$ bezeichnet.*

Beispiel 16.26 (Generator für Poisson-Prozesses und Brown'sche Bewegung).

1. Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Poisson-Prozess mit Parameter λ und $G^{\mathcal{X}}$ sein Generator. Dann gilt
$$(G^{\mathcal{X}} f)(x) = \lambda(f(x+1) - f(x)$$
für $x \in \mathbb{N}$ und $f \in \mathcal{B}(\mathbb{N})$.

   Denn wir berechnen, falls P_t eine Poisson-verteilte Zufallsvariable ist mit Parameter λt
$$\begin{aligned}(G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f(x+P_t) - f(x)]) = \lim_{t\to 0} \frac{1}{t} \sum_{k=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}(f(x+k) - f(x)) \\ &= \lim_{t\to 0} \lambda \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{(k+1)!}(f(x+1+k) - f(x)) \\ &= \lambda(f(x+1) - f(x)\end{aligned}$$
wegen majorisierter Konvergenz.

2. Sei $\mathcal{X} = (X_t)_{t\in I}$ eine Brown'sche Bewegung und $G^{\mathcal{X}}$ ihr Generator. Dann gilt
$$(G^{\mathcal{X}} f)(x) = \tfrac{1}{2} f''(x)$$
für $x \in \mathbb{R}$ und $f \in \mathcal{C}_b^2(\mathbb{R})$, die Menge der beschränkten, zweimal stetig differenzierbaren Funktionen mit beschränkten Ableitungen.

Denn wir berechnen, falls Z eine $N(0,1)$-verteilte Zufallsvariable ist mit der Taylor-Approximation und einer Zufallsvariable Y mit $|Y| \le |Z|$

$$\begin{aligned}
(G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f(x+\sqrt{t}Z) - f(x)]) \\
&= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f'(x)\sqrt{t}Z + \tfrac{1}{2}f''(x)tZ^2 + \tfrac{1}{2}(f''(x+\sqrt{t}Y) - f''(x))tZ^2]) \\
&= \tfrac{1}{2}f''(x) + \lim_{t\to\infty} \mathbf{E}[\tfrac{1}{2}(f''(x+\sqrt{t}Y) - f''(x))Z^2] = \tfrac{1}{2}f''(x)
\end{aligned} \tag{16.5}$$

mit majorisierter Konvergenz.

Analog berechnet man: Ist $\mathcal{X} = (X_t)_{t\in I}$ mit $X_t = (X_t^1, ..., X_t^d)$ eine d-dimensionale Brown'sche Bewegung. Dann gilt

$$(G^{\mathcal{X}} f)(x) = \frac{1}{2}\sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2}(x)$$

für $x \in \mathbb{R}^d$ und $f \in \mathcal{C}_b^2(\mathbb{R}^2)$.

Bemerkung 16.27 (Feller-Halbgruppen und starke Stetigkeit). Ist E zumindest lokal kompakt, so kann man – wenn man $\mathcal{C}_b(E)$ durch $\mathcal{C}_0(E)$, die bei unendlich verschwindenden Funktionen ersetzt – immerhin zeigen, dass jede Feller-Halbgruppe stark stetig ist. Dies erleichtert in einigen Beweisen das Nachprüfen der uniformen Konvergenz für die starke Stetigkeit. Insbesondere sind nach Lemma 16.24 die (Feller-)Halbgruppen des Poisson-Prozesses und der Brown'schen Bewegung stark stetig.

Lemma 16.28 (Zusammenhang zwischen Operator-Halbgruppe und Generator). *Sei $\mathcal{X}$ ein Feller-Prozess mit Operator-Halbgruppe $(T_t^{\mathcal{X}})_{t\in I}$. Weiter sei $G^{\mathcal{X}}$ der Generator von $\mathcal{X}$ und $\mathcal{D} \subseteq \mathcal{D}(G^{\mathcal{X}})$ mit $G^{\mathcal{X}}(\mathcal{D}) \subseteq \mathcal{C}_b(E)$. Für $f \in \mathcal{C}_b(E)$ ist dann $\int_0^t (T_s^{\mathcal{X}} f)ds \in \mathcal{D}(G^{\mathcal{X}})$ mit*

$$(T_t^{\mathcal{X}} f)(x) - f(x) = \Big(G^{\mathcal{X}}\Big(\int_0^t (T_s^{\mathcal{X}} f)ds\Big)\Big)(x) \tag{16.6}$$

und für $f \in \mathcal{D}$ und $t \ge 0$ ist auch $T_t^{\mathcal{X}} f \in \mathcal{D}(G^{\mathcal{X}})$ und es gilt

$$\begin{aligned}
G^{\mathcal{X}}(T_t^{\mathcal{X}} f) &= T_t^{\mathcal{X}}(G^{\mathcal{X}} f), \\
(T_t^{\mathcal{X}} f)(x) - f(x) &= \int_0^t (T_s^{\mathcal{X}}(G^{\mathcal{X}} f))(x)ds,
\end{aligned} \tag{16.7}$$

also

$$\mathbf{E}_x[f(X_t)] = f(x) + \int_0^t \mathbf{E}[(G^{\mathcal{X}} f)(X_s)]ds.$$

Beweis. Für $x \in E$ und $f \in \mathcal{C}_b(E)$ ist $t \mapsto (T_t^{\mathcal{X}} f)(x)$ stetig. Es gilt nämlich wegen der Feller-Eigenschaft von $(T_t^{\mathcal{X}})_{t\in I}$

$$(T_{t+h}^{\mathcal{X}} f)(x) = (T_t^{\mathcal{X}}(T_h^{\mathcal{X}} f))(x) = (T_t^{\mathcal{X}} f)(x).$$

Für die erste Gleichung ist nun

$$\begin{aligned}\frac{1}{h}\mathbf{E}_x\Big[\int_0^t (T_s^{\mathcal{X}} f)(X_h) - (T_s^{\mathcal{X}} f)(x)ds\Big] &= \frac{1}{h}\Big(\int_0^t (T_{s+h}^{\mathcal{X}} f)(x) - (T_s^{\mathcal{X}} f)(x)ds\Big)\\ &= \frac{1}{h}\Big(\int_h^{t+h} (T_s^{\mathcal{X}} f)(x)ds - \int_0^t (T_s^{\mathcal{X}} f)(x)ds\Big)\\ &= \frac{1}{h}\int_t^{t+h} (T_s^{\mathcal{X}} f)(x)ds - \frac{1}{h}\int_0^h (T_s^{\mathcal{X}} f)(x)ds\\ &\xrightarrow{h\to 0} (T_t^{\mathcal{X}} f)(x) - f(x).\end{aligned}$$

Für die anderen Aussagen ist zunächst

$$\begin{aligned}\frac{d}{dt}\mathbf{E}_x[f(X_t)] &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_{t+h}) - f(X_t)]\\ &= (T_t^{\mathcal{X}} \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_h) - f(x)] = (T_t^{\mathcal{X}}(G^{\mathcal{X}} f))(x),\end{aligned}$$

aber auch

$$\begin{aligned}\frac{d}{dt}\mathbf{E}_x[f(X_t)] &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_{t+h}) - f(X_t)]\\ &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[(T_t^{\mathcal{X}} f)(X_h) - (T_t^{\mathcal{X}} f)(x)] = (G^{\mathcal{X}}(T_t^{\mathcal{X}} f))(x),\end{aligned}$$

was die erste Gleichung zeigt. Für die zweite Gleichung bemerken wir, dass $t \mapsto (T_t^{\mathcal{X}}(G^{\mathcal{X}} f))(x)$ nach Voraussetzung stetig ist, also folgt

$$(T_t^{\mathcal{X}} f)(x) - f(x) = \int_0^t \frac{d}{ds}\mathbf{E}_x[f(X_s)]ds = \int_0^t (T_s^{\mathcal{X}}(G^{\mathcal{X}} f))(x)ds.$$

□

Korollar 16.29 (Domain ist dicht). *Sei $\mathcal{X}$, $(T_t^{\mathcal{X}})_{t\in I}$ und $G^{\mathcal{X}}$ wie in Lemma 16.28 und die Voraussetzungen in Lemma 16.28 gelten mit $\mathcal{D} = \mathcal{C}_b(E)$. Weiter sei $(T_t^{\mathcal{X}})_{t\in I}$ stark stetig. Dann ist $\mathcal{D}(G^{\mathcal{X}})$ dicht in $\mathcal{C}_b(E)$ bezüglich der Supremumsnorm, d.h. jedes $f \in \mathcal{C}_b(E)$ lässt sich durch Funktionen aus $\mathcal{D}(G^{\mathcal{X}})$ approximieren.*

Beweis. Für jedes $f \in \mathcal{C}_b(E)$ gilt nach Voraussetzung

$$\tfrac{1}{t}\int_0^t (T_s^{\mathcal{X}} f)ds \xrightarrow{t\to 0} f$$

bezüglich der Supremumsnorm. Da die Funktion auf der linken Seite nach (16.6) in $\mathcal{D}(G^{\mathcal{X}})$ liegen, ist die Behauptung gezeigt. □

Theorem 16.30 (Von Markov-Prozessen abgeleitete Martingale). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein Feller-Prozess mit Generator $G^{\mathcal{X}}$ und Domain $\mathcal{D}(G^{\mathcal{X}})$. Weiter sei $\mathcal{D} \subseteq \mathcal{D}(G^{\mathcal{X}})$ so, dass $G^{\mathcal{X}}(\mathcal{D}) \subseteq \mathcal{C}_b(E)$. Dann sind für $f \in \mathcal{D}$ sowohl*

$$\Big(f(X_t) - \int_0^t (G^{\mathcal{X}} f)(X_s)ds\Big)_{t\in I}$$

als auch, im Falle $(G^{\mathcal{X}}f)/f \in L$

$$\Big(f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_s)}{f(X_s)}ds\Big)\Big)_{t\in I}$$

Martingale.

Beweis. Sei $t \geq s$. Für den ersten Prozess bemerken wir

$$\begin{aligned}
\mathbf{E}\Big[f(X_t) - f(X_s) - \int_s^t (G^{\mathcal{X}}f)(X_r)dr\Big|\mathcal{F}_s\Big]&\\
&= \mathbf{E}\Big[f(X_t) - f(X_s) - \int_s^t (G^{\mathcal{X}}f)(X_r)dr\Big|X_s\Big]\\
&= (T_{t-s}f)(X_s) - f(X_s) - \int_s^t (T_{r-s}(G^{\mathcal{X}}f))(X_s)dr = 0
\end{aligned}$$

nach Lemma 16.28. Weiter ist

$$\begin{aligned}
&\mathbf{E}_x\Big[f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big) - f(X_s)\exp\Big(-\int_0^s \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\Big|\mathcal{F}_s\Big]\\
&\quad = \mathbf{E}_x\Big[f(X_t)\exp\Big(-\int_s^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big) - f(X_s)\Big|X_s\Big]\cdot\exp\Big(-\int_0^s \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)
\end{aligned}$$

und

$$\begin{aligned}
&\frac{d}{dt}\mathbf{E}_{X_s}\Big[f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\Big]\\
&\qquad = \mathbf{E}_{X_s}\Big[(G^{\mathcal{X}}f)(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\\
&\qquad\qquad\qquad - f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\frac{(G^{\mathcal{X}}f)(X_t)}{f(X_t)}\Big] = 0.
\end{aligned}$$

Wieder liefert Integration von s bis t die Behauptung. □

Beispiel 16.31 (Gewöhnliche Differentialgleichung). Sei $\mathcal{X} = (X_t)_{t\geq 0}$ die zeitliche Entwicklung mit Werten in $\mathbb{R}^d$, die durch die gewöhnliche Differentialgleichung

$$\frac{d}{dt}X_t = g(X_t)$$

gegeben ist, wobei $g = (g_i)_{i=1,\dots,d} : \mathbb{R}^d \to \mathbb{R}^d$ eine beschränkte Lipschitz-Funktion ist. Dann ist $\mathcal{X}$ zwar deterministisch, kann aber eben auch als zeitlich homogener (weil g nicht zusätzlich von t abhängt) Markov-Prozess gesehen werden. Den Generator von $\mathcal{X}$ berechnet sich für $f \in \mathcal{C}_b^1(\mathbb{R}^d)$ und $X_0 = x$ als

$$(G^{\mathcal{X}}f)(x) = \lim_{t\to 0}\tfrac{1}{t}(f(X_t) - f(x)) = \frac{d}{dt}(f(X_t))\Big|_{t=0} = \sum_{i=1}^d \frac{\partial f}{\partial x_i}(g(x))\cdot g_i(x) = (\nabla f)(g(x))\cdot g(x).$$

Beispiel 16.32 (Poisson-Prozess und Brown'scher Bewegung). Im folgenden sei $f_n(x) = xe^{-x/n}$, also $f_n \in \mathcal{C}_b(\mathbb{R}_+)$ und $g_n(x) = x^2e^{-x/n}$ so dass $f_n(x) \xrightarrow{n\to\infty} f(x)$ und $g_n(x) \xrightarrow{n\to\infty} g(x)$ mit $f(x) = x$ und $g(x) = x^2$.

1. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Poisson-Prozess mit Rate λ. Damit ist nach Theorem 16.30 und Beispiel 16.26
$$\big(X_t \wedge n - \int_0^t \lambda 1_{X_s \leq n-1} ds\big)_{t\geq 0}$$
ein Martingal. Da X_t integrierbar ist, folgt nach majorisierter Konvergenz auch, dass
$$\big(X_t - \lambda t\big)_{t\geq 0}$$
ein Martingal ist. Analog folgert man (aus der Integrierbarkeit von X_t^2, dass
$$\big(X_t^2 - \lambda \int_0^t (X_s+1)^2 - X_s^2 ds\big)_{t\geq 0}$$
ein Martingal ist. Siehe auch Beispiel 15.4.

2. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung. Aus der Integrierbarkeit von X_t, X_t^2 und $e^{\mu X_t}$ folgert man aus Theorem 16.30, dass wegen $G^{\mathcal{X}} h(x) = \frac{1}{2}h''(x)$
$$\begin{aligned}
&\Big(X_t - \tfrac{1}{2}\int_0^t id''(X_s)ds\Big)_{t\geq 0} = (X_t)_{t\geq 0},\\
&\Big(X_t^2 - \tfrac{1}{2}\int_0^t (id^2)''(X_s)ds\Big)_{t\geq 0} = (X_t^2 - t)_{t\geq 0},\\
&\Big(\exp\Big(\mu X_t - \tfrac{1}{2}\int_0^t \frac{(e^{\mu\cdot})''(X_s)}{e^{\mu X_s}}ds\Big)\Big)_{t\geq 0} = \Big(\exp\Big(\mu X_t - \tfrac{1}{2}\mu^2 t\Big)\Big)_{t\geq 0}
\end{aligned}$$
alles Martingale sind. Siehe auch Beispiel 15.5.

Beispiel 16.33 (Sprungprozess). Mit die einfachsten Markov-Prozesse sind Prozesse, die im Zustandsraum E stückweise konstant sind. Wir beschreiben nun den folgenden Prozess: Gegeben $X_s = x$ springt der Prozess nach einer exponentialverteilten Zeit mit Rate $\lambda(x)$. Der Prozess springt dabei nach dem Markov-Kern $\mu(X_s, .)$, d.h. er springt mit Wahrscheinlichkeit $\mu(X_s, dy)$ nach y.

Gegeben sei also $\lambda \in \mathcal{B}(E)$ mit $0 \leq \lambda \leq \lambda^*$ und der Markov-Kern μ von E nach E. Weiter sei $(Y_k)_{k=0,1,2,...}$ eine Markov-Kette in diskreter Zeit mit $\mathbf{P}(Y_{k+1} \in A | Y_k) = \mu(Y_k, A)$ für alle $A \in \mathcal{B}(E)$. Weiter seien $T_1, T_2, ...$ unabhängig und exp(1)-verteilt. (Wir bemerken dass damit T_k/λ nach exp(λ)-verteilt ist.) Wir definieren den Sprung-Prozess $(X_t)_{t\geq 0}$ durch
$$X_t = \begin{cases} Y_0, & t < \dfrac{T_0}{\lambda(Y_0)},\\[2ex] Y_k, & \displaystyle\sum_{j=0}^{k-1} \frac{T_j}{\lambda(Y_j)} \leq t < \sum_{j=0}^{k} \frac{T_j}{\lambda(Y_j)}. \end{cases} \tag{16.8}$$
Dies ist ein Markov-Prozess wegen der Gedächtnislosigkeit der Exponentialverteilung. Zur Berechnung des Generators von $\mathcal{X}$ bemerken wir, dass die Wahrscheinlichkeit, dass in Zeit t mehr als 2 Sprünge stattfinden, höchstens $1 - e^{-\lambda^* t}(1 + \frac{1}{2}\lambda^* t) = \mathcal{O}(t^2)$ ist. Also gilt für $f \in \mathcal{C}_b(E)$
$$\begin{aligned}
(G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{\mathbf{E}_x[f(X_t) - f(x)]}{t}\\
&= \lim_{t\to 0} \tfrac{1}{t}\Big((e^{-\lambda(x)t} - 1)f(x) + \lambda(x) t e^{-\lambda(x)t}\int \mu(x,dy)f(y)\Big) \qquad (16.9)\\
&= \lambda(x)\int \mu(x,dy)\big(f(y) - f(x)\big)dy.
\end{aligned}$$

Beispiel 16.34 (Master-Gleichung). Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Sprungprozess auf einem endlichen Zustandsraum E, gegeben wie im letzten Beispiel durch die Funktionen λ und den Markov-Kern $\mu(.,.)$. Setzt man in (16.9) die Gleichung die Funktione $f(y) = 1_{y=x}$ (für ein festes x) ein, so erhält man

$$\begin{aligned}\frac{d}{dt}\mathbf{P}(X_t = x) &= \frac{d}{dt}\mathbf{E}[f(X_t)] = \mathbf{E}[(G^{\mathcal{X}}f)(X_t)] \\ &= \mathbf{E}\Big[\lambda(X_t)\int \mu(X_t, dy)(1_{y=x} - 1_{X_t=x})\Big] \\ &= \sum_{z\in E}\mathbf{P}(X_t = z)\lambda(z)\sum_{y\in E}\mu(z,y)(1_{x=y} - 1_{x=z}) \\ &= \Big(\sum_{z\in E}\mathbf{P}(X_t = z)\lambda(z)\mu(z,x)\Big) - \mathbf{P}(X_t = x)\lambda(x).\end{aligned}$$

Diese Gleichung ist also eine Differentialgleichung für $(\mathbf{P}(X_t = x))_{x\in E}$. Die Lösung dieser Gleichung liefert somit die genaue Verteilung von X_t. Wegen ihrer vielfältigen Einsatz-Möglichkeiten heißt diese Gleichung in der Physik auch *Master-Gleichung*.

17 Eigenschaften der Brown'schen Bewegung

Zwar haben wir in Kapitel 14.3 schon die Brown'sche Bewegung eingeführt, jedoch gibt es noch einiges über sie zu berichten. Wir wissen bereits, dass die Brown'sche Bewegung ein Martingal, ein Gauss-Prozess sowie ein starker Markov-Prozess mit unabhängigen und identisch verteilten Zuwächsen ist und stetige Pfade hat. Aus dem bereits gezeigten können wir neue Eigenschaften ablesen, beispielsweise das Blumenthal'sche 0-1-Gesetz, das als Ergänzung zum Kolmogorov'schen 0-1-Gesetz verstanden werden kann.

Theorem 17.1 (Blumenthal'sches 0-1-Gesetz). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung bezüglich eines Wahrscheinlichkeitsraumen $(\Omega, \mathcal{F}, \mathbf{P})$, gestartet in $x \in \mathbb{R}$, und $\mathcal{F}_{0+} := \bigcap_{t>0}\sigma(X_s : s \leq t)$. Dann ist $\mathcal{F}_{0+}$ $\mathbf{P}$-trivial, d.h. es gilt $\mathbf{P}(A) \in \{0, 1\}$ für $A \in \mathcal{F}_{0+}$.*

Sei weiter $\mathcal{T} := \bigcap_{s\geq 0}\sigma(X_t : t \geq s)$ die terminale σ-Algebra von $\mathcal{X}$. Dann ist $\mathcal{T}$ $\mathbf{P}$-trivial.

Beweis. Nach Lemma 16.9 ist die Filtration $(\mathcal{F}_t)_{t\geq 0}$ mit $\mathcal{F}_t = \sigma(X_s : s \leq t)$ rechtsstetig. Aus der Rechtsstetigkeit in 0 folgt $\mathcal{F}_{0+} = \sigma(X_0)$. Da $X_0 = x$ konstant ist, muss also $\mathcal{F}_{0+}$ eine $\mathbf{P}$-triviale σ-Algebra sein.

Weiter ist mit $\mathcal{X}$ nach Theorem 14.19 auch $\mathcal{X}' = (X'_t)_{t\geq 0}$ mit $X'_t = tX_{1/t}$ eine in 0 gestartete Brown'sche Bewegung. Mit dem eben gezeigten gilt, dass $\bigcap_{t\geq 0}\sigma(X'_s : s \leq t)$ $\mathbf{P}$-trivial ist. Daraus folgt aber, dass

$$\bigcap_{s\geq 0}\sigma(X_t : t \geq s) = \bigcap_{s\geq 0}\sigma(tX_{1/t} : t \leq s) = \bigcap_{s\geq 0}\sigma(X'_t : t \leq s)$$

$\mathbf{P}$-trivial ist, also die Behauptung. □

Bemerkung 17.2. Obwohl das Blumenthal'sche 0-1-Gesetz einfach aussieht, mag es dennoch überraschen. Wie wir später zeigen werden, darf man sich die Brown'sche Bewegung – in geeignetem Sinne – als Grenzwert von Irrfahrten vorstellen. Starten wir Irrfahrten in 0, dann ist es ja so, dass Irrfahrtspfade entweder erst nach oben oder erst nach unten springen. Insbesondere verbringen sie für kleine Zeiten entweder mehr Zeit im Positiven oder im Negativen.

Definieren wir analog hierzu für die Brown'sche Bewegun

$$A_t := \Big\{ \int_0^t 1_{X_s>0} ds \geq \int_0^t 1_{X_s<0} ds \Big\}$$

die Menge der Brown'schen Pfade, die bis zur Zeit t mehr Zeit im Positiven verbracht haben, sowie $A := \bigcap_{t>0} \bigcap_{0<s\leq t} A_s$, das ist die Menge der Pfade, die bis zu irgendeiner kleinen Zeit t mehr Zeit im Positiven verbracht haben. Dann ist $A \in \mathcal{F}_{0+}$, also muss aus Symmetriegründen $\mathbf{P}(A) = 0$ gelten. Es gibt also fast sicher keinen Brown'schen Pfad, der in noch so kurzer Zeit immer mehr Zeit im Positiven verbracht hat, obwohl dies für Irrfahrten doch der Fall ist.

Dieses Gesetz ist jedoch nur der Auftakt einer Reihe weiterer Eigenschaften. Wir untersuchen hier die quadratische Variation in Abschnitt 17.1, das auf der starken Markov-Eigenschaft basierende Reflexionsprinzip in Abschnitt 17.2, das Gesetz des iterierten Logarithmus in Abschnitt 17.3, die Konvergenz von Irrfahrten gegen die Brown'sche Bewegung in Abschnitt 17.4 und einen Zusammenhang mit zentrierten Zufallsvariablen in Abschnitt 17.5.

17.1 Quadratische Variation

Die Pfade der Brown'schen Bewegung in den Abbildungen 14.3 und 16.1 sahen – wenn auch stetig – doch sehr *zackig* aus. Diese Eigenschaft soll nun präzisiert werden.

Definition 17.3 (Variation und quadratische Variation). *Sei $f \in \mathcal{D}_{\mathbb{R}}([0,\infty))$, $t \geq 0$ und für $n = 1, 2, ...$ seien $0 = t_{n,0} < t_{n,1} < \cdots < t_{n,k_n} = t$ gegeben. Wir bezeichnen $\zeta_n := \{t_{n,0}, ..., t_{n,k_n}\}$ als n-te Partition (von $[0,t]$). Angenommen $\max_k(t_{n,k} - t_{n,k-1}) \xrightarrow{n\to\infty} 0$, so schöpfen die Partitionen das Intervall $[0,t]$ für $n \to \infty$ immer besser aus. Dann definieren wir die ℓ-Variation von f bezüglich $\zeta = (\zeta_n)_{n=1,2,...}$ als*

$$\nu_{\ell,t,\zeta}(f) := \lim_{n\to\infty} \nu^n_{\ell,t,\zeta}(f)$$

mit

$$\nu^n_{\ell,t,\zeta}(f) = \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^\ell.$$

Falls der Grenzwert unabhängig von ζ ist, ist dies gleich der ℓ-Variation und wird mit $\nu_{\ell,t}(f)$ bezeichnet. Die 1-Variation wird auch als Variation und die 2-Variation auch als quadratische Variation bezeichnet.

Außerdem heißt ζ aufsteigend*, falls $\zeta_n \subseteq \zeta_{n+1}$ für alle $n = 1, 2, ...$ gilt.*

Lemma 17.4 (Elementare Eigenschaften der (quadratischen) Variation). *Sei f stetig und $t \geq 0$. Dann gilt für ζ wie in Definition 17.3*

$$\nu_{\ell,t,\zeta}(f) < \infty \Rightarrow \nu_{\ell+1,t,\zeta}(f) = 0,$$
$$\nu_{\ell+1,t,\zeta}(f) > 0 \Rightarrow \nu_{\ell,t,\zeta}(f) = \infty.$$

Beweis. Es genügt die erste Eigenschaft zu zeigen. Wir schreiben

$$\begin{aligned} 0 &\leq \lim_{n\to\infty} \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^{\ell+1} \\ &\leq \lim_{n\to\infty} \sup_k |f(t_{n,k}) - f(t_{n,k-1})| \cdot \lim_{n\to\infty} \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^{\ell} = 0 \end{aligned}$$

da f auf $[0,t]$ gleichmäßig stetig ist. □

Proposition 17.5 (Quadratische Variation der Brown'schen Bewegung). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung. Dann gilt für ζ wie in Definition 17.3*

$$\nu^n_{2,t,\zeta}(\mathcal{X}) \xrightarrow{n\to\infty}_{L^2} t.$$

Ist ζ aufsteigend, so gilt auch

$$\nu^n_{2,t,\zeta}(\mathcal{X}) \xrightarrow{n\to\infty}_{fs} t.$$

Insbesondere ist die Variation von $\mathcal{X}$ fast sicher unendlich.

Beweis. Wir schreiben $\nu^n_{2,t,\zeta} := \nu^n_{2,t,\zeta}(\mathcal{X})$. Zunächst zur L^2-Konvergenz. Es ist bekannt, dass $X_t - X_s \sim \sqrt{t-s}X_1$ für $s \leq t$ gilt. Also ist

$$\mathbf{E}[\nu^n_{2,\zeta}] = \sum_{k=1}^{k_n} \mathbf{E}[(X_{t_{n,k}} - X_{t_{n,k-1}})^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1})\mathbf{E}[X_1^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1}) = t$$

sowie

$$\mathbf{E}[(\nu^n_{2,\zeta} - t)^2] = \mathbf{V}[\nu^n_{2,\zeta}] = \sum_{k=1}^{k_n} \mathbf{V}[(X_{n,k} - X_{n,k-1})^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1})^2 \mathbf{E}[X_1^4] \xrightarrow{n\to\infty} 0.$$

Für die fast sichere Konvergenz nehmen wir zunächst ohne Beschränkung an, dass es $0 \leq t_1, t_2, ... \leq t$ gibt, so dass $\zeta_n = \{t_1, ..., t_n\}$. Weiter werden wir zeigen, dass $(\nu^{-n}_{2,\zeta})_{n=...,-2,-1}$ ein (Rückwärts-)Martingal ist, also dass

$$\mathbf{E}[\nu^{n-1}_{2,\zeta} - \nu^n_{2,\zeta} | \nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] = 0$$

für alle n gilt. Sind t'_n und t''_n die Zeitpunkte direkt vor und nach t_n in ζ_n, so gilt

$$\begin{aligned} \nu^{n-1}_{2,\zeta} - \nu^n_{2,\zeta} &= (X_{t''_n} - X_{t'_n})^2 - (X_{t''_n} - X_{t_n})^2 - (X_{t_n} - X_{t'_n})^2 \\ &= 2(X_{t''_n} - X_{t_n})(X_{t_n} - X_{t'_n}). \end{aligned}$$

Wir definieren eine zweite Brown'sche Bewegung $(\widetilde{X}_t)_{t\geq 0}$ durch eine unabhängige Zufallsvariable Y mit $\mathbf{P}(Y = 1) = \mathbf{P}(Y = -1) = \frac{1}{2}$ und

$$\widetilde{X}_s = X_{s\wedge t_n} + Y(X_s - X_{s\wedge t_n}).$$

Das bedeutet, dass $(\widetilde{X}_s)_{0\leq s\leq t}$ nach t_n an X_{t_n} gespiegelt ist. Insbesondere gilt $(X_{t_n} - X_{t'_n}) = (\widetilde{X}_{t_n} - \widetilde{X}_{t'_n})$ und $(X_{t''_n} - X_{t_n}) = -(\widetilde{X}_{t''_n} - \widetilde{X}_{t_n})$. Es ist $\nu^k_{2,t,\zeta}(\mathcal{X}) = \nu^k_{2,t,\zeta}(\widetilde{\mathcal{X}})$ für $k = n, n+1, ...$ und damit

$$\mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\mathcal{X}) - \nu^n_{2,t,\zeta}(\mathcal{X})|\nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] = \mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\widetilde{\mathcal{X}}) - \nu^n_{2,t,\zeta}(\widetilde{\mathcal{X}})|\nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...],$$

also

$$\begin{aligned}\mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\mathcal{X}) &- \nu^n_{2,t,\zeta}(\mathcal{X})|\nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] \\ &= \tfrac{1}{2}\big(\mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\mathcal{X}) - \nu^n_{2,t,\zeta}(\mathcal{X})|\nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] + \mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\widetilde{\mathcal{X}}) - \nu^n_{2,t,\zeta}(\widetilde{\mathcal{X}})|\nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...]\big) \\ &= \mathbf{E}[(X_{t''_n} - X_{t_n})(X_{t_n} - X_{t'_n}) + (\widetilde{X}_{t''_n} - \widetilde{X}_{t_n})(\widetilde{X}_{t_n} - \widetilde{X}_{t'_n})|\nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] = 0,\end{aligned}$$

was die gewünschte Martingaleigenschaft zeigt. Nach Theorem 15.36 konvergiert also $(\nu^n_{2,t,\zeta})_{n=1,2,...}$ auch fast sicher gegen t. □

Korollar 17.6 (Brown'sche Bewegung hat nirgends differenzierbare Pfade). *Eine Brown'sche Bewegung $\mathcal{X} = (X_t)_{t\geq 0}$ hat fast sicher nirgends differenzierbare Pfade. Das bedeutet, dass*

$$\mathbf{P}\Big(\lim_{h\to 0}\frac{X_{t+h} - X_t}{h} \text{ existiert für ein } t > 0\Big) = 0.$$

Beweis. Es genügt die Menge der Pfade der Brown'schen Bewegung zu betrachten, deren quadratische Variation in der Zeit $[0, t]$ genau t ist. (Die ist nämlich eine Menge vom Maß 1, wie Proposition 17.5 zeigt.) Jeder Pfad in dieser Menge hat in jedem kleinen Zeitintervall positive quadratische Variation, also nach Lemma 17.4 unendliche Variation. Da Differenzierbarkeit zumindest eine endliche Variation in einem kleinen Zeitintervall voraussetzt, folgt die Behauptung. □

17.2 Starke Markov-Eigenschaft und Reflexionsprinzip

In Beispiel 16.13 haben wir gesehen, dass die Brown'sche Bewegung ein starker Markov-Prozess ist. Dies hat einige nützliche Konsequenzen, die wir nun darstellen werden. Das Reflexionsprinzip ist in Abbildung 17.1 veranschaulicht.

Lemma 17.7 (Reflexionsprinzip). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung und T eine Stoppzeit. Dann ist der reflektierte Prozess $\mathcal{X}' = (X'_t)_{t\geq 0}$ mit*

$$X'_t := X_{t\wedge T} - (X_t - X_{t\wedge T}) = \begin{cases} X_t, & t \leq T, \\ 2X_T - X_t, & t > T \end{cases}$$

ebenfalls eine Brown'sche Bewegung.

Beweis. Zunächst ist wegen der Konstruktion klar, dass $\mathcal{X}'$ stetige Pfade hat. Ohne Einschränkung der Allgemeinheit nehmen wir an, dass $T < \infty$ gilt. Wir definieren $\mathcal{Y} = (Y_t)_{t\geq 0}$ durch $Y_t := X_{t\wedge T}$ sowie $\mathcal{Z} = (Z_t)_{t\geq 0}$ durch $Z_t := X_{T+t} - X_T$. Dann ist $\mathcal{Z}$ wegen der starken Markov-Eigenschaft eine Brown'sche Bewegung, die von $(T, \mathcal{Y})$ unabhängig ist. Damit gilt $(T, \mathcal{Y}, \mathcal{Z}) \stackrel{d}{=} (T, \mathcal{Y}, -\mathcal{Z})$, da $\mathcal{Z} \stackrel{d}{=} -\mathcal{Z}$. Damit folgt auch $(\mathcal{Y}, \mathcal{Z}^T) \stackrel{d}{=} (\mathcal{Y}, -\mathcal{Z}^T)$ mit $\mathcal{Z}^T := (Z^T_t)_{t\geq 0}$, $Z^T_t := Z_{(t-T)^+}$. Daraus folgt

$$\mathcal{X} = \mathcal{Y} + \mathcal{Z}^T \stackrel{d}{=} \mathcal{Y} - \mathcal{Z}^T = \mathcal{X}'.$$

Dies zeigt die Behauptung. □

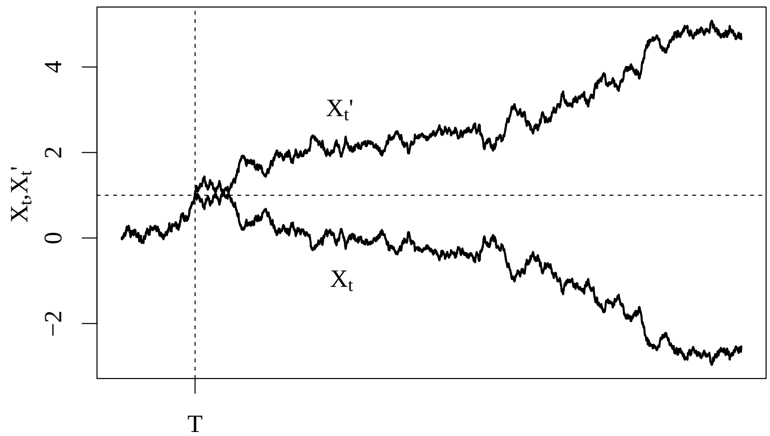


Abbildung 17.1
Das Reflexionsprinzip der Brown'schen Bewegung besagt, dass für eine Brown'sche Bewegung $(X_t)_{t\geq 0}$ der nach T bei $x = X_T$ gespiegelte Prozess $(X'_t)_{t\geq 0}$ ebenfalls eine Brown'sche Bewegung ist.

Als Anwendung des Reflexionsprinzips berechnen wir nun die Verteilung des Maximums einer Brown'schen Bewegung bis zu einer Zeit t. Zunächst bemerken wir jedoch, dass man aus der Doob'schen L^p-Ungleichung, Proposition 15.25, Abschätzungen über die Verteilung des Maximums erhält. Sei hierzu $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung und $\mathcal{M} = (M_t)_{t\geq 0}$ mit $M_t = \sup_{s\leq t} X_s$ der Maximums-Prozess. Dann folgt aus Proposition 15.25 (bzw. der Erweiterung auf zeitstetige Prozesse aus Theorem 15.47) mit $p = 2$

$$\mathbf{P}(M_t \geq x) \leq \tfrac{1}{x^2}\mathbf{E}[X_t^2] = \tfrac{t}{x^2}.$$

Gerade für große x ist diese Wahrscheinlichkeit jedoch in der Tat viel kleiner, wie das nächste Resultat zeigt.

Theorem 17.8 (Maximum der Brown'schen Bewegung). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine (in $X_0 = 0$ gestartete) Brown'sche Bewegung. Wir definieren den Maximums-Prozess $\mathcal{M} = (M_t)_{t\geq 0}$ durch $M_t = \sup_{0\leq s\leq t} X_s$. Dann gilt für*

$$M_t \stackrel{d}{=} M_t - X_t \stackrel{d}{=} |X_t|.$$

Alle drei Zufallsvariablen haben die Dichte

$$x \mapsto \sqrt{\frac{2}{\pi t}} \exp\Big(-\frac{x^2}{2t}\Big) 1_{x\geq 0}.$$

Beweis. Sei $\varphi_t(x) = \frac{1}{\sqrt{2\pi t}} \exp\big(-\frac{x^2}{2t}\big)$ die Dichte der Brown'schen Bewegung zur Zeit t. Dann ist die Dichte von $|X_t|$ gegeben durch $2\varphi_t(x)1_{x\geq 0}$. Also bleibt zu zeigen, dass sowohl M_t and auch $M_t - X_t$ genau diese Dichte besitzen. Hierzu setzen wir $T := T_x = \inf\{s \geq 0 : X_s = x\}$. Für $0, y \leq x$ gilt wegen Lemma 17.7, falls $(X'_t)_{t\geq 0}$ der bei T gespiegelte Prozess ist

$$\mathbf{P}(M_t \geq x, X_t \leq y) = \mathbf{P}(X'_t \geq 2x - y) = \int_{2x-y}^{\infty} \varphi_t(z)dz$$

und damit für $x \geq 0$

$$\begin{aligned}\mathbf{P}(M_t \geq x) &= \mathbf{P}(M_t \geq x, X_t \leq x) + \mathbf{P}(X_t \geq x) \\ &= 2\int_x^\infty \varphi_t(z)dz\end{aligned}$$

woraus folgt, dass $M_t \stackrel{d}{=} |X_t|$. Weiter berechnen wir

$$\begin{aligned}\mathbf{P}(M_t - X_t \geq x) &= \lim_{\varepsilon\to 0} \tfrac{1}{\varepsilon} \int_0^\infty \mathbf{P}(z \leq M_t \leq z+\varepsilon, X_t \leq z-x)dz \\ &= \lim_{\varepsilon\to 0} \tfrac{1}{\varepsilon} \int_0^\infty \mathbf{P}(M_t \geq z, X_t \leq z-x) - \mathbf{P}(M_t \geq z+\varepsilon, X_t \leq z-x)dz \\ &= \lim_{\varepsilon\to 0} \tfrac{1}{\varepsilon} \int_0^\infty 2\varphi_t(z+x)dz = \int_x^\infty 2\varphi(z)dz.\end{aligned}$$

Wieder gilt also $M_t - X_t \stackrel{d}{=} |X_t|$. □

Bemerkung 17.9 (Das pfadwertige Reflexionsprinzip). Das Reflexionsprinzip zeigt nur die Gleichheit der Verteilungen von $|X_t|, M_t, M_t - X_t$ zu einem festen Zeitpunkt t. Es bleibt nun offen, ob etwa auch $(|X_t|)_{t\geq 0} \sim (M_t - X_t)_{t\geq 0}$ gilt. Auch wenn wir das hier nicht zeigen, stellt sich diese Behauptung als richtig heraus. (Nebenbei: Sicher ist $(M_t)_{t\geq 0}$ anders verteilt als $(|X_t|)_{t\geq 0}$ oder $(M_t - X_t)_{t\geq 0}$, da die letzten beiden Prozesse auch fallen können, $(M_t)_{t\geq 0}$ jedoch nicht.)

17.3 Gesetz des iterierten Logarithmus

Wir wollen bestimmen, wie eine Brown'sche Bewegung $\mathcal{X} = (X_t)_{t\geq 0}$ maximal *wächst*. Das bedeutet, dass wir eine Funktion $t \mapsto h_t$ bestimmen wollen, so dass

$$0 < \limsup_{t\to\infty} \frac{X_t}{h_t} < \infty \tag{17.1}$$

gilt. Wir wissen bereits aus dem Gesetz der großen Zahlen, dass $\frac{X_t}{t} \xrightarrow{t\to\infty} 0$. Außerdem gilt

$$\limsup_{t\to\infty} \frac{X_t}{\sqrt{t}} = \infty. \tag{17.2}$$

Denn: Sicher ist $\limsup_{t\to\infty} \frac{X_t}{\sqrt{t}}$ messbar bezüglich der terminalen σ-Algebra von $\mathcal{X}$, also nach Theorem 17.1 fast sicher konstant. Angenommen, $\limsup_{t\to\infty} \frac{X_t}{\sqrt{t}} \xrightarrow{t\to\infty} \gamma$ für ein $0 < \gamma < \infty$. Dann würde insbesondere gelten, dass $\mathbf{P}(\frac{X_t}{\sqrt{t}} > 2\gamma) \xrightarrow{t\to\infty} 0$, im Widerspruch zum zentralen Grenzwertsatz.

Es gilt also nun, eine Funktion $t \mapsto h_t$ zu finden mit $\sqrt{t} \leq h_t \leq t$, so dass (17.1) gilt. Diese wird durch den *iterierten Logarithmus* wiefolgt beschrieben:

Theorem 17.10 (Iterierter Logarithmus für die Brown'sche Bewegung). *Sei* $\mathcal{X} = (X_t)_{t\geq 0}$ *eine Brown'sche Bewegung. Dann gilt*

$$\limsup_{t\to\infty} \frac{X_t}{\sqrt{2t \log\log t}} = \limsup_{t\to 0} \frac{X_t}{\sqrt{2t \log\log 1/t}} = 1 \tag{17.3}$$

fast sicher.

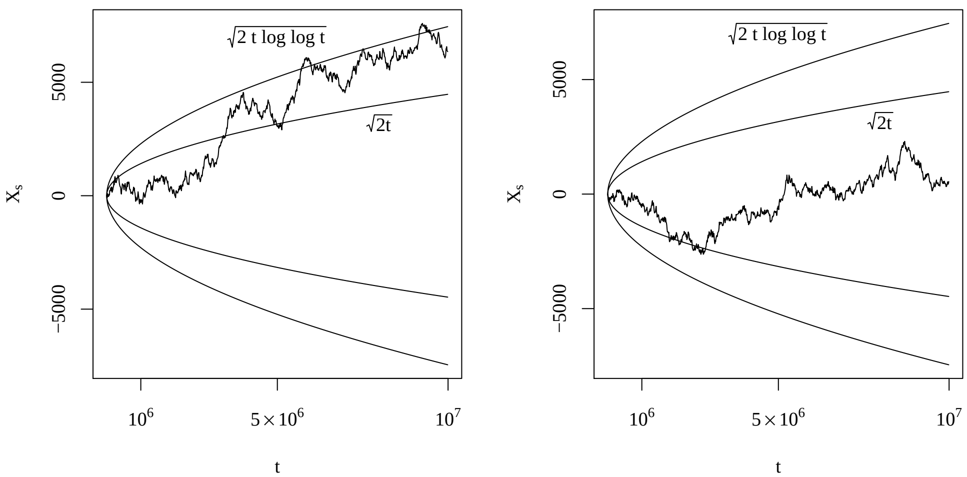


Abbildung 17.2

Hier sind zwei Pfade einer Brown'schen Bewegung angegeben. Wie man sieht, verlassen die beiden Pfade die Kurven $t \mapsto \pm\sqrt{2t}$ deutlich häufiger als die Kurve $t \mapsto \pm h_t$.

Bemerkung 17.11. Aus Symmetriegründen, d.h. weil $-\mathcal{X}$ ebenfalls eine Brow'sche Bewegung ist, gilt ebenfalls

$$\liminf_{t\to\infty} \frac{X_t}{\sqrt{2t\log\log t}} = \liminf_{t\to 0} \frac{X_t}{\sqrt{2t\log\log 1/t}} = -1$$

fast sicher. Zur Veranschaulichung siehe Figur 17.2. Die Tatsache, dass $h_t := \sqrt{2t\log\log t}$ die richtige Funktion ist, bedeutet, dass sich fast jeder Pfad der Brown'schen Bewegung nur endlich oft außerhalb der beiden Kurven $t \mapsto \pm h_t$ befindet, jedoch unendlich oft außerhalb der beiden Kurven $t \mapsto \pm(1-\varepsilon)h_t$, wobei $0 < \varepsilon < 1$ beliebig ist.

Beweis. Zunächst bemerken wir, dass mit Theorem 14.19 auch $(tX_{1/t})_{t\geq 0}$ eine Brown'sche Bewegung ist. Falls wir die Aussage also für den $t \to \infty$ Grenzwert gezeigt haben, so folgt

$$\limsup_{t\to 0} \frac{X_t}{\sqrt{2t\log\log 1/t}} = \limsup_{t\to\infty} \frac{X_{1/t}}{\sqrt{2\frac{1}{t}\log\log t}} = \limsup_{t\to\infty} \frac{tX_{1/t}}{\sqrt{2t\log\log t}} = 1$$

fast sicher. Außerdem schreiben wir $h_t := h(t) := \sqrt{2t\log\log t}$. Den Beweis für $t \to \infty$ erfordert ein paar Abschätzungen. Wir unterteilen den Beweis in drei Schritte.

Schritt 1: Abschätzung der Normalverteilung: Sei $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ die Dichte von X_1. Dann ist

$$\mathbf{P}(X_1 > x) \leq \tfrac{1}{x}\varphi(x), \tag{17.4}$$
$$\mathbf{P}(X_1 > x) \geq \tfrac{x}{1+x^2}\varphi(x), \tag{17.5}$$

Denn: Es gilt $\varphi'(y) = -y\varphi(y)$ und damit

$$\varphi(x) = \int_x^\infty y\varphi(y)dy \geq x\int_x^\infty \varphi(y)dy = x\cdot\mathbf{P}(X > x),$$

was (17.4) zeigt. Für (17.5) schreiben wir, ganz ähnlich, $\left(\frac{\varphi(y)}{y}\right)' = -\frac{1+y^2}{y^2}\varphi(y)$, und damit

$$\frac{\varphi(x)}{x} = \int_x^\infty \frac{1+y^2}{y^2}\varphi(y)dy \leq \frac{1+x^2}{x^2}\int_x^\infty \varphi(y)dy = \frac{1+x^2}{x^2}\cdot \mathbf{P}(X>x).$$

Wir schreiben im folgenden $a(x) \overset{x\to\infty}{\approx} b(x)$, falls $\frac{a(x)}{b(x)} \xrightarrow{x\to\infty} 1$ gilt. Also ist z.B. nach dem eben gezeigtem

$$\mathbf{P}(X_t > x\sqrt{t}) \overset{x\to\infty}{\approx} \tfrac{1}{x}\varphi(x).$$

2. Schritt: Obere Abschätzung: Nach Theorem 17.8 ist für $x>0$

$$\mathbf{P}(\sup_{0\leq s\leq t} X_s > x\sqrt{t}) = 2\cdot \mathbf{P}(X_t > x\sqrt{t}) \overset{x\to\infty}{\approx} \tfrac{2}{x}\varphi(x).$$

Sei nun $r>1$. Wir bemerken zunächst

$$h(r^{n-1}) = \sqrt{\frac{2(\log(n-1)+\log\log r)}{r}}\sqrt{r^n} \overset{n\to\infty}{\approx} \sqrt{\frac{2\log n}{r}}\sqrt{r^n}$$

Nun ist für $c>0$ mit den letzten beiden Abschätzungen

$$\begin{aligned}\mathbf{P}\big(\sup_{0\leq t\leq r^n} X_t > ch(r^{n-1})\big) &\overset{n\to\infty}{\approx} 2\cdot\mathbf{P}\Big(X_{r^n} > c\sqrt{\frac{2\log n}{r}}\sqrt{r^n}\Big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{2r}{\log n}}\varphi(c\sqrt{2\log n^{1/r}})\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{r}{\pi\log n}}\frac{1}{n^{c^2/r}}.\end{aligned}\tag{17.6}$$

Es ist also für $c>1$ und $1<r<c^2$ die rechte Seite der letzten Gleichung summierbar, also folgt mit dem Borel-Cantelli-Lemma

$$\mathbf{P}\Big(\limsup_{t\to\infty}\frac{X_t}{h_t}\geq c\Big) \leq \mathbf{P}\big(\sup_{0\leq t\leq r^n} X_t > ch_{r^{n-1}} \text{ für unendlich viele } n\big) = 0.$$

Damit folgt '$\leq$' in (17.3).

3. Schritt: Untere Abschätzung: Sei $r>1$ (typischerweise groß) und $c>0$ (typischerweise nahe bei 1). Definiere die Ereignisse

$$A_n := \{X_{r^n} - X_{r^{n-1}} > ch(r^n - r^{n-1})\}.$$

Da $X_{r^n} - X_{r^{n-1}} \sim N(0, r^n - r^{n-1})$, gilt nach Schritt 1

$$\begin{aligned}\mathbf{P}(A_n) &= \mathbf{P}\Big(\frac{X_{r^n}-X_{r^{n-1}}}{\sqrt{r^n-r^{n-1}}} > c\frac{h(r^n-r^{n-1})}{\sqrt{r^n-r^{n-1}}}\Big)\\ &= \mathbf{P}\big(X_1 > c\sqrt{2\log\log(r^n-r^{n-1})}\big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\frac{1}{\sqrt{4\pi\log\log(r^n-r^{n-1})}}\exp\big(-c^2\log\log(r^n-r^{n-1})\big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\frac{1}{\sqrt{4\pi\log n}}\frac{1}{n^{c^2}}\end{aligned}$$

Ist $c < 1$, so sind diese Wahrscheinlichkeiten nicht summierbar in n. Da die Ereignisse $A_1, A_2, ...$ unabhängig sind, gilt nach dem Borel-Cantelli Lemma, dass unendlich viele der A_n zutreffen. Also gilt für unendlich viele n, falls $c < 1$

$$X_{r^n} > ch(r^n - r^{n-1}) + X_{r^{n-1}}.$$

Nach der '≤'-Richtung ist $X_{r^{n-1}} > -2h(r^{n-1})$ für fast alle n, d.h. $\liminf_{n\to\infty} \frac{X_{r^{n-1}}}{h(r^n)} \leq -\liminf_{n\to\infty} \frac{h(r^{n-1})}{h(r^n)} = -\frac{1}{\sqrt{r}}$ fast sicher. Weiter ist $h(r^n - r^{n-1})/h(r^n) \xrightarrow{n\to\infty} 1$ und damit

$$\limsup_{t\to\infty} \frac{X_t}{h_t} \geq \limsup_{n\to\infty} \frac{X_{r^n}}{h(r^n)} \geq \limsup_{n\to\infty} \frac{X_{r^n} - X_{r^{n-1}}}{h(r^n - r^{n-1})} - \frac{1}{\sqrt{r}} \geq c - \frac{1}{\sqrt{r}}.$$

Da $0 < c < 1$ und $r > 0$ beliebig waren, folgt '≥' in (17.3). □

17.4 Satz von Donsker

Die Brown'sche Bewegung $\mathcal{X} = (X_t)_{t\geq 0}$ ist ein stochastischer Prozess mit stetigen Pfaden. Paul Lévy betrachtete die Brown'sche Bewegung approximativ als Irrfahrt, wobei

$$X_{t+dt} - X_t = \pm\sqrt{dt}, \text{ jeweils mit Wahrscheinlichkeit } \tfrac{1}{2}.$$

(Natürlich kann dies nur eine formale Darstellung sein, schließlich ist unklar was denn $\sqrt{dt}$ sein soll.) Der hier dargestellte Satz von Donsker macht die Verbindung von Irrfahrten und der Brown'schen Bewegung exakt. Er behauptet die Konvergenz von Irrfahrten gegen die Brown'sche Bewegung in Verteilung.

Bemerkung 17.12 (Irrfahrten und Brown'sche Bewegung). In diesem Abschnitt sind $Y_1, Y_2, ...$ unabhängig und identisch verteilte Zufallsgrößen mit $\mathbf{E}[Y_1] = 0$ und $\mathbf{V}[Y_1] = \sigma^2$ sowie $\widetilde{X}_{n,t} := \frac{Y_1+\cdots+Y_{\lfloor nt\rfloor}}{\sqrt{n\sigma^2}}$ für $t \geq 0$ und $\widetilde{\mathcal{X}}_n = (\widetilde{X}_{n,t})_{t\geq 0}$. Wir wissen aus dem zentralen Grenzwertsatz, dass für $t > 0$

$$\widetilde{X}_{n,t} \xRightarrow{n\to\infty} X_t,$$

wobei $X_t \sim N(0,t)$-verteilt ist. Ganz analog gilt für $0 < t_1 < \cdots < t_k < \infty$

$$\big(\widetilde{X}_{n,t_1}, \widetilde{X}_{n,t_2} - \widetilde{X}_{n,t_1}, ..., \widetilde{X}_{n,t_k} - \widetilde{X}_{n,t_{k-1}}\big) \xRightarrow{n\to\infty} (X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_k} - X_{t_{k-1}}),$$

falls $(X_{t_1}, ..., X_{t_k})$ die Werte einer Brown'schen Bewegung $\mathcal{X}$ zu den Zeitpunkten $t_1, .., t_k$ sind. Bedeutet das nun schon die Konvergenz der Irrfahrt gegen die Brown'sche Bewegung, also $\mathcal{X}_n \xRightarrow{n\to\infty} \mathcal{X}$? Nein! Für diese Konvergenz müssen wir sowohl $\mathcal{X}_n$ und $\mathcal{X}$ als Zufallsvariable mit Werten in einem topologischen Raum – nennen wir ihn $\mathcal{C}$ – auffassen können, wobei die Konvergenz in Verteilung auf der Konvergenz von Erwartungswerten bezüglich stetigen, beschränkten Funktionen $f : \mathcal{C} \to \mathbb{R}$ basiert. Allerdings ist für den überabzählbaren Produktraum die σ-Algebra $\mathcal{B}(\mathbb{R})^{\otimes[0,\infty)}$ nicht die Borel'sche σ-Algebra auf dem Produktraum, und die Theorie schwacher Konvergenz hatten wir nur für den Fall entwickelt in dem wir es mit Wahrscheinlichkeitsmaßen auf einer Borel'schen σ-Algebra zu tun hatten.

Um die Konvergenz in Verteilung gegen die Brown'sche Bewegung formulieren zu können, benötigen wir also zunächst einen geeigneten Zustandsraum. Dieser ist als $\mathcal{C} := \mathcal{C}_{\mathbb{R}}([0,\infty))$, versehen mit der Topologie der kompakten Konvergenz gegeben (siehe Definition 17.13). Um

Konvergenz in diesem Raum zu formulieren, definieren wir die lineare Interpolation der Prozesse $\widetilde{\mathcal{X}}_n$, damit deren Pfade ebenfalls stetig sind. Hierzu setzen wir

$$X_{n,t} := \widetilde{X}_{n,t} + (nt - \lfloor nt \rfloor)\frac{Y_{\lfloor nt \rfloor + 1}}{\sqrt{n\sigma^2}}. \tag{17.7}$$

und $\mathcal{X}_n = (X_{n,t})_{t \geq 0}$. Nun macht es Sinn zu fragen ob

$$\mathcal{X}_n \xRightarrow{n \to \infty} \mathcal{X}$$

gilt, wobei hier die schwache Konvergenz bezüglich der Verteilungen auf $\mathcal{B}(\mathcal{C}_{\mathbb{R}}([0,\infty)))$ gemeint ist.

Definition 17.13 (Uniforme Konvergenz auf Kompakta). *Sei (E, r) ein metrischer Raum. Für $f, f_1, f_2, ... \in \mathcal{C}_E([0,\infty))$ sei $f_n \xrightarrow{n \to \infty} f$* uniform auf Kompakta *genau dann, wenn $\sup_{0 \leq s \leq t} r(f_n(s), f(s)) \xrightarrow{n \to \infty} 0$ für alle $t > 0$.*

Lemma 17.14 ($\mathcal{C}_E([0,\infty))$ ist Polnisch). *Sei E Polnisch mit vollständiger Metrik r. Dann ist die Topologie der uniformen Konvergenz auf Kompakta auf $\mathcal{C}_E([0,\infty))$ separabel. Außerdem, definiert*

$$r_{\mathcal{C}}(f, g) := \int_0^\infty e^{-t} \cdot (1 \wedge \sup_{0 \leq s \leq t} |r(f(s), g(s))|) dt$$

eine vollständige Metrik auf $\mathcal{C}_E([0,\infty))$, die diese Topologie induziert. Insbesondere ist $\mathcal{C}_E([0,\infty))$ Polnisch.

Beweis. Um die Separabilität zu zeigen, genügt es eine abzählbare Klasse von Funktionen zu nennen, die jede Funktion in $\mathcal{C}_E([0,\infty))$ lokal auf Kompakta approximiert. Hierzu sei $D \subseteq E$ dicht und abzählbar. Für jede endliche Folge $x_1, ..., x_n \in D$ und $t_1, ..., t_n$ sei $f = f_{x_1,...,x_n,t_1,...,t_n}$ eine stetige Fuktion mit $f(t_i) = x_i$. Dann ist $\bigcup_n \{f_{x_1,...,x_n,t_1,...,t_n} : x_1, ..., x_n \in D, t_1, ..., t_n \geq 0\}$ abzählbar und dicht in $\mathcal{C}_E([\infty))$.

Nun zur Metrik. Da $t \mapsto \sup_{0 \leq s \leq t} r(f(s), g(s)) \wedge 1$ monoton wachsend ist, gilt $r_{\mathcal{C}}(f_n, f) \xrightarrow{t \to \infty} 0$ genau dann, wenn $\sup_{0 \leq s \leq t} r(f_n(s), f(s)) \xrightarrow{n \to \infty} 0$ für alle t gilt. Dies ist aber genau die kompkte Konvergenz. Sei weiter $f_1, f_2, ...$ eine Cauchy-Folge bezüglich $r_{\mathcal{C}}$. Dann ist für jedes $t > 0$ die Folge $f_1, f_2, ...$, eingeschränkt auf $[0, t]$ eine Cauchy-Folge bezüglich der Supremumsnorm auf $[0, t]$, also uniform konvergent auf $[0, t]$. Die Behauptung folgt nun mittels eines Diagonal-Folgen-Argumentes. □

Zunächst definieren wir zwei Konvergenzarten stochastischer Prozesse, die wir eben kennen gelernt haben.

Definition 17.15 (Konvergenz von stochastischen Prozessen). *Seien $\mathcal{X} = (X_t)_{t \geq 0}, \mathcal{X}^1 = (X_t^1)_{t \geq 0}, \mathcal{X}^2 = (X_t^2)_{t \geq 0}, ...$ stochastische Prozesse mit Zustandsraum E.*

1. *Gilt für jede Wahl von $t_1, ..., t_k, k = 1, 2, ...,$ dass*

$$(X_{t_1}^n, ..., X_{t_k}^n) \xRightarrow{n \to \infty} (X_{t_1}, ..., X_{t_k}),$$

so sagen wir, dass die endlich-dimensionalen Verteilungen von $\mathcal{X}^1, \mathcal{X}^2, ...$ gegen die von $\mathcal{X}$ konvergieren und schreiben

$$\mathcal{X}^n \xRightarrow{n \to \infty}_{fdd} \mathcal{X}.$$

(Hier steht fdd *für* finite dimensional distributions*.)*

2. *Haben die Prozesse $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ Pfade in $\mathcal{C}_E([0,\infty))$ und gilt*

$$\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X},$$

wobei wir hier $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ als Zufallsvariable in $\mathcal{C}_E([0,\infty))$ ansehen, so sagen wir, dass $\mathcal{X}^1, \mathcal{X}^2, ...$ in Verteilung gegen $\mathcal{X}$ konvergiert.

Die fdd-Konvergenz ist schwächer als die schwache Konvergenz von Prozessen. Kann man jedoch zusätzlich die Straffheit (siehe Definition 10.14) der Prozesse zeigen, fallen beide Begriffe zusammen.

Proposition 17.16 (Schwache und fdd-Konvergenz). *Seien $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ Zufallsvariable mit Werten in $\mathcal{C}_E([0,\infty))$. Dann sind äquivalent*

1. $\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X}$
2. $\mathcal{X}^n \xRightarrow{n\to\infty}_{fdd} \mathcal{X}$ *und* $\{\mathcal{X}^n : n = 1,2,...\}$ *ist straff in* $\mathcal{C}_E([0,\infty))$.

Beweis. '1.⇒2.': Zunächst folgt aus der schwachen Konvergenz nach Korollar 10.18 die Straffheit von $\{\mathcal{X}^n : n = 1,2,...\}$. Weiter sind die Abbildungen $f \mapsto (f(t_1),...,f(t_k))$ stetig für $t_1,...,t_k \in [0,\infty)$, also folgt die fdd-Konvergenz nach Theorem 10.10.

'2.⇒1.': Wir definieren die Funktionenklasse

$$\mathcal{M} := \{f \mapsto \varphi(f(t_1),...,f(t_k)) : t_1,...,t_k \in [0,\infty), \varphi \in \mathcal{C}_b(E^k)\} \subseteq \mathcal{C}_b(\mathcal{C}_E([0,\infty))).$$

Klar ist, dass die fdd-Konvergenz $\mathcal{X}^n \xRightarrow{n\to\infty}_{fdd} \mathcal{X}$ äquivalent ist zu $\mathbf{E}[\varphi(\mathcal{X}^n)] \xrightarrow{n\to\infty} \mathbf{E}[\varphi(\mathcal{X})]$ für alle $\varphi \in \mathcal{M}$. Weiter ist $\mathcal{M}$ eine Algebra und trennt Punkte, ist nach Theorem 10.24 also separierend. Nun folgt die schwache Konvergenz aus Proposition 10.27. □

Um die Konvergenz von Prozessen zu zeigen ist nach Proposition 17.16 sowohl die Konvergenz der endlich-dimensionalen Verteilungen als auch die Straffheit zu zeigen. In Anwendungen ist meistens die Überprüfung der Straffheit nicht-trivial. Insbesondere man verstehen muss, wie (relativ-)kompakte Teilmengen von $\mathcal{C}_E([0,\infty))$ charakterisiert werden können. Dies geschieht durch den aus der Analysis bekannten Satz von Arzela-Ascoli, der auf dem Stetigkeitsmodul aufbaut.

Definition 17.17 (Stetigkeitsmodul). *Für $f \in \mathcal{C}_E([0,\infty))$ definieren wir das* Stetigkeitsmodul

$$w(f,\tau,h) := \sup\{r(f(s),f(t)) : s,t \le \tau, |t-s| \le h\}.$$

Theorem 17.18 (Arzela-Ascoli). *Eine Menge $A \subseteq \mathcal{C}_E([0,\infty))$ ist genau dann relativkompakt, wenn $\{f(t) : f \in A\}$ für alle $t \in \mathbb{Q}_+ := [0,\infty) \cap \mathbb{Q}$ relativkompakt in E ist und für alle $\tau > 0$*

$$\lim_{h\to 0} \sup_{f\in A} w(f,\tau,h) = 0 \tag{17.8}$$

gilt.

Beweis. Sei zunächst A relativ-kompakt. Dann müssen $\{f(t) : t \in A\}$ für alle $t \geq 0$ relativ-kompakt sein, sonst wäre es einfach eine divergente Folge zu konstruieren. Außerdem ist A nach Proposition 1.9 total beschränkt. Weiter sei $\tau > 0$, $\varepsilon > 0$ und $f_1, ..., f_N$, so dass $A \subseteq \bigcup_{i=1}^N B_{\varepsilon/3}(f_i)$. Da $f_1, ..., f_N$ auf $[0, \tau]$ gleichmäßig stetig sind, gibt es ein $h > 0$ mit

$$0 \leq s, t \leq \tau, |t - s| < h \qquad \Longrightarrow \qquad r(f_i(t), f_i(s)) \leq \varepsilon/3, \qquad i = 1, ..., N.$$

Also gilt für jedes $f \in A$ und $s, t \leq \tau, |t - s| \leq h$, dass

$$r(f(s), f(t)) \leq \min_{i=1,...,N} r(f(s), f_i(s)) + r(f_i(s), f_i(t)) + r(f_i(t), f(t)) \leq \varepsilon$$

und damit

$$w(f, \tau, h) = \sup\{r(f(t), f(s)) : s, t \leq \tau, |t - s| \leq h\} \leq \varepsilon,$$

unabhängig von f. Daraus folgt also (17.8).

Andersherum gelte (17.8). Es genügt zu zeigen, dass jede Folge in A eine Teilfolge hat, die Cauchy ist. Wegen der Relativ-Kompaktheit von $\{f(t) : f \in A\}$ für $t \in \mathbb{Q}_+$ ist klar, dass es für jede Folge eine Teilfolge $f_1, f_2, ...$ gibt, so dass $f_1(t_i), f_2(t_i), ...$ für alle $t_i \in \mathbb{Q}_+$ eine Cauchy-Folge (also konvergent) ist. Sei nun $\varepsilon > 0$. Nach Voraussetzung gibt es ein $h > 0$, so dass aus $|t - s| \leq h$ und $f \in A$ folgt, dass $r(f(s), f(t)) \leq \varepsilon/3$ gilt. Weiter sei $M = \lceil \tau/h \rceil$ und $0 = t_1, ..., t_M \in \mathbb{Q}_+$, so dass $|t_{i+1} - t_i| \leq h$, $i = 1, ..., M - 1$ und $t_M \geq \tau$. Weiter gibt es ein N, so dass aus $n, m > N$ folgt, dass $\sup_{t=t_1,...,t_M} r(f_n(t), f_m(t)) \leq \varepsilon/3$ gilt. Daraus folgt nun für $0 \leq s \leq t$

$$r(f_n(s), f_m(s)) \leq r(f_n(s), f_n(t_{\lceil s/h \rceil})) + r(f_n(t_{\lceil s/h \rceil}), f_m(t_{\lceil s/h \rceil})) + r(f_m(t_{\lceil s/h \rceil}), f_m(s)) \leq \varepsilon.$$

Daraus folgt, dass $f_1, f_2, ...$ eine Cauchy-Folge bezüglich kompakter Konvergenz auf $[0, t]$ ist, also auf diesem Bereich gleichmäßig konvergiert. Ein Diagonal-Folgen-Argument erweitert diese Aussage auf die kompakte Konvergenz. □

Theorem 17.19 (Straffheit in $\mathcal{C}_{\mathbb{E}}([0, \infty))$). *Seien $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ Zufallsvariablen mit Werten in $\mathcal{C}_E([0, \infty))$. Dann gilt $\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X}$ genau dann, wenn $\mathcal{X}^n \xRightarrow{n\to\infty}_{fdd} \mathcal{X}$ und*

$$\lim_{h\to 0} \limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] = 0 \tag{17.9}$$

für alle $\tau > 0$.

Beweis. Nach Proposition 17.16 genügt es zu zeigen, dass (17.9) äquivalent zur Straffheit der Familie $(\mathcal{X}^n)_{n=1,2,...}$ ist.

Sei zunächst $(\mathcal{X}^n)_{n=1,2,...}$ straff und $\varepsilon > 0$. Dann gibt es eine kompakte Menge $K \subseteq \mathcal{C}_E([0, \infty))$ so dass $\limsup_{n\to\infty} \mathbf{P}(\mathcal{X}^n \notin K) \leq \varepsilon$. Für $\tau > 0$ kann man nach dem Arzela-Ascoli Theorem h klein genug wählen, so dass $w(f, \tau, h) \leq \varepsilon$ für $f \in K$ gilt. Damit ist

$$\limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] \leq \varepsilon + \sup_{n=1,2,...} \mathbf{P}[w(\mathcal{X}^n, \tau, h) > \varepsilon] \leq 2\varepsilon,$$

woraus (17.9) folgt.

Andersherum gelte (17.9) sowie $\mathcal{X}^n \xRightarrow{n\to\infty}_{fdd} \mathcal{X}$. Die Abbildung w ist wachsend in h und $w(\mathcal{X}^n, \tau, h) \xrightarrow{h\to 0} 0$ fast sicher für $n = 1, 2, ...$ Also gilt $\lim_{h\to 0} \sup_{n=1,2,...} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] =$

$\lim_{h\to 0} \sup_{n=k,k+1,\dots} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1]$ für alle k, also auch $\lim_{h\to 0} \sup_{n=1,2,\dots} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] = \lim_{h\to 0} \limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1]$. Also ist (17.9) äquivalent zu

$$\lim_{h\to 0} \sup_{n=1,2,\dots} \mathbf{P}[w(\mathcal{X}^n, \tau, h) > \varepsilon] = 0$$

für alle $\varepsilon > 0$ und $\tau > 0$. Sei $\tau_k = k$ und $\varepsilon > 0$. Dann gibt es $h_1, h_2, \dots > 0$, so dass

$$\sup_{n=1,2,\dots} \mathbf{P}(w(\mathcal{X}^n, \tau_k, h_k) > 2^{-k}) \le 2^{-(k+1)}\varepsilon.$$

Weiter sei $t_1, t_2, \dots$ eine Abzählung von $\mathbb{Q}_+$ und $C_1, C_2, \dots \subseteq \mathbb{R}$ kompakt, so dass

$$\sup_{n=1,2,\dots} \mathbf{P}(X^n(t_k) \notin C_k) \le 2^{-(k+1)}\varepsilon.$$

Nun definieren wir

$$B := \bigcap_{k=1}^{\infty} \big\{ f \in \mathcal{C}_E([0,\infty)) : f(t_k) \in C_k, w(f, \tau_k, h_k) \le 2^{-k} \big\}.$$

Nach dem Satz von Arzela-Ascoli ist $B \subseteq \mathcal{C}_E([0,\infty))$ relativ-kompakt. Außerdem gilt

$$\begin{aligned} \sup_{n=1,2,\dots} \mathbf{P}(\mathcal{X}^n \notin B) &\le \sup_{n=1,2,\dots} \sum_{k=1}^{\infty} \mathbf{P}(X^n(t_k) \notin C_k) + \mathbf{P}(w(\mathcal{X}^n, \tau_k, h_k) > 2^{-k}) \\ &\le \sum_{k=1}^{\infty} 2^{-(k+1)}\varepsilon + 2^{-(k+1)}\varepsilon = \varepsilon. \end{aligned}$$

Daraus folgt, dass $(\mathcal{X}^n)_{n=1,2,\dots}$ straff ist. □

Wir wollen das letzte Resultat anwenden, um die Konvergenz der Irrfahrt gegen die Brown'sche Bewegung zu zeigen. Hierzu benötigen wir noch eine Lemma.

Lemma 17.20. *Seien $Y_1, Y_2, \dots$ unabhängige und identisch verteilte Zufallsgrößen mit $\mathbf{E}[Y_1] = 0$ und $\mathbf{V}[Y_1] = \sigma^2 > 0$ sowie $S_n := Y_1 + \cdots + Y_n$. Dann gilt für $r > 1$*

$$\mathbf{P}(\max_{1\le k\le n} S_k > 2r\sqrt{n}) \le \frac{\mathbf{P}(|S_n| > r\sqrt{n})}{1 - \sigma^2 r^{-2}}.$$

Beweis. Wir definieren $T := \inf\{k : |S_k| > 2r\sqrt{n}\}$. Dann gilt, da $(S_n)_{n=1,2,\dots}$ stark Markov ist,

$$\begin{aligned} \mathbf{P}(|S_n| > r\sqrt{n}) &\ge \mathbf{P}(|S_n| > r\sqrt{n}, \max_{1\le k\le n} S_k > 2r\sqrt{n}) \\ &\ge \mathbf{P}(T \le n, |S_n - S_T| \le r\sqrt{n}) \\ &\ge \mathbf{P}(\max_{1\le k\le n} S_k > 2r\sqrt{n}) \cdot \min_{1\le k\le n} \mathbf{P}(|S_k| \le r\sqrt{n}) \end{aligned}$$

Aus Chebychev's Ungleichung folgt damit

$$\min_{1\le k\le n} \mathbf{P}(|S_k| \le r\sqrt{n}) \ge \min_{1\le k\le n} 1 - \frac{\sigma^2 k}{r^2 n} = 1 - \frac{\sigma^2}{r^2}.$$

□

Theorem 17.21 (Satz von Donsker). *Seien $Y_1, Y_2, \ldots$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathbf{E}[Y_1] = 0$ und $\mathbf{V}[Y_1] = \sigma^2 > 0$, und $\mathcal{X}_n = (X_{n,t})_{t \geq 0}$ gegeben durch*

$$X_{n,t} := \frac{1}{\sqrt{n\sigma^2}}\big(Y_1 + \cdots + Y_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)Y_{\lfloor nt \rfloor + 1}\big)$$

und $\mathcal{X} = (X_t)_{t \geq 0}$ eine Brown'sche Bewegung. Dann gilt

$$\mathcal{X}_n \xRightarrow{n \to \infty} \mathcal{X}.$$

Beweis. Sei oBdA $\sigma^2 = 1$. Wie in Bemerkung 17.12 ausgeführt, gilt $\mathcal{X}_n \xRightarrow{n\to\infty}_{fdd} \mathcal{X}$. Also ist nach Proposition 17.16 noch die Straffheit der Familie $\{\mathcal{X}_n : n \in \mathbb{N}\}$ nachzuweisen, also (17.9) aus Theorem 17.19. Wir schreiben im Folgenden $S_n := Y_1 + \cdots + Y_n$. Mit Lemma 17.20 folgt

$$\begin{aligned}
\lim_{h \to 0} \frac{1}{h} \limsup_{n \to \infty} \mathbf{P}\big(\sup_{0 \leq s \leq h} |X_{n,t+s} - X_{n,t}| > \varepsilon\big) \\
&\leq \lim_{h \to 0} \frac{1}{h} \limsup_{n \to \infty} \mathbf{P}\Big(\sup_{k=1,\ldots,\lceil nh \rceil} |S_k| > \frac{\varepsilon}{\sqrt{h}}\sqrt{nh}\Big) \\
&\leq \lim_{h \to 0} \frac{1}{h} \limsup_{n \to \infty} \mathbf{P}\Big(\frac{|S_{\lceil nh \rceil}|}{\sqrt{nh}} > \frac{\varepsilon}{2\sqrt{h}}\Big) \\
&\leq \lim_{h \to 0} \frac{2}{h} \int_{\varepsilon/(2\sqrt{h})}^{\infty} \varphi(x)dx \\
&= \lim_{h \to 0} \frac{2}{h} \frac{2\sqrt{h}}{\varepsilon} \varphi(\varepsilon/(2\sqrt{h})) = 0
\end{aligned}$$

nach (17.5), wobei φ die Dichte der $N(0,1)$-Verteilung ist. Sei nun $\delta > 0$ und h klein genug für

$$\limsup_{n \to \infty} \mathbf{P}\big(\sup_{0 \leq s \leq h} |X_{n,t+s} - X_{n,t}| > \varepsilon\big) \leq \delta h.$$

Damit können wir schreiben

$$\begin{aligned}
\limsup_{n \to \infty} \mathbf{P}(w(\mathcal{X}_n, \tau, h) > 2\varepsilon) &= \limsup_{n \to \infty} \mathbf{P}(\sup_{0 \leq t \leq \tau - h, 0 \leq s \leq h} |X_{n,t+s} - X_{n,t}| > 2\varepsilon) \\
&\leq \limsup_{n \to \infty} \mathbf{P}(\sup\{|X_{n,kh+s} - X_{n,kh}| : k = 0, 1, \ldots, [\tau/h], 0 \leq s \leq h\} > \varepsilon) \\
&\leq \sum_{k=0}^{[\tau/h]} \limsup_{n \to \infty} \mathbf{P}(\sup\{|X_{n,kh+s} - X_{n,kh}| : 0 \leq s \leq h\} > \varepsilon) \\
&\leq [\tau/h]\delta h \xrightarrow{h \to 0} \tau\delta.
\end{aligned}$$

Da $\delta > 0$ beliebig war, folgt (17.9). □

Wir beenden diesen Abschnitt mit einem Straffheits-Kriterium, das oft anwendbar ist. Es baut auf Theorem 14.8 auf.

Theorem 17.22 (Kolmogorov-Chentsov Kriterium für Straffheit). *Seien $\mathcal{X}_1 = (X_1(t))_{t \geq 0}, \mathcal{X}_2 = (X_2(t))_{t \geq 0}, \ldots$ stochastische Prozesse mit stetigen Pfaden. Angenommen $\{X_n(0) : n \in \mathbb{N}\}$ ist straff und für jedes $\tau > 0$ gibt es Zahlen $\alpha, \beta, C > 0$ mit*

$$\sup_n \mathbf{E}[r(X_n(s), X_n(t))^\alpha] \leq C|t - s|^{1+\beta}$$

für alle $0 \leq s, t \leq \tau$. Dann ist $\{\mathcal{X}_n : n \in \mathbb{N}\}$ straff in $\mathcal{C}_E([\infty))$.

Beweis. Sei $0 < \gamma < \beta/\alpha$ beliebig. Wir benutzen die Notation aus dem Beweis von Theorem 14.8, also z.B. $\xi_{nk} := \max\{r(X_n(s), X_n(t)) : s,t \in D_k, |t-s| = 2^{-k}\}$. OBdA sei $\tau = 1$. Genau wie in (14.1) berechnen wir

$$\sum_{k=0}^{\infty} 2^{\alpha\gamma k} \mathbf{E}[\xi_{nk}^{\alpha}] \leq C \sum_{k=0}^{\infty} 2^{(\alpha\gamma-\beta)k}.$$

Da die rechte Seite nicht von n abhängt, gibt es ein C' mit $\sup_n \mathbf{E}[\xi_{nk}^{\alpha}] \leq C' e^{-\alpha\gamma k}$. Wichtig ist es einzusehen, dass $w(\mathcal{X}_n, 1, 2^{-m}) \leq \sum_{k=m}^{\infty} \xi_{nk}$. Daraus folgt

$$\begin{aligned}\sup_n \mathbf{E}[w(\mathcal{X}_n, 1, 2^{-m})^{\alpha} \wedge 1] &\leq \sup_n \mathbf{E}\Big[\Big(\sum_{k=m}^{\infty} \xi_{nk}\Big)^{\alpha}\Big] \leq \sup_n \Big(\sum_{k=m}^{\infty} \mathbf{E}[\xi_{nk}^{\alpha}]^{1/\alpha}\Big)^{\alpha} \\ &\leq C' \Big(\sum_{k=m}^{\infty} e^{-\gamma k}\Big)^{1/\alpha} \xrightarrow{m\to\infty} 0,\end{aligned}$$

woraus die Behauptung folgt. □

17.5 Der Skorohod'sche Einbettungssatz

Der Name Skorohod fiel bereits beim Zusammenhang zwischen schwacher und fast sicherer Konvergenz, siehe Theorem 10.11. Salopp gesprochen konvergiert eine Folge von Zufallsvariablen genau dann schwach, wenn sie auf einem geeignetem Wahrscheinlichkeitsraum fast sicher konvergiert. Betrachtet man nochmals den Satz von Donsker, kann man sich fragen, wie denn der Wahrscheinlichkeitsraum auszusehen hat, auf dem die Irrfahrt fast sicher gegen eine Brown'sche Bewegung konvergiert. Anders gefragt: wie muss man die Irrfahrt und die Brown'sche Bewegung definieren, so dass beide *immer* nah beisammen sind. Dies beantwortet der Skorohod'sche Einbettungs-Satz, Theorem 17.26. Er lässt weitere Rückschlüsse auf die Irrfahrt zu, etwa das Gesetz des iterierten Logarithmus, Theorem 17.29. Grundlegend ist das folgende Lemma:

Lemma 17.23 (Randomisierung). *Für $w < 0 < z$ sei $Y_{w,z}$ eine Zufallsvariable mit Zustandsraum $\{w, z\}$ mit*

$$\mathbf{P}(Y_{w,z} = w) = \frac{z}{z + |w|}$$

und $Y_{w,z} = 0$ für $w, z = 0$. Weiter sei Y eine reellwertige Zufallsvariable mit $\mathbf{E}[Y] = 0$. Dann gibt es ein Paar von Zufallsvariablen (W, Z) mit $W \leq 0, Z \geq 0$, so dass Y die Verteilung $Y_{W,Z}$ besitzt.

Beweis. Wir setzen $c = \mathbf{E}[Y^+] = \mathbf{E}[Y^-]$. Weiter sei $f : \mathbb{R} \to \mathbb{R}_+$ messbar mit $f(0) = 0$. Dann gilt, falls $Y \sim \mu$,

$$\begin{aligned}c \cdot \mathbf{E}[f(Y)] &= \mathbf{E}[Y^+] \cdot \mathbf{E}[f(-Y^-)] + \mathbf{E}[Y^-] \cdot \mathbf{E}[f(Y^+)] \\ &= \int\int (zf(w) + |w|f(z)) 1_{z\geq 0} 1_{w\leq 0} \mu(dw)\mu(dz) \\ &= \int\int (z + |w|) \mathbf{E}[f(Y_{w,z})] 1_{z\geq 0} 1_{w\leq 0} \mu(dw)\mu(dz).\end{aligned}$$

Das bedeutet, dass wir (W, Z) als Zufallsvariable mit gemeinsamer Verteilung

$$\mu_{W,Z}(dw, dz) = \mu(0)\delta_{0,0}(dw, dz) + \tfrac{1}{c}(z + |w|)1_{w\leq 0}1_{z\geq 0}\mu(dw)\mu(dz)$$

wählen können. (Man prüft leicht nach, dass die Gesamtmasse dieses Maßes 1 ist.) Dann gilt nämlich

$$c\mathbf{E}[f(Y_{W,Z})] = c\mathbf{E}[\mathbf{E}[f(W_{W,Z})|(W,Z)]] = \int\int (z+|w|)\mathbf{E}[f(Y_{w,z})]1_{w\leq 0}1_{z\geq 0}\mu(dw)\mu(dz)$$

und die Behauptung ist gezeigt, da f beliebig war. $\square$

Bemerkung 17.24 (Starke Einbettung). Das Lemma behauptet zunächst nur die Gleichheit in Verteilung, $Y \sim Y_{W,Z}$. Weiter ist es auch möglich, den Wahrscheinlichkeitsraum, auf dem Y definiert ist, zu erweitern um Zufallsvariablen (W, Z) und $Y_{W,Z}$, so dass $Y = Y_{W,Z}$ fast sicher gilt.

Lemma 17.25 (Einbettung einer Zufallsvariable in eine Brown'sche Bewegung). *Sei Y eine reellwertige Zufallsvariable mit $\mathbf{E}[Y] = 0$. Weiter sei (W, Z) so verteilt wie in Lemma 17.23, und $\mathcal{X} = (X_t)_{t\geq 0}$ sei eine unabhängige Brown'sche Bewegung. Dann ist*

$$T_{W,Z} := \inf\{t \geq 0 : X_t \in \{W, Z\}\}$$

eine Stoppzeit bezüglich der Filtration $(\mathcal{F}_t)_{t\geq 0}$ mit $\mathcal{F}_t = \sigma(W, Z; X_s : s \leq t)$. Außerdem gilt

$$X_{T_{W,Z}} \sim Y, \qquad \mathbf{E}[T_{W,Z}] = \mathbf{E}[Y^2].$$

Beweis. Die Brown'sche Bewegung $\mathcal{X}$ ist an $(\mathcal{F}_t)_{t\geq 0}$ adaptiert. Deswegen ist $T_{W,Z}$ nach Proposition 14.30 eine Stoppzeit. Klar nimmt für $w < 0 \leq z$ die Zufallsvariable $X_{T_{w,z}}$ nur die Werte w und z an. Nach Proposition 15.18 ist $(X_{T_{w,z}\wedge t})_{t\geq 0}$ ein Martingal, das nach Theorem 15.21 in L^1 gegen $X_{T_{w,z}}$ konvergiert. Deshalb gilt

$$0 = \mathbf{E}[X_{T_{w,z}}] = w\mathbf{P}(X_{T_{w,z}} = w) + z(1 - \mathbf{P}(X_{T_{w,z}} = w)),$$

also

$$\mathbf{P}(X_{T_{w,z}} = w) = \frac{z}{z + |w|}.$$

Also hat $X_{T_{w,z}}$ dieselbe Verteilung wie $Y_{w,z}$ aus Lemma 17.23 und ist unabhängig von X. Nach dem Lemma folgt also $X_{T_{W,Z}} \sim Y_{W,Z} \sim Y$. Weiter ist $(X_t^2 - t)_{t\geq 0}$ ein Martingal und für $y < 0 \leq z$ ist $(X^2_{T_{w,z}\wedge t} - T_{w,z}\wedge t)_{t\geq 0}$ ein Martingal. Damit gilt mit monotoner und majorisierter Konvergenz

$$\begin{aligned}\mathbf{E}[T_{W,Z}] &= \mathbf{E}[\mathbf{E}[T_{W,Z}|W,Z]] = \mathbf{E}[\lim_{t\to\infty}\mathbf{E}[T_{W,Z}\wedge t]|W,Z] \\ &= \mathbf{E}[\mathbf{E}[X^2_{T_{W,Z}}|W,Z]] = \mathbf{E}[X^2_{T_{W,Z}}] = \mathbf{E}[Y^2].\end{aligned}$$

$\square$

Theorem 17.26 (Skorohod'scher Einbettungssatz). *Seien $Y_1, Y_2, \ldots$ unabhängig und identisch verteilt mit $\mathbf{E}[Y_1] = 0$, sowie $S_n = Y_1 + \cdots + Y_n$. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ mit Filtration $(\mathcal{F}_t)_{t\geq 0}$, sowie eine Brown'sche Bewegung $\mathcal{X} = (X_t)_{t\geq 0}$ auf diesem Wahrscheinlichkeitsraum, die ein $(\mathcal{F}_t)_{t\geq 0}$-Martingal ist und Stoppzeiten $T_1, T_2, \ldots$, so dass folgendes gilt:*

1. $(X_{T_1}, X_{T_2}, ...) \sim S_1, S_2, ...$ *und*

2. $(T_{n+1} - T_n)_{n=0,1,2,...}$ *sind unabhängig mit* $\mathbf{E}[T_{n+1} - T_n] = \mathbf{V}[Y_1]$ *für* $n = 1, 2, ...$

Bemerkung 17.27 (Starke Einbettung). 1. Wie in Bemerkung 17.24 ist es möglich, den Wahrscheinlichkeitsraum, auf dem $Y_1, Y_2, ...$ definiert sind, zu erweitern so dass $(X_{T_1}, X_{T_2}, ...) = S_1, S_2, ...$ fast sicher gilt.

2. Ohne die Einschränkung der Integrierbarkeit von $T_{n+1} - T_n$ wäre die Aussage des Theorems trivial. Dann könnte man nämlich einfach rekursiv $0 = T_0 \leq T_1, ...$ mittels

$$T_n = \inf\{t \geq T_{n-1} : X_t = S_n\}$$

setzen. Jedoch sind diese Wartezeiten nicht integrierbar.

Beweis von Theorem 17.26. Seien die Paare $(W_1, Z_1), (W_2, Z_2), ...$ genauso verteilt wie in Lemma 17.23. Wir erweitern den Wahrscheinlichkeitsraum um eine unabhängige Brown'sche Bewegung $\mathcal{X} = (X_t)_{t\geq 0}$. Wir definieren rekursiv $0 = T_0 \leq T_1 \leq T_2 ...$ durch

$$T_n := \inf\{t \geq T_{n-1} : X_t - X_{T_{n-1}} \in \{W_n, Z_n\}\}.$$

Damit sind $T_1, T_2, ...$ Stoppzeiten bezüglich der Filtration $(\mathcal{F}_t)_{t\geq 0}$ mit $\mathcal{F}_t = \sigma(W_1, Z_1, W_2, Z_2, ...; X_s : s \leq t)$ und $\mathcal{X}$ ist ein Martingal bezüglich $(\mathcal{F}_t)_{t\geq 0}$. Außerdem sind die Paare $(T_{n+1} - T_n, X_{T_{n+1}} - X_{T_n})_{n=0,1,2,...}$ wegen der starken Markov-Eigenschaft der Brown'schen Bewegung voneinander unabhängig. Deswegen folgt aus Lemma 17.25, dass

$$(X_{T_1}, X_{T_2} - X_{T_1}, ...) \sim (Y_1, Y_2, ...),$$

also

$$(X_{T_1}, X_{T_2}, ...) \sim (S_1, S_2, ...),$$

sowie $\mathbf{E}[T_{n+1} - T_n] = \mathbf{E}[Y_n]$. □

Da dank des letzten Theorems der Zusammenhang zwischen der Irrfahrt und der Brown'schen Bewegung gezeigt ist, liegt es auf der Hand, nochmal eine Erweiterung des Satzes von Donsker, Theorem 17.21, zu formulieren.

Korollar 17.28 (Stochastische Konvergenz der Irrfahrt). *Seien* $Y_1, Y_2, ...$ *reellwertige, unabhängige, identisch verteilte Zufallsvariable mit* $\mathbf{E}[Y_1] = 0$, $\mathbf{V}[Y_1] = 1$ *und* $S_n = Y_1 + \cdots + Y_n$. *Dann kann man den Wahrscheinlichkeitsraum erweitern, so dass es eine Brown'sche Bewegung* $\mathcal{X} = (X_t)_{t\geq 0}$ *gibt mit*

$$\sup_{0\leq s\leq t} \left| \frac{1}{\sqrt{n}} S_{[sn]} - \frac{1}{\sqrt{n}} X_{sn} \right| \xrightarrow{n\to\infty}_p 0 \tag{17.10}$$

für alle $t > 0$.

Beweis. Wir verwenden die Konstruktion aus Theorem 17.26 und Bemerkung 17.27. Da $T_{n+1} - T_n$ unabhängig und identisch verteilt sind mit $\mathbf{E}[T_{n+1} - T_n] = 1$, gilt $T_n/n \xrightarrow{n\to\infty} 1$ nach dem Gesetz der großen Zahlen. Damit gilt auch $\frac{1}{n} \sup_{s\leq t} |T_{[sn]} - sn| \xrightarrow{n\to\infty}_{fs} 0$. (Um dies einzusehen, betrachten wir die Menge $\{\frac{1}{n} \sup_{s\leq t} |T_{[sn]} - sn| > \varepsilon\}$ für ein $\varepsilon > 0$.

Auf dieser Menge gibt es $s_1, s_2, ... \leq t$ mit $|T_{[s_n n]} - s_n n| > \varepsilon n$. Dies widerspricht aber $\lim_{n\to\infty} T_{[s_n n]}/[s_n n] = \lim_{n\to\infty} T_n/n = 1$.)

Wir erinnern an die Definition des Stetigkeitsmoduls w aus Definition 17.17. Mit der Skalierungseigenschaft der Brown'schen Bewegung aus Theorem 14.19 folgt, da $S_{[sn]} = X_{T_{[sn]}}$,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbf{P}\Big(\frac{1}{\sqrt{n}} \sup_{0\leq s\leq t} |S_{[sn]} - X_{sn}| > \varepsilon\Big) \\
&\leq \inf_h \limsup_{n\to\infty} \mathbf{P}(w(\mathcal{X}, (t+h)n, nh) > \varepsilon\sqrt{n}) + \mathbf{P}(\sup_{s\leq t} |T_{[sn]} - sn| > nh) \\
&= \inf_h \mathbf{P}(w(\mathcal{X}, t+h, h) > \varepsilon) = 0.
\end{aligned}$$

□

Da nun die Irrfahrt und die Brown'sche Bewegung direkt miteinander in Verbindung stehen, liegt es nahe, Eigenschaften der Brown'schen Bewegung auf Irrfahrten zu übertragen. Dies ist etwa für das Gesetz des iterierten Logarithmus möglich.

Theorem 17.29 (Gesetz des iterierten Logarithmus für Irrfahrten). *Seien $Y_1, Y_2, ...$ reellwertige, unabhängige, identisch verteilte Zufallsvariable mit $\mathbf{E}[Y_1] = 0$, $\mathbf{V}[Y_1] = 1$ und $S_n = Y_1 + \cdots + Y_n$. Dann gilt*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1$$

fast sicher.

Beweis. Wir zeigen nur dass man den Wahrscheinlichkeitsraum so erweitern kann, dass es eine Brown'sche Bewegung $\mathcal{X} = (X_t)_{t\geq 0}$ gibt mit

$$\frac{S_{[t]} - X_t}{\sqrt{2t \log\log t}} \xrightarrow{t\to\infty}_{fs} 0. \tag{17.11}$$

Dann folgt die Aussage nämlich aus dem Gesetz des iterierten Logarithmus für die Brown'sche Bewegung, Theorem 17.10.

Nach Theorem 17.26 gibt es eine Erweiterung des Wahrscheinlichkeitsraumes und Stoppzeiten $0 = T_0, T_1, ...$, so dass $X_{T_n} = S_n$. Wieder gilt $T_n/n \xrightarrow{n\to\infty} 1$ nach dem Gesetz der großen Zahlen, was auch $T_{[t]}/t \xrightarrow{t\to\infty} 1$ impliziert. Sei nun $r > 1$, $c^2 > r - 1$ und $h(t) = \sqrt{2t \log\log t}$. Dann gilt (mit einer ähnlichen Rechnung wie in (17.6)

$$\begin{aligned}
\mathbf{P}\Big(\sup_{r^{n-1}\leq t\leq r^n} |X_t - X_{r^{n-1}}| > ch(r^{n-1})\Big) &= \mathbf{P}\Big(\sup_{0\leq t\leq r^n - r^{n-1}} |X_t| > ch(r^{n-1})\Big) \\
&= 2\mathbf{P}(X_{r^n - r^{n-1}} > ch(r^{n-1})) = 2\mathbf{P}(X_1 > ch(r^{n-1})/\sqrt{r^n - r^{n-1}}) \\
&\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{(r-1)}{\pi \log n}} n^{-c^2/(r-1)},
\end{aligned}$$

da $h(r^{n-1})/\sqrt{r^n - r^{n-1}} \overset{n\to\infty}{\approx} \sqrt{(2\log n)/(r-1)}$. Die rechte Seite ist summierbar ist, also folgt mit dem Borel-Cantelli-Lemma wegen $X_{T_{[t]}} = S_{[t]}$

$$
\begin{aligned}
\mathbf{P}\Big(\limsup_{t\to\infty}\frac{|S_{[t]}-X_t|}{h(t)}=0\Big) &\geq \mathbf{P}\Big(\lim_{r\downarrow 1}\limsup_{t\to\infty}\sup_{t\leq u\leq rt}\frac{|X_u-X_t|}{h(t)}=0\Big)\\
&\geq \mathbf{P}\Big(\lim_{r\downarrow 1}\limsup_{n\to\infty}\sup_{r^{n-1}\leq t\leq r^n}\frac{|X_t-X_{r^{n-1}}|}{h(r^{n-1})}=0\Big).\\
&= \inf_{c>0}\mathbf{P}\Big(\lim_{r\downarrow 1, r<c^2+1}\limsup_{n\to\infty}\sup_{r^{n-1}\leq t\leq r^n}\frac{|X_t-X_{r^{n-1}}|}{h(r^{n-1})}\leq c\Big)=1.
\end{aligned}
$$

Also folgt (17.11). □

18 Stationäre stochastische Prozesse

Unabhängige Familien von Zufallsvariablen stellen eine sehr einfache Struktur dar. Weiter haben wir bereits bestimmte Formen der Abhängigkeit kennen gelernt, insbesondere durch Martingale und Markov-Prozesse. Wir kommen hier zu einer weiteren – handhabbaren – Form von Abhängigkeit, nämlich stationären Familien von Zufallsvariablen. Hierzu sei stets $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und I, abgeschlossen unter Addition (also z.B. $I = \{0, 1, 2, ...\}, I = \mathbb{Z}, I = [0, \infty)$ oder $I = \mathbb{R}$.)

18.1 Begriffe und einfache Beispiele

Wir definieren zunächst die wichtigsten Begriffe. Für spätere Anwendungen ist es sinnvoll, immer den Fall zu betrachten, dass $(X_t)_{t\in I}$ der kanonische Prozess bezüglich $\mathbf{P}$ und ein stationärer Prozess ist, und $\tau((X_t)_{t\in I}) = (X_{t+s})_{t\in I}$ für ein $s \in I$.

Definition 18.1 (Stationäre und ergodische Prozesse). *1. Eine messbare Abbildung $\tau : \Omega \to \Omega$ heißt* maßtreu, *falls $\tau_*\mathbf{P} = \mathbf{P}$.*

2. Für $\tau : \Omega \to \Omega$ messbar heißt das Ereignis $A \in \mathcal{F}$ τ-invariant, falls $\tau^{-1}(A) = A$. Weiter bezeichnen wir mit

$$\mathcal{I} := \mathcal{I}_\tau := \{A \in \mathcal{F} : \tau^{-1}(A) = A\}$$

*die σ-Algebra der τ-*invarianten Ereignisse

3. Ist $\tau : \Omega \to \Omega$ maßtreu und $\mathcal{I}$ trivial (d.h. $\mathbf{P}(A) \in \{0, 1\}$ für $A \in \mathcal{I}_\tau$), dann heißt $(\mathbf{P}, \tau)$ ergodisch*.*

4. Ein stochastischer Prozess $\mathcal{X} = (X_t)_{t\in I}$ heißt stationär, falls für alle $s \in I$

$$(X_{t+s})_{t\in I} \overset{d}{=} (X_t)_{t\in I},$$

oder äquivalent

$$\tau_s(\mathcal{X}) \overset{d}{=} \mathcal{X} \text{ für } \tau_s : (x_t)_{t\in I} \mapsto (x_{t+s})_{t\in I}.$$

Dann ist

$$\mathcal{I}_{\tau_s} = \{A \in \mathcal{B}(E)^{\otimes I} : (X_{t+s})_{t\geq 0} \in A \iff (X_t)_{t\geq 0} \in A\}$$

die σ-Algebra der τ_s-invarianten Ereignisse. Weiter heißt $\mathcal{X}$ ergodisch, wenn $\mathcal{I}_{\tau_s}$ für alle $s \in I$ $\mathbf{P}$-trivial ist.

Beispiel 18.2 (Stationäre und ergodische Prozesse). 1. Ein stochastischer Prozess $\mathcal{X} = (X_t)_{t=0,1,2,...}$ ist genau dann stationär, wenn $(X_t)_{t=0,1,2,...} \stackrel{d}{=} (X_{t+1})_{t=0,1,2,...}$. Er ist genau dann ergodisch, wenn $\mathcal{I}_{\tau_1}$ **P**-trivial ist.

2. Sei $Y : \Omega \to E$ und $\tau : \Omega \to \Omega$ messbar. Weiter definieren wir $\mathcal{X} = (X_t)_{t=0,1,2,...}$ durch
$$X_t(\omega) = Y(\tau^t(\omega)).$$
Dann ist $\mathcal{X}$ genau dann stationär für jede Wahl von Y, wenn τ maßerhaltend ist.
Um dies zu sehen, sei zunächst $\mathcal{X}$ stationär, also $(X_{t+1})_{t=0,1,2,...} \stackrel{d}{=} (X_t)_{t=0,1,2,...}$. Dann gilt $\tau_*\mathbf{E}[Y] = \mathbf{E}[Y \circ \tau] = \mathbf{E}[X_1] = \mathbf{E}[X_0] = \mathbf{E}[Y]$. Da Y beliebig war, folgt, dass τ maßerhaltend ist. Ist andersherum τ maßerhaltend und Y messbar, dann ist $\mathbf{P}(X_0 \in A_0, ..., X_t \in A_t) = \mathbf{P}(Y \circ \tau \in A_0, Y \circ \tau^2 \in A_1, ..., Y \circ \tau^{t+1} \in A_t) = \mathbf{P}(X_1 \in A_0, ..., X_{t+1} \in A_t)$.
Weiter ist $(\mathbf{P}, \tau)$ genau dann ergodisch, wenn der stochastische Prozess X ergodisch ist.

3. Sei $I = 0, 1, 2, ...,$ $\Omega = E^I$, $\tau((\omega_t)_{t\in I}) = (\omega_{t+1})_{t\in I}$ und $Y((\omega_t)_{t\in I}) = \omega_0$. Dann ist mit derselben Notation wie im letzten Beispiel $X_s((\omega_t)_{t\in I}) = (\tau^s(\omega_t)_{t\in I})_0 = ((\omega_{t+s})_{t\in I})_0 = \omega_s$, d.h. $\mathcal{X} = (X_t)_{t\in I}$ ist der kanonische Prozess. Weiter ist $A \in \mathcal{I}$ genau dann, wenn
$$A = \tau^{-t}(A) = \{\omega : \tau^t(\omega) \in A\} = \{\omega : (\omega_t, \omega_{t+1}, ...) \in A\} \in \sigma(X_t, X_{t+1}, ...)$$
für alle $t \in I$. Damit ist
$$\mathcal{I} \subseteq \bigcap_{t\in I} \sigma(X_t, X_{t+1}, ...) = \mathcal{T},$$
wobei $\mathcal{T}$ die terminale σ-Algebra der $X_0, X_1, X_2, ...$ ist.

4. Ist $\mathcal{X} = (X_t)_{t\in I}$ eine unabhängige Familie identisch verteilter Zufallsvariablen, dann ist $\mathcal{X}$ stationär. Nach dem Kolmogorov'schen 0-1-Gesetz, Theorem 9.15, ist $\mathcal{T}$ **P**-trivial und damit ist auch $\mathcal{I}$ **P**-trivial, also ist $\mathcal{X}$ ergodisch.

5. Sei $\mathcal{X} = (X_t)_{t\in I}$ ein zeitlich homogener Markov-Prozess mit Zustandsraum E, Übergangs- und Operator-Halbgruppe $(\mu_t^{\mathcal{X}})_{t\in I}$ und $(T_t^{\mathcal{X}})_{t\in I}$. Angenommen, es gibt ein $\nu \in \mathcal{P}(E)$ mit $\nu\mu_t^{\mathcal{X}} = \nu$, wobei
$$(\nu\mu_t^{\mathcal{X}})(A) = \int \nu(dx)\mu_t^{\mathcal{X}}(x, A),$$
oder äquivalent
$$\int \nu(dx)(T_t^{\mathcal{X}} f)(x) = \int \nu(dx) f(x)$$
für alle $f \in \mathcal{B}(E)$ gilt. Dann ist $\mathcal{X}$ stationär, falls $X_0 \sim \nu$.

6. Ist $\mathcal{X} = (X_t)_{t=0,1,2,...}$ stationär mit Werten in $\mathbb{R}$ und $f : \mathbb{R}^{d+1} \to \mathbb{R}$ messbar, so ist auch $\mathcal{Y} = (Y_t)_{t=0,1,2,...}$ mit
$$Y_t = f(X_t, ..., X_{t+d})$$
stationär. Ist speziell $f(x_0, ..., x_d) = \frac{1}{d+1} \sum_{i=0}^d x_i$, so heißt $\mathcal{Y}$ auch Prozess des gleitendes Mittels.

7. Sind $\mathcal{X} = (X_t)_{t=0,1,2,...}$ identisch verteilt mit dass $X_1 = X_2 = X_3 = ...$ und X_0 unabhängig von $X_1, X_2, ...$, so ist $\mathcal{X}$ nicht stationär.

Wir zeigen nun noch ein einfaches Resultat, um später die Messbarkeit bezüglich $\mathcal{I}$ besser nachprüfen zu können.

Lemma 18.3 (Messbarkeit bezüglich $\mathcal{I}$). *Sei $\tau : \Omega \to \Omega$ messbar. Eine $\mathcal{F}$-messbare Abbildung Y ist genau dann $\mathcal{I}_\tau$-messbar, wenn $Y \circ \tau = Y$.*

Beweis. Sei zunächst $Y = 1_A$. Für '$\Rightarrow$' sei also Y messbar bzgl. $\mathcal{I}_\tau$, d.h. $A \in \mathcal{I}$. Damit ist

$$(Y \circ \tau)(\omega) = 1_A(\tau(\omega)) = 1_{\tau(\omega)\in A} = 1_{\omega\in\tau^{-1}(A)} = 1_{\omega\in A} = Y(\omega).$$

Für '$\Leftarrow$' ist $1_{\tau(\omega)\in A} = 1_{\omega\in A}$, d.h. $\omega \in A \iff \omega \in \tau^{-1}A$ und damit $A = \tau^{-1}(A)$, d.h. $A \in \mathcal{I}$, insbesondere ist Y also $\mathcal{I}_\tau$-messbar. Diese Argumente überträgt man auf einfache Funktionen und schließlich durch ein Approximationsargument auf allgemeine messbare Funktionen. $\square$

18.2 Der Markov-Ketten-Konvergenzsatz

Wir betrachten zunächst den bereits bekannten Fall einer zeitlich homogenen Markov-Kette in diskreter Zeit, d.h. einem Markov-Prozess mit abzählbarem Zustandsraum E und $I = 0, 1, 2, ...$ Wie in den Beispielen 6.10, 12.13 und 16.3 eingeführt, lässt sich eine solche Kette am besten durch die Übergangsmatrix $P = (p(x,y))_{x,y\in E}$ beschreiben, so dass

$$p(x,y) = \mathbf{P}(X_{t+1} = y | X_t = x)$$

gilt. Bekannt ist bereits (siehe die Chapman-Kolmogorov-Gleichungen), dass dann, mit $P^s := (p^s(x,y))_{x,y\in E}$

$$p^s(x,y) = \mathbf{P}(X_{t+s} = y | X_t = x).$$

Weiter schreiben wir $\mathbf{P}_x$ für die Verteilung der Markov-Kette mit $X_0 = x$ fast sicher, und $\mathbf{E}_x$ für den entsprechenden Erwartungswert.

Definition 18.4 (Rekurrenz und Transienz von Zuständen einer Markov-Kette). *Sei $\mathcal{X} = (X_t)_{t\in I}$ eine Markov-Kette mit Übergangsmatrix P,*

$$S_x := \sum_{t=0}^{\infty} 1_{\{X_t = x\}}$$

die Anzahl der Besuche von x,

$$T_x := \inf\{t > 0 : X_t = x\},$$

die erste Treffzeit von x und

$$r_{xy} := \mathbf{P}_x(T_y < \infty) = \mathbf{P}_x(S_y \geq 1).$$

1. *Ein $x \in E$ heißt*

   (a) periodisch *mit Periode $d_x \geq 1$, falls $\{n : p^n(x,x) > 0\}$ den größten gemeinsamen Teiler d_x hat,*

(b) aperiodisch, *falls* $d_x = 1$,
(c) rekurrent, *falls* $r_{xx} = 1$,
(d) positiv rekurrent, *falls* $\mathbf{E}_x[T_x] < \infty$,
(e) nullrekurrent, *falls x rekurrent, aber nicht positiv rekurrent ist,*
(f) transient, *falls* $r_{xx} < 1$.

2. *Die Markov-Kette $\mathcal{X}$ heißt*

   (a) aperiodisch/rekurrent/positiv-rekurrent/null-rekurrent/transient, *falls alle $x \in E$ die Eigenschaft haben*
   (b) irreduzibel, *falls $r_{xy} > 0$ für alle $x, y \in E$.*

3. *Ein Wahrscheinlichkeitsmaß ν auf E heißt* invariant *für $\mathcal{X}$, falls der Zeitshift $\tau : E^I \to E^I$, gegeben durch $\tau((X_t)_{t\in I}) = (X_{t+1})_{t\in I}$, maßtreu für $\mathbf{P}_\nu := \int \nu(dx)\mathbf{P}_x$ ist, d.h.*

$$\mathbf{P}_\nu((X_0, X_1, ...) \in A) = \mathbf{P}_\nu((X_1, X_2, ...) \in A)$$

für alle $A \in \mathcal{B}(E)^I$.

Lemma 18.5 (Okkupationszeiten und Rekurrenzklassen). *Sei $\mathcal{X} = (X_t)_{t=0,1,2,...}$ eine zeitlich homogene Markov-Kette, $x, y \in E$ und $t \geq 0$.*

1. *Es gilt*

$$\begin{aligned}\mathbf{P}_x(S_y \geq n) &= r_{xy} r_{yy}^{n-1},\\ \mathbf{E}_x[S_y] &= \frac{r_{xy}}{1 - r_{yy}}.\end{aligned}$$

2. *Der Zustand x ist genau dann rekurrent wenn S_x unter $\mathbf{P}_x$ fast sicher ∞ ist. Andersfalls ist S_x unter $\mathbf{P}_x$ geometrisch verteilt mit Parameter r_{xx}, und x ist transient.*

3. *Sei x rekurrent und $C_x := \{y \in E : r_{xy} > 0\}$. Dann gilt $r_{yz} = 1$ für alle $y, z \in C_x$. Insbesondere sind alle $y \in C_x$ rekurrent.*

Beweis. 1. Seien $T_x^1 := T_x$ und $T_x^{n+1} := \inf\{t > T_x^n : X_t = x\}$ (für $n = 1, 2, ...$) der Zeitpunkt des $n+1$-ten Besuchs von x. Da $\mathcal{X}$ die starke Markov-Eigenschaft hat (siehe Proposition 16.11) und T_x^n eine Stoppzeit ist, gilt

$$\begin{aligned}\mathbf{P}_x(T_y^{n+1} < \infty) &= \mathbf{E}_x[1_{T_y^n < \infty} \cdot \mathbf{P}_{X_{T_y^n}}(T_y < \infty)]\\ &= \mathbf{P}_x(T_y^n < \infty) \cdot \mathbf{P}_y(T_y < \infty) = \mathbf{P}_x(T_y^n < \infty) \cdot r_{yy}.\end{aligned}$$

Mit Induktion folgt nun

$$\mathbf{P}_x(S_y \geq n) = \mathbf{P}_x(T_y^n < \infty) = r_{xy} r_{yy}^{n-1}$$

und damit

$$\mathbf{E}_x[S_y] = \sum_{n=1}^{\infty} \mathbf{P}_x(S_y \geq n) = \frac{r_{xy}}{1 - r_{yy}}.$$

2. folgt firekt aus 1. Für 3. sei $y \in C_x$. Wegen der starken Markov-Eigenschaft gilt, da x rekurrent ist

$$0 = \mathbf{E}_x[1_{T_y<\infty} \cdot \mathbf{P}_{X_{T_y}}(T_x = \infty)] = \mathbf{P}_x(T_y < \infty) \cdot \mathbf{P}_y(T_x = \infty) = r_{xy}(1 - r_{yx}).$$

Da $r_{xy} > 0$ nach Voraussetzung, muss also $r_{yx} = 1$ sein. Nun wählen wir $r, t \geq 1$ so, dass $p^r(y,x), p^t(x,y) > 0$. Dann gilt

$$\begin{aligned}\mathbf{E}_y[S_y] &\geq \sum_{s=0}^{\infty} p^r(y,x)p^s(x,x)p^t(x,y) = p^r(y,x)p^t(x,y)\sum_{s=0}^{\infty} \mathbf{P}_x(X_s = x)\\ &= p^r(y,x)p^t(x,y)\mathbf{E}_x\Big[\sum_{s=0}^{\infty} 1_{\{X_s=x\}}\Big] = p^r(y,x)p^t(x,y)\mathbf{E}_x[S_x] = \infty.\end{aligned}$$

Mit 1. folgt $S_y = \infty$, also ist y nach 2. rekurrent. Nun gilt auch $r_{yx} = 1$ und damit folgt auch weiter $r_{yz} = 1$ für $y, z \in C_x$. □

Proposition 18.6 (Eigenschaften irreduzibler Ketten). *Sei $\mathcal{X} = (X_t)_{t\in I}$ eine irreduzible homogene Markov-Kette mit Übergangsmatrix P. Dann gilt:*

1. *Entweder sind alle Zustände $x \in E$ positiv rekurrent, oder alle Zustände sind nullrekurrent, oder alle Zustände sind transient.*
2. *Alle $x \in E$ haben dieselbe Periode.*
3. *Ist ν invariant, dann gilt $\nu(x) > 0$ für alle $x \in E$.*

Beweis. Nach Lemma 18.5.3 ist klar, dass entweder alle Zustände rekurrent oder transient sind. Wir zeigen, dass alle Zustände positiv rekurrent sein müssen, wenn es einen positiv rekurrenten Zustand gibt. Angenommen, es gibt also einen positiv rekurrenten Zustand x. Für $y \in E$ sei $t > 0$ so gewählt, dass $\mathbf{P}_x(A) > 0$ für $A := \{X_1, ..., X_{t-1} \neq x, X_t = y\}$. Dann gilt

$$\infty > \mathbf{E}_x[T_x] \geq \mathbf{E}_x[T_x|A] \cdot \mathbf{P}_x(A) \geq (t + \mathbf{E}_y[T_x]) \cdot \mathbf{P}(A).$$

Insbesondere gilt $\mathbf{E}_y[T_x] < \infty$. Andersherum seien $T_x^1, T_x^2, ...$ die Treffzeiten von x und S'_x die Anzahl der Besuche von x vor T_y, und $p := \mathbf{P}_x(T_y \leq T_x) \geq \mathbf{P}_x(A) > 0$. Damit ist S'_x nach oben beschränkt durch eine geometrische Zufallsvariable mit Erfolgswahrscheinlichkeit $p > 0$, hat also endlichen Erwartungswert. Außerdem gilt $T_y \leq T_x^{S'_x+1}$ unter $\mathbf{P}_x$. Da $(T_x^{i+1} - T_x^i)_{i=0,1,2,...}$ unabhängig und identisch verteilt sind, folgt

$$\mathbf{E}_x[T_y] \leq \mathbf{E}_x\Big[T_x^{S'_x+1}\Big] = \mathbf{E}_x\Big[\sum_{i=0}^{S'_x}(T_x^{i+1} - T_x^i)\Big] = \mathbf{E}_x[S'_x]\mathbf{E}_x[T_x^{i+1} - T_x^i] = \mathbf{E}_x[S'_x]\mathbf{E}_x[T_x] < \infty$$

mit Beispiel 15.23. Nun ist $\mathbf{E}_y[T_y] \leq \mathbf{E}_y[T_x] + \mathbf{E}_x[T_y]$, also ist y ebenfalls positiv rekurrent. Daraus folgt nun 1.

Für 2. seien $x, y \in E$ und $r, t \geq 1$ mit $p^r(y,x), p^t(x,y) > 0$. Dann gilt für $s \geq 0$

$$p^{r+s+t}(y,y) \geq p^r(y,x)p^s(x,x)p^t(x,y).$$

Für $s = 0$ folgt damit $p^{r+t}(y,y) > 0$, also gilt $d_y|(r+t)$ (d.h. d_y teilt $r+t$). Gilt weiter $p^s(x,x) > 0$, folgt $d_y|(r+s+t)$ und da $d_y|(r+t)$ auch $d_y|s$. Daraus folgt $d_y \leq d_x$. Aus Symmetriegründen folgt daraus $d_y = d_x$. Für 3. sei $x \in E$. Wir wählen zunächst $y \in E$ mit $\nu(y) > 0$ und $s > 0$ mit $p^s(y,x) > 0$. Dann gilt wegen der Stationarität $\nu(x) = \sum_{z\in E} \nu(z)p^s(z,x) \geq \nu(y)p^s(y,x) > 0$. □

Lemma 18.7 (Invariante Verteilung und Rekurrenz). *Sei $\mathcal{X}$ eine zeitlich homogene Markov-Kette. Falls eine invariate Verteilung $\nu \in \mathcal{P}(E)$ existiert, so sind alle $x \in E$ mit $\nu(x) > 0$ rekurrent.*

Beweis. Sei $x \in E$ mit $\nu(x) > 0$. Zunächst gilt wegen der Invarianz von ν

$$0 < \nu(x) = \int \nu(dy) p^t(y, x).$$

Damit gilt

$$\begin{aligned} \infty = \sum_{t=1}^{\infty} \nu(x) &= \int \nu(dy) \sum_{t=1}^{\infty} p^t(y,x) = \int \nu(dy) \mathbf{E}_y \Big[\sum_{t=1}^{\infty} 1_{\{X_t = x\}} \Big] \\ &= \int \nu(dy) \mathbf{E}_y[S_x] \le \mathbf{E}_x[S_x] \end{aligned}$$

nach Lemma 18.5. Dies ist aber nur möglich, falls $r_{xx} = 1$, x also rekurrent ist. □

Proposition 18.8 (Invariante Verteilungen irreduzibler Ketten). *Sei $\mathcal{X}$ eine irreduzible, positiv rekurrente Markov-Kette. Dann hat $\mathcal{X}$ eine stationäre Verteilung.*

Beweis. Sei $z \in E$. Wir definieren die erwartete Anzahl der Besuche von y vor T_z

$$\widetilde{\nu}(y) := \mathbf{E}_z \Big[\sum_{t=0}^{T_z - 1} 1_{X_t = y} \Big] = \sum_{t=0}^{\infty} \mathbf{P}_z(X_t = y, T_z > t).$$

Klar ist, dass $\widetilde{\nu}(y) \le \mathbf{E}_z[T_z] < \infty$. Wir rechnen nun nach, dass $\widetilde{\nu}$ stationär ist. Hierzu schreiben wir

$$\begin{aligned} \sum_{x \in E} \widetilde{\nu}(x) p(x,y) &= \sum_{t=0}^{\infty} \sum_{x \in E} \mathbf{P}_z(X_t = x, T_z > t) p(x,y) \\ &= \sum_{t=0}^{\infty} \mathbf{P}_z(X_{t+1} = y, T_z > t) \\ &= \sum_{t=1}^{\infty} \mathbf{P}_z(X_t = y, T_z \ge t) \\ &= \widetilde{\nu}(y) - \mathbf{P}_z(X_0 = y, T_z > 0) + \sum_{t=1}^{\infty} \mathbf{P}_z(X_t = y, T_z = t) \\ &= \widetilde{\nu}(y) - \delta_{yz} + \delta_{yz} \sum_{t=1}^{\infty} \mathbf{P}_z(T_z = t) \\ &= \widetilde{\nu}(y). \end{aligned}$$

Außerdem ist

$$\sum_{x \in E} \widetilde{\nu}(x) = \sum_{t=0}^{\infty} \mathbf{P}_z(T_z > t) = \mathbf{E}_z[T_z].$$

Klar ist nun, dass $\nu(x) = \frac{\widetilde{\nu}(x)}{\mathbf{E}_z[T_z]}$ eine invariante Verteilung ist. □

Wir wissen nun, dass irreduzible, positiv rekurrente Markov-Ketten eine stationäre Verteilung besitzen. In Theorem 18.18 werden wir sogar sehen, dass die invariante Verteilung eindeutig ist. Jetzt beschäftigen wir uns jedoch zunächst mit der Konvergenz von Markov-Ketten. Das bedeutet, dass wir untersuchen wollen, wann $X_t \xRightarrow{t\to\infty} \nu$ für ein $\nu \in \mathcal{P}(E)$ gelten kann. Um den Markov-Ketten-Konvergenzsatz zu formulieren, benötigen wir noch den Begriff des Total-Variationsabstandes.

Definition 18.9 (Total-Variationsabstand). *Seien $\mu, \nu \in \mathcal{P}(E)$. Dann ist*

$$d_{TV}(\mu, \nu) := \sup_{A \in \mathcal{B}(E)} |\mu(A) - \nu(A)|$$

der Total-Variationsabstand von μ und ν.

Bemerkung 18.10 (Total-Variationsabstand und schwache Konvergenz). Seien $\mu, \mu_1, \mu_2, ... \in \mathcal{P}(E)$ und $f \in \mathcal{C}_b(E)$. Dann gilt (mit $\mu[f] := \int f d\mu$) für alle $A \in \mathcal{B}(E)$

$$|\mu_n[f] - \mu[f]| \leq ||f|| \cdot (|\mu_n(A) - \mu(A)| + |\mu_n(A^c) - \mu(A^c)|).$$

Daraus folgt insbesondere, dass aus $d_{TV}(\mu_n, \mu) \xrightarrow{n\to\infty} 0$ auch $\mu_n \xRightarrow{n\to\infty} \mu$ folgt.

Theorem 18.11 (Markov-Ketten-Konvergenzsatz). *Sei $\mathcal{X} = (X_t)_{t=0,1,2,...}$ eine zeitlich homogene, irreduzible und aperiodische Markov-Kette mit abzählbarem Zustandsraum E. Dann gilt genau eine der beiden Aussagen:*

1. *$\mathcal{X}$ ist positiv rekurrent und es gibt genau eine invariante Verteilung ν von $\mathcal{X}$. Ist $X_0 \sim \mu$ mit $\mu \in \mathcal{P}(E)$ beliebig, so ist*

$$||(X_t)_*\mathbf{P} - \nu|| \xrightarrow{t\to\infty} 0. \tag{18.1}$$

2. *Es gibt keine invariante Verteilung von $\mathcal{X}$ und es gilt*

$$p^t(x, y) \xrightarrow{t\to\infty} 0 \text{ für alle } x, y \in E. \tag{18.2}$$

   In diesem Fall ist $\mathcal{X}$ entweder transient oder null-rekurrent.

Der Beweis des Markov-Ketten-Konvergenzsatzes benötigt das Werkzeug der Kopplung.

Definition 18.12 (Kopplung).
1. *Seien $\mathcal{X}$ und $\mathcal{Y}$ stochastische Prozesse mit Zustandsräumen E und F. Dann heißt ein stochastischer Prozess $\mathcal{Z}$ eine* Kopplung *von $\mathcal{X}$ und $\mathcal{Y}$, falls $\mathcal{Z}$ Zustandsraum $E \times F$ hat, $\pi_E(\mathcal{Z})$ genauso verteilt ist wie $\mathcal{X}$ und $\pi_F(\mathcal{Z})$ genauso verteilt ist wie $\mathcal{Y}$.*

2. *Sind weiter $\pi_E(\mathcal{Z})$ und $\pi_F(\mathcal{Z})$ unabhängig, so heißt $\mathcal{Z}$ eine* unabhängige Kopplung *von $\mathcal{X}$ und $\mathcal{X}$.*

Lemma 18.13 (Eigenschaften der unabhängigen Kopplung). *Sei $\mathcal{X} = (X_t)_{t=0,1,2,...}$ eine Markov-Kette mit Übergangsmatrix P. Dann ist die Markov-Kette $\mathcal{Z}$ mit Übergangsmatrix Q, gegeben durch $q((x, x'), (y, y')) := p(x, y)p(x', y')$ eine unabhängige Kopplung von $\mathcal{X}$ und $\mathcal{X}$. Ist $\mathcal{X}$ irreduzibel und aperiodisch, dann gilt dies auch für $\mathcal{Z}$. Gibt es in diesem Fall eine invariante Verteilung von $\mathcal{X}$, so ist $\mathcal{Z}$ rekurrent.*

Beweis. Klar ist, dass $\mathcal{Z}$ eine unabhängige Kopplung von $\mathcal{X}$ und $\mathcal{X}$ ist und $\mathcal{Z}$ irreduzibel ist, falls $\mathcal{X}$ irreduzibel ist. Weiter haben nach Proposition 18.6 alle $x \in E$ dieselbe Persiode. Da $p^t(x,x) > 0$ genau dann gilt, wenn $q^t((x,x),(x,x)) > 0$, ist $d_{(x,x)} = 1$. Da wegen der Irreduzibilität auch alle Zustände in $E \times E$ dieselbe Periode haben (siehe Proposition 18.6.2), folgt die Aperiodizität von $\mathcal{Z}$. Ist ν eine invariante Verteilung für $\mathcal{X}$, so ist $\nu \otimes \nu$ invariant für $\mathcal{Z}$ und die Rekurrenz von $\mathcal{Z}$ folgt aus Lemma 18.7, da nach Proposition 18.6 alle Zustände rekurrent sein müssen. □

Beweis von Theorem 18.11. Wir beweisen nun den Markov-Ketten-Konvergenzsatz in zwei Schritten. Falls (18.2) nicht gilt, gibt es nach Schritt 2 eine stationäre Verteilung, und nach Schritt 1 ist diese eindeutig und die in (18.1) behauptete Konvergenz gilt.

Schritt 1: Konvergenz und Eindeutigkeit der stationären Verteilung: Wir nehmen an, dass es eine invariante Verteilung ν gibt. Seien $\mathcal{X} = (X_t)_{t=0,1,2,...}$ und $\mathcal{Y} = (Y_t)_{t=0,1,2,...}$ zwei Markov-Ketten mit Übergangsmatrix P mit $X_0 \sim \mu \in \mathcal{P}(E)$, $Y_0 \sim \mu' \in \mathcal{P}(E)$. Wir zeigen hier, dass dann immer

$$||(X_t)_*\mathbf{P} - (Y_t)_*\mathbf{P}|| \xrightarrow{t\to\infty} 0 \tag{18.3}$$

gilt. Gilt $\mu' = \nu$, so ist dies also gerade die Konvergenz (man beachte, dass dann $(Y_t)_*\mathbf{P} = \nu$) in (18.1). Weiter folgt aus (18.3) die Eindeutigkeit der invarianten Verteilung für $\mathcal{X}$ wie folgt: angenommen, es gibt eine weitere invariante Verteilung ν'. Dann gilt, falls $X_0 \sim \nu', Y_0 \sim \nu$, dass $0 = \lim_{n\to\infty} ||(X_n)_*\mathbf{P} - (Y_n)_*\mathbf{P}|| = ||\nu' - \nu||$, also $\nu' = \nu$.

Es bleibt also (18.3) zu zeigen, und zwar unter der Annahme, dass es eine invariante Verteilung gibt. Aus Lemma 18.13 folgt, dass die unabhängige Kopplung $\mathcal{Z}$ rekurrent ist. Deswegen ist die Stoppzeit

$$T := \min\{t : X_t = Y_t\}$$

fast sicher endlich. Wir betrachten nun eine Kopplung $\widetilde{\mathcal{Z}}$ mit $\widetilde{\mathcal{Z}} = (X_t, \widetilde{Y}_t)_{t=0,1,2,...}$, so dass

$$\widetilde{Y}_t := \begin{cases} Y_t, & t < T, \\ X_t, & t \geq T. \end{cases}.$$

Da $\mathcal{Z}$ stark Markov ist, ist $\widetilde{\mathcal{Z}}$ Markov mit $(\widetilde{Y}_t)_{t=0,1,2,...} \stackrel{d}{=} (Y_t)_{t=0,1,2,...}$. Insbesondere ist $\widetilde{\mathcal{Z}}$ eine Kopplung von $\mathcal{X}$ und $\mathcal{Y}$. Außerdem gilt

$$\begin{aligned}|\mathbf{P}(X_t \in A) - \mathbf{P}(Y_t \in A)| &= |\mathbf{P}(X_t \in A) - \mathbf{P}(\widetilde{Y}_t \in A)| = |\mathbf{P}(X_t \in A) - \mathbf{P}(\widetilde{Y}_t \in A, T >)| \\ &= |\mathbf{P}(X_t \in A, T > t) - \mathbf{P}(\widetilde{Y}_t \in A, T > t)| \leq 2\mathbf{P}(T > t).\end{aligned}$$

Da die rechte Seite nicht von A abhängt, folgt

$$||(X_t)_*\mathbf{P} - (Y_t)_*\mathbf{P}|| \leq 2\mathbf{P}(T > t) \xrightarrow{t\to\infty} 0.$$

Schritt 2: Existenz der stationären Verteilung, falls (18.2) *nicht gilt:* Falls (18.2) nicht gilt, gibt es $\widetilde{x}, \widetilde{y} \in E$ und eine Folge $t_1, t_2, ...$ mit $p^{t_k}(\widetilde{x}, \widetilde{y}) \xrightarrow{k\to\infty} c_{\widetilde{y}} > 0$ und $p^{t_k}(\widetilde{x}, y) \xrightarrow{k\to\infty} c_y \geq 0$ für $y \neq \widetilde{y}$. Damit gilt

$$0 < \sum_{y\in E} c_y \leq \lim_{k\to\infty} \sum_{y\in E} p^{t_k}(\widetilde{x}, y) = 1$$

nach Fatous Lemma, Theorem 4.25. Wir zeigen nun $p^{t_k}(x,y) \xrightarrow{t\to\infty} c_y$ für alle $x \in E$. Hierfür sei $\mathcal{Z}$ die unabhängige Kopplung von $\mathcal{X}$ und $\mathcal{X}$ aus Definition 18.12. Nach Lemma 18.13 ist $\mathcal{Z}$ ebenfalls irreduzibel und aperiodisch und hat die Übergangsmatrix $Q = (q((x,x'),(y,y')))_{(x,x'),(y,y')\in E^2}$ mit $q((x,x'),(y,y')) = p(x,y)p(x',y')$. Die Markov-Kette $\mathcal{Z}$ muss rekurrent sein. Andernfalls wäre sie nämlich nach Proposition 18.6 transient, also müsste

$$\sum_{t=0}^{\infty}(p^2(x,y))^t = \sum_{t=0}^{\infty} q^t((x,x),(y,y)) = \sum_{t=0}^{\infty} \mathbf{P}_{(x,x)}(Z_t = (y,y)) = \mathbf{E}_{(x,x)}[S_{(y,y)}] < \infty$$

gelten, was aber $p^t(x,y) \xrightarrow{t\to\infty} 0$ für alle $x,y \in E$, also (18.2) implizieren würde. Sei $X_0 = x$ und $Y_0 = \widetilde{x}$. Da $\mathcal{Z}$ rekurrent ist, ist $T := \inf\{t : X_t = Y_t\}$ fast sicher endlich. Wie in Schritt 1 folgt daraus, dass $|\mathbf{P}(X_t = y) - \mathbf{P}(Y_t = y)| = |p^t(x,y) - p^t(\widetilde{x},y)| \xrightarrow{t\to\infty} 0$, also folgt $\lim_{k\to\infty} p^{t_k}(x,y) = \lim_{k\to\infty} p^{t_k}(\widetilde{x},y) = c_y$ für alle $x \in E$.

Nun gilt für $x,z \in E$

$$p^{t+1}(x,z) = \sum_{y\in E} p^t(x,y)p(y,z) = \sum_{y\in E} p(x,y)p^t(y,z),$$

also mit dem Lemma von Fatou und majorisierter Konvergenz

$$\sum_{y\in E} c_y p(y,z) \le \sum_{y\in E} p(x,y)c_z = c_z.$$

Da alle Terme positiv sind und $\sum_{z\in E} c_z \le 1$ gilt, folgt durch Summation über z, dass

$$\sum_{y\in E} c_y p(y,z) = c_z.$$

Mit anderen Worten ist die Verteilung ν mit $\nu(x) = c_x/(\sum_{y\in E} c_y)$ invariant. □

18.3 Ergodensätze in diskreter Zeit

Sei $I = \{0,1,2,...\}$ und τ maßerhaltend. Ferner sei $f : \Omega \to \mathbb{R}$ messbar und für jedes $t \in I$

$$X_s(\omega) = f(\tau^s(\omega)). \tag{18.4}$$

wie in Beispiel 18.2.2. Wir behandeln im folgenden den Fall von Gesetzen großer Zahlen für $X = (X_t)_{t\in I}$. Das bedeutet, wir betrachten den Prozess $\mathcal{S} = (S_t)_{t\in I}$ mit

$$S_t = \sum_{s=0}^{t-1} X_s.$$

Als Vorbereitung bringen wir ein grundlegendes Lemma.

Lemma 18.14 (Hopf'sches Maximal-Ergodenlemma). *Sei $X_0 \in \mathcal{L}^1$ und $M_t := \max\{0, S_1, ..., S_t\}$ für $t \in I$. Dann gilt für jedes $t \in I$*

$$\mathbf{E}[X_0; M_t > 0] \ge 0.$$

Beweis. Für $s \leq t$ ist sicher $M_t(\tau(\omega)) \geq S_s(\tau(\omega))$. Damit ist auch $X_0 + M_t \circ \tau \geq X_0 + S_s \circ \tau = S_{s+1}$, d.h. für $s = 1, ..., t$ ist

$$X_0 \geq S_{s+1} - M_t \circ \tau. \tag{18.5}$$

Weiter ist $S_1 = X_0$ und $M_t \circ \tau \geq 0$, d.h. (18.5) gilt auch für $s = 0$. Daraus folgern wir

$$X_0 \geq \max\{S_1, ..., S_t\} - M_t \circ \tau.$$

Außerdem ist, da $M_t \circ \tau \geq 0$ immer gilt,

$$\{M_t > 0\}^c \subseteq \{M_t = 0\} \cap \{M_t \circ \tau \geq 0\} \subseteq \{M_t - M_t \circ \tau \leq 0\}.$$

Aus der Maßtreue von τ und den letzten beiden Gleichungen folgert man

$$\begin{aligned} \mathbf{E}[X_0; M_t > 0] &\geq \mathbf{E}[(\max\{S_1, ..., S_t\} - M_t \circ \tau); M_t > 0] \\ &= \mathbf{E}[(M_t - M_t \circ \tau); M_t > 0] \\ &\geq \mathbf{E}[M_t - M_t \circ \tau] = \mathbf{E}[M_t] - \mathbf{E}[M_t] = 0. \end{aligned}$$

□

Theorem 18.15 (Fast sicherer Ergodensatz, Birkhoff). *Sei $\mathcal{X} = (X_t)_{t \in I}$ wie in (18.4), $\mathcal{I} = \mathcal{I}_\tau$ und $X_0 \in \mathcal{L}^1$. Dann gilt*

$$\frac{1}{t} \sum_{s=0}^{t-1} X_s \xrightarrow{t \to \infty}_{fs} \mathbf{E}[X_0 | \mathcal{I}].$$

Ist speziell X ergodisch, so gilt $\mathbf{E}[X_0|\mathcal{I}] = \mathbf{E}[X_0]$.

Beweis. Der Zusatz ist klar, da $\mathcal{I}$ trivial ist, falls X ergodisch ist. Da $\mathbf{E}[X_0|\mathcal{I}]$ eine $\mathcal{I}$-messbare Zufallsvariable ist, ist $\mathbf{E}[X_0|\mathcal{I}] \circ \tau = \mathbf{E}[X_0|\mathcal{I}]$ nach Lemma 18.3. Damit ist $(X_t - \mathbf{E}[X_0|\mathcal{I}])_{t \in I}$ stationär und wir können Œ annehmen, dass $\mathbf{E}[X_0|\mathcal{I}] = 0$ ist. Setze $S_t = X_0 + \cdots + X_{t-1}$ und $S^* := \limsup_{t \to \infty} \frac{1}{t} S_t$. Für beliebiges $\varepsilon > 0$ werden wir $\mathbf{P}(S^* > \varepsilon) = 0$ zeigen, woraus mittels eines Symmetrieargumentes $\lim_{t \to \infty} \frac{1}{t} S_t = 0$ fast sicher folgt. Offenbar ist $S^* \circ \tau = S^*$, d.h. S^* ist $\mathcal{I}$-messbar, also $\{S^* > \varepsilon\} \in \mathcal{I}$. Wir setzen $A := \{S^* > \varepsilon\}$,

$$\begin{aligned} X_t^\varepsilon &:= (X_t - \varepsilon) 1_A, \\ S_t^\varepsilon &= X_0^\varepsilon + ... + X_{t-1}^\varepsilon, \\ M_t^\varepsilon &:= \max\{0, S_1^\varepsilon, ..., S_t^\varepsilon\}, \\ A_t &:= \{M_t^\varepsilon > 0\}. \end{aligned}$$

Dann ist $A_1 \subseteq A_2 \subseteq ...$ und

$$\bigcup_{t=1}^\infty A_t = \Big\{ \sup_{t \geq 1} S_t^\varepsilon > 0 \Big\} = \Big\{ \sup_{t \geq 1} \tfrac{1}{t} S_t > \varepsilon \Big\} \cap A = A.$$

Der letzte Schritt folgt hier, da aus $\omega \in A$ folgt, dass es ein (und in der Tat sogar unendlich viele) $t \in I$ gibt mit $\frac{1}{t} S_t > \varepsilon$. Damit gilt $A_t \uparrow A$ und $\mathbf{E}[X_0^\varepsilon; A_t] \xrightarrow{t \to \infty} \mathbf{E}[X_0^\varepsilon]$ nach majorisierter Konvergenz. Nach Lemma 18.14 ist $\mathbf{E}[X_0^\varepsilon; A_t] \geq 0$, also, da $A \in \mathcal{I}$,

$$0 \leq \mathbf{E}[X_0^\varepsilon] = \mathbf{E}[(X_0 - \varepsilon); A] = \mathbf{E}[\mathbf{E}[X_0|\mathcal{I}]; A] - \varepsilon \mathbf{P}(A) = -\varepsilon \mathbf{P}(A).$$

Mit anderen Worten, $\mathbf{P}(A) = 0$. □

Lemma 18.16. *Seien $X_0, X_1, \dots$ identisch verteilt und $X_0 \in L^p$ für $p \geq 1$. Dann ist*

$$\Big\{\Big|\tfrac{1}{t}\sum_{s=0}^{t-1} X_s\Big|^p : t = 1, 2, \dots\Big\}$$

gleichgradig integrierbar.

Beweis. Da $\{|X_0|^p\}$ gleichgradig integrierbar ist, gibt es nach Lemma 8.9 eine monoton wachsende, konvexe Funktion f mit $\frac{f(x)}{x} \xrightarrow{x\to\infty} \infty$ und $\mathbf{E}[f(|X_0|^p)] < \infty$. Weiter ist, mit der Jensen'schen Ungleichung

$$\begin{aligned}\sup_{t=1,2,\dots} \mathbf{E}\Big[f\Big(\Big|\tfrac{1}{t}\sum_{s=0}^{t-1} X_s\Big|^p\Big)\Big] &\leq \sup_{t=1,2,\dots} \mathbf{E}\Big[f\Big(\tfrac{1}{t}\sum_{s=0}^{t-1} |X_s|^p\Big)\Big] \\ &\leq \sup_{t=1,2,\dots} \tfrac{1}{t}\sum_{s=0}^{t-1} \mathbf{E}\big[f(|X_s|^p)\big] = \mathbf{E}[f(|X_0|^p)],\end{aligned}$$

woraus die Behauptung (wieder mit Lemma 8.9) folgt. □

Theorem 18.17 (L^p-Ergodensatz, von Neumann). *Sei $\mathcal{X} = (X_t)_{t\in I}$ wie in (18.4), $\mathcal{I} = \mathcal{I}_\tau$ und $X_0 \in L^p$ für $p \geq 1$. Dann gilt*

$$\frac{1}{t}\sum_{s=0}^{t-1} X_s \xrightarrow{t\to\infty}_{L^p} \mathbf{E}[X_0|\mathcal{I}].$$

Beweis. Nach Lemma 18.16 ist

$$\Big\{\Big|\tfrac{1}{t}\sum_{s=0}^{t-1} X_s\Big|^p : t = 1, 2, \dots\Big\}$$

gleichgradig integrierbar. Da die behauptete Konvergenz fast sicher (und damit stochastisch) nach Theorem 18.15 gilt, gilt sie also auch in L^p nach Theorem 8.11. □

Wir wenden nun den Ergodensatz auf irreduzible rekurrente Markov-Ketten an. Diese sind nämlich ergodisch, und wir können zeigen, dass die invariante Verteilung aus Proposition 18.8 eindeutig ist.

Theorem 18.18 (Ergodizität von Markov-Ketten). *Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ eine zeitlich homogene, stationäre, rekurrente und irreduzible Markov-Kette mit abzählbarem Zustandsraum E. Dann ist $\mathcal{X}$ ergodisch und positiv rekurrent und die stationäre Verteilung ν ist eindeutig gegeben durch*

$$\nu(x) = \frac{1}{\mathbf{E}_x[T_x]}, \qquad x \in E. \tag{18.6}$$

Bemerkung 18.19 (Markov-Ketten). Aus dem Ergodensatz liest man ab, dass

$$\frac{1}{t}\sum_{s=0}^{t-1} 1_{\{X_s=x\}} \xrightarrow{t\to\infty}_{fs} \nu(x). \tag{18.7}$$

Man beachte, dass hier auf der linken Seite die mittlere Zeit steht, in der von $\mathcal{X}$ der Wert x angenommen wurde und auf der rechten Seite die Wahrscheinlichkeit, mit der man im Gleichgewicht den Prozess in x beobachtet. Man sagt hier auch, dass das Zeitmittel gleich dem Raummittel wird. Diese Beobachtung wird im Beweis dazu verwendet werden, die stationäre Verteilung eindeutig zu bestimmen.

Beweis von Theorem 18.18. Zunächst zeigen wir, dass $\mathcal{X}$ ergodisch ist. Wir schreiben τ für den Zeitshift, also $\tau((X_t)_{t=0,1,2,\ldots}) = (X_t)_{t=1,2,\ldots}$. Sei $A \in \mathcal{I}$. Nach Beispiel 18.2.3 ist $A \in \mathcal{T} = \bigcap_{t=0}^\infty \sigma(X_t, X_{t+1}, \ldots)$, also für eine fast sicher endliche Stoppzeit T mit der starken Markov-Eigenschaft

$$\mathbf{P}_\nu(\mathcal{X} \in A|\mathcal{F}_T) = \mathbf{P}_\nu(\tau^T \circ \mathcal{X} \in A|\mathcal{F}_T) = \mathbf{P}((X_T, X_{T+1}, \ldots) \in A|\mathcal{F}_T) = \mathbf{P}_{X_T}(\mathcal{X} \in A). \quad (18.8)$$

Für $T = T_x$ gilt also, da $T < \infty$ fast sicher,

$$\mathbf{P}_\nu(\mathcal{X} \in A) = \mathbf{P}_x(\mathcal{X} \in A), \qquad x \in E.$$

Deswegen gilt auch nach Theorem 15.35

$$\begin{aligned}\mathbf{P}_\nu(\mathcal{X} \in A) &= \sum_{x\in E} 1_{\{X_t=x\}} \mathbf{P}_x(\mathcal{X} \in A) = \mathbf{P}_{X_t}(\mathcal{X} \in A)\\ &= \mathbf{P}_\nu(\mathcal{X} \in A|X_0, \ldots, X_t) \xrightarrow{t\to\infty} \mathbf{P}_\nu(\mathcal{X} \in A|X_0, X_1, \ldots) = 1_{\{\mathcal{X}\in A\}}.\end{aligned}$$

Insbesondere ist $\mathbf{P}_\nu(X \in A) \in \{0, 1\}$, d.h. $\mathcal{I}$ ist trivial und $\mathcal{X}$ ist ergodisch.

Wir zeigen nun mit Hilfe des Birkhoff'schen Ergodensatzes, dass

$$\nu(x) = \frac{1}{\mathbf{E}_x[T_x]} \quad (18.9)$$

gelten muss. Daraus folgt dann, dass es mindestens einen positiv rekurrenten Zustand gibt, $\mathcal{X}$ also nach Proposition 18.6.1 positiv rekurrent ist. Weiter zeigt dies die Eindeutigkeit der invarianten Verteilung.

Mit $\mathcal{X}$ ist auch $(1_{\{X_t=x\}})_{t=0,1,2,\ldots}$ ergodisch unter ν. Sei $L_t = \sum_{s=1}^t 1_{\{X_s=x\}} = |\{s \leq t : X_s = x\}|$ die Anzahl der Besuche von x bis zur Zeit t. Dann folgt aus dem Ergodensatz (siehe auch (18.7))

$$\frac{L_t}{t} \xrightarrow{t\to\infty} \nu(x)$$

fast sicher. Andererseits sei T_x^n der Zeitpunkt des n-ten Besuches von x. Da $(T_x^{k+1} - T_x^k)_{k=1,2,\ldots}$ unabhängig und identisch verteilt sind, gilt nach dem starken Gesetz der großen Zahlen

$$\frac{T_x^n}{L_{T_x^n}} = \frac{1}{n}\Big(T_x^1 + \sum_{k=1}^n T_x^{k+1} - T_x^k\Big) \xrightarrow{n\to\infty} \mathbf{E}_x[T_x^2 - T_x^1] = \mathbf{E}_x[T_x]$$

fast sicher unter $\mathbf{P}_\nu$. Deswegen gilt auch

$$\frac{L_t}{t} \xrightarrow{t\to\infty} \frac{1}{\mathbf{E}_x[T_x]}.$$

Da der Limes von L_t/t eindeutig sein muss, folgt (18.9).

□

Als Anwendung des Ergoden-Satzes bringen wir nun noch einen Satz, der zeigt, dass alle zentrierten Irrfahrten auf $\mathbb{Z}$ rekurrent sind. Hier folgt zunächst ein Lemma.

Lemma 18.20 (Range und Fluchtwahrscheinlichkeit). *Sei $\mathcal{Y} = (Y_t)_{t=0,1,2,...}$ stationär (also etwa eine unabhängige Familie) mit Werten in $\mathbb{R}^d$, $S_t = Y_1 + \cdots + Y_t$,*

$$R_t := |\{S_1, ..., S_t\}| \tag{18.10}$$

der Range *von $\mathcal{S} = (S_t)_{t=0,1,2,...}$ und*

$$A := \{S_1, S_2, ... \neq 0\}$$

das Fluchtereignis*. Dann gilt*

$$\frac{1}{t} R_t \xrightarrow{t\to\infty} \mathbf{P}(A|\mathcal{I}).$$

Beweis. Sei τ der Zeitshift um 1. Zunächst ist

$$R_t = |\{r \leq t : S_r \neq S_{r+1}, ..., S_t\}| \geq |\{r \leq t : S_r \neq S_{r+1}, S_{r+2}, ...\}| = \sum_{r=1}^{t} 1_A \circ \tau^r.$$

Der Birkhoff'sche Ergodensatz liefert also

$$\liminf_{t\to\infty} \frac{1}{t} R_t \geq \mathbf{P}(A|\mathcal{I})$$

fast sicher. Für die Umkehrung sei $s \geq 1$ und $A_s := \{S_1, ..., S_s \neq 0\}$, wobei $A = \cap_{s=1}^{\infty} A_s$. Es gilt

$$\begin{aligned} R_t &\leq s + |\{r \leq t - s : S_r \neq S_{r+1}, ..., S_t\}| \\ &\leq s + |\{r \leq t - s : S_r \neq S_{r+1}, ..., S_{r+s}\}| = s + \sum_{r=1}^{t-s} 1_{A_s} \circ \tau^r \end{aligned}$$

Wieder liefert der Birkhoff'sche Ergodensatz

$$\limsup_{t\to\infty} \frac{1}{t} R_t \leq \mathbf{P}(A_s|\mathcal{I}) \xrightarrow{s\to\infty} \mathbf{P}(A|\mathcal{I})$$

fast sicher wegen majorisierter Konvergenz. □

Theorem 18.21 (Satz von Chung-Fuchs). *Seien $Y_1, Y_2, ...$ mit Werte in $\mathbb{Z}$, unabhängig, identisch verteilt mit $\mathbf{E}[Y_1] = 0$, sowie $S_0 = 0$ und $S_t = Y_1 + \cdots + Y_t, t = 1, 2, ...$ Dann ist $\mathcal{S} = (S_t)_{t=0,1,2,...}$ rekurrent.*

Beweis. Sei R_t wie in (18.10) und $A = \{S_1, S_2, ... \neq 0\}$. Wir zeigen $\frac{1}{t} R_t \xrightarrow{t\to\infty}_{fs} 0$. Daraus folgt nach Lemma 18.20, dass $\mathbf{P}(A) = 0$, was die Rekurrenz der Markov-Kette $\mathcal{S}$ impliziert. Da $R_t \leq 1 + \max_{s=1,...,t} S_s - \min_{s=1,...,t} S_s$, genügt es aus Symmetrie zu zeigen, dass $\frac{1}{t} \max_{s=1,...,t} S_s \xrightarrow{t\to\infty}_{fs} 0$. (Falls $\mathbf{E}[Y_1^2] < \infty$ folgt dies aus dem Gesetz des iterierten Logarithmus.) Da $\frac{1}{t} S_t \xrightarrow{t\to\infty} 0$ nach dem Gesetz der großen Zahlen, folgt

$$\limsup_{t\to\infty} \frac{1}{t} \max_{r=1,...,t} S_r = \lim_{s\to\infty} \limsup_{t\to\infty} \frac{1}{t} \max_{r=s,...,t} S_r \leq \lim_{s\to\infty} \sup_{r\geq s} \frac{S_r}{r} = \limsup_{r\to\infty} \frac{S_r}{r} = 0$$

und die Behauptung ist gezeigt. □

18.4 Mischung

Eine Familie von unabhängigen, identisch verteilten Zufallsvariablen sind nach Beispiel 18.2.4 ergodisch. Deshalb kann man sagen, dass Ergodizität eine Abschwächung des Begriffes *Unabhängigkeit* bedeutet. Jedoch erinnern wir an Theorem 18.18, bei dem für eine ergodische Markov-Kette erlaubt war, dass sie periodisch ist und damit einen hohen Grad von Abhängigkeit aufweist. Deshalb führen wir nun den Begriff der Mischung ein. Eine mischende Familie von Zufallsvariablen ist weniger abhängig als eine ergodische, muss aber nicht unabhängig sein.

Wir erinnern zunächst an den Begriff des Cesàro-Grenzwertes.

Bemerkung 18.22 (Cesàro-Grenzwert). Aus der Analysis ist der Begriff des Cesàro-Grenzwertes bekannt. Ist $x_1, x_2, ...$ eine Folge, dann ist, falls existent,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i$$

der Cesàro-Grenzwert dieser Folge. Ist etwa x dieser Grenzwert, schreiben wir im Folgenden

$$C - \lim_{n\to\infty} x_n = x \qquad \text{oder} \qquad x_n \xrightarrow[\text{Cesàro}]{n\to\infty} x.$$

Man erinnere sich daran, dass für konvergente Folgen $(x_n)_{n=1,2,...}$ der Cesàro-Grenzwert ebenfalls existiert und mit dem Grenzwert der Folge übereinstimmt. Andererseits gibt es Folgen, deren Cesàro-Grenzwert existiert, die jedoch selbst nicht konvergieren. Etwa ist der Cesàro-Grenzwert von $y_1 = 0, y_2 = 1, y_3 = 0, y_4 = 1, y_5 = 0, ...$ gerade $\frac{1}{2}$. Weiter gelten für Cesàro-Grenzwerte etwas andere Regeln als für übliche Grenzwerte. Im folgenden ist wichtig einzusehen, dass aus $x_n \xrightarrow[\text{Cesàro}]{n\to\infty} x$ nicht folgt, dass $|x_n - x| \xrightarrow[\text{Cesàro}]{n\to\infty} 0$. (Etwa ist ja $y_n \xrightarrow[\text{Cesàro}]{n\to\infty} \frac{1}{2}$, jedoch $|y_n - \frac{1}{2}| \xrightarrow[\text{Cesàro}]{n\to\infty} \frac{1}{2}$.)

Lemma 18.23 (Ergodizität und Cesàro-Grenzwerte). *Eine maßtreue Abbildung τ ist genau dann ergodisch, wenn*

$$\mathbf{P}(A \cap \tau^{-t} B) \xrightarrow[\textit{Cesàro}]{t\to\infty} \mathbf{P}(A) \cdot \mathbf{P}(B), \qquad A, B \in \mathcal{F}. \tag{18.11}$$

Beweis. '⇒': Sei τ ergodisch. Dann gilt nach dem Birkhoff'schen Ergodensatz $1_B \circ \tau^t \xrightarrow[\text{Cesàro}]{t\to\infty} \mathbf{P}(B)$ fast sicher, also auch $1_A \cdot 1_B \circ \tau^t \xrightarrow[\text{Cesàro}]{t\to\infty} 1_A \cdot \mathbf{P}(B)$. Wegen majorisierter Konvergenz gilt dann auch $\mathbf{P}(A \cap \tau^{-t}(B)) \xrightarrow[\text{Cesàro}]{t\to\infty} \mathbf{P}(A) \cdot \mathbf{P}(B)$.

'⇐': Falls (18.11) gilt, und $A \in \mathcal{I}$, so gilt wegen $A = A \cap \tau^{-t}(A)$, dass $\mathbf{P}(A) = \mathbf{P}(A \cap \tau^{-t}(A)) \xrightarrow[\text{Cesàro}]{t\to\infty} \mathbf{P}(A)^2$. Das ist allerdings nur für $\mathbf{P}(A) \in \{0, 1\}$ möglich, und damit ist τ ergodisch. □

Definition 18.24 (Schwach mischend und mischend). 1. *Sei τ eine maßtreue Abbildung. Dann heißt $(\mathbf{P}, \tau)$* mischend*, falls für alle $A, B \in \mathcal{F}$*

$$\mathbf{P}(A \cap \tau^{-t} B) \xrightarrow{t\to\infty} \mathbf{P}(A) \cdot \mathbf{P}(B)$$

gilt. Falls

$$|\mathbf{P}(A \cap \tau^{-t}B) - \mathbf{P}(A) \cdot \mathbf{P}(B)| \xrightarrow[\text{Cesàro}]{t\to\infty} 0$$

so heißt $(\mathbf{P}, \tau)$ schwach mischend.

2. *Sei* $\mathcal{X} = (X_t)_{t=0,1,2,...}$ *ein stationärer Prozess. Dann heißt* $\mathcal{X}$ mischend, *falls für alle* $A, B, \in \mathcal{B}(E)^I$

$$\mathbf{P}((X_0, X_1, ...) \in A, (X_t, X_{t+1}, ...) \in B) \xrightarrow{t\to\infty} \mathbf{P}(\mathcal{X} \in A) \cdot \mathbf{P}(\mathcal{X} \in B).$$

Er heißt schwach mischend, *falls für alle* $A, B, \in \mathcal{B}(E)^I$

$$|\mathbf{P}((X_0, X_1, ...) \in A, (X_t, X_{t+1}, ...) \in B) - \mathbf{P}(\mathcal{X} \in A) \cdot \mathbf{P}(\mathcal{X} \in B)| \xrightarrow[\text{Cesàro}]{t\to\infty} 0.$$

Theorem 18.25 (Mischend, schwach mischend und ergodisch). *Sei* τ *eine maßtreue Abbildung. Ist* τ *mischend, so ist* τ *auch schwach mischend. Ist* τ *schwach mischend, so ist* τ *auch ergodisch.*

Beweis. Alle Aussagen folgen aus den Rechenregeln für Cesàro-Grenzwerte aus Bemerkung 18.22 sowie aus Lemma 18.23. □

Beispiel 18.26 (Ergodisch aber nicht schwach mischend). Sei $\mathcal{X} = (X_t)_{t=0,1,2,...}$ eine zeitlich homogene Markov-Kette mit Zustandsraum $\{0, 1\}$ und Übergangsmatrix P mit $p(x, 1-x) = 1$, gestartet in ν mit $\nu(0) = \nu(1) = \frac{1}{2}$. Dann ist $\mathcal{X}$ stationär, nach Theorem 18.18 ist $\mathcal{X}$ also ergodisch. Allerdings ist $\mathcal{X}$ nicht schwach mischend. Sei hierzu etwa $A = B = \{X_0 = 1\}$. Dann gilt

$$\mathbf{P}(A \cap \tau^{-t}(B)) = \begin{cases} \frac{1}{2}, & t \text{ gerade},\\ 0, & t \text{ ungerade}. \end{cases}$$

jedoch $\mathbf{P}(A) \cdot \mathbf{P}(B) = \frac{1}{4}$. Also folgt

$$|\mathbf{P}(A \cap \tau^{-t}(B)) - \mathbf{P}(A) \cdot \mathbf{P}(B)| \xrightarrow[\text{Cesàro}]{t\to\infty} \frac{1}{4}.$$

Es gibt Beispiele für maßtreue, schwach mischende Abbildungen, die nicht mischend sind. Auf diese wollen wir hier jedoch nicht eingehen.

Wir haben schon gesehen, dass positiv rekurrente, irreduzible Markov-Ketten ergodisch sind. Nun zeigen wir, dass solche Ketten genau dann mischend sind, wenn sie aperiodisch sind. Wir haben bereits in Theorem 18.11 gesehen, dass solche Ketten konvergieren.

Theorem 18.27 (Mischende Markov-Ketten). *Sei* $\mathcal{X} = (X_t)_{t=0,1,2,...}$ *eine zeitlich homogene, positiv rekurrente, irreduzible Markov-Kette mit abzählbarem Zustandsraum* E*, gestartet in einer (nach Theorem 18.18 eindeutigen) invarianten Verteilung* ν *. Dann ist* $\mathcal{X}$ *genau dann mischend, wenn* $\mathcal{X}$ *aperiodisch ist.*

Beweis. '$\Rightarrow$': Angenommen, $\mathcal{X}$ ist periodisch mit Periode $d \geq 2$. (Die Periode ist für alle Zustände nach Proposition 18.6.2 dieselbe.) Ist $t > 0$ mit $d \not| \, t$, so gilt $p(x,x)^t = 0$. Also gilt $\liminf_{t\to\infty} p^t(x,x) = 0$. Weiter ist $\nu(x) > 0$ für alle $x \in E$ nach Proposition 18.6.3. Damit gilt

$$\liminf_{t\to\infty} \mathbf{P}_\nu(X_0 = x, X_t = x) = \nu(x)p^t(x,x) = 0 \neq \nu(x)^2 = \mathbf{P}_\nu(X_0 = x)^2.$$

Damit ist $\mathcal{X}$ nicht mischend.

'$\Leftarrow$': Sei nun $\mathcal{X}$ aperiodisch. Wir wissen aus Theorem 18.11, dass $p^t(x,y) \xrightarrow{t\to\infty} \nu(y)$ für alle $x, y \in E$ gilt. Angenommen, A hängt nur von den ersten s Zeitpunkten ab, d.h. $A, B \in \mathcal{F}_s$, so gilt wegen der Markov-Eigenschaft

$$\begin{aligned}
\mathbf{P}_\nu(\mathcal{X} \in A, \mathcal{X} \circ \tau^t \in B) &= \sum_{x,y\in E} \mathbf{P}_\nu(\mathcal{X} \in A, X_s = x, X_t = y, \mathcal{X} \circ \tau^t \in B) \\
&= \sum_{x,y\in E} \mathbf{P}_\nu(\mathcal{X} \in A, X_s = x) \cdot p^{t-s}(x,y) \cdot \mathbf{P}_y(\mathcal{X} \in B) \\
&\xrightarrow{t\to\infty} \sum_{x,y\in E} \mathbf{P}_\nu(\mathcal{X} \in A, X_s = x) \cdot \nu(y) \cdot \mathbf{P}_y(\mathcal{X} \in B) \\
&= \mathbf{P}_\nu(\mathcal{X} \in A) \cdot \mathbf{P}_\nu(\mathcal{X} \in B).
\end{aligned}$$

Der allgemeine Fall, d.h. $A \in \mathcal{F}$, folgt mittels eines Approximationsargumentes. $\square$

18.5 Stationäre Prozesse in stetiger Zeit

In diesem Abschnitt erweitern wir die bekannten Ergodensätze 18.15 und 18.17 und formulieren sie im zeitstetigen Fall; siehe Theorem 18.31. Die wichtigste Anwendung ergibt sich mit stationären Markov-Prozessen in stetiger Zeit, und insbesondere mit Markov-Ketten; siehe die Erweiterung des Markov-Konvergenzsatzes (Theorem 18.11) auf den zeitstetigen Fall; siehe Theorem 18.38.

Wir erinnern zunächst an den Begriff der Stationarität im zeit-stetigen Fall. Ein stochastischer Prozess $\mathcal{X} = (X_t)_{t\geq 0}$ heißt stationär, wenn $(X_{t+s})_{t\geq 0} \stackrel{d}{=} (X_t)_{t\geq 0}$ für alle $s \geq 0$ gilt. Anders ausgedrückt: Ist τ_s der Zeit-Shift um s, so gilt $\tau_s(\mathcal{X}) \stackrel{d}{=} \mathcal{X}$.

Definition 18.28 (Fluss). *Sei $I = \mathbb{R}_+$ oder $I = \mathbb{R}$. Eine Familie von Abbildungen $(\tau_s)_{s\in I}$ mit $\tau_s : \Omega \to \Omega$ heißt* Fluss *(oder* Halbgruppe*), falls $\tau_s(\tau_t(\omega)) = \tau_{s+t}(\omega)$. Der Fluss heißt* messbar*, wenn $(\omega, t) \mapsto \tau_t(\omega)$ eine $\mathcal{F} \times \mathcal{B}(I) - \mathcal{F}$-messbare Abbildung ist.*

In diesem Fall nennen wir

$$\mathcal{I} = \{A \subseteq \mathcal{F} : \tau_s^{-1}(A) = A \text{ für alle } s \in I\}$$

die σ-Algebra der $(\tau_s)_{s\in I}$-invarianten Ereignisse.

Beispiel 18.29 (Stationärer zeit-stetiger Prozess). In der Definition 18.28 betrachten wir wieder den Fall, dass $\Omega = E^I$ und $f : E \to \mathbb{R}$. Ist dann $(\tau_s)_{s\in I}$ ein Fluss, so ist

$$X_s := f(\tau_s(\omega)) \tag{18.12}$$

ein stationärer stochastischer Prozess. Weiter ist der Fluss $(\tau_s)_{s\in I}$ genau dann messbar, wenn für jedes messbare f der Prozess $(X_t)_{t\in I}$ messbar bezüglich $\mathcal{B}(E) \times \mathcal{B}(I) - \mathcal{B}(E)$ ist.

Um die Messbarkeit eines stochastischen Prozesses zu gewährleisten, stellen wir einen Zusammenhang zu dem bereits bekannten Begriff der progressiven Messbarkeit her; siehe Definition 14.31.

Lemma 18.30 (Progressiv messbar und messbar). *Sei $\mathcal{X} = (X_t)_{t\in I}$ ein E-wertiger stochastischer Prozess. Ist dieser progressiv messbar, so ist $(x,t) \mapsto X_t$ messbar bezüglich $\mathcal{B}(E) \times \mathcal{B}(I) - \mathcal{B}(E)$.*

Beweis. Die Behauptung ist klar, weil die Einbettung $\mathcal{B}(E) \times \mathcal{B}([0,t]) \to \mathcal{B}(E) \times \mathcal{B}([0,\infty)$ stetig, also messbar ist. □

Theorem 18.31 (Ergodensatz in stetiger Zeit). *Sei $I = [0,\infty)$ und $\mathcal{X} = (X_t)_{t\in I}$ ein stationärer stochastischer Prozess wie in* (18.12)*, $\mathcal{I}$ die invariante σ-Algebra und $X_0 \in L^p$ für $p \geq 1$. Dann gilt*

$$\frac{1}{t}\int_0^t X_s ds \xrightarrow{t\to\infty}_{fs,L^p} \mathbf{E}[X_0|\mathcal{I}]. \tag{18.13}$$

Beweis. Zunächst können wir annehmen, dass $X_t \geq 0$ ist. (Ansonsten schreiben wir $X_t = X_t^+ - X_t^-$ und wenden das Resultat auf X_t^+ und X_t^- an.) Schreibt man $Y_k = \int_{k-1}^k X_s ds$, so folgt die Konvergenz der linken Seite von (18.13) aus Theoremen 18.15 und 18.17, da mit der Jensen'schen Ungleichung gilt, dass

$$\mathbf{E}[Y_1^p] = \mathbf{E}\Big[\Big(\int_0^1 X_s ds\Big)^p\Big] \leq \int_0^1 \mathbf{E}\Big[X_s^p]ds = \mathbf{E}[X_0^p] < \infty.$$

Es geht also nur noch darum, zu zeigen, dass der Grenzwert Z fast sicher gleich $\mathbf{E}[X_0|\mathcal{I}]$ ist. (Dies ist nicht sofort klar, weil in den Ergodensätzen in diskreter Zeit auf der rechten Seite nur die invariante σ-Algebra des Ein-Schritt-Zeitshiftes steht.)

Auf der Menge von Maß 1, auf der (18.13) gilt, ist sicher $\int_{t_1}^{t_2} X_s ds < \infty$ für alle $t_1 < t_2$, da sonst der Grenzwert in nicht existieren würde. Daraus folgt, dass für jedes $r > 0$ (auf dieser Menge)

$$\frac{1}{t}\Big(\int_r^{r+t} X_s ds - \int_0^t X_s ds\Big) \xrightarrow{t\to\infty} 0.$$

Das bedeutet insbesondere

$$Z = \lim_{t\to\infty} \frac{1}{t}\int_r^{r+t} X_s ds = \lim_{t\to\infty} \frac{1}{t}\int_r^{r+t} X_s ds = Z \circ \tau_r.$$

Damit ist Z also messbar bezüglich $\mathcal{I}$ nach Lemma 18.3. Weiter gilt, wegen der L^1-Konvergenz für $A \in \mathcal{I}$

$$\mathbf{E}[X_s, \mathcal{X} \in A] = \mathbf{E}[X_s, \tau_s(\mathcal{X}) \in A] = (\tau_s)_*\mathbf{E}[X_0, \mathcal{X} \in A] = \mathbf{E}[X_0, \mathcal{X} \in A].$$

Daraus folgt (wieder mit $A \in \mathcal{I}$) aus der L^1-Konvergenz

$$\mathbf{E}[Z, \mathcal{X} \in A] = \lim_{t\to\infty} \mathbf{E}\Big[\frac{1}{t}\int_0^t X_s ds, \mathcal{X} \in A\Big] = \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbf{E}[X_s, \mathcal{X} \in A]ds = \mathbf{E}[X_0, \mathcal{X} \in A],$$

woraus $Z = \mathbf{E}[X_0|\mathcal{I}]$ folgt. □

Das wichtigste Beispiel stationärer Prozesse stellen Markov-Prozesse dar. Für diese stellt das nächste Resultat ein einfaches Werkzeug bereits, wie die Stationarität nachgewiesen werden kann.

Proposition 18.32 (Stationäre Verteilungen für Markov-Prozesse). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Markov-Prozess mit Zustandsraum E, Generator $G^{\mathcal{X}}$ und Domain $\mathcal{D}(G^{\mathcal{X}})$. Weiter sei $t \mapsto \mathbf{E}_x[f(X_t)]$ stetig in 0 für alle $x \in E$ und $f \in \mathcal{C}_b(E)$, $x \mapsto \mathbf{E}_x[f(X_t)] \in \mathcal{C}_b(E)$ für $f \in \mathcal{C}_b(E)$ und $G^{\mathcal{X}} f \in \mathcal{C}_b(E)$ für $f \in \mathcal{D}(G^{\mathcal{X}})$. Für $\nu \in \mathcal{P}(E)$ sind äquivalent:*

1. *ν ist eine stationäre Verteilung, das heißt $\mathbf{E}_\nu[f(X_t)] = \mathbf{E}_\nu[f(X_0)]$ für alle $f \in \mathcal{C}_b(E)$.*
2. *Es gibt eine separierende Funktionenklasse $\Pi \subseteq \mathcal{D}(G^{\mathcal{X}})$, so dass $\mathbf{E}_\nu[(G^{\mathcal{X}} f)(\mathcal{X}_0)] = 0$ für alle $f \in \Pi$.*

Beweis. '1. $\Rightarrow$ 2.': Sei $f \in \mathcal{D}(G^{\mathcal{X}})$. Nach Theorem 16.30 ist

$$\Big(f(X_t) - f(X_0) - \int_0^t (G^{\mathcal{X}} f)(X_s)ds\Big)_{t\geq 0} \tag{18.14}$$

ein Martingal. Aus der Stationarität von ν folgt, dass $\mathbf{E}_\nu[g(X_t)] = \mathbf{E}_\nu[g(X_0)]$ für alle $g \in \mathcal{B}(E)$. Daraus folgt

$$0 = \int_0^t \mathbf{E}_\nu[(G^{\mathcal{X}} f)(X_s)]ds = t \cdot \mathbf{E}_\nu[(G^{\mathcal{X}} f)(X_0)]ds$$

für alle $t \geq 0$. Dies ist jedoch nur möglich, falls $\mathbf{E}_\nu[(G^{\mathcal{X}} f)(X_0)] = 0$. Die Aussage folgt nun, da $\mathcal{D}(G)$ dicht in $\mathcal{C}_b(E)$ ist; siehe Korollar 16.29.

'2. $\Rightarrow$ 1.' Für $f \in \Pi$ gilt für alle $t \in I$, da (18.14) ein Martingal ist,

$$\mathbf{E}_\nu[f(X_t)] = \mathbf{E}_\nu[f(X_0)].$$

Da Π separierend ist, folgt $X_t \stackrel{d}{=} X_0 \sim \nu$. □

Definition 18.33 (Markov-Kette in stetiger Zeit). *Wir betrachten das Beispiel 16.33 und die darin verwendete Notation. Ist hier E höchstens abzählbar, so heißt $\mathcal{X}$ eine* Markov-Kette in stetiger Zeit. *Der Prozess $\mathcal{Y} = (Y_k)_{k=0,1,2,\dots}$ heißt die* eingebettete Markov-Kette. *Es heißt $\lambda(x)$ die Sprungrate von x und wir definieren die Übergangsmatrix $P = (p(x,y))_{x,y\in E}$ der eingebetteten Markov-Kette als $p(x,y) := \mu(x,dy)$. Weiter ist $p_t(x,y) := \mu_t(x,dy)$ die Wahrscheinlichkeit der zeit-stetigen Markov-Kette, bei Start in x zur Zeit t in y zu sein.*

Bemerkung 18.34 (Unbeschränkte Sprungsraten). Wir bemerken, dass in Beispiel 16.33 die Sprungrate λ beschränkt ist. Die Konstruktion der Verbindung von $\mathcal{X}$ und $\mathcal{Y}$ funktioniert jedoch auch bei unbeschränktem λ, so lange $\tau := \sum_{j=0}^\infty \frac{T_j}{\lambda(Y_j)} = \infty$ fast sicher gilt (siehe (16.8)). Andernfalls wäre der Prozess $\mathcal{X}$ in endlicher Zeit unendlich oft gesprungen, und die Definition des Prozesses nach τ wäre nicht möglich.

Proposition 18.35 (Stationäre Verteilung von Markov-Ketten). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Markov-Kette in stetiger Zeit und $\mathcal{Y}$ die eingebettete Markov-Kette, sowie $\nu \in \mathcal{P}(E)$. Dann ist ν genau dann stationär für $\mathcal{X}$, wenn $\widetilde{\nu}$ mit $\widetilde{\nu}(x) = \lambda(x)\nu(x)/C$ und $C = \sum_{y\in E} \lambda(y)\nu(y)$ stationär für $\mathcal{Y}$ ist.*

Beweis. Wir wissen aus Proposition 18.32 und (16.9), dass ν genau dann stationär für $\mathcal{X}$ ist, wenn

$$\begin{aligned} 0 &= \sum_{x\in E} \nu(x)(G^{\mathcal{X}} f)(x) = \sum_{x,y\in E} \nu(x)\lambda(x)p(x,y)(f(y)-f(x)) \\ &= C\sum_{x,y} \widetilde{\nu}(x)p(x,y)(f(y)-f(x)) = C(\mathbf{E}_{\widetilde{\nu}}[f(Y_1)] - \mathbf{E}_{\widetilde{\nu}}[f(Y_0)]). \end{aligned}$$

Daraus folgt die Aussage. □

Lemma 18.36 (Treffwahrscheinlichkeiten). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Markov-Kette in stetiger Zeit und $x, y \in E$. Dann ist $p_t(x,x) > 0$ für alle $t \geq 0$ und $x \in E$. Für $x \neq y$ ist entweder $p_t(x,y) > 0$ für alle $t > 0$, oder $p_t(x,y) = 0$ für $t > 0$.*

Beweis. Wir setzen $\tau_k := \sum_{j=0}^k \frac{T_j}{\lambda(Y_j)}$ als k-te Sprungzeit von $\mathcal{X}$. Es gilt für $t > 0$

$$p_t(x,x) = \mathbf{P}_x(X_t = x) \geq \mathbf{P}_x(X_t = x, \tau_1 > t) = e^{-\lambda(x)t} > 0.$$

Weiter sei $\mathcal{Y} = (Y_k)_{k=0,1,2,\dots}$ die eingebettete Markov-Kette mit Übergangsmatrix P und $r_{xy} := \mathbf{P}_x[T_y < \infty]$. Gilt $r_{xy} > 0$, so gibt es ein $n > 0$ mit $p^n(x,y) > 0$. Daraus folgt für jedes $t > 0$

$$p_t(x,y) = \mathbf{P}_x(X_t = y) \geq \mathbf{P}_x(X_t = y, \tau_{n-1} \leq t < \tau_n) = p^n(x,y) \cdot \mathbf{P}_x(\tau_{n-1} \leq t < \tau_n) > 0$$

Ist andersherum $r_{xy} = 0$, so muss $p_t(x,y) = 0$ für alle $t > 0$ gelten. □

Definition 18.37 (Rekurrenz, Transienz). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Markov-Kette in stetiger Zeit. Wir setzen*

$$R_x := \inf\{t \geq 0 : X_t \neq x\} \qquad \textit{und} \qquad T_x := \inf\{t > R_x : X_t = x\}.$$

Dann definieren wir die Begriffe rekurrent, positiv rekurrent, null-rekurrent, transient, irreduzibel *genau wie in Definition 18.4 im zeit-diskreten Fall.*

Nach dem eben gezeigtem Lemma macht für Markov-Ketten in stetiger Zeit der Begriff der Aperiodizität keinen Sinn. Deswegen fassen wir die analogen Aussagen zu den Theorem 18.11 und Theorem 18.18 für Markov-Ketten in stetiger Zeit zusammen.

Theorem 18.38 (Markov-Ketten-Konvergenzsatz in stetiger Zeit). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine zeitlich homogene, irreduzible Markov-Kette in stetiger Zeit mit abzählbarem Zustandsraum E. Dann gilt genau eine der beiden Aussagen*

1. *$\mathcal{X}$ ist positiv rekurrent und es gibt genau eine invariante Verteilung ν von $\mathcal{X}$. Ist $X_0 \sim \mu$ mit $\mu \in \mathcal{P}(E)$ beliebig, so ist*

$$||(X_t)_*\mathbf{P} - \nu|| \xrightarrow{t\to\infty} 0. \tag{18.15}$$

   In diesem Fall ist $\mathcal{X}$ mischend (also auch ergodisch) und die einzige stationäre Verteilung ist $\nu \in \mathcal{P}(E)$ mit

$$\nu(x) = \frac{1}{\lambda(x)\cdot \mathbf{E}_x[T_x]}.$$

2. *Es gibt keine invariante Verteilung von $\mathcal{X}$ und es gilt*
$$p_t(x,y) \xrightarrow{t\to\infty} 0 \text{ für alle } x,y \in E. \tag{18.16}$$
In diesem Fall ist $\mathcal{X}$ entweder transient oder null-rekurrent.

Beweis. Nach Lemma 18.36 ist für jedes $h > 0$ die Markov-Kette $(X_{nh})_{n=0,1,2,...}$ irreduzibel und aperiodisch. Wir beginnen mit der Klassifizierung in die beiden Fälle.

Angenommen, es gibt ein $h > 0$, für das die Markov-Kette $(X_{nh})_{n=0,1,2,...}$ positiv rekurrent ist. Nach Theorem 18.11 gibt es dann eine einzige invariante Verteilung ν. Weiter ist dann $(X_{nh'})_{n=0,1,2,...}$ positiv rekurrent für jedes $h' = 2^{-m}h$, $m = 0,1,2,...$, und zwar – wegen der Eindeutigkeit – mit derselben invarianten Verteilung. Sei $t > 0$. Approximiert man t mittels $\lceil 2^m t/h\rceil 2^{-m}h$ von oben, so folgt wegen der Rechtsstetigkeit von $\mathcal{X}$, dass mit $X_0 \sim \nu$ auch $X_t \sim \nu$, also ν auch invariant für $\mathcal{X}$ ist. Es gilt nun

$$\begin{aligned}
||(X_t)_*\mathbf{P} - \nu|| &\le \sum_{z\in E} |\mathbf{P}(X_t = z) - \nu(z)| = \sum_{z\in E}\Big|\sum_{x\in E}\mu(x)p_t(x,z) - \nu(z)\Big| \\
&\le \sum_{x\in E}\mu(x)\sum_{z\in E}|p_t(x,z) - \nu(z)| \\
&\le \sum_{x\in E}\mu(x)\sum_{y,z\in E}|p_{[t/h]h}(x,y) - \nu(y)|p_{t-[t/h]h}(y,z) \\
&= \sum_{x\in E}\mu(x)\sum_{y\in E}|p_{[t/h]h}(x,y) - \nu(y)| \xrightarrow{t\to\infty} 0
\end{aligned}$$

nach majorisierter Konvergenz mit Theorem 18.11. In diesem Fall folgt genau wie in Theorem 18.27, dass $\mathcal{X}$ mischend ist.

Angenommen, $(X_{nh})_{n=0,1,2,...}$ ist für alle $h > 0$ entweder null-rekurrent oder transient. Für $x, z \in E$ ist dann, wieder mit Theorem 18.11

$$\begin{aligned}
\lim_{t\to\infty} p_t(x,z) &= \lim_{t\to\infty}\sum_{y\in E} p_{t-[t/h]h}(x,y)p_{[t/h]h}(y,z) \\
&\le \lim_{t\to\infty} p_{[t/h]h}(x,z) + \sum_{y\neq x} p_{t-[t/h]h}(x,y) \le \sum_{y\neq x} p_h(x,y) \xrightarrow{h\to 0} 0.
\end{aligned}$$

Wir kommen nun zur Berechnung der stationären Verteilung von $\mathcal{X}$ im Falle positiver Rekurrenz. Sei $T_x^h := \inf\{n > 0 : X_{nh} = x\}$. Wir berechnen, wegen der Rechtsstetigkeit von $\mathcal{X}$,

$$\begin{aligned}
\mathbf{E}_x[T_x] &= \sum_{y\in E} p(x,y)\mathbf{E}_y[T_x] = \lim_{h\to 0}\sum_{y\in E} p(x,y)\mathbf{E}_y[T_x^h]\cdot h = \lim_{h\to 0}\mathbf{E}_x[T_x^h|X_h \neq x]\cdot h \\
&= \lim_{h\to 0}\frac{\mathbf{E}_x[T_x^h]\cdot h - \mathbf{E}_x[T_x^h, X_h = x]\cdot h}{\mathbf{P}(X_h \neq x)} = \lim_{h\to 0}\frac{\mathbf{E}_x[T_x^h]\cdot h - h}{1 - e^{-\lambda(x)h}} = \lim_{h\to 0}\frac{\mathbf{E}_x[T_x^h]}{\lambda(x)}
\end{aligned}$$

Ist nun $(X_{nh})_{n=0,1,2,...}$ für ein $h > 0$ positiv rekurrent, so ist nach Theorem 18.18 die invariante Verteilung ν von $\mathcal{X}$ durch

$$\nu(x) = \frac{1}{\mathbf{E}_x[T_x^h]} = \lim_{m\to\infty}\frac{1}{\mathbf{E}_x[T_x^{2^{-m}h}]} = \lim_{h\to 0}\frac{1}{\mathbf{E}_x[T_x^h]} = \frac{1}{\lambda(x)\cdot\mathbf{E}_x[T_x]}.$$

gegeben. □

Wir beenden diesen Abschnitt mit einem stationären Markov-Prozess, der aber keine Markov-Kette ist.

Beispiel 18.39 (Ornstein-Uhlenbeck-Prozess). Wir lernen nun noch einen speziellen Prozess in stetiger Zeit kennen, der, sowohl ein Gauss-Prozess (siehe Definition 14.15), als auch ein starker Markov-Prozess und stationär ist. Hierzu sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung, $\mu > 0$, und $\mathcal{Y} = (Y_t)_{t\geq 0}$ gegeben durch

$$Y_t = e^{-\mu t} X_{\int_0^t e^{2\mu s} ds} = e^{-\mu t} X_{\frac{1}{2\mu}(e^{2\mu t}-1)}.$$

Klar ist, dass $\mathcal{Y}$ ein Gauss'scher Prozess ist, und zwar mit Covarianz-Funktion

$$\mathbf{COV}(Y_s, Y_t) = e^{-\mu(s+t)} \mathbf{COV}(X_{\frac{1}{2\mu}(e^{2\mu t}-1)}, X_{\frac{1}{2\mu}(e^{2\mu s}-1)}) = \tfrac{1}{2\mu} e^{-\mu(s+t)} (e^{2\mu(s\wedge t)} - 1).$$

Gegeben $Y_0 = y$, folgt daraus, dass $Y_t \sim N(e^{-\mu t} y, \frac{1}{2\mu}(1-e^{-2\mu t}))$. Wir berechnen für $s \leq t \leq u$

$$\begin{aligned} 4\mu^2 \cdot \mathbf{COV}(Y_s, Y_u) \cdot \mathbf{V}(Y_t) &= (e^{-\mu(u-s)} - e^{-\mu(u+s)}) \cdot (1 - e^{-2\mu t}) \\ &= (e^{-\mu(t-s)} - e^{-\mu(t+s)})(e^{-\mu(u-t)} - e^{-\mu(u+t)}) \\ &= 4 \cdot \mathbf{COV}(Y_s, Y_t) \cdot \mathbf{COV}(Y_t, Y_u). \end{aligned}$$

Nach Theorem 16.5 ist $\mathcal{Y}$ also ein Markov-Prozess. Aus Theorem 16.12 können wir außerdem ablesen, dass $\mathcal{Y}$ stark Markov ist.

Damit ist der Generator dieses Markov-Prozesses für $f \in \mathcal{C}_b^2(\mathbb{R})$, mit Hilfe von (16.5), für ein t' mit $|t'-1| \leq |1 - e^{-\mu t}|$

$$\begin{aligned} (G^{\mathcal{Y}} f)(y) &= \lim_{t\to 0} \tfrac{1}{t} \big(\mathbf{E}_y[f(Y_t) - f(y)] \big) \\ &= \lim_{t\to 0} \tfrac{1}{t} \big(\mathbf{E}_y[f(e^{-\mu t} X_{\int_0^t e^{2\mu s} ds}) - f(y)] \big) \\ &= \lim_{t\to 0} \tfrac{1}{t} \big(\mathbf{E}_y[f(X_{\int_0^t e^{2\mu s} ds}) - f(y) \\ &\qquad + f'(X_{\int_0^t e^{2\mu s} ds})(e^{-\mu t} - 1) X_{\int_0^t e^{2\mu s} ds} + \tfrac{1}{2} f''(X_{\int_0^t e^{2\mu s} ds})(t'-1)^2 X^2_{\int_0^t e^{2\mu s} ds}] \big) \\ &= \tfrac{1}{2} f''(y) - \mu y f'(y) \end{aligned}$$

Da $Y_t \sim N(e^{-\mu t} y, \frac{1}{2\mu}(1-e^{-2\mu t}))$, folgt schon, dass $Y_t \xRightarrow{t\to\infty} \nu = N(0, \frac{1}{2\mu})$. Dies legt nahe, dass ν eine stationäre Verteilung ist. Dies kann man mit Hilfe von Proposition (18.32) nachrechnen. Es gilt nämlich, für $f \in \mathcal{C}_b^2(\mathbb{R})$, mittels partieller Integration,

$$\begin{aligned} \mathbf{E}_\nu[(G^{\mathcal{Y}} f)(Y_0)] &= \sqrt{\tfrac{\mu}{\pi}} \int e^{-\mu x^2} (\tfrac{1}{2} f''(x) - \mu x f'(x)) dx \\ &= \sqrt{\tfrac{\mu}{\pi}} \Big(- \int \tfrac{1}{2} f'(x) \frac{d}{dx} e^{-\mu x^2} dx - \mu \int x e^{-\mu x^2} f'(x) dx \Big) = 0, \end{aligned}$$

woraus die Stationarität von ν folgt.

Teil IV
Stochastische Integration

In Definition 15.14 haben wir bereits das diskrete stochastische Integral kennen gelernt. Für $I = \{0, 1, 2, ...\}$ und eine Filtration $(\mathcal{F}_t)_{t\in I}$ war $\mathcal{H} = (H_t)_{t\in I}$ ein reellwertiger und previsibeler stochastischer Prozess (d.h. H_t ist $\mathcal{F}_{t-1}$-messbar) und $\mathcal{X} = (X_t)_{t\in I}$ war reellwertig und adaptiert. Das stochastische Integral $\mathcal{H}\cdot\mathcal{X} = ((\mathcal{H}\cdot\mathcal{X})_t)_{t\in I}$ hatten wir dann definiert als den adaptierten Prozess

$$(\mathcal{H}\cdot\mathcal{X})_t := \sum_{s=1}^{t} H_s(X_s - X_{s-1}).$$

Intuitiv bedeutet das, dass das stochastische Integral $\mathcal{H}\cdot\mathcal{X}$ den Prozess $\mathcal{H}$ relativ zu den Änderungen von $\mathcal{X}$ misst. Weiter haben wir bereits wichtige Eigenschaften dieses Integrals kennengelernt. Etwa ist $\mathcal{H}\cdot\mathcal{X}$ ein Martingal, falls $\mathcal{X}$ eines ist (siehe Proposition 15.15 und Tabelle 19.2, die die Martingal-Eigenschaften von $\mathcal{H}\cdot\mathcal{X}$ zusammenfasst). Weiter fällt auf, dass wir die quadratische Variation sowohl für zeit-diskrete stochastische Integrale (Beispiel 15.16) als auch für die Brown'sche Bewegung definiert haben (siehe Abschnitt 17.1). Auch hier sind Querverbindungen zu erwarten. Ziel dieses Kapitels ist es, die Theorie der stochastischen Integrale auf Prozesse in stetiger Zeit fortzusetzen. Hierbei soll der zeit-diskrete Fall als Vorlage dienen.

Im Folgenden sei stets $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, auf dem alle stochastischen Prozesse definiert sind, und $(\mathcal{F}_t)_{t\geq 0}$ eine Filtration. Soweit nicht anders genannt, sind alle stochastischen Prozesse reellwertig.

19 Einführung

Integrale (basierend etwa auf σ-endlichen Maßen auf $\mathcal{B}(\mathbb{R})$) wurden bereits in der Maßtheorie ausführlich behandelt. Das Integral entsprach dann später in der Wahrscheinlichkeitstheorie dem Erwartungswert einer Zufallsvariable; siehe Bemerkung 7.1.5. Anders ist es nun bei der stochastischen Integration. Hier ist gemeint, dass das Integral selbst eine Zufallsgröße darstellt.

Wir verwenden im Folgenden

$$\mathcal{H}\cdot\mathcal{X} = \int H_s dX_s.$$

als Schreibweise für stochastische Integrale. Dabei wird der Prozess $\mathcal{H} = (H_t)_{t\geq 0}$ als Integrand und $\mathcal{X} = (X_t)_{t\geq 0}$ als Integrator bezeichnet. Die Klasse der möglichen Integratoren in der allgemeinen Theorie der stochastischen Integration sind die Semi-Martingale (das ist die Summe aus einem lokalen Martingal – siehe Definition 19.23 – und einem Prozess von endlicher Variation). Genau wie beim diskreten stochastischen Integral sind die Integranden previsible Prozesse, was insbesondere die adaptierten, linksstetigen Prozesse mit einschließt.

Nachdem wir in Abschnitt 19.1 einige Grundlagen gelegt haben, führen wir im Abschnitt 19.2 stochastischen Integrale zunächst für Prozesse mit endlicher Variation ein, und anschließend – nach einer Einführung in lokale Martingale in Abschnitt 19.3 – für lokale Martingale mit stetigen Pfaden in Abschnitt 19.4. In Tabelle 19.1 sind die wichtigsten Schritte der Konstruktion stochastischer Integrale zusammen gestellt. Die allgemeine Theorie mit möglicherweise unstetigen Semimartingalen als Integratoren werden wir nur am Rande später beleuchten.

19.1 Grundlegendes

Genau wie in der Integrationstheorie, der wir in der Maßtheorie begegnet sind, betrachten wir ale erstes stochastische Integrale von einfachen Integranden. Der Prozess $\mathcal{I}$ aus (19.2) wird dann als Definition des stochastischen Integrals dienen. Nachdem wir wichtige Martingal-Eigenschaften dieses Prozesses in Proposition 19.2 gezeigt haben, folgt die Einführung von previsiblen Prozessen.

Definition 19.1 (Einfache previsible Prozesse). *Ein* einfacher, previsibler Prozess $\mathcal{H} = (H_t)_{t\geq 0}$ *ist von der Form*

$$H_t = \sum_{i=1}^n G_i 1_{(T_i,T_{i+1}]}(t) \tag{19.1}$$

für Stoppzeiten $T_1 \leq ... \leq T_{n+1}$ *und* $G_1, ..., G_n$ *so, dass* G_i *nach* $\mathcal{F}_{T_i}$*-messbar ist,* $i = 1, ..., n$*. Die Menge aller einfachen, previsiblen Prozesse bezeichnen wir mit* $\mathbb{S}$*.*

Folgendes Resultat ist zentral für die allgemeine Definition des stochastischen Integrals. Wichtig ist es zu bemerken, dass hier keine Voraussetzungen an das Martingal $\mathcal{X}$ gestellt werden. Dies legt nahe, das stochastische Integral bezüglich einfacher previsibler Funktionen von der Form (19.1) durch (19.2) zu definieren und den so erhaltenen Integralbegriff anschließend geeignet zu erweitern.

Proposition 19.2 (Martingal-Eigenschaft des stochastischen Integrals). *Sei* $\mathcal{X} = (X_t)_{t\geq 0}$ *ein Martingal und* $\mathcal{H} = (H_t)_{t\geq 0} \in \mathbb{S}$ *ein einfacher, previsibler Prozess wie in* (19.1) *mit* $G_1, ..., G_n \in \mathcal{L}^1$*, sowie* $\mathcal{I} = (I_t)_{t\geq 0}$*, gegeben durch*

$$I_t = \sum_{i=1}^n G_i (X_{T_{i+1}\wedge t} - X_{T_i \wedge t}). \tag{19.2}$$

Es gilt:

1. *Der Prozess* $\mathcal{I}$ *ist ein Martingal.*
2. *Falls außerdem* $\mathcal{H} \leq \ell$ *beschränkt ist, so gilt*

$$\sup_{t\geq 0} ||I_t||_2 \leq \ell \cdot \sup_{t\geq 0} ||X_t||_2.$$

Beweis. Zunächst eine kurze Vorüberlegung, die auf dem Optional Sampling Theorem (Theorem 15.21) beruht. Sind S, T zwei beschränkte Stoppzeiten (und nicht notwendigerweise $S \leq T$) und $\mathcal{X} = (X_t)_{t\geq 0}$ ein Martingal, so gilt

$$\begin{aligned}\mathbf{E}[X_T|\mathcal{F}_S] &= \mathbf{E}[(1_{S\leq T} + 1_{S>T})X_T|\mathcal{F}_S] = 1_{S\leq T}\mathbf{E}[X_{S\vee T}|\mathcal{F}_S] + 1_{S>T}X_T \\ &= 1_{S\leq T}X_S + 1_{S>T}X_T = X_{S\wedge T}.\end{aligned}$$

$\mathcal{H} \cdot \mathcal{X}$	$\mathcal{H} \in \mathbb{S}$	$\mathcal{H}$ hat Pfade in $\mathcal{G}_{\mathbb{R}}([0,\infty))$	$\mathcal{H}$ progressiv	$\mathcal{X}$ hat lokal beschränkte Variation	$\mathcal{X} \in \mathcal{M}^2$ (stetige Pfade)	$\mathcal{X}$ lokales Martingal, stetige Pfaden	definierende Eigenschaft
Definition 19.1, Lemma 19.12	•			•			$(\mathcal{H} \cdot \mathcal{X})_t = \sum G_i(X_{T_{i+1}\wedge t} - X_{T_i \wedge t})$
Definition 19.7.2			•	•			Lebesgue-Stieltjes Integral
Definition 19.19	•				•		$(\mathcal{H} \cdot \mathcal{X})_t = \sum G_i(X_{T_{i+1}\wedge t} - X_{T_i \wedge t})$
Proposition19.20, Definition 19.21		•			•		$\mathcal{M}^2$-Grenzwert
Definition 19.34, Lemma 19.35	•					•	$(\mathcal{H} \cdot \mathcal{X})_t = \sum G_i(X_{T_{i+1}\wedge t} - X_{T_i \wedge t})$
Theorem 19.36			•			•	$[\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}]$

Tabelle 19.1: Das stochastische Integral $\mathcal{H} \cdot \mathcal{X}$ wird für verschiedene Klassen von Prozessen $\mathcal{H}$ und $\mathcal{X}$ eingeführt. In chronologischer Reihenfolge sind hier die entsprechenden Definitionen und Resultate aufgelistet. Es sei bemerkt, dass die definierende Eigenschaft des stochastischen Integrals für $\mathcal{H} \in \mathbb{S}$ (einfache, previsible Prozesse) immer dieselbe ist.

	$\mathcal{X}$ Martingal	$\mathcal{X}$ Martingal, beschränkte Variation, $\sup_{t\geq 0}\mathbf{E}[X_t^2]<\infty$	$\mathcal{X}\in\mathcal{M}^2$ (stetige Pfade)	$\mathcal{X}$ lokales Martingal, stetige Pfade
$\mathcal{H}\in\mathbb{S}, G_1,...,G_n\in L^1$	$\mathcal{H}\cdot\mathcal{X}$ Martingal, Proposition 19.2			
$\mathcal{H}$ beschränkt, Pfade in $\mathcal{G}_{\mathbb{R}}([0,\infty))$		$\mathcal{H}\cdot\mathcal{X}$ Martingal, Proposition 19.15	$\mathcal{H}\cdot\mathcal{X}\in\mathcal{M}^2$, Prop. 19.20	
$\mathcal{H}$ progressiv, $\mathbf{E}[(\mathcal{H}^2\cdot[\mathcal{X}])_t]<\infty$				$\mathcal{H}\cdot\mathcal{X}$ lokales Martingal, Theorem 19.36

Tabelle 19.2: Die Martingal-Eigenschaften des stochastischen Integrals $\mathcal{H}\cdot\mathcal{X}$ sind hier zusammen gefasst.

Nun also zum Beweis von 1. Wegen der Linearität der bedingten Erwartung genügt es, die Behauptung im Falle $H_t = G \cdot 1_{(S,T]}(t)$, also

$$I_t = G(X_{T\wedge t} - X_{S\wedge t})$$

für $\mathcal{F}_S$-messbares G und $S \leq T$ zu zeigen. Für $s \leq t$ schreiben wir

$$\begin{aligned}
\mathbf{E}[I_t|\mathcal{F}_s] &= \mathbf{E}[(1_{T,S>s} + 1_{S\leq s<T} + 1_{T\leq s})G(X_{T\wedge t} - X_{S\wedge t})|\mathcal{F}_s] \\
&= \mathbf{E}[1_{T,S>s}G \cdot \mathbf{E}[X_{T\wedge t} - X_{S\wedge t}|\mathcal{F}_{S\vee s}]|\mathcal{F}_s] \\
&\qquad + 1_{S\leq s<T}G \cdot \mathbf{E}[X_{T\wedge t} - X_{S\wedge t}|\mathcal{F}_s] \\
&\qquad + \mathbf{E}[1_{T\leq s}G(X_{T\wedge s} - X_{S\wedge s})|\mathcal{F}_s] \\
&= \mathbf{E}[1_{T,S>s}G \cdot (X_{T\wedge t\wedge(S\vee s)} - X_{S\wedge t\wedge(S\vee s)})|\mathcal{F}_s] \\
&\qquad + 1_{S\leq s<T}G \cdot (X_{T\wedge s} - X_{S\wedge s}) \\
&\qquad + 1_{T\leq s}G \cdot (X_{T\wedge s} - X_{S\wedge s}) \\
&= (1_{S\leq s<T} + 1_{T\leq s})G \cdot (X_{T\wedge s} - X_{S\wedge s}) = G \cdot (X_{T\wedge s} - X_{S\wedge s}) = I_s,
\end{aligned}$$

wobei wir etwa verwendet haben, dass $1_{S\leq s<T}G$ nach $\mathcal{F}_s$-messbar ist. Damit ist die Behauptung gezeigt.
2. Ist $\mathcal{H} \leq \ell$, so schreiben wir

$$\begin{aligned}
\mathbf{E}[I_t^2] &\leq \mathbf{E}\Big[\sum_{i=1}^n \sum_{j=1}^n G_i G_j (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})(X_{T_{j+1}\wedge t} - X_{T_j\wedge t})\Big] \\
&= \mathbf{E}\Big[\sum_{i=1}^n G_i^2 (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})^2\Big] \\
&\qquad\qquad + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{E}\big[G_i G_j (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})\mathbf{E}[X_{T_{j+1}\wedge t} - X_{T_j\wedge t}|\mathcal{F}_{T_j\wedge t}]\big] \\
&= \mathbf{E}\Big[\sum_{i=1}^n G_i^2 (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})^2\Big] \\
&\leq \ell^2 \sum_{i=1}^n \mathbf{E}\big[(X_{T_{i+1}} - X_{T_i})^2\big] \\
&= \ell^2 \sum_{i=1}^n \mathbf{E}\big[X_{T_{i+1}}^2 - 2\mathbf{E}[X_{T_{i+1}}|\mathcal{F}_{T_i}]X_{T_i} + X_{T_i}^2\big] \\
&= \ell^2 \sum_{i=1}^n \mathbf{E}\big[X_{T_{i+1}}^2 - X_{T_i}^2\big] = \ell^2 \cdot \sup_{t\geq 0} \mathbf{E}[X_t^2].
\end{aligned}$$

□

Der soeben definierte Prozess $\mathcal{I}$ erbt die Stetigkeitseigenschaften des Prozessen $\mathcal{X}$.

Lemma 19.3 (Stetigkeits-Eigenschaften des stochastischen Integrals). *Sei $\mathcal{X}$ ein stochastischer Prozess und $\mathcal{H}$ ein einfacher previsibler Prozess. Dann gilt für den Prozess $\mathcal{I}$ aus Proposition 19.2:*

1. *Hat $\mathcal{X}$ stetige Pfade, dann auch $\mathcal{I}$.*
2. *Hat $\mathcal{X}$ rechtsstetige Pfade, dann auch $\mathcal{I}$.*

Beweis. Klar. □

Die einfachen previsiblen Integranden aus der letzten Proposition bilden noch keine besonders große Klasse von stochastischen Prozessen. Deswegen werden wir durch Approximation von allgemeineren Prozessen durch einfache vorhersagbare stochastische Prozesse versuchen, den Integralbegriff zu erweitern. Dies führt zu den previsiblen stochastischen Prozessen, die wir nun einführen.

Definition 19.4 (Linksstetige Prozesse und previsible σ-Algebra). *1. Wir bezeichnen mit $\mathcal{G}_E([0,\infty))$ die Menge der linksstetigen Funktionen $f : [0,\infty) \to E$ mit rechtsseitigen Grenzwerten.*

2. Die previsible σ-Algebra $\mathcal{V}$ *auf $[0,\infty)\times\Omega$ ist die kleinste σ-Algebra, bezüglich der alle Prozesse mit Pfaden in $\mathcal{G}_E([0,\infty))$ messbar sind. (Das bedeutet, dass $(t,\omega)\mapsto X_t(\omega)$ für $\mathcal{X}=(X_t)_{t\geq 0}$ mit Pfaden in $\mathcal{G}_E([0,\infty))$ messbar bezüglich $\mathcal{V}/\mathcal{B}(E)$ ist.) Ein adaptierter Prozess $\mathcal{X}$ heißt previsibel, wenn er bezüglich $\mathcal{V}$ messbar ist.*

Proposition 19.5 (Previsible Prozesse). *1. Jeder adaptierte Prozess mit Pfaden in $\mathcal{G}_E([0,\infty))$ ist previsibel. (Insbesondere sind Prozesse mit stetigen Pfaden previsibel.)*

2. Jeder previsible Prozess ist progressiv messbar.

3. Hat $\mathcal{X}$ Pfade in $\mathcal{D}_E([0,\infty))$, so ist $\mathcal{X}_- := (X_{t-})_{t\geq 0}$ previsibel.

Beweis. 1. Klar nach Definition der previsiblen σ-Algebra.
2. Genau wie im Beweis von Lemma 14.32 folgert man, dass Prozesse mit Pfaden in $\mathcal{G}_E([0,\infty))$ progressiv messbar sind. Da $\mathcal{V}$ von den Prozessen mit Pfaden in $\mathcal{G}_E([0,\infty))$ erzeugt wird und diese Prozesse progressiv messbar sind, folgt die Aussage.
3. Klar ist, dass $\mathcal{X}_-$ Pfade in $\mathcal{G}_E([0,\infty))$ hat. Also folgt die Aussage aus 1. □

Die Klasse der linksstetigen stochastischen Prozessen ist recht groß. Wir zeigen nun, dass sich solche Prozesse durch einfache vorhersagbare Prozesse sehr gut approximieren lassen.

Lemma 19.6 (Approximationen von Prozessen mit Pfaden in $\mathcal{G}_E([0,\infty))$). *Sei $\mathcal{Y}=(Y_t)_{t\geq 0}$ ein beschränkter Prozess mit Pfaden in $\mathcal{G}_E([0,\infty))$ und $Y_0=0$. Dann gibt es für jedes $t>0$ eine Folge $\mathcal{Y}^1=(Y_t^1)_{t\geq 0}, \mathcal{Y}^2=(Y_t^2)_{t\geq 0},\ldots\in\mathbb{S}$ von einfachen, previsiblen Prozessen und $\varepsilon_1,\varepsilon_2,\ldots>0$ mit $\varepsilon_n\downarrow 0$ und*

$$\sup_{0\leq s\leq t}|Y_s - Y_s^n|\leq\varepsilon_n$$

fast sicher.

Beweis. Wir bemerken, dass $(Y_{t+})_{t\geq 0}$ Pfade in $\mathcal{D}_E([0,\infty))$ hat. Sei $\varepsilon>0$. Wir definieren rekursiv $T_0^\varepsilon=0$,

$$T_{n+1}^\varepsilon := \inf\{t>T_n^\varepsilon : |Y_t - Y_{T_n^\varepsilon +}|>\varepsilon\}.$$

Dann gilt $T_{n+1}^\varepsilon\uparrow\infty$. Weiter setzen wir

$$Y_t^{n,\varepsilon} := \sum_{i=1}^n Y_{T_i^\varepsilon +}\cdot 1_{(T_i^\varepsilon\wedge n, T_{i+1}^\varepsilon\wedge n,]}(t),$$

so dass $\sup_{0\le s\le t\wedge n}|Y_s - Y_s^{n,\varepsilon}| \le \varepsilon$ nach Definition fast sicher gilt. Die Behauptung folgt nun, wenn wir eine Folge $\varepsilon_n \downarrow 0$ betrachten sowie $\mathcal{Y}^n = (Y_t^{n,\varepsilon_n})_{t\ge 0}$. □

19.2 (Stochastische) Stieltjes-Integrale

Die Konstruktion von stochastischen Integralen erfolgt für verschienene Fälle zunächst getrennt. Als erstes konstruieren wir das stochastische Integral im Falle von Integratoren, die stochastische Prozessen mit Pfaden von endlicher Variation sind (siehe Definition 17.3). Dies erfolgt über das das Stieltjes-Integrals, was eine (einfache) Erweiterung des Lebensgue-Integrals darstellt.

Wir erinnern an Proposition 3.18: Ein σ-endliches Maß auf $\mathbb{R}$ wird genau beschrieben durch eine nicht-fallende, rechtsstetige Funktion $g : \mathbb{R} \to \mathbb{R}$ mit $g(0) = 0$. Wir führen zunächst den entsprechenden Integral-Begriff für dieses Maß ein.

Definition 19.7 (Lebesgue-Stieltjes Integral für nicht-fallende Funktionen). *Sei $g : \mathbb{R}_+ \to \mathbb{R}_+$ rechtsstetig und nicht-fallend. Dann definiert g eindeutig ein σ-endliches Maß μ_g. Ist weiter $f : \mathbb{R}_+ \to \mathbb{R}$ Borel-messbar, dann setzen wir, falls existent,*

$$f \cdot g := \int f dg := f \cdot \mu_g = \int f d\mu_g.$$

Hierbei heißt $\int f dg$ das Lebesgue-Stieltjes-Integral von f bezüglich g. (Die Schreibweise $f \cdot \mu_g$ wurde in Definition 5.12 eingeführt.)

Beispiel 19.8 (Lebesgue-Integral, Integration bezüglich eines Poisson-Prozesses).

1. Mit λ bezeichnen wir das (ein-dimensionale) Lebesgue-Maß. Ist in obiger Definition $g(x) = x$, so ist
$$f \cdot g = \int f(x)\lambda(dx).$$

2. Bekanntlich hat der Poisson-Prozess $\mathcal{X} = (X_t)_{t\ge 0}$ monoton nicht-fallende, rechtsstetige Pfade. Deshalb gibt es zu jedem Pfad $(X_t(\omega))_{t\ge 0}$ eines Poisson-Prozesses ein σ-endliches Maß auf $\mathbb{R}_+$. (Dies ist insbesondere ein Zählmaß.) Seien $T_1, T_2, ...$ die Sprungzeiten eines Poisson-Prozesses und $\mathcal{H} = (H_t)_{t\ge 0}$ ein stochastischer Prozess. Dann können wir
$$(\mathcal{H} \cdot \mathcal{X})_t = \int_0^t H_s dX_s = \sum_{k:T_k\le t} H_{T_k}.$$
schreiben. (Um das letzte Gleichheitszeichen einzusehen, sei bemerkt, dass das Maß $\mu_{\mathcal{X}}$ Atome der Größe 1 auf die Zeitpunkte $T_1, T_2, ...$ legt. Integriert man also gegen ein solches Maß, spielen nur diese Atome eine Rolle.) Insbesondere ist durch das Lebesgue-Stieltjes-Integral bereits die Integration bezüglich des Poisson-Prozesses erklärt.

Monoton nicht-fallende Funktionen (und damit auch stochastische Prozesse mit solchen Pfaden) sind selten. Deutlich häufiger sind solche mit lokal endlicher Variation. Dabei ist eine Funktion $f : [0, \infty] \to \mathbb{R}$ lokal von endlicher Variation, falls

$$\nu_{1,t}(f) := \sup_{n,0\le t_0<...<t_n<t} \sum_{k=1}^n |f(t_k) - f(t_{k-1})| < \infty,$$

für alle $t > 0$ gilt. (Vergleiche auch Definition 17.3.) Es gibt folgenden Zusammenhang zwischen nicht-fallenden und Funktionen mit endlicher Variation:

Lemma 19.9 (Darstellung von Funktionen von lokal endlicher Variation). *Eine Funktion $g : [0, \infty) \to \mathbb{R}$ ist genau dann lokal von endlicher Variation, falls sie als Differenz von zwei monoton nicht-fallenden Funktionen dargestellt werden kann.*

Beweis. '$\Leftarrow$': Sei zunächst $g = a - b$, wobei a und b nicht-fallend sind. Dann gilt $\nu_{1,t}(a) = a(t) - a(0)$ und $\nu_{1,t}(b) = b(t) - b(0)$. Außerdem ist

$$\begin{aligned}\nu_{1,t}(g) &= \sup_{n, 0 \leq t_0 < \ldots < t_n < t} \sum_{k=1}^{n} |g(t_k) - g(t_{k-1})| \\ &\leq \sup_{n, 0 \leq t_0 < \ldots < t_n < t} \sum_{k=1}^{n} |a(t_k) - a(t_{k-1})| + |b(t_k) - b(t_{k-1})| \\ &\leq \nu_{1,t}(a) + \nu_{1,t}(b) < \infty.\end{aligned}$$

'$\Rightarrow$': Klar ist, dass sowohl $t \mapsto \nu_{1,t}(g)$ als auch $t \mapsto \nu_{1,t}(g) - g(t)$ nicht-fallend sind. Deswegen erfüllt schon $g(t) = \nu_{1,t}(g) - (\nu_{1,t}(g) - g(t))$ das Gewünschte. $\square$

Definition 19.10 (Stochastisches Lebesgue-Stieltjes Integral).

1. *Sei $g : \mathbb{R}_+ \to \mathbb{R}$ rechtsstetig und von endlicher Variation. Weiter sei $g = a - b$, wobei a, b nicht-fallend sind. Dann setzen wir für messbares f, falls existent,*

$$f \cdot g := \int f da - \int f db.$$

2. *Sei $\mathcal{X}$ ein stochastischer Prozess mit rechtsstetigen Pfaden von endlicher Variation und $\mathcal{H}$ progressiv messbar. Dann bezeichnen wir mit $\mathcal{H} \cdot \mathcal{X}$ das* stochastische Integral von $\mathcal{H}$ bezüglich $\mathcal{X}$ *als den stochastischen Prozess mit*

$$(\mathcal{H} \cdot \mathcal{X})_t(\omega) := (H_s(\omega))_{0 \leq s \leq t} \cdot (X_s(\omega))_{0 \leq s \leq t}.$$

Bemerkung 19.11 (Wohl-Definiertheit und signierte Maße).

1. Das Integral $f \cdot g$ für eine Funktion g von endlicher Variation (und damit auch das stochastische Integral) ist wohldefiniert: Seien nämlich $g = a - b = a' - b'$ zwei Darstellungen der Funktion g. Dann gilt $h := a + b' = a' + b$, also auch $f \cdot a + f \cdot b' = f \cdot a' + f \cdot b$. Daraus folgt die Behauptung.

2. Im Abschnitt Maßtheorie haben wir σ-endliche Maße auf $\mathcal{B}([0, \infty))$ als Abbildungen $\mathcal{B}([0, \infty)) \to [0, \infty]$ kennen gelernt. Dieser Begriff lässt sich zu σ-endlichen, *signierten Maßen* verallgemeinern. Dies sind σ-endliche Abbildungen $\mathcal{B}([0, \infty)) \to (-\infty, \infty]$ oder $\mathcal{B}([0, \infty)) \to [-\infty, \infty)$. (Mengen können also auch negatives Maß haben.) Klar ist, dass die Differenz zweier σ-endlicher Maße $\mu^+ - \mu^-$, von denen mindestens eines endlich ist, ein signiertes Maß ist. Die Umkehrung gilt auch und ist als Jordan'scher Zerlegungssatz bekannt. Insbesondere könnten wir das Integral $f \cdot g$ für eine Funktion g von endlicher Variation als Integral bezüglich eines endlichen, signierten Maßes, schreiben.

Genau wie in der Integrationstheorie, der wir in der Maßtheorie begegnet sind, betrachten wir nun stochastische Integrale von einfachen Integranden.

Lemma 19.12 (Stochastische Integration einfacher Funktionen). *Sei $\mathcal{X} = (X_t)_{t \geq 0}$ ein Prozess mit Pfaden von lokal endlicher Variation und $\mathcal{H} = (H_t)_{t \geq 0}$ ein einfacher previsibler Prozess wie in* (19.1)*. Dann gilt*

$$(\mathcal{H} \cdot \mathcal{X})_t = \sum_{i=1}^{n} G_i (X_{T_{i+1} \wedge t} - X_{T_i \wedge t}). \tag{19.3}$$

Beweis. Es gibt nicht-fallende Prozesse $\mathcal{Y} = (Y_t)_{t \geq 0}$ und $\mathcal{Z} = (Z_t)_{t \geq 0}$ mit $\mathcal{X} = \mathcal{Y} - \mathcal{Z}$. Weiter gilt $(1_{(T_i, T_{i+1}]} \cdot \mathcal{Y})_t = Y_{T_{i+1} \wedge t} - Y_{T_i \wedge t}$ und $(1_{(T_i, T_{i+1}]} \cdot \mathcal{Z})_t = Z_{T_{i+1} \wedge t} - Z_{T_i \wedge t}$ nach Definition des Stieltjes-Integrales, $i = 1, ..., n$. Die Aussage folgt nun wegen der Linearität des Integrals. □

Wir beschäftigen uns nun noch mit dem Spezialfall von Integratoren von beschränkter Variation. In diesem Fall ist das Lebensgue-Stieltjes Integral ebenfalls als Riemann-Integral interpretierbar. Ähnlich wie in Proposition 4.22 gilt nämlich folgendes:

Proposition 19.13 (Riemann-Integrierbarkeit). *Die Funktion $g : \mathbb{R}_+ \to \mathbb{R}$ habe lokal endliche Variation, d.h. für jedes $t > 0$ und $0 = t_{n,0} \leq ... \leq t_{n,k_n} = t$ mit $\max_k |t_{n,k} - t_{n,k-1}| \xrightarrow{n\to\infty} 0$ ist $\sum_{i=1}^{k_n} |g(t_{n,k}) - g(t_{n,k-1})| < \infty$. Dann existiert $f \cdot g$ für eine stetige Funktion f genau dann, wenn*

$$(f \cdot g)_t = \lim_{n\to\infty} \sum_{k=1}^{\infty} f(s_{n,k})(g(t_{n,k}) - g(t_{n,k-1})) \tag{19.4}$$

für beliebiges $t_{n,k-1} \leq s_{n,k} \leq t_{n,k}$ gilt.

Beweis. Analog zum Beweis von Proposition 4.22. □

Hat der Integrator sogar stetige Pfade von endlicher Variation, so gibt es eine Transformationsformel. Wir bemerken bereits hier, dass ähnliche Transformationen im Fall von unbeschränkter Variation einen zusätzlichen Term benötigen.

Theorem 19.14 (Transformationsformel). *Sei $\mathcal{X} = (X_t)_{t \geq 0}$ ein Prozess mit stetigen Pfaden von lokal endlicher Variation und $f \in \mathcal{C}^1(\mathbb{R})$. Dann gilt*

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s.$$

Beweis. Da $(f'(X_t))_{t \geq 0}$ stetig ist, existiert die rechte Seite. Sei weiter $0 = t_{n,1} \leq \cdots \leq t_{n,k_n} = t$ mit $\max_k t_{n,k} - t_{n,k-1} \xrightarrow{n\to\infty} 0$. Dann gilt für geeignete Zufallsvariablen $t_{n,k-1} \leq S_{n,k} \leq t_{n,k}$ nach Proposition 19.13 und dem Mittelwertsatz

$$\begin{aligned} f(X_t) - f(X_0) &= \sum_{i=1}^{k_n} f(X_{t_{n,k}}) - f(X_{t_{n,k-1}}) \\ &= \sum_{i=1}^{k_n} f'(X_{S_{n,k}})(X_{t_{n,k}} - X_{t_{n,k-1}}) \xrightarrow{n\to\infty} \int_0^t f'(X_s) dX_s. \end{aligned}$$

□

Proposition 19.15 (Martingal-Eigenschaft des stochastischen Integrals). *Sei* $\mathcal{X} = (X_t)_{t\geq 0}$ *ein Martingal mit Pfaden von beschränkter Variation,* $\sup_{t\geq 0} \mathbf{E}[X_t^2] < \infty$ *und* $\mathcal{H}$ *ein beschränkter, adaptierter Prozess mit Pfaden in* $\mathcal{G}_{\mathbb{R}}([0,\infty))$*. Dann ist* $\mathcal{H}\cdot\mathcal{X}$ *ein Martingal mit* $\sup_{t\geq 0} \mathbf{E}[(\mathcal{H}\cdot\mathcal{X})_t^2] < \infty$.

Beweis. Nach Lemma 19.6 gibt es $\mathcal{H}^1 = (H_t^1)_{t\geq 0}, \mathcal{H}^2 = (H_t^2)_{t\geq 0}, ... \in \mathbb{S}$ und $\varepsilon_1, \varepsilon_2, ... > 0$ mit $\varepsilon_2 \downarrow 0$ und $\sup_{0\leq s\leq t} |H_s - H_s^n| \leq \varepsilon_n$ fast sicher. Dann gilt auch

$$\mathbf{P}(\lim_{n\to\infty} \sup_{0\leq s\leq t} |((\mathcal{H}^n - \mathcal{H})\cdot\mathcal{X})_s| = 0) \leq \mathbf{P}(\lim_{n\to\infty} \varepsilon_n \nu_{1,s}(\mathcal{X}) = 0) = 1. \tag{19.5}$$

Deshalb gilt $(\mathcal{H}^n\cdot\mathcal{X})_t \xrightarrow{n\to\infty}_{fs} (\mathcal{H}^n\cdot\mathcal{X})_t$, und zwar sogar uniform auf Kompakta. Nach dem Lemma von Fatou gilt, falls $\mathcal{H}, \mathcal{H}^1, \mathcal{H}^2, ... < \ell$, nach Proposition 19.2.2

$$\mathbf{E}[((\mathcal{H}\cdot\mathcal{X})_t)^2] \leq \liminf_{n\to\infty} \mathbf{E}[((\mathcal{H}^n\cdot\mathcal{X})_t)^2] \leq \ell^2 \cdot \sup_{t\geq 0} \mathbf{E}[X_t^2] < \infty.$$

Zuletzt prüfen wir noch die Martingal-Eigenschaft von $\mathcal{H}\cdot\mathcal{X}$ nach. Sei hierzu $s \leq t$ fest. Nach (19.5) und der L^2-Beschränktheit von $\mathcal{H}^n\cdot\mathcal{X}$ gilt, dass $(\mathcal{H}^n\cdot\mathcal{X})_t \xrightarrow{n\to\infty}_{L^1} (\mathcal{H}\cdot\mathcal{X})_t$. Dann ist

$$\mathbf{E}[(\mathcal{H}\cdot\mathcal{X})_t|\mathcal{F}_s] = \lim_{n\to\infty} \mathbf{E}[(\mathcal{H}^n\cdot\mathcal{X})_t|\mathcal{F}_s] = \lim_{n\to\infty} (\mathcal{H}^n\cdot\mathcal{X})_s = (\mathcal{H}\cdot\mathcal{X})_s$$

und alle Behauptungen sind gezeigt. □

Beispiel 19.16 (Integration bezüglich des Poisson-Prozesses). Sei $\mathcal{Y} = (Y_t)_{t\geq 0}$ eine Poisson-Prozess mit Intensität $\lambda > 0$ und $\mathcal{X} = (X_t)_{t\geq 0}$ mit $X_t = Y_t - \lambda t$. Dann ist $\mathcal{X}$ nach Beispiel 15.4 ein Martingal und hat Pfade von lokal beschränkter Variation. Weiter seien $f, g : \mathbb{N}_0 \to \mathbb{R}$ beschränkt und so, dass $g(i) = f(i+1) - f(i)$ für $i = 1, 2, ...$ Da $\mathcal{Y}_- = (Y_{t-})_{t\geq 0}$ ein Prozess mit Pfaden in $\mathcal{G}_{\mathbb{R}}([0,\infty))$ ist, folgt, falls $T_1, T_2, ...$ die Sprungzeiten von $\mathcal{Y}$ sind, dass

$$\begin{aligned}(f(Y_{t-}+1) - f(Y_{t-}))\cdot\mathcal{X} &= g(\mathcal{Y}_-)\cdot\mathcal{X}\\ &= \Big(\sum_{i:T_i\leq t} g(i-1) - \lambda\int_0^t g(Y_{s-})ds\Big)_{t\geq 0}\\ &= \Big(f(Y_t) - f(0) - \lambda\int_0^t f(Y_s+1) - f(Y_s)ds\Big)_{t\geq 0}\end{aligned}$$

ein Martingal ist. In der Tat, diese Martingal-Eigenschaft folgt ebenfalls aus Theorem 16.30 zusammen mit Beispiel 16.26.1.

19.3 L^2-beschränkte stetige Martingale als Integratoren

Obwohl die Definition des stochastischen Integrals mit Prozessen von lokal beschränkter Variation recht einfach verlief, sind wir noch nicht in der Lage, bezüglich stetiger Martingale (etwa der Brown'schen Bewegung) zu integrieren. Wie wir nämlich in Proposition 17.5 gesehen haben, ist die Variation der Pfade einer Brown'schen Bewegung fast sicher unendlich (und nur die quadratische Variation ist fast sicher endlich und positiv).

Das bisherige Vorgehen basierte darauf, dass man eine Funktion von beschränkter Variation als signiertes Maß auffassen kann. Hierfür war grundlegend, dass ein nicht-fallender Prozess eindeutig ein σ-endliches Maß auf $\mathcal{B}(\mathbb{R})$ definiert. Dies ist für Funktionen mit unbeschränkter Variation nicht so.

Um dennoch stetige Martingale mit unbeschränkter Variation als Integratoren zuzulassen, erinnern wir an die Definition des stochastischen Integrals bezüglich einfachen previsiblen Prozessen aus Proposition 19.2. Wir können nämlich (19.2) als Definition des stochastischen Integrals bezüglich einfacher previsibler Prozesse verwenden und anschließend den Integralbegriff mit Hilfe der Martingal-Eigenschaft des stochastischen Integrals aus Proposition 19.2 erweitern. Die Integratoren die wir hierbei betrachten, sind stetige L^2-beschränkte Martingale.

Definition 19.17 (L^2-beschränkte, stetige Martingale). *Wir bezeichnen mit $\mathcal{M}^2$ die Menge der stetigen, L^2-beschränkten Martingale $\mathcal{X} = (X_t)_{t\geq 0}$ mit $X_0 = 0$. (Das bedeutet, dass $\mathcal{X}$ stetige Pfade hat und $\sup_{t\geq 0}\mathbf{E}[X_t^2] < \infty$). Nach Theorem 15.47 (und Theoremen 15.31 und 15.32) existiert für jedes $\mathcal{X} = (X_t)_{t\geq 0} \in \mathcal{M}^2$ ein X_∞, so dass $(X_t)_{0\leq t\leq\infty}$ ein Martingal ist. Wir definieren die Norm $||\mathcal{X}|| := ||X_\infty||_2$ auf $\mathcal{M}^2$ und erinnern daran, dass $||\sup_{t\geq 0} X_t^2||_2 \leq 2||\mathcal{X}||$ nach Proposition 15.25.*

Um stochastische Integrale für allgemeine Integranden approximieren zu können, werden wir sie als L^2-Grenzwerte definieren. Das funktioniert deswegen, weil der Raum L^2 – und damit auch der Raum $\mathcal{M}^2$ – vollständig ist. Die Vollständigkeit von $\mathcal{M}^2$ wird nun gezeigt.

Lemma 19.18. *Der Raum $\mathcal{M}^2$ ist ein Hilbert-Raum mit Skalarprodukt $\langle \mathcal{X}, \mathcal{Y}\rangle := \mathbf{E}[X_\infty Y_\infty]$.*

Beweis. Wir müssen zeigen, dass $\mathcal{M}^2$ vollständig ist. Sei hierzu $\mathcal{X}^1 = (X_t^1)_{0\leq t\leq\infty}, \mathcal{X}^2 = (X_t^2)_{0\leq t\leq\infty}, ...$ eine Cauchy-Folge in $\mathcal{M}^2$. Nach Definition der Norm in $\mathcal{M}^2$ ist $X_\infty^1, X_\infty^2, ...$ eine Cauchy-Folge in L^2. Deshalb existiert der Grenzwert $X_\infty^n \xrightarrow{n\to\infty} X_\infty$. Wir definieren nun $\mathcal{X} = (X_t)_{t\geq 0}$ durch $X_t = \mathbf{E}[X_\infty|\mathcal{F}_t]$ und bemerken, dass $X_t \xrightarrow{t\to\infty}_{fs,L^2} X_\infty$. Weiter gilt

$$||\sup_{t\geq 0}(X_t^n - X_t)||_2 \leq 2||X_\infty^n - X_\infty|| \xrightarrow{n\to\infty} 0$$

nach Propotision 15.25. Durch Übergang zu einer Teilfolge zeigt dies, dass $\mathcal{X}^n$ gleichmäßig gegen $\mathcal{X}$ konvergiert. Insbesondere hat $\mathcal{X}$ stetige Pfade und $X_0 = 0$. □

Definition 19.19 (Stochastisches Integral von einfachen previsiblen Prozessen). *Sei $\mathcal{X} \in \mathcal{M}^2$ und $\mathcal{H}$ ein einfacher previsibler Prozess wie in Definition 19.1. Dann ist das stochastische Integral $\mathcal{H}\cdot\mathcal{X}$ definiert durch*

$$(\mathcal{H}\cdot\mathcal{X})_t := \sum_{i=1}^n G_i(X_{T_{i+1}\wedge t} - X_{T_i\wedge t}),$$

also wie in (19.3). Sind $G_1, ..., G_n \in \mathcal{L}^1$, ist dies nach Proposition 19.2 wiederum ein Martingal und besitzt außerdem ebenfalls stetige Pfade.

Proposition 19.20 (Integration von Prozessen mit Pfaden in $\mathcal{G}_\mathbb{R}([0,\infty))$). *Sei $\mathcal{X} = (X_t)_{t\geq 0} \in \mathcal{M}^2$ und $\mathcal{H}$ ein beschränkter, adaptierter Prozess mit Pfaden in $\mathcal{G}_\mathbb{R}([0,\infty))$. Ist weiter $\mathcal{H}^1, \mathcal{H}^2, ...$ eine Folge einfacher, previsibler Prozesse, die uniform auf Kompakta gegen $\mathcal{H}$ konvergiert (wie in Lemma 19.6), dann konvergiert für jedes $\tau > 0$ die Folge $((\mathcal{H}^n\cdot\mathcal{X})_{t\wedge\tau})_{t\geq 0}$ zu einem Martingal in $\mathcal{M}^2$.*

Definition 19.21 (Stochastisches Integral von Prozessen mit Pfaden in $\mathcal{G}_\mathbb{R}([0,\infty))$). *Sei $\mathcal{X} \in \mathcal{M}^2$ und $\mathcal{H}$ ein beschränkter, adaptierter, stochastischer Prozess mit Pfaden in $\mathcal{G}_\mathbb{R}([0,\infty))$. Dann definieren wir $\mathcal{H} \cdot \mathcal{X}$ als den stochastischen Prozess, für den $(\mathcal{H} \cdot \mathcal{X})_t$ der Grenzwert aus Proposition 19.20 für (ein beliebiges) $\tau \geq t$ ist.*

Beweis von Proposition 19.20. Es genügt, Martingale mit kompakter Indexmenge $t \in [0, \tau]$ zu betrachten. Für die approximierende Folge $\mathcal{H}^1, \mathcal{H}^2, ...$ gibt es $\varepsilon_1, \varepsilon_2, ...$ mit $\varepsilon_n \downarrow 0$ und $\sup_{0 \leq t \leq \tau} |H_t^n - H_t| \leq \varepsilon_n$ fast sicher wegen Lemma 19.6. Weiter ist, wegen $\sup_{0 \leq t \leq \tau} |H_t^n - H_t^m| \leq \varepsilon_m + \varepsilon_n$, wegen Proposition 19.2,

$$||\mathcal{H}^n \cdot \mathcal{X} - \mathcal{H}^m \cdot \mathcal{X}|| = ||(\mathcal{H}^n - \mathcal{H}^m) \cdot \mathcal{X}|| \leq (\varepsilon_m + \varepsilon_n) \cdot ||\mathcal{X}|| \xrightarrow{m,n \to \infty} 0.$$

Deshalb ist $\mathcal{H}^n \cdot \mathcal{X}$ eine Cauchy-Folge in $\mathcal{M}^2$, die wegen der Vollständigkeit zu einem Element $\mathcal{H} \cdot \mathcal{X} \in \mathcal{M}^2$ konvergiert. □

Beispiel 19.22 (Integral bezüglich der Brown'schen Bewegung). Mit Hilfe der letzten Proposition können wir nun Integrale bezüglich der Brown'schen Bewegung definieren. Sei hierzu $\mathcal{X} = (X_t)_{t \geq 0}$ eine Brown'sche Bewegung mit $X_0 = 0$ und $f \in \mathcal{C}_b^2(\mathbb{R})$ beschränkt. Wir werden nun zeigen, dass

$$\int_0^t f'(X_s) dX_s = f(X_t) - f(X_0) - \tfrac{1}{2} \int_0^t f''(X_s) ds.$$

Hierzu definieren wir rekursiv $T_0^\varepsilon = 0$, $T_{i+1}^\varepsilon = \inf\{t > T_i^\varepsilon : |X_t - X_{T_i^\varepsilon}| = \varepsilon\}$. Dann gilt für geeignete $C, C' > 0$

$$\begin{aligned} \mathbf{E}[T_{i+1}^\varepsilon - T_i^\varepsilon] &= \mathbf{E}[T_1^\varepsilon] = C\varepsilon^2, \\ \mathbf{E}[(T_{i+1}^\varepsilon - T_i^\varepsilon)^2] &= \mathbf{E}[(T_1^\varepsilon)^2] = C'\varepsilon^4 \end{aligned}$$

und somit

$$\begin{aligned}
&\mathbf{E}\Big[\Big(\sum_{i=0}^\infty f'(X_{T_i^\varepsilon \wedge t})(X_{T_{i+1}^\varepsilon \wedge t} - X_{T_i^\varepsilon \wedge t}) - \Big(f(X_t) - f(X_0) - \tfrac{1}{2}\int_0^t f''(X_s)ds\Big)\Big)^2\Big] \\
&\quad = \mathbf{E}\Big[\Big(\sum_{i=0}^\infty f'(X_{T_i^\varepsilon \wedge t})(X_{T_{i+1}^\varepsilon \wedge t} - X_{T_i^\varepsilon \wedge t}) - \Big(f(X_{T_{i+1}^\varepsilon \wedge t}) - f(X_{T_i^\varepsilon \wedge t}) - \tfrac{1}{2}\int_{T_i^\varepsilon \wedge t}^{T_{i+1}^\varepsilon \wedge t} f''(X_s)ds\Big)\Big)^2\Big] \\
&\quad = \sum_{i=0}^\infty \mathbf{E}\Big[\Big(f'(X_{T_i^\varepsilon \wedge t})(X_{T_{i+1}^\varepsilon \wedge t} - X_{T_i^\varepsilon \wedge t}) - \Big(f(X_{T_{i+1}^\varepsilon \wedge t}) - f(X_{T_i^\varepsilon \wedge t}) - \tfrac{1}{2}\int_{T_i^\varepsilon \wedge t}^{T_{i+1}^\varepsilon \wedge t} f''(X_s)ds\Big)\Big)^2\Big] \\
&\quad = \tfrac{1}{4}\sum_{i=0}^\infty \mathbf{E}\Big[\Big(f''(X_{S_i^\varepsilon \wedge t})(X_{T_{i+1}^\varepsilon \wedge t} - X_{T_i^\varepsilon \wedge t})^2 - f''(X_{\tilde{S}_i^\varepsilon \wedge t})(T_{i+1}^\varepsilon \wedge t - T_i^\varepsilon \wedge t)\Big)^2\Big] \\
&\quad \leq \tfrac{1}{2}\sum_{i=0}^\infty \mathbf{E}\Big[\Big(f''(X_{T_i^\varepsilon \wedge t})((X_{T_{i+1}^\varepsilon \wedge t} - X_{T_i^\varepsilon \wedge t})^2 - (T_{i+1}^\varepsilon \wedge t - T_i^\varepsilon \wedge t))\Big)^2\Big] \\
&\qquad\qquad + \mathbf{E}[(f''(X_{T_i^\varepsilon \wedge t}) - f''(X_{S_i^\varepsilon \wedge t}))^2 \varepsilon^4] + \mathbf{E}[(f''(X_{T_i^\varepsilon \wedge t}) - f''(X_{\tilde{S}_i^\varepsilon \wedge t}))^2 (T_{i+1}^\varepsilon \wedge t - T_i^\varepsilon \wedge t)^2] \\
&\quad \xrightarrow{\varepsilon \to 0} 0
\end{aligned}$$

für Zufallsvariable $T_i^\varepsilon \leq S_i^\varepsilon, \tilde{S}_i^\varepsilon \leq T_{i+1}^\varepsilon$ nach der Taylor-Formel und dem Mittelwertsatz. (Im letzten Ungleichheitszeichen haben wir die einfache Abschätzung $(ab)^2 \leq 2(a-a')^2b^2 + 2(a')^2b^2$ verwendet.) Damit konvergiert einerseits $\sum_{i=0}^\infty f'(X_{T_i^\varepsilon \wedge t})(X_{T_{i+1}^\varepsilon \wedge t} - X_{T_i^\varepsilon \wedge t}$ in L^2 gegen $\int_0^t f'(X_s)dX_s$ nach Definition des stochastischen Integrals, jedoch auch gegen $f(X_t) - f(X_0) - \frac{1}{2}\int_0^t f''(X_s)ds$. Wegen der Eindeutigkeit des L^2-Grenzwertes folgt die Behauptung.

Es scheint klar zu sein, dass diese Berechnungsmethode für stochastische Integrale zwar machbar, jedoch nicht sonderlich elegant ist. Deswegen werden wir mit der Itô-Formel (Theorem 19.51) eine einfachere Methode zur Berechnung in ähnlichen Fällen kennen lernen.

19.4 Lokale Martingale als Integratoren

In Rechnungen wäre es häufig schön zu wissen, dass X_s für ein Martingal $\mathcal{X} = (X_t)_{t\geq 0}$ nicht *zu groß* sein kann. Hat $\mathcal{X}$ stetige Pfade, ist dies durch Übergang zu einer Folge $\mathcal{X}^1, \mathcal{X}^2, ...$ von gestoppten Martingalen möglich, so dass $\mathcal{X}^n$ immer dann gestoppt wird, wenn $|X_t| = n$. Andersherum sind stetige Prozesse, bei denen es solche Stoppzeiten gibt, nicht unbedingt Martingale. Deswegen benötigen wir eine Klasse von stochastischen Prozessen, die größer ist als die Klasse der Martingale.

Definition 19.23 (Lokales Martingal und gestoppter Prozess).
1. *Ein reellwertiger, stochastischer Prozess $\mathcal{X} = (X_t)_{t\geq 0}$ heißt* lokales Martingal*, falls es Stoppzeiten $T_1, T_2, ...$ mit $T_n \uparrow \infty$ gibt, so dass $(X_{t\wedge T_n})_{t\geq 0}$ für alle n ein Martingal ist. Hier heißt $T_1, T_2, ...$ eine* lokalisierende Folge von Stoppzeiten*.*
2. *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein stochastischer Prozess und T eine Stoppzeit. Dann bezeichnen wir mit $\mathcal{X}^T := (X_{T\wedge t})_{t\geq 0}$ den bei T gestoppten Prozess.*

Anders formuliert bedeutet obige Definition, dass $\mathcal{X}$ genau dann ein lokales Martingal ist, falls $\mathcal{X}^{T_n}$ für eine geeignete Folge von Stoppzeiten $T_n \uparrow \infty$ Martingale sind.

Bemerkung 19.24 (Eigenschaften von lokalen Martingalen).
1. Nach dem Optional Stopping Theorem, Proposition 15.18 und Korollar 15.49, ist jedes Martingal ein lokales Martingal.
2. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein stetiges lokales Martingal. Dann ist $T_n = \inf\{t : |X_t| = n \vee |X_0|\}$ eine lokalisierende Folge von Stoppzeiten.

   Denn: Da $\mathcal{X}$ reellwertig und stetig ist, nimmt jeder Pfad auf $[0,t]$ sein Supremum an. Daraus folgt, dass $T_n \uparrow \infty$ gilt. Weiter sei $S_m \uparrow \infty$ eine lokalisiernde Folge von Stoppzeiten für $\mathcal{X}$. Für $s \leq t$ ist

   $$\mathbf{E}[X_{t\wedge T_n}|\mathcal{F}_s] = \lim_{m\to\infty} \mathbf{E}[X_{t\wedge T_n \wedge S_m}|\mathcal{F}_s] = \lim_{m\to\infty} X_{s\wedge T_n\wedge S_m} = X_{s\wedge T_n},$$

   da $X_{t\wedge T_n}$ beschänkt ist.
3. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein beschränktes lokales Martingal. Dann ist $\mathcal{X}$ ein Martingal.

   Denn: Sei $T_1, T_2, ...$ eine lokalisierende Folge von Stoppzeiten. Nach majorisierter Konvergenz gilt $\mathbf{E}[X_t|\mathcal{F}_s] = \lim_{n\to\infty} \mathbf{E}[X_{t\wedge T_n}|\mathcal{F}_s] = \lim_{n\to\infty} X_{s\wedge T_n} = X_s$ für $s \leq t$.

Beispiel 19.25 (Ein echtes lokales Martingal). Sei $\mathcal{W} = (X_t, Y_t, Z_t)_{t\geq 0}$ eine dreidimensionale Brownsche Bewegung, die in $(x, y, z) \neq 0$ gestartet wird. Wir betrachten nun den Prozess $\mathcal{V} = (V_t)_{t\geq 0}$, gegeben durch

$$V_t = \frac{1}{\sqrt{X_t^2 + Y_t^2 + Z_t^2}}$$

und behaupten, dass $\mathcal{V}$ zwar ein lokales Martingal, jedoch kein Martingal ist.
Denn: Der Generator der Brown'schen Bewegung ist bekanntlich

$$(Gf)(w) = \tfrac{1}{2}\Big(\frac{\partial^2 f(w)}{\partial x^2} + \frac{\partial^2 f(w)}{\partial y^2} + \frac{\partial^2 f(w)}{\partial z^2}\Big).$$

Da für $f(w) = 1/\sqrt{x^2 + y^2 + z^2}$

$$\frac{1}{2}\frac{\partial^2 f(w)}{\partial x^2} f(w) = -\frac{\partial}{\partial x}\frac{x}{(x^2 + y^2 + z^2)^{3/2}} = -\frac{1}{(x^2 + y^2 + z^2)^{3/2}} + 3\frac{x^2}{(x^2 + y^2 + z^2)^{5/2}},$$

ist für $w \neq 0$

$$(Gf)(w) = 0.$$

Sei nun $B_\varepsilon(0)$ der Ball um 0 mit Radius ε und $g^\varepsilon \in \mathcal{C}_b^2(\mathbb{R}^3)$, so dass $g^\varepsilon|_{B_\varepsilon(0)^c} = f|_{B_\varepsilon(0)^c}$ und $T_\varepsilon := \inf\{t > 0 : ||W_t||_2 = \varepsilon\}$.

Nach Theorem 16.30 ist $\Big(g^\varepsilon(V_t) - \int_0^t (Gg^\varepsilon)(V_s)ds\Big)_{t\geq 0}$ ein Martingal, also ist nach dem Optional Stopping Theorem – Proposition 15.18 – auch

$$(f(V_{t\wedge T_\varepsilon}))_{t\geq 0} = (g^\varepsilon(V_{t\wedge T_\varepsilon}))_{t\geq 0} = \Big(g^\varepsilon(V_{t\wedge T_\varepsilon}) - \int_0^{T_\varepsilon\wedge t} (Gg^\varepsilon)(V_s)ds\Big)_{t\geq 0}$$

ein Martingal. Sei nun $T = T_\varepsilon \wedge T_R$ für $\varepsilon < ||W_0||_2 < R$. Dann ist ebenfalls $(f(V_{t\wedge T}))_{t\geq 0}$ ein Martingal, also

$$1/||V_0||_2 = f(V_0) = g^\varepsilon(V_0) = \mathbf{E}[g^\varepsilon(V_T)] = \mathbf{P}(T_\varepsilon > T_R)\frac{1}{R} + (1 - \mathbf{P}(T_\varepsilon > T_R))\frac{1}{\varepsilon}.$$

Daraus folgt

$$\mathbf{P}(T_\varepsilon > T_R) = \frac{1/\varepsilon - 1/||V_0||_2}{1/\varepsilon - 1/R} \xrightarrow{\varepsilon\to 0} 1.$$

Da $T_R \xrightarrow{R\to\infty} \infty$ aufgrund der Stetigkeit der Pfade der Brown'schen Bewegung, gilt damit auch $T_\varepsilon \xrightarrow{\varepsilon\to 0} \infty$.

Wir haben nun gezeigt, dass $(f(V_t))_{t\geq 0}$ ein lokales Martingal ist. Allerdings berechnet man

$$\mathbf{E}[f(V_t)] = \mathbf{E}\Big[\frac{1}{||V_t||_2}\Big] = \frac{1}{\sqrt{t}} \cdot \mathbf{E}\Big[\frac{1}{||V_1||_2}\Big] \xrightarrow{t\to\infty} 0.$$

Da der Erwartungswert eines Martingales konstant ist, kann also $(f(V_t))_{t\geq 0}$ kein Martingal sein.

Wir haben bereits stochastische Integrale bezüglich Prozessen von beschränkter Variation sowie bezüglich stetigen Martingalen. Wir zeigen nun, dass wir dabei keine Fälle doppelt behandelt haben.

Theorem 19.26 (Stetige lokale Martingale von beschränkter Variation sind konstant). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein lokales Martingal mit stetigen Pfaden und beschränkter Variation. Dann ist $t \mapsto X_t$ fast sicher konstant.*

Beweis. Ohne Einschränkung nehmen wir an, dass $X_0 = 0$ und $\mathcal{X}$ ein Martingal ist. (Falls $\mathcal{X}$ nur ein lokales Martingal ist, können wir zeigen, dass alle gestoppten Martingale entlang der lokalisierenden Folge von Stoppzeiten konstant sind, was die Aussage impliziert.) Sei $t \mapsto \nu_{1,t}(\mathcal{X})$ die Variation von $\mathcal{X}$. Wir definieren die Stoppzeiten

$$T_N := \inf\{t \geq 0 : \nu_{1,t}(\mathcal{X}) \geq N\}$$

und stellen fest, dass $(X_{t\wedge T_N})_{t\geq 0}$ ein Martingal ist, dessen Variation durch N beschränkt ist. Außerdem ist $t \mapsto X_t$ genau dann konstant, wenn $t \mapsto X_{t\wedge T_N}$ für ein N konstant ist. Nach Voraussetzung gilt außerdem $T_N \uparrow \infty$. Es genügt also, die Behuaptung im Fall zu zeigen, dass die Variation von $\mathcal{X}$ durch N beschränkt ist. Für $t > 0$ setzen wir $t_{n,k} := tk/n$ und definieren

$$\begin{aligned} Z_n &:= \sum_{k=1}^n (X_{t_{n,k}} - X_{t_{n,k-1}})^2 \leq \max_{1\leq k\leq n} |X_{t_{n,k}} - X_{t_{n,k-1}}| \cdot \sum_{k=1}^n |X_{t_{n,k}} - X_{t_{n,k-1}}| \\ &\leq \max_{1\leq k\leq n} |X_{t_{n,k}} - X_{t_{n,k-1}}| \cdot \nu_{1,t}(\mathcal{X}) \xrightarrow{n\to\infty} 0, \end{aligned}$$

da $\mathcal{X}$ stetige Pfade hat. Nach Definition gilt außerdem $Z_n \leq \nu_{1,t}(\mathcal{X})^2 \leq N^2$. Deshalb folgern wir mit majorisierter Konvergenz, dass

$$\mathbf{E}[X_t^2] = \mathbf{E}\Big[\Big(\sum_{k=1}^n X_{t_{n,k}} - X_{t_{n,k-1}}\Big)^2\Big] = \mathbf{E}[Z_n] \xrightarrow{n\to\infty} 0.$$

Also ist $X_t = 0$ fast sicher. □

Wir fahren fort mit einer häufig anwendbaren Charakterisierung von lokalen Martingalen.

Lemma 19.27 (Charakterisierungen von lokalen Martingalen). *Sei $T_n \uparrow \infty$ und $\mathcal{X}$ ein reellwertiger stochastischer Prozess. Dann ist $\mathcal{X}$ genau dann ein lokales Martingal, falls $\mathcal{X}^{T_n}$ für alle n ein lokales Martingal ist.*

Beweis. Sei zunächst $\mathcal{X}$ ein lokales Martingal und $S_1, S_2, ...$ eine lokalisierende Folge von Stoppzeiten. Für jede Stoppzeit T ist dann – nach dem Optional Stopping Theorem – der Prozess $(\mathcal{X}^{S_n})^T = (\mathcal{X}^T)^{S_n}$ ein Martingal. Das bedeutet, dass $\mathcal{X}^T$ ein lokales Martingal (mit der lokalisierenden Folge von Stoppzeiten $S_1, S_2, ...$) ist.

Sei andersherum für jedes n der Prozess $\mathcal{X}^{T_n}$ ein lokales Martingal mit lokalisierender Folge von Stoppzeiten $S_1^n, S_2^n, ...$ Da $S_k^N \uparrow \infty$ für $k \to \infty$, wählen wir k_n mit

$$\mathbf{P}(S_{k_n}^n < T_n \wedge n) \leq 2^{-n}.$$

Nach dem Borel-Cantelli-Lemma ist $T_n' := S_{k_n}^n \wedge T_n \uparrow \infty$. Um eine aufsteigende Folge von Stoppzeiten zu erhalten, definieren wir $T_n'' := \inf_{m\geq n} T_n'$. Außerdem ist $\mathcal{X}^{T_n''} = (\mathcal{X}^{T_n'})^{T_n''} = ((\mathcal{X}^{T_n})^{S_{k_n}^n})^{T_n''}$. Nach Voraussetzung ist $(\mathcal{X}^{T_n})^{S_{k_n}^n}$ ein Martingal, also auch $\mathcal{X}^{T_n''}$. Insbesondere ist $\mathcal{X}$ ein lokales Martingal. □

Für die Brown'sche Bewegung $\mathcal{X}$ hatten wir bereits gesehen, dass die quadratische Variation durch $[\mathcal{X}]_t = t$ gegeben ist. Da die quadratische Variation (und Co-variation zwischen zwei Prozessen) für die stochastische Integration bezüglich lokaler Martingale eine entscheidende Rolle spielen wird (siehe Theorem 19.36), werden wir diese nun konstruieren.

Proposition 19.28 (Quadratische Variation für beschränkte, stetige Martingale). *Sei $\mathcal{X}$ ein beschränktes Martingal mit stetigen Pfaden und $X_0 = 0$. Weiter sei rekursiv $T_0^n = 0$ sowie*

$$T_{k+1}^n := \inf\{t > T_k^n : |X_t - X_{T_k^n}| > 2^{-n}\}.$$

Dann gibt es einen fast sicher eindeutigen Prozess $[\mathcal{X}] = ([\mathcal{X}]_t)_{t\geq 0}$ mit $[\mathcal{X}]_0 = 0$ und nicht-fallenden Pfaden endlicher Variation, so dass für

$$\mathcal{Q}^n = (Q_t^n)_{t\geq 0} \text{ mit } Q_t^n := \sum_{k=0}^{\infty} (X_{t\wedge T_{k+1}^n} - X_{t\wedge T_k^n})^2$$

gilt, dass $\sup_{t\geq 0} |Q_t^n - [\mathcal{X}]_t| \xrightarrow{n\to\infty}_p 0$. Weiter ist $\mathcal{X}^2 - [\mathcal{X}]$ ein Martingal.

Beweis. Die fast sichere Eindeutigkeit folgt aus Theorem 19.26. (Angenommen, es gäbe zwei Prozesse $[\mathcal{X}]$ und $[\mathcal{X}]'$ mit den geforderten Eigenschaften. Dann wäre $[\mathcal{X}] - [\mathcal{X}]'$ ein Martingal mit Pfaden von endlicher Variation, also fast sicher konstant. Da $[\mathcal{X}]_0 = [\mathcal{X}']_0$, wäre $[\mathcal{X}] = [\mathcal{X}]'$ fast sicher.) Wir definieren den Prozess $\mathcal{H}^n = (H_t^n)_{t\geq 0} \in \mathbb{S}$ mittels

$$H_t^n = \sum_{k=0}^{\infty} X_{T_k^n} 1_{(T_k^n, T_{k+1}^n]}(t).$$

Dann gilt $||\mathcal{H}^n - \mathcal{X}||_2 \leq 2^{-n}$. Nach Proposition 19.2.2 gilt damit auch

$$||\mathcal{H}^n \cdot \mathcal{X} - \mathcal{H}^m \cdot \mathcal{X}||_2 \xrightarrow{m,n\to\infty} 0.$$

Also ist $(\mathcal{H}^n \cdot \mathcal{X})_{n=1,2,\ldots}$ eine $\mathcal{M}^2$-Cauchy Folge, die nach Proposition 19.18 gegen ein Martingal $\mathcal{N} = (N_t)_{t\geq 0}$ konvergiert. Weiter schreiben wir

$$\begin{aligned}
(\mathcal{H}^n \cdot \mathcal{X})_t &= \sum_{k=0}^{\infty} X_{T_k^n}(X_{T_{k+1}^n\wedge t} - X_{T_k^n\wedge t}),\\
X_t^2 &= \sum_{k=0}^{\infty} X_{T_{k+1}^n\wedge t}^2 - X_{T_k^n\wedge t}^2\\
&= \sum_{k=0}^{\infty} (X_{T_{k+1}^n\wedge t} - X_{T_k^n\wedge t})^2 + 2X_{T_k^n\wedge t}(X_{T_{k+1}^n\wedge t} - X_{T_k^n\wedge t})\\
&= Q_t^n + 2(\mathcal{H}^n \cdot \mathcal{X})_t.
\end{aligned}$$

Wir definieren nun $[\mathcal{X}] := \mathcal{X}^2 - 2\mathcal{N}$ (so dass $\mathcal{X}^2 - [\mathcal{X}]$ automatisch ein Martingal ist). Dann gilt

$$\sup_{t\geq 0} |Q_t^n - [\mathcal{X}]_t| = \sup_{t\geq 0} |X_t^2 - 2(\mathcal{H}^n \cdot \mathcal{X})_t - X_t^2 + 2N_t = \sup_{t\geq 0} |2N_t - 2(\mathcal{H}^n \cdot \mathcal{X})_t| \xrightarrow{n\to\infty}_p 0.$$

Da $\mathcal{Q}^n$ nicht-fallende Pfade hat, gilt dasselbe auch für $[\mathcal{X}]$. □

Wir erweitern nun das letzte Resultat von beschränkten Martingalen auf lokale Martingale.

Theorem 19.29 (Quadratische Variation von lokalen Martingalen). *Sei $\mathcal{X}$ ein lokales Martingal mit stetigen Pfaden. Dann gibt es einen fast sicher eindeutigen Prozess $[\mathcal{X}] = ([\mathcal{X}]_t)_{t\geq 0}$ mit $[\mathcal{X}]_0 = 0$ und nicht-fallenden Pfaden endlicher Variation, so dass $\mathcal{X}^2 - [\mathcal{X}]$ ein lokales Martingal ist. Weiter gilt $[\mathcal{X}^T] = [\mathcal{X}]^T$ für jede Stoppzeit T.*

Beweis. Wieder folgt die fast sichere Eindeutigkeit aus Theorem 19.26. Ist außerdem $\mathcal{X}^2 - [\mathcal{X}]$ ein lokales Martingal und T eine fast sicher endliche Stoppzeit, dann ist $(\mathcal{X}^T)^2 - [\mathcal{X}]^T = (\mathcal{X}^2 - [\mathcal{X}])^T$ ein lokales Martingal, also $[\mathcal{X}^T] = [\mathcal{X}]^T$.

Für die Existenz von $[\mathcal{X}]$ definieren wir $T_n := \inf\{t \geq 0 : |X_t| = n\}$, womit $T_n \uparrow \infty$ gilt. Weiter ist $\mathcal{X}^{T_n}$ ein (durch n) berschränktes Martingal mit stetigen Pfaden. Also existiert $[\mathcal{X}^{T_n}]$ nach Proposition 19.28. Da für $m \geq n$ außerdem $(\mathcal{X}^{T_m})^{T_n} = \mathcal{X}^{T_n}$, ist $\mathcal{X}^{T_m} = \mathcal{X}^{T_n}$ auf $[0, T_n]$, also auch $[\mathcal{X}^{T_n}] = [\mathcal{X}^{T_m}]$ auf $[0, T_n]$. Wir definieren $[\mathcal{X}]_t := \lim_{n\to\infty}[\mathcal{X}^{T_n}]_t$, wobei die Konvergenz aus $T_n \uparrow \infty$ folgt. Weiter ist klar, dass $(\mathcal{X}^2 - [\mathcal{X}])^{T_n} = (\mathcal{X}^{T_n})^2 - [\mathcal{X}^{T_n}]$ für jedes n ein Martingal, $\mathcal{X}^2 - [\mathcal{X}]$ also ein lokales Martingal ist. □

Beispiel 19.30 (Von der Brown'schen Bewegung abgeleitete Martingale). Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung. Bekanntermaßen (siehe Beispiel 15.5) ist $\mathcal{Y} = (Y_t)_{t\geq 0}$ mit $Y_t = X_t^2 - t$ ein Martingal. Wir zeigen, dass

$$[\mathcal{Y}]_t = 4\int_0^t X_s^2 ds.$$

Denn: Nach Beispiel 16.26.2 ist auch

$$\Big(X_t^4 - 6\int_0^t X_s^2 ds\Big)_{t\geq 0}$$

ein Martingal. Weiter betrachten wir den Prozess $(t, X_t)_{t\geq 0}$, also den Markov-Prozess mit Zustandsraum $\mathbb{R}_+ \times \mathbb{R}$ und Generator

$$(Gf)(t,x) = \frac{\partial f(t,x)}{\partial t} + \tfrac{1}{2}\frac{\partial^2 f(t,x)}{\partial x^2}.$$

Mit $f(t,x) = tx^2$ folgt aus Theorem 16.30, dass

$$\Big(tX_t^2 - \int_0^t X_s^2 + s ds\Big)_{t\geq 0} = \Big(tX_t^2 - \tfrac{1}{2}t^2 - \int_0^t X_s^2 ds\Big)_{t\geq 0}$$

ein Martingal ist. Also ist

$$\Big(X_t^4 - 6\int_0^t X_s^2 ds - 2tX_t + t^2 + 2\int_0^t X_s^2 ds\Big)_{t\geq 0} = \Big((X_t^2 - t)^2 - 4\int_0^t X_s^2 ds\Big)_{t\geq 0}$$

ebenfalls ein Martingal. Damit folgt die Behauptung.

Theorem 19.31 (Covariation von lokalen Martingalen). *Seien $\mathcal{X}, \mathcal{Y}$ lokale Martingale. Dann gibt es einen fast sicher eindeutigen Prozess $[\mathcal{X}, \mathcal{Y}]$ von lokal beschränkter Variation und $[\mathcal{X}, \mathcal{X}]_0$, so dass $\mathcal{X}\mathcal{Y} - [\mathcal{X}, \mathcal{Y}]$ ein lokales Martingal ist. Weiter ist $(\mathcal{X}, \mathcal{Y}) \mapsto [\mathcal{X}, \mathcal{Y}]$ symmetrisch und bilinear mit*

$$[\mathcal{X}, \mathcal{Y}]^T = [\mathcal{X}, \mathcal{Y}^T] = [\mathcal{X}^T, \mathcal{Y}] = [\mathcal{X}^T, \mathcal{Y}^T]$$

für jede Stoppzeit T.

Beweis. Wir definieren
$$[\mathcal{X},\mathcal{Y}] = \tfrac{1}{4}([\mathcal{X}+\mathcal{Y}] - [\mathcal{X}-\mathcal{Y}]).$$
Dann gilt
$$4(\mathcal{X}\mathcal{Y} - [\mathcal{X},\mathcal{Y}]) = (\mathcal{X}+\mathcal{Y})^2 - (\mathcal{X}-\mathcal{Y})^2 - [\mathcal{X}+\mathcal{Y}] + [\mathcal{X}-\mathcal{Y}],$$
woraus alle Behauptungen mit Theorem 19.29 folgen. □

Proposition 19.32 (Stetigkeit der quadratischen Variation). *Sei $\mathcal{X}^1 = (X_t^1)_{t\ge 0}, \mathcal{X}^2 = (X_t^2)_{t\ge 0}, ...$ eine Folge von lokalen Martingalen mit stetigen Pfaden mit Start in 0. Dann ist $\sup_{t\ge 0} |X_t^n| \xrightarrow{n\to\infty}_p 0$ genau dann, wenn $[\mathcal{X}^n]_\infty \xrightarrow{n\to\infty}_p 0$. Insbesondere ist die Abbildung $\mathcal{X} \mapsto [\mathcal{X}]$ auf dem Raum der lokalen Martingale mit stetigen Pfaden stetig. Analoges gilt für die Covariation.*

Beweis. Sei zunächst $\sup_{t\ge 0} |X_t^n| \xrightarrow{n\to\infty}_p 0$ und $\varepsilon > 0$. Wir definieren $T_n := \inf\{t \ge 0 : |X_t^n| > \varepsilon$, $n = 1, 2, ...$ Dann ist für $\mathcal{Y}^n := (\mathcal{X}^n)^2 - [\mathcal{X}^n]$ der Prozess $(\mathcal{Y}^n)^{T_n}$ ein Martingal, das nach Theorem 15.32 (und Theorem 15.47) zu einem Martingal mit Indexmenge $[0,\infty]$ erweitert werden kann. Da $\mathbf{E}[Y_t^n] = 0$ und $X_{t\wedge T_n}^n \le \varepsilon$, folgt $\mathbf{E}[[\mathcal{X}^n]_{T_n}] \le \varepsilon^2$. Aus der Markov-Ungleichung folgern wir
$$\mathbf{P}([\mathcal{X}^n]_\infty > \varepsilon) \le \mathbf{P}(T_n < \infty) + \frac{1}{\varepsilon}\mathbf{E}[[\mathcal{X}^n]_{T_n}] \le \mathbf{P}(\sup_{t\ge 0} |X_t^n| > \varepsilon) + \varepsilon.$$
Nach Voraussetzung konvergiert die rechte Seite gegen ε, woraus $[\mathcal{X}^n]_\infty \xrightarrow{n\to\infty}_p 0$ folgt.

Gilt andersherum $[\mathcal{X}^n]_\infty \xrightarrow{n\to\infty}_p 0$, so schreiben wir mit $T_n := \inf\{t : [\mathcal{X}^n]_t > \varepsilon\}$
$$\begin{aligned}\mathbf{P}(\sup_{t\ge 0} |X_t^n| > \varepsilon) &\le \mathbf{P}(T_n < \infty) + \mathbf{P}(\sup_{t\ge 0} |X_{t\wedge T_n}^n| > \varepsilon) \\ &\le \mathbf{P}(T_n < \infty) + \frac{\mathbf{E}[[\mathcal{X}^n]_\infty^{T_n}]}{\varepsilon^2} \\ &= \mathbf{P}(T_n < \infty) + \frac{\mathbf{E}[[\mathcal{X}^n]_\infty \wedge \varepsilon]}{\varepsilon^2} \xrightarrow{n\to\infty} 0.\end{aligned}$$
□

Proposition 19.33 (Cauchy-Schwartz Ungleichungen für die Covariation). *Seien $\mathcal{X}, \mathcal{Y}$ lokale Martingale und $\mathcal{H} = (H_t)_{t\ge 0}$ progressiv messbar. Dann gilt*
$$[\mathcal{X},\mathcal{Y}]^2 \le [\mathcal{X}][\mathcal{Y}]$$
sowie
$$(\mathcal{H}\cdot[\mathcal{X},\mathcal{Y}])_t^2 \le (\mathcal{H}^2\cdot[\mathcal{X}])_t[\mathcal{Y}]_t$$
fast sicher.

Beweis. Zunächst bemerken wir folgendes: Sei $A, B \geq 0$. Wenn $a^2A + 2abC + b^2B \geq 0$ für alle a, b, dann ist $C^2 \leq AB$. Setzt man nämlich $a = \pm 1/\sqrt{A}, b = 1/\sqrt{B}$, so folgt $\pm 2C/\sqrt{AB} \geq -2$, also $|C| \leq \sqrt{AB}$ oder $C^2 \leq AB$.

Für alle a, b gilt wegen der Linearität der Covariation

$$0 \leq [a\mathcal{X} + b\mathcal{Y}] = a^2[\mathcal{X}] + 2ab[\mathcal{X}, \mathcal{Y}] + b^2[\mathcal{Y}]$$

fast sicher. (Wir können die Ausnahme-Nullmenge A sogar unabhängig von a, b wählen. Hierzu bemerken wir, dass die Vereinigung der Ausnahme-Nullmengen für $a, b \in \mathbb{Q}$ wieder eine Nullmenge ist. Diese muss gleichzeitig die Ausnahme-Nullmenge für alle $a, b \in \mathbb{R}$ sein, da die quadratischen Variation stetig ist; siehe Proposition 19.32.) Aus der Eingangsbemerkung folgt nun sofort $[\mathcal{X}, \mathcal{Y}]_t^2 \leq [\mathcal{X}]_t[\mathcal{Y}]_t$ für alle $t \geq 0$. Die erste Behauptung ergibt sich, da die rechte Seite eine nicht-fallende Funktion in t ist.

Um die zweite Aussage zu zeigen, nehmen wir o.E. an, dass $\mathcal{H} \geq 0$. Außerdem bemerken wir, dass die erste Aussage analog für Teilintervalle bewiesen wird, also $([\mathcal{X}, \mathcal{Y}]_t - [\mathcal{X}, \mathcal{Y}]_s)^2 \leq ([\mathcal{X}]_t - [\mathcal{X}]_s)([\mathcal{Y}]_t - [\mathcal{Y}]_s)$. Sei nun $\mathcal{H} = \sum_{i=1}^n G_i 1_{I_i}$ für disjunkte offene Intervalle $I_i = (V_i, W_i)$ und $G_1, ..., G_n \geq 0$. Dann gilt auf A^c

$$\begin{aligned}
(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_t &= \sum_{i=1}^n G_i([\mathcal{X}, \mathcal{Y}]_{t\wedge W_i} - [\mathcal{X}, \mathcal{Y}]_{t\wedge V_i}) \\
&\leq \sum_{i=1}^n G_i \sqrt{[\mathcal{X}]_{t\wedge W_i} - [\mathcal{X}]_{t\wedge V_i}} \sqrt{[\mathcal{Y}]_{t\wedge W_i} - [\mathcal{Y}]_{t\wedge V_i}} \\
&\leq \sqrt{\sum_{i=1}^n G_i^2([\mathcal{X}]_{t\wedge W_i} - [\mathcal{X}]_{t\wedge V_i})} \sqrt{\sum_{k=1}^n [\mathcal{Y}]_{t\wedge W_k} - [\mathcal{Y}]_{t\wedge V_k}} \\
&= (\mathcal{H}^2 \cdot [\mathcal{X}])_t^{1/2} [\mathcal{Y}]_t^{1/2}.
\end{aligned}$$

Der Fall für progressiv messbare Prozesse $\mathcal{H}$ ergibt sich aus der letzten Rechnung zunächst durch Approximation messbarer Mengen anstelle der Intervalle $I_1, ..., I_n$. Anschließend erreicht man durch monotoner Konvergenz eine Approximation von $\mathcal{H}$, woraus sich die Aussage durch monotone Konvergenz ergibt. □

Nun beschäftigen wir uns mit dem stochastischen Integral mit lokalen Martingalen als Integratoren.

Definition 19.34 (Stochastisches Integral von previsiblen, einfachen Prozessen). *Sei $\mathcal{X}$ ein lokales Martingal und $\mathcal{H}$ ein einfacher previsibler Prozess wie in Definition 19.1. Dann definieren wir das stochastische Integral $\mathcal{H} \cdot \mathcal{X}$ genauso wie in Definition 19.19.*

Lemma 19.35 (Stochastisches Integral für einfache previsible Prozesse). *Seien $\mathcal{X}, \mathcal{Y}$ lokale Martingale mit stetigen Pfaden, sowie $\mathcal{H} \in \mathbb{S}$. Dann ist $\mathcal{H} \cdot \mathcal{X}$ wieder ein lokales Martingal und es gilt*

$$[\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}]. \tag{19.6}$$

Beweis. Zunächst ist $\mathcal{H} \cdot \mathcal{X}$ ein lokales Martingal (mit derselben lokalisierenden Folge von Stoppzeiten wie für $\mathcal{X}$). Wir müssen zeigen, dass $(\mathcal{H} \cdot \mathcal{X})\mathcal{Y} - \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}]$ ein lokales Martingal ist. Durch Linearität genügt es, $H_t = G(1_{T\wedge t} - 1_{S\wedge t})$ für $\mathcal{F}_S$-messbares G zu betrachten. Da $\mathcal{X}\mathcal{Y} - [\mathcal{X}, \mathcal{Y}]$ ein lokales Martingal ist, folgern wir mit wiederholter Anwendung des Optional Sampling Theorems

$$\begin{aligned}
&\mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t Y_t - (\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_t | \mathcal{F}_s] \\
&\quad = \mathbf{E}\big[G(X_{T\wedge t} Y_t - [\mathcal{X}, \mathcal{Y}]_{T\wedge t} - (X_{S\wedge t} Y_t - [\mathcal{X}, \mathcal{Y}]_{S\wedge t})) | \mathcal{F}_s\big] \\
&\quad = G 1_{T\leq s}\big(X_{T\wedge s} Y_s - [\mathcal{X}, \mathcal{Y}]_{T\wedge s} - (X_{S\wedge s} Y_s - [\mathcal{X}, \mathcal{Y}]_{S\wedge s})\big) \\
&\qquad + G 1_{S\leq s\leq T} \mathbf{E}[X_{T\wedge t} Y_{T\wedge t} - [\mathcal{X}, \mathcal{Y}]_{T\wedge t} - (X_{S\wedge t} Y_s - [\mathcal{X}, \mathcal{Y}]_{S\wedge t})) | \mathcal{F}_s] \\
&\qquad + \mathbf{E}\big[G 1_{s<S\leq t} \mathbf{E}[X_{T\wedge t} Y_{T\wedge t} - [\mathcal{X}, \mathcal{Y}]_{T\wedge t} - (X_{S\wedge t} Y_t - [\mathcal{X}, \mathcal{Y}]_{S\wedge t})) | \mathcal{F}_{S\vee s}] | \mathcal{F}_s\big] \\
&\quad = G 1_{S\leq s}\big(X_{T\wedge s} Y_s - [\mathcal{X}, \mathcal{Y}]_{T\wedge s} - (X_{S\wedge s} Y_s - [\mathcal{X}, \mathcal{Y}]_{S\wedge s})\big) \\
&\quad = G\big((X_{T\wedge s} - X_{S\wedge s}) Y_s - ([\mathcal{X}, \mathcal{Y}]_{T\wedge s} - [\mathcal{X}, \mathcal{Y}]_{S\wedge s})\big) \\
&\quad = (\mathcal{H} \cdot \mathcal{X})_s Y_s - (\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_s.
\end{aligned}$$

Mit Lemma 15.22 folgt die Aussage. □

Um nun das stochastische Integral von Integranden in $\mathbb{S}$ zu progressiven stochastischen Prozessen zu erweitern, nehmen wir (19.6) als definierende Eigenschaft.

Theorem 19.36 (Allgemeines stochastisches Integral). *Sei $\mathcal{X}$ ein lokales Martingal mit stetigen Pfaden und $\mathcal{H}$ ein progressiv messbarer stochastischer Prozess mit $(\mathcal{H}^2 \cdot [\mathcal{X}])_t < \infty$ für alle $t \geq 0$. Dann gibt es ein fast sicher eindeutiges lokales Martingal $\mathcal{H} \cdot \mathcal{X}$ mit $(\mathcal{H} \cdot \mathcal{X})_0 = 0$, so dass für jedes lokale Martingal $\mathcal{Y}$ mit stetigen Pfaden*

$$[\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}]$$

gilt.

Beweis. Zuerst zeigen wir die Eindeutigkeit. Wäre der Prozess $\mathcal{H} \cdot \mathcal{X}$ nicht eindeutig, so gäbe zwei Prozesse $\mathcal{Z}', \mathcal{Z}''$, so dass $[\mathcal{Z}', \mathcal{Y}] = [\mathcal{Z}'', \mathcal{Y}] = [\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}]$ für alle lokale Martingale $\mathcal{Y}$ mit stetigen Pfaden. Dann gilt wegen der Linearität der Covariation mit $\mathcal{Y} = \mathcal{Z}' - \mathcal{Z}''$, dass $[\mathcal{Z}' - \mathcal{Z}''] = 0$. Aus Proposition 19.28 folgt, dass dann $\mathcal{Z}' = \mathcal{Z}''$.

Für die Existenz von $\mathcal{H} \cdot \mathcal{X}$ genügt es wie im Beweis von Theorem 19.29, den Fall $\mathbf{E}[(\mathcal{H}^2 \cdot [\mathcal{X}])_\infty] < \infty$ zu betrachten. (Andernfalls definieren wir $T_n := \inf\{t \geq 0 : (\mathcal{H}^2 \cdot \mathcal{X})_t \leq n\}$. Nach Voraussetzung ist dann $T_n \uparrow \infty$. Dann gibt es stetige lokale Martingale $\mathcal{H} \cdot \mathcal{X}^{T_n}$, so dass $[\mathcal{H} \cdot \mathcal{X}^{T_n}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X}^{T_n}, \mathcal{Y}]$ für alle stetigen lokalen Martingale $\mathcal{Y}$ und $n = 1, 2, ...$ Weiter muss $\mathcal{H} \cdot \mathcal{X}^{T_n} = \mathcal{H} \cdot \mathcal{X}^{T_m}$ für $m \geq n$ auf $[0, T_n]$ gelten, da auf derselben Menge auch $\mathcal{X}^{T_n} = \mathcal{X}^{T_m}$ gilt. Nun können wir $\mathcal{H} \cdot \mathcal{X} = \lim_{n\to\infty} \mathcal{H} \cdot \mathcal{X}^{T_n}$ definieren. Dann gilt $(\mathcal{H} \cdot \mathcal{X})^{T_n} = \mathcal{H} \cdot \mathcal{X}^{T_n}$. Da letzteres für jedes n ein lokales Martingal ist, ist nach Lemma 19.27 auch $\mathcal{H} \cdot \mathcal{X}$ ein lokales Martingal.)

Wir müssen nun noch zeigen, dass $(\mathcal{H} \cdot \mathcal{X})\mathcal{Y} - \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}]$ ein lokales Martingal ist. Zunächst folgt mit Proposition 19.33 und der Cauchy-Schwartz-Ungleichung, dass für $\mathcal{Y} \in \mathcal{M}^2$

$$|\mathbf{E}[(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_\infty]| \leq \mathbf{E}[|\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])|_\infty] \leq (\mathbf{E}[(\mathcal{H}^2 \cdot [\mathcal{X}])_\infty])^{1/2} (\mathbf{E}[[\mathcal{Y}]_\infty])^{1/2} < \infty.$$

Damit ist $\mathcal{Y} \mapsto \mathbf{E}[(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_\infty]$ ein stetiges, lineares Funktional auf $\mathcal{M}^2$. Nach dem Satz von Riesz-Fréchet[22] (angewandt auf den Hilbert-Raum $\mathcal{M}^2$) gibt es also einen eindeutig bestimmten Prozess in $\mathcal{M}^2$, den wir $\mathcal{H} \cdot \mathcal{X}$ nennen, so dass

$$\mathbf{E}[(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_\infty] = \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_\infty \mathcal{Y}_\infty]$$

für alle $\mathcal{Y} \in \mathcal{M}^2$ gilt. Aus dieser Gleichung folgt auch, dass $\mathcal{H} \mapsto \mathcal{H} \cdot \mathcal{X}$ stetig ist, und die Definition von $\mathcal{H} \cdot \mathcal{X}$ für $\mathcal{H} \in \mathbb{S}$ aus Lemma 19.35 erweitert. Weiter gilt mit Theorem 19.31 für eine Stoppzeit T

$$\mathbf{E}[(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_T] = \mathbf{E}[(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}^T])_\infty] = \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_\infty \mathcal{Y}_\infty^T] = \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_T \mathcal{Y}_T].$$

Aus Lemma 15.22 folgt nun, dass $(\mathcal{H} \cdot \mathcal{X})\mathcal{Y} - \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}]$ ein Martingal ist. Daraus folgt die Behauptung. □

Korollar 19.37 (Kettenregel). *Sei $\mathcal{H}, \mathcal{X}, \mathcal{Y}$ wir in Theorem 19.36 und $\mathcal{K}$ progressiv messbar mit $(\mathcal{K}^2 \cdot [\mathcal{Y}])_t < \infty$ für alle $t \geq 0$. Dann gilt*

$$[\mathcal{H} \cdot \mathcal{X}, \mathcal{K} \cdot \mathcal{Y}] = (\mathcal{H}\mathcal{K}) \cdot [\mathcal{X}, \mathcal{Y}].$$

Beweis. Da $\mathcal{K} \cdot \mathcal{Y}$ ein lokales Martingal ist, folgt nach Theorem 19.36 sofort, dass

$$[\mathcal{H} \cdot \mathcal{X}, \mathcal{K} \cdot \mathcal{Y}] = \mathcal{K} \cdot [\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot \mathcal{K} \cdot [\mathcal{X}, \mathcal{Y}].$$

Da $[\mathcal{X}, \mathcal{Y}]$ ein Prozess endlicher Variation ist, gelten für die Integration bezüglich $[\mathcal{X}, \mathcal{Y}]$ dieselben Rechenregeln wie für das Lebesgue-Integral, insbesondere beim Rechnen mit Dichten. Nun folgt die Behauptung aus Lemma 5.13.2. □

Beispiel 19.38 (Quadratische Variation von $X_t^2 - t$). Sei $\mathcal{X} = (X_t)_{t \geq 0}$ eine Brown'sche Bewegung. Nach Beispiel 19.30 gilt, dass für $\mathcal{Y} = (Y_t)_{t \geq 0}$ mit $Y_t = X_t^2 - t$

$$[\mathcal{Y}]_t = 4 \int_0^t X_s^2 ds.$$

Diese Tatsache können wir mit Hilfe des letzten Resultats nochmals nachrechnen: Nach Beispiel 19.22 mit $f(x) = x^2$ gilt, dass

$$2 \int_0^t X_s dX_s = X_t^2 - t$$

gilt. Deshalb schreiben wir

$$[\mathcal{Y}] = [2\mathcal{X} \cdot \mathcal{X}] = 4\mathcal{X}^2 \cdot [\mathcal{X}] = \Big(4 \int_0^t X_s^2 ds\Big)_{t \geq 0}.$$

Zum Abschluss beweisen wir nun eine Stetigkeitseigenschaft des stochastischen Integrals.

Proposition 19.39 (Stetigkeit des stochastischen Integrals). *Seien $\mathcal{X}^1, \mathcal{X}^2, ...$ stetige lokale Martingale und $\mathcal{H}^1, \mathcal{H}^2, ...$ progressiv messbar mit $((\mathcal{H}^n)^2 \cdot [\mathcal{X}^n])_t < \infty$, $t \geq 0, n = 1, 2, ...$ Dann gilt $\sup_{t \geq 0} |(\mathcal{H}^n \cdot \mathcal{X}^n)_t| \xrightarrow{n \to \infty}_p 0$ genau dann, wenn $((\mathcal{H}^n)^2 \cdot [\mathcal{X}^n]) \xrightarrow{n \to \infty}_p 0$.*

Beweis. Nach Korollar 19.37 ist $[\mathcal{H}^n, \mathcal{X}^n] = (\mathcal{H}^n)^2 \cdot [\mathcal{X}^n]$. Nun genügt es, Proposition 19.32 anzuwenden. □

[22] Satz von Riesz-Fréchet: Sei H ein Hilbert-Raum (mit Skalarprodukt $\langle ., . \rangle$) über $\mathbb{R}$ und H' der Raum der linearen, stetigen Abbildungen $H \to \mathbb{R}$. Dann lässt sich $x' \in H'$ schreiben als: $x'(x) = \langle y, x \rangle$ für ein geeignetes $y \in H$.

19.5 Rechenregeln für stochastische Integrale

Es folgen nun wichtige Regeln zum Rechnen mit stochastischen Integralen. Die wichtigsten sind die partielle Integration (Theorem 19.48) und die Itô-Formel (Theorem 19.51). Um jedoch zunächst einen möglichst allgemeinen Rahmen abzustecken, führen wir die Klasse der stetigen Semimartingale ein. Ein Semimartingal $\mathcal{X}$ ist dabei die Summe eines Prozesses von lokal endlicher Variation $\mathcal{A}$ und einem lokalen Martingal $\mathcal{M}$. Integrale bezüglich stetiger Semimartingale sind dann definiert durch die Summe der Integrale bezüglich $\mathcal{A}$ und $\mathcal{M}$.

Definition 19.40 (Semimartingal). 1. *Ein adaptierter Prozess $\mathcal{X}$ heißt Semimartingal, wenn er rechtsstetige Pfade hat und sich als $\mathcal{X} = \mathcal{A} + \mathcal{M}$ schreiben lässt, wobei $\mathcal{A}$ mit $A_0 = 0$ ein Prozess mit lokal endlicher Variation und $\mathcal{M}$ ein lokales Martingal ist.*

2. *Ein stetiges Semimartingal $\mathcal{X}$ ist ein Semimartingal, bei dem $\mathcal{A}$ und $\mathcal{M}$ als stetige Prozesse gewählt werden können. Die Zerlegung $\mathcal{X} = \mathcal{A} + \mathcal{M}$ heißt dann kanonische Zerlegung.*

Lemma 19.41 (Kanonische Zerlegung eindeutig). *Die kanonische Zerlegung eines stetigen Semimartingals ist eindeutig.*

Beweis. Sei $\mathcal{X} = \mathcal{A} + \mathcal{M} = \mathcal{A}' + \mathcal{M}'$ zwei Zerlegungen des stetigen Semimartingals $\mathcal{X}$. Dann gilt $\mathcal{A} - \mathcal{A}' = \mathcal{M} - \mathcal{M}'$. Da die rechte Seite ein stetiges lokales Martingal ist mit $M_0 - M_0' = 0$, gilt nach Theorem 19.26, dass $\mathcal{M} = \mathcal{M}'$. Damit folgt auch $\mathcal{A} = \mathcal{A}'$. □

Beispiel 19.42 (Zerlegung von Semimartingalen nicht endeutig). Zwar ist die Zerlegung $\mathcal{X} = \mathcal{A} + \mathcal{M}$ für stetige Semimartingale eindeutig, nicht jedoch für Semimartingale. Als einfaches Gegenbeispiel sei $\mathcal{X} = (X_t)_{t\geq 0}$ der Poisson-Prozess (mit Parameter λ). Dann lässt sich $\mathcal{X}$ schreiben als

$$X_t = X_t + 0 \qquad \text{und} \qquad X_t = t + (X_t - t).$$

Hierbei sind $A_t = X_t, M_t = 0$ und $A_t' = t, M_t' = X_t - t$ zwei unterschiedliche Zerlegungen von $\mathcal{X}$.

Beispiel 19.43 (Funktionale von stetigen Markov-Prozessen). Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Markov-Prozess mit Zustandsraum E und stetigen Pfaden. Der Generator von $\mathcal{X}$ ist $G^{\mathcal{X}}$ mit $\mathcal{D}(G^{\mathcal{X}}) \subseteq \mathcal{C}_b(E)$. Dann ist $(f(X_t))_{t\geq 0}$ für jedes $f \in \mathcal{D}(G)$ ein Semimartingal. Man kann nämlich wegen Theorem 16.30

$$f(X_t) = A_t + M_t \text{ mit } A_t = \int_0^t (G^{\mathcal{X}} f)(X_s)ds,\ M_t = f(X_t) - \int_0^t (G^{\mathcal{X}} f)(X_s)ds$$

schreiben, was eine Zerlegung in den Prozess mit lokal endlicher Variation $\mathcal{A}$ und das (lokale) Martingal $\mathcal{M}$ ist.

Definition 19.44 (Stetige Semimartingale als Integratoren). *Sei $\mathcal{X} = \mathcal{A} + \mathcal{M}$ ein stetiges Semimartingal. Wie definieren*

$$L(\mathcal{A}) := \{\mathcal{H} \text{ progressiv} : \mathcal{H} \cdot \mathcal{A} \text{ existiert}\}$$

sowie

$$L(\mathcal{M}) := \{\mathcal{H} : \mathcal{H}^2 \in L([\mathcal{M}])\}$$

und $L(\mathcal{X}) := L(\mathcal{A}) \cap L(\mathcal{M})$. Für $\mathcal{H} \in L(\mathcal{X})$ setzen wir

$$\mathcal{H} \cdot \mathcal{X} := \mathcal{H} \cdot \mathcal{A} + \mathcal{H} \cdot \mathcal{M}.$$

Wir definieren die quadratische Variation von $\mathcal{X}$ als

$$[\mathcal{X}] := [\mathcal{M}].$$

Ist $\mathcal{Y} = \mathcal{B} + \mathcal{N}$ die kanonische Zerlegung eines weiteres stetigen Semimartingal, setzen wir die Covariation

$$[\mathcal{X}, \mathcal{Y}] := [\mathcal{M}, \mathcal{N}].$$

Lemma 19.45 (Covariation des stochastischen Integrals). *Seien $\mathcal{X} = \mathcal{A} + \mathcal{M}$ und $\mathcal{Y} = \mathcal{B} + \mathcal{N}$ die kanonischen Zerlegungen von den stetigen Semimartingalen $\mathcal{X}$ und $\mathcal{Y}$. Dann gilt*

$$4[\mathcal{X}, \mathcal{Y}] = [\mathcal{X} + \mathcal{Y}] - [\mathcal{X} - \mathcal{Y}].$$

Ist $\mathcal{H} \in L(\mathcal{X})$, dann ist $\mathcal{H} \cdot \mathcal{X} = \mathcal{H} \cdot \mathcal{A} + \mathcal{H} \cdot \mathcal{M}$ die kanonische Zerlegung des Semimartingales $\mathcal{H} \cdot \mathcal{X}$. Außerdem gilt

$$[\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}].$$

Beweis. Alle Aussagen folgen direkt aus der Definition von $\mathcal{H} \cdot \mathcal{X}$ und $[\mathcal{X}]$ sowie $[\mathcal{X}, \mathcal{Y}]$. □

Wir wiederholen eine einfache Eigenschaft des Integrals bezüglich eines Maßes mit Dichte. Ist nämlich μ ein σ-endliches Maß, $\nu = g \cdot \mu$ das Maß mit Dichte g bezüglich μ und f eine (beschränkte) messbare Funktion. Dann gilt $\nu[f] = g \cdot \mu[f] = \mu[fg]$. Diese Aussage hat ein Analogon für die stochastische Integration bezüglich Prozessen mit lokal endlicher Variation. Hier ein einfaches Beispiel: Sind $\mathcal{H} = (H_t)_{t \geq 0}, \mathcal{K} = (K_t)_{t \geq 0}$ progressiv messbare Prozesse, dann gilt

$$\mathcal{H} \cdot \Big(\int_0^t K_s ds \Big)_{t \geq 0} = \Big(\int_0^t H_s K_s ds \Big)_{t \geq 0}.$$

Dieses Beispiel wird nun wesentlich erweitert.

Proposition 19.46 (Kettenregel). *Sei $\mathcal{X} = (X_t)_{t \geq 0}$ ein stetiges Semimartingal, $\mathcal{K} \in L(\mathcal{X})$. Dann ist $\mathcal{H} \in L(\mathcal{K} \cdot \mathcal{X})$ genau dann, wenn $\mathcal{H}\mathcal{K} \in L(\mathcal{X})$. In diesem Fall gilt*

$$\mathcal{H} \cdot (\mathcal{K} \cdot \mathcal{X}) = \mathcal{H}\mathcal{K} \cdot \mathcal{X}$$

Beweis. Sei $\mathcal{X} = \mathcal{A} + \mathcal{M}$ die kanonische Zerlegung von $\mathcal{X}$. Klar ist $\mathcal{H}(\mathcal{K} \cdot \mathcal{A}) = (\mathcal{H}\mathcal{K}) \cdot \mathcal{A}$, was genau wie im Beweis von Korollar 19.37 aus Lemma 5.13.2. folgt, falls eine der beiden Seiten existiert. Nun bemerken wir, dass $\mathcal{H} \in L(\mathcal{K} \cdot \mathcal{M})$ genau dann, wenn $\mathcal{H}^2 \in L(\mathcal{K}^2 \cdot [\mathcal{M}])$, was genau dann der Fall ist, wenn $\mathcal{H}^2\mathcal{K}^2 \in L([\mathcal{M}])$, also $\mathcal{H}\mathcal{K} \in L(\mathcal{M})$. Analog gilt $\mathcal{H} \in L(\mathcal{K} \cdot \mathcal{A})$ genau dann, wenn $\mathcal{H} \cdot \mathcal{K} \in L(\mathcal{A})$, also $\mathcal{H}\mathcal{K} \in L(\mathcal{A})$. Daraus folgt die behauptete Äquivalenz.

Wir müssen nun noch zeigen, dass $\mathcal{H} \cdot (\mathcal{K} \cdot \mathcal{M}) = (\mathcal{H}\mathcal{K}) \cdot \mathcal{M}$. Hierzu bemerken wir, dass für ein beliebiges lokale Martingal $\mathcal{N}$

$$[(\mathcal{H}\mathcal{K}) \cdot \mathcal{M}, \mathcal{N}] = (\mathcal{H}\mathcal{K}) \cdot [\mathcal{M}, \mathcal{N}] = \mathcal{H} \cdot (\mathcal{K} \cdot [\mathcal{M}, \mathcal{N}]) = \mathcal{H} \cdot [\mathcal{K} \cdot \mathcal{M}, N] = [\mathcal{H} \cdot (\mathcal{K} \cdot \mathcal{M}), \mathcal{N}].$$

Daraus folgt, dass $\mathcal{H} \cdot (\mathcal{K} \cdot \mathcal{M}) - (\mathcal{H}\mathcal{K}) \cdot \mathcal{M}$ ein lokales Martingal mit quadratischer Variation 0 ist. Also folgt die Behauptung aus Theorem 19.26. □

In der Integrationstheorie spielte der Satz von der majorisierten Konvergenz eine wichtige Rolle. Dieser bekommt nun ein Analogon für stochastische Integrale.

Proposition 19.47 (Majorisierte Konvergenz für stochastische Integrale). *Sei $\mathcal{X}$ ein stetiges Semimartingal und $\mathcal{H}, \mathcal{K} = (K_t)_{t\geq 0}, \mathcal{K}^1 = (K_t^1)_{t\geq 0}, \mathcal{K}^2 = (K_t^2)_{t\geq 0}, \ldots \in L(\mathcal{X})$ mit $|\mathcal{K}^n| \leq \mathcal{H}$ und $\sup_{t\geq 0} |K_n(t) - K(t)| \xrightarrow{n\to\infty}_{fs} 0$. Dann gilt $\sup_{0\leq s\leq t} |(\mathcal{K}^n \cdot \mathcal{X} - \mathcal{K} \cdot \mathcal{X})_s| \xrightarrow{n\to\infty}_{p} 0$ für alle $t \geq 0$.*

Beweis. Sei $\mathcal{X} = \mathcal{A} + \mathcal{M}$ die kanonische Zerlegung von $\mathcal{X}$. Da $\mathcal{H} \in L(\mathcal{X})$, gilt $H \in L(\mathcal{A})$ und $H^2 \in L([\mathcal{M}])$. Da wir den Satz von der majorisierten Konvergenz auf Stieltjes-Integrale übertragen können, folgt $((\mathcal{K}^n - \mathcal{K})^2 \cdot [\mathcal{M}])_t \xrightarrow{n\to\infty}_{fs} 0$ für $t \geq 0$. Dies impliziert (durch Stoppen bei t) nach Proposition 19.39, dass $\sup_{0\leq s\leq t} |(\mathcal{K}^n \cdot \mathcal{M} - \mathcal{K} \cdot \mathcal{M})_s| \xrightarrow{n\to\infty}_{p} 0$. Weiter ist $((\mathcal{K}^n - \mathcal{K}) \cdot \mathcal{A})_t \xrightarrow{n\to\infty}_{fs} 0$, wieder nach dem Satz von der majorisierten Konvergenz für Stieltjes-Integrale. Aus den letzten beiden Konvergenzen folgt die Aussage. □

Es folgt eine weitere Regel der Lebesgue-Integration: Ist f messbar und lokal beschränkt und λ das ein-dimensionale Lebesgue-Integral, dann ist $f \cdot \lambda$ von lokal beschränkter Variation. Es gilt wegen Fubini

$$((f \cdot \lambda) \cdot (f \cdot \lambda))_t = ((f(f \cdot \lambda)) \cdot \lambda)_t = \int_0^t f(s) \int_0^s f(r) dr ds = \tfrac{1}{2}(f \cdot \lambda)_t^2,$$

also $2(f \cdot \lambda) \cdot (f \cdot \lambda) = (f \cdot \lambda)^2$. Analog folgern wir, dass für einen Prozess mit lokal endlicher Variation $\mathcal{A}$ gilt, dass

$$2\mathcal{A} \cdot \mathcal{A} = \mathcal{A}^2. \tag{19.7}$$

Dieses Resultat werden wir nun verallgemeinern.

Theorem 19.48 (Partielle Integration). *Seien $\mathcal{X}$ und $\mathcal{Y}$ stetige Semimartingale. Dann gilt*

$$\mathcal{X}\mathcal{Y} = X_0 Y_0 + \mathcal{X} \cdot \mathcal{Y} + \mathcal{Y} \cdot \mathcal{X} + [\mathcal{X}, \mathcal{Y}].$$

Beweis. Wir zeigen zunächst nur den Fall $\mathcal{X} = \mathcal{Y}$, also $\mathcal{X}^2 = X_0^2 + 2\mathcal{X} \cdot \mathcal{X} + [\mathcal{X}]$. Zunächst betrachten wir $\mathcal{X} \in \mathcal{M}^2$ und setzen $T_0^n, T_1^n, \ldots$ und $\mathcal{Q}^n$ wie in Proposition 19.28, sowie $\mathcal{H}^n = (H_t^n)_{t\geq 0} \in \mathbb{S}$ mittels

$$H_t^n = \sum_{k=0}^{\infty} X_{T_k^n} 1_{(T_k^n, T_{k+1}^n]}(t).$$

Genau wie in Propositin 19.28 gilt dann $X_t^2 - X_0^2 = Q_t^n + 2(\mathcal{H}^n \cdot \mathcal{X})_t$. und $\sup_{t\geq 0} |H_t^n - X_t| \leq 2^{-n} \xrightarrow{n\to\infty 0}$. Wegen Proposition 19.47 gilt also

$$2(\mathcal{X} \cdot \mathcal{X})_t = \lim_{n\to\infty} 2(H^n \cdot \mathcal{X})_t = X_t^2 - X_0^2 - \lim_{n\to\infty} Q_t^n = X_t^2 - X_0^2 - [\mathcal{X}]_t.$$

Ist $\mathcal{X}$ ein lokales Martingal, so folgt die Aussage wie für $\mathcal{X} \in \mathcal{M}^2$ durch geeignetes Stoppen von $\mathcal{X}$. Ist andererseits $\mathcal{X} = \mathcal{A}$ ein Prozess mit lokal endlicher Variation, so folgt die Aussage wie in (19.7).

Betrachten wir nun den Fall eines Semimartingals $\mathcal{X}$ mit kanonischer Zerlegung $\mathcal{X} = \mathcal{A} + \mathcal{X}$. Dann ist die Behauptung äquivalent zu

$$\mathcal{M}^2 + 2\mathcal{M}\mathcal{A} + \mathcal{A}^2 = 2\mathcal{M} \cdot \mathcal{M} + 2\mathcal{M} \cdot \mathcal{A} + 2\mathcal{A} \cdot \mathcal{M} + 2\mathcal{A} \cdot \mathcal{A} + [\mathcal{M}].$$

Wir müssen also zeigen, dass

$$\mathcal{M}\mathcal{A} = \mathcal{A} \cdot \mathcal{M} + \mathcal{M} \cdot \mathcal{A}.$$

Wir definieren nun für $t \geq 0$ und $n = 1, 2, ...$ die Prozesse $\mathcal{A}^n = (A_s^n)_{0 \leq s \leq t}$ und $\mathcal{M}^n = (M_s^n)_{0 \leq s \leq t}$ durch

$$A_s^n = A_{(k-1)t/n} \text{ und } M_s^n = M_{kt/n} \text{ für } s \in t(k-1, k]/n.$$

Dann gilt

$$\begin{aligned}
(\mathcal{A}^n \cdot \mathcal{M})_t + (\mathcal{M}^n \cdot \mathcal{A})_t &= \sum_{k=1}^n A_{(k-1)t/n}(M_{kt/n} - M_{(k-1)t/n}) + M_{kt/n}(A_{kt/n} - A_{(k-1)t/n}) \\
&= \sum_{k=1}^n M_{kt/n}A_{kt/n} - M_{(k-1)t/n}A_{(k-1)t/n} = A_t M_t.
\end{aligned}$$

Aus majorisierter Konvergenz für Stieltjes-Integrale (im Term $(\mathcal{M}^n \cdot \mathcal{A})_t$) und aus Proposition 19.47 (im Term $(\mathcal{A}^n \cdot \mathcal{M})_t$) folgt nun die Aussage im Fall $\mathcal{X} = \mathcal{Y}$. Im Fall $\mathcal{X} \neq \mathcal{Y}$ schreiben wir

$$\begin{aligned}
4\mathcal{X}\mathcal{Y} &= (\mathcal{X} + \mathcal{Y})^2 - (\mathcal{X} - \mathcal{Y})^2 \\
&= (X_0 + Y_0)^2 + 2(\mathcal{X} + \mathcal{Y}) \cdot (\mathcal{X} + \mathcal{Y}) + [\mathcal{X} + \mathcal{Y}] - (X_0 - Y_0)^2 - 2(\mathcal{X} - \mathcal{Y}) \cdot (\mathcal{X} - \mathcal{Y}) - [\mathcal{X} - \mathcal{Y}] \\
&= 4X_0Y_0 + 4\mathcal{X} \cdot \mathcal{Y} + 4\mathcal{Y} \cdot \mathcal{X} + 4[\mathcal{X}, \mathcal{Y}]
\end{aligned}$$

und die Aussage ist bewiesen. □

Beispiel 19.49 (Von der Brown'schen Bewegung abgeleitete Semimartingale). Sei $\mathcal{X} = (X_t)_{t \geq 0}$ eine Brown'sche Bewegung. Dann haben wir in Beispiel 19.30 gesehen, dass $2\int_0^t X_s dX_s = X_t^2 - t$ gilt. Dies ist auch genau die Formel der partiellen Integration (wenn man beachtet, dass $[\mathcal{X}]_t = t$).

In Beispiel 19.30 haben wir gesehen, dass für eine Brown'sche Bewegung $\mathcal{X} = (X_t)_{t \geq 0}$ der Prozess

$$\Big(tX_t^2 - \tfrac{1}{2}t^2 - \int_0^t X_s^2 ds\Big)_{t \geq 0}$$

ein Martingal ist. Dies können wir auch mittels partieller Integration einsehen. Es ist nämlich für $\mathcal{Y} = (Y_t)_{t \geq 0}$ und $Y_t = X_t^2 - t$

$$\begin{aligned}
tX_t^2 &= \Big(\mathrm{id} \cdot \mathcal{X}^2 + \mathcal{X}^2 \cdot \mathrm{id}\Big)_t \\
&= \Big(\mathrm{id} \cdot \mathcal{Y} + \mathrm{id} \cdot \mathrm{id} + \mathcal{X}^2 \cdot \mathrm{id}\Big)_t \\
&= \int_0^t s dY_s + \int_0^t s ds + \int_0^t X_s^2 ds
\end{aligned}$$

Also ist obiges Martingal identisch mit $\mathrm{id} \cdot \mathcal{Y}$.

Lemma 19.50 (Die Covariation). *Seien $\mathcal{X}$ und $\mathcal{Y}$ stetige Semimartingale. Für $t > 0$ sei $\zeta_n := \{t_{n,0}, ..., t_{n,k_n}\}$ eine Partition von $[0, t]$. Gilt $\max_k |t_{n,k} - t_{n,k-1}| \xrightarrow{n\to\infty} 0$, so ist*

$$Z_n := \sum_{k=1}^{k_n} (X_{t_{n,k}} - X_{t_{n,k-1}})(Y_{t_{n,k}} - Y_{t_{n,k-1}}) \xrightarrow{n\to\infty}_p [\mathcal{X}, \mathcal{Y}]_t.$$

Beweis. Ähnlich wie im letzten Beweis definieren wir

$$X_s^n = X_{t_{n,k-1}} \text{ und } Y_s^n = Y_{t_{n,k-1}} \text{ für } s \in (t_{n,k-1}, t_{n,k}].$$

Dann gilt

$$\begin{aligned} X_t Y_t &= \sum_{k=1}^{n} X_{t_{n,k}} Y_{t_{n,k}} - X_{t_{n,k-1}} Y_{t_{n,k-1}} \\ &= \sum_{k=1}^{n} X_{t_{n,k-1}} (Y_{t_{n,k}} - Y_{t_{n,k-1}}) + Y_{t_{n,k-1}} (X_{t_{n,k}} - X_{t_{n,k-1}}) \\ &\qquad\qquad + (X_{t_{n,k}} - X_{t_{n,k-1}})(Y_{t_{n,k}} - Y_{t_{n,k-1}}) \\ &= (\mathcal{X}^n \cdot \mathcal{Y})_t + (\mathcal{Y}^n \cdot \mathcal{X})_t + Z_n. \end{aligned}$$

Da $(\mathcal{X}^n \cdot \mathcal{Y})_t \xrightarrow{n\to\infty}_p (\mathcal{X} \cdot \mathcal{Y})_t$ und $(\mathcal{Y}^n \cdot \mathcal{X})_t \xrightarrow{n\to\infty}_p (\mathcal{Y} \cdot \mathcal{X})_t$ nach Proposition 19.47, folgt mit partieller Integration

$$\lim_{n\to\infty} Z_n = X_t Y_t - (\mathcal{X} \cdot \mathcal{Y})_t - (\mathcal{Y} \cdot \mathcal{X})_t = [\mathcal{X}, \mathcal{Y}]_t.$$

□

Theorem 19.51 (Itô-Formel). *Seien $\mathcal{X}^1, ..., \mathcal{X}^d$ stetige Semimartingale und $f \in \mathcal{C}^2(\mathbb{R}^d)$. Mit $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$, $\underline{X}_t = (X_t^1, ..., X_t^d)$ und $f(\underline{\mathcal{X}}) = (f(\underline{X}_t))_{t\geq 0}$ ist*

$$f(\underline{\mathcal{X}}) = f(\underline{X}_0) + \sum_{i=1}^{d} f_i(\underline{\mathcal{X}}) \cdot \mathcal{X}^i + \tfrac{1}{2} \sum_{i,j=1}^{d} f_{ij}(\underline{\mathcal{X}}) \cdot [\mathcal{X}^i, \mathcal{X}^j]$$

wobei $f_i = \partial f / \partial x_i$ und $f_{ij} = \partial^2 f / \partial x_i \partial x_j$. Ist insbesondere $d = 1$, so gilt

$$f(\mathcal{X}) = f(X_0) + f'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2} f''(\mathcal{X}) \cdot [\mathcal{X}]. \tag{19.8}$$

Beweis. Wir zeigen nur den Fall $d = 1$. Der allgemeine Fall lässt sich analog beweisen. Sei $\mathcal{C} \subseteq \mathcal{C}_b(\mathbb{R})$ die Klasse der Funktionen, für die (19.8) gilt. Dann ist $\mathcal{C}$ ein Vektorraum und id$\in \mathcal{C}$. Wir zeigen, dass $\mathcal{C}$ auch unter Multiplikation abgeschlossen ist. Ist nämlich $f, g \in \mathcal{C}$, dann gilt mit Proposition 19.46 und Theorem 19.48, dass

$$\begin{aligned} (fg)(\mathcal{X}) - (fg)(X_0) &= f(\mathcal{X}) \cdot g(\mathcal{X}) + g(\mathcal{X}) \cdot f(\mathcal{X}) + [f(\mathcal{X}), g(\mathcal{X})] \\ &= f(\mathcal{X}) \cdot (g'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2} g''(\mathcal{X}) \cdot [\mathcal{X}]) \\ &\qquad + g(\mathcal{X}) \cdot (f'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2} f''(\mathcal{X}) \cdot [\mathcal{X}]) + [f'(\mathcal{X}) \cdot \mathcal{X}, g'(\mathcal{X}) \cdot \mathcal{X}] \\ &= (fg' + f'g)(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2}(f''g + 2f'g' + fg'')(\mathcal{X}) \cdot [\mathcal{X}] \\ &= (fg)'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2}(fg)''(\mathcal{X}) \cdot [\mathcal{X}]. \end{aligned}$$

Sei nun $f \in \mathcal{C}''(\mathbb{R})$ beliebig und $p_1, p_2, ... \in \mathcal{C}$ Polynome, so dass $\sup_{|x|\leq c} |p_n(x) - f''(x)| \xrightarrow{n\to\infty} 0$ für jedes $c > 0$. Durch Integration erhält man Polynome $f_1, f_2, ...$ mit

$$\sup_{|x|\leq c} |f_n(x) - f(x)| \vee |f_n'(x) - f'(x)| \vee |f_n''(x) - f''(x)| \xrightarrow{n\to\infty} 0.$$

Für die kanonische Zerlegung $\mathcal{X} = \mathcal{A} + \mathcal{M}$ folgt damit mit majorisierter Konvergenz für Stieltjes-Integrale

$$(f_n'(\mathcal{X}) \cdot \mathcal{A} + \tfrac{1}{2} f_n''(\mathcal{X}) \cdot [\mathcal{X}] \xrightarrow{n\to\infty} (f'(\mathcal{X}) \cdot \mathcal{A} + \tfrac{1}{2} f''(\mathcal{X}) \cdot [\mathcal{X}]$$

uniform auf Kompakta. Außerdem gilt

$$((f_n(\mathcal{X}) - f(\mathcal{X}))^2 \cdot [\mathcal{X}])_t \xrightarrow{n\to\infty} 0,$$

also folgt mit Proposition 19.39, dass $f_n(\mathcal{X}) \cdot \mathcal{M} \xrightarrow{n\to\infty}_p f(\mathcal{X}) \cdot \mathcal{M}$ uniform auf Kompakta. Daraus folgt die Behauptung. □

Beispiel 19.52 (Anwendung auf die Brown'sche Bewegung). In Beispiel 19.22 haben wir gesehen, dass für $f' \in \mathcal{C}_b^2(\mathbb{R})$ und eine Brown'sche Bewegung $\mathcal{X} = (X_t)_{t\geq 0}$ gilt, dass

$$f'(\mathcal{X}) \cdot \mathcal{X} = f(\mathcal{X}) - f(X_0) - \tfrac{1}{2} f''(\mathcal{X}) \cdot [\mathcal{X}],$$

da $[\mathcal{X}]_t = t$. Dies ist genau die Itô-Formel, angewandt auf das Semimartingal $\mathcal{X}$.

Bemerkung 19.53 (Itô-Formel als Taylor-Entwicklung). Die Itô-Formel lässt verschiedene Schreiweisen zu. Schreibt man etwa die stochastischen Integrale in (19.8) aus, erhält man

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s + \tfrac{1}{2} \int_0^t f''(X_s) d[\mathcal{X}]_s.$$

Vergleicht man diese Schreibweise mit Theorem 19.14 (das sich mit Prozessen lokal beschränkter Variation beschäftigt), sehen wir einen wichtigen Unterschied im Rechnen mit stetigen Martingalen im Vergleich zu Prozessen mit lokal beschränkter Variation. Man bezeichnet nämlich den Term $\frac{1}{2} \int_0^t f''(X_s) d[\mathcal{X}]_s$ als *Itô-Korrektur-Term*. Diesen erhält man auch, wenn man $f(X_t)$ durch die Taylor-Reihe bis zum zweiten Gied darstellt. In differentieller Schreibweise bedeutet das

$$df(\mathcal{X}) = f'(\mathcal{X}) d\mathcal{X} + \tfrac{1}{2} f''(\mathcal{X}) d[\mathcal{X}].$$

Bemerkung 19.54 (Das Stratonovich-Integral). Neben dem stochastischen Itô-Integral kann man stochastische Integrale auch im Sinne von R. Stratonovich einführen. Dies wird durch Approximation mittels

$$(\mathcal{H} \bullet \mathcal{X})_t = \lim_{n\to\infty} \sum_{i=1}^{k_n} \tfrac{1}{2} (H_{t_{i+1}^n} + H_{t_i^n})(X_{t_{i+1}^n} - X_{t_i^n}) \tag{19.9}$$

für Partitionen $0 = t_0^n < \cdots < t_{k_n}^n$ mit $\max_i |t_{i+1}^n - t_i^n| \xrightarrow{n\to\infty} 0$ eingeführt. Diese Konstruktion ist also ähnlich wie bei der Integration im Sinne von Riemann. Nach Konstruktion des stochastischen Integrals führt dies auf

$$\mathcal{X} \bullet \mathcal{Y} = \mathcal{X} \cdot \mathcal{Y} + \tfrac{1}{2} [\mathcal{X}, \mathcal{Y}]$$

für Semimartingale $\mathcal{X}$ und $\mathcal{Y}$. Analog ergibt sich die Formel

$$f(\mathcal{X}) = f(X_0) + f'(\mathcal{X}) \bullet \mathcal{X}$$

für $f \in \mathcal{C}^2(\mathbb{R})$, was genau die Itô-Formel ohne den Itô-Korrekturterm ist. Es sieht also so aus, als wäre das Stratonovich-Integral die natürliche Erweiterung des Lebesgue-Integrals. Allerdings muss man in (19.9) bei der Integration durch nicht-adaptierte Prozesse approximieren, was wiederum unnatürlich erscheint. In der Stochastik hat sich das Itô-Integral weitestgehend durchgesetzt.

20 Anwendungen der Itô-Formel

Die Itô-Formel gilt als wichtigste Formel der stochastischen Integrationstheorie. Wir werden in diesem Abschnitt einige Anwendungen präsentieren. Vor allem stehen solche im Mittelpunkt, die die Beziehung zwischen allgemeinen stetigen lokalen Martingalen und der Brown'schen Bewegung herstellen. So werden wir in Abschnitt 20.1 sehen, dass mittels einer Zeittransformation jedes stetige lokale Martingal in eine Brown'sche Bewegung übergeführt werden kann (Theorem 20.4), und in Abschnitt 20.2 ebenso als stochastisches Integral bezüglich einer Brown'schen Bewegung (Theorem 20.9). Weiter werden wir in Abschnitt 20.3 nicht-negative Martingale dazu verwenden, einen Maßwechsel durchzuführen, wodurch Semimartingale in Martingale überführt werden können (Theorem 20.14). Schließlich wird in Abschnitt 20.4 der Begriff der Lokalzeit eines Semimartingals eingeführt, der eine Erweiterung von Theorem 17.8 über die Verteilung des Maximums einer Brown'schen Bewegung zulässt.

20.1 Transformationen der Brown'schen Bewegung

Die Itô-Formel stellt eine allgemeine Transformationsformel für stetige lokale Martingale dar. Zentral ist dabei der Begriff der quadratischen Variation. Wir werden zunächst in Theorem 20.3 die Lévy'sche Charakterisierung der Brown'schen Bewegung kennen lernen, die besagt, dass die Brown'sche Bewegung das einzige stetige lokale Martingal $\mathcal{X}$ ist mit $[\mathcal{X}]_t = t$. Dies unterstreicht nicht nur die Bedeutung der Brown'schen Bewegung, sondern öffnet auch das Tor, um allgemeine stetige lokale Martingale mittels der Brown'schen Bewegung darzustellen. Beispielsweise stellt sich heraus, dass jedes stetige lokale Martingal eine zeit-transformierte Brown'sche Bewegung ist (Theorem 20.4). Der Beweis von Theorem 20.3 ist am einfachsten, wenn wir $\mathbb{C}$-wertige lokale Martingale einführen.

Bemerkung 20.1 ($\mathbb{C}$-wertige Martingale). Sei $\mathcal{X} = \mathcal{Y} + i\mathcal{Z}$ ein stochastischer Prozess mit Werten in $\mathbb{C}$, dann nennen wir ihn (lokales) Martingal, wenn sowohl $\mathcal{Y}$ als auch $\mathcal{Z}$ (lokale) Martingale sind.

Lemma 20.2 (Ein exponentielles Martingal). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein lokales Martingal mit $X_0 = 0$. Dann ist $\mathcal{Z} = (Z_t)_{t\geq 0}$ mit*

$$Z_t = \exp(iX_t + \tfrac{1}{2}[\mathcal{X}]_t)$$

ein $\mathbb{C}$-wertiges lokales Martingal und es gilt

$$Z_t = 1 + i(\mathcal{Z} \cdot \mathcal{X})_t$$

fast sicher.

Beweis. Wir wenden die Itô-Formel aus das Semimartingal $(iX_t + \frac{1}{2}[\mathcal{X}]_t)_{t\geq 0}$ und die Funktion $f(z) = e^z$ an. Dies ergibt

$$dZ = Z(idX + \tfrac{1}{2}d[\mathcal{X}] + \tfrac{1}{2}d[i\mathcal{X}]) = iZdX$$

oder $\mathcal{Z} = Z_0 + i(\mathcal{Z} \cdot \mathcal{X})$, woraus die Behauptung folgt. □

Theorem 20.3 (Lévy's Charakterisierung der Brown'schen Bewegung). *Sei* $\underline{\mathcal{X}} = (\underline{X}_t)_{t\geq 0} = (\mathcal{X}^1, ..., \mathcal{X}^d)$, $\mathcal{X}^k = (X_t^k)_{t\geq 0}$ *ein adaptierter stochastischer Prozess mit* $X_0^k = 0, k = 1, ..., d$. *Dann sind äquivalent:*

1. $\underline{\mathcal{X}}$ *ist eine Brown'sche Bewegung, d.h.* $X_t^k - X_s^k$ *ist für alle* k *und* $t - s$ *nach* $N(0, t-s)$*-verteilt und unabhängig von* $\mathcal{F}_s$ *und von* $(\mathcal{X}^l)_{l\neq k}$.

2. $\underline{\mathcal{X}}$ *ist ein stetiges lokales Martingal und*

$$[\mathcal{X}^k, \mathcal{X}^l]_t = \delta_{kl} t.$$

Beweis. 1.⇒2. ist klar nach Beispiel 15.5.
2.⇒1.: Sei $\underline{\gamma} \in \mathbb{R}^d$. Der Prozess $(\langle \underline{\gamma}, \underline{X}_t \rangle)_{t\geq 0}$ ist ein lokales Martingal mit

$$[\langle \underline{\gamma}, \underline{\mathcal{X}} \rangle]_t = \sum_{k=1}^d \gamma_k^2 t = \langle \underline{\gamma}, \underline{\gamma} \rangle t.$$

Nach Lemma 20.2 folgt, dass für jedes $\underline{\gamma} \in \mathbb{R}^d$ der Prozess

$$(\exp(i\langle \underline{\gamma}, \underline{X}_t \rangle + \tfrac{1}{2}\langle \underline{\gamma}, \underline{\gamma} \rangle t))_{t\geq 0} \tag{20.1}$$

ein stetiges Martingal ist. Daraus folgt

$$\mathbb{E}[\exp(i\langle \underline{\gamma}, \underline{X}_t - \underline{X}_s \rangle)|\mathcal{F}_s] = \exp(-\tfrac{1}{2}\langle \underline{\gamma}, \underline{\gamma} \rangle (t-s)),$$

woraus alle Behauptungen folgen. □

Die Bedeutung der Brown'schen Bewegung in der Klasse der lokalen Martingale wird nun noch weiter herausgestellt. Wir zeigen nun, dass wir mittels einer Zeittransformation jedes stetige lokale Martingal in einen Brown'sche Bewegung überführen können.

Theorem 20.4 (Zeit-Transformation, Dubins-Schwartz). *Sei* $\mathcal{Y}$ *ein stetiges lokales Martingal mit* $Y_0 = 0$, *so dass* $[\mathcal{Y}]_t \uparrow \infty$ *mit* $t \to \infty$. *Definiere*

$$T_t := \inf\{u \geq 0 : [\mathcal{Y}]_u \geq t\}, \qquad \mathcal{G}_t := \mathcal{F}_{T_t}, \qquad X_t := Y_{T_t}.$$

Dann ist $\mathcal{X} = (X_t)_{t\geq 0}$ *eine* $(\mathcal{G}_t)_{t\geq 0}$*-adaptierte Brown'sche Bewegung (also insbesondere ein Martingal bezüglich* $(\mathcal{G}_t)_{t\geq 0}$*). Darüberhinaus ist* $[\mathcal{Y}]_t$ *eine* $(\mathcal{G}_t)_{t\geq 0}$*-Stoppzeit und es gilt*

$$Y_t = X_{[\mathcal{Y}]_t}.$$

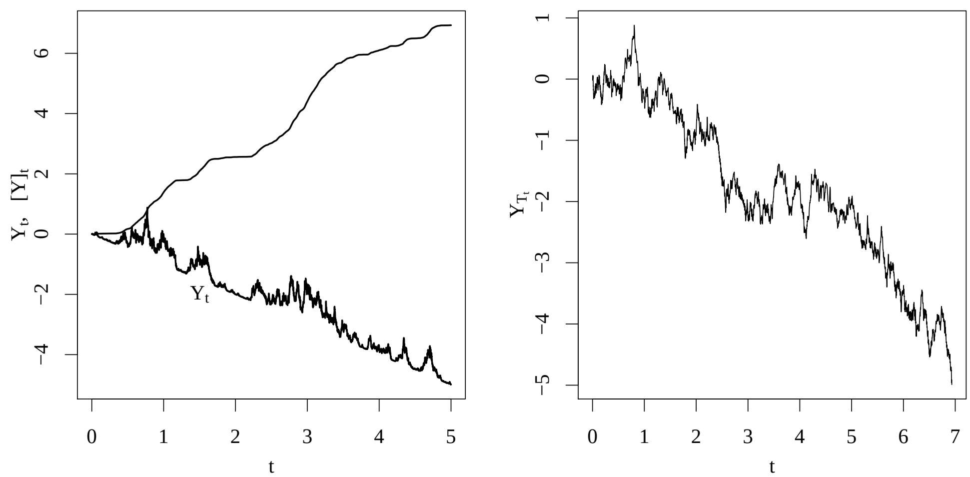


Abbildung 20.1: Skizze zur Zeit-Transformation der Brown'schen Bewegung. Hier ist T_t die erste Zeit, für die die quadratische Variation von $\mathcal{Y}$ gerade t erreicht.

Bemerkung 20.5 (Merkhilfe). Die Beziehungen $X_t := Y_{T_t}$ und $Y_t = X_{[\mathcal{Y}]_t}$ lassen sich mit Hilfe der quadratischen Variation begründen. Schließlich ist, nach Definition von T_t und wegen $[\mathcal{X}]_t = t$ nämlich

$$[\mathcal{X}]_t = t = [\mathcal{Y}]_{T_t}, \qquad [\mathcal{Y}]_t = [\mathcal{X}]_{[\mathcal{Y}]_t}.$$

Beispiel 20.6. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung. Wir wissen nach Beispiel 19.30, dass $\mathcal{Y} = (Y_t)_{t\geq 0}$ mit $Y_t = X_t^2 - t$ ein Martingal ist und $[\mathcal{Y}]_t = 4\mathcal{X}^2 \cdot \lambda$. Das bedeutet, dass $\mathcal{Y}$ durch Zeit-Transformation wieder in eine Brown'sche Bewegung überführt werden kann. Siehe Abbildung 20.1.

Beweis. Wir beginnen den Beweis damit zu zeigen, dass $\mathcal{X}$ stetige Pfade hat. Hierzu müssen wir zeigen, dass folgendes fast sicher gilt: Ist $t \mapsto T_t$ auf einem Intervall kostant, (d.h. $[\mathcal{Y}]_t$ ist konstant) so ist $\mathcal{Y}$ auf demselben Intervall ebenfalls konstant. Es genügt hierzu anzunehmen, dass $\mathcal{X} \in \mathcal{M}^2$; andersfalls gehen wir zu gestoppten Prozessen über. Weiter genügt es, die Aussage für Intervalle mit rationalen Endpunkten zu zeigen. Für

$$S_s := \inf\{t > s : [\mathcal{Y}]_t > [\mathcal{Y}]_s\}$$

berechnen wir mit Hilfe des Optional Sampling Theorems

$$\mathbf{E}[(Y_{S_s})^2 - [\mathcal{Y}]_{S_s} | \mathcal{F}_s] = Y_s^2 - [\mathcal{Y}]_s,$$

also wegen $[\mathcal{Y}]_s = [\mathcal{Y}]_{S_s}$ auch

$$\mathbf{E}[(Y_{S_s} - Y_s)^2 | \mathcal{F}_s] = 0,$$

woraus die Stetigkeit von $\mathcal{X}$ folgt.

Wir zeigen nun, dass sowohl $\mathcal{X}$ als auch $(X_t^2 - t)$ stetige lokale Martingale bezüglich $(\mathcal{G}_t)_{t\geq 0}$ sind. Mit Theorem 20.3 folgt dann die Behauptung. Sei nun

$$U_n := \inf\{t \geq 0 : |\mathcal{Y}_t| > n\}, \qquad V_n := [\mathcal{Y}]_{U_n}.$$

Dann gilt

$$T_{t\wedge V_n} = T_t \wedge U_n.$$

(Ist nämlich $t \leq V_n$, so ist die quadratische Variation bis Zeit U_n bereits größer als t. Somit liegt die Zeit, bis zu der quadratische Variation t aufgesammelt wird, vor U_n also $U_n \geq T_t$. Falls andersherum $t \geq V_n$, so ist bis U_n noch nicht t als quadratische Variation erreicht worden. Bis t erreicht wird, vergeht also noch Zeit, also $T_t \geq U_n$. Klar ist in diesem Falle $T_{V_n} = U_n$.) Da nun $\{V_n \leq t\} = \{U_n \leq T_t\} \in \mathcal{G}_t$, ist V_n eine $(\mathcal{G}_t)_{t\geq 0}$-Stoppzeit. Weiter gilt für eine $(\mathcal{G}_t)_{t\geq 0}$-Stoppzeit T

$$\begin{aligned} \mathbf{E}[X_{t\wedge V_n \wedge T}] &= \mathbf{E}[Y_{T_{t\wedge T\wedge V_n}}] = \mathbf{E}[Y_{T_{t\wedge T}\wedge U_n}] = 0, \\ \mathbf{E}[X^2_{t\wedge V_n\wedge T} - t\wedge V_n \wedge T] &= \mathbf{E}[Y^2_{T_{t\wedge V_n\wedge T}} - t\wedge [\mathcal{Y}]_{U_n} \wedge T] = 0, \end{aligned}$$

Weiter gilt für $s \leq t$

$$\mathbf{E}[X_{t\wedge V_n}|\mathcal{G}_s] = \mathbf{E}[Y_{T_{t\wedge V_n}}|\mathcal{G}_s] = \mathbf{E}[Y_{T_t\wedge U_n}|\mathcal{G}_s] = Y_{T_s\wedge U_n} = X_{s\wedge V_n}.$$

Damit ist $\mathcal{X}$ ein lokales Martingal mit $V_1, V_2, ...$ als lokalisierende Folge von Stoppzeiten. Weiter ist

$$\begin{aligned} \mathbf{E}[X^2_{t\wedge V_n} - t\wedge V_n|\mathcal{G}_s] &= \mathbf{E}[Y^2_{T_t\wedge U_n} - [\mathcal{Y}]_{T_t\wedge U_n} + [\mathcal{Y}]_{T_{t\wedge V_n}} - t\wedge V_n|\mathcal{G}_s] \\ &= Y^2_{T_s\wedge U_n} - [\mathcal{Y}]_{T_s\wedge V_n} = X^2_{s\wedge V_n} - s\wedge V_n. \end{aligned}$$

Da nun $\mathcal{X}^2 - t$ ein lokales Martingal ist, folgt die Behauptung. □

20.2 Martingal-Repräsentationen

Wir beschäftigen uns nun damit, wie man (stetige) lokale Martingale als Integrale bezüglich Brown'scher Bewegungen darstellen kann. Sei etwa $\mathcal{Y}$ ein lokales Martingal. Dann ist also eine Brownsche Bewegung $\mathcal{X}$ gegeben und ein Prozess $\mathcal{H}$ gesucht, dass $\mathcal{Y} = \mathcal{H}\cdot\mathcal{X}$; siehe Theorem 20.9. Andererseits kann auch der Prozess $\mathcal{H}$ gegeben sein, und es wird eine Brown'sche Bewegung $\mathcal{X}$ gesucht, so dass $\mathcal{Y} = \mathcal{H}\cdot\mathcal{X}$ gilt; siehe Theorem 20.10.

Beispiel 20.7. Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung und $\mathcal{Y} = (Y_t)_{t\geq 0}$ mit $Y_t = X_t^2 - t$.

1. Wir wissen bereits, dass $\mathcal{Y} = 2\mathcal{X}\cdot\mathcal{X}$ gilt. Das bedeutet, dass wir einen Prozess $\mathcal{H} = 2\mathcal{X}$ gefunden haben, so dass $\mathcal{Y} = \mathcal{H}\cdot\mathcal{X}$. Der allgemeine Fall hierzu wird in Theorem 20.9 behandelt.

2. Sei $\mathcal{H} = 2\mathcal{X}$. Wir suchen eine Brown'sche Bewegung $\mathcal{X}'$, so dass $\mathcal{Y} = \mathcal{H}\cdot\mathcal{X}'$. Natürlich könnten wir in diesem Beispiel die Brown'sche Bewegung $\mathcal{X}' = \mathcal{X}$ verwenden. Es geht aber auch anders: Wir bemerken, dass $[\mathcal{Y}] = \mathcal{H}^2\cdot\lambda$ nach Beispiel 19.30. Weiter verwenden wir (zumindest für Zeitpunkte t mit $X_t \neq 0$) den Prozess $\mathcal{X}' = \frac{1}{2}\mathcal{X}^{-1}\cdot\mathcal{Y}$. Dieser ist ein stetiges lokales Martingal mit $[\mathcal{X}'] = \frac{1}{4}\mathcal{X}^{-2}\cdot[\mathcal{Y}] = \mathcal{X}^{-2}\mathcal{X}^2\cdot\lambda = \lambda$ nach Beispiel 19.30. Mit Theorem 20.3 folgt, dass $\mathcal{X}'$ eine Brown'sche Bewegung ist. Außerdem gilt $\mathcal{H}\cdot\mathcal{X}' = \frac{1}{2}\mathcal{X}^{-1}2\mathcal{X}\cdot\mathcal{Y} = \mathcal{Y}$ nach Proposition 19.46. Der allgemeine Fall ist in Theorem 20.10 behandelt.

Proposition 20.8 (Darstellung von $Y \in L^2$ durch ein Brown'sches Integral). *Sei $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ eine d-dimensionale Brownsche Bewegung, so dass $(\mathcal{G}_t)_{t\geq 0}$ die vervollständigte erzeugte Filtration von $\underline{\mathcal{X}}$ ist und $Y \in \mathcal{L}^2$ und nach $\mathcal{G}_\infty$-messbar. Dann gibt es* *einen fast sicher eindeutigen Prozess* $\underline{\mathcal{H}} = (\mathcal{H}^1, ..., \mathcal{H}^d)$ *mit* $\mathbf{E}[\int_0^\infty ||H_s||_2^2 ds] < \infty$, *so dass*

$$Y = \mathbf{E}[Y] + \sum_{i=1}^{d} (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty.$$

Beweis. Das Blumenthal'sche 0-1-Gesetz, Theorem 17.1 zeigt, dass $\mathbf{E}[Y|\mathcal{G}_0] = \mathbf{E}[Y]$ gilt. Also genügt es anzunehmen, dass $\mathbf{E}[Y] = 0$. Sei $\mathcal{I}$ der Hilbert-Raum der $\mathcal{G}_\infty$-messbaren Zufallsvariablen Z mit $\mathbf{E}[Z] = 0$ und $\mathbf{E}[Z^2] < \infty$, und $\mathcal{J} \subseteq \mathcal{I}$ der Teilraum von Zufallsvariablen, der die gewünschte Darstellung $Z = \sum_{i=1}^d (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty$ zulässt. Diese Darstellung ist eindeutig: sei nämlich $Z = \sum_{i=1}^d (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty \sum_{i=1}^d (\mathcal{K}^i \cdot \mathcal{X}^i)_\infty$, dann gilt

$$0 = \mathbf{E}\Big[(\sum_{i=1}^{d} (\mathcal{H}^i - \mathcal{K}^i) \cdot \mathcal{X}^i)_\infty^2 \Big] = \mathbf{E}\Big[\Big[\sum_{i=1}^{d} (\mathcal{H}^i - \mathcal{K}^i) \cdot \mathcal{X}^i \Big]_\infty \Big] = \sum_{i=1}^{d} \mathbf{E}\Big[\int_0^\infty (H_s^i - K_s^i)^2 ds \Big],$$

also $\mathcal{H}^i = \mathcal{K}^i$ fast sicher. Analog folgt, dass mit $\mathcal{I}$ auch $\mathcal{J}$ vollständig ist, und dass für $Z = \sum_{i=1}^d (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty \in \mathcal{J}$

$$\mathbf{E}\Big[\int_0^\infty ||H_s||_2 ds \Big] = \mathbf{E}[Z^2] < \infty$$

gilt. Also ist $\mathcal{J} \subseteq \mathcal{I}$ ein abgeschlossener Teilraum und wir müssen zeigen, dass für $Y \in \mathcal{I}$ mit $Y \perp \mathcal{J}$ immer $Z = 0$ gilt. Dann ist nämlich $\mathcal{J} = \mathcal{I}$.

Seien $\underline{h} = (h^1, ..., h^d)$ deterministische Funktionen mit $\int_0^\infty ||\underline{h}_s||_2^2 ds < \infty$. Dann ist $\sum_{k=1}^d h^k \cdot \mathcal{X}^k$ ein lokales Martingal mit $[\sum_{k=1}^d h^k \cdot \mathcal{X}^k]_t = \int_0^t ||h_s||_2^2 ds$. Nach Lemma 20.2 ist $\mathcal{Z} = (Z_t)_{t\geq}$ mit $Z_t = \exp(i \sum_{i=1}^d (h^k \cdot \underline{\mathcal{X}}^k)_t + \frac{1}{2} \int_0^t ||\underline{h}_s||_2^2 ds)$ ein lokales Martingal und $\mathcal{Z} - 1 = i(\mathcal{Z} \cdot \sum_{k=1}^d h^k \cdot \mathcal{X}^k) = i(\sum_{k=1}^d \mathcal{Z} h^k \cdot \mathcal{X}^k)$, wobei das letzte Gleichheitszeichen aus Proposition 19.46 folgt. Damit ist $\mathcal{Z} - 1 \in \mathcal{J}$ und es gilt für $Y \perp \mathcal{J}$, dass

$$0 = \mathbf{E}[Y(Z_\infty - 1)] = \mathbf{E}\Big[Y \exp(i \sum_{i=1}^{d} (h^k \cdot \underline{\mathcal{X}}^k)_\infty + \tfrac{1}{2} \int_0^\infty ||\underline{h}_s||_2^2 ds) \Big]$$

Wählt man speziell Treppenfunktionen $h^1, ..., h^d$, so folgt aus der Eindeutigkeit der charakteristischen Funktion, dass

$$\mathbf{E}[Y, (X_{t_1}, ..., X_{t_n}) \in C] = 0$$

für $t_1, ..., t_n \in [0, \infty)$ und $C \in \mathcal{B}^n$. Diese Aussage lässt sich nun erweitern zu $\mathbf{E}[Y, C] = 0$ für $C \in \mathcal{G}_\infty$. Da Y nach $\mathcal{G}_\infty$ messbar ist, folgt $Y = 0$. □

Theorem 20.9 (Darstellung von lokalen Martingalen durch Brown'sche Integrale).

Sei $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ *eine d-dimensionale Brownsche Bewegung, so dass* $(\mathcal{G}_t)_{t\geq 0}$ *die vervollständigte erzeugte Filtration von* $\underline{\mathcal{X}}$ *ist, sowie* $\mathcal{Y} = (Y_t)_{t\geq 0}$ *ein lokales* $(\mathcal{G}_t)_{t\geq 0}$*-Martingal. Dann hat* $\mathcal{Y}$ *fast sicher stetige Pfade und es gibt einen fast sicher eindeutigen Prozess* $\underline{\mathcal{H}} = (\mathcal{H}^1, ..., \mathcal{H}^d)$, *so dass*

$$\mathcal{Y} = Y_0 + \sum_{i=1}^{d} \mathcal{H}^i \cdot \mathcal{X}^i$$

fast sicher.

Beweis. Der wichtigste Schritt ist es zu zeigen, dass $\mathcal{Y}$ fast sicher stetige Pfade hat. Zunächst genügt es, diese Behauptung zu zeigen, falls $\mathcal{Y}$ uniform integrierbar ist. (Dies kann man durch eine lokalisierende Folge von Stoppzeiten erreichen.) Also schreiben wir $Y_t = \mathbf{E}[Y_\infty|\mathcal{G}_t]$ für eine $\mathcal{G}_\infty$-messbare Zufallsvariable Y_∞. Weiter wählen wir beschränkte Zufallsvariablen $Y_\infty^1, Y_\infty^2, ...$, so dass $\mathbf{E}[|Y_\infty^n - Y_\infty|] \leq 3^{-n}$. Da $Y_\infty^n \in L^2$, können wir wegen Proposition 20.8 Y_∞^n als Integral bezüglich $\underline{\mathcal{X}}$ schreiben. Insbesondere sind mit $Y_t^n := \mathbf{E}[Y_\infty^n|\mathcal{G}_t]$ die Martingale $(Y_t^n)_{t\geq 0}$ stetig. Weiter ergibt sich aus Lemma 15.24, dass

$$\mathbf{P}(\sup_{t\geq 0} |Y_t^n - Y_t| > 2^{-n}) \leq 2^n \mathbf{E}[|Y_\infty^n - Y_\infty|] \leq (2/3)^n.$$

Da die rechte Seite summierbar ist, folgt mit dem Borel-Cantelli-Lemma die uniforme Konvergenz der $\mathcal{Y}^n$ gegen $\mathcal{Y}$. Da der uniforme Grenzwert stetiger Funktionen stetig ist, ist also auch $\mathcal{Y}$ stetig. Da $\mathcal{Y}$ nun stetig ist, gibt es eine Stoppzeit T, so dass $\mathcal{Y}^T$ beschränkt ist. Erneute Anwendung von Proposition 20.8 liefert, dass $\mathcal{Y}^T = (\sum_{k=1}^d \mathcal{H}^k \cdot \mathcal{X}^k)^T$ für geeignete $\mathcal{H}^1, ..., \mathcal{H}^d$. □

Im nächsten Resultat benötigen wir eine Erweiterung des Wahrscheinlichkeitsraumes $(\Omega, \mathcal{F}, \mathbf{P})$. Dies ist ein Raum $(\Omega', \mathcal{F}', \mathbf{P}')$, so dass die Einbettung $\Omega \to \Omega'$ maßerhaltend ist. Wichtig ist es einzusehen, dass auf Ω' weitere stochastische Prozesse definiert sein können.

Theorem 20.10 (Integral-Darstellung von lokalen Martingalen).
Sei $\underline{\mathcal{Y}} = (\mathcal{Y}^1, ..., \mathcal{Y}^d)$ ein Vektor von stetigen lokalen Martingalen mit $Y_0^i = 0, i = 1, ..., d$ und

$$[\mathcal{Y}^i, \mathcal{Y}^j]_t = \sum_{k=1}^n \int_0^t H_s^{i,k} H_s^{j,k} ds$$

für progressiv messbare Prozesse $\mathcal{H}^{i,k} = (H_t^{i,k})_{t\geq 0}$, $i = 1, ..., d$, $k = 1, ..., n$. Dann gibt es eine Erweiterung des Wahrscheinlichkeitsraumes und eine Brown'sche Bewegung $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^n)$ mit

$$\mathcal{Y}^i = \sum_{k=1}^n \mathcal{H}^{i,k} \cdot \mathcal{X}^k, \qquad i = 1, ..., d.$$

Beweis. Wir betrachten die $d \times n$-Matrix $(\mathcal{H}_t^{i,k})_{i=1,...,d,k=1,...,n}$. Sei $N_t \subseteq \mathbb{R}^n$ der Kern und $R_t \subseteq \mathbb{R}^d$ das Bild der entsprechenden linearen Abbildung. Weiter seien $N_t^\perp$ und $R_t^\perp$ die dazu orthogonalen Räume. Projektionen auf diese Teilräume bezeichnen wir mit $\pi_{N_t}, \pi_{N_t^\perp}, \pi_{R_t}$ und $\pi_{R_t^\perp}$. Dann ist $(\mathcal{H}_t^{i,k})_{i=1,...,d,k=1,...,n}$ eine Bijektion $N_t^\perp \to R_t$ und wir bezeichnen dessen Umkehrabbildung mit $((\mathcal{H}_t^{i,k})^{-1})_{i=1,...,d,k=1,...,n}$.

Wir erweitern den Wahrscheinlichkeitsraum um eine unabhängige Brown'sche Bewegung $\underline{\mathcal{Z}} = (\mathcal{Z}^1, ..., \mathcal{Z}^n)$, sowie $\mathcal{G}_t = \sigma(\mathcal{F}_t, (Z_s)_{s\leq t})$. Dann ist $\underline{\mathcal{Y}}$ ein lokales Martingal und $\underline{\mathcal{Z}}$ ein Martingal bezüglich $(\mathcal{G}_t)_{t\geq 0}$. Wir definieren nun

$$\mathcal{X} = \mathcal{H}^{-1}\pi_R \cdot \mathcal{Y} + \pi_N \cdot \mathcal{Z}.$$

Damit gilt

$$d\mathcal{X}_t^k = \begin{cases} \sum_{i=1}^d (H_t^{i,k})^{-1} dY_t^i, & H_t \in R_t \\ dZ_t, & H_t \notin R_t \end{cases},$$

$$d[\mathcal{X}^k, \mathcal{X}^l]_t = \begin{cases} \sum_{i,j=1}^d (H_t^{i,k})^{-1} (H_t^{j,l})^{-1} d[\mathcal{Y}^i, \mathcal{Y}^j]_t, & H_t \in R_t, \\ \delta_{kl} dt, & H_t \notin R_t \end{cases}$$

Da nach Voraussetzung

$$\begin{aligned} \sum_{i,j=1}^d (H_t^{i,k})^{-1} (H_t^{j,l})^{-1} d[\mathcal{Y}^i, \mathcal{Y}^j]_t &= \sum_{m=1}^n \sum_{i,j=1}^d (H_t^{i,k})^{-1} (H_t^{j,l})^{-1} H_t^{i,m} H_t^{j,m} dt \\ &= \sum_{m=1}^n \delta_{km} \delta_{lm} dt = \delta_{kl} dt, \end{aligned}$$

gilt also $d[\mathcal{X}^k, \mathcal{X}^l] = \delta_{kl} \cdot \lambda$. Nach Theorem 20.3 ist $\underline{\mathcal{X}}$ also eine Brown'sche Bewegung. Außerdem gilt mit Proposition 19.46

$$\mathcal{Y}^i = 1 \cdot \mathcal{Y}^i = \sum_{j=1}^d \delta_{ij} \cdot \mathcal{Y}^j = \sum_{k=1}^n \sum_{j=1}^d (\mathcal{H}^{i,k} \cdot (\mathcal{H}^{j,k})^{-1} \cdot \mathcal{Y}^j = \sum_{k=1}^n \mathcal{H}^{i,k} \cdot \mathcal{X}^k.$$

□

20.3 Maßwechsel und Transformationen der Drift

Wir kommen nun dazu, zwei Maße $\mathbf{P}$ und $\mathbf{Q}$ auf dem Messraum $(\Omega, \mathcal{F})$ zu betrachten, auf dem unsere stochastischen Prozesse definiert sind. Wir setzen immer voraus, dass dieser Messraum mit einer Filtration $(\mathcal{F}_t)_{t\geq 0}$ ausgestattet ist, die rechtsstetig und vollständig (bezüglich $\mathbf{P}$) ist; dabei schreiben wir $\mathcal{F}_\infty := \sigma((\mathcal{F}_t)_{t\geq 0})$. Da die Martingal-Eigenschaft eine bedingte Erwartung voraussetzt, und diese wiederum ein Wahrscheinlichkeitsmaß, werden wir nun von $\mathbf{P}$-Martingalen bzw. $\mathbf{Q}$-Martingalen sprechen. Weiter schreiben wir $\mathbf{E_P}[.]$ für Erwartungswerte bezüglich $\mathbf{P}$ und $\mathbf{E_Q}[.]$ für Erwartungswerte bezüglich $\mathbf{Q}$.

Wir wissen nach dem Satz von Radon-Nikodym, Korollar 5.16, dass es im Falle $\mathbf{Q} \ll \mathbf{P}$ (auf $\mathcal{F}$) eine Zufallsvariable Z mit $\mathbf{E_P}[Z] = 1$ gibt, so dass $\mathbf{Q}(A) = \mathbf{E_P}[Z, A]$ gilt. Gilt jedoch nur $\mathbf{Q} \ll \mathbf{P}$ auf $\mathcal{F}_t$, d.h. $\mathbf{P}$-Nullmengen $N \in \mathcal{F}_t$ sind auch $\mathbf{Q}$-Nullmengen, gibt es immerhin eine $\mathcal{F}_t$-messbare Zufallsvariable Z_t, so dass $\mathbf{Q}(A) = \mathbf{E_P}[Z_t, A]$ für $A \in \mathcal{F}_t$ gilt. Diesen Fall wollen wir nun betrachten. Wir erinnern daran, dass $Z \cdot \mathbf{P}$ für eine nicht-negative Zufallsvariable Z mit $\mathbf{E_P}[Z] = 1$ das Wahrscheinlichkeitsmaß mit Dichte Z bezüglich $\mathbf{P}$ ist.

Beispiel 20.11 (Transformation der Brown'schen Bewegung aus Proposition 15.6). Das grundlegende Beispiel eines Maßwechsels haben wir bereits in Proposition 15.6 kennen gelernt. Sei $\mathcal{X}$ eine Brown'sche Bewegung und $\mathbf{P}$ dessen Verteilung und $(\mathcal{F}_t)_{t\geq 0}$ die erzeugte Filtration, sowie $\mathcal{Z} = (Z_t)_{t\geq 0}$ mit $Z_t = \exp(\mu X_t - \frac{1}{2}\mu^2 t)$ und $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$. Dann haben wir gezeigt, dass die Verteilung von $\mathcal{X}$ unter $\mathbf{Q}$ gerade die eines Prozesses $\mathcal{Y}$ ist mit $Y_t = X_t + \mu t$. Also haben wir durch Maßwechsel die *Drift* der Brown'schen Bewegung geändert. Die Änderung der Verteilung eines stochastischen Prozesses unter Maßwechsel ist genau das hier behandelte Thema.

Lemma 20.12 (Maßransformationen und Martingale). *Sei E ein Polnischer Raum und $\mathbf{P}$ ein Wahrscheinlichkeitsmaß auf $(E^I, (\mathcal{B}(E))^I)$ für $I = [0, \infty)$. Weiter sei $(\mathcal{F}_t)_{t\geq 0}$ eine rechtsstetige und vollständige Filtration.*

1. *Ist $\mathcal{Z} = (Z_t)_{t\geq 0}$ mit $Z_t \geq 0$ und $Z_0 = 1$ ein $\mathbf{P} - ((\mathcal{B}(E))_t)_{t\geq 0}$-Martingal, dann gibt es ein Wahrscheinlichkeitsmaß $\mathbf{Q}$ auf $(E^I, (\mathcal{B}(E))^I)$, so dass $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$ gilt, $t \geq 0$.*
2. *Sei $\mathbf{Q}$ ein Wahrscheinlichkeitsmaß, so dass $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$, $t \geq 0$, für geeignete Z_t. Dann ist $\mathcal{Z} = (Z_t)_{t\geq 0}$ ein $\mathbf{P}$-Martingal. Es ist genau dann gleichgradig integrierbar, wenn $\mathbf{Q} \ll \mathbf{P}$ auf $\mathcal{F}_\infty$. Außerdem ist ein Prozess $\mathcal{X}$ genau dann ein $\mathbf{Q}$-Martingal, wenn $\mathcal{X}\mathcal{Z}$ ein $\mathbf{P}$-Martingal ist.*

Beweis. 1. Wir definieren eine projektive Familie von Wahrscheinlichkeitsmaßen auf $E^{[0,\infty)}$ mittels des Martingals $\mathcal{Z}$. Sei hierzu $J = \{t_1, ..., t_n\}$ mit $t_1 < ... < t_n$. Dann definieren wir

$$\mathbf{Q}_J(A_1 \times ... \times A_n) := \mathbf{E}_\mathbf{P}[Z_{t_n}, A_1 \times \cdots \times A_n].$$

Diese Familie ist projektiv, da für $H = \{t_1, ..., t_{n-1}\}$

$$\begin{aligned}(\pi^J_H)_*\mathbf{Q}_J(A_1 \times ... \times A_{n-1}) &:= \mathbf{E}_\mathbf{P}[Z_{t_n}, A_1 \times \cdots \times A_{n-1}] = \mathbf{E}_\mathbf{P}[Z_{t_{n-1}}, A_1 \times \cdots \times A_{n-1}] \\ &= \mathbf{Q}_H(A_1 \times ... \times A_{n-1})\end{aligned}$$

aufgrund der Martingal-Eigenschaft von $\mathcal{Z}$. Damit existiert nach Theorem 6.24 ein Wahrscheinlichkeitsmaß $\mathbf{Q}$ mit den geforderten Eigenschaften.

2. Sei $\mathcal{X} = (X_t)_{t\geq 0}$. Klar ist, dass X_t genau dann in $L^1(\mathbf{Q})$ ist, wenn $X_t Z_t \in L^1(\mathbf{P})$ ist. In diesem Fall ist $\mathbf{E}_\mathbf{Q}[X_s, A] = \mathbf{E}_\mathbf{Q}[X_t, A]$ für $A \in \mathcal{F}_s$ (was bedeutet, dass $\mathcal{X}$ in $\mathbf{Q}$-Martingal ist) genau dann, wenn $\mathbf{E}_\mathbf{P}[Z_s X_s, A] = \mathbf{E}_\mathbf{P}[Z_t X_t, A]$ gilt. Letzteres ist genau dann der Fall für alle $s \leq t$, wenn $\mathcal{X}\mathcal{Z}$ ein $\mathbf{P}$-Martingal ist. Also haben wir die letzte Behauptung gezeigt. Wählt man $\mathcal{X} = 1$, folgt insbesondere, dass $\mathcal{Z}$ ein $\mathbf{P}$-Martingal sein muss.

Es bleibt zu zeigen, dass $\mathbf{Q} \ll \mathbf{P}$ genau dann, wenn $\mathcal{Z}$ gleichgradig integrierbar ist. Sei zunächst $\mathcal{Z}$ gleichgradig integrierbar. Nach Theorem 15.31 (und Theorem 15.47) konvergiert dann $\mathcal{Z}$ in L^1 gegen eine Zufallsvariable Z_∞. Deshalb gilt $\mathbf{Q}(A) = \mathbf{E}[Z_\infty, A]$ für alle $A \in \bigcup_{t\geq 0} \mathcal{F}_t$. Diese Eigenschaft lässt sich wie üblich für alle $A \in \mathcal{F}_\infty$ fortsetzen, was bedeutet, dass $\mathbf{Q} = Z_\infty \cdot \mathbf{P}$. Insbesondere ist $\mathbf{Q} \ll \mathbf{P}$. Gilt andersherum $\mathbf{Q} \ll \mathbf{P}$, so gibt es ein Z mit $\mathbf{Q} = Z \cdot \mathbf{P}$ auf $\mathcal{F}_\infty$. Damit ist notwendigerweise $Z_t = \mathbf{E}_\mathbf{P}[Z|\mathcal{F}_t]$. Außerdem ist $\mathcal{Z}$ nach Lemma 12.5 gleichgradig integrierbar. □

Lemma 20.13 (Maßtransformation zu Stoppzeiten). *Sei $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$, $t \geq 0$. Dann gilt für jede Stoppzeit T*

$$\mathbf{Q} = Z_T \cdot \mathbf{P} \text{ auf } \mathcal{F}_t \cap \{T < \infty\}.$$

Außerdem ist ein Prozess $\mathcal{X}$ genau ein lokales $\mathbf{Q}$-Martingal, wenn $\mathcal{X}\mathcal{Z}$ ein lokales $\mathbf{P}$-Martingal ist.

Beweis. Sei $A \in \mathcal{F}_T$. Dann gilt wegen monotoner Konvergenz und dem Optional Sampling Theorem

$$\begin{aligned}\mathbf{Q}(A \cap \{T < \infty\}) &= \lim_{t\geq 0} \mathbf{Q}(A \cap \{T \leq t\}) = \lim_{t\geq 0} \mathbf{E}_\mathbf{P}[Z_t, A \cap \{T \leq t\}] \\ &= \lim_{t\geq 0} \mathbf{E}_\mathbf{P}[Z_{T\wedge t}, A \cap \{T \leq t\}] = \mathbf{E}_\mathbf{P}[Z_T, A \cap \{T < \infty\}].\end{aligned}$$

Also folgt $\mathbf{Q} = Z_T \cdot \mathbf{P}$ auf $\mathcal{F}_t \cap \{T < \infty\}$. Weiter ist für eine Stoppzeit T der Prozess $\mathcal{X}^T$ genau dann ein $\mathbf{Q}$-Martingal, wenn $\mathcal{X}^T \mathcal{Z}$ ein $\mathbf{P}$-Martingal ist. Schreiben wir $\mathcal{Z} = \mathcal{Z}^T + (\mathcal{Z} - \mathcal{Z}^T)$, ist klar, dass $\mathcal{X}^T(\mathcal{Z} - \mathcal{Z}^T)$ ein $\mathbf{P}$-Martingal ist. Damit ist $\mathcal{X}^T \mathcal{Z}$ genau dann ein $\mathbf{P}$-Martingal, wenn $\mathcal{X}^T \mathcal{Z}^T$ ein $\mathbf{P}$-Martingal ist. Also sind alle Behauptungen gezeigt. □

Theorem 20.14 (Girsanov-Transformation). *1. Sei $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$, $t \geq 0$, wobei $\mathcal{Z} = (Z_t)_{t \geq 0}$ stetige Pfade hat ($\mathbf{P}$-fast sicher). Ist nun $\mathcal{M}$ ein stetiges, lokales $\mathbf{P}$-Martingal, so ist $\mathbf{Q}$-fast sicher der Prozess $\mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$ lokal beschränkt und $\widetilde{\mathcal{M}} := \mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$ ein lokales $\mathbf{Q}$-Martingal.*

2. Ein stetiger Prozess $\mathcal{Z} > 0$ ist genau dann ein lokales $\mathbf{P}$-Martingal, wenn

$$\mathcal{Z} = \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}])$$

für ein stetiges lokales $\mathbf{P}$-Martingal $\mathcal{N}$. In diesem Fall ist $\mathcal{N}$ fast sicher eindeutig, und für jedes stetige lokale $\mathbf{P}$-Martingal $\mathcal{M}$ ist $[\mathcal{M}, \mathcal{N}] = \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$. Ist also $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$, dann ist $\widetilde{M} := \mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}] = \mathcal{M} - [\mathcal{M}, \mathcal{N}]$ ein lokales $\mathbf{Q}$-Martingal.

Beweis. 1. Wir nehmen zunächst an, dass $\mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}] < \infty$ ist. Dann ist $\widetilde{\mathcal{M}}$ ein stetiges $\mathbf{P}$-Semimartingal, und aus der Formel für partielle Integration, Theorem 19.48, folgt

$$\begin{aligned}
\widetilde{\mathcal{M}}\mathcal{Z} - \widetilde{M}_0 Z_0 &= \widetilde{\mathcal{M}} \cdot \mathcal{Z} + \mathcal{Z} \cdot \widetilde{\mathcal{M}} + [\widetilde{\mathcal{M}}, \mathcal{Z}] \\
&= \widetilde{\mathcal{M}} \cdot \mathcal{Z} + \mathcal{Z} \cdot \mathcal{M} - \mathcal{Z} \cdot \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}] + [\widetilde{\mathcal{M}}, \mathcal{Z}] \\
&= \widetilde{\mathcal{M}} \cdot \mathcal{Z} + \mathcal{Z} \cdot \mathcal{M}.
\end{aligned}$$

Deshalb ist $\widetilde{\mathcal{M}}\mathcal{Z}$ ein lokales $\mathbf{P}$-Martingal. Nach Lemma 20.13 ist damit $\widetilde{\mathcal{M}}$ ein lokales $\mathbf{Q}$-Martingal.

Im allgemeinen Fall definieren wir $T_n := \inf\{t \geq 0 : Z_t \leq 1/n\}$. Dann ist genau wie oben $\widetilde{\mathcal{M}}^{T_n}$ ein lokales $\mathbf{Q}$-Martingal. Weiter ist $T_n \uparrow \infty$ unter $\mathbf{Q}$, denn

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{Q}(T_n \leq t) &= \lim_{n\to\infty} \mathbf{E_P}[Z_t, T_n \leq t] = \lim_{n\to\infty} \mathbf{E_P}[Z_{T_n \wedge t}, T_n \leq t] \\
&= \lim_{n\to\infty} \mathbf{E_P}[Z_{T_n}, T_n \leq t] = \lim_{n\to\infty} \mathbf{E_P}[1/n, T_n \leq t] = 0
\end{aligned}$$

für jedes $t \geq 0$. Damit ist $\widetilde{\mathcal{M}}$ nach Lemma 19.27 ein lokales $\mathbf{Q}$-Martingal.

2. Ist $\mathcal{N}$ ein lokales $\mathbf{P}$-Martingal, so gilt mit der Itô-Formel

$$\begin{aligned}
\exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) - \exp(N_0) &= \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) \cdot (\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) + \tfrac{1}{2}\exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) \cdot [\mathcal{N}] \\
&= \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) \cdot \mathcal{N}.
\end{aligned}$$

Insbesondere ist $\exp(\mathcal{N} - \frac{1}{2}[\mathcal{N}])$ ein lokales $\mathbf{P}$-Martingal. Ist andersherum $\mathcal{Z} > 0$, so folgt wieder aus der Itô-Formel

$$\log \mathcal{Z} - \log Z_0 = \mathcal{Z}^{-1} \cdot \mathcal{Z} - \tfrac{1}{2}\mathcal{Z}^{-2} \cdot [\mathcal{Z}] = \mathcal{Z}^{-1} \cdot \mathcal{Z} - \tfrac{1}{2}[\mathcal{Z}^{-1} \cdot \mathcal{Z}],$$

und man liest die Behauptung für $\mathcal{N} = \log Z_0 + \mathcal{Z}^{-1} \cdot \mathcal{Z}$ ab. Die fast sichere Eindeutigkeit von $\mathcal{N}$ folgt aus der Eindeutigkeit von $\log \mathcal{Z}$ und Theorem 19.26. Weiter gilt für ein lokales $\mathbf{P}$-Martingal $\mathcal{M}$

$$[\mathcal{M}, \mathcal{N}] = [\mathcal{M}, \mathcal{Z}^{-1} \cdot \mathcal{Z}] = \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}].$$

□

Die quadratische Variation $[\mathcal{X}]$ eines stetigen Semimartingals $\mathcal{X} = \mathcal{A} + \mathcal{M}$ war als der fast sicher eindeutige Prozess von lokale beschränkter Variation, so dass $\mathcal{M}^2 - [\mathcal{X}]$ ein lokales Martingal ist. Da letztere Eigenschaft vom Wahrscheinlichkeitsmaß $\mathbf{P}$ abhängt, hängt auch $[\mathcal{X}]$ von $\mathbf{P}$ ab. Deswegen schreiben wir nun $[\mathcal{X}]^{\mathbf{P}}$ für die quadratische Variation bezüglich des Wahrscheinlichkeitsmaßes $\mathbf{P}$. Jedoch stellt sich im nächsten Resultat heraus, dass $[\mathcal{X}]$ durch Transformation von $\mathbf{P}$ nicht verändert wird. Weiter haben wir $\mathcal{H} \cdot \mathcal{X}$ durch eine Martingal-Eigenschaft definiert. Deshalb schreiben wir $(\mathcal{H} \cdot \mathcal{X})^{\mathbf{P}}$ für das stochstische Integral bezüglich $\mathbf{P}$. Hier gilt ebenfalls, dass das stochastischen Integral durch einen Maßwechsel nicht beeinflusst wird. (Dies ist für $\mathcal{H} \in \mathbb{S}$ klar.)

Proposition 20.15 (Quadratische Variation ändert sich nicht unter Maßwechsel). *Sei $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$ und $\mathcal{Z} = (Z_t)_{t \geq 0}$ stetig. Dann ist jedes $\mathbf{P}$-Semimartingal $\mathcal{X}$ auch ein $\mathbf{Q}$-Semimartingal und es gilt $[\mathcal{X}]^{\mathbf{P}} = [\mathcal{X}]^{\mathbf{Q}}$. Weiter gilt $\mathbf{Q}$-fast sicher $(\mathcal{H} \cdot \mathcal{X})^{\mathbf{P}} = (\mathcal{H} \cdot \mathcal{X})^{\mathbf{Q}}$ für $\mathcal{H} \in L_{\mathbf{P}}(\mathcal{X})$, also insbesondere $L_{\mathbf{P}}(\mathcal{X}) \subseteq L_{\mathbf{Q}}(\mathcal{X})$. Weiter gilt für jedes $\mathbf{P}$-Martingal $\mathcal{M}$*

$$\mathcal{H} \cdot \mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{H} \cdot \mathcal{M}, \mathcal{Z}] = \mathcal{H} \cdot (\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]).$$

Beweis. Ist $\mathcal{X} = \mathcal{A} + \mathcal{M}$ ein $\mathbf{P}$-Semimartingal, so schreiben wir $\mathcal{X} = (\mathcal{A} + \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]) + (\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}])$. Auf der rechten Seite steht nun die Zerlegung von $\mathcal{X}$ in den Prozess von lokal beschränkter Variation $\mathcal{A} + \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$ und das $\mathbf{Q}$-Martingal $\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$. Damit ist $\mathcal{X}$ auch ein $\mathbf{Q}$-Semimartingal.

Die Behauptung $[\mathcal{X}]^{\mathbf{P}} = [\mathcal{X}]^{\mathbf{Q}}$ folgt aus Lemma 19.50. Gilt nun $\mathcal{H} \in L_{\mathbf{P}}(\mathcal{X})$, so gilt $\mathcal{H} \in L_{\mathbf{P}}(\mathcal{A})$, also auch $\mathcal{H} \in L_{\mathbf{Q}}(\mathcal{A})$. Weiter ist $\mathcal{H}^2 \in L_{\mathbf{P}}([\mathcal{X}])$ und es folgt $\mathcal{H}^2 \in L_{\mathbf{Q}}([\mathcal{X}]) = L_{\mathbf{Q}}([\mathcal{M}]) = L_{\mathbf{Q}}([\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]])$. Es bleibt noch zu zeigen, dass $\mathcal{H} \in L_{\mathbf{Q}}(\mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}])$. Da $\mathbf{Q}$-fast sicher $\mathcal{Z} > 0$ gilt, genügt es zu zeigen, dass $\mathcal{H} \in L_{\mathbf{Q}}([\mathcal{M}, \mathcal{Z}]) = L_{\mathbf{Q}}([\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}], \mathcal{Z}])$. Da $\mathcal{H}^2 \in L_{\mathbf{Q}}([\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}])$, folgt die Behauptung aus Proposition 19.33. □

Bemerkung 20.16 (Vollständigkeit der Filtration). Die üblichen Bedingungen, die für Filtrationen gelten sollen, sind die Rechtsstetigkeit und die Vollständigkeit (bzgl. eines Wahrscheinlichkeitsmaßes $\mathbf{P}$). Letztere werden wir im nächsten Resultat aufgeben, und zwar aus gutem Grund: sei $\mathcal{X}$ eine Brown'sche Bewegung, so gilt

$$\frac{X_t}{t} \xrightarrow{t \to \infty} 0$$

$\mathbf{P}$-fast sicher. Insbesondere ist $\{X_t/t \xrightarrow{t \to \infty} \mu\}$ für jedes $\mu \neq 0$ eine $\mathbf{P}$-Nullmenge. Transformieren wir jedoch nun wie in Proposition 15.6 $\mathbf{Q} = Z_t \cdot \mathbf{P}$, so dass $\mathcal{X}$ unter $\mathbf{Q}$ eine Brown'sche Bewegung mit Drift μ wird, so gilt $\mathbf{Q}\{X_t/t \xrightarrow{t \to \infty} \mu\} = 1$. Insbesondere würde dann nicht $\mathbf{Q} \ll \mathbf{P}$ (auf $\mathcal{F}_\infty$) gelten. Deshalb verzichten wir auf die Vollständigkeit der Filtration $(\mathcal{F}_t)_{t \geq 0}$.

Theorem 20.17 (Deterministische Transformation der Brown'schen Bewegung). *Sei $\mathcal{X} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ mit $\mathcal{X}^i = (X_t^i)_{t \geq 0}$ eine Brown'sche Bewegung in $\mathbb{R}^d$ mit kanonischer Filtration $(\mathcal{F}_t)_{t \geq 0}$, sowie $h : [0, \infty) \to \mathbb{R}^d$ mit $h(0) = 0$. Weiter sei $\mathbf{P}_0$ die Verteilung von $\mathcal{X}$ und $\mathbf{P}_h$ die von $\mathcal{X} + h$. Dann ist $\mathbf{P}_0 \sim \mathbf{P}_h$ auf jedem $\mathcal{F}_t$ genau dann, wenn $h_i = f_i \cdot \lambda$ für Funktionen f_i mit $(f_i^2 \cdot \lambda)_t < \infty$ für alle $t \geq 0$. In diesem Fall gilt*

$$\mathbf{P}_h = \exp\Big(\Big(\sum_{i=1}^d f_i \cdot \mathcal{X}^i - \tfrac{1}{2} f_i^2 \cdot \lambda\Big)_t\Big) \cdot \mathbf{P}_0$$

auf $\mathcal{F}_t$.

Beweis. Sei zunächst $\mathbf{P}_0 \sim \mathbf{P}_h$ auf jedem $\mathcal{F}_t$. Dann gibt es ein $\mathbf{P}_0$-Martingal $\mathcal{Z}$, so dass $\mathbf{P}_h = Z_t \cdot \mathbf{P}_0$ auf $\mathcal{F}_t$. Nach Theorem 20.9 hat $\mathcal{Z}$ stetige Pfade. Um einzusehen, dass $\mathcal{Z} > 0$ gilt, sei T die Treffzeit der 0 von $\mathcal{Z}$. Angenommen, $T < t$ mit positiver Wahrscheinlichkeit für ein $t > 0$, so gäbe es nach Lemma 20.13 eine Menge A mit $\mathbf{P}_h(A) = 0$, $\mathbf{P}_0(A) > 0$, im Widerspruch zu $\mathbf{P}_0 \sim \mathbf{P}_h$. Nach Theorem 20.14.2 gibt es ein lokales Martingal $\mathcal{M}$, so dass $\mathcal{Z} = \exp(\mathcal{M} - \frac{1}{2}[\mathcal{M}])$. Mit Theorem 20.9 gibt es Prozesse $\mathcal{H}^1, ..., \mathcal{H}^d$ mit $((\mathcal{H}^i)^2 \cdot \lambda)_t < \infty$ für alle $t \geq 0$, so dass $\mathcal{M} = \sum_{i=1}^d \mathcal{H}^i \cdot \mathcal{X}^i$. Nun ist jeder der Prozesse $\mathcal{X}^i - [\mathcal{X}^i, \mathcal{M}] = \mathcal{X}^i - \mathcal{H}^i \cdot \lambda$ nach Theorem 20.14 ein stetiges $\mathbf{P}_h$-Martingal mit quadratischem Variationsprozess λ, und verschwindenen Covariationen. Nach Theorem 20.3 handelt es sich also um eine $\mathbf{P}_h$-Brownsche Bewegung. Nun gilt

$$\mathcal{X}^i = (\mathcal{X}^i - \mathcal{H}^i \cdot \lambda) + \mathcal{H}^i \cdot \lambda = (\mathcal{X}^i - h_i) + h_i,$$

was zwei Zerlegungen des $\mathbf{P}_h$-Semimartingals $\mathcal{X}^i$ darstellt. Da diese Zerlegung eindeutig ist, folgt $h_i = \mathcal{H}^i \cdot \lambda$. Damit ist $\mathcal{H}^i$ deterministisch, d.h. $h_i = f^i \cdot \lambda$ für ein geeignetes f^i. Damit gilt auch $\mathcal{M} = \sum_{i=1}^d \mathcal{H}^i \cdot \mathcal{X}^i = \sum_{i=1}^d f^i \cdot \mathcal{X}^i$.

Sei andersherum $h_i = f_i \cdot \lambda$ für geeignete $f_1, ..., f_d$. Man berechnet mit der Itô-Formel, dass $\exp\big(\sum_{i=1}^d f_i \cdot \mathcal{X}^i - \frac{1}{2} f_i^2 \cdot \lambda\big)_{t \geq 0}$ ein $\mathbf{P}_0$-Martingal ist. Nach Lemma 20.12 gibt es also ein Wahrscheinlichkeitsmaß $\mathbf{Q}$ mit $\mathbf{Q} = \exp\big(\sum_{i=1}^d f_i \cdot \mathcal{X}^i - \frac{1}{2} f_i^2 \cdot \lambda\big)_t \cdot \mathbf{P}_0$ auf $\mathcal{F}_t$. Außerdem impliziert Theorem 20.14, dass $\mathcal{X}^i - [\mathcal{X}^i, \sum_{j=1}^d f_j \cdot \mathcal{X}^j] = \mathcal{X}^i - f_i \cdot \lambda = \mathcal{X}^i - h_i$ ein $\mathbf{Q}$-Martingal mit quadratischer Variation λ und verschwindender Covariation ist. Also ist $(\mathcal{X}^1 - h_1 \cdot \lambda, ..., \mathcal{X}^d - h_d \cdot \lambda)$ eine $\mathbf{Q}$-Brown'sche Bewegung, und deshalb $\mathbf{Q} = \mathbf{P}_h$. Insbesondere gilt $\mathbf{P}_0 \sim \mathbf{P}_h$ auf $\mathcal{F}_t$. □

20.4 Lokalzeit

Die Itô-Formel können wir nur für zweimal stetig differenzierbare Funktionen f anwenden. Manchmal ist es jedoch auch wünschenswert, andere Funktionen von stetigen Semimartingalen zu beschreiben. Sei etwa $\mathcal{X}$ eine Brown'sche Bewegung. Ist dann $|\mathcal{X}|$ ein Semimartingal? Die Beantwortung dieser Frage führt auf die Lokalzeit und die Darstellung

$$|X_t| - |X_0| = \int_0^t \operatorname{sgn}(X_s) dX_s + L_t \text{ oder } |\mathcal{X}| - |X_0| = \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{X} + \mathcal{L}, \tag{20.2}$$

wobei $\mathcal{L} = (L_t)_{t \geq 0}$ die Lokalzeit der Brown'schen Bewegung im Punkt 0 genannt wird und

$$\operatorname{sgn}(x) := \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{20.3}$$

Dabei ist $\mathcal{L}$ ein fast sicher nicht-fallender Prozess (und hat somit lokal endliche Variation). Die rechte Seite von (20.2) stellt man sich dabei am besten wiefolgt vor: ist $X_t > 0$, so ist die Änderung von $|\mathcal{X}|$ im Zeitintervall $[t, t+dt]$ identisch mit der Änderung von $\mathcal{X}$. Ist andererseits $X_t < 0$, so ist diese Änderung gerade die negative Änderung von $\mathcal{X}$. Ist außerdem $X_t = 0$, so wächst $|\mathcal{X}|$ in $[t, t + dt]$ sicher. Andererseits spielen solche Zeiten für das Integral bezüglich $\mathcal{X}$ keine Rolle, da diese Zeitpunkte eine Lebesgue-Nullmenge sind. Deswegen ist $\mathcal{L}$ gerade der nicht-fallende Prozess, der gerade dann wächst, wenn $|\mathcal{X}| = 0$ ist, so dass (20.2) stimmt. (Man kann insbesondere zeigen, dass

$$L_t := \lim_{h \to 0} \frac{1}{h} \int_0^t 1_{\{|X_s| < h\}} ds$$

gilt.)

Da die Konstruktion der Lokalzeit nicht nur für die Brown'sche Bewegung, sondern für allgemeine stetige Semimartingale funktioniert, formulieren wir die Existenz der Lokalzeit gleich in diesem Fall.

Theorem 20.18 (Tanaka's Formel). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein stetiges Semimartingal. Dann gibt es einen stetigen, nicht-fallenden, adaptierten Prozess $\mathcal{L} = (L_t)_{t\geq 0}$ mit*

$$|X_t| - |X_0| = \int_0^t sgn(X_s)dX_s + L_t.$$

Dabei gilt fast sicher

$$\int_0^\infty 1_{X_s\neq 0}dL_s = 0,$$

d.h. $\mathcal{L}$ wächst nur wenn $X_s = 0$. Außerdem gilt

$$L_t = 0 \vee (-|X_0| - \inf_{s\leq t}(sgn(\mathcal{X}) \cdot \mathcal{X})_s).$$

Zum Beweis des Theorems benötigen wir eine elementare Aussage.

Lemma 20.19. *Sei $f : [0,\infty) \to \mathbb{R}$ stetig mit $f(0) \geq 0$. Dann gibt es genau eine nicht-fallende, stetige Funktion g mit $g(0) = 0$, so dass $f + g \geq 0$ und $\int 1_{\{f+g>0\}}dg = 0$, nämlich*

$$g(t) := 0 \vee \sup_{s\leq t}(-f(s)).$$

Beweis. OBdA ist $f(0) = 0$. Andernfalls gehen wir zur Funktion $f(t + t_0)$ über, wobei t_0 die erste Zeit ist mit $f(t) = 0$. In diesem Fall behaupten wir also, dass $g(t) := \sup_{s\leq t}(-f(s))$ die einzige Funktion mit den angegebenen Eigenschaften ist. Für diese Funktion ist zunächst

$$f(t) + g(t) \geq f(t) - \inf_{s\leq t} f(s) \geq f(t) - f(t) = 0.$$

Ist außerdem $f+g > 0$ auf dem Intervall $(t, t']$, dann ist $g(t) = -\inf_{s\leq t} f(s) = -\inf_{s\leq t'} f(s) = g(t')$, woraus $\int 1_{\{f+g>0\}}dg = 0$ folgt.

Nun überprüfen wir die Eindeutigkeit. Seien also g und g' zwei Funktionen mit den angegebenen Eigenschaften. Falls $g(t) < g'(t)$ für ein $t \geq 0$, so definieren wir $r := \sup\{s \leq t : g(s) = g'(s)\}$. Dann gilt $f + g' \geq f + g' - f - g = g' - g > 0$ auf $(r, t]$. Aus $\int 1_{\{f+g'>0\}}dg' = 0$ folgt, dass $g'(r) = g'(t)$. Und damit gilt $0 < g'(t) - g(t) \leq g'(r) - g(r) = 0$, was ein Widerspruch ist. □

Beweis von Theorem 20.18. Wir werden die Funktion $f(x) = |x|$ durch $\mathcal{C}^\infty$-Funktionen approximieren, und auf diese die Itô-Formel anwenden. Sei hierzu $\varphi \in \mathcal{C}^\infty(\mathbb{R})$ eine wachsende Funktion, so dass $\varphi(x) = -1$ für $x \leq 0$ und $\varphi(x) = 1$ für $x \geq 1$. Dann wählen wir f_n so, dass $f_n'(x) = \varphi(nx)$ und $f_n(0) = 0$. Dann ist $f_n \uparrow f$ gleichmäßig für alle $x \in \mathbb{R}$. Die Itô-Formel liefert

$$f_n(\mathcal{X}) - f_n(X_0) = f_n'(\mathcal{X}) \cdot \mathcal{X} + \mathcal{L}^n$$

mit $\mathcal{L}^n = \frac{1}{2}f_n''(\mathcal{X}) \cdot [\mathcal{X}]$. Da $f_n''(x) = 0$ für $x \geq 1/n$, gilt $1_{\{\mathcal{X}>1/n\}} \cdot \mathcal{L}^n = 1_{\{\mathcal{X}>1/n\}}f_n''(\mathcal{X}) \cdot [\mathcal{X}] = 0$. Da weiter $\sup_{t\geq 0}|f_n'(X_t) - \text{sgn}(X_t)| \xrightarrow{n\to\infty}_{fs} 0$, folgt mit Proposition 19.47, dass

$f_n'(\mathcal{X}) \cdot \mathcal{X} \xrightarrow{n\to\infty}_p \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{X}$ uniform auf Kompakta. Weiter gilt $f_n(\mathcal{X}) \xrightarrow{n\to\infty} fs|\mathcal{X}|$ uniform für alle t, also konvergiert $\mathcal{L}^n = f_n(\mathcal{X}) - f_n(X_0) - f_n'(\mathcal{X}) \cdot \mathcal{X}$ in Wahrscheinlichkeit gegen $|\mathcal{X}| - |X_0| - \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{X}$. Diesen Grenzwert bezeichnen wir mit $\mathcal{L}$. Da $\mathcal{L}^n$ nicht-fallend ist, ist auch $\mathcal{L}$ nicht fallend. Betrachtet man $\mathcal{L}^n$ als Verteilungsfunktion eines Maßes, so konvergiert $\mathcal{L}^n$ schwach gegen $\mathcal{L}$. Also gilt nach Theorem 10.6(iii), da $\mathcal{X}$ stetig ist,

$$1_{\{\mathcal{X}>1/m\}} \cdot \mathcal{L} \leq \liminf_{n\to\infty} 1_{\{\mathcal{X}>1/m\}} \cdot \mathcal{L}^n = 0.$$

Mit $m \to \infty$ folgt die zweite Behauptung. Die letzte Behauptung folgt, wenn wir Lemma 20.19 für den Prozess $|X_0| + \int_0^t \operatorname{sgn}(X_s) dX_s$ anstelle von anwenden. Es folgt nämlich direkt, dass die Funktion g gerade die Lokalzeit $\mathcal{L}$ sein muss. □

Korollar 20.20. *Für die gleiche Situation wie in Theorem 20.18 gilt*

$$X_t^+ - X_0^+ = \int_0^t 1_{\{X_s>0\}} dX_s + \tfrac{1}{2} L_t,$$
$$X_t^- - X_0^- = \int_0^t 1_{\{X_s<0\}} dX_s + \tfrac{1}{2} L_t.$$

Beweis. Wir bemerken nur, dass $x^+ = \frac{1}{2}(x + |x|)$ und $x^- = \frac{1}{2}(x - |x|)$. □

Als Anwendung der Lokalzeit erinnern wir an Theorem 17.8. Dort haben wir für eine in 0 gestartete Brown'sche Bewegung $(X_t)_{t\geq 0}$ und $M_t = \sup_{s\leq t} X_s$ gezeigt, dass $M_t - X_t \stackrel{d}{=} |X_t|$. In Bemerkung 17.9 haben wir behauptet, dass diese Gleichheit in Verteilung auch entlang des ganzen Pfades gilt. Dies können wir nun nachprüfen.

Proposition 20.21 (Pfadwertiges Reflexionsprinzip). *Sei $\mathcal{X}$ eine in 0 gestartete Brown'sche Bewegung und $\mathcal{L}$ wie in Theorem 20.18 die Lokalzeit in 0. Weiter ist $\mathcal{M} = (M_t)_{t\geq 0}$ mit $M_t = \sup_{s\leq t} X_s$. Dann gilt*

$$(\mathcal{L}, |\mathcal{X}|) \stackrel{d}{=} (\mathcal{M}, \mathcal{M} - \mathcal{X}).$$

Beweis. Wir definieren $\mathcal{X}' = -\operatorname{sgn}\mathcal{X} \cdot \mathcal{X}$. Dann ist $\mathcal{X}'$ ein stetiges lokales Martingal mit $[\mathcal{X}] = (\operatorname{sgn}(\mathcal{X})^2 \cdot [\mathcal{X}] = \lambda$. Damit ist $\mathcal{X}'$ ebenfalls eine Brown'sche Bewegung. Weiter ist

$$M_t' := \sup_{s\leq t} X_s' = -\inf_{s\leq t} (\operatorname{sgn}\mathcal{X} \cdot \mathcal{X})_s = L_t$$

und nach der Tanaka-Formel

$$|\mathcal{X}| = -\mathcal{X}' + \mathcal{L} = \mathcal{M}' - \mathcal{X}'.$$

Daraus folgt $(\mathcal{L}, |\mathcal{X}|) = (\mathcal{M}', \mathcal{M}' - \mathcal{X}') \stackrel{d}{=} (\mathcal{M}, \mathcal{M} - \mathcal{X})$. □

21 Stochastische Differentialgleichungen und Diffusionen

In vielen Anwendungsgebieten, etwa der Biologie, Finanzmathematik, oder den Ingenieurswissenschaften, sind die hier behandelten, durch stochastische Differentialgleichungen beschriebenen Prozesse, besonders wichtig. Diese stellt man sich am besten wie die Lösung von gewöhnlichen Differentialgleichungen vor, bei denen allerdings noch ein Rauschterm integriert ist, so dass die Lösung ein stochastischer Prozess ist.

21.1 Stochastische Differentialgleichungen

In diesem Abschnitt studieren wir Gleichungen der Form

$$dX^i = \mu^i(t, \underline{X}_t)dt + \sum_{j=1}^{n} \sigma^{ij}(t, \underline{X})dW_t^j, \tag{21.1}$$

mit $\underline{X}_t = (X_t^1, ..., X_t^d)$, wobei $\mathcal{W} = (\mathcal{W}^j)_{j=1,...,n}$ mit $\mathcal{W}^j = (W_t^j)_{t\geq 0}$ eine n-dimensionale Brown'sche Bewegung ist, und $t \mapsto \mu^i(t, \underline{X}_t)$, $t \mapsto \sigma^{ij}(t, \underline{X}_t)$, $i = 1, ..., d, j = 1, ..., n$ progressiv messbare Prozesse sind. Solche Gleichungen werden auch als stochastische Differentialgleichungen oder SDEs genannt (von *Stochastic Differential Equations*). Abkürzend schreiben wir auch $\underline{\mu} = (\mu^i)_{i=1,...,d}, \underline{\underline{\sigma}} = (\sigma^{ij})_{i=1,...,d,j=1,...,n}$ sowie

$$d\underline{X} = \underline{\mu}(t, \underline{X})dt + \underline{\underline{\sigma}}(t, \underline{X})d\underline{W}.$$

oder in integraler Schreibweise

$$\underline{X}_t = \underline{X}_0 + \int_0^t \underline{\mu}(s, \underline{X}_s)ds + \int_0^t \underline{\underline{\sigma}}(s, \underline{X})d\underline{W}_s$$

oder noch kürzer

$$\underline{\mathcal{X}} = \underline{\mu}(., \underline{\mathcal{X}}) \cdot \lambda + \underline{\underline{\sigma}}(., \underline{X}) \cdot \underline{\mathcal{W}}.$$

Beispiel 21.1 (Einige einfache Fälle).

1. Ist $\sigma^{ij} = 0$ für alle i, j, so wird (21.1) zu

$$dX^i = \mu^i(t, \underline{X})dt, \qquad \text{oder} \qquad \dot{X}^i = \mu^i(t, \underline{X}).$$

Gleichungen dieser Form werden auch als (gewöhnliche) Differentialgleichungen bezeichnet.

2. Eine besonders einfache Differentialgleichung ist $\dot{X} = \mu X$, die bekanntlich durch $X_t = X_0 e^{\mu t}$ gelöst wird. Wie sieht es aber mit

$$dX = \mu X dt + \sigma X dW$$

aus? Wir formen dies zu $d(lnX) = \mu dt + \sigma dW$ um. Diese Formulierung legt nahe, die Itô-Formel auf (den noch zu benennenden) Prozess $(\ln X_t)_{t\geq 0}$ anzuwenden. Das ergibt (wegen $[\mathcal{X}] = [\sigma\mathcal{X} \cdot \mathcal{W}] = \sigma^2\mathcal{X}^2 \cdot \lambda$)

$$d\ln X = \frac{1}{X}dX - \frac{1}{X^2}d[\mathcal{X}] = \mu dt + \sigma dW - \tfrac{1}{2}\sigma^2 dt,$$

oder auch

$$\ln X_t = \ln X_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t.$$

Auflösen nach X_t ergibt

$$X_t = X_0 \exp\Big((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\Big).$$

Der Prozess $\mathcal{X}$ heißt auch *geometrische Brown'sche Bewegung* und wird im Abschnitt über Finanzmathematik ein Modell für die Entwicklung eines Aktienpreises sein.

Es stellt sich heraus, dass Gleichungen der Form (21.1) verschiedene Lösungsbegriffe zulassen.

Definition 21.2 (Starke und schwache Lösungen einer SDE). *Wir schreiben* $\underline{\mathcal{X}} = (\mathcal{X}^i)_{i=1,\dots,d}$ *mit* $\mathcal{X}^i = (X_t^i)_{t\geq 0}$.

1. *Sei* $(\Omega, \mathcal{F}, \mathbf{P})$ *ein Wahrscheinlichkeitsraum, auf dem eine* n*-dimensionale Brown'sche Bewegung* $\underline{\mathcal{W}}$ *definiert ist, sowie* $(\mathcal{F}_t)_{t\geq 0}$ *die von* $\underline{\mathcal{W}}$ *erzeugte Filtration. Ein* $(\mathcal{F}_t)_{t\geq 0}$*-adaptierter Prozess* $\underline{\mathcal{X}}$*, für den* (21.1) *(mit der gegebenen Brown'schen Bewegung* $\underline{\mathcal{W}}$*) gilt, heißt* starke Lösung der SDE (21.1)*.*

2. *Ein Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbf{P})$*, auf dem eine* n*-dimensionale Brown'sche Bewegung* $\underline{\mathcal{W}}$ *mit erzeugter Filtration* $(\mathcal{F}_t)_{t\geq 0}$*, und ein adaptierter Prozess* $\underline{\mathcal{X}}$ *definiert sind, für die* (21.1) *gelten, heißt* schwache Lösung der SDE (21.1)*. Kurz sagen wir auch, dass* $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$ *eine schwache Lösung ist.*

Der Unterschied zwischen einer starken und einer schwachen Lösung ist also, dass bei starken Lösungen $\mathcal{X}$ die Brown'sche Bewegung $\mathcal{W}$ im vorhinein gegeben ist. Bei schwachen Lösungen wird nicht nur der Prozess $\mathcal{X}$ gesucht, der (21.1) löst, sondern auch die Brown'sche Bewegung $\mathcal{W}$, die auf der rechten Seite in (21.1) auftaucht. Klar ist, dass jede starke Lösung auch eine schwache Lösung zulässt. Die Beispiele aus Beispiel (21.1) waren alles starke Lösungen.

Beispiel 21.3 (Eine SDE, die schwach aber nicht stark lösbar ist). Wir betrachten die SDE

$$dX = \operatorname{sgn}(X)dW, \tag{21.2}$$

wobei sgn so wie in (20.3) definiert ist. Da $(\operatorname{sgn}(X))^2 = 1$ gilt, ganz egal wie der Prozess $\mathcal{X}$ aussieht, folgt, dass die Lösung dieser SDE eine Brown'sche Bewegung sein muss. (Schließlich ist die Lösung ein lokales Martingal mit $[\mathcal{X}]_t = t$; siehe Theorem 20.3.)

Wir zeigen nun, dass es eine schwache Lösung von (21.2) gibt: Sei hierfür $\mathcal{X} = (X_t)_{t\geq 0}$ irgendeine Brown'sche Bewegung und $\mathcal{W} = \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{X}$. Dann gilt nach Proposition (19.46)

$$\mathcal{X} = \operatorname{sgn}(\mathcal{X})^2 \cdot \mathcal{X} = \operatorname{sgn}(\mathcal{X}) \cdot \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{X} = \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{W}.$$

Damit ist $(\mathcal{W}, \mathcal{X})$ eine schwache Lösung der SDE (21.2).

Andererseits zeigen wir nun, dass es keine starke Lösung von (21.2) geben kann. Angenommen, $\mathcal{X}$ ist eine starke Lösung von (21.2) für eine gegebene Brown'sche Bewegung $\mathcal{W}$. Dann ist $\mathcal{X}$ eine Brown'sche Bewegung, für die

$$\mathcal{W} = 1 \cdot \mathcal{W} = (\operatorname{sgn}(\mathcal{X}))^2 \cdot \mathcal{W} = \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{X} = |\mathcal{X}| - |X_0| - \mathcal{L}_{\mathcal{X}}$$

wegen (20.2) gilt, wobei $\mathcal{L}_{\mathcal{X}}$ die Lokalzeit von $\mathcal{X}$ im Punkt 0 ist. Insbesondere folgt daraus, dass $\sigma(X_s : s \leq t) \supsetneq \mathcal{F}_t$ ist. (Schließlich lässt sich aus der rechten Seite der letzten Gleichung nicht ablesen, in welche Richtung $\mathcal{X}$ bei Treffpunkten der 0 gegangen ist.) Dies ist im Widerspruch dazu, dass $\mathcal{X}$ an $(\mathcal{F}_t)_{t\geq 0}$ adaptiert ist. Also kann es keine starke Lösung von (21.2) geben.

Wir kommen nun zum Existenz- und Eindeutigkeitssatz für starke Lösungen von SDEs. Wir werden im folgenden die Notation $\lesssim$ aus Bemerkung 11.13 verwenden.

Theorem 21.4 (Starke Lösungen von SDEs). *Gegeben sei die SDE*

$$X_t^i = X_0^i + \int_0^t \mu^i(s, \underline{X}_s)ds + \sum_{j=1}^n \int_0^t \sigma^{ij}(s, \underline{X}_s)dW_s^j \tag{21.3}$$

auf $[0, \tau]$ *mit einer Brown'schen Bewegung* $\underline{\mathcal{W}} = (\mathcal{W}^j)_{j=1,...,n}$*, mit* $\mu^i, i = 1, ..., d$ *sowie* $\sigma^{ij}, i = 1, ..., d, j = 1, ..., n$*, für die*

$$||\underline{\mu}(t, \underline{x})||_2 + ||\underline{\underline{\sigma}}(t, \underline{x})||_2 \lesssim 1 + ||\underline{x}||_2, \tag{21.4}$$

$$||\underline{\mu}(t, \underline{x}) - \underline{\mu}(t, \underline{y})||_2 + ||\underline{\underline{\sigma}}(t, \underline{x}) - \underline{\underline{\sigma}}(t, \underline{y})||_2 \lesssim ||\underline{x} - \underline{y}||_2 \tag{21.5}$$

für $\underline{x}, \underline{y} \in \mathbb{R}^d, 0 \leq t \leq \tau$ *(wobei* $||\underline{\underline{\sigma}}||_2^2 = \sum(\sigma^{ij})^2$*) gilt. Weiter sei* $\underline{Z}$ *eine* $\mathbb{R}^d$*-wertige Zufallsvariable mit* $\mathbf{E}[||\underline{Z}||_2^2] < \infty$*, die unabhängig von der von* $\underline{\mathcal{W}}$ *generierten* σ*-Algebra ist. Dann gibt es eine eindeutige, an* $(\sigma((\underline{W}_s)_{s\leq t}, Z))_{t\geq 0}$ *adaptierte,* t*-stetige Lösung* $\underline{\mathcal{X}}^{\underline{Z}} = (X_t^{\underline{Z}})_{t\geq 0}$ *von* (21.3) *mit* $\underline{X}_0^{\underline{Z}} = Z$ *und* $\mathbf{E}\big[\int_0^\tau ||\underline{X}_t||_2^2 dt\big] < \infty$*. Außerdem ist die Abbildung* $\underline{z} \mapsto \underline{\mathcal{X}}^{\underline{z}}$ *(für deterministisches* $\underline{z}$*) stetig in* $\underline{z}$ *(wobei der Raum der Pfade von* $\underline{\mathcal{X}}$ *mit der uniformen Konvergenz auf Kompakta ausgestattet ist).*

Bevor wir das Theorem beweisen, brauchen wir zwei Lemmata.

Lemma 21.5. *Wir bezeichnen mit* $(t, \underline{x}) \mapsto S^i(t, \underline{x})$ *die* i*-te Zeile der rechten Seite der SDE* (21.3)*. Dann gilt unter denselben Voraussetzungen wie in Theorem 21.4 für zwei stetige, adaptierte Prozesse* $(\underline{X}_t)_{t\geq 0}$ *und* $(\underline{Y}_t)_{t\geq 0}$

$$\mathbf{E}[\sup_{0\leq s\leq t}(S^i(s, \underline{X}_s) - S^i(s, \underline{Y}_s))^2] \leq c \cdot \mathbf{E}[||\underline{X}_0 - \underline{Y}_0||_2^2] + c(1+t) \cdot \int_0^t \mathbf{E}[||\underline{X}_s - \underline{Y}_s||_2^2]ds,$$

wobei $c > 0$ *eine Konstante ist, die nicht von den beiden Prozessen abhängt.*

Beweis. Wir verwenden die Doob'sche Ungleichung und die Lipschitz-Stetigkeit der Funktionen μ^i und σ^{ij} für

$$\begin{aligned}
\mathbf{E}[\sup_{0\le s\le t}(S^i(\underline{X}_s)-S^i(\underline{Y}_s))^2] &\lesssim \mathbf{E}[(\underline{X}_0^i-\underline{Y}_0^i)^2]+\mathbf{E}\Big[\sup_{0\le s\le t}\Big(\int_0^s \mu^i(r,\underline{X}_r)-\mu^i(r,\underline{Y}_r)dr\Big)^2\Big]\\
&\quad+\mathbf{E}\Big[\sup_{0\le s\le t}\Big(\int_0^s\sum_{j=1}^n \sigma^{ij}(r,\underline{X}_r)-\sigma^{ij}(r,\underline{Y}_r)dW_r^j\Big)^2\Big]\\
&\lesssim \mathbf{E}[||\underline{X}_0-\underline{Y}_0||_2^2]+\mathbf{E}\Big[\Big(\int_0^t |\mu^i(s,\underline{X}_s)-\mu^i(s,\underline{Y}_s)|ds\Big)^2\Big]\\
&\quad+\mathbf{E}\Big[\Big(\int_0^t\sum_{j=1}^n \sigma^{ij}(s,\underline{X}_s)-\sigma^{ij}(s,\underline{Y}_s)dW_s^j\Big)^2\Big]\\
&= \mathbf{E}[||\underline{X}_0-\underline{Y}_0||_2^2]+\mathbf{E}\Big[\Big(\int_0^t |\mu^i(s,\underline{X}_s)-\mu^i(s,\underline{Y}_s)|ds\Big)^2\Big]\\
&\quad+\mathbf{E}\Big[\int_0^t\sum_{j=1}^n (\sigma^{ij}(s,\underline{X}_s)-\sigma^{ij}(s,\underline{Y}_s))^2ds\Big]\\
&\lesssim \mathbf{E}[||\underline{X}_0-\underline{Y}_0||_2^2]+\mathbf{E}\Big[\Big(\int_0^t ||\underline{X}_s-\underline{Y}_s||_1 ds\Big)^2\Big]\\
&\quad+\mathbf{E}\Big[\int_0^t ||\underline{X}_s-\underline{Y}_s||_2^2 ds\Big]\\
&\lesssim \mathbf{E}[||\underline{X}_0-\underline{Y}_0||_2^2]+(1+t)\int_0^t \mathbf{E}\big[||\underline{X}_s-\underline{Y}_s||_2^2\big]ds.
\end{aligned}$$

□

Lemma 21.6 (Gronwall-Lemma). *Sei $f:[0,\infty)\to\mathbb{R}$ eine stetige Funktion mit*

$$f(t)\le a+b\int_0^t f(s)ds \qquad \textit{oder} \qquad \dot f\le bf$$

für $a,b\ge 0$. Dann gilt

$$f(t)\le ae^{bt}$$

für alle $t\ge 0$.

Beweis. Aus der Voraussetzung folgt sofort

$$\frac{d}{dt}\Big(e^{-bt}\int_0^t f(s)ds\Big)=\Big(f(t)-b\int_0^t f(s)ds\Big)e^{-bt}\le ae^{-bt}.$$

Integration liefert

$$\int_0^t f(s)ds\le e^{bt}a\int_0^t e^{-bs}ds=\frac{a}{b}(e^{bt}-1).$$

Durch erneutes Ableiten bei t folgt das Resultat. □

Beweis von Theorem 21.4. Zunächst zeigen wir die Eindeutigkeit. Angenommen, es gibt zwei Lösungen $\underline{\mathcal{X}}^{\underline{Z}}=(X_t)_{t\ge 0}$ und $\underline{\mathcal{Y}}^{\underline{Z}}=(Y_t)_{t\ge 0}$ von (21.3) mit demselben Startwert $\underline{Z}$. Dann folgt mit Lemma 21.5, dass

$$\begin{aligned}\mathbf{E}[\sup_{0\le s\le t}||\underline{X}_s-\underline{Y}_s||_2^2] &\le \sum_{i=1}^d \mathbf{E}[\sup_{0\le s\le t}(S^i(s,\underline{X}_s)-S^i(s,\underline{Y}_s))^2]\\ &\lesssim (1+t)\int_0^t \mathbf{E}[\sup_{0\le s\le t}||\underline{X}_s-\underline{Y}_s||_2^2].\end{aligned}\tag{21.6}$$

Also gilt für die Funktion $v : t \mapsto \mathbf{E}[\sup_{0\le s\le t}||\underline{X}_s-\underline{Y}_s||_2^2|||]$ (eingeschränkt auf $t\in[0,\tau]$), dass $\dot v \le cv$ für eine Konstante c ist und $v(0)=0$. Mit dem Gronwall-Lemma folgt, dass $v=0$ gilt. Damit ist insbesondere $\underline{\mathcal{X}}^{\underline{Z}}=\underline{\mathcal{Y}}^{\underline{Z}}$.

Kommen wir nun zur Existenz der starken Lösung. Hierzu definieren wir rekursiv Prozesse $\underline{\mathcal{X}}^0=(\underline{X}_t^0)_{t\ge0}, \underline{\mathcal{X}}^1=(\underline{X}_t^1)_{t\ge0},\ldots$ mit $\underline{\mathcal{X}}^0=\underline{Z}$, $\underline{X}_t^{n+1}=\underline{S}(t,\underline{X}_t^n)$ mit $\underline{S}=(S^i)_{i=1,\ldots,d}$ (und bemerken, dass eine Lösung von (21.3) genau $\underline{X}_t=\underline{S}(t,\underline{X}_t)$ erfüllt). Dann gilt

$$\begin{aligned}\mathbf{E}[||\underline{X}_t^1-\underline{X}_t^0||_2^2]=\mathbf{E}[||\underline{S}(t,\underline{Z})-\underline{Z}||_2^2] &= \mathbf{E}\Big[\int_0^t \underline{\mu}(s,\underline{Z})ds+\int_0^t \underline{\underline{\sigma}}(s,\underline{Z})d\underline{W}_s\Big]\\ &\lesssim t(1+\mathbf{E}[||\underline{Z}||_2^2]).\end{aligned}$$

wegen (21.4), sowie

$$\begin{aligned}\mathbf{E}[||\underline{X}_t^{k+1}-\underline{X}_t^k||_2^2] &= \mathbf{E}[||\underline{S}(t,\underline{X}_t^k)-\underline{S}(t,\underline{X}_t^{k-1})||_2^2]\\ &\lesssim (1+t)\int_0^t \mathbf{E}[||\underline{X}_s^k-\underline{X}_s^{k-1}||_2^2]ds\end{aligned}$$

wegen Lemma 21.5. Mit Induktion folgt nun die Existenz einer Konstanten $c>0$, so dass

$$\mathbf{E}[||\underline{X}_t^{k+1}-\underline{X}_t^k||_2^2]\le\frac{c^{k+1}t^k}{(k+1)!}$$

für $k=0,1,2,\ldots$ und $0\le t\le\tau$. Nun folgt, wieder mit Lemma 21.5

$$\begin{aligned}\mathbf{P}(\sup_{0\le t\le\tau}||\underline{X}_t^{k+1}-\underline{X}_t^k||_2>2^{-k}) &\le 4^k\cdot\mathbf{E}[\sup_{0\le t\le\tau}||\underline{S}(t,\underline{X}_t^k)-\underline{S}(t,\underline{X}_t^{k-1})||_2^2]\\ &\lesssim 4^k\int_0^\tau \mathbf{E}[||\underline{X}_t^k-\underline{X}_t^{k-1}||_2^2]dt\le\frac{(4c\tau)^k}{k!}.\end{aligned}$$

Da die rechte Seite summierbar ist, folgt aus dem Borel-Cantelli-Lemma, dass $\sup_{0\le t\le\tau}||\underline{X}_t^l-\underline{X}_t^k||_2 \xrightarrow{k,l\to\infty}_{fs} 0$. Deshalb konvergiert $\underline{\mathcal{X}}^k$ uniform auf $[0,\tau]$ gegen einen Prozess $\underline{\mathcal{X}}$. Mit Lemma 21.5 zeigt man außerdem, dass die Konvergenz ebenfalls in L^2 gilt, so dass $\mathcal{X}$ die Eigenschaft $\mathbf{E}\big[\int_0^\tau||\underline{X}_t||_2^2dt\big]<\infty$ besitzt. Da $\underline{X}^k, k=1,2,\ldots$ an $(\sigma((\underline{W}_s)_{s\le t},Z))_{t\ge0}$ adaptiert und stetig ist, gilt dasselbe für $\mathcal{X}$. Es bleibt zu zeigen, dass $\mathcal{X}$ die SDE (21.3) löst. Dies ist jedoch klar, da $\underline{X}_t=\lim_{k\to\infty}\underline{X}_t^k=\lim_{k\to\infty}\underline{X}_t^{k+1}=\lim_{k\to\infty}\underline{S}(t,\underline{X}_t^k)=\underline{S}(t,\underline{X}_t)$ wegen der Stetigkeit von $\underline{S}$, die aus Lemma 21.5 folgt. Mit einer zu (21.6) analogen Rechnung folgt schließlich auch die Stetigkeit von $\underline{z}\mapsto\mathcal{X}^{\underline{z}}$ uniform auf Kompakta. □

Bemerkung 21.7 (Itô-Diffusionen). Gegeben sei eine SDE (21.3), wobei $\underline{\mu}$ und $\underline{\underline{\sigma}}$ nicht von t abhängt, also $\underline{\mu}(t,\underline{x})=\underline{\mu}(\underline{x})$ und $\underline{\underline{\sigma}}(t,\underline{x})=\underline{\underline{\sigma}}(\underline{x})$. In diesem Fall heißt die Lösung der SDE eine *Itô-Diffusion*. (Allgemeiner heißt ein starker Markov-Prozess mit stetigen Pfaden eine Diffusion. Es gibt auch noch andere Diffusionen als Itô-Diffusionen, etwa ist $|\mathcal{X}|$ eine Diffusion, wenn $\mathcal{X}$ eine Brown'sche Bewegung ist.)

In gewissem Sinn sind alle Lösungen von SDEs Itô-Diffusionen: führt man nämlich eine neue Variable $X_t^{d+1} := t$ ein und ergänzt die SDE (21.3) um $dX_t^{d+1} = dt$, und schreibt man $\underline{\mu}(t; x_1, ..., x_d) = \underline{\mu}(x_{d+1}; x_1, ..., x_d)$ sowie $\underline{\underline{\sigma}}(t; x_1, ..., x_d) = \underline{\underline{\sigma}}(x_{d+1}; x_1, ..., x_d)$, stellt man fest, dass die rechten Seiten der letzten beiden Gleichungen unabhängig von t sind. Aus diesem Grund werden wir uns im Folgenden nur mit Itô-Diffusionen beschäftigen, also mit SDEs der Form

$$d\underline{X} = \underline{\mu}(\underline{X}_t)dt + \underline{\underline{\sigma}}(\underline{X}_t)d\underline{W}_t. \tag{21.7}$$

21.2 Martingalprobleme

Gegeben sei eine (schwache) Lösung $\underline{\mathcal{X}}$ der SDE (21.7), sowie der Operator $G^{\underline{\mathcal{X}}} : \mathcal{C}_b^{(2)}(\mathbb{R}^d) \to \mathcal{C}(\mathbb{R}^d)$ (hier ist $\mathcal{C}_b^{(2)}(\mathbb{R}^d)$ der Raum der reellwertigen, zweimal beschränkt stetig differenzierbaren Funktionen auf $\mathbb{R}^d$), gegeben durch

$$(G^{\underline{\mathcal{X}}}f)(\underline{x}) = \sum_{i=1}^{d} \mu^i(\underline{x})\frac{\partial f}{\partial x_i}(\underline{x}) + \tfrac{1}{2}\sum_{i,j}(\sigma^{ij})^2(\underline{x})\frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}) \tag{21.8}$$

mit $(\sigma^{ij})^2 = \sum_{k=1}^{n} \sigma^{ik}\sigma^{jk}$. Wenden wir die Itô-Formel an, so erhalten wir

$$\begin{aligned} df(\underline{X}_t) &= \sum_{i=1}^{d} \frac{\partial f}{\partial X_t^i} dX^i + \tfrac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_j} d[\mathcal{X}^i, \mathcal{X}^j]_t \\ &= \sum_{i=1}^{d} \mu^i(\underline{X}_t)\frac{\partial f}{\partial X_t^i}dt + \tfrac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_j}(\sigma^{ij})^2(\underline{X}_t)dt + dM_t^f \end{aligned}$$

für das lokale Martingal $\mathcal{M}^f = (M_t^f)_{t\geq 0}$, wobei

$$M_t^f = \sum_{i=1}^{d}\sum_{j=1}^{n}\int_0^t \frac{\partial f}{\partial x_i}(\underline{X}_s)\sigma^{ij}(\underline{X}_s)dW_s^j.$$

Mit anderen Worten ist also

$$\mathcal{M}^f := \Big(f(\underline{X}_t) - f(\underline{X}_0) - \int_0^t (G^{\underline{\mathcal{X}}}f)(\underline{X}_s)ds\Big)_{t\geq 0} \tag{21.9}$$

ein lokales Martingal für alle $f \in \mathcal{C}_b^{(2)}(\mathbb{R}^d)$. Das Konzept der (lokalen) Martingalprobleme stellt genau diese Eigenschaft heraus. Es wird also gefragt: Gegeben ein Operator $G^{\underline{\mathcal{X}}}$ (z.B. der Form (21.8)), gibt es dann einen Wahrscheinlichkeitsraum mit einem stochastischen Prozess $\underline{\mathcal{X}}$, so dass $\mathcal{M}^f$ aus (21.9) für alle $f \in \mathcal{D}(G^{\mathcal{X}})$ ein lokales Martingal ist?

Definition 21.8 (Lokales Martingal-Problem). *Sei E ein Polnischer Raum, $\mathbf{P}_0$ ein Wahrscheinlichkeitsmaß auf E und G ein Operator mit Domain $\mathcal{D}(G) \subseteq \mathcal{B}(E)$ (der Menge der beschränkten messbaren Funktionen). Sind für einen stochastischen Prozess $\mathcal{X} = (X_t)_{t\geq 0}$ mit Pfaden in $\mathcal{C}_E[0,\infty)$ und Zustandsraum E alle abgeleiteten Prozesse $\mathcal{M}^f = (M_t^f)_{t\geq 0}$,*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (Gf)(X_s)ds \tag{21.10}$$

lokale Martingale, und ist $X_0 \stackrel{d}{=} \mathbf{P}_0$, so heißt $\mathcal{X}$ Lösung des lokalen $(\mathbf{P}_0, G, \mathcal{D})$ Martingalproblems (auf $\mathcal{C}_E[0,\infty)$).

Sind alle $\mathcal{M}^f$ sogar Martingale, dann sagt man, dass $\mathcal{X}$ das $(\mathbf{P}_0, G, \mathcal{D})$ Martingalproblem löst. (Ist keine Verteilung für X_0 spezifiziert, spricht man auch vom (lokalen) $(G,\mathcal{D})$-Martingalproblem.) Gibt es genau eine Lösung $\mathcal{X}$ für das (lokale) $(\mathbf{P}_0, G, \mathcal{D})$ Martingalproblem, so sagt man, dass es gut gestellt sei.

Beispiel 21.9 (Ein einfaches Martingalproblem). Sei $E = \mathbb{R}^d$, $\mathbf{P}_0 = \delta_{\underline{0}}$ und G der Operator auf $\mathcal{D}(G) = \{\underline{x} \mapsto x_i,\ \underline{x} \mapsto x_i x_j;\ i,j = 1,...,d\}$ mit $Gx_i = 0, Gx_ix_j = \delta_{ij}$. Dann löst nur die d-dimensionale Brown'sche Bewegung das $(\mathbf{P}_0, G, \mathcal{D})$-Martingalproblem. Insbesondere ist dieses gut gestellt.

Denn: Nach Theorem 20.3 ist ein Prozess $\underline{\mathcal{X}} = (\underline{X}_t)_{t\geq 0}, \underline{X}_t = (X_t^i)_{i=1,...,d}$ genau dann eine d-dimensionale Brown'sche Bewegung, wenn $(X_t^i)_{t\geq 0}$ als auch $(X_t^iX_t^j - \delta_{ij}t)_{t\geq 0}$ lokale Martingale sind. Letzteres ist genau dann der Fall, wenn $\mathcal{X}$ das $(G,\mathcal{D})$-Martingalproblem löst.

Schwache Lösungen von SDEs der Form (21.3) und Lösungen von lokalen Martingalproblemen für Generatoren der Form (21.8) hängen eng zusammen, wie folgendes Resultat zeigt.

Theorem 21.10 (Lokale Martingalprobleme und schwache Lösungen von SDEs). *Seien $\underline{\mu} = (\mu^1, ..., \mu^d)$ und $\underline{\underline{\sigma}} = (\sigma^{ij})_{i=1,...,d,j=1,...,n}$ stetige Funktionen und $\mathbf{P}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{C}_{\mathbb{R}^d}[0,\infty)$. Dann hat die SDE (21.3) genau dann eine schwache Lösung mit Verteilung $\mathbf{P}$, wenn ein nach $\mathbf{P}$ verteilter Prozess das lokale Martingalproblem für $(G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$ mit G wie in (21.8) löst.*

Beweis. Ist $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$ eine schwache Lösung von (21.3), dann haben wir bereits am Anfang dieses Abschnitts unter (21.8) gezeigt, dass die Verteilung von $\mathcal{X}$ das lokale $(G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$-Martingalproblem löst.

Sei andersherum $\underline{\mathcal{X}}$ eine Lösung des lokalen $(G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$ Martingalproblems. Setzt man $\underline{x} \mapsto x_i$ in G ein, so sieht man, dass $\mathcal{M}^i = (M_t^i)_{t\geq 0}$ mit

$$M_t^i = X_t^i - X_0^i - \int_0^t \mu^i(\underline{X}_s)ds$$

lokale Martingale sind. (Formal ist es erst einmal nur erlaubt, beschränkte Funktionen in G einzusetzen. Daraus folgt durch Einsetzen von beschränkten Funktionen f mit $f(\underline{x}) = x_i$ auf $A_n := \{\underline{x} : ||\underline{x}||_2 \leq n\}$ erstmal, dass $(M_{t\wedge T}^i)_{t\geq 0}$ ein lokales Martingal ist, wobei T die Treffzeit von A_n ist. Aus Lemma 19.27, folgt, dass $\mathcal{M}^i$ ein lokales Martingal ist.) Setzt man weiter $\underline{x}_i x_j$ in G ein, sieht man analog, dass $\mathcal{M}^{ij} = (M_t^{ij})_{t\geq 0}$ mit

$$M_t^{ij} = X_t^iX_t^j - X_0^iX_0^j - \int_0^t X_s^i\mu^j(\underline{X}_s) + X_s^j\mu^i(\underline{X}_s) - (\sigma^{ij})^2(\underline{X}_s)ds$$

ein lokales Martingal ist. Partielle Integration liefert außerdem

$$\begin{aligned}\mathcal{M}^{ij} &= \mathcal{X}^i \cdot \mathcal{X}^j + \mathcal{X}^j \cdot \mathcal{X}^i + [\mathcal{X}^i, \mathcal{X}^j] - (\mathcal{X}^i\mu^j + \mathcal{X}^j\mu^i + (\sigma^{ij})^2) \cdot \lambda \\ &= \mathcal{X}^i \cdot \mathcal{M}^j + \mathcal{X}^j \cdot \mathcal{M}^i + [\mathcal{M}^i, \mathcal{M}^j] - (\sigma^{ij})^2 \cdot \lambda.\end{aligned}$$

Daraus lesen wir ab, dass

$$[\mathcal{M}^i, \mathcal{M}^j] = (\sigma^{ij})^2 \cdot \lambda = \sum_{k=1}^{n} \sigma^{ik}(\underline{\mathcal{X}})(\sigma^{ik})(\underline{\mathcal{X}}) \cdot \lambda.$$

Aus Theorem 20.10 folgt, dass es eine Erweiterung des Wahrscheinlichkeitsraumes und eine Brown'sche Bewegung $\underline{\mathcal{W}} = (\mathcal{W}^1, ..., \mathcal{W}^n)$ gibt mit

$$\mathcal{X}^i - \mu^i(\underline{\mathcal{X}}) \cdot \lambda = \mathcal{M}^i = \sum_{k=1}^{n} \sigma^{ik}(\underline{\mathcal{X}}) \cdot \mathcal{W}^k.$$

Das bedeutet, dass $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$ die SDE (21.3) löst. □

Schwache Lösungen sind nur von der gemeinsamen Verteilung der zugrunde liegenden Brown'schen Bewegung $\underline{\mathcal{W}}$ und der Lösung der SDE $\underline{\mathcal{X}}$ abhängig. Deshalb ist es möglich, schwache Lösungen von SDEs durch Maßtransformationen, also einem Wechsel der Verteilung der Prozesse, zu erzeugen. Dies geschieht mit der aus Theorem 20.14 bekannten Girsanov-Transformation.

Theorem 21.11 (Die Girsanov-Transformation von Lösungen von SDEs). *Auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ mit Filtration $(\mathcal{F}_t)_{t\geq 0}$ sei $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ mit $\mathcal{X}^i = (X^i_t)_{t\geq 0}$ Lösung der SDE*

$$d\underline{X}_t = \underline{\mu}(\underline{X}_t)dt + \underline{\underline{\sigma}}(\underline{X}_t)d\underline{W}_t. \tag{21.11}$$

für eine $\mathbf{P}$-Brown'sche Bewegung $\underline{\mathcal{W}} = (\mathcal{W}^1, ..., \mathcal{W}^d)$. Weiter sei $\underline{\gamma} : \mathbb{R}^d \to \mathbb{R}^d$ stetig und es gibt einen progressiv messbaren Prozess $\underline{\mathcal{H}}$ in $\mathbb{R}^n$ mit $\underline{\underline{\sigma}}(\underline{\mathcal{X}})\underline{\mathcal{H}} = \underline{\gamma}(\underline{\mathcal{X}})$ und so, dass

$$\mathcal{Z} = \exp\Big(\sum_{j=1}^{n} \mathcal{H}^j \cdot \mathcal{W}^j - \tfrac{1}{2}(\mathcal{H}^j)^2 \cdot \lambda\Big) \tag{21.12}$$

ein Martingal ist. Definiere dann das Maß $\mathbf{Q}$ durch

$$\mathbf{Q} = Z_t \cdot \mathbf{P} \ \textit{auf}\ \mathcal{F}_t.$$

Dann ist der Prozess $\widetilde{\underline{\mathcal{W}}} := \underline{\mathcal{W}} - \underline{\mathcal{H}} \cdot \lambda$ eine Brown'sche Bewegung bezüglich $\mathbf{Q}$ und

$$d\underline{X}_t = (\underline{\mu}(\underline{X}_t) + \underline{\gamma}(\underline{X}_t))dt + \underline{\underline{\sigma}}(\underline{X}_t)d\widetilde{\underline{W}}_t. \tag{21.13}$$

Mit anderen Worten: das Paar $(\widetilde{\underline{W}}, \underline{\mathcal{X}})$ ist schwache Lösung von (21.13) (unter $\mathbf{Q}$).

Beweis. Der Beweis ist eine Anwendung von Theorem 20.14. Zunächst bemerken wir, dass nach diesem Theorem der Prozess $\underline{\mathcal{W}} - [\underline{\mathcal{W}}, \sum \mathcal{H}^j \mathcal{W}^j] = \underline{\mathcal{W}} - \underline{\mathcal{H}} \cdot \lambda$ ein stetiges lokales $\mathbf{Q}$-Martingal ist. Nach Proposition 20.15 hat diese denselben Covariations-Prozess wie eine Brown'sche Bewegung und ist damit nach Theorem 10.33 eine Brown'sche Bewegung unter $\mathbf{Q}$. Außerdem gilt

$$(\underline{\mu}(\underline{\mathcal{X}}) + \underline{\gamma}(\underline{\mathcal{X}})) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{X}}) \cdot \widetilde{\underline{\mathcal{W}}} = (\underline{\mu}(\underline{\mathcal{X}}) + \underline{\gamma}(\underline{\mathcal{X}})) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{X}}) \cdot \underline{\mathcal{W}} - \underline{\underline{\sigma}}(\underline{\mathcal{X}})\underline{\mathcal{H}} \cdot \lambda = \underline{\mathcal{X}}.$$

Damit sind alle Behauptungen gezeigt. □

Korollar 21.12. *Sei* $\underline{\mathcal{X}}, \underline{\mathcal{W}}, \underline{\gamma}, \underline{\mathcal{H}}, \mathcal{Z}$ *wie in Theorem 21.11 und* $\underline{\mathcal{Y}} = (\mathcal{Y}^1, ..., \mathcal{Y}^d)$ *mit* $\mathcal{Y}^i = (Y_t^i)_{t\geq 0}$ *Lösung der SDE*

$$d\underline{Y}_t = (\underline{\mu}(\underline{Y}_t) - \underline{\gamma}(\underline{Y}_t))dt + \underline{\underline{\sigma}}(\underline{Y}_t)d\underline{W}_t. \tag{21.14}$$

Dann gilt

$$d\underline{Y}_t = \underline{\mu}(\underline{Y}_t)dt + \underline{\underline{\sigma}}(\underline{Y}_t)d\underline{\widetilde{W}}_t. \tag{21.15}$$

Mit anderen Worten: Sowohl $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$*, als auch* $(\underline{\widetilde{\mathcal{W}}}, \underline{\mathcal{Y}})$ *lösen die SDE* (21.11).

Beweis. Es ist nur noch zu zeigen, dass $(\underline{\widetilde{\mathcal{W}}}, \underline{\mathcal{Y}})$ die SDE (21.15) löst. Hierzu schreiben wir

$$\underline{\mu}(\underline{\mathcal{Y}}) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{Y}}) \cdot \underline{\widetilde{\mathcal{W}}} = \underline{\mu}(\underline{\mathcal{Y}}) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{Y}}) \cdot \underline{\mathcal{W}} - \underline{\underline{\sigma}}(\underline{\mathcal{Y}})\underline{\mathcal{H}} \cdot \lambda = \underline{\mathcal{Y}}.$$

□

Bemerkung 21.13 (Wann ist $\mathcal{Z}$ ein Martingal?). Nach Theorem 20.14.2 ist $\mathcal{Z}$ aus (21.12) immer ein lokales Martingal. In Anwendungen ist es oft wichtig zu zeigen, dass der durch $\mathcal{Z}$ in (21.12) definierte Prozess ein Martingal ist, da nur durch (echte, d.h. keine lokalen) Martingale ein Maßwechsel durchgeführt werden können. Hierzu genügt es etwa (siehe Bemerkung 19.24.3), wenn $\mathcal{Z}$ beschränkt ist. Eine andere Bedingung für die Martingal-Eigenschaft von $\mathcal{Z}$ heißt Novikov-Bedingung und ist erfüllt, wenn

$$\mathbf{E}\Big[\exp\Big(\tfrac{1}{2}\Big(\sum_{j=1}^n \int_0^t (H_s^j)^2 ds\Big)\Big)\Big] < \infty$$

für alle $t > 0$. Dass hieraus die Martingal-Eigenschaft von $\mathcal{Z}$ folgt, werden wir hier jedoch nicht zeigen.

Beispiel 21.14 (Transformation der geometrischen Brown'schen Bewegung). Folgendes Beispiel spielt in der Finanzmathematik eine wichtige Rolle. Sei $\mathcal{W} = (W_t)_{t\geq 0}$ eine Brown'sche Bewegung. In Beispiel 21.1.2 haben wir gesehen, dass $\mathcal{X} = (X_t)_{t\geq 0}$ mit

$$X_t = \exp\big(\sigma W_t + (\mu - \tfrac{1}{2}\sigma^2)t\big)$$

Lösung der SDE

$$dX = \mu X dt + \sigma X dW$$

ist. Führt man eine Maßtransformation mittels $\mathcal{Z} = (Z_t)_{t\geq 0}$,

$$Z_t = \exp(-\tfrac{\mu}{\sigma} W_t - \tfrac{1}{2}\tfrac{\mu^2}{\sigma^2} dt)$$

zu einem Maß $\mathbf{Q}$ durch (d.h. $\mathbf{Q} = Z_t \cdot \mathbf{P}$ auf $\mathcal{F}_t$), so ist $\widetilde{\mathcal{W}} = \mathcal{W} + \frac{\mu}{\sigma} t$ eine Brown'sche Bewegung unter $\mathbf{Q}$ und $\mathcal{X}$ eine Lösung von

$$dX = \sigma X d\widetilde{W}.$$

Modelliert nun $\mathcal{X}$ einen Aktienpreisprozess, so gibt es das Maß $\mathbf{Q}$, bezüglich dem dieser Preisprozess ein Martingal ist. Ein solches Maß heißt auch Martingalmaß (für den Aktienpreisprozess $\mathcal{X}$).

21.3 Markov-Eigenschaft

Die Lösung einer SDE der Form (21.1) stellt man sich meist so vor, dass die μ^i die Richtung des Prozesses $\mathcal{X}^i$, und σ^{ij} die Fluktuationen angeben, die zur Zeit t nur von $\underline{X}_t$ und dem infinitesimalen Inkrement der Brown'schen Bewegungen abhängen. In dieser Beschreibung wird klar, dass die zukünftige Entwicklung des Prozesses nach t nur von t und $\underline{X}_t$ abhängt. Mit anderen Worten würden wir erwarten, dass $\underline{\mathcal{X}}$ ein Markov-Prozess ist. Wir zeigen nun, dass dieses Bild richtig ist.

Theorem 21.15 (Markov-Eigenschaft bei Martingalproblemen und SDEs). *Seien $\underline{\mu} = (\mu^1, ..., \mu^d)$ und $\underline{\underline{\sigma}} = (\sigma_{ij})_{i=1,...,d,j=1,...,n}$ stetige Funktionen, so dass das lokale $(\mathbf{P}_0, G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$-Martingalproblem mit G wie in (21.8) für jedes Wahrscheinlichkeitsmaß $\mathcal{P}_0$ auf $\mathbb{R}^d$ wohl-gestellt ist mit Lösung $\underline{\mathcal{X}}$. Dann ist $\underline{\mathcal{X}}$ ein Markov-Prozess. Ist außerdem $\underline{x} \mapsto \mathbf{P}_{\delta_{\underline{x}}}$ stetig, so ist $\mathcal{X}$ stark Markov. Der Generator $G^{\underline{\mathcal{X}}}$ von $\underline{\mathcal{X}}$ hat Domain $\mathcal{D}(G^{\underline{X}}) \supseteq \mathcal{C}_b^{(2)}(\mathbb{R}^d)$ und hat die Form (21.8) für $f \in \mathcal{C}_b^{(2)}(\mathbb{R}^d)$.*

Beweis. Sei $\mathcal{M}^f$ wie in (21.10) und $s, t \geq 0$ sowie θ_s der Zeit-Shift-Operator (d.h. $\theta_s((\omega_t)_{t\geq 0}) = (\omega_{t+s})_{t\geq 0}$. Dann gilt für $A \in \mathcal{F}_s$

$$\mathbf{E}[((M_t^f - M_0^f) \circ \theta_s, A] = \mathbf{E}[((M_{t+s}^f - M_s^f), A] = 0,$$

da $\mathcal{M}^f$ ein Martingal ist. Daraus folgt sofort

$$\mathbf{E}[((M_t^f - M_0^f) \circ \theta_s | \mathcal{F}_s] = 0,$$

woraus auch

$$\mathbf{E}[((M_t^f - M_0^f) \circ \theta_s | X_s] = 0,$$

sofort klar ist. Da das lokale Martingalproblem eine eindeutige Lösung hat, legt die gemeinsame Verteilung von $M_t^f - M_0^f$ (für alle f) den Prozess $\mathcal{X}$ (und damit X_t) eindeutig fest. Also gilt für $t \geq 0$

$$\mathbf{P}(X_{t+s} \in .|\mathcal{F}_s) = \mathbf{P}(X_t \circ \theta_s \in .|\mathcal{F}_s) = \mathbf{P}(X_t \circ \theta_s \in .|X_s) = \mathbf{P}(X_{t+s} \in .|X_s).$$

Dies ist jedoch gerade die Markov-Eigenschaft von $\mathcal{X}$. Die starke Markov-Eigenschaft im Falle der stetigen Abbildung $\underline{x} \mapsto \mathbf{P}_{\underline{x}}$ folgt aus Theorem 16.12. Da $\underline{\mathcal{X}}$ nun ein Markov-Prozess ist, gilt mit der Itô-Formel für $f \in \mathcal{C}_b^{(2)}(\mathbb{R}^d)$

$$\begin{aligned}
(G^{\underline{\mathcal{X}}} f)(\underline{x}) &= \lim_{t\to 0} \frac{1}{t} \mathbf{E}[f(\underline{X}_t) - f(\underline{X}_0) | \underline{X}_0 = \underline{x}] \\
&= \lim_{t\to 0} \frac{1}{t} \mathbf{E}\Big[\int_0^t \sum_{i=1}^d \mu^i(\underline{X}_s) \frac{\partial f}{\partial x_i}(\underline{X}_s) ds + \tfrac{1}{2} \sum_{i,j=1}^d (\sigma^{ij})^2 (\underline{X}_s) \frac{\partial^2 f}{\partial x_i \partial x_j} \Big| \underline{X}_0 = \underline{x}\Big] \\
&= \sum_{i=1}^d \mu^i(\underline{x}) \frac{\partial f}{\partial x_i}(\underline{x}) + \tfrac{1}{2} \sum_{i,j=1}^d (\sigma^{ij})^2(\underline{x}) \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x})
\end{aligned}$$

wegen der Stetigkeit von $\underline{\mu}$ und $\underline{\underline{\sigma}}^2$. □

Bemerkung 21.16 (Infinitesimaler Erwartungswert und (Co-)Varianz). Setzen wir $f(\underline{x}) = x^i$ in $G^{\underline{\mathcal{X}}}$ ein, so erhalten wir

$$\lim_{t\to 0} \frac{1}{t}\mathbf{E}[X_t^i - X_0^i | \underline{X} = \underline{x}] = \mu^i(\underline{x}).$$

Mit $f(\underline{x}) = x^i x^j$ gilt

$$\begin{aligned}
&\lim_{t\to 0} \frac{1}{t}\mathbf{E}[(X_t^i - X_0^i)(X_t^j - X_0^j)|\underline{X} = \underline{x}] \\
&\qquad = \lim_{t\to 0} \frac{1}{t}\mathbf{E}[(X_t^i X_t^j - X_0^i X_0^j - X_0^i(X_t^j - X_0^j) - X_0^j(X_t^i - X_0^i)|\underline{X} = \underline{x}] \\
&\qquad = x_i\mu^j(\underline{x}) + x_j\mu^i(\underline{x}) + \sigma_{ij}^2(\underline{x}) - x_j\mu^i(\underline{x}) - x_i\mu^j(\underline{x}) \\
&\qquad = \sigma_{ij}^2(\underline{x}).
\end{aligned}$$

Deshalb heißt $\underline{\mu}(\underline{X})$ der Vektor des *infinitesimalen Erwartungswertes* von $\mathcal{X}$ und $\sigma_{ij}^2(\mathcal{X})$ die Matrix der *infinitesimalen Covarianz* von $\mathcal{X}$. Die Diagonalelemente $(\sigma_{ii}^2(\underline{X}))_{i=1,\dots,d}$ nennt man entsprechend auch *infinitesimale Varianz.*

Interessanterweise können Itô-Diffusionen hilfreich sein, partielle Differentialgleichungen, die durch einen Operator der Form (21.8) gegeben sind, eindeutig zu lösen. Das nächste Resultat zeigt diesen Zusammenhang auf.

Theorem 21.17 (Rückwärts-Gleichung). *Sei $\underline{\mathcal{X}}$ ein Markov-Prozess mit Generator G der Form* (21.8) *und $f \in \mathcal{C}^{(2)}(\mathbb{R})$. Für das Randwertproblem*

$$\frac{\partial u}{\partial t} = Gu, \qquad t > 0, \underline{x} \in \mathbb{R}^d, \tag{21.16}$$

$$u(0, \underline{x}) = f(\underline{x}); \qquad \underline{x} \in \mathbb{R}^d. \tag{21.17}$$

(wobei Gu die Anwendung von G auf die $\underline{x}$-Koordinaten von u beschreibt) ist

$$u(t, \underline{x}) := \mathbf{E}[f(\underline{X}_t)|\underline{X}_0 = \underline{x}]$$

die einzige Lösung.

Beweis. Es ist zu zeigen, dass das gegebene u das Randwertproblem löst, sowie die Eindeutigkeit der Lösung. Sei also $u(t, \underline{x})$ wie angegeben. Dann gilt $u(0, \underline{x}) = f(\underline{x})$ nach Definition und (21.16) gilt nach Definition des Generators. Nun verwenden wir Lemma 16.28, speziell (16.7). Ist nämlich $(T_t^{\underline{\mathcal{X}}})$ die von $\underline{\mathcal{X}}$ erzeugte Halbgruppe, so ist also $u(t, \underline{x}) := (T_t^{\underline{\mathcal{X}}} f)(\underline{x})$. Damit folgt (21.16) direkt aus (16.7).

Nun zur Eindeutigkeit: Sei hierzu w eine Lösung des Randwertproblems, $t > 0$ und $\underline{x} \in \mathbb{R}^d$. Es ist zu zeigen, dass $w(t, \underline{x}) = \mathbf{E}[f(\underline{X}_t)|\underline{X}_0 = \underline{x}]$ gilt. Wir definieren den Prozess $\underline{\mathcal{Y}} = (\underline{Y}_s)_{0\le s\le t}$ mittels $\underline{Y}_s = (t - s, \underline{X}_s)$, wobei $\underline{\mathcal{X}} = (\underline{X}_s)_{0\le s\le t}$ in $\underline{X}_0 = \underline{x}$ startet. Dann ist der Generator von $\mathcal{Y}$ gegeben durch

$$(G^{\underline{Y}} f)(s, \underline{x}) = -\frac{\partial f}{\partial s} + (Gf)(s, \underline{x}),$$

wobei G im letzten Term auf $\underline{x}$ wirkt. Wählen wir speziell $f = w$, so gilt also $(G^{\underline{Y}} w) = 0$, da w das Randwertproblem löst. Damit gilt

$$w(t, \underline{x}) = w(\underline{Y}_0) = \mathbf{E}[w(\underline{Y}_t)] = \mathbf{E}[w(0, \underline{X}_t)|\underline{X}_0 = \underline{x}] = \mathbf{E}[f(\underline{X}_t)|\underline{X}_0 = \underline{x}] = u(t, \underline{x}).$$

$\square$

21.4 Ein-dimensionale Diffusionen

Diffusionen sind zeit-homogene, starke Markov-Prozesse mit stetigen Pfaden. Wie aus Theorem 21.15 und Theorem 21.10 klar wurde, sind beispielsweise eindeutige Lösungen von SDEs, die stetig in den Anfangsbedingungen sind, Diffusionen. Wir hatten solche Prozesse Itô-Diffusionen genannt. Wir wenden uns dem speziellen Fall ein-dimensionaler Itô-Diffusionen zu, d.h. Lösungen von Gleichungen der Form

$$dX = \mu(X)dt + \sigma(X)dW. \tag{21.18}$$

In diesem Abschnitt werden wir durchgehend folgende Annahmen treffen:

1. Die Gleichung (21.18) hat stetiges μ und σ und ist eindeutig lösbar durch einen starken Markov-Prozess,
2. Die Lösung verlässt einen Bereich $[l, r] \subseteq [-\infty, \infty]$ nicht und $\sigma > 0$ auf (l, r).

In Hinblick auf Theorem 21.15 oder auch Theorem 21.4 ist die erste Voraussetzungen sicher dann erfüllt, wenn μ und σ lokal Lipschitz stetig sind, und nicht schneller als linear wachsen. Die zweite Voraussetzung ($\sigma > 0$) bedeutet, dass die Lösung überall in (l, r) *rauscht*, d.h. dass $\mathcal{X}$ nie im Inneren von (l, r) stehen bleibt.

Wie in Bemerkung 21.16 erklärt, heißt μ der infinitesimale Erwartungswert und σ^2 die infinitesimale Varianz von $\mathcal{X}$. Wir haben bereits gesehen, dass die Lösung dieser SDE ein starker Markov-Prozess ist mit Generator

$$(G^{\mathcal{X}} f)(x) = \mu(x) f'(x) + \tfrac{1}{2}\sigma^2(x) f''(x),$$

wobei $f \in \mathcal{C}^{(2)}(\mathbb{R})$. Wir beginnen mit einem Beispiel, das uns in diesem Kapitel begleiten wird.

Beispiel 21.18 (Ornstein-Uhlenbeck Prozess). Wir betrachten die Lösung der SDE $\mathcal{X} = (X_t)_{t\geq 0}$ mit

$$dX = -\mu X dt + dW$$

mit Generator

$$(Gf)(x) = -\mu x f'(x) + \tfrac{1}{2} f''(x).$$

Aus Beispiel 21.18 folgt, dass $\mathcal{X}$ ein Gauss-Prozess ist und $X_t \xRightarrow{t\to\infty} N(0, 1/2\mu)$.

Die Rückwärts-Gleichung besagt im Falle von ein-dimensionalen SDE, dass das Randwertproblem

$$\begin{aligned}\frac{\partial u}{\partial t} &= \mu(x)\frac{\partial u}{\partial x} + \tfrac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2},\\ u(0, x) &= f(x)\end{aligned}$$

gerade die Lösung $u(t, x) = \mathbf{E}[f(X_t)|X_0 = x]$ hat, wobei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Itô-Diffusion ist mit infinitesimalem Erwartungswert μ und infinitesimaler Varianz σ^2. Es gibt noch eine zweite partielle Differentialgleichung, die eine wichtige Rolle spielt. Hierzu erinnern wir daran, dass jeder homogene Markov-Prozess mit Zustandsraum $\mathbb{R}$ eine Familie von Übergangskernen $(\mu_t^{\mathcal{X}})_{t\geq 0}$ definiert.

Theorem 21.19 (Vorwärts-Gleichung). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Itô-Diffusion mit infinitesimalem Erwartungswert μ und infinitesimaler Varianz σ^2. Haben die Übergangskerne $(\mu_t^{\mathcal{X}})_{t\geq 0}$ Dichten, gilt also*

$$\mu_t^{\mathcal{X}}(x, B) = \int_B p_t^{\mathcal{X}}(x, y)dy$$

für eine Familie $(p_t^{\mathcal{X}})_{t\geq 0}$, so dass $t \mapsto p_t^{\mathcal{X}}(x, y)$ zweimal stetig differenzierbar ist, so erfüllen diese die Gleichung

$$\frac{\partial p_t^{\mathcal{X}}(x, y)}{\partial t} = -\frac{\partial}{\partial y}\big(\mu(y)p_t^{\mathcal{X}}(x, y)\big) + \tfrac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma^2(y)p_t^{\mathcal{X}}(x, y)\big).$$

Beweis.

$$\begin{aligned}\frac{\partial p_t^{\mathcal{X}}(x, z)}{\partial t} &= \lim_{s\to 0}\frac{\partial}{\partial s}\int p_t^{\mathcal{X}}(x, y)p_s^{\mathcal{X}}(y, z)dy \\ &= \lim_{s\to 0}\int p_t^{\mathcal{X}}(x, y)\frac{\partial p_s^{\mathcal{X}}(y, z)}{\partial s}dy \\ &= \lim_{s\to 0}\int p_t^{\mathcal{X}}(x, y)\Big(\mu(y)\frac{\partial p_s^{\mathcal{X}}(y, z)}{\partial y} + \tfrac{1}{2}\sigma^2(y)\frac{\partial^2 p_s^{\mathcal{X}}(y, z)}{\partial y^2}\Big)dy \\ &= \lim_{s\to 0}\int -p_s^{\mathcal{X}}(y, z)\Big(\frac{\partial}{\partial y}\big(p_t^{\mathcal{X}}(x, y)\mu(y)\big) - \tfrac{1}{2}\frac{\partial^2}{\partial y^2}\big(p_t^{\mathcal{X}}(x, y)\sigma^2(y)\big)\Big) \\ &= -\frac{\partial}{\partial z}\big(\mu(z)p_t^{\mathcal{X}}(x, z)\big) + \tfrac{1}{2}\frac{\partial^2}{\partial z^2}\big(\sigma^2(z)p_t^{\mathcal{X}}(x, z)\big).\end{aligned}$$

□

Beispiel 21.20 (Stationäre Verteilung des Ornstein-Uhlenbeck Prozesses). Sei $\mathcal{X} = (X_t)_{t\geq 0}$ ein Ornstein-Uhlenbeck-Prozess mit $X_0 = x$. Wie wir bereits in Beispiel 18.39 gesehen haben, ist $X_t \stackrel{d}{=} N(e^{-\mu t}x, \frac{1}{2\mu}(1 - e^{-2\mu t}))$. Insbesondere ist $N(0, \frac{1}{2\mu})$ die stationäre Verteilung des Ornstein-Uhlenbeck Prozesses. Dies werden wir nun nochmal mit Hilfe der Vorwärts-Gleichung, Theorem 21.19 nachrechnen. Wir berechnen mit

$$\begin{aligned}p(y) := p_t(x, y) &= \sqrt{\frac{\mu}{\pi}}\exp\big(-\mu y^2\big), \\ \frac{\partial p}{\partial y}(y) &= -2\mu y p(y), \\ \frac{\partial p^2}{\partial y^2} &= (4\mu^2 y^2 - 2\mu)p(y),\end{aligned}$$

dass

$$\begin{aligned}-\frac{\partial}{\partial y}\big(\mu(y)p(y)\big) + \tfrac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma^2(y)p(y)\big) &= \frac{\partial}{\partial y}\mu y p(y) + \tfrac{1}{2}\frac{\partial^2}{\partial y^2}p(y) \\ &= (\mu - 2\mu^2 y^2)p(y) + \tfrac{1}{2}(4\mu^2 y^2 - 2\mu)p(y) = 0.\end{aligned}$$

Ist also $X_0 \stackrel{d}{=} N(0, \frac{1}{2\mu})$, so ist nach der Vorwärtsgleichung

$$\frac{\partial p_t(x, y)}{\partial t} = 0.$$

Das bedeutet, dass sich die Dichte der Verteilung von X_t in t nicht ändert. Wir haben also (nochmals), dass $N(0, \frac{1}{2\mu})$ eine stationäre Verteilung von $\mathcal{X}$ ist.

Es stellt sich heraus, dass ein-dimensionale Itô-Diffusionen sehr explizite Rechnungen ermöglichen. Hierzu benötigen wir die Skalenfunktion einer ein-dimensionalen Itô-Diffusion.

Proposition 21.21 (Skalenfunktion). *Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Lösung von (21.18) mit Zustandsraum $[l, r]$. Wir erinnern an die Annahme $\sigma > 0$ auf (l, r) und definieren*

$$S(x) := \int^x \exp\Big(-2\int^y \frac{\mu(z)}{\sigma^2(z)}dz\Big), \qquad x \in (l, r), \tag{21.19}$$

wobei die unteren Grenzen der Integrale beliebig in (l, r) sein können. Dann ist $S(\mathcal{X}) = (S(X_t))_{t\geq 0}$ ein lokales Martingal.

Beweis. Zunächst bemerken wir, dass S die gewöhnliche Differentialgleichung

$$\frac{1}{2}\sigma^2 S'' + \mu S' = 0$$

löst. Deswegen erhalten wir mit der Itô-Formel

$$\begin{aligned} S(\mathcal{X}) - S(X_0) &= S'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2}S''(\mathcal{X}) \cdot [\mathcal{X}] \\ &= (S'(\mathcal{X})\mu(\mathcal{X}) + \tfrac{1}{2}\sigma^2(\mathcal{X})S''(\mathcal{X})) \cdot \lambda + S'(\mathcal{X}) \cdot \mathcal{W} = S'(\mathcal{X}) \cdot \mathcal{W}. \end{aligned}$$

□

Sei nun τ_x die Treffzeit in $x \in [l, r]$. Wir erinnern nun an folgende Rechnung für die Brown'sche Bewegung:

Beispiel 21.22 (Treffwahrscheinlichkeiten für die Brown'sche Bewegung). Sei $\mathcal{X} = (X_t)_{t\geq 0}$ eine Brown'sche Bewegung mit $X_0 = x$ und $a < x < b$. Dann gilt wegen des Optional Sampling und Optional Stopping Theorems

$$x = \mathbf{E}_x[X_{\tau_a\wedge\tau_b}] = a\mathbf{P}_x(\tau_a < \tau_b) + b(1 - \mathbf{P}_x(\tau_a > \tau_b)),$$

also

$$\mathbf{P}_x(\tau_a < \tau_b) = \frac{b - x}{b - a}.$$

Diese Rechnung verallgemeinern wir nun mit Hilfe der Skalenfunktion.

Theorem 21.23 (Treffwahrscheinlichkeiten). *Sei $\mathcal{X}$ und S wie in Proposition 21.21, sowie $l < a < x < b < r$. Dann gilt*

$$\mathbf{P}_x(\tau_a < \tau_b) = \frac{S(b) - S(x)}{S(b) - S(a)}$$

sowie

$$\mathbf{P}_x(\tau_b < \tau_a) = \frac{S(x) - S(a)}{S(b) - S(a)}.$$

Beweis. Die Funktion S ist auf (l, r) strikt monoton wachsend. Bezeichnet man mit τ_y^S die erste Treffzeit von $S(\mathcal{X})$ von y, so gilt damit

$$\mathbf{P}_x(\tau_a < \tau_b) = \mathbf{P}_x(\tau^S_{S(a)} < \tau^S_{S(b)}).$$

Da der Prozess $S(\mathcal{X}^{\tau_a \wedge \tau_b})$ ein beschränktes Martingal ist, konvergiert er fast sicher und es gilt

$$\begin{aligned} S(x) &= S(a)\mathbf{P}(\tau^S_{S(a)} < \tau^S_{S(b)} | S(X_0) = S(x)) + S(b)\mathbf{P}(\tau^S_{S(b)} < \tau^S_{S(a)} | S(X_0) = S(x)) \\ &= S(a)\mathbf{P}_x(\tau_a < \tau_b) + S(b)(1 - \mathbf{P}_x(\tau_a < \tau_b)). \end{aligned}$$

Auflösen nach $\mathbf{P}_x(\tau_a < \tau_b)$ ergibt die Behauptung. □

Wir wollen neben den Treffwahrscheinlichkeiten des letzten Theorems auch erwartete Treffzeiten berechnen. Wieder beginnen wir mit einem Beispiel für die Brown'sche Bewegung.

Beispiel 21.24 (Erwartete Treffzeiten der Brown'schen Bewegung). Sei $\mathcal{X} = (X_t)_{t \geq 0}$ eine Brown'sche Bewegung, d.h. eine Itô-Diffusion mit $\mu = 0$ und $\sigma = 1$ und $\tau_a := \inf\{t \geq 0 : X_t = a\}$. Da $(X_t^2 - t)_{t \geq 0}$ ein Martingal ist, gilt

$$\begin{aligned} \mathbf{E}_x[\tau_a \wedge \tau_b] &= \mathbf{E}_x[X^2_{\tau_a \wedge \tau_b}] - \mathbf{E}_x[X^2_{\tau_a \wedge \tau_b} - \tau_a \wedge \tau_b] = a^2\mathbf{P}(\tau_a < \tau_b) + b^2\mathbf{P}_x(\tau_b < \tau_a) - x^2 \\ &= (a^2 - x^2)\frac{b-x}{b-a} + (b^2 - x^2)\frac{x-a}{b-a} \\ &= \frac{(x-a)(b-x)}{b-a}(-a - x + b + x) = (x-a)(b-x). \end{aligned}$$

Dieses Beispiel wollen wir nun verallgemeinern.

Theorem 21.25 (Erwartete Treffzeiten). *Sei $l < a < x < b < r$, sowie $\mathcal{X}$ und S wie in Theorem 21.23. Sei $\tau := \tau_a \wedge \tau_b$ und $g \in \mathcal{C}^{(2)}((l, r))$ so, dass $x \mapsto \mathbf{E}_x\Big[\int_0^\tau g(X_s)ds\Big]$ zweimal stetig differenzierbar ist. Dann gilt*

$$\mathbf{E}_x\Big[\int_0^\tau g(X_s)ds\Big] = \int_a^b G(x, y)g(y)dy$$

mit

$$G(x, y) = \begin{cases} 2\dfrac{S(x) - S(a)}{S(b) - S(a)} \cdot \big(S(b) - S(y)\big)m(y), & x \leq y \leq b, \\[2ex] 2\dfrac{S(b) - S(x)}{S(b) - S(a)} \cdot \big(S(y) - S(a)\big)m(y), & a \leq y \leq x, \end{cases}$$

wobei

$$m(x) := \frac{1}{\sigma^2(x)S'(x)}.$$

Hierbei heißt G die Green-Funktion *von $\mathcal{X}$.*

Bemerkung 21.26 (Interpretation). Der Wert $G(x, y)dy$ sollte man interpretieren als die Zeit, die im Mittel von einem in x gestarteten Pfad in $[y, y + dy)$ verbracht wird, bevor a oder b getroffen wird.

Beispiel 21.27 (Erwartete Treffzeiten der Brown'schen Bewegung). Für $\mu = 0, \sigma = 1$ berechnen wir

$$G(x, y) = \begin{cases} 2\dfrac{x-a}{b-a} \cdot (b - y), & x \leq y \leq b, \\[2ex] 2\dfrac{b-x}{b-a} \cdot (y - a), & a \leq y \leq x, \end{cases}$$

Also folgt

$$\begin{aligned}\mathbf{E}_x[\tau_a \wedge \tau_b] &= 2\frac{b-x}{b-a}\int_a^x (y-a)dy + 2\frac{x-a}{b-a}\int_x^b (b-x)dy\\ &= \frac{(b-x)(x-a)^2}{b-a} + \frac{(x-a)(b-x)^2}{b-a} = \frac{(b-x)(x-a)}{b-a}(x-a+b-x) = (x-a)(b-x),\end{aligned}$$

genau wie in Beispiel 21.24.

Beweis von Theorem 21.25. Wir setzen

$$w(x) = \mathbf{E}_x\Big[\int_0^T g(X_s)ds\Big].$$

Dann gilt mit der Markov-Eigenschaft von $\mathcal{X}$

$$w(x) = \mathbf{E}_x\Big[\int_0^h g(X_s)ds\Big] + \mathbf{E}_x\big[w(X_h)\big].$$

Für kleine h schreiben wir nun, da w zweimal stetig differenzierbar ist,

$$\begin{aligned}\mathbf{E}_x\Big[\int_0^h g(X_s)ds\Big] &= hg(x) + o(h^2),\\ \mathbf{E}_x[w(X_h)] &= \mathbf{E}_x[w(x) + (X_h - x)w'(x) + \tfrac{1}{2}(X_h - x)^2 w''(x) + o(h^2)]\\ &= w(x) + h\big(\mu(x)w'(x) + \tfrac{1}{2}\sigma^2(x)w''(x) + o(h)\big).\end{aligned}$$

Daraus lesen wir ab, dass

$$\mu(x)w'(x) + \tfrac{1}{2}\sigma^2(x)w''(x) = -g(x), \qquad w(a) = w(b) = 0$$

oder äquivalent

$$\begin{aligned}&\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big)\frac{2\mu(x)}{\sigma^2(x)}w'(x) + \exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big)w''(x) = -\frac{2g(x)}{\sigma^2(x)}\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big),\\ &\frac{d}{dx}\Big(\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big)w'(x)\Big) = -\frac{2g(x)}{\sigma^2(x)}\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big).\end{aligned}$$

Es gilt

$$m(x) = \frac{1}{\sigma^2(x)S'(x)} = \frac{1}{\sigma^2(x)}\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big),$$

woraus folgt, dass

$$\frac{d}{dx}\Big(\frac{w'(x)}{S'(x)}\Big) = -2m(x)g(x).$$

Integriert man beide Seiten, so erhält man

$$\begin{aligned}\frac{w'(x)}{S'(x)} &= -2\int_a^x m(y)g(y)dy + \beta,\\ w(x) &= -2\int_a^x S'(\eta)\int_a^\eta m(y)g(y)dyd\eta + \beta\int_a^x S'(\eta)d\eta + \alpha\\ &= -2\int_a^x (S(x)-S(y))m(y)g(y)dy + \beta(S(x)-S(a)) + \alpha\end{aligned}$$

für $\alpha, \beta \in \mathbb{R}$. Da $w(a) = 0$, folgt $\alpha = 0$ und wegen $w(b) = 0$ gilt

$$\beta = \frac{2}{S(b) - S(a)} \int_a^b \big(S(b) - S(y)\big) m(y) g(y) dy$$

Daraus folgt also

$$\begin{aligned} w(x) = 2 \int_a^x & \frac{(S(x) - S(a))(S(b) - S(y)) - (S(x) - S(y))(S(b) - S(a))}{S(b) - S(a)} m(y) g(y) dy \\ & + 2 \int_x^b \frac{(S(x) - S(a))(S(b) - S(y))}{S(b) - S(a)} m(y) g(y) dy. \end{aligned}$$

Vereinfachen zeigt nun das Resultat. □

Beispiel 21.28 (Geometrische Brown'sche Bewegung und der Feller'sche Verzweigungsprozess).

1. Wie wir wissen, ist die Lösung der SDE $dX = \sigma X dW$ das Martingal $(\exp(\sigma W_t - \frac{1}{\sigma}{}^2 t))_{t \geq 0}$. Dieser Prozess, die geometrische Brown'sche Bewegung, trifft nicht die 0. Mit Hilfe des letzten Resultats können wir zumindest zeigen, dass die erwartete Treffzeit unendlich ist. Es gilt nämlich $S(x) = x$ und mit $X_0 = 1$

$$\begin{aligned} \mathbf{E}[\tau_\varepsilon] &= \lim_{R \to \infty} \mathbf{E}[\tau_\varepsilon \wedge \tau_R] \\ &= \lim_{R \to \infty} 2 \int_\varepsilon^1 \frac{(R-1)(y-\varepsilon)}{R-\varepsilon} \frac{1}{\sigma^2 y^2} dy + 2 \int_1^R \frac{(1-\varepsilon)(R-y)}{R-\varepsilon} \frac{1}{\sigma^2 y^2} dy \\ &= 2 \int_\varepsilon^1 \frac{y-\varepsilon}{\sigma^2 y^2} dy + \int_1^\infty \frac{1-\varepsilon}{\sigma^2 y^2} dy \xrightarrow{\varepsilon \to 0} \infty. \end{aligned}$$

   Da der erste Term der Green-Funktion solche Pfade mit $\tau_\varepsilon < \tau_R$ betrachtet, zeigt man analog

$$\mathbf{E}[\tau_0, \tau_0 < \tau_R] = \lim_{\varepsilon \to 0} 2 \int_\varepsilon^1 \frac{(R-1)(y-\varepsilon)}{R-\varepsilon} \frac{1}{\sigma^2 y^2} dy = 2 \int_0^1 \frac{(R-1)y}{R} \frac{1}{\sigma^2 y^2} dy = \infty.$$

2. Die geometrische Brown'sche Bewegung ist ein nicht-negativer Prozess, der nie in endlicher Zeit die 0 erreicht. Dies gilt nicht für den Feller'schen Verzweigungsprozess, die Lösung der SDE

$$dX = \sqrt{X} dW.$$

   (Die Existenz und Eindeutigkeit der Lösung folgt übrigens hier nicht aus Theorem 21.4, sondern aus anderen Überlegungen.) Hier gilt ebenfalls $S(x) = x$, aber mit $x = 1$

$$\mathbf{E}[\tau_0, \tau_0 < \tau_R] = \lim_{\varepsilon \to 0} \mathbf{E}[\tau_\varepsilon, \tau_\varepsilon < \tau_R] = 2 \int_0^1 \frac{(R-1)y}{R} \frac{1}{\sigma^2 y} dy < \infty.$$

   Das bedeutet, dass die 0 von den Pfaden, die R nicht treffen, in fast sicher endlicher Zeit getroffen wird.

Teil V
Finanzmathematik

Wir wollen nun den in Kapitel 13 gegebenen Ausblick über Anwendungen von stochastischen Prozessen in der Finanzmathematik vertiefen. Im Gegensatz zu Kapitel 13 werden die nun behandelten Modelle zeit-stetig sein. Unser erworbenes Wissen über stochastische Integration wird dabei zum Einsatz kommen.

22 Einführung

22.1 Begriffe

Um Modelle für Finanzmärkte mathematisch sauber einzuführen, benötigen wir für das folgende einige Definitionen. Hierzu ist stets ein Wahrscheinlichkeitsaum $(\Omega, \mathcal{F}, \mathbf{P})$ und eine n-dimensionale Brown'sche Bewegung $\underline{\mathcal{W}} = (\mathcal{W}^1, ..., \mathcal{W}^n), \mathcal{W}^j = (W_t^j)_{0 \leq t \leq \tau}$ für ein $\tau > 0$ gegeben. Weiter sei $(\mathcal{F}_t)_{0 \leq t \leq \tau}$ die von $\underline{\mathcal{W}}$ erzeugte Filtration mit $\mathcal{F} = \mathcal{F}_\tau$. Wir modellieren also einen Aktienmarkt zwischen Zeitpunkten 0 und $\tau > 0$.

Definition 22.1 (Finanzmarkt). *Ein* (Finanz-)Markt *ist ein* $(\mathcal{F}_t)_{0 \leq t \leq \tau}$*-adaptierter,* $d+1$*-dimensionaler Prozess* $\underline{\mathcal{X}} = (\mathcal{X}^0, ..., \mathcal{X}^d)$*,* $\mathcal{X}^i = (X_t^i)_{0 \leq t \leq \tau}$ *der Form*

$$\begin{aligned} \mathcal{X}^0 &= X_0^0 + \mathcal{R}\mathcal{X}^0 \cdot \lambda, \\ \mathcal{X}^i &= X_0^i + \mu^i \mathcal{X}^i \cdot \lambda + \mathcal{X}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k, \qquad i = 1, ..., d \end{aligned} \tag{22.1}$$

für adaptierte Prozesse $\mathcal{R} = (R_t)_{0 \leq t \leq \tau}, \mu^i = (\mu_t^i)_{0 \leq t \leq \tau}, \sigma^{ik} = (\sigma_t^{ik})_{0 \leq t \leq \tau}, i = 1, ..., d, k = 1, ..., n$*. Hierbei heißt* $\mathcal{X}^0$ risikolose Geldanlage*. Der Markt heißt* normalisiert*, wenn* $\mathcal{X}_t^0 = 1$ *für alle* $0 \leq t \leq \tau$*. Unter der* Normalisierung *des Marktes* $\underline{\mathcal{X}}$ *versteht man den Prozess* $\underline{\widetilde{\mathcal{X}}} = (\widetilde{\mathcal{X}}^0, ..., \widetilde{\mathcal{X}}^d)$ *mit* $\widetilde{X}_t^i = X_t^i / X_t^0$*.*

Bemerkung 22.2 (Interpretation).

1. In obiger Definition mag es verwundern, dass $\mathcal{X}^0$ als risikolos eingestuft wird, obwohl $\mathcal{R}$ ein stochastischer Prozess sein kann. Wenigstens folgt aus (22.1), dass – im Gegensatz zu $\mathcal{X}^1, ..., \mathcal{X}^n$ die quadratische Variation von $\mathcal{X}^0$ verschwindet.

2. In Anwendungen werden wir meistens annehmen, dass $\mathcal{X}^0, ..., \mathcal{X}^d$ Lösung der SDE

$$\begin{aligned} dX^0 &= r(t, \underline{X}_t) X_t^0 dt, \\ dX^i &= \mu^i(t, \underline{X}_t) X_t^i dt + X_t^i \sum_{k=1}^n \sigma^{ik}(t, \underline{X}_t) dW_t^k, \qquad i = 1, ..., n. \end{aligned} \tag{22.2}$$

ist, wobei $r, \mu^i, \sigma^{ik}, i = 1, ..., d, k = 1, ..., n$ lokal Lipschitz-stetige Funktionen sind. Dieses Marktmodell heißt auch *verallgemeinertes Black-Scholes-Modell.* Da das Modell (22.1) auch Prozesse μ^i, σ^{ik} zulässt, die zur Zeit t nicht nur von $\underline{X}_t$, sondern vom ganzen Pfad $(\underline{X}_s)_{0 \leq s \leq t}$ abhängen, ist dies eine starke – jedoch praktikable – Einschränkung.

3. Die Prozesse $\mathcal{X}^1, ..., \mathcal{X}^d$ sind Preise risikobehafteter Geldanlagen, also etwa Aktien. Diese Preise sind korreliert, wie man aus (22.1) abliest. Es gilt etwa

$$\begin{aligned}[][\mathcal{X}^i, \mathcal{X}^j] &= \Big[\mathcal{X}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k, \mathcal{X}^j \sum_{l=1}^n \sigma^{jl} \cdot \mathcal{W}^l\Big] = \mathcal{X}^i \mathcal{X}^j \sum_{k,l=1}^n \sigma^{ik}\sigma^{jl} \cdot [\mathcal{W}^k, \mathcal{W}^l] \\ &= \mathcal{X}^i \mathcal{X}^j \sum_{k=1}^n \sigma^{ik}\sigma^{jk} \cdot \lambda.\end{aligned}$$

Daraus berechnet sich beispielsweise

$$\mathbf{COV}[X_t^i, X_t^j] = \mathbf{E}\Big[\sum_{k=1}^n \int_0^t X_s^i \sigma_s^{ik} X_s^j \sigma_s^{jk} ds\Big].$$

4. Das Modell (22.1) hat die Eigenschaft, dass alle Preise stetige Semimartingale sind. Um die in der Praxis auch auftretenden großen Kursschwankungen in kleiner Zeit modellieren zu können, kann man auch zu unstetigen Semimartingalen übergehen. Da wir jedoch die Theorie der stochastischen Integration nur für stetige Integatoren entwickelt haben, werden wir solche Modelle hier nicht betrachten.

Definition 22.3 (Portfolio). *In einem Markt der Form* (22.1) *ist ein* Portfolio *ein adaptierter stochastischer Prozess* $\underline{\Delta} = (\Delta^0, ..., \Delta^d)$, $\Delta^i = (\Delta_t^i)_{0 \le t \le \tau}$. *Hierbei ist* Δ_t^i *die Anzahl der Wertpapiere vom Typ* i *zur Zeit* t. *Der* Wert *des Portfolios* $\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) = V_t(\underline{\Delta}, \underline{\mathcal{X}})$ *zur Zeit* t *ist*

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) := \sum_{i=0}^d \Delta_t^i X_t^i =: \langle \underline{\Delta}_t, \underline{X}_t \rangle.$$

Ein Portfolio heißt selbstfinanzierend *(bezüglich* $\underline{\mathcal{X}}$*), wenn*

$$\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) = V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^d \Delta^i \cdot \mathcal{X}^i, \tag{22.3}$$

d.h.

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) - V_0(\underline{\Delta}, \underline{\mathcal{X}}) := \sum_{i=0}^d \int_0^t \Delta_s^i dX_s^i$$

und

$$\begin{aligned}\int_0^t |\Delta_s^0 R_s X_s^0| ds &< \infty, \\ \int_0^t |\Delta_s^i \mu_s^i X_s^i| ds &< \infty, \qquad i = 1, ..., d \\ \int_0^t (\Delta_s^i X_s^i \sigma_s^{ik})^2 ds &< \infty, \qquad i = 1, ..., d, k = 1, ..., n.\end{aligned}$$

Bemerkung 22.4 (Interpretation). Der Begiff des Portfolios ist selbsterklärend. Für den Wert eines (nicht notwendigerweise selbstfinanzierenden) Portfolios, das aus Semimartingalen zusammen gesetzt ist, gilt wegen der Formel für partielle Integration, Theorem 19.48,

$$\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) - V_0(\underline{\Delta}, \underline{\mathcal{X}}) = \sum_{i=0}^{d} \Delta^i \cdot \mathcal{X}^i + \mathcal{X}^i \cdot \Delta^i + [\mathcal{X}^i, \Delta^i].$$

Um selbst-finanzierende Portfolios zu verstehen, ist es hilfreich, zunächst die Zeit zu diskretisieren. Unterteilen wir das Intervall $[0, t]$ in gleich große Intervalle der Länge t/n mit großem n, so ist ein Portfolio selbstfinanzierend, wenn

$$V_{(k+1)t/n}(\underline{\Delta}, \underline{\mathcal{X}}) - V_{kt/n}(\underline{\Delta}, \underline{\mathcal{X}}) = \sum_{i=0}^{d} \Delta^i_{kt/n}(X_{(k+1)t/n} - X_{kt/n}).$$

Das bedeutet, dass Änderungen im Wert des Portfolios nur dadurch zustande kommen, dass sich die Preise der Wertpapiere ändern. Nach erfolgter Preisänderung kann dann eine Veränderung des Portfolios, d.h. eine Um-Investition in eine andere Zusammensetzung von Wertpapieren, erfolgen.

Oftmals ist es praktisch, statt im Marktmodell $\underline{\mathcal{X}}$ im normalisierten Markt $\underline{\widetilde{\mathcal{X}}}$ zu rechnen. Es folgt ein einfaches Lemma, das zeigt, wie das geht und dass selbstfinanzierende Portfolios unter Normalisierung selbstfinanzierend bleiben.

Lemma 22.5 (Normalisierung von Märkten und selbstfinanzierende Strategien). *Seien $\underline{\mathcal{X}}$ und $\underline{\widetilde{\mathcal{X}}}$ wie in Definition 22.1, d.h. insbesondere dass $\widetilde{X}^i_t = X^i_t / X^0_t$. Dann gilt*

$$\begin{aligned}
&\widetilde{\mathcal{X}}^0 = 1,\\
&\widetilde{\mathcal{X}}^i = X^i_0 + (\mu^i - \mathcal{R})\widetilde{\mathcal{X}}^i \cdot \lambda + \widetilde{\mathcal{X}}^i \sum_{k=1}^{n} \sigma^{ik} \cdot \mathcal{W}^k, \qquad i = 1, ..., d.
\end{aligned}$$

Ist weiter $\underline{\Delta}$ ein Portfolio. Dann gilt

$$\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = \frac{1}{\mathcal{X}^0} \mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}})$$

und $\underline{\Delta}$ ist genau dann selbstfinanzierend bezüglich $\underline{\mathcal{X}}$, wenn $\underline{\Delta}$ selbstfinanzierend bezüglich $\underline{\widetilde{\mathcal{X}}}$ ist.

Beweis. Die Fomel $\widetilde{\mathcal{X}}^0 = 1$ ist klar nach Definition. Weiter gilt

$$X^0_t = \exp\Big(\int_0^t R_s ds\Big) = 1 + \int_0^t R_s \exp\Big(\int_0^s R_r dr\Big) ds, \qquad \mathcal{X}^0 = 1 + \mathcal{R}\mathcal{X}^0 \cdot \lambda, \tag{22.4}$$

also auch

$$\widetilde{X}^i_t = X^i_t \exp\Big(-\int_0^t R_s ds\Big), \qquad \frac{1}{\mathcal{X}^0} = 1 - \frac{\mathcal{R}}{\mathcal{X}^0} \cdot \lambda.$$

Mit Hilfe der Formel für partielle Integration schreiben wir (da die quadratische Variation von $\mathcal{X}^0$ verschwindet)

$$\begin{aligned}\widetilde{\mathcal{X}}^i &= X_0^i + \frac{1}{\mathcal{X}^0} \cdot \mathcal{X}^i + \mathcal{X}^i \cdot \frac{1}{\mathcal{X}^0} \\ &= X_0^i + \frac{1}{\mathcal{X}_0} \cdot \Big(\mu^i \mathcal{X}^i \cdot \lambda + \mathcal{X}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k\Big) - \mathcal{X}^i \cdot \mathcal{R}\frac{1}{\mathcal{X}^0} \cdot \lambda \\ &= X_0^i + \widetilde{\mathcal{X}}^i(\mu^i - \mathcal{R}) \cdot \lambda + \widetilde{\mathcal{X}}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k.\end{aligned} \tag{22.5}$$

Sei nun $\underline{\Delta}$ ein (bezüglich $\underline{\mathcal{X}}$) selbstfinanzierendes Portfolio, also $\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) := \sum_i \Delta^i \mathcal{X}^i = \sum_i \Delta^i \cdot \mathcal{X}^i$. Dann gilt, wieder mit partieller Integation,

$$\begin{aligned}\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) &= \sum_{i=0}^d \Delta^i \widetilde{\mathcal{X}}^i = \sum_{i=0}^d \frac{\Delta^i \mathcal{X}^i}{\mathcal{X}^0} = \frac{1}{\mathcal{X}^0} \sum_{i=0}^d \Delta^i \mathcal{X}^i = \sum_{i=0}^d \frac{1}{\mathcal{X}^0}(V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \Delta^i \cdot \mathcal{X}^i) \\ &= V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^d \frac{1}{\mathcal{X}^0}\Big(\Delta^i \mu^i \mathcal{X}^i \cdot \lambda + \Delta^i \mathcal{X}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k\Big) \\ &= V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^d \Delta^i \mu^i \widetilde{\mathcal{X}}^i \cdot \lambda + \Delta^i \widetilde{\mathcal{X}}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k - \Delta^i \mathcal{X}^i \frac{\mathcal{R}}{\mathcal{X}^0} \cdot \lambda \\ &= V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^d \Delta^i \cdot \widetilde{\mathcal{X}}^i.\end{aligned}$$

Die umgekehrte Richtung zeigt man analog. □

Im Folgenden werden wir uns mit Finanzderivaten und deren fairen Preisen befassen, wobei die zugrunde liegenden Aktienpreise bestimmten stochastischen Prozessen folgen. Wir definieren hier noch, was Finanzderivate sind, und was ein fairer Preis ist. Einen solchen kann es jedoch nur in Arbitrage-freien Märkten geben.

Definition 22.6 (Arbitrage). *Wir verwenden die Notation aus Definitionen 22.1 und 22.3.*

1. *Ein Portfolio $\underline{\Delta}$ ist* zulässig, *falls der Wert $\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}})$ fast sicher von unten beschränkt ist.*

2. *Ein zulässiges, selbstfinanzierendes Portfolio $\underline{\Delta}$ heißt* Arbitrage, *falls fast sicher*

$$V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0, \qquad V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0 \qquad \textit{und} \qquad \mathbf{P}(V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) > 0) > 0.$$

Bemerkung 22.7 (Interpretation).

1. Zulässig in obigem Sinne sind nur Potfolios, bei denen der Investor beschränkt viele Schulden aufnehmen kann.

2. Die Definition von Arbitrage haben wir bereits in Bemerkung 13.6 kennen gelernt. Grob gesagt ist eine Arbitrage eine Möglichkeit, risikolos und mit begrenzten finanziellen Mitteln Gewinn zu erzielen. Wir bemerken, dass die Existenz von Arbitrage von den Preispozessen $\underline{\mathcal{X}}$ abhängt. Gibt es keine Arbitrage, sagen wir auch, der Markt $\underline{\mathcal{X}}$ sei arbitrage-frei.

Definition 22.8 (Derivate, erreichbare Strategien und der faire Preis). *Wir verwenden die Notation aus Definitionen 22.1 und 22.3.*

1. *Ein* (Finanz-)Derivat *ist eine $\mathcal{F}_\tau$-messbare, von unten beschränkte, quadratintegierbare Zufallsvariable S.*
2. *Das Derivat S heißt* erreichbar*, falls es ein $h \in \mathbb{R}$ und selbstfinanzierendes Portfolio $\underline{\Delta} = (\Delta^0, ...\Delta^d)$ mit $\mathcal{V}_0(\underline{\Delta}, \underline{\mathcal{X}}) = h$, so dass*
$$S = \mathcal{V}_\tau(\underline{\Delta}, \underline{\mathcal{X}})$$
fast sicher. Ist weiter der Markt arbitrage-frei, so heißt $\mathcal{V}_t(\underline{\Delta}, \underline{\mathcal{X}})$ der faire Preis des Derivats S zur Zeit t. Weiter heißt $\underline{\Delta}$ ein Hedging-Portfolio *von S.*

Bemerkung 22.9 (Der faire Preis). Der Begriff des *fairen Preises* des Derivats S ist durch die Forderung der *Arbitrage-Freiheit* des Marktes zu erklären. Würde – etwa zum Zeitpunkt $t = 0$ – ein Derivat mit fairem Preis h zu einem Preis $\ell > h$ verkauft werden, so könnte man dieses Derivat zum Preis ℓ verkaufen, und mit $h < \ell$ Startkapital eine selbstfinanzierendes Portfolio aufsetzen, das zu jeder Zeit fast sicher denselben Wert besitzt wie das verkaufte Derivat. Sicherlich ist zu Zeit τ noch $\ell - h$ vom Startkapital übrig, und zwar ohne Risiko. Für $\ell < h$ führt eine ähnliche Vorgehensweise auf ein Arbitrage-Möglichkeit. Wir werden auf den Begriff der Arbitrage im weiteren öfter zurück kommen.

22.2 Das Black-Scholes Modell

Das Black-Scholes Modell ist das einfachste zeitstetige Modell eines Aktienmarktes der Form, die in Definition 22.1 gegeben ist. Es besteht aus nur zwei Wertpapieren, wobei der Prozess $\mathcal{R}$ sowie μ^1 und σ^{11} konstant sind. Es gilt also
$$\begin{aligned} dX^0 &= rX^0 dt, \\ dX^1 &= \mu X^1 dt + \sigma X^1 dW \end{aligned} \tag{22.6}$$
für $r, \mu \in \mathbb{R}, \sigma > 0$ und eine Brown'sche Bewegung $\mathcal{W} = (W_t)_{0 \le t \le \tau}$. Diese SDE lässt sich explizit lösen, nämlich durch
$$\begin{aligned} X_t^0 &= X_0^0 e^{rt}, \\ X_t^1 &= X_0^1 \exp\big((\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t\big), \end{aligned} \tag{22.7}$$
wobei X_0^1 der (deterministische) Startpreis der Aktie zur Zeit 0 ist und hier als 1 angenommen wird. Siehe Beispiele 21.1.2 und 21.14. Durch (22.7) sehen wir, dass X_t^1 für jedes t einer logarithmischen Normalverteilung folgt. Weiter sehen wir, dass X_s^1 und $(X_t^1 - X_s^1)/X_s^1$ für $0 \le s \le t$ unabhängig sind. Außerdem gilt mit $s \le t \le u$
$$\frac{X_t^1 - X_s^1}{X_s^1} \stackrel{d}{=} \frac{X_{t-s}^1 - X_0^1}{X_0^1},$$
d.h. die relativen Inkremente sind unabhängig und identisch verteilt. Darüber hinaus sehen wir durch Lemma 22.5, dass die Nomalisierung von (22.6), $\widetilde{\mathcal{X}}^1$ durch
$$d\widetilde{X}^1 = (\mu - r)\widetilde{X}^1 dt + \sigma \widetilde{X}^1 dW, \qquad \widetilde{X}_t^1 = X_0^1 \exp\big((\mu - r - \tfrac{1}{2}\sigma^2)t + \sigma W_t\big) \tag{22.8}$$
gegeben ist. Wir beginnen mit einer wichtigen Beobachtung über den normalisierten Prozess $\widetilde{\mathcal{X}}^1$, die wir eigentlich schon in Beispiel (21.14) gesehen haben. Im ganzen Abschnitt verwenden wir die oben eingeführte Notation.

Proposition 22.10 (Ein Maß, unter dem $\widetilde{\mathcal{X}}^1$ ein Martingal ist). *Wir definieren ein Wahrscheinlichkeitsmaß* $\mathbf{Q}$ *mittels*

$$\mathbf{Q} = Z_\tau \cdot \mathbf{P}, \qquad Z_\tau = \exp\Big(- \frac{\mu - r}{\sigma} W_\tau - \frac{1}{2}\frac{(\mu - r)^2}{\sigma^2}\tau\Big).$$

Dann ist $\widetilde{\mathcal{X}}^1$ *ein* $\mathbf{Q}$*-Martingal, gegeben durch*

$$\widetilde{\mathcal{X}}^1 = \sigma \widetilde{\mathcal{X}}^1 \cdot \widetilde{\mathcal{W}},$$

$\widetilde{\mathcal{W}} = \mathcal{W} + \frac{\mu - r}{\sigma} t$ *ist eine Brown'sche Bewegung bezüglich* $\mathbf{Q}$ *und*

$$\widetilde{X}_t^1 = X_0^1 \exp\Big(\sigma \widetilde{W}_t - \tfrac{1}{2}\sigma^2 t\Big). \tag{22.9}$$

Beweis. Siehe Beispiel 21.14. □

Wie wir in Definition 22.8 gesehen haben, lässt sich der Preis eines Derivats nur dann sinnvoll bestimmen, wenn der Markt arbitrage-frei ist. Dies wollen wir nun für das Black-Scholes Modell zeigen.

Theorem 22.11 (Arbitrage-Freiheit und Preise im Black-Scholes-Modells). *Das Black-Scholes-Modell ist arbitrage-frei. Das bedeutet, dass für jedes von unten beschränkte, selbstfinanzierende Portfolio* $\underline{\Delta}$ *mit* $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$ *und* $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ *fast sicher gilt, dass* $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = 0$*. Weiter ist der faire Preis eines Derivats* S *zur Zeit* t *gegeben durch*

$$\mathbf{E}_{\mathbf{Q}}[e^{-r(\tau - t)} S | \mathcal{F}_t], \tag{22.10}$$

wobei $\mathbf{Q}$ *das Wahrscheinlichkeitsmaß aus Proposition 22.10 ist.*

Beweis. Zunächst zur Arbitrage-Freiheit: Sei $\underline{\Delta}$ ein Portfolio mit $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$ und $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ fast sicher. Da $\underline{\Delta}$ nach Lemma 22.5 auch selbstfinanzierend bezüglich $\underline{\widetilde{\mathcal{X}}}$ ist, gilt

$$\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = \sum_{i=0}^{d} \Delta^i \cdot \widetilde{\mathcal{X}}^i.$$

Da nach Proposition 22.10 die Prozesse $\widetilde{\mathcal{X}}^i, i = 0, 1$ Martingale bezüglich $\mathbf{Q}$ sind, ist $\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})$ ebenfalls ein $\mathbf{Q}$-Martingal. Nach Voraussetzung ist $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ $\mathbf{P}$-fast sicher. Da $\mathbf{P}$ und $\mathbf{Q}$ äquivalent sind, ist also auch $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ $\mathbf{Q}$-fast sicher. Wegen der Martingaleigenschaft von $\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})$ ist außerdem $\mathbf{E}_{\mathbf{Q}}[V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})] = 0$. Dies ist jedoch nur möglich falls $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 0$ $\mathbf{Q}$-fast sicher. Wieder wegen der Äquivalenz von $\mathbf{P}$ und $\mathbf{Q}$ folgt $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 0$ $\mathbf{P}$-fast sicher, und die Arbitragefreiheit ist gezeigt.

Wir zeigen nun die allgemeine Preisformel (22.10). Zunächst ist klar, dass $\mathcal{M} = (M_t)_{0 \leq t \leq \tau}$, gegeben durch $M_t := \mathbf{E}_{\mathbf{Q}}[e^{-r\tau} S | \mathcal{F}_t]$, ein $\mathbf{Q}$-Martingal ist. Nach Theorem 20.9 gibt es einen Prozess $\mathcal{H} = (H_t)_{0 \leq t \leq \tau}$ mit

$$M_t = \mathbf{E}_{\mathbf{Q}}[e^{-r\tau} S] + \int_0^t H_s d\widetilde{W}_s.$$

Nun schreiben wir $\Delta_t^1 := H_t/(\sigma \widetilde{X}_t^1), \Delta_t^0 = M_t - H_t/\sigma$, so dass

$$V_t(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = \Delta_t^0 + \Delta_t^1 \widetilde{X}_t^1 = M_t,$$

$$\mathbf{E_Q}[e^{-r\tau} S] + (\Delta^0 \cdot \widetilde{\mathcal{X}}^0 + \Delta^1 \cdot \widetilde{\mathcal{X}}^1)_t = \mathbf{E_Q}[e^{-r\tau} S] + \int_0^t H_s d\widetilde{W}_s = M_t.$$

Also ist $\underline{\Delta} = (\Delta^1, \Delta^2)$ selbstfinanzierend bezüglich $\underline{\widetilde{\mathcal{X}}}$, nach Lemma 22.5 also auch selsbtfinanzierend bezüglich $\underline{\mathcal{X}}$. Wegen der Definition von $\mathcal{M}$ gilt außerdem $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = M_\tau = e^{-r\tau} S$ fast sicher, also – wieder mit Lemma 22.5 –

$$V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = S$$

fast sicher. Nach Definition ist nunmehr der faire Preis zur Zeit t gegeben durch

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) = e^{rt} V_t(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = e^{rt} M_t = \mathbf{E_Q}[e^{-r(\tau - t)} S | \mathcal{F}_t].$$

Damit sind alle Behauptungen gezeigt. □

Während das letzte Theorem ein allgemeines Resultat für den fairen Preis eines Derivats angibt, benötigt man in der Praxis noch ein Hedging-Portfolio für das Derivat. Dieses Problem wird nun für das Black-Scholes Modell gelöst.

Proposition 22.12 (Hedging-Portfolio im Black-Scholes Modell). *Sei S ein Derivat mit Preis*

$$\mathbf{E_Q}[e^{-r(\tau-t)} S | \mathcal{F}_t] = F(X_t^1, t)$$

für eine glatte Funktion F. Dann ist das Portfolio $\underline{\Delta} = (\Delta_t^0, \Delta_t^1)_{0 \le t \le \tau}$ mit

$$\Delta_t^1 = \frac{\partial F(X_t^1, t)}{\partial x}, \qquad \Delta_t^0 = e^{-rt} F(X_t^1, t) - \Delta_t^1 \widetilde{X}_t^1$$

ein Hedging-Portfolio für S.

Beweis. Wir berechnen sofort den Wert des Portfolios als

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) = \Delta_t^0 X_t^0 + \Delta_t^1 X_t^1 = F(X_t^1, t) - \Delta_t^1 \widetilde{X}_t^1 e^{rt} + \Delta_t^1 X_t^1 = F(X_t^1, t).$$

Bleibt also zu zeigen, dass das Portfolio selbstfinanzierend ist. Hierzu genügt es wegen Lemma 22.5 zu zeigen, dass das Portfolio selbstfinanzierend für $\underline{\widetilde{\mathcal{X}}}$ ist. Wir setzen

$$\widetilde{F}(x, t) := e^{-rt} F(x e^{rt}, t),$$

so dass mit der Itô-Formel

$$\begin{aligned}\widetilde{F}(\widetilde{X}_t^1, t) &= \widetilde{F}(\widetilde{X}_0^1, 0) + \int_0^t \frac{\partial \widetilde{F}(\widetilde{X}_s^1, s)}{\partial x} dX_s^1 + \int_0^t \frac{\partial \widetilde{F}(\widetilde{X}_s^1, s)}{\partial s} ds + \tfrac{1}{2} \int_0^t \frac{\partial^2 \widetilde{F}(\widetilde{X}_s^1, s)}{\partial x^2} d[\widetilde{\mathcal{X}}^1]_s \\ &= \widetilde{F}(\widetilde{X}_0^1, 0) + \int_0^t \frac{\partial \widetilde{F}(\widetilde{X}_s^1, s)}{\partial x} d\widetilde{X}_s^1 + \int_0^t K_s ds\end{aligned}$$

für einen Prozess $(K_t)_{0 \le t \le \tau}$ gilt. Da $(\widetilde{F}(\widetilde{X}_t^1, t))_{0 \le t \le \tau} = e^{-r\tau} (\mathbf{E_Q}[S|\mathcal{F}_t])_{0 \le t \le \tau}$ ein $\mathbf{Q}$-Martingal ist, muss nach Theorem 19.26 $(K_t)_{0 \le t \le \tau}$ verschwinden. Es gilt also, da $\frac{\partial \widetilde{F}(\widetilde{X}_t, t)}{\partial x} = \frac{\partial F(X_t, t)}{\partial x}$,

$$\begin{aligned}V_t(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = \widetilde{F}(\widetilde{X}_t^1, t) &= F(X_0^1, 0) + \int_0^t \frac{\partial F(X_s^1, s)}{\partial x} d\widetilde{X}_s^1 \\ &= V_0(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) + \int_0^t \Delta_s^0 d\widetilde{X}_s^0 + \int_0^t \Delta_s^1 d\widetilde{X}_s^1,\end{aligned}$$

womit $\underline{\Delta}$ selbstfinanzierend für $\underline{\widetilde{\mathcal{X}}}$ ist, was zu zeigen war. □

22.3 Preis und Hedging euopäischer Optionen: Die Black-Scholes Formel

In diesem Abschnitt berechnen wir den fairen Preis von europäischen Call- und Put-Optionen im Black-Scholes Modell. Dies führt auf die bekannte Black-Scholes-Formel zur Optionspreisbewertung.

Definition 22.13 (Europäische Call- und Put-Option). *1. Eine europäische Call-Option (auf eine Aktie mit Preisprozess $\mathcal{X}^1$) mit Maturität τ und Ausübungspreis K ist ein Derivat $S^c := (X^1_\tau - K)^+$.*

2. Eine europäische Put-Option (auf eine Aktie mit Preisprozess $\mathcal{X}^1$) mit Maturität τ und Ausübungspreis K ist ein Derivat $S^p := (K - X^1_\tau)^+$.

Bemerkung 22.14 (Interpretation). 1. Die europäische Option verleiht dem Besitzer wie in Abschnitt 13 das Recht (aber nicht die Pflicht) die Aktie (mit Preisprozess $\mathcal{X}^1$) zur Zeit τ zum Preis von K zu *kaufen*. Der Wert zur Zeit τ ist damit gerade der mögliche Gewinn bei Ausübung der Option. Verkauft der Besitzer die für K gekaufte Aktie sofort wieder zum Preis $X^1_\tau > K$, so erzielt er einen Gewinn von $X^1_\tau - K$. Ist $K > X^1_\tau$, so ist es günstiger, die Option verfallen zu lassen. Insgesamt ergibt sich der Wert von $S^c = (X^1_\tau - K)^+$ zur Zeit τ.

2. Die europäische Put-Option verleiht dem Besitzer das Recht, die Aktie zur Zeit τ zum Preis von K zu *verkaufen*. Dies lohnt sich natürlich nur dann, wenn $X^1_\tau < K$. In diesem Fall ist dann – ähnlich wie bei der call-Option – der Gewinn $K - X^1_\tau$. (Man beachte, dass S^p damit sogar beschränkt ist.)

Da sich europäische Call- und Put-Optionen in gewissem Sinne ergänzen, gilt für ihre Preise folgender Zusammenhang.

Lemma 22.15 (Put/Call-Parität für europäische Optionen). *Seien S^c_t und S^p_t die Preise von europäischen Call- und Put-Optionen auf eine Aktie mit Preisprozess $\mathcal{X}^1$ zur selben Maturität τ und Ausübungspreis K im Black-Scholes Modell. Dann gilt*

$$S^c_t - S^p_t = X^1_t - Ke^{-r(\tau-t)}.$$

Beweis. Es gilt $(X^1_t - K)^+ - (K - X^1_t)^+ = (X^1_t - K)^+ + (X^1_t - K)^- = X^1_t - K$. Also folgt mit Theorem 22.11 mit dem äquivalenten Maß $\mathbf{Q}$

$$\begin{aligned} S^c_t - S^p_t &= \mathbf{E}_{\mathbf{Q}}[e^{-r(\tau-t)}(X^1_\tau - K)|\mathcal{F}_t] = e^{rt}\mathbf{E}_{\mathbf{Q}}[\widetilde{X}^1_\tau|\mathcal{F}_t] - Ke^{-r(\tau-t)} \\ &= e^{rt}\widetilde{X}^1_t - K = X^1_t - Ke^{-r(T-t)}. \end{aligned}$$

□

Wir verwenden nun Theorem 22.11 für die Berechnung des Preises von europäischen Call- und Put-Optionen im Black-Scholes Modell. Dies führt auf die berühmte Black-Scholes-Formel zur Optionspreisberechnung.

Theorem 22.16 (Black-Scholes-Formel). *1. Der faire Preis zur Zeit t einer europäischen Call-Option mit Maturität τ und Ausübungspreis K im Black-Scholes Modell ist gegeben durch*

$$S^c_t = F(X^1_t, \tau - t)$$

mit

$$F(x,t) = x\Phi(d_1(x,t)) - Ke^{-rt}\Phi(d_2(x,t)),$$

wobei Φ *die Verteilungsfunktion der Standard-Normalverteilung ist sowie*

$$d_1(x,t) = \frac{\log(x/K) + (r + \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}}, \qquad d_2(x,t) = d_1(x,t) - \sigma\sqrt{t}.$$

Das Portfolio $\underline{\Delta} = (\Delta_t^0, \Delta_t^1)_{0\le t\le\tau}$ *mit*

$$\Delta_t^1 = \Phi(d_1(X_t^1, t)), \qquad \Delta_t^0 = e^{-rt}F(X_t^1, t) - \Delta_t^1 \widetilde{X}_t^1$$

ist ein Hedging-Portfolio für S^c.

2. *Weiter ist der faire Preis zur Zeit* t *einer europäischen Put-Option mit Maturität* τ *und Ausübungspreis* K *gegeben durch*

$$S_t^p = Ke^{-r(\tau-t)}\Phi(-d_2(X_t^1, \tau - t)) - X_t^1\Phi(-d_1(X_t^1, \tau - t)).$$

Das Portfolio $\underline{\Delta} = (\Delta_t^0, \Delta_t^1)_{0\le t\le\tau}$ *mit*

$$\Delta_t^1 = -\Phi(-d_1(X_t^1, t)), \qquad \Delta_t^0 = e^{-rt}F(X_t^1, t) - \Delta_t^1 \widetilde{X}_t^1$$

ist ein Hedging-Portfolio für S^p.

Beweis. Da $\Phi(-x) + \Phi(x) = 1$ für alle $x \in \mathbb{R}$, genügt es wegen Lemma 22.15, den Preis der europäischen Call-Option zu berechnen. Auch der Beweis des Hedging-Portfolios ist analog.

Um den Preis einer Call-Option zu berechnen, müssen wir $S_t^c = \mathbf{E}_{\mathbf{Q}}[e^{-r(\tau-t)}(X_\tau^1 - K)^+ | \mathcal{F}_t]$ für das äquivalente Maß $\mathbf{Q}$ mit Dichte Z_τ aus Proposition 22.10 berechnen. Wegen (22.9) ist

$$\begin{aligned} S_t^c &= \mathbf{E}_{\mathbf{Q}}\Big[e^{-r(\tau-t)}\Big(X_t^1 \frac{\widetilde{X}_\tau^1}{\widetilde{X}_t^1} e^{r(\tau-t)} - K\Big)^+ \Big| \mathcal{F}_t\Big] \\ &= \mathbf{E}_{\mathbf{Q}}[e^{-r(\tau-t)}(X_t^1 e^{\sigma(\widetilde{W}_\tau^1 - \widetilde{W}_t^1) + \left(r - \frac{1}{2}\sigma^2\right)(\tau-t)} - K)^+ | \mathcal{F}_t] \\ &= \mathbf{E}_{\mathbf{Q}}\big[\big(X_t^1 e^{\sigma Z\sqrt{\tau-t} - \frac{1}{2}\sigma^2(\tau-t)} - Ke^{-r(\tau-t)}\big)^+ | \mathcal{F}_t\big]. \end{aligned}$$

für eine Zufallsvariable $Z \stackrel{d}{=} N(0,1)$, die unabhängig von $\mathcal{F}_t$ ist. Nun gilt

$$xe^{\sigma\sqrt{t}z - \frac{1}{2}\sigma^2 t} > Ke^{-rt} \iff z > \frac{1}{\sigma\sqrt{t}}\Big(\tfrac{1}{2}\sigma^2 t + \log\frac{Ke^{-rt}}{x}\Big) = \frac{-\log\frac{x}{K} - (r - \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}} = -d_2(x,t).$$

Deshalb ist $S_t^c = F(X_t^1, t)$ mit

$$\begin{aligned} F(x,t) &= \frac{1}{\sqrt{2\pi}} \int_{-d_2(x,\tau-t)}^\infty (xe^{\sigma z\sqrt{\tau-t} - \frac{1}{2}\sigma^2(\tau-t)} - Ke^{-r(\tau-t)})e^{-\frac{z^2}{2}}\,dz \\ &\stackrel{y=z-\sigma\sqrt{\tau-t}}{=} \frac{1}{\sqrt{2\pi}}\Big(x\int_{-d_1(x,\tau-t)}^\infty e^{\sigma y\sqrt{\tau-t} + \frac{1}{2}\sigma^2(\tau-t)} e^{-(\frac{y^2}{2} + \sigma y\sqrt{\tau-t} + \frac{1}{2}\sigma^2(\tau-t))} dy \\ &\qquad\qquad - Ke^{-r(\tau-t)}\Big)\int_{-d_2(x,\tau-t)}^\infty e^{-\frac{z^2}{2}}\,dz\Big) \\ &= x\Phi(d_1(x,\tau-t)) - Ke^{-r(\tau-t)}\Phi(d_2(x,\tau-t)) \end{aligned}$$

und die Optionspreisformel ist gezeigt. Für das Hedging-Portfolio verwenden wir Proposition 22.12 und berechnen direkt

$$\begin{aligned}
\frac{\partial F(x,t)}{\partial x} &= \Phi(d_1(x,t)) + \frac{1}{\sqrt{2\pi}}\Big(x\exp\big(-\tfrac{1}{2}d_1^2(x,t)\big) - Ke^{-rt}\exp\big(-\tfrac{1}{2}(d_1(x,t)-\sigma\sqrt{t})^2\big)\frac{1}{x\sigma\sqrt{t}}\\
&= \Phi(d_1(x,t)) + \frac{1}{x\sigma\sqrt{2\pi t}}\exp\big(-\tfrac{1}{2}d_1^2(x,t)\big)\Big(x - Ke^{-rt}\exp\big(d_1(x,t)\sigma\sqrt{t}-\tfrac{1}{2}\sigma^2 t\big)\frac{1}{x\sigma\sqrt{t}}\\
&= \Phi(d_1(x,t)) + \frac{1}{\sqrt{2\pi}}\exp\big(-d_1^2(x,t)\big)\Big(x - Ke^{-rt}\exp\big(\log(x/K)+rt\big)\Big)\frac{1}{x\sigma\sqrt{t}}\\
&= \Phi(d_1(x,t)).
\end{aligned}$$

□

Bemerkung 22.17 (Volatilität). Es fällt auf, dass der Preis der europäischen Call- und Put-Optionen nicht mehr von μ, sondern nur von r und σ, was auch als Volatilität bezeichnet wird, abhängt. Normalerweise ist r als Zinssatz bekannt. Um die Volatilität eines bestimmten Wertpapieres zu bestimmen, gibt es zwei verschiedene Möglichkeiten:

1. Die *historische Methode* geht von den Beobachtungen $\log(X_t^1/X_0^1)$, $\log(X_{2t}^1/X_t^1)$, $\log(X_{3t}^1/X_{2t}^1), \ldots$ aus. Diese Zufallsgrößen sind im Black-Scholes-Modell unabhängig und nach $N((\mu-\frac{1}{2}\sigma^2)t, \sigma^2 t)$ verteilt. Aus der empirischen Varianz dieser Größen lässt sich also die Volatilität σ schätzen.
2. Die *implizite Methode* beobachtet am Markt gehandelte (europäische) Optionen sowie deren Preise und verwendet die Preise um zu berechnen, welche Volatilität angesetzt wurde, so dass mittels der Black-Scholes-Formel der gehandelte Preis entsteht.

Bemerkung 22.18 (The Greeks). In der Praxis interessiert man sich für die Abhängigkeit des Preises einer Option von den Modell-Parametern. Die Größe

$$\frac{\partial F(x,t)}{\partial x}$$

gibt die Abhängigkeit des Preises der Option vom momentanen Preis der Aktie an und heißt das *Delta* der Option. (Wir haben eben im Beweis berechnet, dass es $\Phi(d_1(x,t))$ für eine europäische Call-Option beträgt.) Um weiter die Sensitivität bezüglich des momentanen Preises zu bestimmen, berechnet man $\frac{\partial^2 F(x,t)}{\partial x^2}$, was auch als *gamma* bezeichnet. Schließlich ist *theta* gegeben als $\frac{\partial F(x,t)}{\partial t}$.

22.4 Arbitrage-Freiheit und Vollständigkeit

Ziel dieses Abschnittes ist es, die zwei fundamentalen Sätze der Optionspreisbewertung herzuleiten. Der erste besagt, dass Arbitrage-Freiheit äquivalent zur Existenz eines sogenannten äquivalenten Martingalmaßes ist. (Dieses ist ein Maß, bezüglich dem alle Wertpapiere Martingal sind.) Das zweite besagt, dass genau dann alle Derivate gehedgt werden können, wenn das äquivalente Martingalmaß eindeutig ist. Wir werden diese beiden Theoreme für den Finanzmarkt aus Definition 22.1 zeigen, d.h. die Aktienpreise folgen den Preisprozessen $\underline{\mathcal{X}} = (\mathcal{X}^0, \ldots, \mathcal{X}^d)$, $\mathcal{X}^i = (X_t^i)_{0\le t\le\tau}$ der Form

$$\begin{aligned}
\mathcal{X}^0 &= 1 + \mathcal{R}\mathcal{X}^0\cdot\lambda,\\
\mathcal{X}^i &= X_0^i + \mu^i\mathcal{X}^i\cdot\lambda + \mathcal{X}^i\sum_{k=1}^n \sigma^{ik}\cdot\mathcal{W}^k, \qquad i=1,\ldots,d
\end{aligned} \tag{22.11}$$

mit einer n-dimensionalen Brown'schen Bewegung $\underline{\mathcal{W}}$ und davon erzeugter Filtration $(\mathcal{F}_t)_{0\le t\le\tau}$, so dass $\mathcal{R} = (R_t)_{0\le t\le\tau}, \mu^i = (\mu^i_t)_{0\le t\le\tau}, \sigma^{ik} = (\sigma^{ik}_t)_{0\le t\le\tau}, i = 1,...,d, k = 1,...,n$ adaptiert sind. Weiter erinnern wir an die normalisierten Prozesse $\underline{\widetilde{\mathcal{X}}} = (\widetilde{\mathcal{X}}^0,...,\widetilde{\mathcal{X}}^d)$ mit $\widetilde{\mathcal{X}}^i = \mathcal{X}^i/\mathcal{X}^0$.

Wir sehen aus Lemma 22.5, dass die normalisierten Aktienpreisprozesse den Gleichungen

$$\begin{aligned} &\widetilde{\mathcal{X}}^0 = 1,\\ &\widetilde{\mathcal{X}}^i = X^i_0 + (\mu^i - \mathcal{R})\widetilde{\mathcal{X}}^i \cdot \lambda + \widetilde{\mathcal{X}}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k, \qquad i = 1,...,d \end{aligned} \tag{22.12}$$

genügen.

Definition 22.19 (Äquivalentes Martingalmaß). *Ein (zu* $\mathbf{P}$*) äquivalentes Martingalmaß ist ein Wahrscheinlichkeitsmaß* $\mathbf{Q}$*, das (i) äquivalent zu* $\mathbf{P}$ *ist (d.h. dieselben Nullmengen hat) und (ii) bezüglich dem alle normalisierten Prozesse* $\widetilde{\mathcal{X}^i}$ *Martingale sind.*

Proposition 22.20. *Gibt es einen Prozess* $\underline{\mathcal{H}} = (\mathcal{H}^1,...,\mathcal{H}^n)$*, so dass*

$$\sum_{k=1}^n \sigma^{ik}\mathcal{H}^k = \mu^i - \mathcal{R}, \tag{22.13}$$

und

$$\mathcal{Z} = \exp\Big(-\sum_{k=1}^n \mathcal{H}^k \cdot \mathcal{W}^k - \tfrac{1}{2}(\mathcal{H}^k)^2 \cdot \lambda\Big)$$

ein Martingal ist, dann ist

$$\mathbf{Q} = Z_\tau \cdot \mathbf{P}$$

ein äquivalentes Martingalmaß mit

$$\widetilde{\mathcal{X}}^i = X^i_0 + \widetilde{\mathcal{X}}^i \sum_{k=1}^n \sigma^{ik} \cdot \widetilde{\mathcal{W}}^k,$$

wobei $\widetilde{\mathcal{W}}^k = \mathcal{W}^k + \mathcal{H}^k \cdot \lambda, k = 1,...,n$ *Brown'sche Bewegungen bezüglich* $\mathbf{Q}$ *sind.*

Beweis. Alle Aussagen folgen analog zu Theorem 21.11 mit $\gamma = -\mu + \mathcal{R}$. □

Die Gleichungen (22.13) heißen auch Risikogleichungen des Marktes. Wir stellen nun eine Verbindung von Arbitrage und der Existenz eines äquivalenten Martingalmaßes her. Zur Definition von Arbitrage siehe Definition 22.6.

Theorem 22.21 (Erstes fundamentales Theorem der Anlagebewertung). *Existiert für einen Finanzmarkt der Form* (22.11) *ein äquivalentes Martingalmaß, so ist dieser arbitrage-frei.*

Beweis. Der Beweis ist fast analog zum Beweis der Arbitrage-Freiheit des Black-Scholes-Marktes aus Theorem 22.11:

Sei $\underline{\Delta}$ ein Portfolio mit $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$ und $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \ge 0$ fast sicher. Da $\underline{\Delta}$ nach Lemma 22.5 auch selbstfinanzierend bezüglich $\underline{\widetilde{\mathcal{X}}}$ ist, gilt

$$\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = V_0(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) + \sum_{i=0}^d \Delta^i \cdot \widetilde{\mathcal{X}}^i.$$

Da nach Voraussetzung die Prozesse $\widetilde{\mathcal{X}}^i, i = 0, ..., d$ Martingale bezüglich $\mathbf{Q}$ sind, ist $\mathcal{V}(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}})$ ebenfalls ein $\mathbf{Q}$-Martingal. Nach Voraussetzung ist $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ $\mathbf{P}$-fast sicher. Da $\mathbf{P}$ und $\mathbf{Q}$ äquivalent sind, ist also auch $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ $\mathbf{Q}$-fast sicher. Wegen der Martingaleigenschaft von $\mathcal{V}(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}})$ ist außerdem $\mathbf{E}_\mathbf{Q}[V_\tau(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}})] = 0$. Dies ist jedoch nur möglich falls $V_\tau(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}}) = 0$ $\mathbf{Q}$-fast sicher. Wieder wegen der Äquivalenz von $\mathbf{P}$ und $\mathbf{Q}$ folgt $V_\tau(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}}) = 0$ $\mathbf{P}$-fast sicher, und die Arbitragefreiheit ist gezeigt. □

Beispiel 22.22 (Ein nicht-arbitrage-freier Markt). Wir geben ein einfaches Beispiel, in dem kein äquivalentes Martingalmaß existiert, aber Arbitrage-Möglichkeiten existieren. Hierzu sei $i = 2$ und $k = 1$, d.h. zwei Aktien werden von derselben Brown'schen Bewegung angetrieben. Wir nehmen an, dass $\mathcal{R} =: r, \mu^i, \sigma^{i1} =: \sigma^i$ konstant sind und $X^1(0) = X^2(0) = 1$. Dann sind die Risikogleichungen des Marktes gegeben durch

$$\sigma^1 \mathcal{H} = \mu^1 - r,$$
$$\sigma^2 \mathcal{H} = \mu^2 - r.$$

Dies hat genau dann eine eindeutige Lösung, wenn

$$\frac{\mu^1 - r}{\sigma^1} = \frac{\mu^2 - r}{\sigma^2}.$$

Nehmen wir an, dass $a := \frac{\mu^1 - r}{\sigma^1} - \frac{\mu^2 - r}{\sigma^2} > 0$. Wir setzen

$$\Delta^1 = \frac{1}{\sigma^1 \mathcal{X}^1}, \qquad \Delta^2 = -\frac{1}{\sigma^2 \mathcal{X}^2}.$$

Weiter sei Δ^0 so gewählt, dass das Portfolio selbstfinanzierend ist mit $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$. Dann muss gelten

$$\begin{aligned}
\Delta^0 &= \frac{1}{\sigma^2} - \frac{1}{\sigma^1}, \\
\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) &= \frac{1}{\sigma^1 \mathcal{X}^1} \cdot \mathcal{X}^1 - \frac{1}{\sigma^2 \mathcal{X}^2} \cdot \mathcal{X}^2 + r\Big(\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) - \frac{1}{\sigma^1} + \frac{1}{\sigma^2}\Big) \cdot \lambda \\
&= \frac{\mu^1 - r}{\sigma^1} \cdot \lambda - \frac{\mu^2 - r}{\sigma^2} \cdot \lambda + r\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) \cdot \lambda + \mathcal{W} - \mathcal{W}
\end{aligned}$$

sowie (mit (22.5))

$$\begin{aligned}
\mathcal{V}(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}}) &= V_0 + \frac{1}{\sigma^1 \mathcal{X}^1} \cdot \widetilde{\mathcal{X}}^1 - \frac{1}{\sigma^2 \mathcal{X}^2} \cdot \widetilde{\mathcal{X}}^2 \\
&= \Big(\frac{\mu^1 - r}{\sigma^1} - \frac{\mu^2 - r}{\sigma^2}\Big) \frac{1}{\mathcal{X}^0} \cdot \lambda = \frac{a}{\mathcal{X}^0} \cdot \lambda > 0
\end{aligned}$$

für $t > 0$. Das bedeutet, dass $\underline{\Delta}$ ein selbstfinanzierendes Portfolio ist mit $V_0(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}}) = 0$ und $V_\tau(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}}) > 0$ fast sicher, also ist der Markt nicht arbitrage-frei.

Um Derivate zu Bepreisen, ist es wichtig, Hedging-Portfolios zu kennen. Wir kommen nun zu Bedingungen, wann solche genau existieren.

Definition 22.23 (Vollständiger Markt). *Ein Markt* (22.11) *heißt* vollständig*, wenn jedes Derivat gehedgt werden kann, d.h. für jede quadratisch integrierbare Zufallsvariable S gibt es eine selbstfinanzierende Strategie mit festem Startwert $V_0(\underline{\Delta}, \underline{\mathcal{X}})$ und $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = S$.*

Theorem 22.24 (Zweites fundamentales Theorem der Anlagebewertung). *Hat ein Finanzmarkt der Form (22.5) ein äquivalentes Martingalmaß, so ist er genau dann vollständig, wenn das Martingalmaß eindeutig ist.*

Bevor wir mit dem beweis beginnen, bringen wir noch ein elementares Lemma aus der linearen Algebra.

Lemma 22.25. *Sei $\underline{\underline{A}}$ eine $d \times n$-Matrix, $\underline{b}$ ein d-dimensionaler Vektor, so dass*

$$\underline{\underline{A}}\,\underline{x}^\top = \underline{b}^\top$$

eine eindeutige Lösung $\underline{x} \in \mathbb{R}^n$ hat. Dann hat

$$\underline{y}\,\underline{\underline{A}} = \underline{c}$$

für jeden n-dimensionalen Vektor $\underline{c}$ mindestens eine Lösung $\underline{y} \in \mathbb{R}^d$.

Beweis. Wir interpretieren $\underline{\underline{A}}$ als Abbildung $\underline{\underline{A}} : \mathbb{R}^n \to \mathbb{R}^d$. Da die Lösung von $\underline{\underline{A}}\,\underline{x}^\top = \underline{b}^\top$ eindeutig ist, muss $\ker(A) = \underline{\underline{0}}$ gelten. (Denn mit $\underline{x}$ ist auch $\underline{x}+\underline{x}_0$ für $\underline{x}_0 \in \ker(A)$ eine Lösung.) Damit ist $n = \dim(\ker(A)) + \dim(\mathrm{im}(A)) = \dim(\mathrm{im}(A)) \leq d$, also $\mathrm{rg}(A) = \dim(\mathrm{im}(A)) = n$. Wir schreiben die Gleichung $\underline{y}\,\underline{\underline{A}} = \underline{c}$ um zu $\underline{\underline{A}}^\top\,\underline{y}^\top = \underline{c}^\top$. Da $\mathrm{rg}(A^\top) = \mathrm{rg}(A) = n$, liegt jeder $\underline{c} \in \mathbb{R}^n$ in $\mathrm{im}(A^\top)$, insbesondere hat $\underline{\underline{A}}^\top\,\underline{y}^\top = \underline{c}^\top$ mindestens eine Lösung. □

Beweis von Theorem 22.24. Sei der Markt zunächst vollständig und $\mathbf{Q}$ und $\mathbf{R}$ seien zwei äquivalente Martingalmaße. Ziel ist es, $\mathbf{Q} = \mathbf{R}$ zu zeigen. Sei hierfür $A \in \mathcal{F}_\tau = \mathcal{F}$ und $S = 1_A X^0_\tau$. Da der Markt vollständig ist, kann S gehedgt werden, es gibt also selbstfinanzierendes Portfolio $\underline{\Delta}$ mit $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = S$, oder auch $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 1_A$. Da $\widetilde{\mathcal{X}}^i$ sowohl ein $\mathbf{Q}$- als auch ein $\mathbf{R}$-Martingal ist, $i = 0, ..., d$, folgt

$$\mathbf{E}_\mathbf{Q}[1_A] = \mathbf{E}_\mathbf{Q}[V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})] = \mathbf{E}_\mathbf{Q}\Big[\sum_{i=0}^d \Delta_0^i X_0^i + (\Delta^i \cdot \widetilde{\mathcal{X}}^i)_\tau\Big] = \sum_{i=0}^d \Delta_0^i X_0^i,$$

und analog

$$\mathbf{E}_\mathbf{Q}[1_A] = \sum_{i=0}^d \Delta_0^i X_0^i.$$

Insbesondere gilt $\mathbf{Q}(A) = \mathbf{R}(A)$, und da A beliebig war, folgt $\mathbf{Q} = \mathbf{R}$.

Andersherum sei $\mathbf{Q}$ das einzige äquivalente Martingalaß. Dann besitzen die Risikogleichungen des Marktes, (22.13), genau eine Lösung $\mathcal{H}^1, ..., \mathcal{H}^n$. (Denn: Zunächst ist $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ für eine Zufallsvariable $Z_\tau > 0$ nach dem Satz von Radon-Nikodym. Es ist $\mathcal{Z} = (Z_t)_{0 \leq t \leq \tau}$ mit $Z_t := \mathbf{E}_\mathbf{P}[Z_\tau | \mathcal{F}_t]$ ein positives $\mathbf{P}$-Martingal, also folgt nach Theorem 20.14, dass $\mathcal{Z} = \exp(\mathcal{N} - \frac{1}{2}[\mathcal{N}])$ für ein (eindeutiges) lokales $\mathbf{P}$-Martingal $\mathcal{N}$, das man nach dem Martingal-Repräsentations-Theorem 20.9 als Integral von Prozessen $\mathcal{H}^1, ..., \mathcal{H}^n$ bezüglich der Brown'schen Bewegungen $\mathcal{W}^1, ...\mathcal{W}^n$ schreiben kann. Damit nun $\widetilde{\mathcal{X}}^i$ $\mathbf{Q}$-Martingale sind, müssen notwendigerweise die Gleichungen (22.13) erfüllt sein. Da diese Argumentation für jedes äquivalente Martingalmaß angewendet werden kann, dieses jedoch eindeutig ist nach Voraussetzung, hat (22.13) genau eine Lösung.) Wir setzen $\underline{\underline{\sigma}} := (\sigma^{ik})_{i=1,...,d;k=1,...,n}$, $\underline{\mu} := (\mu^i)_{i=1,...,d}$, $\underline{\mathcal{H}} := (\mathcal{H}^k)_{k=1,...,n}$. Dann können wir diese Gleichungen als lineares Gleichungssysystem

$$\underline{\underline{\sigma}}\,\underline{\mathcal{H}}^\top = \underline{\mu}^\top - \mathcal{R}^\top \tag{22.14}$$

(mit Lösung $\underline{\mathcal{H}}$) schreiben. Diese Gleichung hat (für fast alle ω und alle t) also genau eine Lösung.

Wir müssen zeigen, dass jedes Derivat, gegeben durch eine quadrat-integrierbare Zufallsvariable S, gehedgt werden kann. Zunächst bemerken wir, dass

$$X_t^0 = \exp\Big(\int_0^t R_s ds\Big)$$

gelten muss. Wir definieren den Prozess $\mathcal{V} = (V_t)_{0\le t\le\tau}$ mit

$$V_t := \mathbf{E_Q}[e^{-\int_t^\tau R_s ds} S|\mathcal{F}_t],$$

so dass $\mathcal{V}/\mathcal{X}^0$ wegen

$$V_t/X_t^0 = \mathbf{E_Q}[e^{-\int_0^\tau R_s ds} S|\mathcal{F}_t]$$

für alle $0 \le t \le \tau$ ein $\mathbf{Q}$-Martingal ist. Mit $\widetilde{\mathcal{W}}^k = \mathcal{W}^k - \mathcal{H}^k \cdot \lambda$ und dem Martingal-Repräsentationstheorem 20.9 gibt es Prozesse $\mathcal{L}^1, ..., \mathcal{L}^n$ mit

$$\mathcal{V}/\mathcal{X}^0 = V_0 + \sum_{k=1}^n \mathcal{L}^k \cdot \widetilde{\mathcal{W}}^k.$$

Um S zu hedgen, genügt es ein selbstfinanzierendes Portfolio $\underline{\Delta}$ zu finden mit $V_t(\underline{\Delta}, \underline{\mathcal{X}}) = V_t/X_t^0$. Dies impliziert wegen $\widetilde{\mathcal{X}}^i = X_0^i + \widetilde{\mathcal{X}}^i \sum_k \sigma^{ik} \cdot \widetilde{\mathcal{W}}^k$

$$V_0 + \sum_{k=1}^n \int_0^t L_s^k d\widetilde{W}_s^k = V_t/X_t^0 = V_0 + \sum_{k=1}^n \int_0^t \Delta_s^i \widetilde{X}_s^i \sigma_s^{ik} d\widetilde{W}_s^k.$$

Um S also zu hedgen, müssen wir $\underline{\Delta}$ finden, so dass

$$\mathcal{L}^k = \Delta^i \widetilde{\mathcal{X}}^i \sigma^{ik},$$

oder äquivalent

$$\underline{\mathcal{L}} = \underline{\Delta\widetilde{\mathcal{X}}}\,\underline{\underline{\sigma}}$$

mit $\underline{\Delta\widetilde{\mathcal{X}}} = (\Delta^i \widetilde{\mathcal{X}}^i)_{i=1,...d}$ und $\underline{\mathcal{L}} = (\mathcal{L}^k)_{k=1,...,n}$. Da (22.14) nach Voraussetzung genau eine Lösung hat, hat dieses Gleichungssystem nach Lemma 22.25 mindestens eine Lösung $\underline{\Delta}$. Wie soeben gezeigt, liefert dieses $\underline{\Delta}$ ein Hedging-Portfolio für S. $\square$

23 Zinsstrukturmodelle

Im Black-Scholes Modell ist der Wert des risikolosen Wertpapieres, $\mathcal{X}^0$, deterministisch. Diese Bedingung werden wir nun aufgeben, dafür aber keine weiteren Wertpapiere betrachten. Der Markt ist also bestimmt durch die Gleichung

$$\mathcal{X}^0 = 1 + \mathcal{R}\mathcal{X}^0 \cdot \lambda, \tag{23.1}$$

wobei wir stets annehmen, dass die Filtration, an die $\mathcal{R} = (R_t)_{0\leq t\leq\tau}$ und damit $\mathcal{X}^0 = (X_t^0)_{0\leq t\leq\tau}$ adaptiert sind, von einer Brown'schen Bewegung $\mathcal{W}$ erzeugt ist. Klar ist, dass

$$X_t^0 = \exp\Big(\int_0^t R_s ds\Big).$$

Ziel dieses Abschnittes ist es, den fairen Preis von Zero-Bonds zu berechnen. Dies sind Derivate, die zu einem Zeitpunkt $0 \leq t \leq \tau$ ausgegeben werden und deren Besitzer zu einem Zeitpunkt $t \leq u \leq \tau$ eine Geldeinheit bekommen. Die Frage ist also, was es zur Zeit t wert ist, zur Zeit u sicher eine Geldeinheit zu haben, wenn der Zinsprozess durch $\mathcal{R}$ gegeben ist. Den Preis des Zero-Bonds mit Maturität u zur Zeit t bezeichnen wir mit S_t^u, was zu einem Prozess $(S_t^u)_{0\leq t\leq u}$ mit $S_u^u = 1$ führt.

23.1 Grundlegendes

Wieder werden wir mit Hilfe von äquivalenten Martingalmaßen den fairen Preis von Zero-Bonds berechnen; siehe Proposition 23.1. Anschließend können wir die Preisprozesse $(S_t^u)_{0\leq t\leq u}$ mittels einer stochastischen Differentialgleichung beschreiben; siehe Proposition 23.4.

Proposition 23.1 (Preis eines Zero-Bonds). *Gibt es im Markt* (23.1) *ein äquivalentes Martingalmaß* $\mathbf{Q}$, *(d.h. ein Maß, bezüglich dem* $(S_t^u/X_t^0)_{0\leq t\leq u}$ *für alle* $0 \leq u \leq \tau$ *ein Martingal ist) dann ist*

$$S_t^u = \mathbf{E_Q}\Big[\exp\Big(-\int_t^u R_s ds\Big)\Big|\mathcal{F}_t\Big]$$

der faire Preis des Zero-Bonds mit Maturität u *zur Zeit* t.

Beweis. Klar ist, dass $S_u^u/X_u^0 = \exp\Big(-\int_0^u R_s ds\Big)$. Da $(S_t^u/X_t^0)_{0\leq t\leq u}$ ein $\mathbf{Q}$-Martingal ist, folgt schon

$$S_t^u/X_t^0 = \mathbf{E_Q}[S_u^u/X_u^0|\mathcal{F}_t] = \mathbf{E_Q}\Big[\exp\Big(-\int_0^u R_s ds\Big)\Big|\mathcal{F}_t\Big],$$

also folgt die Behauptung nach Multiplikation mit X_t^0. □

Da die Filtration $(\mathcal{F}_t)_{0\leq t\leq\tau}$ von der Brown'schen Bewegung erzeugt wird, können wir äquivalente Martingalmaße im Markt (23.1) gut beschreiben.

Proposition 23.2 (Charakterisierung äquivalenter Maße). *Sei* $\mathbf{Q}$ *äquivalent zu* $\mathbf{P}$. *Dann gibt es eine Zufallsvariable* $Z_\tau > 0$ *und einen Prozess* $\mathcal{H} = (H_t)_{0\leq t\leq\tau}$, *so dass* $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ *und*

$$Z_\tau = \exp\Big(-\int_0^\tau H_s dW_s - \tfrac{1}{2}\int_0^\tau H_s^2 ds\Big). \tag{23.2}$$

Beweis. Nach dem Satz von Radon-Nikodym gibt es eine Dichte Z_τ von $\mathbf{Q}$ bezüglich $\mathbf{P}$, so dass $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$. Klar ist, dass $\mathcal{Z} = (Z_t)_{0\le t\le\tau}$ mit $Z_t = \mathbf{E_P}[Z_\tau|\mathcal{F}_t]$ ein $\mathbf{P}$-Martingal ist. Nach Theorem 20.14.2 gibt es ein lokales $\mathbf{P}$-Martingal $\mathcal{N}$, so dass $\mathcal{Z} = \exp(\mathcal{N} - \frac{1}{2}[\mathcal{N}])$. Da die Filtration $(\mathcal{F}_t)_{0\le t\le\tau}$ von $\mathcal{W}$ erzeugt wird, gibt es nach Theorem 20.9 einen Prozess $\mathcal{H}$ mit $\mathcal{N} = -\mathcal{H}\cdot\mathcal{W}$. Daraus folgt die Behauptung. □

Korollar 23.3 (Preis eines Zero-Bonds). *Gibt es im Markt* (23.1) *ein äquivalentes Martingalmaß* $\mathbf{Q}$ *mit* $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ *für*

$$Z_\tau = \exp\Big(-\int_0^\tau H_s dW_s - \tfrac{1}{2}\int_0^\tau H_s^2 ds\Big).$$

und einen Prozess $\mathcal{H} = (H_t)_{0\le t\le\tau}$*, dann ist*

$$S_t^u = \mathbf{E_P}\Big[\exp\Big(-\int_t^u R_s ds - \int_t^u H_s dW_s - \frac{1}{2}\int_t^u H_s^2 ds\Big)\Big|\mathcal{F}_t\Big]$$

der faire Preis des Zero-Bonds mit Maturität u *zur Zeit* t*.*

Beweis. Es genügt, die Aussage für $u = \tau$ zu zeigen. Sie folgt sofort aus den letzten beiden Propositionen, wenn man beachtet, dass für jede nicht-negative Zufallsvariable X für $0 \le t \le \tau$ und $A \in \mathcal{F}_t$

$$\mathbf{E_P}[Z_t\mathbf{E_Q}[X|\mathcal{F}_t], A] = \mathbf{E_Q}[\mathbf{E_Q}[X|\mathcal{F}_t], A] = \mathbf{E_Q}[X, A] = \mathbf{E_P}[Z_\tau X, A] = \mathbf{E_P}[\mathbf{E_P}[Z_\tau X|\mathcal{F}_t], A],$$

was soviel bedeutet, dass

$$\mathbf{E_Q}[X|\mathcal{F}_t] = \frac{\mathbf{E_P}[Z_\tau X|\mathcal{F}_t]}{Z_t}$$

gilt. □

Proposition 23.4 (Eine SDE für den Preisprozess von Zero-Bonds). *Gibt es im Markt* (23.1) *ein äquivalentes Martingalmaß* $\mathbf{Q}$ *mit* $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ *für*

$$Z_\tau = \exp\Big(-\int_0^\tau H_s dW_s - \tfrac{1}{2}\int_0^\tau H_s^2 ds\Big).$$

und einen Prozess $\mathcal{H} = (H_t)_{0\le t\le\tau}$*, so gibt es für jedes* $0 \le u \le \tau$ *einen adaptierten stochastischen Prozess* $\mathcal{K}^u = (K_t^u)_{0\le t\le u}$*, so dass*

$$dS_t^u = (R_t + K_t^u H_t)S_t^u dt + K_t^u S_t^u dW.$$

Bemerkung 23.5 (Interpretation). Wir bemerken sofort, dass zwar der Prozess $\mathcal{R}$ endliche Variation besitzt, jedoch die Prozesse $(S_t^u)_{0\le t\le u}$ nicht. Weiter stellen wir fest, dass sich der erwartete Zins, der sich mit einem Zero-Bond berechnen lässt, vom Prozess $\mathcal{R}$ unterscheidet. Ist $\mathcal{H}$ positiv (was der Normalfall ist), so ist der erwartete Zins größer als bei $\mathcal{X}^0$. Dieser größere Zins wird jedoch durch ein höheres Risiko verbunden mit einer positiven quadratischen Variation im Prozess $(S_t^u)_{0\le t\le u}$ erkauft.

Beweis von Proposition 23.4. Da der Prozess $(S_t^u/X_t^0)_{t\geq 0}$ ein positives **Q**-Martingal ist, ist nach Lemma 20.12 $(Z_t S_t^u/X_t^0)_{t\geq 0}$ ein positives **P**-Martingal. Also gibt es nach Theorem 20.14 und Theorem 20.9 einen adaptierten Prozess $\mathcal{L}^u = (L_t^u)_{0\leq t\leq u}$ mit

$$Z_t S_t^u/X_t^0 = S_0^u \exp\Big(\int_0^t L_s^u dW_s - \tfrac{1}{2}\int_0^t (L_s^u)^2 ds\Big),$$

oder auch

$$S_t^u = S_0^u \exp\Big(\int_0^t L_s^u + H_s dW_s + \int_0^t R_s - \tfrac{1}{2}((L_s^u)^2 - H_s^2) ds\Big).$$

Genau wie in Beispiel 21.1 sehen wir, dass $(S_t^u)_{0\leq t\leq u}$ die SDE

$$\begin{aligned} dS_t^u &= \big(R_t - \tfrac{1}{2}((L_t^u)^2 - (H_t)^2) + \tfrac{1}{2}(L_t^u + H_t)^2\big)S_t^u ds + (L_t^u + H_t)S_t^u dW \\ &= \big(R_t + L_t^u H_t + H_t^2\big)S_t^u ds + (L_t^u + H_t)S_t^u dW \end{aligned}$$

löst. Nun folgt die Aussage mit $K_t^u = L_t^u + H_t$. □

23.2 Das Vasicek-Modell

Wie wir in Abschnitt 23.1, insbesondere in Proposition 23.1 und Korollar 23.3, gesehen haben, können wir den Preis S_t^u eines Zero-Bonds berechnen, wenn wir den Preisprozess $\mathcal{X}^0$ unter dem äquivalenten Martingalmaß **Q**, oder sowohl $\mathcal{X}^0$ als auch $\mathcal{H}$ unter **P** kennen. Das (einfaktorielle) Vasicek-Modell ist ein einfaches Modell, in dem wir $\mathcal{R}$ durch einen Ornstein-Uhlenbeck-Prozess beschreiben. Diesen Prozess wiederholen wir nun kurz.

Beispiel 23.6 (Ornstein-Uhlenbeck Prozess). Wir betrachten die Lösung der SDE

$$dX = -\mu X dt + \sigma dW$$

und behaupten, dass

$$X_t = e^{-\mu t}\Big(X_0 + \sigma\int_0^t e^{\mu s} dW_s\Big) \tag{23.3}$$

starke Lösung dieser SDE ist. Nach der Formel für partielle Integration gilt nämlich für den so definierten Prozess und $\mathcal{A} = (A_t)_{t\geq 0}$ mit $A_t = e^{-\mu t}$ (und damit $\mathcal{A} = 1 - \mu\mathcal{A}\cdot\lambda$) und $\mathcal{B} = (B_t)_{t\geq 0}$ mit $B_t = X_0 + \sigma\int_0^t e^{\mu s} dW_s$, d.h. $\mathcal{B} = X_0 + \sigma\mathcal{A}^{-1}\cdot\mathcal{W}$

$$\mathcal{X} - X_0 = \mathcal{A}\cdot\mathcal{B} + \mathcal{B}\cdot\mathcal{A} = \sigma\cdot\mathcal{W} - \mu\mathcal{A}\mathcal{B}\cdot\lambda = -\mu\mathcal{X}\cdot\lambda + \sigma\cdot\mathcal{W}.$$

Weiter behaupten wir, dass

$$Y_t = b + e^{-\mu t}\Big(Y_0 - b + \sigma\int_0^t e^{\mu s} dW_s\Big) \tag{23.4}$$

starke Lösung der SDE

$$dY = \mu(b - Y)dt + \sigma dW \tag{23.5}$$

ist. (Denn: Wenn $\mathcal{X}$ die Gleichung (23.3) löst, dann löst $\mathcal{Y} := \mathcal{X} + b$ die Gleichung (23.4).)

Im Vasicek-Modell wird der Prozess $\mathcal{R}$ aus (23.1) durch die Lösung von (23.4) modelliert. Weiter wird davon ausgegangen, dass das äquivalente Martingalmaß aus Proposition (23.4) die Form $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ mit Z_τ aus (23.2) und $\mathcal{H} = \rho$ für ein $\rho \in \mathbb{R}$ besitzt. In diesem Fahrwasser geben wir einige elementare Resultate an.

Lemma 23.7 (Modellierung im Vasicek-Modell). *Sei $\mathcal{Y} = Y_{0\leq t\leq\tau}$ die Lösung von (23.5) und $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ mit $Z_\tau = \exp(-\rho W_\tau - \frac{1}{2}\rho^2 t)$. Dann ist $\widetilde{\mathcal{W}} = \mathcal{W} + \rho t$ eine Brown'sche Bewegung bezüglich $\mathbf{Q}$ und $\mathcal{Y}$ ist die einzige starke Lösung von*

$$dY = \mu(b^* - Y_t)dt + \sigma\widetilde{W}_t$$

mit $b^* = b - \rho\sigma/\mu$.

Beweis. Dies folgt unmittelbar aus der Girsanov-Transformation von Theorem 21.11. □

Theorem 23.8 (Preise von Zero-Bonds im Vasicek-Modell). *Wir verwenden die Notation aus Proposition 23.1. Sei $\mathcal{R} = R_{0\leq t\leq\tau}$ die Lösung von (23.5) und $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ mit $Z_\tau = \exp(-\rho W_\tau - \frac{1}{2}\rho^2 t)$. Dann ist*

$$S_t^u = F(R_t, u - t)$$

mit $b^* = b - \rho\sigma/\mu$ und

$$F(x,t) = \exp\Big(-\Big(b^* - \frac{\sigma^2}{2\mu^2}\Big)t + \frac{1}{\mu}\Big(b^* - \frac{\sigma^2}{2\mu^2} - x\Big)(1 - e^{-\mu t}) - \frac{\sigma^2}{4\mu^3}(1 - e^{-\mu t})^2\Big)$$

der faire Preis des Zero-Bonds mit Maturität u zur Zeit t.

Bemerkung 23.9 (Der langfristig erreichbare Zins). Für große t gilt approximativ

$$F(x,t) \approx C\exp\Big(-\Big(b^* - \frac{\sigma^2}{2\mu^2}\Big)t\Big)$$

für eine Konstante C. Dies bedeutet, dass $b^* - \sigma^2/(2\mu^2)$ der im Vasicek-Modell langfristig erreichbare Zins ist. In der Praxis geht man jedoch davon aus, dass der langfristig erreichbare Zins vom momentanen Zins, R_t, abhängt.

Beweis von Theorem 23.8. Wegen Proposition 23.1 ist lediglich $\mathbf{E}_\mathbf{Q}[\exp(-\int_t^u R_s ds)|\mathcal{F}_t]$ zu berechnen. Da $\mathcal{R}$ die SDE

$$dR = \mu(b^* - R_t)dt + \sigma d\widetilde{W} \tag{23.6}$$

mit der $\mathbf{Q}$-Brown'schen Bewegung $\widetilde{\mathcal{W}}$ und $b^* = b - \rho\sigma/\mu$ löst, und damit ein Markov-Prozess ist, schreiben wir

$$\mathbf{E}_\mathbf{Q}\Big[\exp\Big(-\int_t^u R_s ds\Big)\Big|\mathcal{F}_t\Big] = F(X_t, u-t)$$

mit

$$F(x,t) = \mathbf{E}\Big[\exp\Big(-\int_0^t R_s ds\Big)\Big|R_0 = x\Big], \tag{23.7}$$

wobei $\mathcal{R} = (R_t)_{0\le t\le \tau}$ die SDE (23.6) löst. Wir können dabei verwenden, dass

$$R_t = b^* + e^{-\mu t}\Big(x - b^* + \sigma \int_0^t e^{\mu s} d\widetilde{W}_s\Big).$$

Da R_s in (23.7) für jedes s normalverteilt ist, ist auch $\int_0^t R_s ds$ normalverteilt. Wir berechnen Erwartungswert und Varianz als

$$\begin{aligned}
\widehat{\mu} := \mathbf{E}\Big[\int_0^t R_s ds\Big] &= \int_0^t b^* + e^{-\mu s}(x-b^*)ds = b^* t + \frac{x-b^*}{\mu}(1-e^{-\mu t}),\\
\widehat{\sigma}^2 := \mathbf{V}\Big[\int_0^t R_s ds\Big] &= 2\int_0^t \int_0^s \mathbf{COV}[R_s, R_r] dr ds\\
&= 2\sigma^2 \int_0^t \int_0^s e^{-\mu r} e^{-\mu s} \mathbf{E}\Big[\int_0^s e^{\mu s} d\widetilde{W}_s \int_0^r e^{\mu r} d\widetilde{W}_r\Big] dr ds\\
&= 2\sigma^2 \int_0^t \int_0^s \int_0^r e^{-\mu r} e^{-\mu s} e^{2\mu v} dv ds dr\\
&= \frac{\sigma^2}{\mu} \int_0^t e^{-\mu s} \int_0^s e^{\mu r} - e^{-\mu r} dr ds\\
&= \frac{\sigma^2}{\mu^2} \int_0^t 1 - 2e^{-\mu s} + e^{-2\mu s} ds\\
&= \frac{\sigma^2 t}{\mu^2} - \frac{2\sigma^2}{\mu^3}(1-e^{-\mu t}) + \frac{\sigma^2}{2\mu^3}(1-e^{-2\mu t})\\
&= \frac{\sigma^2 t}{\mu^2} - \frac{\sigma^2}{\mu^3}(1-e^{-\mu t}) - \frac{\sigma^2}{2\mu^3}(1-e^{-\mu t})^2.
\end{aligned}$$

Da nach Beispiel 7.13.3 für eine nach $N(\widehat{\mu}, \widehat{\sigma}^2)$-vereilte Zufallsvariable Z gilt, dass $\mathbf{E}[e^{-Z}] = e^{-\widehat{\mu}+\widehat{\sigma}^2/2}$, folgt

$$\begin{aligned}
F(x,t) &= e^{-\widehat{\mu}+\widehat{\sigma}^2/2}\\
&= \exp\Big(-\Big(b^* - \frac{\sigma^2}{2\mu^2}\Big)t + \frac{1}{\mu}\Big(b^* - \frac{\sigma^2}{2\mu^2} - x\Big)(1-e^{-\mu t}) - \frac{\sigma^2}{4\mu^3}(1-e^{-\mu t})^2\Big)
\end{aligned}$$

und damit die Behauptung. $\square$