Wahrscheinlichkeitstheorie

Fakultät für Mathematik
Bergische Universität Wuppertal

Martin Friesen[1]

Letzte Version vom
7. Februar 2017

[1] friesen@math.uni-wuppertal.de

Inhaltsverzeichnis

1 Grundbegriffe der Wahrscheinlichkeitstheorie

1.1 Wahrscheinlichkeitsräume, Zufallsvariablen und Erwartungswerte

Definition 1.1. *Ein Wahrscheinlichkeitsraum ist ein Maßraum* $(\Omega, \mathcal{F}, \mathbb{P})$ *mit* $\mathbb{P}(\Omega) = 1$.

Hier und im Folgenden bezeichnet $(\Omega, \mathcal{F}, \mathbb{P})$ einen Wahrscheinlichkeitsraum. Wir führen die Folgende Notation ein:

- Ω – Ereignisraum
- $\omega \in \Omega$ – Elementarereignis
- $A \in \mathcal{F}$ – Ereignis
- $\mathbb{P}$ – Wahrscheinlichkeitsmaß
- $\mathbb{P}(A)$ – Wahrscheinlichkeit von A.
- A^c – Ereignis A tritt nicht ein.
- $A \cup B$ – A oder B treten ein.
- $A \cap B$ – A und B treten ein.
- $A \subset B$ – A impliziert B.

Man beachte

$$0 \leq \mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = 1$$

sowie

$$\mathbb{P}(A^c) = \mathbb{P}(\Omega \backslash A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A).$$

Für eine Folge von Mengen $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ gilt

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

Eine Menge $N \in \mathcal{F}$ heißt Nullmenge, falls $\mathbb{P}(N) = 0$ gilt. Wir sagen, dass eine Eigenschaft fast sicher gilt, falls es eine Menge Nullmenge $N \in \mathcal{F}$ gibt und die Eigenschaft für alle $\omega \notin N$ erfüllt ist.

Definition 1.2. *Eine (reellwertige) Zufallsvariable, ist eine messbare Abbildung* $X : \Omega \longrightarrow \mathbb{R}$. *Für eine nicht-negative Zufallsvariable (d.h.* $X \geq 0$ *fast sicher), ist der Erwartungswert definiert über*

$$\mathbb{E}(X) := \int_{\Omega} X d\mathbb{P} := \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

Eine Zufallsvariable X *heißt integrierbar, falls*

$$\mathbb{E}(|X|) = \int_{\Omega} |X| d\mathbb{P} < \infty.$$

Es sei $\mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathbb{P}) = \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ die Menge aller reellwertigen, integrierbaren Zufallsvariablen. Für $X \in \mathcal{L}^1$ ist der Erwartungswert definiert durch

$$\mathbb{E}(X) := \int_\Omega X d\mathbb{P}.$$

Man beachte, dass für $X = 1$ gilt

$$\mathbb{E}(1) = \mathbb{P}(\Omega) = 1.$$

Es ist nicht schwer zu zeigen, dass $\mathcal{L}^1$ ein Vektorraum ist und der Erwartungswert die folgenden Eigenschaften hat (siehe Maßtheorie)

(a) Der Erwartungswert ist linear, d.h. sind $X, Y \in \mathcal{L}^1$ und $a, b \in \mathbb{R}$, so gilt

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

(b) Der Erwartungswert ist monoton, d.h. sind $X, Y \in \mathcal{L}^1$ mit $X \leq Y$ fast sicher, so gilt

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

(c) Sei $X \in \mathcal{L}^1$ und $N \in \mathcal{F}$ eine Nullmenge, dann gilt

$$\mathbb{E}(X \mathbb{1}_N) = 0.$$

Insbesondere: Ist Y eine weitere Zufallsvariable mit $X = Y$ fast sicher, so gilt

$$\mathbb{E}(X) = \mathbb{E}(Y).$$

Analog definieren wir die $\mathcal{L}^p$-Räume für $p \in (0, \infty)$. Genauer, sei $\mathcal{L}^p$ der Vektorraum aller Zufallsvariablen mit

$$\|X\|_{\mathcal{L}^p} := (\mathbb{E}(|X|^p))^{\frac{1}{p}}.$$

Für $p = \infty$ sei $\mathcal{L}^\infty$ der Raum der wesentlich beschränkten Zufallsvariablen. Eine Zufallsvariable heißt wesentlich beschränkt falls

$$\|X\|_{\mathcal{L}^\infty} := \operatorname{ess\,sup}_{\omega \in \Omega} |X(\omega)| := \inf_{N \in \mathcal{F},\ \mathbb{P}(N)=0} \sup_{\omega \in \Omega \setminus N} |X(\omega)| < \infty.$$

Dann ist $\mathcal{L}^\infty$ ein Vektorraum.

Beispiel 1.3. *Sei $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P} = m$ das Lebesgue Maß und für $r \in \mathbb{R}$ sei*

$$X(\omega) := \begin{cases} \frac{1}{\omega^r}, & \omega \neq 0 \\ 0, & \omega = 0 \end{cases}.$$

$r = 0$: Dann ist $X \in \mathcal{L}^p$ für alle $p \in [1, \infty]$.
$r \neq 0$: Dann ist $X \notin \mathcal{L}^\infty$. Für $p \neq \infty$ gilt

$$\mathbb{E}(|X|^p) = \int_{(0,1]} \omega^{-rp} m(d\omega).$$

Das rechte Integral ist endlich genau dann, wenn $rp < 1$, d.h. $X \in \mathcal{L}^p$ falls $rp < 1$.

Lemma 1.4. *Sei $h : \mathbb{R} \longrightarrow \mathbb{R}$ eine konvexe Funktion und $X \in \mathcal{L}^1$ mit $h(X) \in \mathcal{L}^1$. Dann gilt*

$$h(\mathbb{E}(X)) \leq \mathbb{E}(h(X)).$$

Beweis. Da h konvex ist, gibt es eine affin lineare Funktion g mit $g \leq h$ und $g(\mathbb{E}(X)) = h(\mathbb{E}(X))$ (Stützgerade). Es folgt

$$h(\mathbb{E}(X)) = g(\mathbb{E}(X)) = \mathbb{E}(g(X)) \leq \mathbb{E}(h(X)).$$

□

Korollar 1.5. *Für alle $1 \leq p \leq q \leq \infty$ gilt*

$$\|X\|_{\mathcal{L}^p} \leq \|X\|_{\mathcal{L}^q}$$

und somit ist $\mathcal{L}^q \subset \mathcal{L}^p$ ein Teilvektorraum.

Beweis. Die Funktion $h(x) := |x|^{\frac{q}{p}}$ ist für $1 \leq p \leq q < \infty$ konvex. Es folgt mit $Y := |X|^p$

$$(\mathbb{E}(|X|^p))^{\frac{q}{p}} = h(\mathbb{E}(Y)) \leq \mathbb{E}(h(Y)) = \mathbb{E}(|Y|^{\frac{q}{p}}) = \mathbb{E}(|X|^q)$$

und somit die Behauptung im Fall $q < \infty$. Für $1 \leq p < q = \infty$ folgt aus der Monotonie des Erwartungswertes

$$\|X\|_{\mathcal{L}^p} = (\mathbb{E}(|X|^p))^{\frac{1}{p}} \leq \|X\|_{\mathcal{L}^\infty}.$$

□

Satz 1.6. *(Markovsche Ungleichung) Sei X eine Zufallsvariable und $h \geq 0$ monoton steigend. Dann gilt für alle $c > 0$*

$$h(c)\mathbb{P}(X \geq c) \leq \mathbb{E}(h(X)). \tag{1.1}$$

Insbesondere gilt die Markovsche Ungleichung

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}(|X|^p), \quad \varepsilon > 0. \tag{1.2}$$

Weiterhin, sei $X \geq 0$ fast sicher mit $\mathbb{E}(X) = 0$. Dann gilt $X = 0$ fast sicher.

Beweis. Die Ungleichung (1.1) folgt aus

$$h(c)\mathbb{P}(X \geq c) \leq h(c)\mathbb{P}(h(X) \geq h(c)) = \mathbb{E}(h(c)\mathbb{1}_{h(X)\geq h(c)}) \leq \mathbb{E}(h(X)).$$

Die Ungleichung (1.2) ist eine Folgerung für den Spezialfall $h(x) = |x|^p$. Wir zeigen die letzte Behauptung. Aus (1.2) mit $p = 1$ und $\varepsilon = \frac{1}{n}$ folgt für alle $n \geq 1$ mit $|X| = X$ fast sicher

$$\mathbb{P}\left(X \geq \frac{1}{n}\right) \leq n\mathbb{E}(X) = 0.$$

Daraus folgt

$$\mathbb{P}(X > 0) = \mathbb{P}\left(\bigcup_{n\geq 1}\left\{X \geq \frac{1}{n}\right\}\right) \leq \sum_{n=1}^{\infty}\mathbb{P}\left(X \geq \frac{1}{n}\right) = 0$$

und somit $\mathbb{P}(X = 0) = 1$. □

Lemma 1.7. *Sei $1 \leq p \leq \infty$. Dann ist $\|\cdot\|_{\mathcal{L}^p}$ eine Halbnorm, d.h. für $X, Y \in \mathcal{L}^p$ und $a \in \mathbb{R}$ gilt*

(a) $\|aX\|_{\mathcal{L}^p} = |a| \|X\|_{\mathcal{L}^p}$

(b) $\|X + Y\|_{\mathcal{L}^p} \leq \|X\|_{\mathcal{L}^p} + \|Y\|_{\mathcal{L}^p}$.

(c) Ist $\|X\|_{\mathcal{L}^p} = 0$ *so ist* $X = 0$ *fast sicher.*

Beweis. Eigenschaft (a) folgt aus der Linearität des Erwartungswerts. Eigenschaft (b) ist die Minkowski Ungleichung, siehe Maßtheorie oder Funktionalanalysis. Wir zeigen daher nur (c).
$p < \infty$: In diesem Fall folgt aus $\|X\|_{\mathcal{L}^p} = 0$ auch

$$\mathbb{E}(|X|^p) = 0.$$

Damit folgt $|X|^p = 0$ fast sicher und somit $X = 0$ fast sicher.
$p = \infty$: In diesem Fall gibt es zu jedem $n \in \mathbb{N}$ ein $N_n \in \mathcal{F}$ mit $\mathbb{P}(N_n) = 0$ und

$$\sup_{\omega \in \Omega \setminus N_n} |X(\omega)| < \frac{1}{n}.$$

Es sei $N := \bigcup_{n \in \mathbb{N}} N_n$, dann ist $N \in \mathcal{F}$ und es gilt

$$\mathbb{P}(N) \leq \sum_{n=1}^{\infty} \mathbb{P}(N_n) = 0.$$

Ferner gilt für alle $\omega \in \Omega \setminus N$ bereits

$$|X(\omega)| < \frac{1}{n}, \quad \forall n \geq 1$$

und damit $X = 0$ fast sicher. $\square$

Es sei

$$\mathcal{N} := \{X \text{ Zufallsvariable mit } X = 0 \text{ fast sicher }\}.$$

Dann definiert

$$X \sim Y \Longleftrightarrow X - Y \in \mathcal{N}$$

eine Äquivalenzrelation auf $\mathcal{L}^p$ für alle $1 \leq p \leq \infty$. Es sei $L^p := \mathcal{L}^p / \mathcal{N}$ der Raum der Äquivalenzklassen von Zufallsvariablen. Jedes $[X] \in L^p$ hat demnach einen Repräsentanten $X \in \mathcal{L}^p$. Der Erwartungswert ist definiert über

$$\mathbb{E}([X]) := \mathbb{E}(X)$$

und sei

$$\|[X]\|_{L^p} := \|X\|_{\mathcal{L}^p}.$$

Da keine Verwechslung besteht, schreiben wir oft auch X anstelle von $[X]$, wo X ein Repräsentant von X ist. Man beachte, dass eine Äquivalenzklasse $[X]$ nicht punktweise ausgewertet werden kann. D.h. $[X](\omega)$ macht (im Allgemeinen) keinen Sinn. Wählen wir hingegen einen Repräsentanten X von $[X]$, so ist $X(\omega)$ wohldefiniert.
Warnung: Ist X' ein weiterer Repräsentant, so ist $X'(\omega) \neq X(\omega)$ möglich. Die Menge aller solchen ω ist jedoch eine Nullmenge.

Satz 1.8. *Sei* $1 \leq p \leq \infty$. *Dann ist* $(L^p, \|\cdot\|_{L^p})$ *ein Banachraum. Ist* $p = 2$, *so ist* L^2 *ein Hilbertraum mit Skalarprodukt*

$$\langle [X], [Y] \rangle_{L^2} := \mathbb{E}([X \cdot Y]).$$

Satz 1.9. *(Cauchy-Schwartz Ungleichung) Sind* $X, Y \in \mathcal{L}^2$, *so gilt* $X \cdot Y \in \mathcal{L}^1$ *und*

$$\mathbb{E}(|X \cdot Y|) \leq \sqrt{\mathbb{E}(|X|^2)}\sqrt{\mathbb{E}(|Y|^2)}.$$

Diesselbe Aussage gilt für L^2 *und* L^1 *anstelle von* $\mathcal{L}^2$ *und* $\mathcal{L}^1$.

Satz 1.10. *(Hölder Ungleichung) Seien* $1 \leq p, q \leq \infty$ *mit* $\frac{1}{p} + \frac{1}{q} = 1$ *(wo* $\frac{1}{\infty} := 0$ *) und* $X \in \mathcal{L}^p, Y \in \mathcal{L}^q$. *Dann gilt* $X \cdot Y \in \mathcal{L}^1$ *und*

$$\|X \cdot Y\|_{\mathcal{L}^1} \leq \|X\|_{\mathcal{L}^p} \|Y\|_{\mathcal{L}^q}.$$

Diesselbe Aussage gilt für L^p, L^q, L^1 *anstelle von* $\mathcal{L}^p, \mathcal{L}^q, \mathcal{L}^1$.

1.2 Konvergenzbegriffe

Wir wollen einige Resultate aus der Integrationstheorie wiederholen.

Definition 1.11. *Seien* $(X_n)_{n\in\mathbb{N}}$ *und* X *Zufallsvariablen.*

(a) $(X_n)_{n\in\mathbb{N}}$ *konvergiert fast sicher gegen* X, *falls*

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

(b) $(X_n)_{n\in\mathbb{N}}$ *konvergiert in Wahrscheinlichkeit gegen* X, *falls für alle* $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0.$$

(c) $(X_n)_{n\in\mathbb{N}}$ *konvergiert im p-ten Mittel (bzw. in* $\mathcal{L}^p$*) gegen* X *(wo* $p \in (0, \infty)$*), falls*

$$\mathbb{E}(|X_n - X|^p) \longrightarrow 0, \quad n \to \infty.$$

Bemerkung 1.12. *(a) Gilt* $X_n \longrightarrow X$ *in* $\mathcal{L}^1$, *so konvergiert auch die Folge der Erwartungswerte* $(\mathbb{E}(X_n))_{n\in\mathbb{N}}$ *gegen* $\mathbb{E}(X)$. *Die Umkehrung ist im Allgemeinen falsch.*

(b) Gilt $X_n \longrightarrow X$ *in* $\mathcal{L}^q$, *so gilt auch* $X_n \longrightarrow X$ *in* $\mathcal{L}^p$ *für alle* $1 \leq p \leq q \leq \infty$.

Der nächste Satz zeigt, dass unter der zusätzlichen Bedingung, dass die Folge $(X_n)_{n\in\mathbb{N}}$ eine Majorante besitzt, fast sichere Konvergenz die Konvergenz in $\mathcal{L}^1$ impliziert.

Satz 1.13. *(Satz von Lebesgue, Satz der dominierten Konvergenz) Sei* $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1$ *eine Folge von Zufallsvariablen und* $X \in \mathcal{L}^1$. *Es gebe* $Y \in \mathcal{L}^1$ *mit*

$$|X_n| \leq Y, \quad n \geq 1 \quad \textit{fast sicher}$$

und $X_n \longrightarrow X$ *fast sicher. Dann gilt* $X_n \longrightarrow X$ *in* $\mathcal{L}^1$.

Korollar 1.14. *(Satz von Lebesgue für $\mathcal{L}^p$) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariblen mit $X_n \longrightarrow X$ fast sicher. Es gebe ein $Y \in \mathcal{L}^p$ mit $p \in [1,\infty)$ und $|X_n| \leq Y$ fast sicher. Dann gilt $X_n \longrightarrow X$ in $\mathcal{L}^p$.*

Beweis. Es gilt $|X_n - X|^p \longrightarrow 0$ fast sicher. Folglich gibt es $M \in \mathcal{F}$ mit $\mathbb{P}(M) = 1$ und $n_0 \in \mathbb{N}$ mit

$$|X(\omega)| \leq |X(\omega) - X_n(\omega)| + |X_n(\omega)| \leq 1 + Y(\omega), \quad \omega \in M, \quad n \geq n_0.$$

Somit gilt für $n \geq n_0$, $\omega \in M$ mit $(a+b)^p \leq 2^{p-1}(a^p + b^p)$

$$|X_n - X|^p \leq (|X_n| + |X|)^p \leq (1+2Y)^p \leq 2^{p-1}(1 + 2^p Y^p) \in \mathcal{L}^1.$$

Die Behauptung folgt aus dem Satz von Lebesgue für $p = 1$. □

Satz 1.15. *(Satz von Beppo-Levi, Satz der monotonen Konvergenz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge nicht-negativer Zufallsvariablen mit $X_n \leq X_{n+1}$ und $X_n \nearrow X$ fast sicher, wo X eine weitere Zufallsvariable ist. Dann gilt*

$$\mathbb{E}(X_n) \longrightarrow \mathbb{E}(X), \quad n \to \infty.$$

Korollar 1.16. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge nicht-negativer Zufallsvariablen. Dann gilt*

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n). \tag{1.3}$$

Hierbei ist $\infty = \infty$ zugelassen und die linke Seite ist endlich genau dann, wenn die rechte Seite endlich ist. Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge reellwertiger Zufallsvariablen mit $\sum_{n=1}^{\infty} \mathbb{E}(|X_n|) < \infty$ so gilt (1.3) und beide Seiten sind endlich.

Beweis. Übung. □

Lemma 1.17. *(Lemma von Fatou) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge nicht-negativer Zufallsvariablen. Dann gilt*

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

Satz 1.18. *Seien $(X_n)_{n\in\mathbb{N}}$ und X, Y Zufallsvariablen und $p \in (0,\infty)$. Dann gelten die folgenden Aussagen:*

(a) Angenommen $X_n \longrightarrow X$ fast sicher, dann gilt $X_n \longrightarrow X$ in Wahrscheinlichkeit.

(b) Angenommen $X_n \longrightarrow X$ in $\mathcal{L}^p$, dann gilt $X_n \longrightarrow X$ in Wahrscheinlichkeit.

(c) Angenommen $X_n \longrightarrow X$ und $X_n \longrightarrow Y$ in Wahrscheinlichkeit. Dann gilt $X = Y$ fast sicher.

Beweis. (a) Sei $Y_n := \mathbb{1}_{|X_n - X| \geq \varepsilon}$ mit $\varepsilon > 0$ fest. Dann gilt $Y_n \leq 1$ und $Y_n \longrightarrow 0$ fast sicher. Aus dem Satz von Lebesgue folgt

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}(Y_n) \longrightarrow 0, \quad n \to \infty.$$

(b) Die Markovsche Ungleichung impliziert

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}(|X_n - X|^p) \longrightarrow 0, \quad n \to \infty.$$

(c) Für alle $\varepsilon > 0$ gilt

$$\begin{aligned}\mathbb{P}(|X - Y| \geq \varepsilon) &\leq \mathbb{P}(|X_n - X| + |X_n - Y| \geq \varepsilon)\\ &\leq \mathbb{P}\left(\left\{|X_n - X| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|X_n - Y| \geq \frac{\varepsilon}{2}\right\}\right)\\ &\leq \mathbb{P}\left(|X_n - X| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - X| \geq \frac{\varepsilon}{2}\right) \longrightarrow 0, \quad n \to \infty.\end{aligned}$$

□

Beispiel 1.19. *Alle anderen Implikationen sind im Allgemeinen falsch. Sei hierzu* $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), m(dx))$.

(a) Sei $X_n = n^{\frac{1}{p}}\mathbb{1}_{[0,\frac{1}{n}]}$. *Dann ist* $X_n \in \mathcal{L}^p$ *mit* $X_n \longrightarrow 0$ *fast sicher und damit auch in Wahrscheinlichkeit. Aber* $\mathbb{E}(|X_n|^p) = \frac{1}{n}(n^{\frac{1}{p}})^p = 1$ *zeigt, dass* $(X_n)_{n\in\mathbb{N}}$ *nicht in* $\mathcal{L}^p$ *konvergiert.*

(b) Sei

$$X_n := \mathbb{1}_{k2^{-m},(k+1)2^{-m})}, \quad n = 2^m + k, \quad m \in \mathbb{N}_0,\ 0 \leq k < 2^m.$$

Dann gilt

$$\mathbb{P}(|X_n| \geq \varepsilon) \leq \mathbb{P}(|X_n| > 0) = 2^{-m} \longrightarrow 0, \quad n \to \infty$$

und

$$\mathbb{E}(|X_n|^p) = \mathbb{P}(|X_n| > 0) = 2^{-m}.$$

Folglich gilt $X_n \longrightarrow 0$ *in* $\mathcal{L}^p$ *und in Wahrscheinlichkeit. Wir zeigen, dass* $(X_n(\omega))_{n\in\mathbb{N}}$ *für kein* $\omega \in [0,1)$ *konvergiert. Sei* $\omega \in [0,1)$ *fixiert. Für jedes* $m \in \mathbb{N}$ *finden wir genau ein* $k = k(\omega, m)$ *mit* $\omega \in [k2^{-m}, (k+1)2^{-m})$. *Also gilt* $X_{k(\omega,m)+2^m}(\omega) = 0$. *Für alle anderen* $0 \leq k' < 2^m$ *ist* $X_{k'+2^m}(\omega) = 0$. *Damit kann* $(X_n(\omega))_{n\in\mathbb{N}}$ *nicht konvergieren.*

Im Folgenden betrachten wir die Konvergenz in Wahrscheinlihkeitstheorie und fast sichere Konvergenz genauer.

Lemma 1.20. *(Erstes Borel-Cantelli Lemma) Sei* $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ *mit* $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$. *Dann gilt*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 0.$$

Hierbei ist $\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1}\bigcup_{k\geq n} A_k$.

Beweis. Seien $B_n := \bigcup_{k\geq n} A_k$. Dann ist B_n eine fallende Folge (d.h. $B_{n+1} \subset B_n$ für alle $n \geq 1$) und es gilt $\bigcap_{n\geq 1} B_n = \limsup_{n\to\infty} A_n$, d.h. $B_n \searrow \limsup_{n\to\infty} A_n$. Es folgt

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty}\mathbb{P}(B_n) \leq \lim_{n\to\infty}\sum_{k=n}^{\infty}\mathbb{P}(A_k) = 0.$$

□

Satz 1.21. *Sei $(X_n)_{n\in\mathbb{N}}$ mit $X_n \longrightarrow X$ in Wahrscheinlichkeit. Dann gibt es eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $X_{n_k} \longrightarrow X$ fast sicher.*

Beweis. Da $X_n \longrightarrow X$ in Wahrscheinlichkeit, finden wir zu jedem $k \in \mathbb{N}$ ein $n_k \in \mathbb{N}$ mit

$$\mathbb{P}\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \frac{1}{k^2}.$$

Ohne Einschränkung der Allgemeinheit sei $n_k < n_{k+1}$ für alle k. Setze $M := \limsup_{k\to\infty} A_k$ mit $A_k := \left\{|X_{n_k} - X| \geq \frac{1}{k}\right\}$. Nach dem ersten Borel-Cantelli Lemma folgt $\mathbb{P}(M) = 0$. Für $\omega \in M^c = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k^c$, d.h. es gibt ein $n(\omega) \geq 1$ sodass für alle $k \geq n(\omega)$ gilt: $|X_{n_k}(\omega) - X(\omega)| < \frac{1}{k}$. Also gilt

$$\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega), \quad \omega \in M^c.$$

□

Satz 1.22. *(Vollständigkeit bei fast sicherer Konvergenz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit*

$$|X_n - X_m| \longrightarrow 0, \quad n, m \to \infty \quad \textit{fast sicher.} \tag{1.4}$$

Dann gibt es eine Zufallsvariable X mit $X_n \longrightarrow X$ fast sicher.

Beweis. Sei $M \in \mathcal{F}$ mit $\mathbb{P}(M) = 1$ derart, dass (1.4) für jedes $\omega \in M$ gilt. Da $\mathbb{R}$ vollständig ist, gibt es für jedes $\omega \in M$ einen Grenzwert $X(\omega) := \lim_{n\to\infty} X_n(\omega)$. Für $\omega \notin M$ setze $X(\omega) := 0$. Wir zeigen, dass X messbar ist, dann folgt $X_n \longrightarrow X$ fast sicher und somit die Behauptung.

Sei $X|_M : M \longrightarrow \mathbb{R}$, $\omega \longrightarrow X(\omega) = \lim_{n\to\infty} X_n(\omega)$. Dann ist $X|_M$ als Grenzwert messbarer Abbildungen messbar bezüglich

$$\mathcal{F} \cap M := \{A \cap M \mid A \in \mathcal{F}\}.$$

Wegen $M \in \mathcal{F}$ gilt $\mathcal{F} \cap M \subset \mathcal{F}$ und damit ist $X|_M$ auch messbar bezüglich $\mathcal{F}$. Die Darstellung

$$X = \mathbb{1}_M X|_M + \mathbb{1}_{M^c} \cdot 0$$

zeigt, dass X messbar ist. □

Satz 1.23. *(Vollständigkit bei Konvergenz in Wahrscheinlichkeit) Sei $(X_n)_{n\in\mathbb{N}}$ mit*

$$\lim_{n,m\to\infty} \mathbb{P}(|X_n - X_m| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Dann gibt es eine Zufallsvariable X mit $X_n \longrightarrow X$ in Wahrscheinlichkeit.

Beweis. Sei $(X_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge mit

$$\mathbb{P}\left(|X_{n_k} - X_{n_{k+1}}| \geq \frac{1}{k^2}\right) \leq \frac{1}{k^2}.$$

Setze $A_k = \{|X_{n_k} - X_{n_{k+1}}| \geq \frac{1}{k^2}\}$ und $M := \limsup_{n\to\infty} A_n$. Nach dem ersten Borel-Cantelli Lemma gilt $\mathbb{P}(M) = 0$, also $\mathbb{P}(M^c) = 1$. Für jedes $\omega \in M^c = \bigcup_{n\geq 1}\bigcap_{k\geq n} A_k^c$ gibt es ein $n(\omega) \geq 1$ sodass für alle $k \geq n(\omega)$

$$|X_{n_k}(\omega) - X_{n_{k+1}}(\omega)| < \frac{1}{k^2}.$$

Damit ist $(X_{n_k}(\omega))_{k\in\mathbb{N}}$ eine Cauchy Folge in $\mathbb{R}$. Sei $X(\omega) := \lim_{k\to\infty} X_{n_k}(\omega)$ für alle $\omega \in M^c$. Für $\omega \in M$ setze $X(\omega) := 0$. Dann ist X eine Zufallsvariable mit $X_{n_k} \longrightarrow X$ fast sicher. Es bleibt zu zeigen, dass $X_n \longrightarrow X$ in Wahrscheinlichkeit konvergiert. Sei $n \in \mathbb{N}$ und $n_k \geq n$, dann ist

$$\begin{aligned}
\mathbb{P}(|X_n - X| \geq \varepsilon) &\leq \mathbb{P}(|X_n - X_{n_k}| + |X_{n_k} - X| \geq \varepsilon) \\
&\leq \mathbb{P}\left(\left\{|X_n - X_{n_k}| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|X_{n_k} - X| \geq \frac{\varepsilon}{2}\right\}\right) \\
&\leq \mathbb{P}\left(|X_{n_k} - X| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - X_{n_k}| \geq \frac{\varepsilon}{2}\right).
\end{aligned}$$

Mit $n, k \to \infty$ folgt die Behauptung. □

1.3 Beispiele von Wahrscheinlichkeitsräumen

Diskreter Wahrscheinlichkeitsräume

Wir beginnen mit einem einfachen Beispiel mit einer "diskreten" Wahrscheinlichkeitsverteilung.

Beispiel 1.24. *(Punktmaß) Es sei $\omega \in \Omega$ fest. Dann defininiert*

$$\delta_\omega(A) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

ein Wahrscheinlichkeitsmaß auf Ω. Die σ-Algebra $\mathcal{F}$ kann hierbei beliebig gewählt werden. Für eine Zufallsvariable X auf Ω ist der Erwartungswert (für $\mathbb{P} = \delta_\omega$) gegeben durch

$$\mathbb{E}(X) = X(\omega).$$

Der folgende Satz erweitert obiges Beispiel auf abzählbar viele Punktmaße.

Satz 1.25. *Sei $(\omega_n)_{n\in\mathbb{N}} \subset \Omega$ und $(p_n)_{n\in\mathbb{N}} \subset [0,1]$. Dann definiert für jede σ-Algebra $\mathcal{F}$ auf Ω*

$$\mathbb{P} = \sum_{n=1}^{\infty} p_n \delta_{\omega_n}, \tag{1.5}$$

ein Maß auf Ω. Dieses ist ein Wahrscheinlichkeitsmaß genau dann, wenn

$$1 = \mathbb{P}(\Omega) = \sum_{n=1}^{\infty} p_n.$$

Sei X eine Zufallsvariable. Dann ist $X \in \mathcal{L}^1$ genau dann, wenn $\sum\limits_{n=1}^{\infty} |X(\omega_n)| p_n < \infty$ gilt. In diesem Fall ist das Integral bezüglich $\mathbb{P}$ gegeben durch

$$\int\limits_{\Omega} X d\mathbb{P} = \sum_{n=1}^{\infty} X(\omega_n) p_n. \tag{1.6}$$

Beweis. $\mathbb{P}$ ist eine positive Linearkombination von Maßen und damit wieder ein Maß. Die zweite Behauptung ist offensichtlich. Für die letzte Behauptung zeigen wir zuerst (1.6) für alle nicht-negativen, messbaren Funktionen.

Fall 1: Sei $X = \sum\limits_{k=1}^{N} a_k \mathbb{1}_A$ mit $A_k \in \mathcal{F}$ und $a_k \geq 0$. Dann gilt

$$\begin{aligned}\int\limits_{\Omega} X d\mathbb{P} &= \sum_{k=1}^{N} a_k \int\limits_{\Omega} \mathbb{1}_{A_k} d\mathbb{P} = \sum_{k=1}^{N} a_k \mathbb{P}(A_k) \\ &= \sum_{k=1}^{N} a_k \sum_{n=1}^{\infty} p_n \delta_{\omega_n}(A_k) = \sum_{n=1}^{\infty} p_n \sum_{k=1}^{N} a_k \mathbb{1}_{A_k}(\omega_n) = \sum_{n=1}^{\infty} X(\omega_n) p_n.\end{aligned}$$

Fall 2: Sei $X \geq 0$ messbar, dann gibt es eine Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ wie in Fall 1 mit $X_n \nearrow X$ für $n \to \infty$. Folglich gilt

$$\mathbb{E}(X_n) = \sum_{k=1}^{\infty} X_n(\omega_k) p_k, \quad n \geq 1.$$

Beide Seiten sind monoton steigend in n. Aus dem Satz der monotonen Konvergenz folgt (1.6) für alle Zufallsvariablen $X \geq 0$. Insbesondere ist die linke Seite endlich genau dann, wenn die rechte Seite endlich ist.

Fall 3: Es sei X eine beliebige Zufallsvariable. Wenden wir (1.6) für $|X|$ an, so folgt $X \in \mathcal{L}^1$ genau dann wenn $\sum\limits_{n=1}^{\infty} |X(\omega_n)| p_n < \infty$. Zerlegen wir $X = X^+ - X^-$ mit $X^{\pm} \geq 0$ in Positiv- und Negativteil, so folgt

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) = \sum_{n=1}^{\infty} X^+(\omega_n) p_n - \sum_{n=1}^{\infty} X^-(\omega_n) p_n = \sum_{n=1}^{\infty} X(\omega_n) p_n.$$

□

Das folgende Beispiel ist ein Spezialfall von (1.5).

Beispiel 1.26. *(diskrete Gleichverteilung) Sei Ω ist eine endliche Menge mit Potenz-σ-Algebra $\mathcal{F} = 2^{\Omega}$. Die diskrete Gleichverteilung ist dadurch festgelegt, dass jedes Elementarereignis dieselbe Wahrscheinlichkeit hat, d.h. $\mathbb{P}(\{\omega\}) = a$ für ein geeignetes $a \in [0,1]$ und alle $\omega \in \Omega$ gilt. Wegen $\mathbb{P}(\Omega) = 1$ folgt $a = \frac{1}{|\Omega|}$, d.h.*

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}, \quad \omega \in \Omega,$$

wo $|\Omega| := \sum\limits_{\omega\in\Omega} 1$ *die Anzahl der Elemente in* Ω *bezeichnet. Für eine Menge* $A \subset \Omega$ *gilt demnach*

$$\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}) = \frac{|A|}{|\Omega|}.$$

Das so definierte Maß hat die Darstellung

$$\mathbb{P}(A) = \frac{1}{|\Omega|} \sum_{\omega\in\Omega} \delta_\omega(A).$$

Für eine Zufallsvariable X *gilt demnach*

$$\int\limits_\Omega X d\mathbb{P} = \frac{1}{|\Omega|} \sum_{\omega\in\Omega} X(\omega).$$

Das Nachfolgende ist eine Verallgemeinerung und folgt aus Satz 1.25.

Korollar 1.27. *Sei* Ω *eine höchstens abzählbare Menge und* $\mathcal{F}$ *die Potenz-*σ*-Algebra of* Ω*. Dann gelten die folgenden Aussagen.*

(a) Sei $p : \Omega \longrightarrow [0,1]$ *mit* $\sum\limits_{\omega\in\Omega} p(\omega) = 1$*. Dann definiert*

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega), \quad A \subset \Omega \tag{1.7}$$

ein Wahrscheinlichkeitsmaß auf Ω*.*

(b) Jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ *auf* Ω *ist von der Form (1.7) für eine geeignete Funktion* $p : \Omega \longrightarrow [0,1]$*.*

Für eine Zufallsvariable X *mit* $\sum\limits_{\omega\in\Omega} |X(\omega)|p(\omega) < \infty$ *gilt*

$$\int\limits_\Omega X d\mathbb{P} = \sum_{\omega\in\Omega} X(\omega)p(\omega). \tag{1.8}$$

Beweis. (a) $\mathbb{P}$ gegeben durch (1.7) hat die Darstellung

$$\mathbb{P}(A) = \sum_{\eta\in\Omega} p(\eta)\delta_\eta(A).$$

Die Behauptung folgt aus Satz 1.25
(b) Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf Ω und $A \subset \Omega$. Dann gilt

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{\omega\in A} \{\omega\}\right) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}).$$

Also gilt (1.7) mit $p(\omega) := \mathbb{P}(\{\omega\})$. Die Identität (1.8) folgt aus Satz 1.25. □

Beispiel 1.28. *(Münzwurf) Sei $\Omega = \{0,1\}$ mit $\mathcal{F} = 2^\Omega$. Hierbei steht 0 f ür Kopf und 1 für Zahl. Unter der Annahme, dass die Münze fair ist (beide Ausgängesind gleich wahrscheinlich), wählen wir $\mathbb{P}$ als diskrete Gleichverteilung auf Ω, d.h.*

$$\mathbb{P} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1.$$

Führen wir $n \geq 1$ Münzwürfe mit derselben Münze aus, so wählen wir $\mathcal{F} = 2^\Omega$ mit

$$\Omega = \{0,1\}^n = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega_k \in \{0,1\},\ k = 1, \ldots, n\}.$$

Das Wahrscheinlichkeitsmaß ist demnach gegeben durch die Gleichverteilung auf Ω, d.h.

$$\mathbb{P}(\{\omega\}) = \frac{1}{2^n},$$

siehe (1.7). *Die Zufallsvariable*

$$X(\omega) = \sum_{k=1}^{n} \omega_k$$

gibt demnach die absolute Häufigkeit von Zahl an. Um den Erwartungswert zu bestimmen betrachten wir die disjunkte Zerlegung

$$\Omega = \bigcup_{k=0}^{n} A_k, \quad A_k = \{\omega \in \Omega \mid X(\omega) = k\}.$$

Dann gilt $|A_k| = \binom{n}{k}$ und demnach

$$\mathbb{E}(X) = \frac{1}{2^n} \sum_{\omega \in \Omega} X(\omega) = \frac{1}{2^n} \sum_{k=0}^{n} \sum_{\omega \in A_k} X(\omega) = \frac{1}{2^n} \sum_{k=0}^{n} k \binom{n}{k} = \frac{n}{2}.$$

Beispiel 1.29. *(unfairer Münzwurf) Führen wir $n \geq 1$ Münzwürfe mit derselben Münze aus, so wählen wir $\mathcal{F} = 2^\Omega$ mit*

$$\Omega = \{0,1\}^n = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega_k \in \{0,1\},\ k = 1, \ldots, n\}.$$

Wir nehmen an, dass die Wahrscheinlichkeit Kopf (also 0) zu werfen gegeben ist durch $p \in [0,1]$. Die Wahrscheinlichkeit Zahl (also 1) zu werfen, ist demnach $1-p$. Das Wahrscheinlichkeitsmaß ist gegeben durch

$$\mathbb{P}(\{\omega\}) = p^{\sum\limits_{k=1}^{n}(1-\omega_k)} (1-p)^{\sum\limits_{k=1}^{n}\omega_k}, \quad \omega \in \Omega.$$

Wir betrachten wir die Ereignisse

$$A_k = \left\{ \omega \in \Omega \mid \sum_{j=1}^{n} \omega_j = k \right\}, \quad k = 1, \ldots, n.$$

Dann gilt $|A_k| = \binom{n}{k}$. Die Wahrscheinlichkeit k-mal Zahl zu werfen ist gegeben durch

$$\mathbb{P}(A_k) = \sum_{\omega \in A_k} p^{\sum\limits_{j=1}^{n}(1-\omega_j)} (1-p)^{\sum\limits_{j=1}^{n}\omega_j} = \binom{n}{k} p^{n-k}(1-p)^k.$$

Beispiel 1.30. *(Binomial Verteilung) Sei $n \geq 1$ und $p \in [0,1]$. Wähle $\Omega = \{0, \ldots, n\}$, $\mathcal{F} = 2^\Omega$ und*

$$\mathbb{P} = \sum_{k=0}^{n} \binom{n}{k} p^{n-k}(1-p)^k \delta_k$$

Dann heißt $\mathbb{P}$ Binomialverteilung auf $\{0, \ldots, n\}$.

Beispiel 1.31. *(Poisson Verteilung) Sei $\Omega = \mathbb{R}_+$ mit $\mathcal{F} = \mathcal{B}(\mathbb{R}_+)$. Die Poisson Verteilung mit Parameter $\lambda > 0$ ist gegeben durch*

$$\mathbb{P} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta_k.$$

Kontinuierliche Wahrscheinlichkeitsräume auf $\mathbb{R}$

In diesem Abschnitt wollen wir Beispiele für Wahrscheinlichkeitsräume auf $\mathbb{R}$ betrachten. Hierfür betrachten wir das Lebesgue Maß $m(dx)$ auf $\mathbb{R}$ mit Borel-σ-Algebra $\mathcal{B}(\mathbb{R})$ erzeugt durch die offenen Mengen im $\mathbb{R}$.

Beispiel 1.32. *(uniforme Verteilung) Es sei $\Omega \subset \mathbb{R}$ messbar und*

$$\mathcal{F} = \{A \in \mathcal{B}(\mathbb{R}) \mid A \subset \Omega\}$$

die Menge aller Borel messbaren Mengen in Ω. Die uniforme Verteilung auf Ω ist gegeben durch

$$\mathbb{P}(A) := \frac{1}{m(\Omega)} \int_{\Omega} \mathbb{1}_A(x) m(dx) = \frac{m(A)}{m(\Omega)}.$$

Obige Definition kürzen wir ab mit $\mathbb{P}(dx) = \frac{1}{m(\Omega)} \mathbb{1}_\Omega(x) m(dx)$.

Beispiel 1.33. *(Cauchy Verteilung) Es sei $\Omega = \mathbb{R}$ und $\mathbb{P}(dx) = p(x)m(dx)$ mit $c > 0$, $a \in \mathbb{R}$ und*

$$p(x) = \frac{1}{\pi} \frac{c}{c^2 + (x-a)^2}, \quad x \in \mathbb{R},$$

d.h.

$$\mathbb{P}(A) = \int_A p(x) m(dx), \quad A \in \mathcal{B}(\mathbb{R}).$$

Dann ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß und es gilt

$$\int_{\mathbb{R}} |x|^k p(x) m(dx) = \infty, \quad k \geq 1.$$

Beispiel 1.34. *(Normalverteilung auf $\mathbb{R}$) Es sei $\Omega = \mathbb{R}$ und $\mathbb{P}(dx) = p(x)m(dx)$ mit $\mu \in \mathbb{R}$, $\sigma^2 > 0$ und*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Dann ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß mit

$$\int_{\mathbb{R}} xp(x)m(dx) = \mu$$

und

$$\int_{\mathbb{R}} (x-\mu)^2 p(x)m(dx) = \sigma^2.$$

Obige Situation lässt sich abstrakt in folgender Definition zusammenfassen.

Definition 1.35. *Seien $\mathbb{P}, \mathbb{Q}$ zwei Wahrscheinlichkeitsmaße auf einem Maßraum $(\Omega, \mathcal{F})$. $\mathbb{Q}$ heißt absolut stetig bezüglich $\mathbb{P}$, falls für jedes $A \in \mathcal{F}$ gilt*

$$\mathbb{P}(A) = 0 \Longrightarrow \mathbb{Q}(A) = 0.$$

Bemerkung 1.36. *Obige Definition ist äquivalent zu: Für alle $\varepsilon > 0$ gibt es ein $\delta > 0$ mit*

$$\mathbb{P}(A) < \varepsilon \Longrightarrow \mathbb{Q}(A) < \delta, \quad \forall A \in \mathcal{F}.$$

Sei $g : \Omega \longrightarrow \mathbb{R}_+$ integrierbar bezüglich $\mathbb{P}$ und sei

$$\mathbb{Q}(A) := \int_A g(\omega)\mathbb{P}(d\omega), \quad A \in \mathcal{F}.$$

Dann ist $\mathbb{Q}$ ein Maß und $\mathbb{Q}$ ist absolut stetig bezüglich $\mathbb{P}$. Die Umkehrung ist im nachfolgenden Satz formuliert.

Satz 1.37. *(Satz von Radon-Nikodym) Seien $\mathbb{P}, \mathbb{Q}$ zwei Wahrscheinlichkeitsmaße auf Ω. Die folgenden Aussagen sind äquivalent:*

(a) $\mathbb{Q}$ ist absolut stetig bezüglich $\mathbb{P}$.

(b) Es gibt eine bezüglich $\mathbb{P}$ integrierbare Funktion $\frac{d\mathbb{P}}{d\mathbb{Q}} \geq 0$ mit

$$\mathbb{Q}(A) = \int_A \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P}, \quad A \in \mathcal{F}. \tag{1.9}$$

Ist g eine weitere bezüglich $\mathbb{P}$ integrierbare Funktion mit (1.9). Dann gilt

$$\mathbb{P}\left(\left\{\omega \in \Omega \mid g(\omega) = \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega)\right\}\right) = 1.$$

Die Funktion $\frac{d\mathbb{P}}{d\mathbb{Q}}$ wird Radon-Nikodym Ableitung (bzw. Dichte) von $\mathbb{P}$ bezüglich $\mathbb{Q}$ genannt.

Definition 1.38. *Ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathbb{R}$ hat eine Dichte, falls es absolut stetig bezüglich $m(dx)$ ist. Die Dichte p ist definiert als die Radon-Nikodym Ableitung $p = \frac{d\mathbb{P}}{dm}$*

Satz 1.39. *Sei $g \geq 0$ messbar und*

$$\mathbb{Q}(A) := \int_A g(\omega) d\mathbb{P}(\omega), \quad A \in \mathcal{F}.$$

Dann gilt für jede Zufallsvariable X

$$X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{Q}) \iff \int_\Omega |X(\omega)| g(\omega) d\mathbb{P}(\omega) < \infty.$$

In diesem Fall gilt

$$\mathbb{E}_\mathbb{Q}(X) := \int_\Omega X(\omega) d\mathbb{Q}(\omega) = \int_\Omega X(\omega) g(\omega) d\mathbb{P}(\omega).$$

Beweis. Übung. □

Kontinuierliche Wahrscheinlichkeitsräume auf $\mathbb{R}^d$

Es sei $|a| := \sqrt{\sum_{k=1}^d a_k^2}$ die euklidische Norm von $a = (a_1, \dots, a_d) \in \mathbb{R}^d$. Eine messbare Abbildung $X : \Omega \longrightarrow \mathbb{R}^d$ heißt $\mathbb{R}^d$-werige Zufallsvariable. Häufig lassen wir den Zusatz $\mathbb{R}^d$-wertig weg, sofern dieses aus dem Zusammenhang hervorgeht. Eine solche Abbildung hat eine Darstellung in Komponenten über

$$X(\omega) = (X_1(\omega), \dots, X_d(\omega)),$$

wo $X_j : \Omega \longrightarrow \mathbb{R}$ Abbildungen sind.

Bemerkung 1.40. *Sei $X = (X_1, \dots, X_d) : \Omega \longrightarrow \mathbb{R}^d$ eine Abbildung. Dann sind äquivalent:*

(a) X ist eine $\mathbb{R}^d$-wertige Zufallsvariable.

(b) $X_1, \dots, X_d$ sind Zufallsvariablen.

Lemma 1.41. *Es sei $X = (X_1, \dots, X_d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable und $p \in [1, \infty]$. Dann gilt*

$$|X| \in \mathcal{L}^p \iff X_k \in \mathcal{L}^p, \quad k = 1, \dots, d.$$

Beweis. Es reicht zu zeigen, dass

$$\int_\Omega |X|^p d\mathbb{P} < \infty \iff \int_\Omega |X_k|^p d\mathbb{P} < \infty, \quad k = 1, \dots, d.$$

Es gilt $|X|^p = \left(\sum_{k=1}^d |X_k|^2 \right)^{\frac{p}{2}}$. Daraus folgt

$$|X_k|^p = (|X_k|^2)^{\frac{p}{2}} \leq \left(\sum_{k=1}^d |X_k|^2 \right)^{\frac{p}{2}} = |X|^p.$$

Andererseits gilt auch

$$|X|^p \leq \left(d\left(\max_{k=1,\dots,d}|X_k|\right)^2\right)^{\frac{p}{2}} = d^{\frac{p}{2}}\left(\max_{k=1,\dots,d}|X_k|\right)^p \leq d^{\frac{p}{2}}\sum_{k=1}^{d}|X_k|^p.$$

□

Für $p \in (0,\infty)$ sei $\mathcal{L}^p(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^d) = \mathcal{L}^p(\Omega;\mathbb{R}^d)$ der Vektorraum der $\mathbb{R}^d$-wertigen Zufallsvariablen mit

$$\|X\|_{\mathcal{L}^p} := \mathbb{E}(|X|^p) < \infty.$$

Für $p = \infty$ sei $\mathcal{L}^\infty(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^d) = \mathcal{L}^\infty(\Omega;\mathbb{R}^d)$ der Vektorraum der $\mathbb{R}^d$-wertigen Zufallsvariablen mit

$$\|X\|_{\mathcal{L}^\infty} := \inf_{N\in\mathcal{F},\ \mathbb{P}(N)=0}\ \sup_{\omega\in\Omega\setminus N}|X(\omega)| < \infty.$$

Analog zu dem eindimensionalen Fall lassen sich die Räume $L^p(\Omega;\mathbb{R}^d)$ definieren. Diese sind Banachräume.

Bemerkung 1.42. *Es sei $X \in \mathcal{L}^2(\Omega;\mathbb{R}^d)$. Dann gilt*

$$|X_k \cdot X_j| \leq \frac{|X_k|^2 + |X_j|^2}{2} \leq \frac{1}{2}\sqrt{\sum_{k=1}^{d}|X_k|^2} = \frac{1}{2}|X|^2 \in \mathcal{L}^1.$$

Also gilt $X_j \cdot X_k \in \mathcal{L}^1$ für alle $j,k = 1,\dots,d$. Umgekehrt, gilt $X_k \cdot X_j \in \mathcal{L}^1$ für alle $j,k = 1,\dots,d$, so ist offensichtlich $X \in \mathcal{L}^2(\Omega;\mathbb{R}^d)$.

Definition 1.43. *Sei $X = (X_1,\dots X_d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable mit $X \in \mathcal{L}^1(\Omega;\mathbb{R}^d)$. Dann ist der Erwartungswert definiert durch*

$$\mathbb{E}(X) := (\mathbb{E}(X_1),\dots,\mathbb{E}(X_d)).$$

Zufallsvariablen mit Werten in messbaren Räumen

1.4 Charakteristiken von Zufallsvariablen

Definition 1.44. *Sei $(\Omega,\mathcal{F},\mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\Omega',\mathcal{F}')$ ein messbarer Raum und $\varphi : \Omega \longrightarrow \Omega'$ messbar. Das Bildmaß $\mathbb{P}\circ\varphi^{-1}$ auf $(\Omega',\mathcal{F}')$ ist definiert über*

$$(\mathbb{P}\circ\varphi^{-1})(A) = \mathbb{P}(\varphi^{-1}(A)), \quad A \in \mathcal{F}'. \tag{1.10}$$

Der nächste Satz folgt aus standard Argumenten der Maß und Integrationstheorie.

Satz 1.45. *Sei $(\Omega,\mathcal{F},\mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\Omega',\mathcal{F}')$ ein messbarer Raum und $\varphi : \Omega' \longrightarrow \Omega$ messbar. Dann ist $f : \Omega' \longrightarrow \mathbb{R}$ genau dann integrierbar bezüglich $\mathbb{P}\circ\varphi^{-1}$, wenn $f\circ\varphi$ integrierbar ist bezüglich $\mathbb{P}$. In diesem Fall gilt*

$$\int_\Omega f(\varphi(\omega))\mathbb{P}(d\omega) = \int_{\Omega'} f(\omega')(\mathbb{P}\circ\varphi^{-1})(d\omega').$$

Definition 1.46. *(multivariate Normalverteilung)*

(a) Die Standardnormalverteilung auf $\mathbb{R}^d$ *ist gegeben durch ein Wahrscheinlichkeitsmaß* $\mathbb{P}$ *auf* $\mathbb{R}^d$ *mit Dichte*

$$p(x) := (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}}, \quad x \in \mathbb{R}^d.$$

(b) Es sei A *eine reelle* $d \times d$*-Matrix und* $b \in \mathbb{R}^d$*. Sei* $T(x) := Ax + b$ *die dazugehörige affine Abbildung. Dann heißt* $\mathbb{P} \circ T^{-1}$ *(multivariate) Normalverteilung, wo* $\mathbb{P}$ *gegeben ist wie in (a).*

Satz 1.47. *Es gelten die folgenden Aussagen.*

(a) Sei A *invertierbar. Dann hat* $\mathbb{P} \circ T^{-1}$ *eine Dichte gegeben durch*

$$\frac{d(\mathbb{P} \circ T^{-1})}{dm}(x) = (2\pi)^{-\frac{d}{2}} (\det(AA^t))^{-\frac{1}{2}} e^{-\frac{1}{2}\langle (x-b), (AA^t)^{-1}(x-b)\rangle}, \quad x \in \mathbb{R}^d.$$

Hierbei bezeichnet A^t *die zu* A *transponierte Matrix und* $\langle \cdot, \cdot \rangle$ *das Skalarprodukt im* $\mathbb{R}^d$*.*

(b) Ist A *nicht invertierbar, so ist* $\mathrm{Im}(T)$ *eine Nullmenge bezüglich* m *und* $\mathbb{P} \circ T^{-1}$ *ist nicht absolut stetig bezüglich* m*.*

(c) Die Klasse der Normalverteilungen ist invariant unter affin linearen Transformationen $\mathbb{R}^d \longrightarrow \mathbb{R}^d$*. Genauer, sei* $T(x) = Ax + b$ *mit* A *eine* $d \times d$*-Matrix,* $b \in \mathbb{R}^d$ *und sei* ν *eine (multivariate) Normalverteilung. Dann ist* $\nu \circ T^{-1}$ *eine (multivariate) Normalverteilung auf* $\mathbb{R}^d$*.*

Beweis. (a) Sei A invertierbar. Dann ist auch $\Sigma := AA^t$ invertierbar mit $\Sigma^{-1} = (A^t)^{-1}A$ und es gilt $T^{-1}(y) = A^{-1}(y - b) = A^{-1}y - A^{-1}b$. Aus der Translationsinvarianz von m und dem Transformationssatz folgt für jedes Rechteck im $\mathbb{R}^d$

$$(m \circ T^{-1})(R) = m(T^{-1}(R)) = \det(A^{-1})m(R) = \det(A)^{-1}m(R).$$

Da Rechtecke einen $\cap$-stabilen Erzeuger der Borel-σ-Algebra bilden folgt

$$(m \circ T^{-1})(B) = \det(A)^{-1}m(B), \quad B \in \mathcal{B}(\mathbb{R}^d).$$

Also ist $m \circ T^{-1}$ absolut stetig bezüglich m und es gilt mit dem Determinanten Produktsatz

$$\frac{d(m \circ T^{-1})}{dm} = \det(A^{-1}) = (\det(\Sigma))^{-\frac{1}{2}}.$$

Die Behauptung folgt aus

$$\begin{aligned}(\mathbb{P}\circ T^{-1})(B) &= \int\limits_{\mathbb{R}^d} \mathbb{1}_B(x)(\mathbb{P}\circ T^{-1})(dx) = \int\limits_{\mathbb{R}^d} \mathbb{1}_B(T(x))\mathbb{P}(dx)\\ &= \int\limits_{\mathbb{R}^d} \mathbb{1}_B(T(x))(2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|T^{-1}Tx|^2} m(dx)\\ &= \int\limits_{\mathbb{R}^d} \mathbb{1}_B(y)(2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|T^{-1}y|^2} (m\circ T^{-1})(dy)\\ &= \int\limits_B (2\pi)^{-\frac{d}{2}} e^{-\langle y-b,\Sigma^{-1}(y-b)\rangle} (\det(\Sigma))^{-\frac{1}{2}} m(dy).\end{aligned}$$

□

Definition 1.48. *Ein Polnischer Raum E ist ein topologischer Raum E derart dass eine Metrik auf E existiert sodass*

- *Die Topologie auf E stimmt mit der durch die Metrik induzierte Topologie überein.*
- *E ist seperabel und vollständig bezüglich dieser Metrik.*

Im folgenden Betrachten wir einige Beispiele für Polnische Räume.

Beispiel 1.49. *Kontinuierliche Polnische Räume sind $E = \mathbb{R}_+$ und $E = \mathbb{R}^d$. Diskrete Polnische Räume sind $E = \mathbb{Z}_+$ oder $E = \mathbb{Z}^d$ mit $d \geq 1$. Die Metrik ist in allen Fällen gegeben durch*

$$d(x,y) := \sqrt{\sum_{k=1}^{d} |x_k - y_k|^2}, \quad x,y \in E.$$

Beachte, dass im diskreten Fall jede Teilmenge $A \subset E$ offen ist. Abzählbare Produkte von Polnischen Räumen sind wieder Polnisch, sofern diese mit der Produkttopologie versehen sind. Insbesondere ist also

$$\mathbb{R}^{\mathbb{Z}_+} := \{(x_n)_{n\geq 0} \mid x_n \in \mathbb{R}\}$$

ein Polnischer Raum.

Von besonderem Interesse sind hierbei Funktionenräume. Zufallsvariablen mit Werten in solchen Räumen können als stochastische Prozesse bzw. stochastische Felder interpretiert werden. Man beachte, dass jeder seperable Banachraum ein Polnischer Raum ist.

Beispiel 1.50. *Sei $1 \leq p < \infty$. Der Raum der Äquivalenzklassen Funktionen f mit*

$$\|f\|_{L^p} := \left(\int\limits_{\mathbb{R}^d} |f(x)|^p m(dx)\right)^{\frac{1}{p}}$$

wird mit $L^p(\mathbb{R}^d, m(dx))$ bezeichnet. Dieser ist Ein seperabler Banachraum und folglich ein Polnischer Raum, siehe [Wer00, Seite 33]. Insbesondere ist der Raum der quadratintegrierbaren Funktionen $E = L^2(\mathbb{R}^d, m(dx))$ ein Polnischer Raum.

Beispiel 1.51. *Ist T ein kompakter metrischer Raum, so ist*

$$C(T) := \{f : T \longrightarrow \mathbb{R} \mid f \text{ ist stetig } \}$$

mit

$$\|f\|_\infty := \sup_{t \in T} |f(t)|$$

ein seperabler topologischer Raum, siehe [Wer00, Seite 33]. Man wähle zum Beispiel $T = [0,1]$. Zufallsvariablen mit Werten in $C([0,1])$, d.h. $X : \Omega \longrightarrow C([0,1])$ können somit als zufällige Funktionen $X_t : \Omega \longrightarrow \mathbb{R}$ aufgefasst werden.

Beispiel 1.52. *Der Raum $C([0,\infty))$ mit der durch die Metrik*

$$d(f,g) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{d_k(f,g)}{1 + d_k(f,g)}, \quad d_k(f,g) := \sup_{x \in [0,k]} |f(x) - g(x)|$$

erzeugten Topologie ist ein Polnischer Raum.

Zufallsvariablen mit Werten in Räumen von Maßen treten in der Theorie der Punktprozesse häufig auf. Genauer Betrachten wir Räume von Punktmaßen.

Beispiel 1.53. *Der Raum der endlichen Punktmaße (bzw. der Raum aller endlichen Punktkonfigurationen) auf $\mathbb{R}^d$ ist definiert durch*

$$\ddot{\Gamma}_0 := \{\eta : \mathcal{B}(\mathbb{R}^d) \longrightarrow \mathbb{N}_0 \mid \eta \text{ ist endliches Maß}\}.$$

Behauptung: *Jedes $\eta \in \ddot{\Gamma}_0$ hat eine eindeutige Darstellung der Form*

$$\eta = \sum_{k=1}^{N} n_k \delta_{x_k},$$

wo $N \in \mathbb{N}$, $n_1, \ldots, n_N \in \mathbb{N}$ und $x_1, \ldots, x_N \in \mathbb{R}^d$ paarweise verschieden sind. Die Punkte $x_1, \ldots, x_N$ heissen Atome von η.

Beweis. Sei $A := \{x \in \mathbb{R}^d \mid \eta(\{x\}) \geq 1\}$. Angenommen A enthält unendlich viele Elemente. Dann gibt es eine Folge paarweise verschiedener Elemente $(x_n)_{n \in \mathbb{N}} \subset A$. Für diese Folge gilt

$$\infty = \sum_{n=1}^{\infty} \eta(\{x_n\}) = \eta\{\{x_n \mid n \in \mathbb{N}\}) \leq \eta(A) \leq \eta(\mathbb{R}^d) < \infty.$$

Also enthält A nur endlich viele Elemente. Sei $A = \{x_1, \ldots, x_N\}$ für paarweise verschiedene $x_1, \ldots, x_N \in \mathbb{R}^d$ und $N \in \mathbb{N}_0$. Definiere

$$\mu := \eta - \sum_{k=1}^{N} n_k \delta_{x_k}$$

mit $n_k := \eta(\{x_k\})$. Dann fölgt für $B \in \mathcal{B}(\mathbb{R}^d)$ aus $\eta(B \cap \{x_k\}) = n_k\delta_{x_k}(B)$ bereits

$$\begin{aligned}\mu(B) = \eta(B) - \sum_{k=1}^{N} n_k\delta_{x_k}(B) &\geq \eta(A \cap B) - \sum_{k=1}^{N} n_k\delta_{x_k}(B)\\ &= \sum_{k=1}^{N} \eta(B \cap \{x_k\}) - \sum_{k=1}^{N} n_k\delta_{x_k}(B) = 0.\end{aligned}$$

Damit ist $\mu \geq 0$ und folglich ein Maß mit Werten in $\mathbb{N}_0$ derart dass $\mu(\{x\}) = 0$ für alle $x \in \mathbb{R}^d$ gilt. Wir zeigen

$$\mu(B) = 0, \quad B \in \mathcal{B}(\mathbb{R}^d)$$

woraus die Behauptung folgt. Es gilt $\mu(B) = \lim\limits_{R\to\infty} \mu(B \cap B_R)$, wo $B_R = \{x \in \mathbb{R}^d \mid |x| \leq R\}$. Da $\mu(B \cap B_R) \in \mathbb{N}_0$, gibt es ein $R > 0$ derart dass $\mu(B) = \mu(B \cap B_R)$. Für jedes $x \in \mathbb{R}^d$ gilt $\mu(B_\varepsilon(x)) \searrow \mu(\{x\}) = 0$ für $\varepsilon \to 0$. Wegen $\mu(B_\varepsilon(x)) \in \mathbb{N}_0$ gibt es ein $\varepsilon(x) > 0$ derart, dass $\mu(B_{\varepsilon(x)}(x)) = 0$ gilt. Da $\overline{B \cap B_R}$ beschränkt und abgeschlossen ist, ist diese Menge auch kompakt. Wegen

$$\overline{B \cap B_R} \subset \bigcup_{x \in \overline{B \cap B_R}} U_{\varepsilon(x)}(x)$$

mit $U_{\varepsilon(x)}(x) = \{y \in \mathbb{R}^d \mid |y - x| < \varepsilon(x)\}$ gibt es endlich viele Punkte $x_1, \dots, x_m$ derart, dass

$$\overline{B \cap B_R} \subset \bigcup_{j=1}^{m} U_{\varepsilon(x_j)}(x_j) \subset \bigcup_{j=1}^{m} B_{\varepsilon(x_j)}(x_j).$$

Daraus folgt

$$\mu(B) = \mu(B \cap B_R) \leq \mu(\overline{B \cap B_R}) \leq \sum_{j=1}^{m} \mu(B_{\varepsilon(x_j)}(x_j)) = 0.$$

□

Der Raum der endlichen Teilmengen von $\mathbb{R}^d$ (bzw. der Raum der endlichen Konfigurationen) ist gegeben durch

$$\Gamma_0 := \left\{ \eta \subset \mathbb{R}^d \mid |\eta| := \sum_{x \in \eta} 1 < \infty \right\}.$$

Jedes $\eta \in \Gamma_0$ kann vermöge $\sum\limits_{x\in\eta} \delta_x$ mit einem Element in $\ddot{\Gamma}_0$ identifiziert werden.

Es sei $C_c(\mathbb{R}^d)$ der Raum der stetigen Funktionen $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ mit kompaktem Träger. Eine Funktion $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ hat kompakten Träger, falls es ein $R > 0$ gibt mit $f(x) = 0$ für alle $|x| \geq R$.

Beispiel 1.54. *Der Raum aller lokal endlichen Teilmengen von $\mathbb{R}^d$ (bzw. lokal endlichen Konfigurationen) ist definiert durch*

$$\Gamma := \{\gamma \subset \mathbb{R}^d \mid |\gamma \cap K| < \infty \quad K \subset \mathbb{R}^d \text{ kompakt } \}.$$

Jedes $\gamma \in \Gamma$ kann vermöge $\sum_{x \in \gamma} \delta_x$ mit einem lokal endlichen Punktmaß auf $\mathbb{R}^d$ identifiziert werden. Für jedes $f \in C_c(\mathbb{R}^d)$ sei

$$I_f(\gamma) := \sum_{x \in \gamma} f(x) = \int_\Gamma f(x)\gamma(dx)$$

das Integral von f bezüglich des Punktmaßes γ. Die Topologie auf Γ ist definiert als die kleinste Topologie derart, dass alle Abbildungen I_f stetig sind. Es lässt sich zeigen, dass Γ bezüglich dieser Topologie ein Polnischer Raum ist, siehe [KK06].

Die Erweiterung auf alle lokal endlichen Punktmaße wird im nächsten Beispiel behandelt.

Beispiel 1.55. *Der Raum aler lokal endlichen Punktmaße ist gegeben durch*

$$\ddot{\Gamma} := \{\widehat{\gamma} : \mathcal{B}(\mathbb{R}^d) \longrightarrow \mathbb{N}_0 \cup \{\infty\} \mid \widehat{\gamma}(K) < \infty, \quad K \subset \mathbb{R}^d \textit{ kompakt } \}.$$

Behauptung: *Jedes $\widehat{\gamma} \in \ddot{\Gamma}$ hat die Darstellung*

$$\widehat{\gamma} = \sum_{x \in \gamma} n_k \delta_x,$$

wo $n_x \in \mathbb{N}_0$ und $\gamma \in \Gamma$.

Beweis. Sei $B_N := \{x \in \mathbb{R}^d \mid |x| \leq N\}$ wo $N \in \mathbb{N}$. Dann ist

$$\mathbb{R}^d = \bigcup_{N=0}^{\infty} (B_{N+1} \backslash B_N)$$

eine disjunkte Zerlegung von $\mathbb{R}^d$, wo $B_0 := \varnothing$. Folglich gilt für jedes $B \in \mathcal{B}(\mathbb{R}^d)$

$$\widehat{\gamma}(B) = \sum_{N=0}^{\infty} \widehat{\gamma}((B_{N+1} \backslash B_N) \cap B).$$

Sei $\widehat{\gamma}_N(B) := \widehat{\gamma}((B_{N+1} \backslash B_N) \cap B)$. Dann ist $\widehat{\gamma} = \sum_{N=0}^{\infty}$ und es ist nicht schwer zu sehen, dass $\widehat{\gamma}_N \in \ddot{\Gamma}_0$ für alle $N \geq 0$. Folglich gibt es $n_1^{(N)}, \ldots, n_{k(N)}^{(N)} \in \mathbb{N}_0$ sowie $x_1^{(N)}, \ldots, x_{k(N)}^{(N)} \in \mathbb{B}_{N+1} \backslash B_N$ mit $k(N) \in \mathbb{N}$ und $N \in \mathbb{N}_0$ derart, dass

$$\widehat{\gamma}_N = \sum_{j=1}^{k(N)} n_j^{(N)} \delta_{x_j^{(N)}}, \quad N \geq 0.$$

□

2 Unabhängigkeit und bedingte Erwartungswerte

2.1 Unabhängigkeit

Definition 2.1. *Sei I eine beliebige Indexmenge.*

(a) Eine Familie von Ereignissen $\{A_i \mid i \in I\} \subset \mathcal{F}$ heißt unabhängig, wenn

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$$

für alle endlichen Teilmengen $J \subset I$ gilt.

(b) Die Familie heißt paarweise unabhängig, wenn

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j), \quad \forall i, j \in I, \quad i \neq j.$$

Bemerkung 2.2. *Unabhängigkeit impliziert paarweise unabhängigkeit. Die Umkehrung ist im Allgemeinen falsch.*

Beispiel 2.3. *(a) Sei $\Omega = \{1, 2, 3, 4\}$ mit $\mathcal{F} = 2^\Omega$ und Gleichverteilung $\mathbb{P}$ auf Ω. Seien*

$$A = \{1, 2\}, \quad B = \{2, 3\}, \quad C = \{1, 3\}.$$

Dann gilt

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$$

und

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}.$$

D.h. die Ereignisse sind paarweise unabhängig. Die Ereignisse sind nicht unabhängig, da

$$\mathbb{P}(A \cap B \cap C) = 0 \neq \left(\frac{1}{2}\right)^3 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

(b) Sei $I = \{1, 2\}$ mit Mengen A_1, A_2. Dann ist Unabhängigkeit äquivalent zu

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

(c) Sei $I = \{1, 2, 3\}$ mit Mengen A_1, A_2, A_3. Dann ist Unabhängigkeit äquivalent zu $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$, $\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_3)$, $\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2)\mathbb{P}(A_3)$ und

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

Definition 2.4. *Sei I eine Indexmenge und für $i \in I$ seien Mengensysteme $\mathcal{E}_i \subset \mathcal{F}$ gegeben. Die Familie $\{\mathcal{E}_i \mid i \in I\}$ heißt unabhängig, wenn für jede Wahl $A_i \in \mathcal{E}_i$ die Familie $\{A_i \mid i \in I\}$ von Ereignissen unabhängig ist.*
Äquivalent: *Für jede endliche Teilmenge $J \subset I$ und jede Wahl $A_j \in \mathcal{E}_j$ mit $j \in J$ gilt*

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Bemerkung 2.5.

(a) *Seien* $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ *mit* $i = 1, \dots, n$ *Wahrscheinlichkeitsräume. Setze* $\Omega = \Omega_1 \times \dots \times \Omega_n$ *mit* $\mathcal{F} = \bigotimes_{i=1}^{n} \mathcal{F}_i$ *und* $\mathbb{P} = \bigotimes_{i=1}^{n} \mathbb{P}_i$*. Seien Mengen* A_i *gegeben durch*

$$A_1 := B_1 \times \Omega_2 \times \dots \times \Omega_n,\ B_1 \in \mathcal{F}_1$$
$$\dots$$
$$A_n := \Omega_1 \times \dots \times B_n,\ B_n \in \mathcal{F}_n.$$

Dann ist die Familie $\{A_1, \dots, A_n\}$ *unabhängig unter* $\mathbb{P}$*.*

(b) $\{\mathcal{E}_i \mid i \in I\}$ *ist unabhängig genau dann, wenn* $\{\mathcal{E}_i \mid i \in J\}$ *für alle endlichen Teilmengen* $J \subset I$ *unabhängig ist.*

(c) *Ist* $\mathcal{A}_i \subset \mathcal{E}_i$ *für alle* $i \in I$*, so gilt*

$$\{\mathcal{E}_i \mid i \in I\} \text{ ist unabhängig } \Longrightarrow \{\mathcal{A}_i \mid i \in I\} \text{ ist unabhängig.}$$

Die Umkehrung ist im Allgemeinen falsch.

Satz 2.6. *Sei* $\{\mathcal{E}_i \mid i \in I\}$ *eine unabhängige Familie. Dann ist auch die Familie der erzeugten Dynkin-Systeme* $\{d(\mathcal{E}_i) \mid i \in I\}$ *unabhängig.*

Beweis. Wegen Bemerkung 2.5 sei ohne Einschränkung I endlich mit $I = \{1, \dots, n\}$. Für $i_0 \in I$ sei

$$\mathcal{D}_{i_0} := \{A \in \mathcal{F} \mid \{\mathcal{E}_1, \dots, \mathcal{E}_{i_0-1}, \{A\}, \mathcal{E}_{i_0+1}, \dots, \mathcal{E}_n\} \text{ ist unabhängige Familie}\}.$$

Wir zeigen, dass $\mathcal{D}_{i_0}$ ein Dynkin-System ist. Hierfür sei $\{i_1, \dots, i_k\} \subset I \backslash \{i_0\}$ beliebig und wähle $A_{i_\nu} \in \mathcal{E}_{i_\nu}$, wo $\nu = 1, \dots, k$.

1. Da die kleinere Auswahl von Mengen auch unabhängig ist, folgt

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k} \cap \Omega) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}) \cdot 1.$$

   Also ist $\Omega \in \mathcal{D}_{i_0}$.

2. Sei $A \in \mathcal{D}_{i_0}$. Dann ist

$$\begin{aligned}\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k} \cap A^c) &= \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k} \cap \Omega) - \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k} \cap A) \\ &= \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})(1 - \mathbb{P}(A)) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})\mathbb{P}(A^c)\end{aligned}$$

   und folglich $A^c \in \mathcal{D}_{i_0}$.

3. Seien $B_l \in \mathcal{D}_{i_0}$ mit $l \in \mathbb{N}$ disjunkt. Dann ist

$$\begin{aligned}&\mathbb{P}\left(A_{i_1} \cap \dots \cap A_{i_k} \cap \bigcup_{l=1}^{\infty} B_l\right) = \mathbb{P}\left(\bigcup_{l=1}^{\infty} A_{i_1} \cap \dots \cap A_{i_k} \cap B_l\right) \\ &= \sum_{l=1}^{\infty} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k} \cap B_l) = \sum_{l=1}^{\infty} \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}) \cdot \mathbb{P}(B_l) \\ &= \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}) \mathbb{P}\left(\bigcup_{l=1}^{\infty} B_l\right)\end{aligned}$$

und folglich $\bigcup\limits_{l=1}^{\infty} B_l \in \mathcal{D}_{i_0}$.

Damit ist $\mathcal{D}_{i_0}$ ein Dynkin System und $\{\mathcal{E}_1, \dots, \mathcal{E}_{i_0-1}, \mathcal{D}_{i_0}, \mathcal{E}_{i_0+1}, \dots, \mathcal{E}_n\}$ ist eine unabhängige Familie. Es gilt $\mathcal{E}_{i_0} \subset \mathcal{D}_{i_0}$ und da $\mathcal{D}_{i_0}$ ein Dynkin System ist, folgt $\mathcal{E}_{i_0} \subset d(\mathcal{E}_{i_0}) \subset \mathcal{D}_{i_0}$. Nach Bemerkung 2.5 ist $\{\mathcal{E}_1, \dots, \mathcal{E}_{i_0-1}, d(\mathcal{E}_{i_0}), \mathcal{E}_{i_0+1}, \dots, \mathcal{E}_n\}$ eine unabhängige Familie. Da $i_0 \in I$ beliebig war, folgt die Behauptung durch Induktion. □

Korollar 2.7. *Sei $\{\mathcal{E}_i \mid i \in I\}$ eine unabhängige Familie von $\cap$-stabilen Mengensystemen mit $\mathcal{E}_i \subset \mathcal{F}$. Dann ist $\{\sigma(\mathcal{E}_i) \mid i \in I\}$ eine unabhängige Familie von σ-Algebren.*

Satz 2.8. *Sei $\{A_i \mid i \in I\}$ unabhängig mit $A_i \in \mathcal{F}$. Für jedes $i \in I$ seien $A_i^* \in \{\emptyset, A_i, A_i^c, \Omega\}$. Dann ist $\{A_i^* \mid i \in I\}$ auch unabhängig.*

Beweis. Sei $\mathcal{E}_i = \{A_i\}$, dann ist nach Voraussetzung $\{\mathcal{E}_i \mid i \in I\}$ eine unabhängige Familie von $\cap$-stabilen Mengensystemen. Nach Korollar 2.7 ist auch $\{\sigma(\mathcal{E}_i) \mid i \in I\}$ unabhängig. Die Behauptung folgt aus $\sigma(\mathcal{E}_i) = \{\emptyset, A_i, A_i^c, \Omega\}$. □

Lemma 2.9. *(Gruppieren) Sei $\{\mathcal{E}_i \mid i \in I\}$ unabhängig und $\cap$-stabil mit $\mathcal{E}_i \subset \mathcal{F}$. Sei $I = \bigcup\limits_{j \in J} I_j$ eine disjunkte Zerlegung und $\mathcal{A}_j = \sigma\left(\bigcup\limits_{i \in I_j} \mathcal{E}_j\right)$. Dann ist $\{\mathcal{A}_j \mid j \in J\}$ eine unabhängige Familie.*

Beweis. Für $k \in J$ sei

$$\mathcal{D}_k = \{A_{i_1} \cap \dots \cap A_{i_n} \mid A_{i_l} \in \mathcal{E}_{i_l},\ \ i_1, \dots, i_n \in I_k,\ n \in \mathbb{N}\}$$

die Menge aller endlichen Schnitte von Mengen aus $\bigcup\limits_{i \in I_k} \mathcal{E}_i$. Dann ist $\mathcal{D}_k$ per Definition $\cap$-stabil. Weiterhin ist $\{\mathcal{D}_k \mid k \in J\}$ unabhängig, da die $\mathcal{E}_i$ unabhängig sind. Wegen

$$\mathcal{A}_k = \sigma\left(\bigcup_{i \in I_k} \mathcal{E}_i\right) = \sigma(\mathcal{D}_k), \quad k \in J$$

und Korollar 2.7 ist $\{\mathcal{A}_k \mid k \in J\}$ unabhängig. □

Satz 2.10. *(Kolmogorovsches 0-1-Gesetz) Sei $(\mathcal{F}_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger σ-Algebren mit $\mathcal{F}_n \subset \mathcal{F}$. Sei*

$$\mathcal{A}_n := \bigvee_{k \geq n} \mathcal{F}_k := \sigma\left(\bigcup_{k \geq n} \mathcal{F}_k\right)$$

und $\mathcal{T}_\infty := \bigcap\limits_{n \geq 1} \mathcal{A}_n$ die terminale σ-Algebra. Dann gilt

$$A \in \mathcal{T}_\infty \Longrightarrow \mathbb{P}(A) \in \{0, 1\}.$$

Notation: $\mathcal{E}_1 \perp \mathcal{E}_2 \Longleftrightarrow \{\mathcal{E}_1, \mathcal{E}_2\}$ unabhängige Familie.

Beweis. **Idee:** $A \in \mathcal{T}_\infty$, dann ist A unabhängig von A, d.h. $\mathbb{P}(A)^2 = \mathbb{P}(A)$.
Nach Korollar 2.7 und Lemma 2.9 gilt für $n \in \mathbb{N}$

$$\mathcal{A}_{n+1} \perp \bigvee_{k=1}^{n} \mathcal{F}_k.$$

Insbesondere also $\mathcal{T}_\infty = \bigcap_{k \geq n+1} \mathcal{A}_k \perp \bigvee_{k=1}^{n} \mathcal{F}_k$. Daraus folgt

$$\mathcal{T}_\infty \perp \bigcup_{n \geq 1} \bigvee_{k=1}^{n} \mathcal{F}_k,$$

da zu jeder Menge aus $\bigcup_{n \geq 1} \bigvee_{k=1}^{n} \mathcal{F}_k$ es ein $n \in \mathbb{N}$ gibt, sodasss diese in $\bigvee_{k=1}^{n} \mathcal{F}_k$ liegt. Korollar 2.7 impliziert

$$\mathcal{T}_\infty \perp \bigvee_{n=1}^{\infty} \mathcal{F}_n = \sigma\left(\bigcup_{n \geq 1} \bigvee_{k=1}^{n} \mathcal{F}_k \right).$$

Es ist $\mathcal{T}_\infty \subset \mathcal{A}_k \subset \mathcal{A}_1 = \bigvee_{k=1}^{\infty} \mathcal{F}_k$ und somit $\mathcal{T}_\infty \perp \mathcal{T}_\infty$. □

Satz 2.11. *Sei $\mathcal{T} \subset \mathcal{F}$ eine σ-Algebra mit $\mathbb{P}(A) \in \{0,1\}$ für alle $A \in \mathcal{T}$. Sei $Z : \Omega \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ eine $\mathcal{T}$ messbare Zufallsvariable. Dann gibt es ein $c \in \mathbb{R} \cup \{\pm\infty\}$ mit $\mathbb{P}(Z = c) = 1$.*

Beweis. Nach Voraussetzung ist $\{Z \leq t\} \in \mathcal{T}$ für alle $t \in \mathbb{R}$. Also ist $F(t) := \mathbb{P}(Z \leq t)$ monoton wachsend mit $F(t) \in \{0,1\}$ für alle t.

Fall 1. $F(t) = 0$ für alle $t \in \mathbb{R}$. Dann ist $\mathbb{P}(Z > n) = 1$ für alle $n \in \mathbb{N}$ und somit

$$\mathbb{P}(Z = +\infty) = \lim_{n \to \infty} \mathbb{P}(Z > n) = 1.$$

Also $c = +\infty$.

Fall 2. $F(t) = 1$ für alle $t \in \mathbb{R}$. Dann ist

$$\mathbb{P}(Z = -\infty) = \lim_{n \to \infty} \mathbb{P}(Z \leq -n) = 1.$$

Also $c = -\infty$.

Fall 3. Es gibt ein $t_0 \in \mathbb{R}$ mit

$$F(t) = \begin{cases} 0, & t < t_0 \\ 1, & t \geq t_0 \end{cases}.$$

Dann ist

$$1 = F\left(t_0 + \frac{1}{n}\right) - F\left(t_0 - \frac{1}{n}\right) = \mathbb{P}\left(t_0 - \frac{1}{n} < Z \leq t_0 + \frac{1}{n}\right), \quad n \in \mathbb{N}.$$

Also gilt

$$\mathbb{P}(Z = t_0) = \lim_{n \to \infty} \mathbb{P}\left(t_0 - \frac{1}{n} < Z \leq t_0 + \frac{1}{n}\right) = 1.$$

□

Definition 2.12. *Sei I eine Indexmenge und X_i Zufallsvariablen mit Werten in einem Maßraum $(E_i, \mathcal{E}_i)$ für alle $i \in I$. Die Familie $\{X_i \mid i \in I\}$ heißt unabhängig, falls die σ-Algebren*

$$\sigma(X_i) := X_i^{-1}(\mathcal{E}_i) = \{X_i^{-1}(A) \mid A \in \mathcal{E}_i\}$$

eine unabhängige Familie $\{\sigma(X_i) \mid i \in I\}$ bilden.

Bemerkung 2.13. *(a) Falls X_i messbar bezüglich $\mathcal{D}_i$-$\mathcal{E}_i$ sind und $\{\mathcal{D}_i \mid i \in I\}$ unabhängig sind, so ist $\{X_i \mid i \in I\}$ unabhängig.*

(b) $\{X_i \mid i \in I\}$ ist eine Familie von unabhängigen Zufallsvariablen genau dann, wenn für alle $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ paarweise verschieden und $A_{i_1} \in \mathcal{E}_{i_1}, \ldots, A_{i_n} \in \mathcal{E}_{i_n}$ gilt

$$\mathbb{P}(X_{i_1} \in A_{i_1}, \ldots, X_{i_n} \in A_{i_n}) = \prod_{j=1}^{n} \mathbb{P}(X_{i_j} \in A_{i_j}). \tag{2.1}$$

Wegen $\mathbb{P}(X_{i_1} \in A_{i_1}, \ldots, X_{i_n} \in A_{i_n}) = \mathbb{P} \circ (X_{i_n}, \ldots, X_{i_n})^{-1}(A_{i_1} \times \cdots \times A_{i_n})$ ist folglich $\{X_i \mid i \in I\}$ genau dann unabhängig, wenn alle endlich dimensionalen Verteilungen $\mathbb{P} \circ (X_{i_n}, \ldots, X_{i_n})^{-1}$ Produktmaße sind, d.h.

$$\mathbb{P} \circ (X_{i_n}, \ldots, X_{i_n})^{-1} = \mathbb{P} \circ X_{i_1}^{-1} \otimes \cdots \otimes \mathbb{P} \circ X_{i_n}^{-1}.$$

(c) Seien $\{X_i \mid i \in I\}$ unabhängig und $\varphi_i : E_i \longrightarrow W_i$ messbar. Dann ist $\{\varphi_i \circ X_i \| \ i \in I\}$ unabhängig.

Lemma 2.14. *Seien X_i Zufallsvariablen mit Werten in $\mathbb{R}$. Genau dann ist $\{X_i \mid i \in I\}$ unabhängig, wenn für alle $n \in \mathbb{N}$, alle paarweise verschiedenen $i_1, \ldots, i_n \in I$ und alle $t_1, \ldots, t_n \in \mathbb{R}$*

$$\mathbb{P}(X_{i_1} \leq t_1, \ldots, X_{i_n} \leq t_n) = \prod_{j=1}^{n} \mathbb{P}(X_{i_j} \leq t_j). \tag{2.2}$$

Beweis. Ist $\{X_i \mid i \in I\}$ unabhängig, so folgt (2.2) direkt aus der Definition. Es gelte also (2.2). Sei $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ paarweise verschieden. Es reicht (2.1) zu zeigen. Diese gilt wegen (2.2) für alle Mengen der Form $A_{i_k} = (-\infty, t_k]$ mit $t_k \in \mathbb{R}$ und $k = 1, \ldots, n$. Da Mengen dieser Form ein $\cap$-stabiles Erzeugendensystem bilden, folgt die Behauptung. □

Lemma 2.15. *Seien X, Y unabhängige $\mathbb{R}$-wertige Zufallsvariablen. Sei $X = X^+ - X^-$ und $Y = Y^+ - Y^-$ die Zerlegung in Positiv und Negativteil. Dann sind unabhängig:*

- X^+, Y^+.
- X^-, Y^-.
- X^+, Y^-.
- X^-, Y^+.

Beweis. Wegen $X^+ = \max\{0, X\}$ und $X^- = \max\{0, -X\}$ sind $X^\pm$ messbar ist bezüglich $\sigma(X)$. Also gilt $\sigma(X^\pm) \subset \sigma(X)$. Analog sehen wir $\sigma(Y^\pm) \subset \sigma(Y)$. Da $\sigma(X) \perp \sigma(Y)$ folgt die Behauptung. $\square$

Satz 2.16. *Seien $X, Y \in \mathcal{L}^1$.*

(a) Sind X, Y unabhängig, so gilt $X \cdot Y \in \mathcal{L}^1$ und $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

(b) Sind X, Y unabhängig mit $X, Y \in \mathcal{L}^2$, so gilt $\operatorname{cov}(X, Y) = 0$ und

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y).$$

(c) X, Y sind genau dann unabhängig, wenn für alle messbaren, beschränkten Funktionen $f, g : \mathbb{R} \longrightarrow \mathbb{C}$ gilt

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)). \tag{2.3}$$

Beweis. Teil (b) folgt direkt aus (a). Wir zeigen zunächst (a). Wir zeigen die Behauptung für $X, Y \geq 0$. Der allgemeine Fall ist eine Konsequenz aus Lemma 2.15 und der Zerlegung in Positiv- und Negativteil.
Fall 1. $X = \mathbb{1}_A$ und $Y = \mathbb{1}_B$. Dann gilt $0 \leq X \cdot Y \leq 1$, also $X \cdot Y \in \mathcal{L}^1$. Ferner gilt

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(\mathbb{1}_{A \cap B}) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}(X)\mathbb{E}(Y).$$

Fall 2. $X = \sum\limits_{n=1}^{N} a_n \mathbb{1}_{A_n}$ und $Y = \sum\limits_{m=1}^{M} b_m \mathbb{1}_{B_m}$ mit $a_n, b_m \geq 0$, $N, M \in \mathbb{N}$ und $A_n, B_m \in \mathcal{F}$. Dann gilt

$$X \cdot Y = \sum_{n=1}^{N} \sum_{m=1}^{M} a_n b_m \mathbb{1}_{A_n \cap B_m} \in \mathcal{L}^1.$$

Ferner gilt

$$\begin{aligned}
\mathbb{E}(X \cdot Y) &= \sum_{n=1}^{N} \sum_{m=1}^{M} a_n b_m \mathbb{P}(A_n \cap B_m) = \sum_{n=1}^{N} \sum_{m=1}^{M} a_n b_m \mathbb{P}(A_n)\mathbb{P}(B_m) \\
&= \left(\sum_{n=1}^{N} a_n \mathbb{P}(A_n) \right) \left(\sum_{m=1}^{M} b_m \mathbb{P}(B_m) \right) = \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

Fall 3. Seien $X, Y \in \mathcal{L}^1$, X_n, Y_n Folgen von Elementarfunktionen wie in Fall 2 mit $X_n \nearrow X$ und $Y_n \nearrow Y$. Dann gilt $X_n \cdot Y_n \in \mathcal{L}^1$, $X_n \cdot Y_n \nearrow X \cdot Y$ und $\mathbb{E}(X_n \cdot Y_n) = \mathbb{E}(X_n)\mathbb{E}(Y_n)$. Aus dem Satz der monotonen Konvergenz folgt

$$\mathbb{E}(X \cdot Y) = \lim_{n \to \infty} \mathbb{E}(X_n \cdot Y_n) = \lim_{n \to \infty} \mathbb{E}(X_n)\mathbb{E}(Y_n) = \mathbb{E}(X)\mathbb{E}(Y) < \infty.$$

Es bleibt (c) zu zeigen. Sind X, Y unabhängig, so sind auch $f(X), g(Y)$ unabhängig und folglich gilt (2.3). Umgekehrt gelte (2.3). Seien $A \in \sigma(X)$ und $B \in \sigma(Y)$, dann gibt es A' und B' mit $A = X^{-1}(A')$ und $B = X^{-1}(B')$. Es gilt $\mathbb{1}_{A'}(X) = \mathbb{1}_A$ und $\mathbb{1}_{B'}(Y) = \mathbb{1}_B$ und folglich

$$\mathbb{P}(A \cap B) = \mathbb{E}(\mathbb{1}_{A'}(X)\mathbb{1}_{B'}(Y)) = \mathbb{E}(\mathbb{1}_{A'}(X))\mathbb{E}(\mathbb{1}_{B'}(Y)) = \mathbb{P}(A)\mathbb{P}(B).$$

$\square$

2.2 Bedingte Erwartungswerte

Im folgenden betrachten wir stets $\mathbb{R}$-wertige Zufallsvariablen. Für $\mathbb{R}^d$-wertige Zufallsvariablen lassen sich alle Resultate durch komponentenweise Anwendung übertragen.

Definition 2.17. *Seien $A, B \in \mathcal{F}$ und $\mathbb{P}(B) > 0$. Die bedingte Wahrscheinlichkeit von A gegeben B ist definiert über*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Bemerkung 2.18. *$A \longmapsto \mathbb{P}(A|B)$ ist für festes B mit $\mathbb{P}(B) > 0$ ein Wahrscheinlichkeitsmaß. Für eine Zufallsvariable $X \in \mathcal{L}^1$ ist der bedingte Erwartungswert von X gegeben B definiert über*

$$\mathbb{E}(X|B) := \int_\Omega X(\omega)\mathbb{P}(d\omega|B).$$

Satz 2.19. *(Formel der totalen Wahrscheinlichkeit) Sei $(B_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ eine Folge disjunkter Mengen mit $\bigcup_{n\geq 1} B_n = \Omega$ und $\mathbb{P}(B_n) > 0$ für alle $n \geq 1$. Dann gilt für alle $A \in \mathcal{F}$*

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

Beweis. Es gilt

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n\geq 1} A \cap B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A \cap B_n) = \sum_{n=1}^{\infty} \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

□

Definition 2.20. *Sei X eine Zufallsvariable mit höchstens abzählbar vielen Werten $\{x_k\}_{k\in\mathbb{N}}$. Definiere*

$$\mathbb{P}(A|X = x_k) = \begin{cases} \frac{\mathbb{P}(A\cap\{X=x_k\})}{\mathbb{P}(X=x_k)}, & \mathbb{P}(X = x_k) > 0 \\ 0, & \textit{sonst} \end{cases}.$$

Der bedingte Erwartungswert einer weiteren Zufallsvariable Y gegeben $X = x_k$ ist in diesem Fall definiert als

$$\mathbb{E}(Y|X = x_k) = \int_\Omega Y(\omega)\mathbb{P}(d\omega|X = x_k),$$

sofern Y bezüglich $\mathbb{P}(\cdot|X = x_k)$ integrierbar ist.

Wir wollen diese Definition auf Zufallsvariablen mit kontinuierlichem Bild erweitern. Ist $\mathbb{P}(X \in B) > 0$, so ist

$$\mathbb{P}(A|X \in B) = \frac{\mathbb{P}(A \cap \{X \in B\})}{\mathbb{P}(X \in B)}$$

wohldefiniert. Für $B = \{x\}$ mit $x \in \mathbb{R}^d$ ist jedoch in vielen Fällen $\mathbb{P}(X = x) = 0$. Die Idee ist

$$\mathbb{P}(A|X = x) = \lim_{h\to 0} \frac{\mathbb{P}(A \cap \{X \in (x-h, x+h)\})}{\mathbb{P}(X \in (x-h, x+h))}.$$

Problem: Der Limes muss nicht existieren!
Lösung: Wir benutzen den Satz von Radon-Nikodym.

Definition 2.21. *Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra und $X \in \mathcal{L}^1$. Der bedingte Erwartungswert von X gegeben $\mathcal{A}$ wird mit $\mathbb{E}(X|\mathcal{A})$ bezeichnet. Dieser ist definiert über die folgenden Eigenschaften*

- $\mathbb{E}(X|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$.
- *Für alle $A \in \mathcal{A}$ gilt*

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X|\mathcal{A})) \tag{2.4}$$

Kurz: Der bedingte Erwartungswert ist die bezüglich $\mathcal{A}$ messbare Zufallsvariable mit (2.4).

Satz 2.22. *(Existenz und Eindeutigkeit vom bedingten Erwartungswert) Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra und $X \in \mathcal{L}^1$. Dann gibt es den bedingten Erwartungswert $\mathbb{E}(X|\mathcal{A})$. Dieser ist eindeutig in dem folgenden Sinn:*

Ist $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ gegeben mit

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot g), \quad A \in \mathcal{A},$$

so gilt $g = \mathbb{E}(X|\mathcal{A})$ fast sicher.

Beweis. Wir zeigen zunächst die Existenz.
Fall 1: Sei $X \geq 0$ fast sicher. Dann definiert

$$\mathcal{A} \ni A \longmapsto \int_A X d\mathbb{P} = \mathbb{E}(\mathbb{1}_A \cdot X) \geq 0$$

ein Maß auf $\mathcal{A}$. Ist A derart, dass $\mathbb{P}(A) = 0$. So gilt auch $\mathbb{E}(\mathbb{1}_A \cdot X) = 0$. Der Satz von Radon-Nikodym liefert die Existenz einer Zufallsvariable $\mathbb{E}(X|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ mit

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot g), \quad A \in \mathcal{A}.$$

Dieses zeigt die Existenz in diesem Fall.
Fall 2: Sei $X \in \mathcal{L}^1$ beliebig. Betrachte die Zerlegung $X = X^+ - X^-$ in Positiv- und Negativteil. Dann sind $X^\pm \in \mathcal{L}^1$ nicht-negativ. Aus Fall 1 folgt die Existenz vom bedingten Erwartungswert $\mathbb{E}(X^\pm|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Setze $\mathbb{E}(X|\mathcal{A}) := \mathbb{E}(X^+|\mathcal{A}) - \mathbb{E}(X^-|\mathcal{A})$. Dann gilt $\mathbb{E}(X|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ und für alle $A \in \mathcal{A}$

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot X^+) - \mathbb{E}(\mathbb{1}_A \cdot X^-) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X^+|\mathcal{A})) - \mathbb{E}(\mathbb{1}_A \cdot \mathbb{E}(X^-|\mathcal{A})) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X|\mathcal{A})).$$

Dieses zeigt die Existenz im allgemeinen Fall. Wir zeigen die Eindeutigkeit. Seien dazu $g_1, g_2 \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ gegeben mit

$$\mathbb{E}(\mathbb{1}_A \cdot g_1) = \mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot g_2), \quad A \in \mathcal{A}. \tag{2.5}$$

Dann ist $A_1 := \{g_1 - g_2 > 0\} \in \mathcal{A}$ und $A_2 := \{g_2 - g_1 > 0\} \in \mathcal{A}$. Aus (2.5) folgt $\mathbb{E}(\mathbb{1}_{A_j} \cdot g_1) = \mathbb{E}(\mathbb{1}_{A_j} \cdot g_2)$ mit $j = 1, 2$. Daraus folgt

$$\mathbb{E}(\mathbb{1}_{A_1}(g_1 - g_2)) = 0 = \mathbb{E}(\mathbb{1}_{A_2}(g_2 - g_1)).$$

Da die Integranden nicht-negativ sind, sind diese fast sicher Null. Wegen der Wahl der Mengen A_j ist $\mathbb{1}_{A_1} = 0 = \mathbb{1}_{A_2}$ fast sicher, d.h. $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 0$. Insbesondere ist

$$\mathbb{P}(g_1 \neq g_2) = \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = 0.$$

□

Bemerkung 2.23. *Die Eindeutigkeit folgt aus dem Satz von Radon-Nikodym und hätte daher nicht gesondert gezeigt werden müssen.*

Das nächste Lemma sammelt einige wichtige Eigenschaften des bedingten Erwartungswertes.

Lemma 2.24. *Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra und $X \in \mathcal{L}^1$. Es gelten die folgenden Eigenschaften:*

(a) Ist X messbar bezüglich $\mathcal{A}$, so gilt $\mathbb{E}(X|\mathcal{A}) = X$.

(b) Ist X unabhängig von $\mathcal{A}$, so gilt $\mathbb{E}(X|\mathcal{A}) = \mathbb{E}(X)$.

(c) Es gilt $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}(X)$.

(d) Für $A \in \mathcal{F}$ mit $\mathbb{P}(A) \in (0,1)$ sei $\mathcal{G} = \{A, A^c, \emptyset, \Omega\}$. Dann gilt

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X|A)\mathbb{1}_A + \mathbb{E}(X|A^c)\mathbb{1}_{A^c}.$$

(e) Ist $Y \in \mathcal{L}^1$ und $a, b \in \mathbb{R}$ so gilt

$$\mathbb{E}(aX + bY|\mathcal{A}) = a\mathbb{E}(X|\mathcal{A}) + b\mathbb{E}(Y|\mathcal{A}).$$

(f) Sei $\mathcal{G} \subset \mathcal{A}$ eine weitere Teil-σ-Algebra. Dann gilt

$$\mathbb{E}(\mathbb{E}(X|\mathcal{A})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}).$$

(g) Es gilt $\mathbb{E}(\mathbb{E}(X|\mathcal{A})) = \mathbb{E}(X)$.

(h) Sei $Y \in \mathcal{L}^1$ mit $Y \leq X$. Dann gilt $\mathbb{E}(Y|\mathcal{A}) \leq \mathbb{E}(X|\mathcal{A})$.

(i) Es gilt $|\mathbb{E}(X|\mathcal{A})| \leq \mathbb{E}(|X||\mathcal{A})$.

Beweis. Übung. □

Wir schreiben $\mathbb{E}(X|Y)$ für $\mathbb{E}(X|\sigma(Y))$ für zwei Zufallsvariablen X, Y.

Satz 2.25. *Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra. Dann gelten die folgenden Aussagen.*

(a) Sei $0 \leq X_n \in \mathcal{L}^1$ und $0 \leq X \in \mathcal{L}^1$. Es gelte $X_n \nearrow X$ fast sicher für $n \to \infty$. Dann gilt auch

$$\mathbb{E}(X_n|\mathcal{A}) \nearrow \mathbb{E}(X|\mathcal{A}), \quad n \to \infty$$

fast sicher.

(b) Sei $X_n \in \mathcal{L}^1$ und $X \in \mathcal{L}^1$. Sei $\varphi \in \mathcal{L}^1$ mit $|X_n| \leq \varphi$ für alle $n \geq 1$ und es gelte $X_n \longrightarrow X$ fast sicher. Dann gilt

$$\mathbb{E}(X_n|\mathcal{A}) \longrightarrow \mathbb{E}(X|\mathcal{A}), \quad n \to \infty$$

fast sicher.

(c) Seien X, Y Zufallsvariablen mit $X \cdot Y, Y \in \mathcal{L}^1$. Ferner sei X messbar bezüglich $\mathcal{A}$. Dann gilt

$$\mathbb{E}(X \cdot Y|\mathcal{A}) = X\mathbb{E}(Y|\mathcal{A})$$

fast sicher.

Beweis. (a) Aus der Monotonie des bedingten Erwartungswertes folgt $0 \leq \mathbb{E}(X_n|\mathcal{A}) \leq \mathbb{E}(X_{n+1}|\mathcal{A})$. Es sei $Y := \lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{A}) = \sup_{n\in\mathbb{N}} \mathbb{E}(X_n|\mathcal{A}) \in \mathbb{R} \cup \{\pm\infty\}$. Dann ist Y messbar bezüglich $\mathcal{A}$. Für jedes $A \in \mathcal{A}$ gilt

$$\mathbb{E}(\mathbb{1}_A \cdot X_n) = \mathbb{E}(\mathbb{1}_A \cdot \mathbb{E}(X_n|\mathcal{A})), \quad n \geq 1.$$

Aus dem Satz der monotonen Konvergenz folgt

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot Y), \quad A \in \mathcal{A}.$$

(b) Wir betrachten zuerst den Fall $X_n \geq 0$ fast sicher. In diesem Fall ist auch $X \geq 0$ fast sicher. Angenommen X_n ist monoton. Wir betrachten den Fall, dass X_n monoton steigend ist. Der Fall einer monoton fallenden Folge geht analog. Ist X_n monoton steigend, so folgt aus (a) $\mathbb{E}(X_n|\mathcal{A}) \nearrow \mathbb{E}(X|\mathcal{A})$ für $n \to \infty$. Ist X_n nicht monoton, so betrachten wir $X_n' := \inf_{m\geq n} X_m$ und $X_n'' = \sup_{m\geq n} X_m$. Dann gilt $X_n', X_n'' \leq \varphi$. Diese Folgen sind monoton mit $X_n' \nearrow \liminf_{n\to\infty} X_n = X$ und $X_n'' \searrow \limsup_{n\to\infty} X_n = X$ fast sicher. Aus dem ersten Fall folgt

$$\lim_{n\to\infty} \mathbb{E}(X_n'|\mathcal{A}) = \mathbb{E}(X|\mathcal{A}) = \lim_{n\to\infty} \mathbb{E}(X_n''|\mathcal{A}).$$

Wegen $X_n' \leq X_n \leq X_n''$ folgt in diesem Fall die Behauptung aus

$$\mathbb{E}(X_n'|\mathcal{A}) \leq \mathbb{E}(X_n|\mathcal{A}) \leq \mathbb{E}(X_n''|\mathcal{A}), \quad n \geq 1.$$

Im allgemeinen Fall, sei $X_n = X_n^+ - X_n^-$ sowie $X = X^+ - X^-$ die Zerlegung in Positiv- und Negativteil. Dann folgt

$$\mathbb{E}(X_n|\mathcal{A}) = \mathbb{E}(X_n^+|\mathcal{A}) - \mathbb{E}(X_n^-|\mathcal{A}) \longrightarrow \mathbb{E}(X^+|\mathcal{A}) - \mathbb{E}(X^-|\mathcal{A}) = \mathbb{E}(X|\mathcal{A})$$

fast sicher.
(c) Fall 1. Sei $X = \mathbb{1}_B$ mit $B \in \mathcal{A}$. Dann gilt für jede Menge $A \in \mathcal{A}$

$$\mathbb{E}(\mathbb{1}_A X \cdot Y) = \mathbb{E}(\mathbb{1}_{A\cap B} Y) = \mathbb{E}(\mathbb{1}_{A\cap B}\mathbb{E}(Y|\mathcal{A})) = \mathbb{E}(\mathbb{1}_A X \mathbb{E}(Y|\mathcal{A})).$$

Fall 2. Aufgrund der Linearität obiger Gleichung golt die Behauptung für alle Elementarfunktionen $X = \sum\limits_{n=1}^{N} a_n \mathbb{1}_{B_n}$ mit $a_n \in \mathbb{R}$ und $B_n \in \mathcal{A}$.
Fall 3. Seien $X, Y \geq 0$. Dann gibt es eine Folge von Elementarfunktionen X_n mit $X_n \nearrow X$. Dann gilt $X_n \cdot Y \nearrow X \cdot Y$ und wir erhalten aus (a)

$$\mathbb{E}(X \cdot Y|\mathcal{A}) = \lim_{n\to\infty} \mathbb{E}(X_n \cdot Y|\mathcal{A}) = \lim_{n\to\infty} X_n \mathbb{E}(Y|\mathcal{A}) = X\mathbb{E}(Y|\mathcal{A}).$$

Fall 4. Seien X, Y wie in der Behauptung. Durch Zerlegung in Positiv- und Negativteil erhalten wir

$$\mathbb{E}(X \cdot Y|\mathcal{A}) = X\mathbb{E}(Y|\mathcal{A}).$$

□

Satz 2.26. *(Jensen Ungleichung für bedingte Erwartungswerte, [Bil99]) Sei $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ konvex und X eine Zufallsvariable mit $X, \varphi \circ X \in \mathcal{L}^1$. Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra. Dann gilt*

$$\varphi(\mathbb{E}(X|\mathcal{A})) \leq \mathbb{E}(\varphi \circ X|\mathcal{A}).$$

Definition 2.27. *Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra, $(E, \mathcal{E})$ ein Maßraum und X eine E-wertige Zufallsvariable. Eine Abbildung $Q : \Omega \times \mathcal{E} \longrightarrow [0,1]$ heißt regulärer bedingter Erwartungswert von X gegeben $\mathcal{A}$, wenn die folgenden Eigenschaften gelten:*

(a) Für jedes $\omega \in \Omega$ ist $Q(\omega, \cdot) : \mathcal{E} \longrightarrow [0,1]$ ein Wahrscheinlichkeitsmaß auf E.

(b) Für jedes $B \in \mathcal{E}$ ist $Q(\cdot, B) : \Omega \longrightarrow [0,1]$ messbar bezüglich $\mathcal{A}$.

(c) Für jedes $B \in \mathcal{E}$ gilt $\mathbb{E}(\mathbb{1}_{X\in B}|\mathcal{A}) = Q(\cdot, B)$ fast sicher.

Satz 2.28. *[KS07] Sei E ein Polnischer Raum. Ferner sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra. Für jede E-wertige Zufallsvariable X gibt es einen regulären bedingten Erwartungswert von X gegeben $\mathcal{A}$. Dieser ist eindeutig in dem folgenden Sinn:*

Gegeben zwei reguläre bedingte Erwartungswerte Q, Q'. Dann ist gibt es eine Menge $M \in \mathcal{F}$ mit $\mathbb{P}(M) = 1$ derart, dass

$$Q(\omega, B) = Q'(\omega, B), \quad B \in \mathcal{E},\ \omega \in M.$$

Satz 2.29. *Sei E ein Polnischer Raum. Ferner sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra. Sei X eine E-wertige Zufallsvariable und $Q(\omega, dx)$ der reguläre bedingte Erwartungswert. Sei $\varphi : E \longrightarrow \mathbb{R}$ messbar mit $\varphi \circ X \in \mathcal{L}^1$. Dann gilt*

$$\int_E |\varphi(x)| Q(\omega, dx) < \infty \quad \textit{fast sicher}$$

und

$$\mathbb{E}(\varphi(X)|\mathcal{A})(\omega) = \int\limits_E \varphi(x)Q(\omega, dx) \tag{2.6}$$

für fast alle ω.

Beweis. Fall 1. Sei $\varphi = \mathbb{1}_B$ mit $B \in \mathcal{B}(E)$. Dann ist fast sicher

$$\mathbb{E}(\varphi(X)|\mathcal{A})(\omega) = \mathbb{E}(\mathbb{1}_{X\in B}|\mathcal{A})(\omega) = Q(\omega, B).$$

Fall 2. Da (2.6) linear in φ ist, gilt die Behauptung fdirur alle Elementarfunktionen.
Fall 3. Sei $\varphi \geq 0$. Dann gibt es eine Folge von Elementarfunktionen φ_n mit $\varphi_n \nearrow \varphi$. Dann ist

$$\int\limits_E \varphi_n(x)Q(\omega, dx) = \mathbb{E}(\varphi_n(X)|\mathcal{A}) \nearrow \mathbb{E}(\varphi(X)|\mathcal{A})$$

fast sicher. Aus dem Sart der monotonen Konvergenz folgt

$$\int\limits_E \varphi_n(x)Q(\omega, dx) \nearrow \int\limits_E \varphi(x)Q(\omega, dx) = \mathbb{E}(\varphi(X)|\mathcal{A}) < \infty.$$

Fall 4. Zerlege in Positiv- und Negativteil und benutze Fall 3. □

Beispiel 2.30. *Es sei $E = \mathbb{R}$ und $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra. Sei $X \in \mathcal{L}^1$ und $\varphi(x) = x$. Sei $Q(\omega, dx)$ der reguläre bedingte Erwartungswert. Dann gilt für fast alle $\omega \in \Omega$*

$$\int\limits_{\mathbb{R}} |x|Q(\omega, dx) < \infty$$

und

$$\mathbb{E}(X|\mathcal{A})(\omega) = \int\limits_{\mathbb{R}} xQ(\omega, dx).$$

3 Schwache Konvergenz

Sei (E, d) ein metrischer Raum und sei $\mathcal{P}(E)$ der Raum der Wahrscheinlichkeitsmaße auf E. Im ersten Teil untersuchen wir die Konvergenz von Wahrscheinlichkeitsmaßen.
Idee: Mengenweise Konvergenz, d.h.

$$\mu_n(A) \longrightarrow \mu(A), \quad A \in \mathcal{E}, \ n \to \infty. \tag{3.1}$$

Problem: Für viele praktische Anwendungen ist diese Art der Konvergenz zu stark. Dazu betrachten wir die Folge $\mu_n(dx) = n\mathbb{1}_{[0,\frac{1}{n}]}(x)m(dx)$ und $A = (0, 1]$. Dann gilt $\mu_n(A) = 1$. Es liegt nahe anzunehmen, dass $\mu_n \longrightarrow \delta_0$ konvergiert. Es gilt jedoch $\delta_0(A) = 0$. Eine mengenweise Konvergenz liegt in diesem Beispiel nicht vor.

Wir brauchen daher eine abgeschwächte Form von (3.1). Die Idee ist es Konvergenz über die Konvergenz der dazugehörigen Momente bzw. Erwartungswerte zu beschreiben. Genauer sei $M(E)$ eine geeignete Familie von messbaren Funktionen auf E. Die Konvergenz $\mu_n \longrightarrow \mu$ wird daher definiert über

$$\int_E f(x)d\mu_n(x) \longrightarrow \int_E f(x)d\mu(x), \quad n \to \infty \tag{3.2}$$

für alle $f \in M(E)$.

Bemerkung 3.1. *Es sei $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Genau dann gilt* (3.1), *wenn* (3.2) *für alle beschränkten messbaren Funktionen $f : E \longrightarrow \mathbb{R}$ gilt.*

Beweis. Gilt (3.2), so wählen wir $f = \mathbb{1}_A$ für $A \in \mathcal{E}$. Umgekehrt gelte (3.1). Dann gilt (3.2) für alle Indikatiorfunktionen $f = \mathbb{1}_A$ und mittels Linearität auch für alle Elementarfunktionen. Es reicht (3.2) für alle $f \geq 0$ zu zeigen. Sei dazu $f \geq 0$ beschränkt und messbar und $\varepsilon > 0$ beliebig. Dann gibt es eine Elementarfunktion g mit $0 \leq g \leq f$ und $\sup_{x\in E}(f(x) - g(x)) < \varepsilon$. Daraus folgt

$$\begin{aligned}
&\left|\int_E f(x)d\mu(x) - \int_E f(x)d\mu_n(x)\right| \leq \left|\int_E f(x)d\mu(x) - \int_E g(x)d\mu(x)\right| \\
&+ \left|\int_E g(x)d\mu(x) - \int_E g(x)d\mu_n(x)\right| + \left|\int_E g(x)d\mu_n(x) - \int_E f(x)d\mu_n(x)\right| \\
&\leq 2\sup_{x\in E}|f(x) - g(x)| + \left|\int_E g(x)d\mu(x) - \int_E g(x)d\mu_n(x)\right| \\
&\leq 2\varepsilon + \left|\int_E g(x)d\mu(x) - \int_E g(x)d\mu_n(x)\right|
\end{aligned}$$

welches mit $n \to \infty$ die Behauptung zeigt. □

Es liegt daher nahe die Menge $M(E)$ geeignet einzuschränken. In diesem Kapitel betrachten wir den Fall $M(E) = C_b(E)$, wo $C_b(E)$ der Raum der stetigen, beschränkten Funktionen $f : E \longrightarrow \mathbb{R}$ ist. Dieser ist ein Banachraum mit der Supremumsnorm

$$\|f\|_\infty := \sup_{x\in E}|f(x)|.$$

Der Zusammenhang zu Zufallsvariablen ist dann gegeben durch $\mu_n := \mathbb{P}\circ X_n^{-1}$ und $\mu := \mathbb{P}\circ X^{-1}$. Dieses führt zu dem Begriff der Konvergenz in Verteilung.

3.1 Konvergenz von Wahrscheinlichkeitsmaßen

Sei im Folgenden (E, d) ein metrischer Raum.

Definition 3.2. *Eine Folge $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ konvergiert schwach gegen $\mu \in \mathcal{P}(E)$, falls für alle $f \in C_b(E)$*

$$\int_E f(x)d\mu_n(x) \longrightarrow \int_E f(x)d\mu(x), \quad n \to \infty.$$

Das nächste Lemma zeigt, dass der Grenzwert bei schwacher Konvergenz eindeutig ist.

Lemma 3.3. *Sei $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$. Angenommen es gibt $\mu, \nu \in \mathcal{P}(E)$ derart, dass μ_n schwach gegen μ konvergiert. Dann ist $\mu = \nu$.*

Für den Beweis ist die folgende Funktion hilfreich. Sei $A \subset E$. Der Abstand von $x \in E$ zu A ist definiert über

$$d(x, A) := \inf\{d(x, a) \mid a \in A\}.$$

Es gilt

$$|d(x, A) - d(y, A)| \le d(x, y), \quad x, y \in E$$

und $d(x, A) = 0$ genau dann, wenn $x \in \overline{A}$. Hierbei bezeichnet $\overline{A}$ den Abschluss von A.

Beweis. Aus der schwachen Konvergenz folgt

$$\int_E f(x)d\mu(x) = \int_E f(x)d\nu(x), \quad f \in C_b(E).$$

Es reicht zu zeigen, dass $\mu(K) = \nu(K)$ für alle abgeschlossenen Mengen $K \subset E$ gilt. Es sei $K \subset E$ abgeschlossen. Definiere $\varphi(t) := \begin{cases} 1-t, & t \in [0,1] \\ 0, & t \ge 1 \end{cases}$ und sei

$$f_n(x) := \varphi(nd(x, K)), \quad x \in E.$$

Dann ist $0 \le f_n \le 1$ und f_n ist stetig für jedes $n \ge 1$. Ferner gilt $f_n(x) \longrightarrow \mathbb{1}_K(x)$ für $n \to \infty$ und jedes $x \in E$. Damit folgt

$$\mu(K) = \lim_{n\to\infty} \int_E f_n(x)d\mu(x) = \lim_{n\to\infty} \int_E f_n(x)d\nu(x) = \nu(K).$$

□

Beispiel 3.4. *(a) Sei $(x_n)_{n\in\mathbb{N}} \subset E$ und $x \in E$. Falls $d(x_n, x) \longrightarrow 0$ für $n \to \infty$ gilt, so konvergiert δ_{x_n} schwach gegen δ_x.*

(b) Sei $\mu_n(dx) = p_n(x)m(dx)$ mit

$$p_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{n}{2}x^2}, \quad x \in \mathbb{R}.$$

Dann konvergiert μ_n schwach gegen δ_0.

(c) Sei $\mu_n(dx) = n\mathbb{1}_{[0,\frac{1}{n}]}(x)m(dx)$. Dann konvergiert μ_n schwach gegen δ_0.

Beweis. Übung. $\square$

Satz 3.5. *(Portmanteau-Theorem) Sei (E, d) ein metrischer Raum mit Borel-σ-Algebra $\mathcal{B}(E)$. Sei $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Dann sind die folgenden Aussagen äquivalent.*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) Für alle gleichmässig stetigen Funktionen $f \in C_b(E)$ gilt

$$\int_E f(x)d\mu_n(x) \longrightarrow \int_E f(x)d\mu(x), \quad n \to \infty.$$

(c) Für alle abgeschlossenen Mengen $F \subset E$ gilt

$$\limsup_{n\to\infty} \mu_n(F) \le \mu(F).$$

(d) Für alle offenen Mengen $O \subset E$ gilt

$$\liminf_{n\to\infty} \mu_n(O) \ge \mu(O).$$

(e) Für alle Mengen $A \in \mathcal{B}(E)$ mit $\mu(\partial A) = 0$ gilt

$$\lim_{n\to\infty} \mu_n(A) = \mu(A).$$

Bevor wir den Beweis führen brauchen wir das folgende Lemma.

Lemma 3.6. *Sei $g : E \longrightarrow \mathbb{R}$ eine stetige Funktion und $a \in \mathbb{R}$. Dann gilt*

$$\partial\{x \in E \mid g(x) \le a\} \subset \{x \in E \mid g(x) = a\}.$$

Beweis. Sei $x \in \partial\{x \in R \mid g(x) \le a\}$. Dann gilt für alle $\varepsilon > 0$

$$U_\varepsilon(x) \cap \{y \in E \mid g(y) > a\} \neq \emptyset$$

sowie

$$U_\varepsilon(x) \cap \{y \in E \mid g(y) \le a\} \neq \emptyset.$$

Hierbei bezeichnet $U_\varepsilon(x) := \{y \in E \mid d(x,y) < \varepsilon\}$ den offenen Ball mit Radius $\varepsilon > 0$ und Mittelpunkt x. Für $\varepsilon = \frac{1}{n}$ mit $n \in \mathbb{N}$ erhalten wir zwei Folgen $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ mit $d(x, x_n) < \frac{1}{n}$ sowie $d(x, y_n) < \frac{1}{n}$ derart, dass $g(x_n) > a$ und $g(y_n) \le a$. Weiterhin gilt $x_n, y_n \longrightarrow x$ und somit

$$a \le \lim_{n\to\infty} g(x_n) = g(x) = \lim_{n\to\infty} g(y_n) \le a.$$

$\square$

Wir kommen zum Beweis von Satz 3.5.

Beweis. $(c) \iff (d)$: klar durch Komplementbildung.
$(a) \implies (b)$: trivial.
$(b) \implies (c)$: Sei $F \subset E$ abgeschlossen und setze

$$G_m := \left\{ x \in E \mid d(x,F) < \frac{1}{m} \right\}.$$

Dann ist G_m offen, $G_{m+1} \subset G_m$ und $F = \bigcap_{m \geq 1} G_m$. Folglich gilt $\mu(G_m) \longrightarrow \mu(F)$ für $m \to \infty$. Sei

$$\varphi(t) := \begin{cases} 1-t, & t \in [0,1] \\ 0, & t \geq 1 \end{cases}$$

und setze $f_m(x) := \varphi(md(x,F))$. Dann gilt $0 \leq f_m \leq 1$, $f_m|_{G_m^c} = 0$, $f_m|_F = 1$ und f_m ist Lipschitz stetig. Insbesondere ist f_m gleichmässig stetig mit $\mathbb{1}_F \leq f_m \leq \mathbb{1}_{G_m}$. Daraus folgt

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \int_E f_m d\mu_n = \int_E f_m d\mu \leq \mu(G_m) \longrightarrow \mu(F).$$

$(c) \implies (e)$: Sei $A \in \mathcal{B}(E)$ mit $\mu(\partial A) = 0$. Dann gilt

$$\begin{aligned} \mu(A) = \mu(\mathring{A}) &\leq \liminf_{n\to\infty} \mu_n(\mathring{A}) \leq \liminf_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(A) \\ &\leq \limsup_{n\to\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}) = \mu(A). \end{aligned}$$

$(e) \implies (c)$: Sei $F \subset E$ abgeschlossen. Auf Lemma 3.6 folgt für alle $\delta > 0$

$$\partial\{x \in E \mid d(x,F) \leq \delta\} \subset \{x \in E \mid d(x,F) = \delta\}. \tag{3.3}$$

Da $\{x \in E \mid d(x,F) = \delta\}$ für alle $\delta > 0$ eine abgeschlossene Menge ist, liegt sie in $\mathcal{B}(E)$. Weiterhin gilt

$$\{x \in E \mid d(x,F) = \delta\} \cap \{x \in E \mid d(x,F) = \delta'\} = \varnothing, \quad \delta \neq \delta'. \tag{3.4}$$

Für $n \in \mathbb{N}$ sei

$$D_n := \left\{ \delta > 0 \mid \mu\left(\left\{ x \in E \mid d(x,F) \geq \frac{1}{n} \right\} \right) \right\}.$$

Behauptung: D_n enthält für jedes $n \in \mathbb{N}$ nur endlich viele Elemente.

Beweis. Angenommen nicht. Dann gibt es ein $n \in \mathbb{N}$ und eine Folge paarweise verschiedener Elemente $(\delta_k)_{k\in\mathbb{N}} \subset D_n$. Dann folgt aus (3.4)

$$1 \geq \mu\left(\bigcup_{k\geq 1} \{x \in E \mid d(x,F) = \delta_k\} \right) = \sum_{k=1}^{\infty} \mu(\{x \in E \mid d(x,F) = \delta_k\}) \geq \sum_{k=1}^{\infty} \frac{1}{n} = \infty.$$

□

Also ist $D = \bigcup_{n\geq 1} D_n$ abzählbar. Folglich gibt es eine Nullfolge $(\delta_k)_{k\in\mathbb{N}} \subset [0,\infty)\backslash D$. Aus (3.3) folgt

$$\mu(\partial F_k) \leq \mu(\{x \in E \mid d(x,F) = \delta_k\}) = 0$$

für $F_k := \{x \in E \mid d(x,F) \leq \delta_k\}$ und alle $k \in \mathbb{N}$. Wegen $F_k \searrow \{x \in E \mid d(x,F) = 0\} = F$ folgt

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F_k) = \mu(F_k) \longrightarrow \mu(F), \quad k \to \infty.$$

$(c) \Longrightarrow (a)$: Sei $f \in C_b(E)$. Es reicht zu zeigen, dass

$$\limsup_{n\to\infty} \int_E f(x) d\mu_n(x) \leq \int_E f(x) d\mu(x).$$

Denn, daraus folgt für $(-f)$ anstelle von f

$$-\liminf_{n\to\infty} \int_E f(x) d\mu_n(x) = \limsup_{n\to\infty} (-f(x)) d\mu_n(x) \leq \int_E (-f(x)) d\mu(x)$$

und somit

$$\liminf_{n\to\infty} \int_E f(x) d\mu_n(x) \geq \int_E f(x) d\mu(x).$$

Für $f = 0$ ist nichts zu zeigen. Sei also $f \neq 0$. Wir können ohne Einschränkung annehmen, dass $0 \leq f \leq 1$. Anderenfalls betrachten wir $\frac{f+\|f\|_\infty}{2\|f\|_\infty}$. Sei $k \in \mathbb{N}$ fest und für $i \in \mathbb{N}$ definiere

$$F_i := \left\{ f \geq \frac{i}{k} \right\}.$$

Dann ist F_i abgeschlossen, $F_{i+1} \subset F_i$ und es gilt

$$F_i \backslash F_{i+1} = \left\{ \frac{i}{k} \leq f < \frac{i+1}{k} \right\}. \tag{3.5}$$

Lemma 3.7. *Es gilt*

$$\frac{1}{k} \sum_{i=1}^{k} \mathbb{1}_{F_i} = \sum_{i=1}^{k} \frac{i}{k} \mathbb{1}_{F_i \backslash F_{i+1}} \leq f \leq \sum_{i=0}^{k} \frac{i+1}{k} \mathbb{1}_{F_i \backslash F_{i+1}} = \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \mathbb{1}_{F_i}.$$

Beweis. Es gilt $E = \bigcup_{i=0}^{k} F_i \backslash F_{i+1}$ mit $F_0 = E$ und $F_{k+1} = \emptyset$, wobei die Vereinigung disjunkt ist. Daraus, und aus (3.5) ergibt sich die mittlere Ungleichung. Für die beiden Identitäten beachte

$$\begin{aligned}
\sum_{i=0}^{k} (i+1) \mathbb{1}_{F_i \backslash F_{i+1}} &= \sum_{i=0}^{k} (i+1) \mathbb{1}_{F_i} - \sum_{i=0}^{k-1} (i+1) \mathbb{1}_{F_{i+1}} \\
&= \sum_{i=0}^{k} \mathbb{1}_{F_i} + \sum_{i=0}^{k} i \mathbb{1}_{F_i} - \sum_{i=1}^{k} i \mathbb{1}_{F_i} = 1 + \sum_{i=1}^{k} \mathbb{1}_{F_i}.
\end{aligned}$$

Dieses impliziert die rechte Identität. Für die linke Identität beachte $1 = \sum_{i=0}^{k} \mathbb{1}_{F_i \setminus F_{i+1}}$ woraus mit obiger Rechnung $\sum_{i=1}^{k} i \mathbb{1}_{F_i \setminus F_{i+1}} = \sum_{i=1}^{k} \mathbb{1}_{F_i}$ folgt. $\square$

Die Behauptung folgt aus

$$\begin{aligned}\limsup_{n\to\infty} \int_E f d\mu_n - \frac{1}{k} &\leq \frac{1}{k} \limsup_{n\to\infty} \sum_{i=1}^{k} \mu_n(F_i) \\ &\leq \frac{1}{k} \sum_{i=1}^{k} \limsup_{n\to\infty} \mu_n(F_i) \leq \frac{1}{k} \sum_{i=1}^{k} \mu(F_i) \leq \int_E f d\mu.\end{aligned}$$

$\square$

Satz 3.8. *Seien (E,d) und (E',d') metrische Räume, $\mu_n, \mu \in \mathcal{P}(E)$ und $h : E \longrightarrow E'$ stetig. Angenommen $\mu_n \longrightarrow \mu$ schwach, dann ist $\mu_n \circ h^{-1} \longrightarrow \mu \circ h^{-1}$ schwach auf E'.*

Beweis. Sei $f \in C_b(E')$. Dann ist $f \circ h \in C_b(E)$. $\square$

Beispiel 3.9. *Auf die Voraussetzung, dass h stetig ist kann im Allgemeinen nicht verzichtet werden. Denn sei $(x_n)_{n\in\mathbb{N}} \subset E$ mit $x_n \longrightarrow x$ und $x_n \neq x$. Dann gilt $\delta_{x_n} \longrightarrow \delta_x$ schwach. Wähle $h(y) = \mathbb{1}_{\{x\}}(y)$, dann gilt*

$$\delta_{x_n} \circ h^{-1}(A) = \delta_{x_n}(h^{-1}(A)) = \begin{cases} 1, & h(x_n) \in A \\ 0, & h(x_n) \notin A \end{cases} = \delta_{h(x_n)}(A) = \delta_0(A).$$

Analog zeigen wir $\delta_x \circ h^{-1} = \delta_{h(x)} = \delta_1$.

Wir müssen daher eine Bedingung an die Stetigkeitsstellen von h stellen.

Satz 3.10. *Seien (E,d), (E',d') metrische Räume, $h : E \longrightarrow E'$ messbar und $\mu_n, \mu \in \mathcal{P}(E)$ mit $\mu_n \longrightarrow \mu$ schwach. Sei*

$$D_h := \{x \in E \mid h \text{ ist nicht stetig in } x\}.$$

Gilt $\mu(D_h) = 0$, so folgt $\mu_n \circ h^{-1} \longrightarrow \mu \circ h^{-1}$ schwach.

Beweis. Wegen

$$D_h := \bigcup_{n\geq 1} \bigcap_{m\geq 1} \left\{ x \in E \mid \exists y \in E : \; d(x,y) < \frac{1}{m} \; d'(h(y), h(x)) \geq \frac{1}{n} \right\}$$

folgt $D_h \in \mathcal{B}(E)$. Sei $F \subset E'$ abgeschlossen. Dann gilt

$$\limsup_{n\to\infty} \mu_n(h^{-1}(F)) \leq \limsup_{n\to\infty} \mu_n(\overline{h^{-1}(F)}) \leq \mu(\overline{h^{-1}(F)}).$$

Sei $x \in \overline{h^{-1}(F)}$, dann gibt es eine Folge $(x_n)_{n\in\mathbb{N}} \subset h^{-1}(F)$ mit $x_n \longrightarrow x$. Ist h stetig in x, so folgt $h(x_n) \longrightarrow h(x)$ und da F abgeschlossen ist, folgt aus $h(x_n) \in F$ bereits $h(x) \in F$. In diesem Fall ist $x \in h^{-1}(F)$. Falls h nicht stetig in x ist, so ist $x \in D_h$ und wir erhalten $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$. Daraus folgt aus der schwachen Konvergenz $\mu_n \longrightarrow \mu$ und dem Satz von Portmanteau

$$\limsup_{n\to\infty} \mu_n(h^{-1}(F)) \leq \limsup_{n\to\infty} \mu_n(\overline{h^{-1}(F)}) \leq \mu(\overline{h^{-1}(F)}) \leq \mu(h^{-1}(F)) + \mu(D_h) = \mu(h^{-1}(F)).$$

Nach dem Satz von Portmanteau impliziert dieses die Behauptung. □

Satz 3.11. *[KS07, Theorem 8.5] Seien $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R})$ und $\mu \in \mathcal{P}(\mathbb{R})$ gegeben und bezeichne mit $F_n(t) = \mu_n((-\infty, t])$ und $F(t) = \mu((-\infty, t])$ die dazugehörigen Verteilungsfunktionen. Dann sind äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) $F_n(t) \longrightarrow F(t)$ für $n \to \infty$ und alle Stetigkeitsstellen t von F.

Als nächstes wollen wir Familien von Wahrscheinlichkeitsmaßen betrachten, welche die Bedingung der Folgenkompaktheit bezüglich der schwachen Konvergenz erfüllen.

Definition 3.12. *Sei $\Gamma \subset \mathcal{P}(E)$ eine Familie von Wahrscheinlichkeitsmaßen.*

(a) Γ heißt relativ kompakt, falls jede Folge in Γ eine schwach konvergente Teilfolge hat.

(b) Γ heißt straff, falls für alle $\varepsilon > 0$ es eine kompakte Menge $K_\varepsilon \subset E$ gibt mit

$$\mu(K_\varepsilon) \geq 1 - \varepsilon, \quad \forall \mu \in \Gamma.$$

Lemma 3.13. *Sei $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Dann sind äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) Für jede Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ von $(\mu_n)_{n\in\mathbb{N}}$ gibt es eine weitere Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ mit $\mu_{n_{k_l}} \longrightarrow \mu$ schwach für $l \to \infty$.

Beweis. $(a) \implies (b)$: Konvergiert eine Folge, so konvergiert auch jede Teilfolge gegen denselben Grenzwert.
$(b) \implies (a)$: Angenommen (a) gilt nicht. Dann gibt es $f \in C_b(E)$, $\varepsilon > 0$ und eine Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ mit

$$\left| \int_E f(x) d\mu_{n_k}(x) - \int_E f(x) d\mu(x) \right| \geq \varepsilon, \quad k \in \mathbb{N}. \tag{3.6}$$

Nach Voraussetzung gibt es eine weitere Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ mit $\mu_{n_{k_l}} \longrightarrow \mu$ schwach für $l \to \infty$. Für diese Teilfolge gilt

$$\int_E f(x) d\mu_{n_{k_l}}(x) \longrightarrow \int_E f(x) d\mu(x), \quad l \to \infty$$

welches im Wiederspruch zu (3.6) steht. □

Bemerkung 3.14. *(a) ISt E kompakt, so ist $\Gamma = \{\mu\}$ für jedes $\mu \in \mathcal{P}(E)$ straff. Dann sei $A_n \subset E$ eine Folge von abgeschlossen Mengen mit $A_n \nearrow E$. Dann ist jedes A_n als abgeschlossene Menge einer kompakten Menge wieder kompakt. Die Behauptung folgt aus $\mu(A_n) \nearrow \mu(E) = 1$ für $n \to \infty$.*

(b) Es gebe eine Folge von kompakten Mengen $(K_n)_{n\in\mathbb{N}} \subset E$ mit $\bigcup_{n\geq 1} K_n = E$. Dann ist $\bigcup_{n=1}^{N} K_n \nearrow E$, und wegen $\mu\left(\bigcup_{n=1}^{N} K_n\right) \nearrow \mu(E) = 1$ ist $\Gamma = \{\mu\}$ für jedes $\mu \in \mathcal{P}(E)$ straff.

(c) $\Gamma = \mathcal{P}(\mathbb{R}^d)$ ist nicht straff.

Der nächste Satz stellt den Zusammenhang zu Straffheit und relativer Kompaktheit her.

Satz 3.15. *(Satz von Prokhorov, [KS07, Theorem 8.9]) Sei E ein metrischer Raum. Ist $\Gamma \subset \mathcal{P}(E)$ straff, so ist Γ relativ kompakt. Ist E zusätzlich ein Polnischer Rau, so ist Γ genau dann straff, wenn es relativ kompakt ist.*

Im folgenden führen wir einige Resulate ohne Beweis auf. Hierbei ist E ein metrischer Raum.

Definition 3.16. *Es sei $\rho : \mathcal{P}(E) \times \mathcal{P}(E) \longrightarrow [0, \infty)$ gegeben durch*

$$\rho(\mu, \nu) := \inf\{\varepsilon > 0 \mid \mu(F) \leq \nu(F^\varepsilon) + \varepsilon, \quad F \text{ ist abgeschlossen } \}$$

mit $F^\varepsilon := \{y \in E \mid d(y, F) < \varepsilon\}$. Die Abbildung ρ wird Prokhorov Metrik genannt.

Lemma 3.17. *[EK86] Die Prokhorov Metrik ρ definiert eine Metrik auf E.*

Satz 3.18. *[EK86] Sei E ein seperabler metrischer Raum und $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$, $\mu \in \mathcal{P}(E)$. Dann sind die folgenden Aussagen äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) $\rho(\mu_n, \mu) \longrightarrow 0$ für $n \to \infty$.

Satz 3.19. *[EK86] Es gelten die folgenden Aussagen.*

(a) Ist E seperabel, so ist auch $\mathcal{P}(E)$ seperabel.

(b) Ist E vollständig, so ist auch $\mathcal{P}(E)$ volltändig.

Insbesondere, ist E ein Polnischer Raum, so ist auch $\mathcal{P}(E)$ ein Polnischer Raum.

Zum Schluß betrachten wir noch eine zusätzliche Eigenschaft von Wahrscheinlichkeitsmaßen auf metrischen Räumen.

Definition 3.20. *$\mu \in \mathcal{P}(E)$ heißt regulär, falls für jedes $A \in \mathcal{B}(E)$*

$$\mu(A) = \sup\{\mu(K) \mid K \subset A \text{ kompakt}\} = \inf\{\mu(U) \mid A \subset U \text{ offen }\}.$$

Eine äquivalente Definition ist gegeben durch: Für jedes $A \in \mathcal{B}(E)$ und jedes $\varepsilon > 0$ gibt es eine kompakte Menge K_ε und eine offene Menge U_ε mit $K_\varepsilon \subset A \subset U_\varepsilon$ und

$$\mu(U_\varepsilon \backslash K_\varepsilon) < \varepsilon. \tag{3.7}$$

Satz 3.21. *Sei E ein metrischer Raum. Dann gilt für jedes $\mu \in \mathcal{P}(E)$*

$$\mu(A) = \sup\{\mu(K) \mid K \subset A \textit{ abgeschlossen } \} = \inf\{\mu(U) \mid A \subset U \textit{ offen } \}.$$

Beweis. Es sei

$$\mathcal{K} = \{A \in \mathcal{B}(E) \mid \forall \varepsilon > 0 \exists U_\varepsilon \text{ offen}, \exists K_\varepsilon \text{ abgeschlossen mit } K_\varepsilon \subset A \subset U_\varepsilon \text{ und } (3.7)\}.$$

Dann enthält $\mathcal{K}$ alle abgeschlossenen Mengen. Denn ist A abgeschlossen, so wählen wir

$$U_n := \left\{x \in E \mid d(x, A) < \frac{1}{n}\right\}.$$

Dann ist $A \subset U_n$ und $A = \bigcap_{n \geq 1} U_n$. Folglich gilt $\lim_{n \to \infty} \mu(U_n) = \mu(A)$, also $A \in \mathcal{K}$.

Wir müssen zeigen, dass $\mathcal{K} = \mathcal{B}(E)$. Hierfür reicht es zu zeigen, dass $\mathcal{K}$ ein Dynkin-System ist.

- Sei $\varepsilon > 0, x_0 \in E$ fest und $U_\varepsilon = E$. Die Mengen der Form

  $$K_\alpha = \{x \in E \mid d(x, x_0) \leq \alpha\}$$

  sind abgeschlossen und es gilt $K_\alpha \nearrow E$. Folglich gilt $\mu(K_\alpha) \nearrow \mu(E) = 1$ für $\alpha \to \infty$. Also gibt es ein $\alpha(\varepsilon) > 0$ mit $\mu(K_{\alpha(\varepsilon)}) > 1 - \varepsilon$. Dann gilt $K_{\alpha(\varepsilon)} \subset A \subset U_\varepsilon$ und (3.7) für diese Mengen.

- Sei $A \in \mathcal{K}$ und $\varepsilon > 0$. Seien $K_\varepsilon, U_\varepsilon$ die dazugehörigen Mengen. Dann ist $K' := E \backslash K$ offen, $U' := E \backslash U$ abgeschlossen und es gilt $U' \subset E \backslash A \subset K'$ und

  $$\mu(K' \backslash U') = \mu(U \backslash K) < \varepsilon.$$

- Sei $(A_n)_{n \in \mathbb{N}} \subset \mathcal{K}$ eine Folge disjunkter Mengen. Sei $\varepsilon > 0$. Dann gilt

  $$0 \leq \sum_{n=1}^{\infty} \mu(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \mu(E) \leq 1,$$

  d.h. $(\mu(A_n))_{n \geq 1}$ ist summierbar. Folglich gibt es ein $n_0 \geq 2$ mit $\sum_{n=n_0}^{\infty} \mu(A_n) < \frac{\varepsilon}{2}$. Wähle abgeschlossene Mengen K_e^n und offene Mengen U_ε^n mit $K_\varepsilon^n \subset A_n \subset U_\varepsilon^n$ und

  $$\mu(U_\varepsilon^n \backslash K_\varepsilon^n) < \frac{\varepsilon}{2^{n+1}}, \quad n \in \mathbb{N}.$$

Dann ist $K_\varepsilon := \bigcup_{n=1}^{n_0-1} K_\varepsilon^n$ abgeschlossen und $U_\varepsilon := \bigcup_{n=1}^{\infty} U_\varepsilon^n$ offen. Ferner gilt

$$\mathcal{K}_\varepsilon \subset \bigcup_{n=1}^{\infty} K_\varepsilon^n \subset \bigcup_{n=1}^{\infty} A_n \subset U_\varepsilon$$

und

$$\begin{aligned}
\mu(U_\varepsilon \backslash K_\varepsilon) &\leq \sum_{n=1}^{\infty} \mu(U_\varepsilon^n \backslash K_\varepsilon) = \sum_{n=1}^{n_0-1} \mu(U_\varepsilon^n \backslash K_\varepsilon) + \sum_{n=n_0}^{\infty} \mu(U_\varepsilon^n \backslash K_\varepsilon) \\
&\leq \sum_{n=1}^{n_0-1} \mu(U_\varepsilon^n \backslash K_\varepsilon^n) + \sum_{n=n_0}^{\infty} \mu(U_\varepsilon^n) \leq \varepsilon \sum_{n=1}^{n_0-1} \frac{1}{2^{n+1}} + \sum_{n=n_0}^{\infty} \left(\frac{\varepsilon}{2^{n+1}} + \mu(K_\varepsilon^n) \right) \\
&\leq \frac{\varepsilon}{2} \sum_{n=1}^{\infty} \frac{1}{2^n} + \sum_{n=n_0}^{\infty} \mu(A_n) \leq \varepsilon.
\end{aligned}$$

Damit ist $\bigcup_{n=1}^{\infty} A_n \in \mathcal{K}$.

$\square$

Korollar 3.22. *Sei $\mu \in \mathcal{P}(E)$ derart, dass $\{\mu\}$ straff ist. Dann ist μ regulär.*

Für den nächsten Satz brauchen wir folgendes Lemma zur Charakterisierung von Kompaktheit in metrischen Räumen.

Lemma 3.23. *Sei (E, d) ein vollständiger metrischer Raum und $K \subset E$ abgeschlossen. Dann ist K genau dann kompakt, wenn es total beschränkt ist, d.h. für jedes $\varepsilon > 0$ gibt es endlich viele Punkte $x_1, \ldots, x_m \in E$ mit*

$$K \subset \bigcup_{k=1}^{m} B_\varepsilon(x_k), \tag{3.8}$$

wo $B_\varepsilon(x_k)$ den abgeschlossenen Ball um x_k mit Radius ε bezeichnet.

Beweis. Sei K kompakt und $\varepsilon > 0$. Dann ist

$$K \subset \bigcup_{x \in K} U_\varepsilon(x)$$

eine offene Überdeckung von K. Es gibt daher endlich viele Punkte $x_1, \ldots, x_m \in K$ mit (3.8).

Umgekehrt gelte (3.8) für jedes $\varepsilon > 0$. Wir zeigen, dass jede Folge in K eine konvergente Teilfolge enthält. Sei $(x_n)_{n \in \mathbb{N}} \subset K$. Für jedes $m \geq 1$ gibt es endlich viele Bälle mit Radius $\frac{1}{m}$, welche K überdecken. Mindestens einer dieser Bälle enthält unendlich viele Folgenglieder von $(x_n)_{n \in \mathbb{N}}$. Für $m = 1$ sei B_1 der Ball mit der Eigenschaft, dass $N_1 := \{n \mid x_n \in B_1$ unendlich viele Elemente hat. Sei $n_1 \in N_1$ beliebig. Für $m = 2$ sei B_2 der Ball mit der Eigenschaft, dass

$N_2 := \{n > n_1 \mid x_n \in B_1 \cap B_2\}$ unendlich viele Elemente hat. Wähle $n_2 \in N_2$ beliebig. Durch Iteration erhalten wir Folgen B_m, N_m und $(n_m)_{m\in\mathbb{N}}$ sodass

$$N_{m+1} = \left\{ n > n_m \mid x_n \in \bigcap_{k=1}^{m} B_k \right\}$$

unendlich viele Elemente enthält und $n_{m+1} \in N_{m+1}$. Dann ist $(x_{n_m})_{m\in\mathbb{N}}$ eine Teilfolge und es gilt $x_{n_k} \in B_m$ für alle $k \geq m$. Insbesondere gilt

$$d(x_{n_k}, x_{n_l}) \leq \frac{1}{m}, \quad k, l \geq m,$$

d.h. $(x_{n_m})_{m\geq 1}$ ist eine Cauchy Folge. Da E vollständig ist, hat diese einen Grenzwert $x \in E$. Da K abgeschlossen ist, liegt der Grenzwert in K. □

Satz 3.24. *Sei E ein Polnischer Raum. Dann ist $\{\mu\}$ straff für jedes $\mu \in \mathcal{P}(E)$. Insbesondere ist jedes $\mu \in \mathcal{P}(E)$ regulär.*

Beweis. Sei $(a_n)_{n\in\mathbb{N}} \subset E$ eine abzählbar dichte Teilmenge. Dann gilt für jedes $\delta > 0$

$$\bigcup_{n\geq 1} B_\delta(a_n) = E.$$

Damit folgt

$$1 = \mu(E) = \mu\left(\bigcup_{n\geq 1} B_\delta(a_n)\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} B_\delta(a_k)\right).$$

Sei $\varepsilon > 0$. Dann gibt es für jedes $m \geq 1$ ein $n_m \in \mathbb{N}$ mit

$$\mu\left(\bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)\right) \geq 1 - 2^{-m}\varepsilon.$$

Sei

$$K := \bigcap_{m\geq 1} \bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k).$$

Dann ist K abgeschlossen. Für jedes $\delta > 0$ sei $m \geq 1$ mit $\frac{1}{m} \leq \delta$. Dann gilt

$$K \subset \bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k) \subset \bigcup_{k=1}^{n_m} B_\delta(a_k).$$

Also ist K nach Lemma 3.23 kompakt. Die Behauptung folgt aus

$$\begin{aligned} \mu(K^c) &= \mu\left(\bigcup_{m\geq 1} \bigcap_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)^c\right) \leq \sum_{m=1}^{\infty} \mu\left(\bigcap_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)^c\right) \\ &= \sum_{m=1}^{\infty} \left(1 - \mu\left(\bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)\right)\right) \leq \varepsilon \sum_{m=1}^{\infty} 2^{-m} = \varepsilon, \end{aligned}$$

wo wir

$$\bigcap_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)^c = \bigcap_{k=1}^{n_m}(E \backslash B_{\frac{1}{m}}(a_k)) = E \backslash \left(\bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k) \right)$$

benutzt haben. □

3.2 Konvergenz von Zufallsvariablen

Sei (E, d) ein metrischer Raum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine E-wertige Zufallsvariable. Dann ist $\omega \longmapsto (X_n(\omega), X(\omega))$ messbar bezüglich $\mathcal{F} - -\mathcal{B}(E) \otimes \mathcal{B}(E)$. Ferner ist $(x, y) \longmapsto d(x, y)$ stetig, also $\mathcal{B}(E \times E) - -\mathcal{B}(\mathbb{R})$ messbar. Beachte, dass $E \times E$ mit $d_{E\times E}((x, y), (x', y')) := d(x, y) + d(x', y')$ wieder ein metrischer Raum ist. Um die Konvergenz $X_n \longrightarrow X$ zu untersuchen, betrachten wir die Abbildung

$$\Omega \longrightarrow [0, \infty), \quad \omega \longmapsto d(X_n(\omega), X(\omega)).$$

Diese ist messbar, sofern $\mathcal{B}(E \times E) \subset \mathcal{B}(E) \otimes \mathcal{B}(E)$ gilt. Ohne weiteres ist dieses jedoch nicht der Fall.

Lemma 3.25. *Sei E seperabel. Dann ist $E \times E$ seperabel und es gilt*

$$\mathcal{B}(E \times E) = \mathcal{B}(E) \otimes \mathcal{B}(E).$$

Hier und im folgenden ist (E, d) stets ein seperabler, metrischer Raum.

Definition 3.26. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X ein weitere E-wertige Zufallsvariable.*

(a) $X_n \longrightarrow X$ is Wahrscheinlichkeit, falls $d(X_n, X) \longrightarrow 0$ in Wahrscheinlichkeit.

(b) $X_n \longrightarrow X$ fast sicher, falls $d(X_n, X) \longrightarrow 0$ fast sicher.

(c) $X_n \longrightarrow X$ in $\mathcal{L}^p$, falls $d(X_n, X) \longrightarrow 0$ in $\mathcal{L}^p$.

(d) $X_n \longrightarrow X$ in Verteilung, falls $\mu_n := \mathbb{P} \circ X_n^{-1}$ schwach gegen $\mu := \mathbb{P} \circ X^{-1}$ konvergiert.

Konvergenz in Verteilung ist, per Definition vom Bildmaß, äquivalent zu

$$\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X)), \quad n \to \infty$$

für alle $f \in C_b(E)$.

Satz 3.27. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine weitere E-wertige Zufallsvariable. Angenommmen es gilt $X_n \longrightarrow X$ in Verteilung und X ist fast sicher konstant. Dann gilt $X_n \longrightarrow X$ in Wahrscheinlichkeit.*

Beweis. Es sei $\alpha \in E$ derart, dass $\mathbb{P}(X = \alpha) = 1$. Sei $A \in \mathcal{B}(E)$ beliebig, dann gilt

$$\mathbb{P} \circ X^{-1}(A) = \mathbb{P}(\{X \in A\}) = \mathbb{P}(\{X \in A\} \cap \{X = \alpha\}) = \delta_\alpha(A).$$

Folglich ist $\mathbb{P} \circ X^{-1} = \delta_\alpha$. Sei $\varepsilon > 0$ beliebig. Dann ist $\{y \in E \mid d(y, \alpha) \geq \varepsilon\}$ abgeschlossen. Aus dem Satz von Portmanteau folgt

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}(d(X_n, X) \geq \varepsilon) &= \limsup_{n\to\infty} \mathbb{P}(d(X_n, \alpha) \geq \varepsilon) \\
&= \limsup_{n\to\infty} \mathbb{P} \circ X_n^{-1} \left(\{y \in E \mid d(y, \alpha) \geq \varepsilon\}\right) \\
&= \mathbb{P} \circ X^{-1} \left(\{y \in E \mid d(y, \alpha) \geq \varepsilon\}\right) = 0,
\end{aligned}$$

da $\alpha \notin \{y \in E \mid d(y, \alpha) \geq \varepsilon\}$. $\square$

Satz 3.28. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine E-wertige Zufallsvariable. Gilt $X_n \longrightarrow X$ in Wahrscheinlichkeit, so folgt $X_n \longrightarrow X$ in Verteilung.*

Beweis. Sei $f \in C_b(E)$ gleichmässig stetig und $\varepsilon > 0$. Dann gibt es ein $\delta > 0$, sodass für alle $x, y \in E$ mit $d(x, y) < \delta$ bereits $|f(x) - f(y)| < \varepsilon$ gilt. Dann gilt

$$\begin{aligned}
|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| &\leq \int\limits_{d(X_n, X) < \delta} |f(X_n) - f(X)| d\mathbb{P} + \int\limits_{d(X_n, X) \geq \delta} |f(X_n) - f(X)| d\mathbb{P} \\
&\leq \varepsilon + 2\|f\|_\infty \mathbb{P}\left(d(X_n, X) \geq \delta\right).
\end{aligned}$$

$\square$

Satz 3.29. *Seien $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ Zufallsvariablen auf E und $\mu \in \mathcal{P}(E)$. Angenommen $\mathbb{P} \circ X_n^{-1} \longrightarrow \mu$ und $d(X_n, Y_n) \longrightarrow 0$ in Wahrscheinlichkeit. Dann gilt*

$$\mathbb{P} \circ Y_n^{-1} \longrightarrow \mu, \quad n \to \infty.$$

Beweis. Sei $F \subset E$ abgeschlossen. Für $\varepsilon > 0$ sei $F^\varepsilon = \{x \in E \mid d(x, F) \leq \varepsilon\}$ und $F_0^\varepsilon = \{x \in E \mid d(x, F) < \varepsilon\}$. Dann gilt

$$\begin{aligned}
\{Y_n \in F\} &= (\{Y_n \in F\} \cap \{d(X_n, Y_n) < \varepsilon\}) \cup (\{Y_n \in F\} \cap \{d(X_n, Y_n) \geq \varepsilon\}) \\
&= (\{Y_n \in F\} \cap \{d(X_n, Y_n) < \varepsilon\} \cap \{X_n \in F_0^\varepsilon\}) \cup (\{Y_n \in F\} \cap \{d(X_n, Y_n) \geq \varepsilon\}) \\
&\subset \{X_n \in F_0^\varepsilon\} \cup \{d(X_n, Y_n) \geq \varepsilon\} \subset \{X_n \in F^\varepsilon\} \cup \{d(X_n, Y_n) \geq \varepsilon\}.
\end{aligned}$$

Da F^ε abgeschlossen ist folgt

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}(Y_n \in F) &\leq \limsup_{n\to\infty} \mathbb{P}(d(Y_n, X_n) \geq \varepsilon) + \limsup_{n\to\infty} \mathbb{P}(X_n \in F^\varepsilon) \\
&= \limsup_{n\to\infty} \mathbb{P} \circ X_n^{-1}(F^\varepsilon) \leq \mathbb{P} \circ X^{-1}(F^\varepsilon).
\end{aligned}$$

Wegen $F^\varepsilon \searrow F$ für $\varepsilon \to 0$ folgt die Behauptung. $\square$

Wir zeigen noch einen alternativen Beweis.

Beweis. Wegen dem Satz von Portmanteau reicht es zu zeigen, dass für jede gleichmässig stetige Funktion $f : E \longrightarrow \mathbb{R}$

$$\int_E f(y)(\mathbb{P} \circ Y_n^{-1})(dy) \longrightarrow \int_E f(y) d\mu(y), \quad n \to \infty$$

gilt. Sei $\varepsilon > 0$ beliebig und wähle $\delta > 0$ derart, dass

$$|f(x) - f(y)| < \frac{\varepsilon}{3}, \quad \forall x, y \in E \text{ mit } d(x,y) < \delta.$$

Wähle $n_0 \in \mathbb{N}$ derart, dass

$$\left| \int_E f(y)(\mathbb{P} \circ X_n^{-1})(dy) - \int_E f(y) d\mu(y) \right| < \frac{\varepsilon}{3}$$

sowie $\mathbb{P}(d(X_n, Y_n) \geq \delta) < \frac{\varepsilon}{3(2\|f\|_\infty + 1)}$ für alle $n \geq n_0$ gilt. Dann folgt

$$\begin{aligned}
&\left| \int_E f(y)(\mathbb{P} \circ Y_n^{-1})(dy) - \int_E f(y) d\mu(y) \right| \\
&\leq |\mathbb{E}(f(X_n)) - \mathbb{E}(f(Y_n))| + \left| \int_E f(y)(\mathbb{P} \circ X_n^{-1})(dy) - \int_E f(y) d\mu(y) \right| \\
&\leq \int_{d(X_n,Y_n) \geq \delta} |f(X_n) - f(Y_n)| d\mathbb{P} + \int_{d(X_n,Y_n) < \delta} |f(X_n) - f(Y_n)| d\mathbb{P} + \frac{\varepsilon}{3} \\
&\leq 2\|f\|_\infty \mathbb{P}(d(X_n, Y_n) \geq \delta) + \frac{2\varepsilon}{3} \leq \varepsilon.
\end{aligned}$$

□

4 Charakteristische Funktionen

Im letzten Kapitel haben wir Eigenschaften von Wahrscheinlichkeitsmaßen auf metrischen und Polnischen Räumen untersucht. Dieses Kapitel widmet sich dem Studium von Wahrscheinlichkeitsmaßen auf $E = \mathbb{R}^d$. Für diesen Spezialfall ist es möglich jedem endlichen Maß μ auf $\mathbb{R}^d$ eine komplex-wertige Funktion (charakteristische Funktion) $\widehat{\mu}(\xi)$ zuzuordnen. Schwache Konvergenz von Wahrscheinlichkeitsmaßen sowie Existenz von Momenten können mithilfe der charakteristischen Funktion beschrieben werden.

4.1 Fouriertransformation von Funktionen

Für $\xi, x \in \mathbb{R}^d$ sei $\xi \cdot x := \sum_{k=1}^{d} \xi_k x_k = \langle \lambda, A\lambda \rangle$ das euklidische Skalarprodukt. hier und im folgenden bezeichnet $\mathcal{L}^1(\mathbb{R}^d)$ den Raum aller messbaren, integrierbaren Funktionen auf $\mathbb{R}^d$, d.h. für $g \in \mathcal{L}^1(\mathbb{R}^d)$ gilt

$$\int_{\mathbb{R}^d} |g(x)| m(dx) < \infty.$$

Definition 4.1. *Sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Die Fouriertransformierte von g ist für $\xi \in \mathbb{R}^d$ definiert durch*

$$\widehat{g}(\xi) := \int_{\mathbb{R}^d} e^{i\xi \cdot x} g(x) m(dx) := \int_{\mathbb{R}^d} \cos(\xi \cdot x) g(x) m(dx) + i \int_{\mathbb{R}^d} \sin(\xi \cdot x) g(x) m(dx).$$

Die Definition ist wegen $|e^{i\xi \cdot x}| = 1$ für alle $\xi \in \mathbb{R}^d$ wohldefiniert.

Definition 4.2. *Eine Funktion $g : \mathbb{R}^d \longrightarrow \mathbb{C}$ heißt positiv (semi-)definit, wenn für alle $n \in \mathbb{N}$ und alle $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$ die Matrix $(g(\xi_i - \xi_j))_{1 \leq i,j \leq n}$ positiv (semi-)definit ist.*

Diese Definition ist äquivalent zu: Für alle $n \in \mathbb{N}$, alle $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$ und alle $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$ gilt

$$\sum_{i,j=1}^{n} \lambda_j \overline{\lambda}_i g(\xi_j - \xi_i) \geq 0. \tag{4.1}$$

Denn setze $A := (g(\xi_j - \xi_i))_{1 \leq i,j \leq n}$, d.h. $A_{ij} = g(\xi_j - \xi_i)$. Dann gilt $(A\lambda)_i = \sum_{j=1}^{n} A_{ij} \lambda_j$ und folglich

$$\langle \lambda, A\lambda \rangle = \sum_{i=1}^{n} \overline{\lambda}_i (A\lambda)_i = \sum_{i,j=1}^{n} \overline{\lambda}_i \lambda_j A_{ij} = \sum_{i,j=1}^{n} \overline{\lambda}_i \lambda_j g(\xi_j - \xi_i).$$

Man beachte, dass die Matrix A insbesondere hermitsch ist, d.h. $\overline{A} = A^T$. Es sei g eine positiv (semi-)definite Funktion. Dann gilt:

- Für $n = 1$ und $\lambda \in \mathbb{C}$ folgt aus (4.1) $|\lambda|^2 g(0) \geq 0$, also $g(0) \geq 0$.
- Für $n = 2$ mit $\xi_1 = 0$ und $\xi_2 = \xi \in \mathbb{R}^d$ folgt, dass die Matrix $\begin{pmatrix} g(0) & g(\xi) \\ g(-\xi) & g(0) \end{pmatrix}$ positiv (semi-)definit ist. Daraus folgt $g(-\xi) = \overline{g(\xi)}$. Ferner muss die Determinante nicht-negativ sein, also
$$g(0)^2 - g(\xi)g(-\xi) = g(0)^2 - |g(\xi)|^2 \geq 0$$
und somit $|g(\xi)| \leq g(0)$.

Es sei $C_c^\infty(\mathbb{R}^d)$ der Raum der glatten Funktionen mit kompaktem Träger. D.h. $f \in C_c^\infty(\mathbb{R}^d)$, falls f beliebig oft differenzierbar ist und $f \in C_c(\mathbb{R}^d)$.

Lemma 4.3. *[Wer00] Für jedes $g \in \mathcal{L}^1(\mathbb{R}^d)$ und jedes $\varepsilon > 0$ gibt es eine Funktion $f_\varepsilon \in C_c^\infty(\mathbb{R}^d)$ mit*

$$\|f_\varepsilon - g\|_{\mathcal{L}^1} := \int_{\mathbb{R}^d} |f_\varepsilon(x) - g(x)| m(dx) < \varepsilon.$$

Die wichtigsten Eigenschaften der Fouriertransformierten von g sind im nächsten Satz zusammengefasst.

Satz 4.4. *Es sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Es gelten die folgenden Eigenschaften.*

(a) $|\widehat{g}(\xi)| \le \|g\|_{\mathcal{L}^1}$ *und* $\widehat{g}(0) = \int_{\mathbb{R}^d} g(x) m(dx)$.

(b) $\widehat{g}$ *ist gleichmässig stetig.*

(c) $\widehat{g}$ *verschwindet im Unendlichen, d.h. für alle $\varepsilon > 0$ gibt es eine kompakte Menge $K_\varepsilon \subset \mathbb{R}^d$ derart, dass*

$$|\widehat{g}(\xi)| < \varepsilon, \quad \xi \notin K_\varepsilon.$$

(d) Ist g fast überall nicht-negativ, so ist $\widehat{g}$ positiv (semi-)definit.

Beweis. (a) Klar.
(b) Seien $\xi_1, \xi_2 \in \mathbb{R}^d$ und $\varepsilon > 0$. Wähle eine kompakte Menge $K_\varepsilon \subset \mathbb{R}^d$ derart, dass

$$\int_{K_\varepsilon^c} |g(x)| m(dx) < \frac{\varepsilon}{4}.$$

Sei $\delta > 0$ mit $\delta < \frac{\varepsilon}{2(1+\sup_{x \in K_\varepsilon} |x|)(1+\|g\|_{\mathcal{L}^1})}$. Dann gilt für alle $|\xi_1 - \xi_2| < \delta$

$$|\widehat{g}(\xi_1) - \widehat{g}(\xi_2)| \le \int_{\mathbb{R}^d} |e^{i\xi_1 \cdot x} - e^{i\xi_2 \cdot x}||g(x)| m(dx) \le \int_{K_\varepsilon} |e^{i\xi_1 \cdot x} - e^{i\xi_2 \cdot x}||g(x)| m(dx) + \frac{\varepsilon}{2}.$$

Es gilt

$$|e^{i\xi_1 \cdot x} - e^{i\xi_2 \cdot x}| = \left| i \int_{\xi_1 \cdot x}^{\xi_2 \cdot x} e^{it} dt \right| \le |(\xi_2 - \xi_1) \cdot x| \le |x| \cdot |\xi_2 - \xi_1|$$

woraus folgt

$$|\widehat{g}(\xi_1) - \widehat{g}(\xi_2)| \le \frac{\varepsilon}{2} + \|g\|_{\mathcal{L}^1} \sup_{x \in K_\varepsilon} |x| \cdot |\xi_1 - \xi_2| \le \varepsilon.$$

(c) Wir zeigen die Behauptung zuerst für $g \in C_c^\infty(\mathbb{R}^d)$. Wähle $R > 0$ derart, dass $g(x) = 0$ für alle $|x| \ge R$. Wähle $R' > R$ und sei $|\xi| \ge R$. Dann finden wir ein $j \in \{1, \dots, d\}$ mit $|\xi_j| \ge \frac{R'}{\sqrt{d}}$.

Mittels partieller Integration, wobei der Randterm wegen $g \in C_c^\infty(\mathbb{R}^d)$ verschwidet, erhalten wir

$$\begin{aligned}\widehat{g}(\xi) &= \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} g(x) m(dx) = \int\limits_{|x|\le R'} e^{i\xi\cdot x} g(x) m(dx) \\ &= \frac{1}{i\xi_j} \int\limits_{|x|\le R'} \frac{\partial e^{i\xi\cdot x}}{\partial x_j} g(x) m(dx) = -\frac{1}{i\xi_j} \int\limits_{|x|\le R'} e^{i\xi\cdot x} \frac{\partial g(x)}{\partial x_j} m(dx).\end{aligned}$$

Daraus folgt

$$|\widehat{g}(\xi)| \le \frac{1}{|\xi_j|} \left\| \frac{\partial g}{\partial x_j} \right\|_{\mathcal{L}^1} \le \frac{\sqrt{d}}{R'} \left\| \frac{\partial g}{\partial x_j} \right\|_{\mathcal{L}^1}.$$

Die Behauptung folgt aus $R' \longrightarrow \infty$. Für den allgemeinen Fall sei $g \in \mathcal{L}^1(\mathbb{R}^d)$, $\varepsilon > 0$ und $f_\varepsilon \in C_c^\infty(\mathbb{R}^d)$ derart, dass $\|g - f_\varepsilon\|_{\mathcal{L}^1} < \varepsilon$. Die Behauptung folgt aus

$$|\widehat{g}(\xi)| \le |\widehat{g}(\xi) - \widehat{f}_\varepsilon(\xi)| + |\widehat{f}_\varepsilon(\xi)| \le \|g - f_\varepsilon\|_{\mathcal{L}^1} + |\widehat{f}_\varepsilon(\xi)|.$$

(d) Es gilt

$$\int\limits_{\mathbb{R}^d} \left| \sum_{j=1}^n \lambda_j e^{i\xi_j\cdot x} \right|^2 g(x) m(dx) = \int\limits_{\mathbb{R}^d} \sum_{i,j=1}^n \lambda_j \overline{\lambda}_i e^{ix\cdot(\xi_j - \xi_i)} g(x) m(dx) = \sum_{i,j=1}^n \lambda_j \overline{\lambda}_i \widehat{g}(\xi_j - \xi_i).$$

□

Wir betrachten das Beispiel einer Normalverteilung.

Beispiel 4.5. *Es sei $g(x) := e^{-\alpha|x|^2}$ mit $\alpha > 0$. Dann ist*

$$\widehat{g}(\xi) = \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} e^{-\frac{|\xi|^2}{4\alpha}}. \tag{4.2}$$

Beachte, für $\alpha = \frac{1}{2}$ ist $\widehat{g}(\xi) = (2\pi)^{\frac{d}{2}} g(\xi)$.

Beweis. Es gilt

$$\widehat{g}(\xi) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} e^{-\alpha|x|^2} m(dx) = \prod_{k=1}^d \int\limits_{\mathbb{R}} e^{i\xi_k\cdot x_k} e^{-\alpha x_k^2} m(dx).$$

Also reicht es aus die Behauptung für $d = 1$ zu beweisen. Für diesen Fall erhalten wir für $\xi, x \in \mathbb{R}$

$$e^{-\alpha x^2} e^{i\xi x} = e^{-\alpha\left(x^2 - 2x\frac{i\xi}{2\alpha} + \left(\frac{i\xi}{2\alpha}\right)^2 - \left(\frac{i\xi}{2\alpha}\right)^2\right)} = e^{-\frac{\xi^2}{4\alpha}} e^{-\alpha\left(x - \frac{i\xi}{2\alpha}\right)^2}.$$

Daraus folgt

$$\int\limits_{\mathbb{R}} e^{i\xi x} e^{-\alpha x^2} m(dx) = e^{-\frac{\xi^2}{4\alpha}} \int\limits_{\mathbb{R}} e^{-\alpha\left(x - \frac{i\xi}{2\alpha}\right)^2} m(dx).$$

Für das Integral erhalten wir mittels Cauchys Integralsatz (siehe [FL80]) und der Substitution $y = \sqrt{\alpha}x$

$$\int_{\mathbb{R}} e^{-\alpha\left(x-\frac{i\xi}{2\alpha}\right)^2} m(dx) = \int_{\mathbb{R}} e^{-\alpha x^2} m(dx) = \frac{1}{\sqrt{\alpha}} \int_{\mathbb{R}} e^{-x^2} m(dx) = \sqrt{\frac{\pi}{\alpha}}.$$

□

Als nächstes wollen wir die Inverse Transformation finden.

Lemma 4.6. *Es seien $g, h \in \mathcal{L}^1(\mathbb{R}^d)$. Dann gilt*

$$\int_{\mathbb{R}^d} \widehat{g}(\xi) h(\xi) m(d\xi) = \int_{\mathbb{R}^d} g(\xi) \widehat{h}(\xi) m(d\xi).$$

Beweis. Mittels Fubini erhalten wir

$$\int_{\mathbb{R}^d} \widehat{g}(\xi) h(\xi) m(d\xi) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{i\xi\cdot x} g(x) h(\xi) m(dx) m(d\xi) = \int_{\mathbb{R}^d} g(\xi) \widehat{h}(\xi) m(d\xi).$$

Fubini ist anwendbar, da $(x, \xi) \longmapsto g(x)h(\xi)$ integrierbar ist bezüglich $m(dx)m(d\xi)$. □

Lemma 4.7. *Für jedes $g \in C_b(\mathbb{R}^d)$ und jedes $x \in \mathbb{R}^d$ gilt*

$$\int_{\mathbb{R}^d} g(y) \frac{e^{-\frac{|x+y|^2}{4\alpha}}}{(4\pi\alpha)^{\frac{d}{2}}} m(dy) \longrightarrow g(-x), \quad \alpha \to 0.$$

Beweis. Übung. □

Lemma 4.8. *Es sei $g \in C_b(\mathbb{R}^d) \cap \mathcal{L}^1(\mathbb{R}^d)$ derart, dass $\widehat{g} \in \mathcal{L}^1(\mathbb{R}^d)$. Dann gilt*

$$\widehat{\widehat{g}}(x) = (2\pi)^d g(-x), \quad x \in \mathbb{R}^d.$$

Beweis. Sei $\alpha > 0$ und setze $\varphi_x(\xi) := e^{i\xi\cdot x} e^{-\alpha|\xi|^2}$. Für diese Funktion gilt nach (4.2)

$$\widehat{\varphi}_x(\xi) = \int_{\mathbb{R}^d} e^{i(\xi+x)\cdot y} e^{-\alpha|y|^2} m(dy) = \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} e^{-\frac{|\xi+x|^2}{4\alpha}}.$$

Damit erhalten wir

$$\begin{aligned}\int_{\mathbb{R}^d} \widehat{g}(\xi) e^{i\xi\cdot x} e^{-\alpha|\xi|^2} m(d\xi) &= \int_{\mathbb{R}^d} g(\xi) \widehat{\varphi}_x(\xi) m(d\xi) \\ &= \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} \int_{\mathbb{R}^d} g(\xi) e^{-\frac{|\xi+x|^2}{4\alpha}} m(d\xi) = (2\pi)^d \int_{\mathbb{R}^d} g(\xi) \frac{e^{-\frac{|\xi+x|^2}{4\alpha}}}{(4\pi\alpha)^{\frac{d}{2}}} m(d\xi).\end{aligned}$$

Betrachten wir den Grenzwert $\alpha \to 0$, so folgt die Behauptung aus Lemma 4.7. □

Satz 4.9. *Sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Definiere die Abbildung*

$$\breve{g}(x) := \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x} g(\xi) m(d\xi), \quad x \in \mathbb{R}^d.$$

Dann gilt für alle $g \in C_b(\mathbb{R}^d) \cap \mathcal{L}^1(\mathbb{R}^d)$ derart, dass $\widehat{g} \in \mathcal{L}^1(\mathbb{R}^d)$

$$\breve{\widehat{g}} = g = \widehat{\breve{g}}.$$

Beweis. Beachte, dass $\breve{g}(x) = \frac{1}{(2\pi)^d}\widehat{g}(-x)$ für alle $g \in \mathcal{L}^1(\mathbb{R}^d)$ gilt. Sei $g \in C_b(\mathbb{R}^d) \cap \mathcal{L}^1(\mathbb{R}^d)$ mit $\widehat{g} \in \mathcal{L}^1(\mathbb{R}^d)$. Aus Lemma 4.8 folgt $\breve{\widehat{g}}(x) = \frac{1}{(2\pi)^d}\widehat{\widehat{g}}(-x) = g(x)$ für alle $x \in \mathbb{R}^d$. Analog lässt sich die andere Identität zeigen. □

4.2 Charakteristische Funktion für Maße

Definition 4.10. *Sei μ ein endliches Maß auf $\mathbb{R}^d$. Die charakteristische Funktion von μ ist die Fouriertransformierte von μ gegeben durch*

$$\widehat{\mu}(\xi) := \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} \mu(dx), \quad \xi \in \mathbb{R}^d.$$

Bemerkung 4.11. *(a) Sei $\mu = \sum_{k=1}^{\infty} a_k \delta_{x_k}$ mit $a_k \geq 0$ und $x_k \in \mathbb{R}^d$ derart, dass $\sum_{k=1}^{\infty} a_k = 1$. Dann ist μ ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$ und es gilt*

$$\widehat{\mu}(\xi) = \sum_{k=1}^{\infty} a_k e^{i\xi\cdot x_k}, \quad \xi \in \mathbb{R}^d.$$

(b) Hat μ eine Dichte, d.h. $\mu(dx) = p(x)m(dx)$, so ist

$$\widehat{\mu}(\xi) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} p(x) m(dx) = \widehat{p}(x).$$

Die wichtigsten Eigenschaften sind im nächsten Satz zusammengefasst.

Satz 4.12. *Es sei μ ein endliches Maß auf $\mathbb{R}^d$. Dann gelten die folgenden Eigenschaften.*

(a) $|\widehat{\mu}(\xi)| \leq \mu(\mathbb{R}^d) = \widehat{\mu}(0)$ für alle $\xi \in \mathbb{R}^d$.

(b) $\widehat{\mu}$ ist gleichmässig stetig.

(c) $\widehat{\mu}$ ist positiv (semi-)definit.

Beweis. Übung. □

Wir betrachten einige Beispiele.

Beispiel 4.13. *(a) Benoulli Verteilung:* $\mu = p\delta_1 + (1-p)\delta_0$. *Dann ist*

$$\widehat{\mu}(\xi) = pe^{i\xi} + (1-p) = 1 + p(1-e^{i\xi}), \quad \xi \in \mathbb{R}.$$

(b) Binomialverteilung: $\mu = \sum\limits_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}\delta_k$. *Dann ist*

$$\widehat{\mu}(\xi) = (1-p+pe^{i\xi})^n, \quad \xi \in \mathbb{R}.$$

(c) Standardnormalverteilung: Im eindimensionalen Fall sei $\mu(dx) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} m(dx)$. *Dann ist*

$$\widehat{\mu}(\xi) = e^{-\frac{\xi^2}{2}}, \quad \xi \in \mathbb{R}.$$

Für den d-dimansionalen Fall sei $\mu(dx) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}}$. *Dann gilt*

$$\widehat{\mu}(\xi) = e^{-\frac{|\xi|^2}{2}}, \quad \xi \in \mathbb{R}^d.$$

(d) Normalverteilung im $\mathbb{R}^m$*: Sei* μ *eine Normalverteilung auf dem* $\mathbb{R}^m$. *Dann gibt es eine* $m \times d$*-Matrix* A *und* $b \in \mathbb{R}^m$ *mit* $\mu = \nu \circ T^{-1}$, *wo* $T(x) = Ax + b$ *und*

$$\nu(dx) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}} m(dx).$$

Damit folgt

$$\begin{aligned}\widehat{\mu}(\xi) &= \int\limits_{\mathbb{R}^m} e^{i\xi\cdot x}\mu(dx) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot T(x)}\nu(dx)\\ &= e^{i\xi\cdot b}\int\limits_{\mathbb{R}^d} e^{i(A^T\xi)\cdot x}\nu(dx) = e^{i\xi\cdot b}e^{-\frac{1}{2}|A^t\xi|^2} = e^{i\xi\cdot b}e^{-\frac{1}{2}\langle\xi,(AA^t)\xi\rangle}.\end{aligned}$$

(e) Uniforme Verteilung: Sei $\mu(dx) = \mathbb{1}_{(-a,a)}(x)\frac{1}{2a}m(dx)$ *mit* $a > 0$. *Dann ist*

$$\widehat{\mu}(\xi) = \begin{cases} \frac{\sin(\xi a)}{\xi a}, & \xi \neq 0 \\ 1, & \textit{sonst} \end{cases}, \quad \xi \in \mathbb{R}.$$

(f) Cauchy Verteilung: Sei $\mu(dx) = \frac{1}{\pi}\frac{c}{c^2+x^2}m(dx)$ *mit* $c > 0$. *Dann gilt*

$$\widehat{\mu}(\xi) = e^{-c|\xi|}, \quad \xi \in \mathbb{R}.$$

Der nächste Satz zeigt, dass die charakteristische Funktion das Maß eindeutig festlegt.

Satz 4.14. *Seien* μ, ν *zwei endliche Maße auf* $\mathbb{R}^d$ *mit* $\widehat{\mu} = \widehat{\nu}$. *Dann gilt* $\mu = \nu$.

Beweis. Es gilt $\mu(\mathbb{R}^d) = \widehat{\mu}(0) = \widehat{\nu}(0) = \nu(\mathbb{R}^d)$. Ist $\mu(\mathbb{R}^d) = 0$, so ist nichts zu zeigen. Sei also $\mu(\mathbb{R}^d) > 0$. Definiere neue Maße $\mu' := \frac{1}{\mu(\mathbb{R}^d)}\mu$ sowie $\nu' := \frac{1}{\nu(\mathbb{R}^d)}\nu$. Dann gilt $\widehat{\mu'} = \widehat{\nu'}$. Es reicht daher aus die Behauptung für Wahrscheinlichkeitsmaße zu beweisen.

Die kompakten Mengen bilden ein $\cap$-stabiles Erzeugendensystem der Borel-σ-Algebra $\mathcal{B}(\mathbb{R}^d)$. Es genügt also $\mu(K) = \nu(K)$ für alle kompakten Mengen $K \subset \mathbb{R}^d$ zu zeigen. Sei $K \subset \mathbb{R}^d$ kompakt und

$$f_m(x) = \begin{cases} 1, & x \in K \\ 0, & d(x, K) \geq \frac{1}{m} \\ 1 - md(x, K), & \text{sonst} \end{cases}.$$

Dann gilt $0 \leq f_m \leq 1$, f_m ist stetig und $f_m \searrow \mathbb{1}_K$ für $m \to \infty$. Wir zeigen

$$\int\limits_{\mathbb{R}^d} f_m(x) d\mu(x) = \int\limits_{\mathbb{R}^d} f_m(x) d\nu(x), \quad m \in \mathbb{N}. \tag{4.3}$$

Dann folgt mit dem Satz von Lebesgue $\nu(K) = \mu(K)$. Allgemeiner zeigen wir (4.3) für alle $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ mit

- $0 \leq f \leq 1$.
- $f \in C_c(\mathbb{R}^d)$.

Sei $\varepsilon > 0$ beliebig. Da f kompakten Träger hat, können wir $N \in \mathbb{N}$ wählen mit

$$\overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}} \subset [-N, N]^d =: B_N$$

und $\mu(B_N^c) < \varepsilon$, $\nu(B_N^c) < \varepsilon$. NAch dem Weierstraßschen Approximationssatz (siehe [Wer00] für den Beweis in einer Dimension) gibt es eine Funktion

$$g_\varepsilon(x) = \sum_{j=1}^{m} c_j^\varepsilon e^{i \frac{\pi}{N} \langle \xi_j^\varepsilon, x \rangle}$$

mit $c_j^\varepsilon \in \mathbb{C}$, $\xi_j^\varepsilon \in \mathbb{Z}^d$, $\|g_\varepsilon\|_\infty \leq \|f\|_\infty$ und

$$\sup_{x \in B_N} |f(x) - g(x)| \leq \varepsilon.$$

Es gilt mit $f|_{B_N^c} = 0$ und $\mu(B_N) \leq 1$

$$\begin{aligned} \left| \int\limits_{\mathbb{R}^d} (f - g_\varepsilon) d\mu \right| &\leq \int\limits_{B_N} |f - g_\varepsilon| d\mu + \int\limits_{B_N^c} |f| + |g_\varepsilon| d\mu \\ &\leq \varepsilon\mu(B_N) + \|g_\varepsilon\|_\infty \mu(B_N^c) \leq (1 + \|f\|_\infty)\varepsilon. \end{aligned}$$

Analog lässt sich zeigen, dass

$$\left| \int\limits_{\mathbb{R}^d} (f - g_\varepsilon) d\nu \right| \leq (1 + \|f\|_\infty)\varepsilon.$$

Insgesamt erhalten wir

$$\begin{aligned}\left|\int\limits_{\mathbb{R}^d} f d\mu - \int\limits_{\mathbb{R}^d} f d\nu\right| &\leq \left|\int\limits_{\mathbb{R}^d} (f - g_\varepsilon) d\mu\right| + \left|\int\limits_{\mathbb{R}^d} g_\varepsilon d\mu - \int\limits_{\mathbb{R}^d} g_\varepsilon d\nu\right| + \left|\int\limits_{\mathbb{R}^d} (g_\varepsilon - f) d\nu\right| \\ &\leq 2\varepsilon + \left|\int\limits_{\mathbb{R}^d} g_\varepsilon d\mu - \int\limits_{\mathbb{R}^d} g_\varepsilon d\nu\right|.\end{aligned}$$

Die Behauptung folgt aus

$$\int\limits_{\mathbb{R}^d} g_\varepsilon d\mu = \sum_{j=1}^m c_j^\varepsilon \int\limits_{\mathbb{R}^d} e^{i\frac{\pi}{N}\langle \xi_j^\varepsilon, x\rangle} d\mu(x) = \sum_{j=1}^m c_j^\varepsilon \int\limits_{\mathbb{R}^d} e^{i\frac{\pi}{N}\langle \xi_j^\varepsilon, x\rangle} d\nu(x) = \int\limits_{\mathbb{R}^d} g_\varepsilon d\nu.$$

□

Satz 4.15. *[Jac01, Theorem 3.5.7] Eine Funktion $g : \mathbb{R}^d \longrightarrow \mathbb{C}$ ist genau dann die charakteristische Funktion von einem endlichen Maß μ auf $\mathbb{R}^d$, wenn die folgenden Eigenschaften gelten:*

(i) g ist stetig.

(ii) g ist positiv (semi-)definit.

In diesem Fall ist $g(0) = \mu(\mathbb{R}^d)$. Insbesondere ist μ ein Wahrscheinlichkeitsmaß genau dann, wenn $g(0) = 1$.

Im Folgenden wollen wir weitere Eigenschaften von μ aus der charakteristischen Funktion $\widehat{\mu}$ ableiten.

Lemma 4.16. *Für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ gibt es eine Folge $(\varphi_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ mit $\widehat{\varphi}_n \in \mathcal{L}^1(\mathbb{R}^d)$ derart, dass $\varphi_n(x) \longrightarrow \varphi(x)$ für alle $x \in \mathbb{R}^d$.*

Beweis. Sei $p_\alpha(x) = (4\pi\alpha)^{-\frac{d}{2}} e^{-\frac{|x|^2}{4\alpha}}$ und

$$\varphi_\alpha(x) := \int\limits_{\mathbb{R}^d} \varphi(x-y) p_\alpha(y) m(dy).$$

Dann gilt

$$|\varphi_\alpha(y)| \leq \|\varphi\|_\infty \int\limits_{\mathbb{R}^d} p_\alpha(y) m(dy) = \|\varphi\|_\infty.$$

Da $p_\alpha(y)m(dy)$ schwach gegen δ_0 konvergiert, folgt $\varphi_\alpha(y) \longrightarrow \varphi(y)$ für $alpha \to \infty$ und alle $y \in \mathbb{R}^d$. Ferner gilt $\widehat{\varphi}_\alpha = \widehat{\varphi} \cdot \widehat{p}_\alpha$. Da $\widehat{\varphi}$ beschränkt ist und $\widehat{p}_\alpha \in \mathcal{L}^1(\mathbb{R}^d)$ folgt die Behauptung. □

Satz 4.17. *Es sei μ ein endliches Maß auf $\mathbb{R}^d$. Es gelten die folgenden Aussagen.*

(a) Ist $\widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$, so hat μ eine Dichte $0 \le p \in C_b(\mathbb{R}^d)$, welche im Unendlichen verschwindet und gegeben ist durch

$$p(x) := \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x} \widehat{\mu}(\xi) m(d\xi), \quad x \in \mathbb{R}^d. \tag{4.4}$$

(b) Ist $\widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$ derart, dass für ein $m \ge 1$

$$\int\limits_{\mathbb{R}^d} |\xi|^m |\widehat{\mu}(\xi)| m(d\xi) < \infty.$$

Dann hat p stetige Ableitungen bis zur Ordnung m. In diese Fall gilt für alle $0 \le k \le m$ und alle $1 \le j_1, \ldots, j_k \le d$

$$\frac{\partial^k p(x)}{\partial x_{j_1} \cdots \partial x_{j_k}} = \frac{(-i)^k}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x} \xi_{j_1} \cdots \xi_{j_k} \widehat{\mu}(\xi) m(d\xi).$$

(c) Ist

$$\int\limits_{\mathbb{R}^d} |x|^m \mu(dx) < \infty$$

für ein $m \ge 1$. Dann hat $\widehat{\mu}$ stetige partielle Ableitungen bis zur Ordnung m und es gilt für alle $0 \le k \le m$ und $1 \le j_1, \ldots, j_k \le d$

$$\frac{\partial^k \widehat{\mu}(xi)}{\partial \xi_{j_1} \cdots \partial \xi_{j_k}} = i^k \int\limits_{\mathbb{R}^d} x_{j_1} \cdots x_{j_k} e^{i\xi\cdot\cdot x} \mu(dx).$$

Beweis. (a) Für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d)$ gilt $\varphi \cdot \widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$ und folglich

$$\int\limits_{\mathbb{R}^d} \varphi(x)\widehat{\mu}(x) m(dx) = \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} \varphi(x) e^{i\xi\cdot x} \mu(d\xi) m(dx) = \int\limits_{\mathbb{R}^d} \widehat{\varphi}(\xi)\mu(d\xi).$$

Man beachte, dass alle Integrale existieren und somit der Satz von Fubini anwendbar ist. Offensichtlich definiert (4.4) eine stetige, im Unendlich verschwindende Funktion p mit $|p(x)| \le \|\widehat{\mu}\|_{\mathcal{L}^1}$. Wir erhalten für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ mit $\widehat{\varphi} \in \mathcal{L}^1(\mathbb{R}^d)$

$$\begin{aligned}(2\pi)^d \int\limits_{\mathbb{R}^d} \varphi(x)p(x)m(dx) &= \int\limits_{\mathbb{R}^d} \varphi(x)\widehat{\widehat{\mu}}(-x)m(dx) = \int\limits_{\mathbb{R}^d} \varphi(-x)\widehat{\widehat{\mu}}(x)m(dx) \\ &= \int\limits_{\mathbb{R}^d} \widehat{\varphi(-\cdot)}(x)\widehat{\mu}(x)m(dx) = \int\limits_{\mathbb{R}^d} \widehat{\widehat{\varphi(-\cdot)}}(x)\mu(dx) = (2\pi)^d \int\limits_{\mathbb{R}^d} \varphi(x)\mu(dx).\end{aligned}$$

Aus Lemma 4.16 und dem Satz von Lebesgue folgt für alle $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$

$$\int\limits_{\mathbb{R}^d} \varphi(x)p(x)m(dx) = \int\limits_{\mathbb{R}^d} \varphi(x)\mu(dx).$$

Die Behauptung folgt aus $C_c(\mathbb{R}^d) \subset \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$.
(b) Die Behauptung folgt aus der Darstellung (4.4) und denselben Argumenten wir für Teil (c).
(c) Wir zeigen nur den Fall $m = 1$. Der allgemeine Fall folgt dann über Induktion. Sei e_j der Basisvektor im $\mathbb{R}^d$ in Richtung j. Dann ist für $h > 0$

$$\frac{\widehat{\mu}(\xi + he_j) - \widehat{\mu}(xi)}{h} = \int\limits_{\mathbb{R}^d} \frac{e^{i(\xi+he_j)\cdot x} - e^{i\xi\cdot x}}{h}\mu(dx) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x}\frac{e^{ix_jh} - 1}{h}\mu(dx).$$

Es gilt $\frac{e^{ix_jh}-1}{h} \longrightarrow ix_j$ und nach Lemma 8.1 (siehe Appendix) gibt es eine stetige Funktion Θ_1 mit $|\Theta_1(x)| \leq 1$ derart, dass $e^{ix_kh} = 1 + ix_jh\Theta_1(x_jh)$. Daraus folgt

$$\frac{e^{ix_jh} - 1}{h} = ix_j\Theta_1(x_jh)$$

und somit $\left|\frac{e^{ix_jh}-1}{h}\right| \leq |x_j|$ für alle $h > 0$. Dieses liefert eine integrierbare Majorante, also folgt aus dem Satz von Lebesgue

$$\frac{\partial\widehat{\mu}(\xi)}{\partial\xi_j} = i\int\limits_{\mathbb{R}^d} x_je^{i\xi_j\cdot x}\mu(dx).$$

Die Ableitung ist stetig, ebenfalls aus dem Satz von Lebesgue. □

Bemerkung 4.18. *Die Momente von μ haben die Darstellung*

$$\frac{1}{i^k}\frac{\partial^k\widehat{\mu}(0)}{\partial\xi_{j_1}\cdots\partial\xi_{j_k}} = \int\limits_{\mathbb{R}^d} x_{j_1}\cdots x_{j_k}\mu(dx).$$

Für $x \in \mathbb{R}^d$ setze $g_x(y) := \langle x, y\rangle = x \cdot y$.

Satz 4.19. *(Satz von Cramer-Wald) Seien $\mu_n, \mu \in \mathcal{P}(\mathbb{R}^d)$. Dann sind äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) $\mu_n \circ g_x^{-1} \longrightarrow \mu \circ g_x^{-1}$ schwach für $n \to \infty$ und alle $x \in \mathbb{R}^d$.

Beweis. $(a) \Longrightarrow (b)$: Folgt aus der Stetigkeit von g_x.
$(b) \Longrightarrow (a)$: Die Familie $\{\mu_n \circ g_x^{-1} \mid n \in \mathbb{N}\}$ ist nach Voraussetzung für jedes $x \in \mathbb{R}^d$ relativ kompakt und folglich straff. Seien $e_1, \ldots, e_d \in \mathbb{R}^d$ die Einheitsvektoren und $\varepsilon > 0$ beliebig. Dann gibt es für jedes $i = 1, \ldots, d$ eine kompakte Menge $K_i \subset \mathbb{R}$ mit

$$(\mu_n \circ g_{e_i}^{-1})(K_i) \geq 1 - \frac{\varepsilon}{d}, \quad n \in \mathbb{N}.$$

Setze $K := \bigcap_{i=1}^{d} g_{e_i}^{-1}(K_i)$. Es ist nicht schwer zu zeigen, dass K abgeschlossen und beschränkt ist, also kompakt. Ferner gilt

$$\mu_n(K^c) = \mu_n\left(\bigcup_{i=1}^{d}(g_{e_i}^{-1}(K_i))^c\right) \leq \sum_{i=1}^{d}(\mu_n \circ g_{e_i}^{-1})(K_i^c) \leq d\frac{\varepsilon}{d} = \varepsilon$$

und folglich ist $\{\mu_n \mid n \in \mathbb{N}\}$ straff. Nach dem Satz von Prokhorov ist die Folge auch relativ kompakt. Sei $(\mu_n')_{n\in\mathbb{N}}$ eine konvergente Folge in $\{\mu_n \mid n \in \mathbb{N}\}$ und sei $\mu' \in \mathcal{P}(\mathbb{R}^d)$ der Grenzwert. Dann gilt auch $\mu_n' \circ g_x^{-1} \longrightarrow \mu' \circ g_x^{-1}$ schwach für alle $x \in \mathbb{R}^d$. Da $\mu_n' \circ g_x^{-1}$ eine Teilfolge von $\mu_n \circ g_x^{-1}$ ist, konvergiert $\mu_n' \circ g_x^{-1}$ gegen denselben Grenzwertwie $\mu_n \circ g_x^{-1}$, d.h. $\mu \circ g_x^{-1} = \mu' \circ g_x^{-1}$ für alle $x \in \mathbb{R}^d$. Wegen

$$\widehat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{i\xi\cdot x}\mu(dx) = \int_{\mathbb{R}} e^{it}\mu \circ g_\xi(dt) = \widehat{\mu'}(\xi), \quad \xi \in \mathbb{R}^d$$

folgt $\mu = \mu'$. Die Behauptung folgt aus Lemma 3.13. □

Satz 4.20. *(Levys Stetigkeitssatz) Seien $(\mu_n)_{n\in\mathbb{N}}$ Wahrscheinlichkeitsmaße auf $\mathbb{R}^d$. Dann gelten die folgenden Aussagen.*

(a) Ist $\mu_n \longrightarrow \mu$ schwach, so gilt auch $\widehat{\mu_n} \longrightarrow \widehat{\mu}$ punktweise auf $\mathbb{R}^d$.

(b) Falls es eine in 0 stetige Funktion $f : \mathbb{R}^d \longrightarrow \mathbb{C}$ gibt derart, dass $\widehat{\mu}_n(\xi) \longrightarrow f(\xi)$ für alle $\xi \in \mathbb{R}^d$. Dann gibt es ein Wahrscheinlichkeitsmaß μ auf $\mathbb{R}^d$ mit $\widehat{\mu} = f$ und $\mu_n \longrightarrow \mu$ schwach.

Insbesondere gilt für jede Folge $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$ und $\mu \in \mathcal{P}(\mathbb{R}^d)$

$$\mu_n \longrightarrow \mu \text{ schwach } \iff \widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi), \quad \forall \xi \in \mathbb{R}^d.$$

Beweis. (a) Klar, da $x \longmapsto e^{i\xi\cdot x}$ stetig und beschränkt ist.
(b) Wir zeigen nur den Fall $d = 1$.
Schritt 1: $\{\mu_n \mid n \in \mathbb{N}\}$ ist straff.
Für alle $\alpha > 0$ gilt

$$\begin{aligned}\int_{-\alpha}^{\alpha} \widehat{\mu}_n(t)dt &= \int_{-\alpha}^{\alpha}\int_{\mathbb{R}} e^{itx}\mu_n(dx)dt = \int_{\mathbb{R}}\int_{-\alpha}^{\alpha} e^{itx}dt\mu_n(dx) \\ &= \int_{\mathbb{R}} \frac{e^{i\alpha x} - e^{-i\alpha x}}{ix}\mu_n(dx) = \int_{\mathbb{R}} \frac{2\sin(\alpha x)}{x}\mu_n(dx).\end{aligned}$$

Es folgt

$$\begin{aligned}\int_{-\alpha}^{\alpha}(1 - \widehat{\mu}_n(t))dt &= 2\alpha - \int_{\mathbb{R}} \frac{2\sin(\alpha x)}{x}\mu_n(dx) \\ &= 2\alpha\left(1 - \int_{\mathbb{R}} \frac{\sin(\alpha x)}{\alpha x}\mu_n(dx)\right) = 2\alpha\int_{\mathbb{R}}\left(1 - \frac{\sin(\alpha x)}{\alpha x}\right)\mu_n(dx).\end{aligned}$$

Der Integrand ist nicht-negativ und es gilt

$$1 - \frac{\sin(\alpha x)}{\alpha x} \geq \frac{1}{2}, \quad |\alpha x| \geq 2.$$

Daraus folgt

$$\int_{-\alpha}^{\alpha} (1 - \widehat{\mu}_n(t)) dt \geq \alpha \mu_n \left(\left[-\frac{2}{\alpha}, \frac{2}{\alpha} \right]^c \right).$$

Mit $\beta = \frac{2}{\alpha}$ folgt

$$\mu_n([-\beta, \beta]^c) \leq \frac{1}{\alpha} \int_{-\alpha}^{\alpha} (1 - \widehat{\mu}_n(t)) dt = \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu}_n(t)) dt.$$

Nach Voraussetzung ist f stetig in 0 und $f(0) = \lim_{n\to\infty} \widehat{\mu}_n(0) = 1$. Also gibt es zu $\varepsilon > 0$ ein $\beta > 0$ mit

$$|1 - f(t)| \leq \frac{\varepsilon}{4}, \quad |t| \leq \frac{2}{\beta}.$$

Dann ist

$$\lim_{n\to\infty} \frac{2}{\beta} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu}_n(t)) dt = \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - f(t)) dt \leq \frac{\varepsilon}{4} \frac{\beta}{2} \frac{4}{\beta} = \frac{\varepsilon}{2}$$

und es existiert ein $n_0(\varepsilon) = n_0 \in \mathbb{N}$ mit

$$\left| \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu}_n(t)) dt - \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - f(t)) dt \right| < \frac{\varepsilon}{2}, \quad n \geq n_0.$$

Damit ist also $\frac{2}{\beta} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu}_n(t)) dt < \varepsilon$ für alle $n \geq n_0$, d.h.

$$\mu_n([-\beta, \beta]^c) \leq \varepsilon, \quad n \geq n_0.$$

Für $n < n_0$ gibt es ein $\beta_n > 0$ mit $\mu_n([-\beta_n, \beta_n]^c) \leq \varepsilon$. Sei $\beta_0 := \max\{\beta, \beta_1, \ldots, \beta_{n_0-1}\}$, so gilt

$$\mu_n([-\beta_0, \beta_0]^c) \leq \varepsilon, \quad n \geq 1.$$

Schritt 2. Sei $(\mu_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge von $(\mu_n)_{n\in\mathbb{N}}$. Dann ist $\{\mu_{n_k} \mid k \in \mathbb{N}\}$ straff und damit nach dem Satz von Prokhorov auch relativ kompakt. Folglich gibt es eine Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ welche schwach gegen μ' konvergiert. Damit folgt

$$f(\xi) = \lim_{l\to\infty} \widehat{\mu}_{n_{k_l}}(\xi) = \widehat{\mu'}(\xi), \quad \xi \in \mathbb{R}.$$

Die Behauptung folgt aus Lemma 3.13. □

4.3 Charakteristische Funktion für Zufallsvariablen

Definition 4.21. *Für eine Zufallsvariable X mit Werten in $\mathbb{R}^d$ ist die charakteristische Funktion definiert durch*

$$\varphi_X(\xi) := \mathbb{E}(e^{i\xi\cdot X}) = \widehat{\mathbb{P}\circ X^{-1}}(\xi), \quad \xi \in \mathbb{R}^d.$$

Korollar 4.22. *Sei X eine Zufallsvariable mit Werten in $\mathbb{R}^d$. Dann gelten die folgenden Aussagen:*

(a) Ist $\varphi_X \in \mathcal{L}^1(\mathbb{R}^d)$, so hat X eine Dichte $p \in C_b(\mathbb{R}^d)$, welche im Unendlichen verschwindet und gegeben ist durch

$$p(x) = \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x}\varphi_X(\xi)m(d\xi).$$

(b) Ist $\varphi_X \in \mathcal{L}^1(\mathbb{R}^d)$ und gilt für ein $m \geq 1$

$$\int\limits_{\mathbb{R}^d} |\xi|^m|\varphi_X(\xi)|m(d\xi) < \infty,$$

so hat p stetige Ableitungen bis zur Ordnung m und für $0 \leq k \leq m$, $1 \leq j_1, \dots, j_k \leq d$ gilt

$$\frac{\partial^k p(x)}{\partial x_{j_1}\cdots\partial x_{j_k}} = \frac{(-i)^k}{(2\pi)^d}\int\limits_{\mathbb{R}^d} \xi_{j_1}\cdots\xi_{j_k}e^{-i\xi\cdot x}\varphi_X(\xi)m(d\xi).$$

(c) Es gebe ein $m \in \mathbb{N}$ derart, dass

$$\mathbb{E}(|X|^m) < \infty.$$

Dann hat φ_X stetige Ableitungen bis zur Ordnung m und es gilt für alle $0 \leq k \leq m$, $1 \leq j_1, \dots, j_k \leq d$

$$\frac{\partial^k \varphi_X(\xi)}{\partial\xi_{j_1}\cdots\partial\xi_{j_k}} = i^k\mathbb{E}(X_{j_1}\cdots X_{j_k}e^{i\xi\cdot X}).$$

Insbesondere gilt für die Momente von X die Darstellung

$$\mathbb{E}(X_{j_1}\cdots X_{j_k}) = \frac{1}{i^k}\frac{\partial^k\varphi_X(0)}{\partial\xi_{j_1}\cdots\partial\xi_{j_k}}.$$

Satz 4.23. *Seien X, Y unabhängige Zufallsvariablen mit Werten in $\mathbb{R}^d$. Dann gilt*

$$\varphi_X(\xi)\varphi_Y(\xi) = \varphi_{X+Y}(\xi), \quad \xi \in \mathbb{R}^d.$$

Beweis. Klar. □

Satz 4.24. *Sei X eine Zufallsvariable mit Werten in $\mathbb{R}^d$ sowie Y eine Zufallsvariable mit Werten in $\mathbb{R}^m$. Seien $\varphi_X, \varphi_Y, \varphi_{(X,Y)}$ die dazugehörigen charakteristischen Funktionen. Genau dann sind X, Y unabhängig, wenn*

$$\varphi_X(\xi_1)\varphi_Y(\xi_2) = \varphi_{(X,Y)}(\xi_1,\xi_2), \quad \xi_1 \in \mathbb{R}^d,\ \xi_2 \in \mathbb{R}^m. \tag{4.5}$$

Beweis. Seien X, Y unabhängig. Dann ist $\mathbb{P} \circ (X,Y)^{-1} = \mathbb{P} \circ X^{-1} \otimes \mathbb{P} \circ Y^{-1}$ und folglich

$$\begin{aligned}
\varphi_{(X,Y)}(\xi_1, \xi_2) &= \int\limits_{\mathbb{R}^d \times \mathbb{R}^m} e^{i\xi_1 \cdot x} e^{i\xi_2 \cdot x} \mathbb{P} \circ (X,Y)^{-1}(dx, dy) \\
&= \int\limits_{\mathbb{R}^d} e^{i\xi_1 \cdot x} \mathbb{P} \circ X^{-1}(dx) \int\limits_{\mathbb{R}^m} e^{i\xi_2 \cdot y} \mathbb{P} \circ Y^{-1}(dy) \\
&= \varphi_X(\xi_1) \varphi_Y(\xi_2).
\end{aligned}$$

Umgekehrt gelte (4.5). Seien X', Y' unabhängige Zufallsvariablen mit charakteristischen Funktionen φ_X sowie φ_Y. Sei $\varphi_{(X',Y')}$ die charakteristische Funktion von (X', Y'). Dann gilt

$$\varphi_{(X',Y')}(\xi_1, \xi_2) = \varphi_{X'}(\xi_1)\varphi_{Y'}(\xi_2) = \varphi_X(\xi_1)\varphi_Y(\xi_2) = \varphi_{(X,Y)}(\xi_1, \xi_2)$$

und folglich

$$\mathbb{P} \circ (X,Y)^{-1} = \mathbb{P} \circ (X', Y')^{-1} = \mathbb{P} \circ X'^{-1} \otimes \mathbb{P} \circ Y'^{-1} = \mathbb{P} \circ X^{-1} \otimes \mathbb{P} \circ Y^{-1}.$$

□

Korollar 4.25. *Sei $X = (X_1, \dots, X_d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable. Genau dann sind $X_1, \dots, X_d$ unabhängig, wenn*

$$\varphi_X(\xi) = \varphi_{X_1}(\xi_1) \cdots \varphi_{X_d}(\xi_d), \quad \forall \xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d.$$

Bemerkung 4.26. *Sei X eine $\mathbb{R}^d$-wertige Zufallsvariable, A eine $m \times d$ Matrix und $b \in \mathbb{R}^m$. Dann hat $Y := AX + b$ die charakteristische Funktion*

$$\varphi_Y(\xi) = e^{i\xi \cdot b} \varphi_X(A^t \xi), \quad \xi \in \mathbb{R}^m$$

Beispiel 4.27. *(a) Cauchy-Verteilung: Seien X, Y unabhängig und Cauchy verteilt mit Parameter $c > 0$, d.h. diese haben die Verteilung*

$$\mu(dx) = \frac{1}{\pi} \frac{c}{c^2 + x^2} m(dx).$$

Sei $\lambda \in (0,1)$. Dann gilt

$$\varphi_{\lambda X + (1-\lambda)Y}(\xi) = \varphi_X(\lambda \xi) \varphi_Y((1-\lambda)\xi) = e^{-c\lambda|\xi|} e^{-c(1-\lambda)|\xi|} = e^{-c|\xi|}.$$

Also ist $\lambda X + (1-\lambda)Y$ wieder Cauchy verteilt mit Parameter c.

(b) Seien X, Y unabhängig und normalverteilt, d.h. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ mit $\mu_X, \mu_Y \in \mathbb{R}$ und $\sigma_X^2, \sigma_Y^2 > 0$. Dann ist

$$\varphi_{X+Y}(\xi) = \varphi_X(\xi)\varphi_Y(\xi) = e^{i\mu_X \xi - \sigma_X^2 \frac{\xi^2}{2}} e^{i\mu_Y \xi - \sigma_Y^2 \frac{\xi^2}{2}} = e^{i(\mu_X + \mu_Y)\xi - (\sigma_X^2 + \sigma_Y^2)\frac{\xi^2}{2}}.$$

Also ist $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Als nächstes zeigen wir, dass die Kovarianzmatrix und der Erwartungswert die Normalverteilung eindeutig festlegen.

Satz 4.28. *Für jedes $b \in \mathbb{R}^d$ und jede positiv (semi-)definite, symmetrische Matrix Σ gibt es genau eine Normalverteilung mit Erwartungswert b und Kovarianzmatrix Σ.*

Für den Existenzbeweis brauchen wir folgendes Lemma aus der Linearen Algebra.

Lemma 4.29. *Sei Σ eine symmetrische, positiv (semi-)definite reellwertige $d \times d$-Matrix. Dann gibt es eine reellwertige Matrix A mit $\Sigma = AA^t$.*

Beweis. Jede symmetrische, reellwertige Matrix ist orthogonal diagonalisierbar. Es gibt daher eine orthogonale Matrix S mit

$$S\Sigma S^t = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$$

wo $\lambda_1, \ldots, \lambda_d$ die Eigenwerte von Σ bezeichnen. Da Σ positiv (semi-)definit ist, sind die Eigenwerte nicht-negativ. Definiere eine neue Matrix

$$D^{\frac{1}{2}} = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d}).$$

Dann gilt $(D^{\frac{1}{2}})^t = D^{\frac{1}{2}}$ und $D^{\frac{1}{2}}(D^{\frac{1}{2}})^t = D$. Sei $A := SD^{\frac{1}{2}}S^t$. Dann gilt

$$AA^t = SD^{\frac{1}{2}}S^t(SD^{\frac{1}{2}}S^t)^t = SD^{\frac{1}{2}}S^t(S^t)^t(D^{\frac{1}{2}})^tS^t = SDS^t = \Sigma^T = \Sigma.$$

$\square$

Beweis. (Satz 4.28) Ist μ eine Normalverteilung mit Erwartungswert b und Kovarianzmatrix Σ, so gilt

$$\widehat{\mu}(\xi) = e^{i\langle \xi, b\rangle} e^{-\frac{1}{2}\langle \xi, \Sigma\xi\rangle}, \quad \xi \in \mathbb{R}^d.$$

Dieses zeigt die Eindeutigkeit. Für die Existenz wähle A mit $\Sigma = AA^t$ und benutze die Definition der Normalverteilung. $\square$

Satz 4.30. *Sei $X = (X_1, \ldots, X_d)$ eine normalverteilte Zufallsvariable mit Erwartungswert $b \in \mathbb{R}^d$ und Kovarianzmatrix Σ. Dann gelten die folgenden Aussagen:*

(a) Sei $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$ und definiere

$$Y := \sum_{k=1}^{d} a_k X_k = \langle a, X\rangle.$$

Dann ist Y eine normalverteilte Zufallsvariable mit Erwartungswert $\langle a, b\rangle$ und Varianz $\langle a, \Sigma a\rangle$.

(b) X_j sind für $j = 1, \ldots, d$ normalverteilt mit Erwartungswert b_j und Varianz Σ_{jj}.

(c) Sind die X_j paarweise unkorreliert, so ist $(X_1, \ldots, X_d)$ unabhängig.

Umgekehrt, seien $X_1, \ldots, X_d$ unabhängige, normalverteilte Zufallsvariablen mit Erwartungswert $b_j \in \mathbb{R}$ und Varianz σ_j^2. Dann ist $X = (X_1, \ldots, X_d)$ normalverteilt mit Erwartungswert $b = (b_1, \ldots, b_d) \in \mathbb{R}^d$ un Kovarianzmatrix $\Sigma_{ij} = \delta_{ij}\sigma_j^2$.

Beweis. (a) $\square$

5 Grenzwertsätze

5.1 Gesetze der großen Zahlen

Sei $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1$ eine Folge von Zufallsvariablen, $m_n := \mathbb{E}(X_n)$, $M_n := \sum\limits_{k=1}^{n} m_k$ die Erwartungswerte und $S_n := \sum\limits_{k=1}^{n} X_k$.

Definition 5.1. *Die Folge $(X_n)_{n\in\mathbb{N}}$ erfüllt das schwache Gesetz der großen Zahlen, wenn*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}|S_n - M_n| \geq \varepsilon\right) = 0, \quad \forall\varepsilon > 0.$$

Die Folge X_n erfüllt das starke Gesetz der großen Zahlen, wenn $\frac{S_n - M_n}{n} \longrightarrow 0$ fast sicher.

Sehen wir jede Variable X_n als ein Zufallsexperiment mit Ausgang X_n an, so ist $\frac{1}{n}S_n$ der Mittelwert der Ausgänge der ersten n Experimente. Folglich ist $\frac{1}{n}M_n$ der erwartete (deterministische!) Mittelwert von $\frac{1}{n}S_n$. Das Gesetz der großen Zahlen besagt, dass nach hinreichend vielen Experimenten die Ausgänge der Experimente sich im Mittel deterministisch Verhalten.

Satz 5.2. *Seien $X_n \in \mathcal{L}^2$ unkorreliert und es gelte $\operatorname{var}(X_n) \leq C < \infty$ für alle $n \in \mathbb{N}$ und eine Konstante $C > 0$. Dann gilt*

$$\frac{1}{n^2}\mathbb{E}((S_n - M_n)^2) \longrightarrow 0, \quad n \to \infty.$$

Insbesondere gilt das schwache Gesetz der großen Zahlen.

Beweis. Es gilt $\mathbb{E}(S_n) = M_n$ und folglich

$$\begin{aligned}\frac{1}{n^2}\mathbb{E}((S_n - M_n)^2) &= \frac{1}{n^2}\mathbb{E}\left((S_n - \mathbb{E}(S_n))^2\right) = \frac{1}{n}\operatorname{var}(S_n)\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\operatorname{var}(X_k) \leq \frac{C}{n}.\end{aligned}$$

Aus der Tschebyscheff Ungleichung folgt

$$\mathbb{P}\left(|S_n - M_n| \geq \varepsilon n\right) \leq \frac{1}{\varepsilon^2 n^2}\mathbb{E}((S_n - M_n)^2)$$

und somit die Behauptung. □

Im folgenden wollen wir zeigen, dass unter denselben Voraussetzungen bereits das starke Gesetz der großen Zahlen gilt.

Lemma 5.3. *(stochastisch schnelle Konvergenz) Sei X_n eine Folge von Zufallsvariablen mit*

$$\sum_{k=1}^{\infty}\mathbb{P}(|X_k| \geq \varepsilon) < \infty, \quad \forall\varepsilon > 0.$$

Dann gilt $\mathbb{P}\left(\lim\limits_{n\to\infty} X_n = 0\right) = 1$.

Beweis. Nach dem ersten Borel-Cantelli Lemma gilt

$$\mathbb{P}\left(\limsup_{n\to\infty}\{|X_n| \geq \varepsilon\}\right) = \mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{k\geq n}\{|X_k| \geq \varepsilon\}\right) = 0.$$

Insbesondere gilt

$$1 = \mathbb{P}\left(\left(\limsup_{n\to\infty}\{|X_n| \geq \varepsilon\}\right)^c\right) = \mathbb{P}\left(\bigcup_{n\geq 1}\bigcap_{k\geq n}\{|X_k| < \varepsilon\}\right).$$

D.h. für fast alle ω gibt es ein $n(\omega) \in \mathbb{N}$ sodass für alle $k \geq n(\omega)$ gilt $|X_k(\omega)| < \varepsilon$. □

Satz 5.4. *Seien $X_n \in \mathcal{L}^2$ paarweise unkorreliert mit $\mathrm{var}(X_n) \leq C < \infty$ für ein $C > 0$ und alle $n \geq 1$. Dann gilt das starke Gesetz der großen Zahlen.*

Beweis. Wir betrachten zuerst den Fall $\mathbb{E}(X_n) = 0$. In diesem Fall ist $M_n = 0$.
Schritt 1. Aus der Tschebyscheff Ungleichung folgt

$$\mathbb{P}(|S_{k^2}| \geq \varepsilon k^2) \leq \frac{1}{\varepsilon^2}\frac{1}{k^4}\mathrm{var}((S_{k^2})^2) = \frac{1}{\varepsilon^2 k^4}\sum_{j=1}^{k^2}\mathrm{var}(X_j) \leq \frac{C}{\varepsilon^2}\frac{1}{k^2}$$

Insbesondere gilt

$$\sum_{k=1}^{\infty}\mathbb{P}(|S_{k^2}| \geq k^2\varepsilon) \leq \frac{C}{\varepsilon^2}\sum_{k=1}^{\infty}\frac{1}{k^2} < \infty, \quad \forall \varepsilon > 0.$$

Nach Lemma 5.3 gibt es $N_1 \in \mathcal{F}$ mit $\mathbb{P}(N_1) = 0$ und

$$\lim_{k\to\infty} S_{k^2}(\omega) = 0, \quad \omega \in \Omega\backslash N_1.$$

Schritt 2. Definiere eine neue Zufallsvariable

$$D_{k^2} := \max_{k^2 < l \leq (k+1)^2}\left|\sum_{j=1}^{l} X_j - \sum_{j=1}^{k^2} X_j\right| = \max_{k^2 < l \leq (k+1)^2}\left|\sum_{j=k^2+1}^{l} X_j\right|.$$

Für jedes $\varepsilon > 0$ folgt aus der Tschebyscheff Ungleichung

$$\begin{aligned}
\mathbb{P}\left(\frac{D_{k^2}}{k^2} \geq \varepsilon\right) &= \mathbb{P}\left(\bigcup_{l=k^2+1}^{(k+1)^2}\left\{\left|\sum_{j=k^2+1}^{l} X_j\right| \geq \varepsilon k^2\right\}\right)\\
&\leq \sum_{l=k^2+1}^{(k+1)^2}\mathbb{P}\left(\left|\sum_{j=k^2+1}^{l} X_j\right| \geq \varepsilon k^2\right) \leq \frac{1}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2}\mathrm{var}\left(\sum_{j=k^2+1}^{l} X_j\right)\\
&\leq \frac{C}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2}(l - k^2) \leq \frac{C}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2}\left((k+1)^2 - k^2\right)\\
&= \frac{C}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2}(2k+1) \leq \frac{4C}{\varepsilon^2 k^3}\left((k+1)^2 - k^2\right) = \frac{8C}{\varepsilon^2 k^2}.
\end{aligned}$$

Nach Lemma 5.3 gibt es $N_2 \in \mathcal{F}$ mit $\mathbb{P}(N_2) = 0$ und

$$\lim_{k\to\infty} \frac{D_{k^2}(\omega)}{k^2} = 0, \quad \omega \in \Omega \backslash N_2.$$

Schritt 3. Für $n \in \mathbb{N}$ sei $k = k(n) \in \mathbb{N}$ mit $k^2 \le n < (k+1)^2$. Dann gilt für alle $\omega \in \Omega \backslash (N_1 \cup N_2)$

$$|S_n(\omega)| \le \frac{1}{n}\left|\sum_{j=1}^{n} X_k(\omega)\right| \le S_{k^2}(\omega) + \frac{D_{k^2}(\omega)}{k^2}.$$

Schritt 4. Für den allgemeinen Fall sei $Y_n := X_n - \mathbb{E}(X_n)$. Dann ist $Y_n \in \mathcal{L}^2$ mit $\mathbb{E}(Y_n) = 0$ und

$$\operatorname{var}(Y_n) = \operatorname{var}(X_n) \le C < \infty, \quad n \in \mathbb{N}.$$

Ferner gilt $\operatorname{cov}(Y_n, Y_m) = \operatorname{cov}(X_n, X_m)$ und

$$S_n - M_n = \sum_{k=1}^{n} (X_k - \mathbb{E}(X_k)) = \sum_{k=1}^{n} Y_k.$$

Die Behauptung folgt aus Schritt 1 – 3. □

Im folgenden wollen wir die Bedingung an die Momente

$$\operatorname{var}(X_n) \le C < \infty, \quad n \in \mathbb{N}$$

weiter abschwächen. Hierfür brauchen wir das zweite Borel-Cantelli Lemma.

Lemma 5.5. *(Borel-Cantelli Lemma 2) Sei $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ eine Folge unabhängiger Mengen und es gelte $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. Dann gilt $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\bigcap_{n\ge 1} \bigcup_{k\ge n} A_k\right) = 1$.*

Beweis. Es gilt

$$\mathbb{P}((\limsup_{n\to\infty} A_n)^c) = \mathbb{P}\left(\bigcup_{n\ge 1} \bigcap_{k\ge n} A_k^c\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\ge n} A_k^c\right).$$

Folglich reicht es zu zeigen, dass $\mathbb{P}\left(\bigcap_{k\ge n} A_k^c\right) = 0$ für $n \ge 1$. Dieses folgt aus

$$\mathbb{P}\left(\bigcap_{k\ge n} A_k^c\right) = \prod_{k=n}^{\infty} \mathbb{P}(A_k^c) = \prod_{k=n}^{\infty} (1 - \mathbb{P}(A_k)) \le \prod_{k=n}^{\infty} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{\infty} \mathbb{P}(A_k)} = 0.$$

□

Lemma 5.6. *(Kolmogorov Ungleichung) Seien $X_n \in \mathcal{L}^2$ unabhängig. Dann gilt*

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k - M_k| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2} \sum_{k=1}^{n} \operatorname{var}(X_k) = \frac{1}{\varepsilon^2} \operatorname{var}(S_n).$$

Beweis. Betrachte die Ereignisse

$$C_k = \left\{ \left| \sum_{j=1}^{i} X_j - \sum_{j=1}^{i} m_j \right| < \varepsilon,\ 1 \le i < k,\ \left| \sum_{j=1}^{k} X_j - \sum_{j=1}^{k} m_j \right| \ge \varepsilon \right\}.$$

Dann gilt $C_k \cap C_j = \varnothing$ für $k \neq j$ und $C = \bigcup_{k=1}^{n} C_k$ erfüllt

$$\mathbb{P}\left(\max_{1 \le k \le n} |S_k - M_k| \ge \varepsilon \right) = \mathbb{P}(C).$$

Es gilt

$$\begin{aligned}
\operatorname{var}(S_n) &= \int_\Omega \left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i \right)^2 d\mathbb{P} \\
&\ge \int_C \left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i \right)^2 d\mathbb{P} = \sum_{k=1}^{n} \int_{C_k} \left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i \right)^2 d\mathbb{P} \\
&= \sum_{k=1}^{n} \int_{C_k} \left(\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} m_i \right)^2 d\mathbb{P} \\
&\quad + 2 \sum_{k=1}^{n} \int_{C_k} \left(\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} m_i \right) \left(\sum_{i=k+1}^{n} X_i - \sum_{i=k+1}^{n} m_i \right) d\mathbb{P} \\
&\quad + \sum_{k=1}^{n} \int_{C_k} \left(\sum_{i=k+1}^{n} X_i - \sum_{i=k+1}^{n} m_i \right)^2 d\mathbb{P}.
\end{aligned}$$

Die Zufallsvariablen $\eta_1 = \mathbb{1}_{C_k} \left(\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} m_i \right)$ und $\eta_2 = \sum_{i=k+1}^{n} X_i - \sum_{i=k+1}^{n}$ sind nach Voraussetzung unabhängig. Es gilt folglich

$$\mathbb{E}(\eta_1 \eta_2) = \mathbb{E}(\eta_1)\mathbb{E}(\eta_2) = 0.$$

Damit erhalten wir

$$\operatorname{var}(S_n) \ge \sum_{k=1}^{n} \int_{C_k} \left(\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} m_i \right)^2 d\mathbb{P} \ge \varepsilon^2 \sum_{k=1}^{n} \mathbb{P}(C_k) = \varepsilon^2 \mathbb{P}(C).$$

□

Lemma 5.7. *Es gilt*

$$\sum_{k:\ 2^k \ge i} \frac{1}{2^{2k}} \le \frac{16}{3} \frac{1}{i^2}, \quad i \in \mathbb{N}.$$

Beweis. Sei $q \in (0,1)$ und $n_0 \in \mathbb{N}$. Dann gilt

$$\sum_{n=n_0}^{\infty} q^n = q^{n_0} \sum_{n=n_0}^{\infty} q^{n-n_0} = \frac{q^{n_0}}{1-q}.$$

Für $q = \frac{1}{4}$ und $n_0 \in \mathbb{N}$ mit $\frac{\ln(i)}{\ln(2)} - 1 < n_0 \leq \frac{\ln(i)}{\ln(2)}$ folgt

$$\sum_{k:\ 2^k \geq i} \frac{1}{2^{2k}} \leq \sum_{k=n_0}^{\infty} 4^{-k} = \frac{4^{-n_0}}{1-\frac{1}{4}} \leq \frac{4}{3} \exp\left(-\left(\frac{\ln(i)}{\ln(2)} - 1\right) 2\ln(2)\right) = \frac{16}{3} \frac{1}{i^2}.$$

$\square$

Satz 5.8. *(Erstes Kolmogorov Theorem) Seien $X_n \in \mathcal{L}^2$ unabhängig und es gelte*

$$\sum_{n=1}^{\infty} \frac{\operatorname{var}(X_n)}{n^2} < \infty.$$

Dann gilt das starke Gesetz der großen Zahlen.

Beweis. Wir können ohne Einschränkung der Allgemeinheit annehmen, dass $\mathbb{E}(X_n) = 0$ für alle $n \in \mathbb{N}$ gilt. Sei $\varepsilon > 0$ und

$$\begin{aligned} B(\varepsilon) &= \{\omega \in \Omega \mid \exists N(\varepsilon) \in \mathbb{N} : \quad |S_n(\omega)| < n\varepsilon, \quad \forall n \geq N(\varepsilon)\} \\ &= \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|S_n| < n\varepsilon\}. \end{aligned}$$

Setze $B_k(\varepsilon) = \left\{ \max\limits_{2^{k-1} \leq n < 2^k} \frac{1}{n} |S_n| \geq \varepsilon \right\}$. Die Kolmogorov Ungleichung impliziert

$$\begin{aligned} \mathbb{P}(B_k(\varepsilon)) &= \mathbb{P}\left(\max_{2^{k-1} \leq n < 2^k} \frac{1}{n} \left| \sum_{i=1}^{n} X_i \right| \right) \leq \mathbb{P}\left(\max_{2^{k-1} \leq n < 2^k} \left| \sum_{i=1}^{n} X_i \right| \geq \varepsilon 2^{k-1} \right) \\ &\leq \mathbb{P}\left(\max_{1 \leq n \leq 2^k} \left| \sum_{i=1}^{n} X_i \right| \geq \varepsilon 2^{k-1} \right) \leq \frac{1}{\varepsilon^2 2^{2k-2}} \sum_{i=1}^{2^k} \operatorname{var}(X_i). \end{aligned}$$

Daraus folgt

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}(B_k(\varepsilon)) &\leq \sum_{k=1}^{\infty} \frac{1}{\varepsilon^2 2^{2k-2}} \sum_{i=1}^{2^k} \operatorname{var}(X_i) \\ &= \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \operatorname{var}(X_i) \sum_{k:\ 2^k \geq i} \frac{1}{2^{2k-2}} \leq \frac{32}{3\varepsilon^2} \sum_{i=1}^{\infty} \frac{\operatorname{var}(X_i)}{i^2}. \end{aligned}$$

Nach dem ersten Borel-Cantelli Lemma gibt es für fast alle ω ein $k_0(\omega) \in \mathbb{N}$ mit

$$\max_{2^{k-1} \leq n < 2^k} \frac{1}{n} |S_n(\omega)| < \varepsilon, \quad k \geq k_0(\omega).$$

Insbesondere gilt $\mathbb{P}(B(\varepsilon)) = 1$ für alle $\varepsilon > 0$. Also gilt auch $\mathbb{P}(B(m^{-1})) = 1$ für alle $m \geq 1$ und somit $\mathbb{P}\left(\bigcap_{m\geq 1} B(m^{-1})\right) = 1$. Für jedes $\omega \in \bigcap_{m\geq 1} B(m^{-1})$ gibt es ein $N(\omega, m) \in \mathbb{N}$ sodass für alle $n \geq N(\omega, m)$ gilt $\frac{1}{n}|S_n(\omega)| < \frac{1}{m}$. □

Eine Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ heißt identisch verteilt, wenn $\mathbb{P} \circ X_n^{-1} = \mathbb{P} \circ X_1^{-1}$ für alle $n \geq 1$ gilt. In diesem Fall setze $m := \mathbb{E}(X_n) = \mathbb{E}(X_1)$. Das starke Gesetz der großen Zahlen ist äquivalent zu

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow m, \quad n \to \infty$$

fast sicher. Wollen wir die Integrabilität noch weiter abschwächen, z.B. $X_n \in \mathcal{L}^1$, so müssen wir annehmen dass die Folge X_n identisch verteilt ist.

Satz 5.9. *(Zweites Kolmogorov Theorem, [KS07]) Seien $X_n \in \mathcal{L}^1$ unabhängig und identisch verteilt. Dann gilt das starke Gesetz der großen Zahlen.*

Satz 5.10. *(Satz von Etemadi, 1981, [Ete81]) Seien $X_n \in \mathcal{L}^1$ paarweise unabhängig und identisch verteilt. Dann gilt das starke Gesetz der großen Zahlen.*

5.2 Zentraler Grenzwertsatz

Sei $X_n \in \mathcal{L}^2$ eine Folge von unabhängigen Zufallsvariablen mit Erwartungswert $m_k := \mathbb{E}(X_k)$ und Varianz $\sigma_i^2 := \mathrm{var}(X_i)$. Definiere $S_n := \sum_{i=1}^{n} X_i$, $M_n := \sum_{i=1}^{n} m_i$ und $D_n^2 := \mathrm{var}(S_n) = \sum_{i=1}^{n} \sigma_i^2$.

Definition 5.11. *(Lindeberg Bedingung) Die Folge $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^2$ erfüllt die Lindeberg Bedingung, wenn für jedes $\varepsilon > 0$ gilt*

$$\lim_{n\to\infty} \frac{1}{D_n^2} \sum_{i=1}^{n} \int\limits_{|X_i - m_i| \geq \varepsilon D_n} (X_i - m_i)^2 d\mathbb{P} = 0.$$

Man beachte, ist $D_n = 0$ für alle $n \in \mathbb{N}$, so ist $X_n = 0$ fast sicher für alle $n \in \mathbb{N}$. Um diesen Fall auszuschließen, können wir annehmen, dass $D_n > 0$ für hinreichend große $n \in \mathbb{N}$ gilt. Wir werden später, unter geeigneten Bedingungen, zeigen dass $D_n \longrightarrow \infty$ gilt. Vorher brauchen wir jedoch das nächste Lemma.

Lemma 5.12. *Sei* ln *die Fortsetzung des natürlichen Logarithmus auf das Gebiet*

$$G = \left\{1 + z \in \mathbb{C} \mid |z| < \frac{1}{4}\right\},$$

d.h. es gilt

$$\ln(1 + z) = \sum_{k=1}^{n} (-1)^{k+1} \frac{z^k}{k}, \quad |z| < \frac{1}{4}.$$

Dann gibt es eine stetige Funktion $\Theta : \{z \in \mathbb{C} \mid |z| < \frac{1}{4}\} \longrightarrow \mathbb{C}$ mit $|\Theta(z)| \leq 1$ und

$$\ln(1 + z) = z + \Theta(z)|z|^2, \quad |z| < \frac{1}{4}.$$

Beweis. Für $z = 0$ ist nichts zu zeigen. Wir betrachten $0 < |z| < \frac{1}{4}$. Dann gilt

$$\ln(1+z) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{z^k}{k} = z + |z|^2 \sum_{k=2}^{\infty} (-1)^{k+1} \frac{z^{k-2}}{k} \frac{z^2}{|z|^2}.$$

Setze $\Theta(z) := \sum_{k=2}^{\infty} (-1)^{k+1} \frac{z^{k-2}}{k} \frac{z^2}{|z|^2}$. Wir zeigen, dass die Reihe absolut konvergiert

$$\sum_{k=2}^{\infty} \left| (-1)^{k+1} \frac{z^{k-2}}{k} \frac{z^2}{|z|^2} \right| = \sum_{k=0}^{\infty} \frac{|z|^k}{k+2} \leq \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{4^k} = \frac{2}{3} < 1.$$

□

Satz 5.13. *(zentraler Grenzwertsatz) Sei $X_n \in \mathcal{L}^2$ eine Folge von unabhängigen Zufallsvariablen und es gelte die Lindeberg Bedingung. Dann konvergiert $S_n^* := \frac{S_n - M_n}{D_n}$ schwach gegen die Normalverteilung $\mathcal{N}(0,1)$ auf $\mathbb{R}$.*

Für den Beweis können wir ohne Einschränkung der Allgemeinheit annehmen, dass $m_i = 0$. Anderseits definiere $Y_i := X_i - m_i$. Für diese Zufallsvariablen ist die Lindeberg Bedingung immernoch erfüllt. Im Beweis zeigen wir die schwache Konvergenz mittels Konvergenz der charakteristischen Funktionen. Sei φ_n die charakteristische Funktion von X_n. Das nächste Lemma liefert hierfür eine geeignete Restgliedabschätzung.

Lemma 5.14. *Es gibt Zahlen $a_k^{(n)}$ derart, dass für $1 \leq k \leq n$ gilt*

$$\varphi_k\left(\frac{\xi}{D_n}\right) = 1 - \frac{\xi^2 \sigma_k^2}{2 D_n^2} + a_k^{(n)}$$

sowie

$$\lim_{n\to\infty} \sum_{k=1}^{n} |a_k^{(n)}(\xi)| = 0, \quad \forall \xi \in \mathbb{R}.$$

Beweis. Es gibt stetige Funktionen Θ_1, Θ_2 mit $|\Theta_1(x)|, |\Theta_2(x)| \leq 1$ sowie

$$e^{ix} = 1 + ix + \frac{\Theta_1(x) x^2}{2},$$
$$e^{ix} = 1 + ix - \frac{x^2}{2} + \frac{\Theta_2(x) x^3}{6}.$$

für alle $x \in \mathbb{R}$. Aus diesen Entwicklungen erhalten wir

$$\begin{aligned}
\varphi_k\left(\frac{\xi}{D_n}\right) &= \int\limits_{\Omega} e^{i\frac{\xi}{D_n}X_k} d\mathbb{P} \\
&= \int\limits_{|X_k|\geq\varepsilon D_n} \left(1 + \frac{i\xi}{D_n}X_k + \frac{\Theta_1(X_k)(\xi X_k)^2}{2D_n^2}\right) d\mathbb{P} \\
&\quad + \int\limits_{|X_k|<\varepsilon D_n} \left(1 + \frac{i\xi}{D_n}X_k - \frac{\xi^2}{2D_n^2}X_k^2 + \frac{\Theta_2(X_k)(\xi X_k)^3}{6D_n^3}\right) d\mathbb{P} \\
&= 1 - \frac{\xi^2\sigma_k^2}{2D_n^2} + \frac{\xi^2}{2D_n^2} \int\limits_{|X_k|\geq\varepsilon D_n} (1+\Theta_1(X_k))X_k^2 d\mathbb{P} + \frac{\xi^3}{6D_n^3} \int\limits_{|X_k|<\varepsilon D_n} \Theta_2(X_k)X_k^3 d\mathbb{P}.
\end{aligned}$$

Definiere

$$a_k^{(n)} := \frac{\xi^2}{2D_n^2} \int\limits_{|X_k|\geq\varepsilon D_n} (1+\Theta_1(X_k))X_k^2 d\mathbb{P} + \frac{\xi^3}{6D_n^3} \int\limits_{|X_k|<\varepsilon D_n} \Theta_2(X_k)X_k^3 d\mathbb{P}.$$

Betrachten wir $\sum\limits_{k=1}^{n} |a_k^{(n)}|$ so erhalten wir für den ersten Term aus der Lindeberg Bedingung

$$\frac{\xi^2}{2D_n^2} \sum_{k=1}^{n} \int\limits_{|X_k|\geq\varepsilon D_n} (1+|\Theta_1(X_k)|)X_k^2 d\mathbb{P} \leq \frac{\xi^2}{D_n^2} \sum_{k=1}^{n} \int\limits_{|X_k|\geq\varepsilon D_n} X_k^2 d\mathbb{P} \longrightarrow 0, \quad n \to \infty.$$

Für den zweiten Term erhalten wir

$$\frac{|\xi|^3}{6D_n^3} \sum_{k=1}^{n} \int\limits_{|X_k|<\varepsilon D_n} |\Theta_2(X_k)||X_k|^3 d\mathbb{P} \leq \frac{|\xi|^3}{6D_n^3} \sum_{k=1}^{n} \int\limits_{|X_k|<\varepsilon D_n} \varepsilon D_n |X_k|^2 d\mathbb{P} \leq \frac{\varepsilon|\xi|^3}{6D_n^2} \sum_{k=1}^{n} \sigma_k^2 = \frac{\varepsilon|\xi|^3}{6}.$$

Da ε beliebig war, kann dieser Term beliebig klein gemacht werden. □

Eine weitere wichtige Ungleichung ist im nächsten Lemma formuliert.

Lemma 5.15. *Es gilt*

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{\sigma_k^2}{D_n^2} = 0. \tag{5.1}$$

Beweis. Für jedes $\varepsilon > 0$ gilt

$$\max_{1\leq k\leq n} \frac{\sigma_k^2}{D_n^2} \leq \max_{1\leq k\leq n} \frac{1}{D_n^2} \int\limits_{|X_k|\geq\varepsilon D_n} X_k^2 d\mathbb{P} + \max_{1\leq k\leq n} \frac{1}{D_n^2} \int\limits_{|X_k|\leq\varepsilon D_n} X_k^2 d\mathbb{P}.$$

Der erste Term konvergiert für $n \to \infty$ wegen der Lindeberg Bedingung gegen 0. Für den zweiten Term gilt

$$\max_{1 \le k \le n} \frac{1}{D_n^2} \int\limits_{|X_k| \le \varepsilon D_n} X_k^2 d\mathbb{P} \le \varepsilon^2.$$

Da ε beliebig war, folgt die Behauptung. $\square$

Kommen wir zum Beweis vom zentralen Grenzwertsatz.

Beweis. (Satz 5.13) Es reicht zu zeigen, dass

$$\varphi_{\tau_n}(\xi) \longrightarrow e^{-\frac{\xi^2}{2}}, \quad \xi \in \mathbb{R}.$$

Fixiere $\xi \in \mathbb{R}$. Dann gilt

$$\varphi_{\tau_n}(\xi) = \mathbb{E}(e^{i\xi\tau_n}) = \mathbb{E}\left(e^{i\frac{\xi}{D_n}\sum\limits_{k=1}^{n} X_k}\right) = \prod_{k=1}^{n} \varphi_k\left(\frac{\xi}{D_n}\right).$$

Daraus folgt

$$\ln\left(\varphi_{\tau_n}(\xi)\right) = \sum_{i=1}^{n} \ln\left(\varphi_i\left(\frac{\xi}{D_n}\right)\right)$$

und somit reicht es

$$\sum_{i=1}^{n} \ln\left(\varphi_i\left(\frac{\xi}{D_n}\right)\right) \longrightarrow -\frac{\xi^2}{2}, \quad n \to \infty \tag{5.2}$$

zu zeigen. Für hinreichend große $n \in \mathbb{N}$ ist $|z_{k,n}| < \frac{1}{4}$ für $z_{k,n} = -\frac{\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}$ (siehe (5.1). Daraus folgt

$$\begin{aligned}
\sum_{k=1}^{n} \ln\left(\varphi_k\left(\frac{\xi}{D_n}\right)\right) &= \sum_{k=1}^{n} \ln\left(1 - \frac{\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}\right) \\
&= -\frac{\xi^2}{2D_n^2}\sum_{k=1}^{n}\sigma_k^2 + \sum_{k=1}^{n} a_k^{(n)} + \sum_{k=1}^{n}\Theta_{k,n}\left|\frac{-\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}\right|^2 \\
&= -\frac{\xi^2}{2} + \sum_{k=1}^{n} a_k^{(n)} + \sum_{k=1}^{n}\Theta_{k,n}\left|\frac{-\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}\right|^2
\end{aligned}$$

mit $\Theta k, n = \Theta(z_{k,n})$, $|\Theta_{k,n}| \le 1$. Es reicht daher zu zeigen, dass der letzte Term gegen Null konvergiert. Es gilt

$$\sum_{k=1}^{n} |\Theta_{k,n}| \left|\frac{-\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}\right|^2 \le \max_{1 \le k \le n}\left(\frac{\xi^2\sigma_k^2}{2D_n^2} + |a_k^{(n)}|\right)\sum_{k=1}^{n}\left(\frac{\xi^2\sigma_k^2}{2D_n^2} + |a_k^{(n)}|\right).$$

Die Summe ist wegen

$$\sum_{k=1}^{n}\left(\frac{\xi^2\sigma_k^2}{2D_n^2}+|a_k^{(n)}|\right)=\frac{\xi^2}{2}+\sum_{k=1}^{n}|a_k^{(n)}|$$

offensichtlich beschränkt. Der erste Faktor konvergiert gegen Null, welches die Behauptung zeigt. □

Bemerkung 5.16. *Sei $K \subset \mathbb{R}$ kompakt. Dann gilt mit denselben Notationen wie im letzten Beweis*

$$\lim_{n\to\infty}\sup_{\xi\in K}\left|\varphi_{\tau_n}(\xi)-e^{-\frac{\xi^2}{2}}\right|=0.$$

Der Beweis ist Übung.

Korollar 5.17. *Sei $X_n \in \mathcal{L}^2$ ein Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $m := \mathbb{E}(X_1)$ und $0 < \sigma^2 := \operatorname{var}(X_1)$. Dann konvergiert $\frac{S_n - nm}{\sqrt{n\sigma^2}}$ schwach gegen die Normalverteilung.*

Beweis. Da alle X_n die gleiche Verteilung haben folgt $\sigma_i^2 = \sigma_1^2$ und somit $D_n = \sqrt{n\sigma^2}$. Ebenso gilt $M_n = mn$ und

$$\int\limits_{|X_k-m_k|\geq\varepsilon D_n}(X_k-m_k)^2 d\mathbb{P}=\int\limits_{|X_1-m_1|}(X_1-m_1)^2 d\mathbb{P}.$$

Die Lindeberg Bedingung folgt daher aus

$$\frac{1}{D_n^2}\sum_{k=1}^{n}\int\limits_{|X_k-m_k|\geq\varepsilon D_n}(X_k-m_k)^2 d\mathbb{P}=\frac{1}{\sigma^2}\int\limits_{|X_1-m_1|\geq\varepsilon\sqrt{\sigma^2 n}}(X_1-m_1)^2 d\mathbb{P}$$

und dem Satz von Lebesgue wegen $X_1 - m_1 \in \mathcal{L}^2$. □

Definition 5.18. *(Lyapunov Bedingung) Eine Folge unabhängiger Zufallsvariablen $X_n \in \mathcal{L}^2$ erfüllt die Lyapunov Bedingung, wenn es ein $\delta > 0$ gibt derart, dass*

$$\lim_{n\to\infty}\frac{1}{D_n^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}(|X_k-m_k|^{2+\delta})=0.$$

Korollar 5.19. *(zentraler Grenzwertsatz, Lyapunov Bedingung) Sei $X_n \in \mathcal{L}^2$ eine Folge unabhängiger Zufallsvariablen, welche die Lyapunov Bedingung erfüllen. Dann konvergiert $\frac{S_n - M_n}{D_n}$ schwach gegen die Normalverteilung $\mathcal{N}(0,1)$.*

Beweis. Sei $\varepsilon > 0$ und $\delta > 0$ gegeben durch die Lyapunov Bedingung. Dann gilt

$$\begin{aligned}\frac{1}{D_n^2}\int\limits_{|X_k-m_k|\geq\varepsilon D_n}(X_k-m_k)^2 d\mathbb{P} &\leq \frac{1}{D_n^2(\varepsilon D_n)^\delta}\int\limits_{|X_k-m_k|\geq\varepsilon D_n}(X_k-m_k)^{2+\delta} d\mathbb{P}\\ &\leq \varepsilon^{-\delta}\frac{\mathbb{E}(|X_k-m_k|^{2+\delta})}{D_n^{2+\delta}}.\end{aligned}$$

Folglich gilt die Lindeberg Bedingung. □

Satz 5.20. *(Lindeberg-Feller, [KS07]) Sei $X_n \in \mathcal{L}^2$ eine Folge unabhängiger Zufallsvariablen mit*

$$\lim_{n\to\infty} \max_{1\le i\le n} \frac{\sigma_i^2}{D_n^2} = 0.$$

Dann sind die folgenden Aussagen äquivalent

- *Die Folge X_n erfüllt die Lindeberg Bedingung.*
- *Die Folge $\frac{S_n - M_n}{D_n}$ konvergiert schwach gegen die Normalverteilung $\mathcal{N}(0,1)$.*

5.3 Weitere Grenzwertsätze

Sei $p(x) \geq 0$ eine Wahrscheinlichkeitsdichte. Wir schreiben

$$p(x) \sim q(x), \quad |x| \longrightarrow \infty$$

falls es ein $R > 0$ und eine beschränkte Funktion g gibt mit $g(x) \longrightarrow 0$ für $|x| \longrightarrow \infty$ und

$$p(x) = q(x)(1 + g(x)), \quad |x| \geq R.$$

Lemma 5.21. *Sei p eine Wahrscheinlichkeitsdichte mit $p(x) = p(-x)$,*

$$p(x) \sim \frac{c}{|x|^{\alpha+1}}, \quad |x| \longrightarrow \infty$$

und $0 < \alpha < 2$, $c > 0$. Sei φ die charakteristische Funktion von p. Dann gilt

$$\varphi(\xi) = 1 - c_1|\xi|^\alpha + o(|\xi|^\alpha), \quad \xi \to 0. \tag{5.3}$$

Die Behauptung besagt, dass es eine Konstante c_1 gibt derart, dass

$$\frac{\varphi(\xi) - 1 + c_1|\xi|^\alpha}{|\xi|^\alpha} \longrightarrow 0, \quad \xi \to 0.$$

Äquivalent, es gibt eine Funktion h mit $\frac{h(\xi)}{|\xi|^\alpha} \longrightarrow 0$ für $\xi \to 0$ und es gilt

$$\varphi(\xi) = 1 + c_1|\xi|^\alpha + h(\xi)$$

für alle hinreichend kleinen $\xi \in \mathbb{R}$. Wegen

$$\varphi(-\xi) = \int_{\mathbb{R}} p(x)e^{-i\xi x} m(dx) = \int_{\mathbb{R}} p(-x)e^{i\xi x} m(dx) = \varphi(\xi)$$

reicht es die Asymptotik nur für positive ξ zu betrachten. Im folgenden schreiben wir der einfachhalber Lebesgue Integrale wie gewöhnliche Riemann Integrale.

Beweis. Sei $R > 0$ derart, dass

$$p(x) = \frac{c}{|x|^{\alpha+1}}(1 + g(x)), \quad |x| \geq R$$

für eine beschränkte Funktion g mit $g(x) \longrightarrow 0$ für $|x| \to \infty$. Der einfachhalber betrachten wir $\xi > 0$. Für $0 < \xi < \frac{1}{R}$ gilt

$$\begin{aligned}
\varphi(\xi) &= \int_{-\infty}^{-\frac{1}{\xi}} p(x)e^{i\xi x}dx + \int_{-\frac{1}{\xi}}^{-R} p(x)e^{i\xi x}dx + \int_{-R}^{R} p(x)e^{i\xi x}dx + \int_{R}^{\frac{1}{\xi}} p(x)e^{i\xi x}dx + \int_{\frac{1}{\xi}}^{\infty} p(x)e^{i\xi x}dx \\
&= I_1(\xi) + I_2(\xi) + I_3(\xi) + I_4(\xi) + I_5(\xi).
\end{aligned}$$

Für I_3 sehen wir, dass I_3 beliebig oft stetig differenzierbar ist mit $I_3(0) = \int_{-R}^{R} p(x)dx$ und wegen $p(x) = p(-x)$ ist auch

$$I_3'(0) = \int_{-R}^{R} p(x)ixdx = 0.$$

Daraus folgt

$$I_3(\xi) = I_3(0) + I_3'(0)\xi + O(\xi^2) = \int_{-R}^{R} p(x)dx + O(\xi^2), \quad \xi \to 0.$$

Da für jede Funktion $h \in O(\xi^2)$ bereits $h \in o(|\xi|^\alpha)$ gilt, folgt

$$I_3(\xi) = \int_{-R}^{R} p(x)dx + o(|\xi|^\alpha), \quad \xi \to 0.$$

Für I_1 erhalten wir mit $y := \xi x$ und $dy = \xi dx$

$$\begin{aligned}
I_1(\xi) &= \int_{-\infty}^{-\frac{1}{\xi}} p(x)e^{i\xi x}dx = \int_{-\infty}^{-\frac{1}{\xi}} \frac{c(1+g(x))}{|x|^{\alpha+1}}e^{i\xi x}dx = c\xi^\alpha \int_{-\infty}^{-1} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}e^{iy}dy \\
&= c\xi^\alpha \int_{-\infty}^{-1} \frac{e^{iy}}{|y|^{\alpha+1}}dy + c\xi^\alpha \int_{-\infty}^{-1} \frac{e^{iy}g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}dy.
\end{aligned}$$

Da g beschränkt ist und $g\left(\frac{y}{\xi}\right) \longrightarrow 0$ für $\xi \to 0$, folgt mit dem Satz von Lebesgue

$$I_1(\xi) = c\xi^\alpha \int_{-\infty}^{-1} \frac{e^{iy}}{|y|^{\alpha+1}}dy + o(|\xi|^\alpha).$$

Analog erhalten wir

$$I_5(\xi) = \int_{\frac{1}{\xi}}^{\infty} p(x)e^{i\xi x}dx = \int_{\frac{1}{\xi}}^{\infty} \frac{c(1+g(x))}{|x|^{\alpha+1}}e^{i\xi x}dx = c\xi^{\alpha}\int_{1}^{\infty} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}e^{iy}dy$$

$$= c\xi^{\alpha}\int_{1}^{\infty} \frac{e^{iy}}{|y|^{\alpha+1}}dy + c\xi^{\alpha}\int_{1}^{\infty} \frac{e^{iy}g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}dy = c\xi^{\alpha}\int_{1}^{\infty} \frac{e^{iy}}{|y|^{\alpha+1}}dy + o(|\xi|^{\alpha}).$$

Für die letzten zwei Terme erhalten wir aus $p(x) = p(-x)$

$$I_2(\xi) + I_4(\xi) = \int_{-\frac{1}{\xi}}^{-R} p(x)(e^{i\xi x} - 1 - i\xi x)dx + \int_{R}^{\frac{1}{\xi}} p(x)(e^{i\xi x} - 1 - i\xi x)dx$$

$$+ \int_{-\frac{1}{\xi}}^{-R} p(x)dx + \int_{R}^{\frac{1}{\xi}} p(x)dx.$$

Der dritte Term liefert

$$\int_{-\frac{1}{\xi}}^{-R} p(x)dx = \int_{-\infty}^{-R} p(x)dx - \int_{-\infty}^{-\frac{1}{\xi}} \frac{c(1+g(x))}{|x|^{\alpha+1}}dx = \int_{-\infty}^{-R} p(x)dx - c\xi^{\alpha}\int_{-\infty}^{-1} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}dy$$

$$= \int_{-\infty}^{-R} p(x)dx - c\xi^{\alpha}\int_{-\infty}^{-1} \frac{1}{|y|^{\alpha+1}}dy + o(|\xi|^{\alpha}).$$

Für den vierten Term erhalten wir analog

$$\int_{R}^{\frac{1}{\xi}} p(x)dx = \int_{R}^{\infty} p(x)dx - \int_{\frac{1}{\xi}}^{\infty} \frac{c(1+g(x))}{|x|^{\alpha+1}}dx = \int_{R}^{\infty} p(x)dx - c\xi^{\alpha}\int_{1}^{\infty} \frac{1}{|y|^{\alpha+1}}dy + o(|\xi|^{\alpha}).$$

Für den ersten Term benutzen wir $e^{iy} - 1 - iy = \frac{\Theta_1(y)}{2}y^2$ mit $|\Theta_1(y)| \le 1$. Weiterhin ist $\alpha - 1 \in$

$(-1,1)$ und somit ist $\frac{\Theta_1(y)}{|y|^{\alpha-1}}$ auf $[-1,0]$ integrierbar. Daraus folgt

$$\begin{aligned}
\int_{-\frac{1}{\xi}}^{-R} p(x)(e^{i\xi x}-1-i\xi x)dx &= c\xi^\alpha \int_{-1}^{-\xi R} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}(e^{iy}-1-iy)dy \\
&= \frac{c\xi^\alpha}{2}\int_{-1}^{-\xi R} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}\Theta_1(y)|y|^2 dy \\
&= \frac{c\xi^\alpha}{2}\int_{-1}^{0} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha-1}}\Theta_1(y)dy - \frac{c\xi^\alpha}{2}\int_{-\xi R}^{0} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha-1}}\Theta_1(y)dy \\
&= \frac{c\xi^\alpha}{2}\int_{-1}^{0} \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy + o(|\xi|^\alpha).
\end{aligned}$$

Analog zeigen wir

$$\int_{R}^{\frac{1}{\xi}} p(x)(e^{i\xi x}-1-i\xi x)dx = \frac{c\xi^2}{2}\int_0^1 \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy + o(|\xi|^\alpha).$$

Setze

$$c_1 := c\left(2\int_1^\infty \frac{\cos(y)-1}{|y|^{\alpha+1}}dy + \frac{1}{2}\int_0^1 \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy + \frac{1}{2}\int_{-1}^0 \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy\right).$$

Dann gilt wegen $\int_{\mathbb{R}} p(x)m(dx) = 1$

$$\varphi(\xi) = 1 + c_1\xi^\alpha + o(|\xi|^\alpha).$$

□

Lemma 5.22. *Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge reeller Zahlen mit $x_n \longrightarrow x$ für $n \to \infty$. Dann gilt*

$$\lim_{n\to\infty}\left(1+\frac{x_n}{n}\right)^n = e^x.$$

Beweis. Sei $A > 0$ mit $|x_n|, |x| \le A$ für alle $n \in \mathbb{N}$. Wegen

$$\left|\left(1+\frac{x_n}{n}\right)^n - e^x\right| \le \left|\left(1+\frac{x_n}{n}\right)^n - \left(1+\frac{x}{n}\right)^n\right| + \left|\left(1+\frac{x}{n}\right)^n - e^x\right|$$

reicht es zu zeigen, dass der erste Term auf der rechten Seite gegen Null konvergiert. Das Letztere folgt aus

$$\begin{aligned}
\left|\left(1+\frac{|x_n|}{n}\right)^n - \left(1+\frac{x}{n}\right)^n\right| &\le \sum_{k=1}^n \left|\left(1+\frac{x_n}{n}\right) - \left(1+\frac{x}{n}\right)\right| \cdot \left(1+\frac{A}{n}\right)^{n-k}\left(1+\frac{A}{n}\right)^{k-1} \\
&= |x-x_n|\left(1+\frac{A}{n}\right)^{n-1}.
\end{aligned}$$

□

Satz 5.23. *Sei X_n eine Folge von unabhängigen Zufallsvariablen mit Verteilung*

$$\mathbb{P} \circ X_n^{-1}(dx) = p(x)dx, \quad n \in \mathbb{N}.$$

Es gelte $p(x) = p(-x)$ und für ein $c > 0$, $0 < \alpha < 2$

$$p(x) \sim \frac{c}{|x|^{\alpha+1}}, \quad |x| \to \infty.$$

Setze $S_n := \frac{1}{n^{\frac{1}{\alpha}}} \sum\limits_{k=1}^{n} X_k$. Dann gilt

$$\mathbb{P} \circ S_n^{-1} \longrightarrow \nu, \quad n \to \infty$$

schwach und die Verteilung ν ist bestimmt durch die charakteristische Funktion $\psi(\xi) = e^{-c_1|\xi|^\alpha}$.

Beweis. Es sei $\varphi(\xi) = \int\limits_{\mathbb{R}} e^{i\xi x} p(x) m(dx)$. Dann gilt wegen Unabhängigkeit

$$\mathbb{E}(e^{i\xi S_n}) = \left(\varphi\left(\frac{\xi}{n^{\frac{1}{\alpha}}}\right)\right)^n.$$

Für jedes $\xi \in \mathbb{R}$ gibt nach es (5.3) eine Folge $a_n(\xi)$ mit $na_n(\xi) \longrightarrow 0$ für $n \to \infty$ derart, dass

$$\varphi\left(\frac{\xi}{n^{\frac{1}{\alpha}}}\right) = 1 - \frac{c_1|\xi|^\alpha}{n} + a_n(\xi).$$

Wir erhalten somit

$$\begin{aligned}\left(\varphi\left(\frac{\xi}{n^{\frac{1}{\alpha}}}\right)\right)^n &= \left(1 - \frac{c_1|\xi|^\alpha}{n} + a_n(\xi)\right)^n \\ &= \left(1 + \frac{na_n(\xi) - c_1|\xi|^\alpha}{n}\right)^n.\end{aligned}$$

Wegen Lemma 5.22 mit $x_n := na_n(\xi) - c_1|\xi|^\alpha \longrightarrow -c_1|\xi|^\alpha =: x$ gilt

$$\mathbb{E}(e^{i\xi S_n}) \longrightarrow e^{-c_1|\xi|^\alpha}, \quad n \to \infty.$$

Die Behauptung folgt somit aus Levys Stetigkeitssatz. □

Bemerkung 5.24. *(a) Sei $X_n \in \mathcal{L}^2$ eine Folge unabhängig, identisch verteilter Zufallsvariablen mit $\mathbb{E}(X_1) = 0$ und $\mathbb{E}(X_1^2) = 1$. Sei $M_n := \max\left\{\frac{|X_k|}{\sqrt{n}}, \dots, \frac{|X_k|}{\sqrt{n}}\right\}$. Dann gilt*

$$\mathbb{P}(M_n \leq t) = \mathbb{P}\left(\frac{|X_k|}{\sqrt{n}} \leq t, \quad k = 1, \dots, n\right) = \prod_{k=1}^{n} \mathbb{P}\left(\frac{|X_k|}{\sqrt{n}} \leq t\right) = \mathbb{P}(|X_1| \leq \sqrt{n}t)^n.$$

Weiterhin erhalten wir

$$\begin{aligned}
\mathbb{P}(|X_k| > \sqrt{n}t) &= \int_{\mathbb{R}} \mathbb{1}_{|y|>\sqrt{n}t} \mathbb{P} \circ X_1^{-1}(dy) \\
&\leq \int_{\mathbb{R}} \mathbb{1}_{|y|>\sqrt{n}t} \frac{y^2}{nt^2} \mathbb{P} \circ X_1^{-1}(dy) \\
&= \frac{1}{nt^2} \int_{|y|>\sqrt{n}t} y^2 \mathbb{P} \circ X_1^{-1}(dy) = \frac{\Theta_n(t)}{n}
\end{aligned}$$

mit $\Theta_n(t) \longrightarrow 0$ *für alle* $t > 0$. *Daraus folgt*

$$\begin{aligned}
\mathbb{P}(M_n \leq t) = \mathbb{P}(|X_1| \leq \sqrt{n}t)^n &= (1 - \mathbb{P}(|X_1| > \sqrt{n}t))^n \\
&\geq \left(1 - \frac{\Theta_n(t)}{n}\right)^n \longrightarrow 1, \quad n \to \infty.
\end{aligned}$$

Also gilt $\lim\limits_{n\to\infty} \mathbb{P}(M_n \leq t) = 1$ *für alle* $t > 0$ *und somit*

$$\mathbb{P} \circ M_n^{-1} \longrightarrow \delta_0, \quad n \to \infty$$

schwach.
Konsequenz: *Der wesentliche Beitrag im zentralen Grenzwertsatz kommt von der Summe und nicht von dem maximalen Summanden.*

(b) Sei X_n *wie in Satz 5.23. Für* $t > 0$ *gilt dann*

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq k\leq n} |X_k| \geq tn^{\frac{1}{\alpha}}\right) &= 1 - \mathbb{P}\left(\max_{1\leq k\leq n} |X_k| < tn^{\frac{1}{\alpha}}\right) \\
&= 1 - \left(\mathbb{P}\left(|X_1| < tn^{\frac{1}{\alpha}}\right)\right)^n \\
&= 1 - \left(1 - \mathbb{P}\left(|X_1| \geq tn^{\frac{1}{\alpha}}\right)\right)^n.
\end{aligned}$$

Ferner gilt für hinreichend große $n \in \mathbb{N}$

$$\mathbb{P}(|X_1| \geq tn^{\frac{1}{\alpha}}) \sim \int_{|x|\geq tn^{\frac{1}{\alpha}}} \frac{c}{|x|^{\alpha+1}} m(dx) = \frac{2c}{\alpha t^\alpha n}.$$

Daher gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{1\leq k\leq n} |X_k| \geq tn^{\frac{1}{\alpha}}\right) = \lim_{n\to\infty} \left(1 - \left(1 - \frac{2c}{\alpha t^\alpha n}\right)^n\right) = 1 - e^{-\frac{2c}{\alpha t^\alpha}} > 0.$$

Konsequenz: *Der meximale Term hat in diesem Fall einen Einfluss auf die Konvergenz in Satz 5.23.*

Definition 5.25. *Ein Wahrscheinlichkeitsmaß $\mu(dx)$ auf $\mathbb{R}$ heißt Poissonverteilt, wenn es ein $\lambda > 0$ gibt mit*

$$\mu(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots,$$

d.h. $\mu = e^{-\lambda} \sum\limits_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta_k$. Eine Zufallsvariable X heißt Poissonverteilt, wenn ihre Verteilung $\mathbb{P} \circ X^{-1}$ Poissonverteilt ist.

Sei $n \in \mathbb{N}$. Wir betrachten n unabhängige Münzwürfe, wo 1 für Kopf und 0 für Zahl steht. Ein Ausgang ist daher ein n-Tupel $\omega = (\omega_1, \ldots, \omega_n) \in \{0,1\}^n$. Wir wählen

$$\Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) \in \{0,1\}^n\}$$

mit $\mathcal{F} = 2^{\Omega_n}$ und

$$\mathbb{P}_n(\{\omega\}) = p_n^{\sum\limits_{k=1}^{n} \omega_k} (1 - p_n)^{\sum\limits_{k=1}^{n} (1-\omega_k)}, \quad \omega \in \Omega_n.$$

Hier ist $p_n \in [0,1]$ die Wahrscheinlichkeit bei einem der Münzwürfe Kopf zu werfen. Demnach ist $1 - p_n$ die Wahrscheinlichkeit Zahl zu werfen.

Satz 5.26. *(Poissonscher Grenzwertsatz) Sei $p_n \in [0,1]$ mt $np_n \longrightarrow \lambda$ für ein $\lambda > 0$. Dann gilt*

$$\lim_{n\to\infty} \mathbb{P}_n\left(\left\{\omega \in \Omega_n \mid \sum_{j=1}^{n} \omega_j = k\right\}\right) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0.$$

Beweis. Sei $k \in \mathbb{N}_0$ fest. Wegen $np_n \longrightarrow \lambda$ folgt $p_n \longrightarrow 0$. Wir können daher annehmen, dass $p_n < 1$ gilt. Die Anzahl der Elemente in

$$A_k := \left\{\omega \in \Omega_n \mid \sum_{j=1}^{n} \omega_j = k\right\}$$

ist gleichbedeutend mit der Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Diese hat demnach $\binom{n}{k}$ Elemente. Für jedes $\omega \in A_k$ gilt

$$\mathbb{P}_n(\{\omega\}) = p_n^k (1 - p_n)^{n-k}$$

und somit

$$\mathbb{P}_n(A_k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} p_n^k \left(1 - \frac{np_n}{n}\right)^{n-k}.$$

Wegen $\left(1 - \frac{np_n}{n}\right)^k \longrightarrow 1$ für $n \to \infty$ folgt

$$\lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^{n-k} = \frac{\lim\limits_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^n}{\lim\limits_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^k} = e^{-\lambda}.$$

Weiterhin gilt

$$\lim_{n\to\infty} n(n-1)\cdots(n-k+1) p_n^k = \lim_{n\to\infty} (np_n)^k = \lambda^k.$$

Daraus folgt die Behauptung. $\square$

Korollar 5.27. *Sei X_n eine Folge von unabhängigen Zufallsvariablen mit*

$$\mathbb{P}(X_n = 1) = p_n, \quad \mathbb{P}(X_n = 0) = 1 - p_n, \quad n \in \mathbb{N}$$

für eine Folge $p_n \in [0,1]$ mit $np_n \longrightarrow \lambda > 0$. Sei $S_n = \sum_{j=1}^{n} X_j$. Dann gilt

$$\mathbb{P}(S_n = k) \longrightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}.$$

Die Verteilung von S_n konvergiert also gegen die Poissonverteilung.

Zum Schlußbetrachten wir ein Beispiel aus der Thermodynamik.

Beispiel 5.28. *Es sei $V_R \subset \mathbb{R}^3$ eine Box mit Kantenlänge $R > 0$ und uniformer Verteilung $\frac{1}{R^3} m(dx)$ auf V_R. Dann beschreibt $x \in V_R$ die Position eines einzelnen Atoms. Wir betrachten ein ideales Gas aus $n(R) \in \mathbb{N}$ Atomen welche nicht miteinander wechselwirken, d.h. unabhängig voneinander sind. Die Positionen der Atome seien $x_1, \ldots, x_{n(R)} \in V_R$. Der gemeinsame Zustandsraum ist $\Omega := V_R^{n(R)}$ und wir nehmen an dass alle Atome uniform verteilt sind, d.h.*

$$\mathbb{P} = \frac{1}{(R^3)^{n(R)}} m^{\otimes n(R)}.$$

Ist $U \subset V_R$ eine weitere Box, so bezeichnet

$$X_U(x_1, \ldots, x_{n(R)}) = \sum_{k=1}^{n(R)} \mathbb{1}_U(x_k)$$

die Anzahl der Atome in U. Im thermodynamischen Limes betrachten wir den Grenzfall $R \to \infty$ mit $n(R) \to \infty$ geeignet gewählt. Ziel ist es die Verteilung von X_U in diesem Limes zu bestimmen. Dieses kann als Verteilung von sehr vielen Atomen in einem großen Volumen interpretiert werden.

Die Wahrscheinlichkeit 1 Atom in U zu finden ist gegeben durch $\mathbb{P}(X_U = 1) = \frac{m(U)}{R^3} =: p_n(R)$. Es gilt

$$n(R) p_{n(R)} = \frac{n(R) m(U)}{R^3} = m(U) \frac{n(R)}{R^3}.$$

Wir betrachten die Annahme

$$\frac{n(R)}{R^3} \longrightarrow \rho, \quad R \to \infty.$$

In diesem Fall sei $\lambda = \rho m(U)$. Der Poissonsche Grenzwertsatz zeigt

$$\lim_{R \to \infty} \mathbb{P}(X_U = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0.$$

Der Wert $\rho > 0$ kann als Dichte des Gases interpretiert werden. Damit sit λ die mittlere Anzahl der Atome in dem Gebiet U.

6 Grundlagen stochastischer Prozesse

Sei $(E, \mathcal{E})$ ein messbarer Raum.

Definition 6.1. *Sei T eine Indexmenge. Ein stochastischer Prozess ist eine Familie von Zufallsvariablen $X := (X_t)_{t \in T}$ mit Werten in E.*

(a) Für $T = \mathbb{Z}$ oder $T = \mathbb{Z}_+$ heist X (diskreter) stochastischer Prozess.

(b) Ist T ein Intervall $[a, b] \subset \mathbb{R}$ oder $T = \mathbb{R}$ oder $T = \mathbb{R}_+$, so heißt X (Zeit kontinuierlicher) stochastischer Prozess.

(c) Ist T ein mehrdimensionaler Raum, so heißt X auch zufälliges Feld (random field).

Für jedes $\omega \in \Omega$ heißt $t \longmapsto X_t(\omega)$ Trajektorie/ Realisierung bzw. Pfad von X.

6.1 Konstruktion stochastischer Prozesse

Sei E ein metrischer Raum und $\widetilde{\Omega} := \widetilde{\Omega}(T, E)$ der Pfadraum, d.h. die Menge aller Abbildungen $\widetilde{\omega} : T \longrightarrow E$. Für $k \in \mathbb{N}$, $t_1, \ldots, t_k \in T$ und $A \in \mathcal{B}(E^k)$ heißt

$$C(t_1, \ldots, t_k; A) := \left\{ \widetilde{\omega} \in \widetilde{\Omega} \mid (\widetilde{\omega}(t_1), \ldots, \widetilde{\omega}(t_k)) \in A \right\}$$

Zylindermenge mit Basis A. Dann ist

$$\mathcal{B}_{t_1, \ldots, t_k} := \{ C(t_1, \ldots, t_k; A) \mid A \in \mathcal{B}(E_k) \}$$

eine σ-Algebra auf $\widetilde{\Omega}$. Es sei $\mathcal{B}(T, E) := \sigma\left(\mathcal{B}_{t_1, \ldots, t_k} \mid t_1, \ldots, t_k \in T,\ k \in \mathbb{N}\right)$ die kleinste σ-Algebra welche alle $\mathcal{B}_{t_1, \ldots, t_k}$ enthält. Dann ist $(\widetilde{\Omega}(T, E), \mathcal{B}(T, E)))$ ein messbarer Raum.

Lemma 6.2. *Es sei $X = (X_t)_{t \in T}$ ein stochastischer Prozess. Die Abbildung*

$$I_X : \Omega \longrightarrow \widetilde{\Omega}(T, E), \quad \omega \longmapsto (t \longmapsto X_t(\omega))$$

ist messbar.

Beweis. Sei $k \in \mathbb{N}$, $A \in \mathcal{B}(E^k)$ und $t_1, \ldots, t_k \in T$. Dann ist $\omega \longmapsto (X_{t_1}(\omega), \ldots, X_{t_k}(\omega))$ messbar. Folglich gilt

$$\begin{aligned} I_X^{-1}(C(t_1, \ldots, t_k; A)) &= \{ \omega \in \Omega \mid I_X(\omega) \in C(t_1, \ldots, t_k; A) \} \\ &= \{ \omega \in \Omega \mid (X_{t_1}(\omega), \ldots, X_{t_k}(\omega)) \in A \} \in \mathcal{F}. \end{aligned}$$

□

Definition 6.3. *Sei $X = (X_t)_{t \in T}$ ein stochastischer Prozess. Dann heißt $\mathbb{P} \circ I_X^{-1}$ Verteilung von $(X_t)_{t \in T}$. Die Wahrscheinlichkeitsmaße $\mathbb{P} \circ (X_{t_1}, \ldots, X_{t_k})^{-1}$ auf E^k mit $k \in \mathbb{N}$ und $t_1, \ldots, t_k \in T$ heissen endlichdimensionale Verteilungen von X.*

Lemma 6.4. *Seien $(X_t)_{t \in T}$ und $(Y_t)_{t \in T}$ zwei stochastische Prozesse. Dann sind äquivalent:*

(a) *X und Y haben diesselbe Verteilung.*

(b) *X und Y haben diesselben endlichdimensionalen Verteilungen.*

Beweis. $(a) \Longrightarrow (b)$: Sei $k \in \mathbb{N}$, $t_1, \ldots, t_k \in T$ und $A \in \mathcal{B}(E^k)$. Dann gilt

$$\begin{aligned}\mathbb{P} \circ (X_{t_1}, \ldots, X_{t_k})^{-1}(A) &= \mathbb{P}\left(\{\omega \in \Omega \mid (X_{t_1}(\omega), \ldots, X_{t_k}(\omega)) \in A\}\right) \\ &= (\mathbb{P} \circ I_X^{-1})(C(t_1, \ldots, t_k; A)) = (\mathbb{P} \circ I_Y^{-1})(C(t_1, \ldots, t_k; A)) \\ &= \mathbb{P} \circ (Y_{t_1}, \ldots, Y_{t_k})^{-1}(A).\end{aligned}$$

$(b) \Longrightarrow (a)$: Obige Rechnung zeigt dass $\mathbb{P} \circ I_X^{-1}$ und $\mathbb{P} \circ I_Y^{-1}$ auf allen Zylindermengen übereinstimmen. Da diese einen $\cap$-stabilen Erzeuger bilden, folgt die Behauptung. □

Satz 6.5. *Sei μ ein Wahrscheinlichkeitsmaß auf $(\widetilde{\Omega}(T, E), \mathcal{B}(T, E))$. Definiere*

$$X_t(\widetilde{\omega}) := \widetilde{\omega}(t), \quad t \in E, \quad \widetilde{\omega} \in \widetilde{\Omega}(T, E). \tag{6.1}$$

Dann ist $(X_t)_{t \in T}$ ein stochastischer Prozess auf $(\widetilde{\Omega}(T, E), \mathcal{B}(T, E), \mu)$ mit Verteilung μ.

Beweis. Für den ersten Teil reicht es zu zeigen, dass X_t messbar ist für jedes $t \in T$. Sei $t \in T$ fest und $A \in \mathcal{B}(E)$. Dann gilt

$$\begin{aligned}X_t^{-1}(A) &= \{\widetilde{\omega} \in \widetilde{\Omega}(T, E) \mid X_t(\widetilde{\omega}) \in A\} \\ &= \{\widetilde{\omega} \in \widetilde{\Omega}(T, E) \mid \widetilde{\omega}(t) \in A\} = C(t; A) \in \mathcal{B}(T, E).\end{aligned}$$

Für den zweiten Teil betrachten wir $k \in \mathbb{N}$, $t_1, \ldots, t_k \in T$ und $A \in \mathcal{B}(E^k)$. Dann gilt

$$\begin{aligned}\mu \circ I_X^{-1}(C(t_1, \ldots, t_k; A)) &= \mu\left(\left\{\widetilde{\omega} \in \widetilde{\Omega}(T, E) \mid (X_{t_1}(\widetilde{\omega}), \ldots, X_{t_k}(\widetilde{\omega})) \in A\right\}\right) \\ &= \mu\left(\left\{\widetilde{\omega} \in \widetilde{\Omega}(T, E) \mid (\widetilde{\omega}(t_1), \ldots, \widetilde{\omega}(t_k)) \in A\right\}\right) \\ &= \mu(C(t_1, \ldots, t_k; A))\end{aligned}$$

und somit $\mu \circ I_X^{-1} = \mu$. □

Die Darstellung (6.1) heißt kanonische Darstellung eines stochastischen Prozesses. Insbesondere ist ein stochastischer Prozess eindeutig durch seine Verteilung festgelegt. Wir haben gesehen, dass eine solche Verteilung eindeutig durch die endlichdimensionalen Verteilungen charakterisiert ist. Im folgenden wollen wir untersuchen, wann eine Familie von endlichdimensionalen Verteilungen ein Wahrscheinlichkeitsmaß auf $\widetilde{\Omega}(T, E)$ festlegt, und damit einen stochastischen Prozess liefert.

Definition 6.6. *Sei $\{\mathbb{P}_{t_1, \ldots, t_k} \in \mathcal{P}(E^k) \mid k \in \mathbb{N}, \quad t_1, \ldots, t_k \in T\}$ eine Familie von Wahrscheinlichkeitsmaßen. Diese Familie heißt konsisten, wenn sie die folgenden Bedingungen erfüllt.*

(a) *Für jedes $k \in \mathbb{N}$, jede Permutation π von k-Elementen und jedes $A = A_1 \times A_k \in \mathcal{B}(E^k)$ gilt*

$$\mathbb{P}_{t_1, \ldots, t_k}(A) = \mathbb{P}_{\pi(t_1, \ldots, t_k)}(A_{\pi_1} \times \cdots \times A_{\pi_k}),$$

wo $\pi(t_1, \ldots, t_k) = (\pi_1, \ldots, \pi_k)$.

(b) *Für alle $t_1, \dots, t_k, t_{k+1}$ und alle $A \in \mathcal{A}(E^k)$ gilt*

$$\mathbb{P}_{t_1,\dots,t_k}(A) = \mathbb{P}_{t_1,\dots,t_{k+1}}(A \times E).$$

Seien $t_1, \dots, t_k \in T$ mit $k \in \mathbb{N}$ gegeben. Definiere die Projektion

$$\mathrm{pr}_{t_1,\dots,t_k} : \widetilde{\Omega}(T,E) \longrightarrow E^k, \quad \mathrm{pr}_{t_1,\dots,t_k}(\widetilde{\omega}) = (\widetilde{\omega}(t_1), \dots, \widetilde{\omega}(t_k)).$$

Satz 6.7. *(Kolmogorov's Erweiterungssatz) Sei E ein Polnischer Raum.*

(a) *Sei $\widetilde{\mathbb{P}}$ ein Wahrscheinlichkeitsmaß auf $(\widetilde{\Omega}(T,E), \mathcal{B}(T,E))$. Dann definiert*

$$\mathbb{P}_{t_1,\dots,t_k}(A) := \widetilde{\mathbb{P}} \circ \mathrm{pr}^{-1}_{t_1,\dots,t_k}(A), \quad t_1, \dots, t_k \in T, \quad k \in \mathbb{N}, \; A \in \mathcal{B}(E^k) \tag{6.2}$$

eine konsistente Familie von Wahrscheinlichkeitsmaßen.

(b) *Sei $\{\mathbb{P}_{t_1,\dots,t_k} \in \mathcal{P}(E^k) \mid k \in \mathbb{N}, \;\; t_1, \dots, t_k \in T\}$ eine konsistente Familie von Wahrscheinlichkeitsmaßen. Dann gibt es genau ein Wahrscheinlichkeitsmaß $\widetilde{\mathbb{P}}$ auf $(\widetilde{\Omega}(T,E), \mathcal{B}(T,E))$ mit* (6.2).

Beispiel 6.8. *(Unabhängig, identisch verteilte Zufallsvariablen) Sei $T = \mathbb{Z}_+$ und $E = \mathbb{R}$. Wähle ein Wahrscheinlichkeitsmaß μ auf $\mathbb{R}$. Sei die Familie von Wahrscheinlichkeitsmaßen $\mathbb{P}_{t_1,\dots,t_n}$ gegeben durch*

$$\mathbb{P}_{t_1,\dots,t_n}(A) = \prod_{j=1}^{n} \mu(A_j), \quad A = A_1 \times \dots \times A_n \in \mathcal{B}(\mathbb{R}^n).$$

Dann ist $\{\mathbb{P}_{t_1,\dots,t_n} \mid t_1, \dots, t_n \in \mathbb{Z}_+, \; n \in \mathbb{N}\}$ eine konsistente Familie. Es sei $\widetilde{\mathbb{P}}$ das eindeutig bestimmte Wahrscheinlichkeitsmaß auf $\widetilde{\Omega}(\mathbb{Z}_+, \mathbb{R})$ mit endlichdimensionalen Verteilungen $\mathbb{P}_{t_1,\dots,t_n}$. Jedes $\widetilde{\omega} \in \widetilde{\Omega}(\mathbb{Z}_+, \mathbb{R})$ ist eine Folge $(\widetilde{\omega}_n)_{n \in \mathbb{N}} \subset \mathbb{R}$. Der dazugehörige Prozess in der kanonischen Darstellung ist gegeben durch

$$X_n(\widetilde{\omega}) = \widetilde{\omega}_n, \quad n \in \mathbb{Z}_+.$$

Dann ist $(X_n)_{n \in \mathbb{Z}_+}$ eine Folge von unabhängig, identisch verteilten Zufallsvariablen mit Werten in $\mathbb{R}$.

Beispiel 6.9. *(Gauß–Prozesse) Ein stochastischer Prozess $X = (X_t)_{t \geq 0}$ heißt Gauß–Prozess, falls für alle $t_1, \dots, t_n \geq 0$ die endlichdimensionalen Verteilungen $\mathbb{P} \circ (X_{t_1}, \dots, X_{t_n})^{-1}$ Normalverteilt sind. Wie wir bereits gesehen haben, ist die Normalverteilung eindeutig festgelegt durch den Erwartungswert $m(t_i) = \mathbb{E}(X_{t_i})$ und die Konvarianzmatrix B mit Einträgen*

$$B(t_i, t_j) = \mathbb{E}\left((X_{t_i} - m(t_i))(X_{t_j} - m(t_j))\right).$$

Die endlich-dimensionalen Verteilungen sind also festgelegt duch eine Funktion $m(t)$ und eine symmetrische Funktion $B(t,s)$ derart, dass für alle $t_1, \dots, t_n \geq 0$ die Matrix $(B(t_i, t_j))_{i,j=1,\dots,n}$ positiv semi-definit ist.

Umgekehrt gibt es zu je zwei Funktionen $m(t), B(t,s)$ mit obigen Eigenschaften genau einen Gauß–Prozess $(X_t)_{t \geq 0}$. Wählen wir speziell $B(t,s) = \min\{s,t\}$ und $m(t) = 0$, so heißt der dazugehörige Gauß–Prozess auch Brownsche Bewegung. Diese wird häufig auch mit $(W_t)_{t \geq 0}$ bezeichnet.

Satz 6.10. *Sei $(W_t)_{t\geq 0}$ ein stochastischer Prozess mit Werten in $\mathbb{R}$. Genau dann ist $(W_t)_{t\geq 0}$ eine Brownsche Bewegung, wenn die folgenden Eigenschaften gelten.*

(a) Es gilt $W_0 = 0$ fast sicher.

(b) Für alle $0 \leq s \leq t$ ist $W_t - W_s$ normalverteilt mit Erwartungswert 0 und Varianz $t - s$.

(c) Die Zufallsvariablen $W_{t_0}, W_{t_1} - W_{t_0}, \dots, W_{t_k} - W_{t_{k-1}}$ sind für jedes $k \geq 1$ und alle $0 = t_0 \leq t_1 \leq \dots \leq t_k$ unabhängig.

Beweis. Es sei $(W_t)_{t\geq 0}$ eine Brownsche Bewegung. Dann gilt $\mathbb{E}(W_0^2) = \min\{0,0\} = 0$ und somit $W_0 = 0$ fast sicher. Seien $0 \leq s \leq t$. Dann ist (W_s, W_t) normalverteilt und somit ist auch $W_t - W_s$ normalverteilt. Daraus folgt

$$\mathbb{E}(W_t - W_s)^2 = \mathbb{E}(W_t^2) - 2\mathbb{E}(W_t W_s) + \mathbb{E}(W_s^2) = \min\{t,t\} - 2\min\{t,s\} + \min\{s,s\} = t - s.$$

Sei $k \geq 1$ und $0 = t_0 \leq t_1 \leq \dots \leq t_k$. Dann ist $(W_{t_0}, \dots, W_{t_k})$ normalverteilt. Also ist auch $(W_{t_0}, W_{t_1} - W_{t_0}, \dots, W_{t_k} - W_{t_{k-1}})$ normalverteilt. Also reicht es zu zeigen, dass die Komponenten unkorreliert sind. Sei $1 \leq i < j \leq k$, dann gilt

$$\begin{aligned}\mathbb{E}((W_{t_i} - W_{t_{i-1}})(W_{t_j} - W_{t_{j-1}})) &= \mathbb{E}(W_{t_i} W_{t_j}) - \mathbb{E}(W_{t_i} W_{t_{j-1}}) - \mathbb{E}(W_{t_{i-1}} W_{t_j}) + \mathbb{E}(W_{t_{i-1}} W_{t_{j-1}}) \\ &= \min\{t_i, t_j\} - \min\{t_i, t_{j-1}\} - \min\{t_{i-1}, t_j\} + \min\{t_{i-1}, t_{j-1}\} \\ &= t_j - t_{j-1} - t_j + t_{j-1} = 0.\end{aligned}$$

Umgekehrt gelten die Eigenschaften (a),(b),(c). Dann ist $(W_{t_0}, W_{t_1} - W_{t_0}, \dots, W_{t_k} - W_{t_{k-1}})$ normalverteilt und die einzelnen Komponenten sind unabhängig. Es gibt eine invertierbare Matrix A derart, dass

$$(W_{t_0}, W_{t_1}, \dots, W_{t_k})^t = A(W_{t_0}, W_{t_1} - W_{t_0}, \dots, W_{t_k} - W_{t_{k-1}})^t.$$

Folglich ist auch $(W_{t_0}, W_{t_1}, \dots, W_{t_k})$ normalverteilt. Es gilt $m(t) = \mathbb{E}(W_t) = \mathbb{E}(W_t - W_0) = 0$ und für $0 \leq s \leq t$

$$B(t,s) = \mathbb{E}(W_t W_s) = \mathbb{E}((W_t - W_s)(W_s - W_0)) + \mathbb{E}((W_s - W_0)^2) = \mathbb{E}(W_s^2) = s.$$

□

Sei $\mathcal{B}(\mathbb{Z}_+, \mathbb{R})$ die kleinste σ-Algebra erzeugt urch die Zylindermengen mit $T = \mathbb{Z}_+$ und $E = \mathbb{R}$. Ferner sei $\mathcal{B}(\mathbb{R}_+, \mathbb{R})$ die kleinste σ-Algebra erzeugt durch die Zylindermengen mit $T = \mathbb{R}_+$ und $E = \mathbb{R}$.

Satz 6.11. *Genau dann ist $A \in \mathcal{B}(\mathbb{R}_+, \mathbb{R})$, wenn es eine Folge $(t_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+$ und ein $B \in \mathcal{B}(\mathbb{Z}_+, \mathbb{R})$ gibt mit*

$$A = \left\{\tilde{\omega} \in \tilde{\Omega}(\mathbb{Z}_+, \mathbb{R}) \mid (\tilde{\omega}(t_1), \tilde{\omega}(t_2), \dots) \in B\right\}.$$

Korollar 6.12. *Sei $E = \mathbb{R}$, $T = \mathbb{R}_+$ und $C > 0$. Dann sind*

$$\left\{\tilde{\omega} \in \tilde{\Omega}(\mathbb{R}_+, \mathbb{R}) \mid |\tilde{\omega}(t)| < C, \quad t \in \mathbb{R}_+\right\}$$

und $\{\tilde{\omega} \in \tilde{\Omega}(\mathbb{R}_+, \mathbb{R}) \mid \tilde{\omega}$ ist stetig $\}$ keine Elemente von $\mathcal{B}(\mathbb{R}_+, \mathbb{R})$. Weiterhin ist

$$\tilde{\Omega}(\mathbb{R}_+, \mathbb{R}) \times \mathbb{R}_+ \ni (\tilde{\omega}, t) \longmapsto \tilde{\omega}(t)$$

nicht messbar.

6.2 Eigenschaften stochastischer Prozesse

Definition 6.13. *Ein stochastischer Prozess $(X_t)_{t\in T}$ heißt messbar, wenn*

$$(\omega, t) \longmapsto X_t(\omega)$$

messbar ist bezüglich $\mathcal{F} \otimes \mathcal{B}(T) - -\mathcal{B}(E)$.

Sei $X := (X_t)_{t\geq 0}$ ein zeitkontinuierlicher stochastischer Prozess.

- *X ist linksstetig, falls $t \longmapsto X_t$ fast sicher linksstetig ist.*
- *X ist rechtsstetig, falls $t \longmapsto X_t$ fast sicher rechtsstetig ist.*
- *X ist càdlàg, falls $t \longmapsto X_t$ fast sicher rechtsstetig ist und die linksseitigen Grenzwerte existieren.*
- *X ist ist stetig, falls $t \longmapsto X_t$ fast sicher stetig ist.*

Lemma 6.14. *Sei $X := (X_t)_{t\in T}$ ein zeitkontinuierlicher stochastischer Prozess. Angenommen X ist linksstetig bzw. rechtsstetig. Dann ist X messbar.*

Beweis. Wir zeigen den Fall, dass X rechtsstetig ist mit $T = [0, \infty)$. Der allgemeine Fall geht analog. Definiere einen neuen Prozess durch

$$Y_t^n(\omega) := X_{\frac{k}{2^n}}(\omega), \quad \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \tag{6.3}$$

mit $t \geq 0$, $n \in \mathbb{N}$ und $k \in \mathbb{N}_0$. Dann ist $(Y_t^n)_{t\geq 0}$ messbar für jedes $n \in \mathbb{N}$. Denn für $A \in \mathcal{B}(E)$ gilt

$$\{(t, \omega) \in [0, \infty) \times \Omega \mid Y_t^n(\omega) \in A\} = \bigcup_{k\geq 0} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) \times \left\{\omega \in \Omega \mid X_{\frac{k}{2^n}}(\omega) \in A\right\}.$$

Da X rechtsstetig ist, gilt fast sicher $\lim_{n\to\infty} Y_t^n = X_t$ für alle $t \geq 0$. Damit ist auch X messbar. □

Definition 6.15. *Seien $(X_t)_{t\in T}$ und $(Y_t)_{t\in T}$ zwei stochastische Prozesse auf demselben Wahrscheinlichkeitsraum.*

(a) X ist eine Modifikation von Y, falls

$$\mathbb{P}(X_t = Y_t) = 1, \quad t \in T.$$

(b) X und Y sind ununterscheidbar, falls

$$\mathbb{P}(X_t = Y_t, \quad t \in T) = 1.$$

Bemerkung 6.16. *(a) Ist T abzählbar, so ist X eine Modifikation von Y genau dann, wenn X und Y ununterscheidbar sind.*

(b) Ist T nicht abzählbar, so gilt immerhin: Sind X, Y ununterscheidbar, so ist X eine Modifikation von Y. Aber auch Y ist eine Modifikation von X.

Lemma 6.17. *Seien $X := (X_t)_{t\in T}$ und $Y := (Y_t)_{t\in T}$ zwei zeitkontinuierliche stochastische Prozesse, welche beide links oder rechtsstetig sind. Ist X eine Modifikation von Y, so sind X und Y ununterscheidbar.*

Beweis. Sei $\Omega' \in \mathcal{F}$ mit $\mathbb{P}(\Omega') = 1$ derart, dass für alle $\omega \in \Omega'$ $t \longmapsto X_t(\omega)$, $t \longmapsto Y_t(\omega)$ beide links oder rechtsstetig sind. Sei $S \subset T$ eine dichte Teilmenge. Dann gibt es für jedes $t \in S$ ein $\Omega_t \in \mathcal{F}$ mit $\mathbb{P}(\Omega_t) = 1$ derart, dass

$$X_t(\omega) = Y_t(\omega), \quad t \in S, \quad \omega \in \Omega_t.$$

Wähle $\Omega'' := \Omega' \cap \bigcap_{t\in S} \Omega_t \in \mathcal{F}$, dann gilt

$$\mathbb{P}(\Omega''^c) \leq \mathbb{P}(\Omega'^c) + \sum_{t\in S} \mathbb{P}(\Omega_t^c) = 0$$

und somit $\mathbb{P}(\Omega'') = 1$. Für jedes $\omega \in \Omega''$ und $t \in T$ sei $(t_n)_{n\in\mathbb{N}} \subset S$ eine Folge mit $t_n \longrightarrow t$. Sind X und Y linksstetig, so wählen wir $(t_n)_{n\in\mathbb{N}}$ aufsteigend, andererseits wählen wir die Folge absteigend. Dann gilt

$$X_t(\omega) = \lim_{n\to\infty} X_{t_n}(\omega) = \lim_{n\to\infty} Y_{t_n}(\omega) = Y_t(\omega).$$

□

Definition 6.18. *Seien $X^1 = (X_t^1)_{t\in T}, \ldots, X^m = (X_t^m)_{t\in T}$ stochastische Prozesse auf einem gemeinsamen Wahrscheinlichkeitsraum. Wir sagen dass $X^1, \ldots X^m$ unabhängig sind, falls die σ-Algebren*

$$\mathcal{F}^{X^j} := \sigma\left(X_t^j \mid t \in T\right), \quad j = 1, \ldots, m$$

unabhängig sind.

Definition 6.19. *Sei $I \subset [0, \infty)$ eine Indexmenge. Eine Filtration ist eine Familie von σ-Algebren $(\mathcal{F}_t)_{t\in I}$ mit den folgenden Eigenschaften:*

- *$\mathcal{F}_t \subset \mathcal{F}$ für alle $t \in I$.*
- *$\mathcal{F}_s \subset \mathcal{F}_t$ für alle $s, t \in I$ mit $s \leq t$.*

Der Raum $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, \mathbb{P})$ heißt filtrierter Raum.

Im folgenden betrachten wir stets den Fall $I = \mathbb{R}_+$ oder $I = \mathbb{N}_0$. Im Falle $I = \mathbb{R}_+$ ist die folgende Definition sinnvoll.

Definition 6.20. *Eine Filtration $(\mathcal{F}_t)_{t\geq 0}$ heißt rechtsstetig, falls $\mathcal{F}_t = \bigcup_{u>t} \mathcal{F}_u$.*

Ist eine Filtration $(\mathcal{F}_t)_{t\geq 0}$ gegeben, so ist per Definition $\mathcal{F}_{t+} := \bigcap_{u>t} \mathcal{F}_t$ rechtsstetig. Insbesondere ist $\mathcal{F}_t$ rechtsstetig genau dann, wenn $\mathcal{F}_t = \mathcal{F}_{t+}$ für alle $t \geq 0$.

Beispiel 6.21. *Sei $(X_t)_{t\geq 0}$ ein stochastischer Prozess. Die von X erzeugte Filtration ist gegeben durch*

$$\mathcal{F}_t := \sigma\left(X_s \mid s \leq t\right), \quad t \geq 0.$$

Definition 6.22. *Sei $(X_t)_{t\geq 0}$ ein stochastischer Prozess mit Werten in E.*

(a) $(X_t)_{t\geq 0}$ heißt adaptiert, falls X_t messbar bezüglich $\mathcal{F}_t$ ist für alle $t \geq 0$.

(b) $(X_t)_{t\geq 0}$ heißt progressiv messbar, falls für alle $T > 0$

$$[0,T] \times \Omega \ni (t,\omega) \longmapsto X_t(\omega)$$

messbar bezüglich $\mathcal{B}([0,T]) \otimes \mathcal{F}_T - \mathcal{B}(E)$ ist.

Satz 6.23. *Sei $(X_t)_{t\geq 0}$ ein stochastischer Prozess und $f : E \longrightarrow \mathbb{R}$ beschränkt und messbar.*

(a) Ist X progressiv messbar, so ist X adaptiert und messbar.

(b) Ist X adaptiert und messbar, so gibt es eine Modifikation Y von X derart, dass X progressiv messbar ist.

(c) Ist X adaptiert und rechtsstetig, so ist X progressiv messbar.

(d) Ist X progressiv messbar, so ist $f \circ X$ progressiv messbar und

$$Y_t := \int_0^t f(X_s)ds$$

adaptiert.

(e) Ist X messbar und adaptiert, so ist $f \circ X$ adaptiert und messbar und Y_t hat eine progressiv messbare Modifikation.

Beweis. (a) klar.
(b) Siehe ...
(c) Approximiere X durch (6.3).
(d) Sei $A \in \mathcal{B}(\mathbb{R})$, dann gilt für alle $T > 0$

$$\{(\omega,t) \in \Omega \times [0,T] \mid f(X_t(\omega)) \in A\} = \{(\omega,t) \in \Omega \times [0,T] \mid X_t(\omega) \in f^{-1}(A)\} \in \mathcal{F} \otimes \mathcal{B}([0,T])$$

und folglich ist $f \circ X$ progressiv messbar. Für festes t ist $f(X_s)$ messbar bezüglich $\mathcal{F}_s \subset \mathcal{F}_t$ für alle $s \in [0,t]$, welches die Behauptung zeigt.

(e) $f \circ X$ ist messbar, als Verkettung messbarer Funktionen. Sei $t \geq 0$ und $A \in \mathcal{B}(E)$. Dann gilt

$$\{\omega \in \Omega \mid f(X_t(\omega)) \in A\} = \{\omega \in \Omega \mid X_t(\omega) \in f^{-1}(A)\} \in \mathcal{F}_t$$

und folglich ist $f \circ X$ adaptiert. Folglich ist Y_t messbar und adaptiert. Die Behauptung folgt aus (b). $\square$

6.3 Stoppzeiten

Definition 6.24. *Eine Abbildung $T : \Omega \longrightarrow [0,\infty]$ heißt Stoppzeit, falls*

$$\{T \leq t\} \in \mathcal{F}_t, \quad t \geq 0.$$

Die Stoppzeit T heißt diskret, falls $\{T \in D\} = \Omega$ für eine abzählbare Menge $D \subset [0,\infty]$ gilt.

Für $s < t$ auch $\{T \leq s\} \in \mathcal{F}_s \subset \mathcal{F}_t$ und folglich

$$\{T < t\} = \bigcup_{n \geq 1} \left\{T \leq t - \frac{1}{n}\right\} \in \mathcal{F}_t \tag{6.4}$$

und $\{T = t\} = \{T \leq t\} \backslash \{T < t\} \in \mathcal{F}_t$. Ist T zusätzlich diskret, so ist T eine Stoppzeit genau dann, wenn $\{T = t\} \in \mathcal{F}_t$ für alle $t \in D \cap [0,\infty)$.

Lemma 6.25. *Sei $(T_n)_{n\in\mathbb{N}}$ eine Folge von Stoppzeiten, T, S weitere Stoppzeiten.*

(a) $\min\{T, S\}$ *und $\alpha \cdot T$ sind Stoppzeiten für $\alpha > 1$.*

(b) T ist eine Stoppzeit bezüglich $(\mathcal{F}_{t+})_{t\geq 0}$ genau dann, wenn $\{T < t\} \in \mathcal{F}_t$ für alle $t \geq 0$.

(c) $\sup\limits_{n\in\mathbb{N}} T_n$ *ist eine Stoppzeit.*

(d) $\min\limits_{k\leq n} T_k$ *ist eine Stoppzeit für alle $n \geq 1$.*

(e) Ist $(\mathcal{F}_t)_{t\geq 0}$ rechtsstetig, so sind auch $\inf\limits_{n\in\mathbb{N}} T_n$, $\liminf\limits_{n\to\infty} T_n$ *und* $\limsup\limits_{n\to\infty} T_n$ *Stoppzeiten.*

Beweis. (a) Es gilt

$$\{\min\{T,S\} \leq t\} = \{T \leq t\} \cap \{S \leq t\} \in \mathcal{F}_t.$$

Weiterhin gilt

$$\{\alpha \cdot T \leq t\} = \left\{T \leq \frac{t}{\alpha}\right\} \in \mathcal{F}_{\frac{t}{\alpha}} \subset \mathcal{F}_t.$$

(b) Ist $\{T < t\} \in \mathcal{F}_t$ für alle $t \geq 0$, so gilt $\{T < t + n^{-1}\} \in \mathcal{F}_{t+m^{-1}}$ für alle $n \geq m$ und folglich

$$\{T \leq t\} = \bigcap_{n\geq 1} \{T < t + n^{-1}\} \in \bigcap_{m \geq 1} \mathcal{F}_{t+m^{-1}} = \mathcal{F}_{t+}.$$

Die Umkehrung folgt aus (6.4).
(c) Es gilt $\{\sup\limits_{n\geq 1} T_n \leq t\} = \bigcap\limits_{n\geq 1} \{T_n \leq t\} \in \mathcal{F}_t$.
(d) Es gilt $\{\min\limits_{k\leq n} T_k \leq t\} = \bigcup\limits_{k\leq n} \{T_k \leq t\} \in \mathcal{F}_t$.
(e) Es gilt

$$\{\inf_{n\geq 1} T_n < t\} = \bigcup_{n\geq 1} \{T_n < t\} \in \mathcal{F}_t$$

und da $(\mathcal{F}_t)_{t\geq 0}$ rechtsstetig ist, folgt die Behauptung aus (b). Wegen $\liminf\limits_{n\to\infty} T_n = \sup\limits_{m\geq 1} \inf\limits_{n\geq m} T_n$ und $\limsup\limits_{n\to\infty} T_n = \inf\limits_{m\geq 1} \sup\limits_{n\geq m} T_n$ folgt die Behauptung. $\square$

Insbesondere lässt sich jede Stoppzeit T durch eine Folge von beschränkten Stoppzeiten $T_n := \max\{T, n\}$ von unten approximieren. Eine weitere nützliche Approximationseigenschaft ist im folgenden Satz formuliert.

Satz 6.26. *Für $n \in \mathbb{N}$ sei $0 = t_0^{(n)} < t_1^{(n)} \dots$ eine Folge mit $\lim\limits_{k\to\infty} t_k^{(n)} = \infty$ und*

$$\lim_{n\to\infty} \sup_{k\in\mathbb{N}} (t_{k+1}^{(n)} - t_k^{(n)}) = 0.$$

Sei T eine $(\mathcal{F}_{t+})_{t\geq 0}$ Stoppzeit und

$$T_n := \begin{cases} t_{k+1}^{(n)}, & t_k^{(n)} \leq T < t_{k+1}^{(n)},\ k \geq 0 \\ \infty, & T = \infty \end{cases}.$$

Dann ist T_n eine Stoppzeit und es gilt $\lim\limits_{n\to\infty} T_n = T$. Ist zusätzlich $(t_k^{(n)})_{k\in\mathbb{N}} \subset (t_k^{(n+1)})_{k\in\mathbb{N}}$ für alle $n \geq 1$, so gilt $T_{n+1} \leq T_n$.

Beweis. Sei $a_n(t) := \max\{t_k^{(n)} \mid t_k^{(n)} \leq t\}$. Dann gilt

$$\{T_n \leq t\} = \{T_n = a_n(t)\} = \{T < a_n(t)\} \in \mathcal{F}_{a_n(t)} \subset \mathcal{F}_t.$$

Der zweite Teil der Behauptung ist klar. □

Definition 6.27. *Sei T eine Stoppzeit. Definiere*

$$\mathcal{F}_T := \{A \in \mathcal{F} \mid A \cap \{T \leq t\} \in \mathcal{F}_t, \quad \forall t \geq 0\}$$

und

$$\mathcal{F}_{T+} := \{A \in \mathcal{F} \mid A \cap \{T \leq t\} \in \mathcal{F}_{t+}, \quad \forall t \geq 0\}.$$

Es ist nicht schwer zu zeigen, dass $\mathcal{F}_T$ und $\mathcal{F}_{T+}$ σ-Algebren sind. Ebenfalls ist es nicht schwer zu zeigen, dass T messbar bezüglich $\mathcal{F}_T$ ist, denn für $r \geq 0$ gilt

$$\{T \leq c\} \cap \{T \leq t\} = \{T \leq r \wedge t\} \in \mathcal{F}_t$$

und somit $\{T \leq c\} \in \mathcal{F}_T$.

Lemma 6.28. *Seien T, S Stoppzeiten. Dann gelten die folgenden Eigenschaften.*

(a) $\min\{T, S\}$ *ist messbar bezüglich $\mathcal{F}_T$.*

(b) Ist $T \leq S$, so gilt $\mathcal{F}_T \subset \mathcal{F}_S$.

Beweis. (a) Für alle $s, t \geq 0$ gilt

$$\begin{aligned} \{S \wedge T \leq s\} \cap \{T \leq t\} &= \{S \wedge T \leq s \wedge t\} \cap \{T \leq t\} \\ &= (\{S \leq s \wedge t\} \cup \{T \leq s \wedge t\}) \cap \{T \leq t\} \in \mathcal{F}_t. \end{aligned}$$

Folglich gilt $\{S \wedge T \leq s\} \in \mathcal{F}_T$ und damit ist $S \wedge T$ messbar.
(b) Sei $A \in \mathcal{F}_T$ und $t \geq 0$ beliebig. Dann gilt

$$A \cap \{S \leq t\} = A \cap \{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t$$

da $A \cap \{T \leq t\} \in \mathcal{F}_t$ und $\{S \leq t\} \in \mathcal{F}_t$. □

Sei $X = (X_t)_{t\geq 0}$ ein stochastischer Prozess und T eine Stoppzeit. Definiere

$$X_T : \Omega \longrightarrow E, \quad \omega \longmapsto X_{T(\omega)}(\omega),$$

wo $X_\infty := x_0$ mit einem beliebigen (festen) Punkt $x_0 \in E$. Der zum Zeitpunkt T gestoppte Prozess ist definiert durch

$$(X^T)_t := X_{T\wedge t} = X_t \mathbb{1}_{t<T} + X_T \mathbb{1}_{T\leq t}.$$

Betrachte die dazugehörigen σ-Algebren $\mathcal{F}_T$ sowie

$$\mathcal{F}_{T\wedge t} = \{A \in \mathcal{F} \mid A \cap \{\min\{T, t\} \leq s\} \in \mathcal{F}_s, \quad \forall s \geq 0\}.$$

Satz 6.29. *Es sei $(X_t)_{t\geq 0}$ progressiv messbar bezüglich der Filtration $(\mathcal{F}_t)_{t\geq 0}$. Dann gelten die folgenden Eigenschaften.*

(a) X_T ist messbar bezüglich $\mathcal{F}_T$.

(b) X^T ist $(\mathcal{F}_{T\wedge t})_{t\geq 0}$ sowie $(\mathcal{F}_t)_{t\geq 0}$ progressiv messbar, wobei $\mathcal{F}_{T\wedge t}$ eine Filtration ist.

(c) $(X_{T+t})_{t\geq 0}$ ist $(\mathcal{F}_{T+t})_{t\geq 0}$ progressiv messbar, wo

$$\mathcal{F}_{T+t} = \{A \in \mathcal{F} \mid A \cap \{T + t \leq s\} \in \mathcal{F}_s, \quad \forall s \geq 0\}.$$

eine Filtration ist.

Beweis. (a) Sei $t \geq 0$, dann ist $T \wedge t$ messbar bezüglich $\mathcal{F}_t$. Also ist auch $\omega \longmapsto (T(\omega) \wedge t, \omega)$ messbar bezüglich $\mathcal{F}_t - -\mathcal{B}([0, t]) \otimes \mathcal{F}_t$. Da X progressiv messbar ist, ist auch die Komposition $\omega \longmapsto X_{T(\omega)\wedge t}(\omega)$ messbar bezüglich $\mathcal{F}_t$. Folglich gilt für $A \in \mathcal{B}(E)$

$$\{X_T \in A\} \cap \{T \leq t\} = \{X_{T\wedge t} \in A\} \cap \{T \leq t\} \in \mathcal{F}_t$$

und somit $\{X_T \in A\} \in \mathcal{F}_T$.
(b), (c) siehe [EK86, Chapter 2, Proposition 1.4]. □

Wir wollen uns abschließend mit Beispielen von Stoppzeiten beschäftigen. Es sei $(X_t)_{t\geq 0}$ ein stoch. Prozess mit Werten in E. Gegeben $A \in \mathcal{B}(E)$, dann ist die Ersteintrittszeit von X in A definiert durch

$$T_A = \inf\{t \geq 0 \mid X_t \in A\}$$

wo $\inf \varnothing := +\infty$. Es sei σ eine $[0, \infty]$-wertige Zufallsvariable. Dann ist die Ersteintrittszeit von X in A nach σ definiert durch

$$T_{\sigma,A} = \inf\{t \geq \sigma \mid X_t \in A\}.$$

Beachte, für $\sigma = 0$ ist $T_{\sigma,A} = T_A$.

Satz 6.30. *Es sei $X = (X_t)_{t\geq 0}$ ein rechtsstetiger stoch. Prozess, $(\mathcal{F}_t)_{t\geq 0}$ adaptiert und σ eine $(\mathcal{F}_t)_{t\geq 0}$ Stoppzeit. Ist $A \subset E$ offen, so ist $T_{\sigma,A}$ eine Stoppzeit bezüglich $(\mathcal{F}_{t+})_{t\geq 0}$.*

Beweis. Es gilt
$$\{T_{\sigma,A} < t\} = \bigcup_{n\geq 1} \{X_{t-\frac{1}{n}} \in A\} \cap \left\{\sigma < t + \frac{1}{n}\right\} \in \mathcal{F}_t.$$
□

Es gilt ein stärkeres Resultat. Hierfür brauchen wir, dass der Wahrscheinlichkeitsraum vollständig ist.

Definition 6.31. *Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt vollständig, falls $\mathcal{F}$ alle Nullmengen enthält. D.h. ist $A \in \mathcal{F}$ mit $\mathbb{P}(A) = 0$, so gilt für alle $B \subset A$ bereits $B \in \mathcal{F}$.*

Satz 6.32. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei*
$$\mathcal{N} = \{A \subset \Omega \mid \exists B \in \mathcal{F} : \quad A \subset B,\ \mathbb{P}(B) = 0\}.$$
Dann ist $(\Omega, \mathcal{F}^*, \mathbb{P}^*)$ ein vollständiger Wahrscheinlichkeitsraum, wo
$$\mathcal{F}^* := \{A \cup N \mid A \in \mathcal{F},\ N \in \mathcal{N}\}$$
eine σ-Algebra und
$$\mathbb{P}^*(A \cup N) := \mathbb{P}(A), \quad A \in \mathcal{F}, \quad N \in \mathcal{N}$$
ein Wahrscheinlichkeitsmaß auf $\mathcal{F}^*$ ist.

Definition 6.33. *Ein filtrierter Raum $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ erfüllt die üblichen Bedingungen, falls gilt:*

- *$(\Omega, \mathcal{F}, \mathbb{P})$ ist vollständig.*
- *$\mathcal{N} \subset \mathcal{F}_0$.*
- *$(\mathcal{F}_t)_{t\geq 0}$ ist rechtsstetig.*

Satz 6.34. *Sei $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ ein filtrierter Raum mit den üblichen Bedingungen, $(X_t)_{t\geq 0}$ ein bezüglich $(\mathcal{F}_t)_{t\geq 0}$ progressiv messbarer stochastischer Prozess. Dann ist für jedes $A \in \mathcal{B}(E)$ die Ersteintrittszeit T_A eine Stoppzeit bezüglich $(\mathcal{F}_t)_{t\geq 0}$.*

7 Martingale

Definition 7.1. *Es sei $(M_t)_{t\geq 0}$ ein reellwertiger, stochastischer Prozess adaptiert an eine Filtration $(\mathcal{F}_t)_{t\geq 0}$ mit $\mathbb{E}(|M_t|) < \infty$ für alle $t \geq 0$. Dann ist M_t ein*

- *Submartingal, falls $M_s \leq \mathbb{E}(M_t|\mathcal{F}_s)$ für alle $s \leq t$.*
- *Supermartingal, falls $M_s \geq \mathbb{E}(M_t|\mathcal{F}_s)$ für alle $s \leq t$.*
- *Martingal, falls $M_s = \mathbb{E}(M_t|\mathcal{F}_s)$ für alle $s \leq t$.*

Sei $(M_t)_{t\geq 0}$ ein Submartingal, $s \leq t$ und $A \in \mathcal{F}_s$. Dann gilt

$$\int\limits_A M_s d\mathbb{P} \leq \int\limits_A \mathbb{E}(M_t|\mathcal{F}_s) d\mathbb{P} = \int\limits_A M_t d\mathbb{P}.$$

Umgekehrt $(M_t)_{t\geq 0}$ ist genau dann ein Submartingal, wenn obige Ungleichung gilt. Analog ist $(M_t)_{t\geq 0}$ genau dann ein Supermartingal, wenn für alle $A \in \mathcal{F}_s$

$$\int\limits_A M_s d\mathbb{P} \geq \int\limits_A M_t d\mathbb{P}.$$

$(M_t)_{t\geq 0}$ ist genau dann ein Martingal, wenn für alle $A \in \mathcal{F}_s$

$$\int\limits_A M_s d\mathbb{P} = \int\limits_A M_t d\mathbb{P}.$$

Das nächste Lemma folgt aus der Jensen Ungleichung für bedingte Erwartungswerte.

Lemma 7.2. *(a) Es sei $(M_t)_{t\geq 0}$ ein Martingal und $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ konvex mit $\mathbb{E}(|\varphi(M_t)|) < \infty$ für alle $t \geq 0$. Dann ist $(\varphi \circ M_t)_{t\geq 0}$ ein Submartingal.*

(b) Es sei $(M_t)_{t\geq 0}$ ein Submartingal und $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ konvex und monoton steigend mit $\mathbb{E}(|\varphi(M_t)|) < \infty$ für alle $t \geq 0$. Dann ist $(\varphi \circ M_t)_{t\geq 0}$ ein Submartingal.

Beweis. (a) Für $s \leq t$ folgt die Behauptung aus

$$\varphi(M_s) = \varphi\left(\mathbb{E}(M_t|\mathcal{F}_s)\right) \leq \mathbb{E}(\varphi(M_t)|\mathcal{F}_s).$$

(b) Für $s \leq t$ folgt die Behauptung aus

$$\varphi(M_s) \leq \varphi\left(\mathbb{E}(M_t|\mathcal{F}_s)\right) \leq \mathbb{E}(\varphi(M_t)|\mathcal{F}_s).$$

□

Bemerkung 7.3. *Es sei $(M_t)_{t\geq 0}$ ein an die Filtration $(\mathcal{F}_t)_{t\geq 0}$ adaptierter stochastischer Prozess. Dann gelten:*

(a) Ist M_t ein Submartingal, so $\mathbb{E}(M_s) \leq \mathbb{E}(M_t)$.

(b) Ist M_t ein Supermartingal, so $\mathbb{E}(M_s) \geq \mathbb{E}(M_t)$.

(c) Ist M_t ein Martingal, so $\mathbb{E}(M_s) = \mathbb{E}(M_t)$.

Bevor wir die allgemeine Theorie der Martingale uns ansehen, wollen wir einige Beispiele Betrachten. Wir beginnen mit einem diskreten Beispiel.

Beispiel 7.4. *Sei* $(Z_n)_{n\in\mathbb{N}_0}$ *eine Folge unabhängig, identisch verteilter Zufallsvariablen mit* $Z_1 \in \mathcal{L}^1$ *und* $\mathbb{E}(Z_1) = 0$*. Definiere eine Filtration durch*

$$\mathcal{F}_n := \sigma(Z_k \mid 0 \leq k \leq n).$$

Dann ist

$$M_n := \sum_{k=1}^{n} Z_k, \quad n \geq 0$$

ein zeitdiskretes Martingal in dem folgenden Sinne:

- M_n *ist messbar bezüglich* $\mathcal{F}_n$ *und es gilt* $\mathbb{E}(|M_n|) < \infty$.
- *Für alle* $k < n$ *gilt*
$$\mathbb{E}(M_n|\mathcal{F}_k) = M_k.$$

Die erste Eigenschaft ist klar. Für die zweite Eigenschaft erhalten wir

$$\mathbb{E}(M_n|\mathcal{F}_k) = \sum_{j=0}^{k} Z_j + \sum_{j=k+1}^{n} \mathbb{E}(Z_j|\mathcal{F}_k) = M_k + \sum_{j=k+1}^{n} \mathbb{E}(Z_j) = M_k.$$

Das nächste Beispiel ist eine zeitkontinuierliche Version.

Beispiel 7.5. *Es sei* $(Z_t)_{t\geq 0}$ *ein stetiger stochastischer Prozess, mit* $0 \leq Z_t \in \mathcal{L}^1$, $\mathbb{E}(Z_t) = 0$ *sowie*

$$\int_0^t \mathbb{E}(Z_s)ds < \infty$$

für alle $t \geq 0$*. Definiere eine Filtration* $\mathcal{F}_t := \sigma(Z_s \mid s \leq t)$*. Dann ist*

$$M_t(\omega) := \int_0^t Z_s(\omega)ds, \quad t \geq 0$$

ein Submartingal.

Beweis. Da Z_t stetig und bezüglich $\mathcal{F}_t$ adaptiert ist, ist $(Z_t)_{t\geq 0}$ progressiv messbar und folglich ist M_t adaptiert. Ferner gilt nach Voraussetzung

$$\mathbb{E}(M_t) = \int_0^t \mathbb{E}(Z_s)ds < \infty.$$

Es sei $0 \leq s \leq t$, dann gilt

$$\mathbb{E}(M_t|\mathcal{F}_s) = \int_0^s \mathbb{E}(Z_r|\mathcal{F}_s)dr + \int_s^t \mathbb{E}(Z_r|\mathcal{F}_s)dr = \int_0^s Z_r dr + \int_s^t \mathbb{E}(Z_r|\mathcal{F}_s)dr \geq M_s.$$

□

Beispiel 7.6. *(Brownsche Bewegung) Es sei $(B_t)_{t\geq 0}$ die Brownsche Bewegung. Definiere eine Filtration*

$$\mathcal{F}_t := \sigma(B_s \mid s \leq t).$$

Dann ist $(B_t)_{t\geq 0}$ ein Martingal bezüglich $(\mathcal{F}_t)_{t\geq 0}$.

Beweis. Per Definition ist $(B_t)_{t\geq 0}$ adaptiert. Da B_t für $t \geq 0$ Normalverteilt ist mit Varianz t, folgt

$$\mathbb{E}(|B_t|) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} |x| e^{-\frac{|x|^2}{2t}} dx < \infty.$$

Es sei $0 \leq s \leq t$, dann ist $B_t - B_s$ wegen Unabhängigkeit der Inkremente und der Wahl der Filtration unabhängig von $\mathcal{F}_s$. Daraus folgt

$$\mathbb{E}(B_t|\mathcal{F}_s) = \mathbb{E}(B_t - B_s|\mathcal{F}_s) + \mathbb{E}(B_s|\mathcal{F}_s) = \mathbb{E}(B_t - B_s) + B_s = B_s.$$

□

Eine allgemeine Brownsche Bewegung definieren wir wie folgt.

Definition 7.7. *Es sei $(B_t)_{t\geq 0}$ ein stochastischer Prozess und $(\mathcal{F}_t)_{t\geq 0}$ eine Filtration. Dann heißt $(B_t)_{t\geq 0}$ eine Brownsche Bewegung bezüglich $(\mathcal{F}_t)_{t\geq 0}$, falls die folgenden Eigenschaften gelten:*

- *$B_0 = 0$ fast sicher.*
- *$B_t - B_s$ ist unabhängig von $\mathcal{F}_s$ und Normalverteilt mit Erwartungswert 0 und Varianz $t - s$ für alle $0 \leq s \leq t$.*

Obiges Beispiel zeigt, dass $(B_t)_{t\geq 0}$ ein Martingal bezüglich $(\mathcal{F}_t)_{t\geq 0}$ ist. Allgemeiner ist der nächste Satz.

Satz 7.8. *Es sei $(M_t)_{t\geq 0}$ ein stochastischer Prozess adaptiert and die Filtration $(\mathcal{F}_t)_{t\geq 0}$ mit $\mathbb{E}(|M_t|) < \infty$ für alle $t \geq 0$. Es gelte*

- *$\mathbb{E}(M_0) = 0$.*
- *$M_t - M_s$ sei unabhängig von $\mathcal{F}_s$ für alle $0 \leq s \leq t$.*

Dann sind die folgenden stochastischen Prozesse Martingale:

(a) $(M_t)_{t\geq 0}$.

(b) Ist $\mathbb{E}(M_t^2) < \infty$ für alle $t \geq 0$, so ist $M_t^2 - \mathbb{E}(M_t^2)$ ein Martingal.

Beweis. (a) Es gilt

$$\mathbb{E}(M_t|\mathcal{F}_s) = \mathbb{E}(M_t - M_s|\mathcal{F}_s) + M_s = \mathbb{E}(M_t - M_s) + M_s = M_s.$$

(b) Da $(M_t)_{t\geq 0}$ ein Martingal ist, folgt

$$\begin{aligned}\mathbb{E}((M_t - M_s)^2|\mathcal{F}_s) &= \mathbb{E}(M_t^2 - 2M_tM_s + M_s^2|\mathcal{F}_s)\\ &= \mathbb{E}(M_t^2|\mathcal{F}_s) + M_s^2 + 2M_s\mathbb{E}(M_t|\mathcal{F}_s) = \mathbb{E}(M_t^2 - M_s^2|\mathcal{F}_s).\end{aligned}$$

Daraus folgt

$$\begin{aligned}\mathbb{E}(M_t^2|\mathcal{F}_s) &= \mathbb{E}(M_t^2 - M_s^2|\mathcal{F}_s) + M_s^2 = \mathbb{E}((M_t - M_s)^2|\mathcal{F}_s) + M_s^2\\ &= \mathbb{E}(M_t^2 - M_s^2) + M_s^2\end{aligned}$$

und somit

$$\mathbb{E}(M_t^2 - \mathbb{E}(M_t^2)|\mathcal{F}_s) = M_s^2 - \mathbb{E}(M_s^2).$$

□

7.1 Martingale in diskreter Zeit

Definition 7.9. *Sei $(M_n)_{n\in\mathbb{N}_0}$ eine Folge von Zufallsvariablen auf einem filtrierten Raum $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}_0}, \mathbb{P})$ mit M_n ist messbar bezüglich $\mathcal{F}_n$ und $\mathbb{E}(|M_n|) < \infty$. Dann heißt $(M_n)_{n\in\mathbb{N}_0}$ zeitdiskretes*

- *Submartingal, falls $M_k \leq \mathbb{E}(M_n|\mathcal{F}_k)$*
- *Supermartingal, falls $M_k \geq \mathbb{E}(M_n|\mathcal{F}_k)$*
- *Martingal, falls $M_k = \mathbb{E}(M_n|\mathcal{F}_k)$.*

für alle $k < n$ gilt.

Lemma 7.10. *Sei $(M_n)_{n\in\mathbb{N}_0}$ eine Folge von Zufallsvariablen auf einem filtrierten Raum $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}_0}, \mathbb{P})$ mit M_n ist messbar bezüglich $\mathcal{F}_n$ und $\mathbb{E}(|M_n|) < \infty$. Dann gilt:*

(a) M_n ist ein Submartingal $\Longleftrightarrow M_{n-1} \leq \mathbb{E}(M_n|\mathcal{F}_{n-1})$

(b) M_n ist ein Supermartingal $\Longleftrightarrow M_{n-1} \geq \mathbb{E}(M_n|\mathcal{F}_{n-1})$

(c) M_n ist ein Martingal $\Longleftrightarrow M_{n-1} = \mathbb{E}(M_n|\mathcal{F}_{n-1})$.

Beweis. Wir zeigen nur den Fall eines Submartingals. Die anderen Fälle gehen analog. Die Richtung $\Longrightarrow$ ist klar. Für die Umkehrung sei $k < n$.

$$M_k \leq \mathbb{E}(M_{k+1}|\mathcal{F}_k) \leq \mathbb{E}(\mathbb{E}(M_{k+2}|\mathcal{F}_{k+1})|\mathcal{F}_k) = \mathbb{E}(M_{k+2}|\mathcal{F}_k).$$

Iteration liefert nach endlich vielen Schritten die Behauptung. □

Der erste Satz zeigt, dass jedes Submartingal zerlegt werden kann in die Summe von einem Martingal und einem monoton steigenden stochastischen Prozess.

Satz 7.11. *(Doob Zerlegung) Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal bezüglich einer Filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Dann gibt es zwei stochastische Prozesse $(M_n)_{n\in\mathbb{N}}$ und $(A_n)_{n\in\mathbb{N}}$ mit den folgenden Eigenschaften*

- $X_n = M_n + A_n$, $n \geq 0$.
- $(M_n)_{n\in\mathbb{N}_0}$ *ist ein Martingal bezüglich* $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$.
- $A_0 = 0$, A_n *ist messbar bezüglich* $\mathcal{F}_{n-1}$ *für alle* $n \geq 1$.
- A_n *ist monoton steigen, d.h.*
$$A_n(\omega) \leq A_{n+1}(\omega)$$
für fast alle ω *und alle* $n \geq 0$.

Gegeben ein weiteres Paar M'_n *und* A'_n *mit denselben Eigenschaften. Dann gilt* $M_n = M'_n$ *sowie* $A_n = A'_n$ *fast sicher.*

Beweis. Angenommen wir haben M_n und A_n bereits konstruiert. Dann gilt für $n \geq 1$

$$\begin{aligned} X_{n-1} &= M_{n-1} + A_{n-1} \\ X_n &= M_n + A_n. \end{aligned}$$

Betrachten wir die Differenz dieser Gleichungen und bedingen auf $\mathcal{F}_{n-1}$, so folgt

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} = A_n - A_{n-1}.$$

Also ist A_n eindeutig durch $(X_n)_{n\in\mathbb{N}_0}$ sowie A_{n-1} festgelegt. Wegen der Bedingung $M_n = X_n - A_n$ ist dann auch M_n eindeutig festgelegt. Beginnen wir mit $M_0 = X_0$ und $A_0 = 0$, so lässt sich die Eindeutigkeit über Induktion nachweisen.

Für die Existenz setze $M_0 := X_0$, $A_0 := 0$ und für $n \geq 1$

$$A_n := \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} - A_{n-1}, \quad M_n := X_n - A_n.$$

Es bleibt zu zeigen, dass die so definierten Prozesse alle Bedingungen erfüllen. Die erste Bedingung folgt direkt aus der Konstruktion. Für die zweite Bedingung rechnen wir nach

$$\begin{aligned} \mathbb{E}(M_n|\mathcal{F}_{n-1}) &= \mathbb{E}(X_n - A_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - A_n \\ &= X_{n-1} - A_{n-1} = M_{n-1}. \end{aligned}$$

Die dritte Bedingung folgt aus der Definition von A_n. Die letzte Bedingung folgt aus

$$A_n = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} + A_{n-1} \geq A_{n-1}$$

da $(X_n)_{n\in\mathbb{N}_0}$ ein Submartigal ist. □

Satz 7.12. *(Optional Sampling) Sei* $(X_n)_{n\in\mathbb{N}_0}$ *ein Submartingal und* T, S *zwei zeitdiskrete Stoppzeiten mit* $S \leq T \leq k$ *für ein* $k \in \mathbb{N}$. *Dann gilt*

$$X_S \leq \mathbb{E}(X_T|\mathcal{F}_S). \tag{7.1}$$

Ist $(X_n)_{n\in\mathbb{N}_0}$ *ein Martingal bzw. ein Supermartingal, so gilt* $=$ *bzw.* $\geq$ *in* (7.1)

Beweis. Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal, so ist $(-X_n)_{n\in\mathbb{N}_0}$ ein Submartingal. Wir zeigen daher den Fall dass $(X_n)_{n\in\mathbb{N}_0}$ ein Submartigal ist. Sei $A \in \mathcal{F}_S$ und $1 \le m \le n$. Setze $A_m := A \cap \{S = m\}$. Dann ist $A_m \cap \{T > n\} \in \mathcal{F}_n$ und folglich (da $(X_n)_{n\in\mathbb{N}_0}$ ein Submartigal ist)

$$\int\limits_{A_m\cap\{T>n\}} X_n d\mathbb{P} \le \int\limits_{A_m\cap\{T>n\}} X_{n+1} d\mathbb{P}.$$

Daraus folgt

$$\begin{aligned}\int\limits_{A_m\cap\{T\ge n\}} X_n d\mathbb{P} &\le \int\limits_{A_m\cap\{T>n\}} X_n d\mathbb{P} + \int\limits_{A_m\cap\{T=n\}} X_n d\mathbb{P} \\ &\le \int\limits_{A_m\cap\{T>n\}} X_{n+1} d\mathbb{P} + \int\limits_{A_m\cap\{T=n\}} X_n d\mathbb{P} \\ &= \int\limits_{A_m\cap\{T\ge n+1\}} X_{n+1} d\mathbb{P} + \int\limits_{A_m\cap\{T=n\}} X_n d\mathbb{P}\end{aligned}$$

und somit

$$\int\limits_{A_m\cap\{T\ge n\}} X_n d\mathbb{P} - \int\limits_{A_m\cap\{T\ge n+1\}} X_{n+1} d\mathbb{P} \le \int\limits_{A_m\cap\{T=n\}} X_n d\mathbb{P}.$$

Summieren über $n = m$ bis k liefert mittels Teleskopsumme

$$\begin{aligned}\int\limits_{A_m} X_m d\mathbb{P} &= \int\limits_{A_m\cap\{T\ge m\}} X_m d\mathbb{P} - \int\limits_{A_m\cap\{T\ge k+1\}} X_{k+1} d\mathbb{P} \\ &= \sum_{n=m}^{k} \left(\int\limits_{A_m\cap\{T\ge n\}} X_n d\mathbb{P} - \int\limits_{A_m\cap\{T\ge n+1\}} X_{n+1} d\mathbb{P} \right) \\ &\le \sum_{n=m}^{k} \int\limits_{A_m\cap\{T=n\}} X_n d\mathbb{P} = \int\limits_{A_m} X_T d\mathbb{P}.\end{aligned}$$

Bilden wir jetzt die Summe von $m = 1$ bis k, so folgt

$$\int\limits_A X_S d\mathbb{P} \le \int\limits_A X_T d\mathbb{P}$$

und somit (7.1). □

Definition 7.13. *Eine Familie von Zufallsvariablen $(X_t)_{t\in T}$, wo T eine beliebige Menge ist, heißt gleichgradig integrierbar, falls*

$$\lim_{n\to\infty} \sup_{t\in T} \mathbb{E}(\mathbb{1}_{|X_t|>n}|X_t|) = 0.$$

Lemma 7.14. *Eine Familie von Zufallsvariablen $(X_t)_{t\in T}$ ist genau dann gleichgradig integrierbar, wenn gilt*

- $\sup_{t\in T} \mathbb{E}(|X_t|) < \infty$.
- *Für alle $\varepsilon > 0$ gibt es ein $\delta > 0$ mit*
  $$\mathbb{E}(\mathbb{1}_A |X_t|) \le \varepsilon, \quad t \in T$$
  für alle A mit $\mathbb{P}(A) \le \delta$.

Beweis. Übung. □

Bemerkung 7.15. *Optional Sampling gilt im allgemeinen nur für beschränkte Stoppzeiten $S \le T \le k$. Ist hingegen $(X_n)_{n\in\mathbb{N}}$ gleichgradig integrierbar, so gilt Optional Sampling für beliebige Stoppzeiten $S \le T$.*

Satz 7.16. *(Doob Ungleichung) Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal. Dann gilt für alle $n \in \mathbb{N}$ und jedes $\varepsilon > 0$*

$$\mathbb{P}(\max_{0\le i\le n} X_i \ge \varepsilon) \le \int\limits_{\max\limits_{0\le i\le n} X_i \ge \varepsilon} X_n d\mathbb{P} \le \frac{1}{\varepsilon}\mathbb{E}((X_n)^+),$$

wo $(X_n)^+ = \max\{X_n, 0\}$.

Beweis. Sei

$$S = \min\{k \in \mathbb{N}_0 \mid X_k \ge \varepsilon\}$$

mit $\min \varnothing := \infty$. Dann sind $S \wedge n \le n =: T$ beschränkte Stoppzeiten. Ferner gilt für $m \le n$

$$\{\max_{0\le k\le n} X_k \ge \varepsilon\} \cap \{S \wedge n \le m\} = \{\max_{1\le k\le m} X_k \ge \varepsilon\} \in \mathcal{F}_m.$$

Für $m > n$ ist $\{\max_{0\le k\le n} X_k \ge \varepsilon\} \in \mathcal{F}_n \subset \mathcal{F}_m$ und somit

$$\{\max_{0\le k\le n} X_k \ge \varepsilon\} \cap \{S \wedge n \le m\} \in \mathcal{F}_m.$$

Daraus folgt $\{\max_{0\le k\le n} X_k \ge \varepsilon\} \in \mathcal{F}_{S\wedge n}$ und mit Optional Sampling

$$\varepsilon\mathbb{P}\left(\max_{0\le k\le n} X_k \ge \varepsilon\right) \le \int\limits_{\max\limits_{1\le k\le n} X_k \ge \varepsilon} X_{S\wedge n} d\mathbb{P} \le \int\limits_{\max\limits_{0\le k\le n} X_k \ge \varepsilon} X_n d\mathbb{P} \le \mathbb{E}((X_n)^+).$$

□

Korollar 7.17. *Sei $(M_n)_{n\in\mathbb{N}_0}$ ein Martingal oder ein positives Submartingal. Dann gilt für jedes $\varepsilon > 0$ und $p \in [1, \infty)$*

$$\mathbb{P}\left(\sup_{0\le k\le n} |M_k|^p \ge \varepsilon\right) \le \frac{1}{\varepsilon^p}\mathbb{E}(|M_n|^p)$$

Beweis. Da $x \longmapsto |x|^p$ konvex ist, ist $(|M_n|^p)_{n\in\mathbb{N}}$ ein Submartingal. Folglich gilt

$$\varepsilon^p \mathbb{P}\left(\max_{0\le i\le n} |M_n|^p \ge \varepsilon\right) \le \int\limits_{\max\limits_{0\le i\le n} |M_n|^p\ge\varepsilon} |M_n|^p d\mathbb{P} \le \mathbb{E}(|M_n|^p).$$

□

Satz 7.18. *[EK86] Sei $(M_n)_{n\in\mathbb{N}_0}$ ein Martingal oder ein positives Submartingal. Dann gilt für alle $n\in\mathbb{N}$ und $p>1$*

$$\mathbb{E}\left(\sup_{0\le k\le n} |M_k|^p\right) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left(|M_n|^p\right).$$

Beispiel 7.19. *Sei $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ und $Y\in\mathcal{L}^1$. Dann definiert*

$$M_n := \mathbb{E}(Y|\mathcal{F}_n), \quad n\in\mathbb{N}_0$$

ein Martingal.

Definition 7.20. *Ein Martingal $(X_n)_{n\in\mathbb{N}_0}$ heißt abgeschlossen, falls es ein $M_\infty\in\mathcal{L}^1$ gibt mit $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$.*

Definiere die σ-Algebra

$$\mathcal{F}_\infty := \sigma(\mathcal{F}_n \mid n\ge 0).$$

Man beachte, dass $\mathcal{F}_\infty$ nicht mit $\mathcal{F}$ übereinstimmen muss. Sei $(M_n)_{n\in\mathbb{N}_0}$ ein abgeschlossenes Martingal mit $M_\infty\in\mathcal{L}^1$ und definiere

$$M'_\infty := \mathbb{E}(M_\infty \mid \mathcal{F}_\infty).$$

Dann gilt

$$\mathbb{E}(M'_\infty|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(M_\infty|\mathcal{F}_\infty)|\mathcal{F}_n) = \mathbb{E}(M_\infty|\mathcal{F}_n) = M_n.$$

Daher können wir stets annehmen, dass M_∞ messbar ist bezüglich $\mathcal{F}_\infty$.

Satz 7.21. *Ein Martingal ist genau dann abgeschlossen, wenn es gleichgradig integrierbar ist.*

Wir zeigen nur die gleichgradige integrierbarkeit. Die Umkehrung lässt sich in [Dud02] nachlesen.

Beweis. Sei $(M_n)_{n\in\mathbb{N}}$ ein abgeschlossenes Martingal mit $M_\infty\in\mathcal{L}^1$ messbar bezüglich $\mathcal{F}_\infty$. Wir müssen zeigen, dass

$$\lim_{m\to\infty}\sup_{n\in\mathbb{N}} \mathbb{E}(\mathbb{1}_{|M_n|>m}|M_n|) = 0.$$

Da $|\cdot|$ eine konvexe Funktion ist folgt

$$|M_n| = |\mathbb{E}(M_\infty|\mathcal{F}_n)| \le \mathbb{E}(|M_\infty||\mathcal{F}_n).$$

Daraus folgt $\mathbb{E}(|M_n|) \le \mathbb{E}(|M_\infty|) < \infty$ für alle $n \in \mathbb{N}$ und

$$\int\limits_{|M_n|>m} |M_n| d\mathbb{P} \le \int\limits_{|M_n|>m} |M_\infty| d\mathbb{P}.$$

Sei $\varepsilon > 0$ beliebig. Wegen $\mathbb{1}_{|M_\infty| \le N} \longrightarrow 1$ fast sicher für $N \to \infty$ und dem Satz von Lebesgue gibt es ein N mit $\mathbb{E}(\mathbb{1}_{|M_\infty|>N} |M_\infty|) < \varepsilon$. Daraus folgt

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \mathbb{E}(\mathbb{1}_{|M_n|>m} |M_n|) &\le \sup_{n\in\mathbb{N}} \mathbb{E}(\mathbb{1}_{|M_n|>m} |M_\infty|) \\
&\le \sup_{n\in\mathbb{N}} \mathbb{E}(\mathbb{1}_{|M_n|>m} \mathbb{1}_{|M_\infty|\le N} |M_\infty|) + \sup_{n\in\mathbb{N}} \mathbb{E}(\mathbb{1}_{|M_n|>m} \mathbb{1}_{|M_\infty|>N} |M_\infty|) \\
&\le N \sup_{n\in\mathbb{N}} \mathbb{P}(|M_n| > m) + \mathbb{E}(\mathbb{1}_{|M_\infty|>N} |M_\infty|).
\end{aligned}$$

Der zweite Term ist kleiner ε und für den ersten gilt

$$\sup_{n\in\mathbb{N}_0} \mathbb{P}(|M_n| > m) \le \sup_{n\in\mathbb{N}} \frac{\mathbb{E}(|M_n|)}{m} \le \frac{1}{m} \mathbb{E}(|M_\infty|).$$

□

Satz 7.22. *(Doob) Sei $(M_n)_{n\in\mathbb{N}_0}$ ein abgeschlossenes Martingal. Dann gilt*

$$M_n \longrightarrow M_\infty, \quad n \to \infty$$

fast sicher und in $\mathcal{L}^1$.

Lemma 7.23. *Es sei F die Menge aller $Y \in \mathcal{L}^1$ derart dass Y messbar bezüglch $\mathcal{F}_n$ für ein $n \in \mathbb{N}$ ist. Für jedes $\varepsilon > 0$ und $X \in \mathcal{L}^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$ gibt es $Y \in F$ mit*

$$|\mathbb{E}(X - Y)| < \varepsilon.$$

Beweis. Sei $\mathcal{K} = \bigcup\limits_{n\in\mathbb{N}_0} \mathcal{F}_n$ und

$$\mathcal{G} = \{A \mid \forall \varepsilon > 0 \; \exists B \in \mathcal{K}: \; \mathbb{P}((A \backslash B) \cup (B \backslash A)) < \varepsilon\}.$$

Dann ist $\mathcal{K}$ ein $\cap$-stabiler Erzeuger von $\mathcal{F}_\infty$ und $\mathcal{G}$ ist ein Dynkin System.

$$\mathcal{F}_\infty = \sigma(\mathcal{K}) = d(\mathcal{K}) \subset \mathcal{G}.$$

Die letzte Inklusion gilt da $\mathcal{F}_n \in \mathcal{G}$ für alle $n \ge 1$.
Schritt 1. Ist $X = \mathbb{1}_A$, so wähle $B \in \mathcal{G}$ mit $\mathbb{P}(A \Delta B) < \varepsilon$. Setze $Y := \mathbb{1}_B \in F$. Dann gilt

$$\mathbb{E}(|X - Y|) = \mathbb{P}(A \Delta B) < \varepsilon.$$

Schritt 2. Sei $X = \sum\limits_{n=1}^{N} a_n \mathbb{1}_{A_n}$ mit paarweise disjunkten Mengen $A_n \in \mathcal{F}_\infty$ und $a_n \ge 0$. Wähle $B_n \in \mathcal{G}$ mit

$$\mathbb{P}(A_n \Delta B_n) < \frac{\varepsilon}{N(1 + \max\limits_{k=1,\dots,N} |a_k|)}.$$

Dann ist $Y := \sum_{k=1}^{N} a_k \mathbb{1}_{B_k} \in F$ und es gilt

$$\mathbb{E}(|X-Y|) \leq \mathbb{E}\left(\sum_{k=1}^{N} |a_k||\mathbb{1}_{A_k} - \mathbb{1}_{B_k}|\right) \leq \max_{k=1,\ldots,N} |a_k| \cdot \sum_{k=1}^{N} \mathbb{P}(A_k \Delta B_k) < \varepsilon.$$

Schritt 3. Sei $X \in \mathcal{L}^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$ mit $0 \leq X \leq M$, wo $M > 0$ eine Konstante. Wähle $X' \geq 0$ Elementarfunktion mit $X' \leq X$ und

$$\mathbb{E}(|X'-X|) < \frac{\varepsilon}{2}.$$

Nach Schritt 2 gibt es $Y \in F$ mit $\mathbb{E}(|Y - X'|) < \frac{\varepsilon}{2}$. Daraus folgt

$$\mathbb{E}(|X-Y|) \leq \mathbb{E}(|X-X'|) + \mathbb{E}(|X'-Y|) < \varepsilon.$$

Schritt 4. Sei $X \in \mathcal{L}^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$ nicht-negativ. Dann gibt es $M > 0$ mit $\mathbb{E}(\mathbb{1}_{|X| \geq M}|X|) < \frac{\varepsilon}{2}$. Ferner gibt es nach Schritt 3 ein $Y \in F$ mit $\mathbb{E}(|Y - \mathbb{1}_{|X| \leq M} X|) < \frac{\varepsilon}{2}$. Daraus folgt

$$\mathbb{E}(|X-Y|) \leq \mathbb{E}(\mathbb{1}_{|X| \geq M}|X|) + \mathbb{E}(|X \mathbb{1}_{|X| \leq M} - Y|) < \varepsilon.$$

Der letzte Schritt folgt durch Zerlegen in Positiv und Negativteil. □

Beweis. Zu $\varepsilon > 0$ und $M_\infty \in \mathcal{L}^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$ wähle $Y_\infty \in F$ mit

$$\mathbb{E}(|M_\infty - Y_\infty|) \leq \varepsilon^2. \tag{7.2}$$

Sei $Y_n := \mathbb{E}(Y_\infty | \mathcal{F}_n)$. Dann ist $(M_n - Y_n)_{n \in \mathbb{N}_0}$ ein Martingal. Also ist $(|M_n - Y_n|)_{n \in \mathbb{N}_0}$ ein Submartingal mit

$$|M_n - Y_n| = |\mathbb{E}(M_\infty - Y_\infty | \mathcal{F}_n)| \leq \mathbb{E}(|M_\infty - Y_\infty| \mid \mathcal{F}_n).$$

Daraus folgt

$$\mathbb{E}(|M_n - Y_n|) \leq \mathbb{E}(|M_\infty - Y_\infty|).$$

Aus Doobs Ungleichung folgt

$$\mathbb{P}\left(\sup_{n \in \mathbb{N}} |M_n - Y_n| > \varepsilon\right) \leq \frac{1}{\varepsilon} \sup_{n \in \mathbb{N}} \mathbb{E}(|M_n - Y_n|) \leq \frac{1}{\varepsilon} \mathbb{E}(|M_\infty - Y_\infty|) \leq \varepsilon. \tag{7.3}$$

Nach der Wahl von F gilt $Y_\infty = Y_n$ für hinreichend große $n \in \mathbb{N}$. Also folgt aus (7.3)

$$\mathbb{P}\left(\limsup_{n \to \infty} (M_n - Y_\infty) > \varepsilon\right) < \varepsilon, \quad \mathbb{P}\left(\liminf_{n \to \infty} (M_n - Y_n) < -\varepsilon\right) < \varepsilon. \tag{7.4}$$

Aus Chebyscheff und (7.2) folgt

$$\mathbb{P}\left(|M_\infty - Y_\infty| > \varepsilon\right) \leq \frac{1}{\varepsilon} \mathbb{E}(|M_\infty - Y_\infty|) \leq \varepsilon.$$

Wir erhalten aus (7.4)

$$\mathbb{P}\left(\limsup_{n\to\infty}(M_n - M_\infty) > 2\varepsilon\right) \le \mathbb{P}\left(\limsup_{n\to\infty}(M_n - Y_\infty > \varepsilon\right) + \mathbb{P}\left(\limsup_{n\to\infty}(Y_\infty - M_\infty) > \varepsilon\right) \le 2\varepsilon$$

und analog

$$\mathbb{P}\left(\liminf_{n\to\infty}(M_n - M_\infty) < -2\varepsilon\right) \le 2\varepsilon.$$

Da $\varepsilon > 0$ beliebig war folgt $M_n \longrightarrow M_\infty$ fast sicher für $n \to \infty$.

Für die Konvergenz in $\mathcal{L}^1$ sei $\varepsilon > 0$ fest. Da $(M_n)_{n\in\mathbb{N}}$ gleichgradig integrierbar ist gibt es $\delta > 0$ derart, dass für alle $A \in \mathcal{F}$ mit $\mathbb{P}(A) \le \delta$

$$\sup_{n\in\mathbb{N}} \int_A |M_n| d\mathbb{P} < \varepsilon, \quad \int_A |M_\infty| d\mathbb{P} < \varepsilon.$$

Da $M_n \longrightarrow M_\infty$ fast sicher konvergiert, gilt Konvergenz in Wahrscheinlichkeit, d.h.

$$\mathbb{P}(|M_n - M_\infty| > \varepsilon) \le \delta$$

für hinreichend große $n \in \mathbb{N}$. Für solche großen n erhalten wir

$$\begin{aligned}\mathbb{E}(|M_n - M_\infty|) \le \mathbb{E}(|M_n| \mathbb{1}_{|M_n - M_\infty| > \varepsilon}) + \mathbb{E}(|M_\infty| \mathbb{1}_{|M_n - M_\infty| > \varepsilon}) \\ + \mathbb{E}(|M_n - M_\infty| \mathbb{1}_{|M_n - M_\infty| \le \varepsilon}) \le \varepsilon.\end{aligned}$$

Da ε beliebig war, folgt die Behauptung. □

Satz 7.24. *(Doob) Sei $(M_n)_{n\in\mathbb{N}}$ ein Martingal mit*

$$\sup_{n\in\mathbb{N}} \mathbb{E}(|M_n|) < \infty.$$

Dann gibt es $Y \in \mathcal{L}^1$ mit $M_n \longrightarrow Y$ fast sicher.

Korollar 7.25. *Sei $(X_n)_{n\in\mathbb{N}}$ ein Submartingal mit*

$$\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|) < \infty.$$

Dann gibt es $Y \in \mathcal{L}^1$ mit $X_n \longrightarrow Y$ fast sicher. Ist $(X_n)_{n\in\mathbb{N}}$ gleichgradig integrierbar, so gilt diese Konvergenz auch in $\mathcal{L}^1$.

Beweis. Sei $X_n = M_n + A_n$ die Doob Zerlegung von X_n. Da $A_n \ge 0$ monoton steigend ist, gibt es $A_\infty := \lim_{n\to\infty} A_n \in [0, \infty]$. Aus dem Satz der monotonen Konvergenz folgt

$$\begin{aligned}\mathbb{E}(A_\infty) &= \lim_{n\to\infty} \mathbb{E}(A_n) \le \sup_{n\in\mathbb{N}} \mathbb{E}(A_n) \\ &\le \sup_{n\in\mathbb{N}} \mathbb{E}(X_n - M_n) = \sup_{n\in\mathbb{N}} \mathbb{E}(X_n - M_1) \le \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|) + \mathbb{E}(|M_1|) < \infty.\end{aligned}$$

Insbesondere gilt $A_n \nearrow A$ in $\mathcal{L}^1$. Ferner gibt es $M_\infty \in \mathcal{L}^1$ mit $M_n \longrightarrow M_\infty$ fast sicher. Daraus folgt $X_n = M_n + A_n \longrightarrow M_\infty + A_\infty$ fast sicher. Ist $(X_n)_{n\in\mathbb{N}}$ gleichgradig integrierbar, so ist $M_n \longrightarrow M_\infty$ in $\mathcal{L}^1$ und damit auch $X_n \longrightarrow M_\infty + A_\infty$ in $\mathcal{L}^1$. □

7.2 Martingale in stetiger Zeit

In diesem Abschnitt betrachten wir Martingale $(M_t)_{t\geq 0}$. Hier und im folgenden ist $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ ein filtrierter Raum mit den üblichen Bedingungen. Bisher haben wir nichts an die Pfadeigenschaften von Martingalen gefordert. Dieses ist in der Regel auch nicht notwendig wie der nächste Satz zeigt.

Satz 7.26. *[KS91] Sei $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ ein filtrierter Raum mit den üblichen Bedingungen und $(X_t)_{t\geq 0}$ ein Submartingal. Genau dann hat $(X_t)_{t\geq 0}$ eine Modifikation $(Y_t)_{t\geq 0}$ wo $(Y_t)_{t\geq 0}$ cadlag ist. Eine solche Modifikation kann derart gewählt werden, dass Y_t an $\mathcal{F}_t$ adaptiert ist.*

Man beachte, dass im obigen Fall $(Y_t)_{t\geq 0}$ wieder ein Submartingal ist. Insbesondere besitzt jedes Martingal eine Modifikation mit rechtsstetigen Pfaden. Im Folgenden betrachten wir stets (Sub-)Martingale mit rechtsstetigen Pfaden.

Definition 7.27. *(Poisson-Prozess) Ein Poisson Prozess mit Intensität $\lambda > 0$ ist ein stochastischer Prozess $(N_t)_{t\geq 0}$ mit den folgenden Eigenschaften.*

- *$N_0 = 0$ fast sicher.*
- *$(N_t)_{t\geq 0}$ ist rechtsstetig und $(\mathcal{F}_t)_{t\geq 0}$-adaptiert.*
- *$N_t - N_s$ ist unabhängig von $\mathcal{F}_s$ fü alle $0 \leq s \leq t$ und es gilt*

$$\mathbb{P}(N_t - N_s = n) = \frac{\lambda^n (t-s)^n}{n!} e^{-\lambda}, \quad n \geq 0.$$

Der Prozess $\widetilde{N}_t := N_t - \lambda t$ heißt kompensierter Poisson Prozess.

Beispiel 7.28. *Ist $(N_t)_{t\geq 0}$ ein Poisson Prozess, so ist $\widetilde{N}_t$ ein Martingal. Nach Definition ist $\widetilde{N}_t$ messbar bezüglich $\mathcal{F}_t$. Wegen*

$$\begin{aligned}\mathbb{E}(N_t) = \mathbb{E}(N_t - N_0) &= \sum_{n=0}^{\infty} \mathbb{P}(N_t - N_0 = n) n \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} n = \lambda t e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1} t^{n-1}}{(n-1)!} = \lambda t.\end{aligned}$$

ist $N_t \in \mathcal{L}^1$ und somit auch $\widetilde{N}_t \in \mathcal{L}^1$. Sei $0 \leq s < t$, dann gilt

$$\begin{aligned}\mathbb{E}(\widetilde{N}_t | \mathcal{F}_s) &= \mathbb{E}(N_t - N_s | \mathcal{F}_s) + \mathbb{E}(N_s - \lambda t | \mathcal{F}_s) \\ &= \mathbb{E}(N_t - N_s) + N_s - \lambda t = \lambda t - \lambda s - \lambda t + N_s = \widetilde{N}_s.\end{aligned}$$

Lemma 7.29. *Sei $(M_t)_{t\geq 0} \subset \mathcal{L}^2$ ein Martingal. Dann gilt*

$$\mathbb{E}((M_t - M_s)^2 \mid \mathcal{F}_s) = \mathbb{E}(M_t^2 - M_s^2 \mid \mathcal{F}_s).$$

Beweis. Es gilt

$$\begin{aligned}\mathbb{E}((M_t - M_s)^2 \mid \mathcal{F}_s) &= \mathbb{E}(M_t^2 - 2M_tM_s + M_s^2 \mid \mathcal{F}_s)\\ &= \mathbb{E}(M_t^2 \mid \mathcal{F}_s) - 2M_s\mathbb{E}(M_t \mid \mathcal{F}_s) + M_s^2\\ &= \mathbb{E}(M_t^2 \mid \mathcal{F}_s) - M_s^2) = \mathbb{E}(M_t^2 - M_s^2 \mid \mathcal{F}_s).\end{aligned}$$

□

Beispiel 7.30. *Sei $(N_t)_{t\geq 0}$ ein Poisson Prozess mit Intensität $\lambda > 0$ und $\widetilde{N}_t$ der kompensierte Poisson Prozess. Dann ist $\widetilde{N}_t^2 - \lambda t =: M_t$ ein Martingal bezüglich $\mathcal{F}_t := \sigma(N_s | \, s \leq t)$.*

Denn M_t ist per Definition messbar bezüglich $\mathcal{F}_t$ und es gilt

$$\mathbb{E}(\widetilde{N}_t^2) \leq 2\mathbb{E}(N_t^2) + 2\lambda t < \infty.$$

Folglich ist $M_t \in \mathcal{L}^2$. Sei $0 \leq s < t$, dann gilt

$$\begin{aligned}\mathbb{E}(\widetilde{N}_t^2 - \widetilde{N}_s^2 \mid \mathcal{F}_s) &= \mathbb{E}((\widetilde{N}_t - \widetilde{N}_s)^2 \mid \mathcal{F}_s) = \mathbb{E}((N_t - N_s - \lambda(t-s))^2 \mid \mathcal{F}_s)\\ &= \mathbb{E}((N_t - N_s - \lambda(t-s))^2) = \operatorname{var}(N_t - N_s) = \lambda(t-s).\end{aligned}$$

Daraus folgt die Behauptung.

Das nächste Beispiel zeigt, dass ein solcher Prozess existiert.

Beispiel 7.31. *Sei $(T_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger exponentialverteilter Zufallsvariablen mit Parameter $\lambda > 0$, d.h.*

$$\mathbb{P} \circ T_n^{-1}(dt) = \lambda e^{-\lambda t}\mathbb{1}_{[0,\infty)}(t)m(dt).$$

Die Zeiten T_n werden als Ankunftszeiten interpretiert. Setze $S_0 = 0$ und $S_n := \sum\limits_{k=1}^{n} T_k$. Dann beschreibt S_n den Zeitpunkt der n-ten Ankunft. Definiere einen $\mathbb{N}_0$-wertigen stochastischen Prozess durch

$$N_t := \max\{n \geq 0 \mid S_n \leq t\}$$

mit Filtration $\mathcal{F}_t := \sigma(N_s \mid s \leq t)$. Dann ist $(N_t)_{t\geq 0}$ ein Poisson Prozess. Für einen Beweis siehe [Chu74].

Im Folgenden brauchen wir einige Sätze aus der Integrationstheorie.

Satz 7.32. *(Satz von Vitali) Es sei $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1$ und $X \in \mathcal{L}^1$. Dann sind äquivalent*

(a) $X_n \longrightarrow X$ in $\mathcal{L}^1$.

(b) $X_n \longrightarrow X$ in Wahrscheinlichkeit und $(X_n)_{n\in\mathbb{N}}$ ist gleichgradig integrierbar.

Beweis. Siehe Maß und Integrationstheorie. □

Lemma 7.33. *(Scheffe's Lemma) Sei $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1$ und X eine weitere Zufallsvariable mit $X_n \longrightarrow X$ fast sicher. Dann sind die folgenden Aussgen äquivalent:*

(a) $(X_n)_{n\in\mathbb{N}}$ *ist gleichgradig integrierbar.*

(b) Es gilt $X \in \mathcal{L}^1$ *und*

$$\mathbb{E}(|X_n|) \longrightarrow \mathbb{E}(|X|), \quad n \to \infty.$$

Insbesondere gilt $X_n \longrightarrow X$ *in* $\mathcal{L}^1$.

Beweis. Es gilt fast sicher $|X_n| \leq |X_n - X| + |X|$ und somit

$$0 \leq |X_n - X| + |X| - |X_n| \leq |X_n| + |X| + |X| - |X_n| = 2|X|. \tag{7.5}$$

(a) $\Longrightarrow$ *(b)* : Sei $m \in \mathbb{N}$ derart, dass $\mathbb{E}(\mathbb{1}_{|X_n|\geq m}|X_n|) \leq 1$ für alle $n \in \mathbb{N}$. Dann folgt aus dem Lemma von Fatou

$$\begin{aligned}\mathbb{E}(|X|) &\leq \liminf_{n\to\infty} \mathbb{E}(|X_n|)\\ &\leq \liminf_{n\to\infty} \mathbb{E}(\mathbb{1}_{|X_n|\geq m}|X_n|) + \liminf_{n\to\infty} \mathbb{E}(\mathbb{1}_{|X_n|<m}|X_n|) \leq 1 + m.\end{aligned}$$

Aus (7.5) folgt mit dem Satz von Lebesgue

$$\mathbb{E}(|X_n - X| + |X| - |X_n|) \longrightarrow 0, \quad n \to \infty. \tag{7.6}$$

Nach dem Satz von Vitali folgt $\mathbb{E}(|X_n - X|) \longrightarrow 0$ und somit $\mathbb{E}(|X_n|) \longrightarrow \mathbb{E}(|X|)$.
(b) $\Longrightarrow$ *(a)* : Aus (7.5) und dem Satz von Lebesgue folgt erneut (7.6). Wegen $\mathbb{E}(|X_n|) \longrightarrow \mathbb{E}(|X|)$ folgt die Behauptung. □

Satz 7.34. *(Optional Sampling) Sei* $(X_t)_{t\geq 0}$ *ein cadlag Submartigal bezüglich* $(\mathcal{F}_t)_{t\geq 0}$. *Seien* $S \leq T$ *beschränkte Stoppzeiten. Dann gilt*

$$X_S \leq \mathbb{E}(X_T|\mathcal{F}_S). \tag{7.7}$$

Ist $(X_t)_{t\geq 0}$ *ein Martingal bzw. ein Supermartingal, so gilt* $=$ *bzw.* $\geq$ *in (7.7).*

Beweis. Definiere Stoppzeiten

$$S_n(\omega) = \begin{cases} \infty, & S(\omega) = +\infty \\ \frac{k}{2^n}, & \frac{k-1}{2^n} \leq S(\omega) < \frac{k}{2^n} \end{cases}.$$

Dann ist S_n eine Stoppzeit mit $S \leq S_n$ und $S_n \searrow S$. Definiere T_n analog. Dann gilt $S_n \leq T_n$.

Lemma 7.35. *(ohne Beweis)* $(X_{S_n})_{n\in\mathbb{N}}$ *sowie* $(X_{T_n})_{n\in\mathbb{N}}$ *sind gleichgradig integrierbar.*

In diesem Fall folgt aus $X_{S_n} \longrightarrow X_S$ sowie $X_{T_n} \longrightarrow X_T$ fast sicher mit dem Satz von Vitali $X_S, X_T \in \mathcal{L}^1$ sowie $X_{S_n} \longrightarrow X_S$, $X_{T_n} \longrightarrow X_T$ in $\mathcal{L}^1$. Ist $A \in \mathcal{F}_S$, so folgt aus $S \leq S_n$ auch $\mathcal{F}_S \subset \mathcal{F}_{S_n}$, und damit $A \in \mathcal{F}_{S_n}$. Aus dem diskreten Optional Sampling Theorem erhalten wir

$$\int_A X_{S_n} d\mathbb{P} \leq \int_A X_{T_n} d\mathbb{P}, \quad n \geq 1.$$

Durch Grenzübergang $n \to \infty$ folgt

$$\int_A X_S d\mathbb{P} \leq \int_A X_T d\mathbb{P}.$$

$\square$

Bemerkung 7.36. *Ist $(X_t)_{t\geq 0}$ gleichgradig integrierbar, so gilt Optional Sampling für alle Stoppzeiten $S \leq T$.*

Korollar 7.37. *Sei $(M_t)_{t\geq 0}$ ein cadlag Martingal und T eine Stoppzeit. Dann ist $M^T := (M_{T\wedge t})_{t\geq 0}$ ein cadlag Martingal bezüglich $(\mathcal{F}_t)_{t\geq 0}$. Ist M stetig, so ist auch M^T stetig.*

Beweis. Wir haben bereits gesehen, dass M^T adaptiert ist. Die Integrierbarkeit wurde im letzten Satz bewiesen. Sei $0 \leq s < t$, dann gilt

$$\mathbb{E}(M_t^T - M_s^T | \mathcal{F}_s) = \mathbb{E}(M_{T\wedge t} - M_{T\wedge s} | \mathcal{F}_s) = \mathbb{E}((M_{(T\wedge t)\vee s} - M_s) | \mathcal{F}_s).$$

Wegen $s \leq (T \wedge t) \vee s \leq T \wedge t \leq t$ impliziert Optional Sampling

$$\mathbb{E}((M_{(T\wedge t)\vee s} - M_s) | \mathcal{F}_s) = \mathbb{E}(M_{(T\wedge t)\vee s} | \mathcal{F}_s) - M_s = M_s - M_s = 0.$$

Die letzte Behauptung ist offensichtlich. $\square$

Korollar 7.38. *Sei $(X_t)_{t\geq 0}$ ein cadlag Submartingal und T eine Stoppzeit. Dann ist $X^T := (X_{T\wedge t})_{t\geq 0}$ ein cadlag Submartingal bezüglich $(\mathcal{F}_t)_{t\geq 0}$. Ist X stetig, so ist auch X^T stetig.*

Beweis. Ersetze an geeigneter Stelle $=$ durch $\geq$. Die Details sind Übung. $\square$

Satz 7.39. *(Doob-Ungleichung) Sei $(X_t)_{t\geq 0}$ ein cadlag Submartingal. Dann gilt für alle $\varepsilon > 0$ und $T > 0$*

$$\mathbb{P}\left(\sup_{t\in[0,T]} X_t \geq \varepsilon\right) \leq \frac{1}{\varepsilon}\mathbb{E}(X_T^+) \tag{7.8}$$

und

$$\mathbb{P}\left(\inf_{t\in[0,T]} X_t \leq -\varepsilon\right) \leq \frac{\mathbb{E}(X_T^+) - \mathbb{E}(X_0)}{\varepsilon}. \tag{7.9}$$

Ist X nicht-negativ, so gilt für jedes $p > 1$

$$\mathbb{E}\left(\sup_{t\in[0,T]} X_t^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_T^p).$$

Beweis. Wir zeigen zuerst (7.8). Da X rechtsstetig ist gilt $\sup_{t\in[0,T]} X_t = \sup_{t\in[0,T]\cap\mathbb{Q}} X_t$. Wähle eine aufsteigende Folge von Mengen $(F_n)_{n\in\mathbb{N}}$ mit $[0,T]\cap\mathbb{Q} = \bigcup_{n\geq 1} F_n$. Dann gilt für $0<\delta<\varepsilon$

$$\mathbb{P}\left(\sup_{t\in[0,T]\cap\mathbb{Q}} X_t \geq \varepsilon\right) \leq \lim_{n\to\infty} \mathbb{P}\left(\max_{t\in F_n} X_t \geq \delta\right) \leq \frac{1}{\delta}\mathbb{E}(X_T^+).$$

Für $\delta\to\varepsilon$ folgt (7.8). Die Ungleichung (7.9) folgt durch eine analoge Beweisidee. Für die letzte Behauptung betrachten wir zuerst

$$\mathbb{E}(\left(\sup_{t\in F_n} X_t^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_T^p), \quad n\geq 1,$$

welches aus der Doob Ungleichung für diskrete Submartingale folgt. Die linke Seite ist monoton wachsend in n mit $\sup_{t\in F_n} X_t^p \nearrow \sup_{t\in[0,T]\cap\mathbb{Q}} X_t^p = \sup_{t\in[0,T]} X_t^p$. Die Behauptung folgt mit $n\to\infty$. $\square$

Korollar 7.40. *(Doob-Ungleichung) Sei $(M_t)_{t\geq 0}$ ein cadlag Martingal. Dann gilt für alle $\varepsilon>0$ und $T>0$*

$$\mathbb{P}\left(\sup_{t\in[0,T]} |M_t| \geq \varepsilon\right) \leq \frac{1}{\varepsilon}\mathbb{E}(|M_T|)$$

und für $p>1$

$$\mathbb{E}\left(\sup_{t\in[0,T]} |M_t|^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|M_T|^p).$$

Wir kommen zu den Martingal Konvergenzsätzen.

Satz 7.41. *(Konvergenz von Submartingalen) Sei $(X_t)_{t\geq 0}$ ein cadlag Submartingal mit*

$$\sup_{t\geq 0} \mathbb{E}(X_t^+) < \infty.$$

Dann gibt es $X_\infty \in \mathcal{L}^1$ und es gilt $X_t \longrightarrow X_\infty$ für $t\to\infty$ fast sicher.

Satz 7.42. *(Konvergenz von Submartingalen) Sei $(X_t)_{t\geq 0}$ ein cadlag Submartingal. Betrachte die Bedingungen:*

(a) $(X_t)_{t\geq 0}$ ist gleichgradig integrierbar.

(b) Es gibt $X_\infty \in \mathcal{L}^1$ mit $X_t \longrightarrow X_\infty$ in $\mathcal{L}^1$.

(c) Es gibt $X_\infty \in \mathcal{L}^1$ mit $X_t \longrightarrow X_\infty$ fast sicher und es gilt $X_t \leq \mathbb{E}(X_\infty|\mathcal{F}_t)$.

Dann gilt (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c). Ist ferner X nicht-negativ, so sind alle Bedingungen äquivalent.

Satz 7.43. *(Konvergenz von Supermartingalen) Sei $(X_t)_{t\geq 0}$ ein nicht-negatives cadlag Supermartingal. Dann gibt es eine Zufallsvariable X_∞ mit $X_t \longrightarrow X_\infty$ für $t\to\infty$ fast sicher.*

Es sei $\mathcal{F}_\infty = \sigma\left(\bigcup_{t\geq 0} \mathcal{F}_t\right)$.

Satz 7.44. *(Konvergenz von Martingalen) Sei $(M_t)_{t\geq 0}$ ein cadlag Martingal. Dann sind äquivalent:*

(a) $(M_t)_{t\geq 0}$ ist gleichgradig integrierbar.

(b) Es gibt $M_\infty \in \mathcal{L}^1$ mit $M_t \longrightarrow M_\infty$ in $\mathcal{L}^1$ für $t \to \infty$.

(c) Es gibt $M_\infty \in \mathcal{L}^1$ mit $M_t \longrightarrow M_\infty$ fast sicher für $t \to \infty$ und

$$M_t = \mathbb{E}(M_\infty | \mathcal{F}_\infty).$$

(d) Es gibt $Y \in \mathcal{L}^1$ mit $M_t = \mathbb{E}(Y|\mathcal{F}_\infty)$ für alle $t \geq 0$.

In diesem Fall gilt $M_\infty = \mathbb{E}(Y|\mathcal{F}_\infty)$.

Zum Schluß wollen wir noch die kontinuierliche Version der Doob-Meyer Zerlegung beweisen. Hierfür brauchen wir einige Vorbereitung.

Definition 7.45. *Die σ-Algebra*

$$\mathcal{P} = \sigma\left((H_t)_{t\geq 0} \mid (H_t)_{t\geq 0} \text{ ist linksstetig und adaptiert }\right)$$

heißt die σ-Algebra der vorhersehbaren Mengen. Ein stochastischer Prozess $(X_t)_{t\geq 0}$ heißt vorhersehbar (predictable), falls

$$\Omega \times [0,\infty) \ni (\omega, t) \longmapsto X_t(\omega)$$

messbar ist bezüglich $\mathcal{P}$.

Lemma 7.46. *Es gilt*

$$\mathcal{P} = \sigma\left(\{A \times \{0\}|\ A \in \mathcal{F}_0\} \cup \{A \times (s,t] \mid s < t,\ A \in \mathcal{F}_s\}\right).$$

Insbesondere ist jeder vorhersehbare beschränkte stochastische Prozess $(X_t)_{t\geq 0}$ Grenzwert von Elementarfunktionen der Form

$$X_t(\omega) = a_0 + \mathbb{1}_{A_0}(\omega) + \mathbb{1}_{\{0\}}(t) + \sum_{n=1}^{N} a_n \mathbb{1}_{A_n}(\omega)\mathbb{1}_{(s_n,t_n]}(t) \tag{7.10}$$

mit $a_n \in \mathbb{R}$, $N \in \mathbb{N}$, $0 < s_n < t_n$, $A_n \in \mathcal{F}_{s_n}$.

Beweis. Es sei $\mathcal{P}' = \sigma\left(\{A \times \{0\}|\ A \in \mathcal{F}_0\} \cup \{A \times (s,t] \mid s < t,\ A \in \mathcal{F}_s\}\right)$. Dann ist jede bezüglich $\mathcal{P}'$ messbare Zufallsvariable Grenzwert von Zufallsvariablen der Form (7.10). Ferner sind diese Elementarfunktionen linksstetig und adaptiert, woraus $\mathcal{P}' \subset \mathcal{P}$ folgt. Umgekehrt, sei X linksstetig und adaptiert. Dann ist

$$X_t^{(N)} := X_0 \mathbb{1}_{\{0\}}(t) + \sum_{k=0}^{N-1} X_{\frac{k}{N}} \mathbb{1}_{\left(\frac{k}{N}, \frac{k+1}{N}\right]}(t)$$

messbar bezüglich $\mathcal{P}'$ und es gilt $X_t^{(N)} \longrightarrow X_t$ fast sicher für alle t und $N \to \infty$. Also ist auch X messbar bezüglich $\mathcal{P}'$. Dieses zeigt $\mathcal{P} \subset \mathcal{P}'$. ☐

Definition 7.47. *Sei $(X_t)_{t\geq 0}$ ein stochastischer Prozess.*

(a) $(X_t)_{t\geq 0}$ ist von der Klasse (D), falls die Familie

$$\{X_T \mid T \;\; \textit{Stoppzeit mit } \mathbb{P}(T<\infty)=1\}$$

gleichgradig integrierbar ist.

(b) $(X_t)_{t\geq 0}$ ist von der Klasse (DL), falls für alle $m \in \mathbb{N}$ die Familie

$$\{X_T \mid T \leq m \;\; \textit{Stoppzeit}\ \}$$

gleichgradig integrierbar ist.

Lemma 7.48. *Sei $(X_t)_{t\geq 0}$ ein cadlag Prozess. Es gelten die folgenden Aussagen.*

(a) Ist $(X_t)_{t\geq 0}$ ein nicht-negatives Submartingal, so ist $(X_t)_{t\geq 0}$ von der Klasse (DL).

(b) Ist $(X_t)_{t\geq 0}$ von der Klasse (D), so ist $(X_t)_{t\geq 0}$ von der Klasse (DL).

(c) Ist $(X_t)_{t\geq 0}$ ein gleichgradig integrierbares Submartingal, so ist $(X_t)_{t\geq 0}$ von der Klasse (D).

(d) Ist $X_t \in \mathcal{L}^p$ mit $\sup\limits_{t\in[0,T]} \mathbb{E}(X_t^p) < \infty$ und $p>1$, so ist $(X_t)_{t\in[0,T]}$ gleichgradig integrierbar.

Beweis. (a) Sei T eine beschränkte Stoppzeit mit $T \leq m \in \mathbb{N}$. Aus Optional Sampling folgt $X_T \leq \mathbb{E}(X_m|\mathcal{F}_T)$. Ferner ist $\{X_T > n\} \in \mathcal{F}_T$ und somit

$$\mathbb{E}(\mathbb{1}_{\{X_T>n\}}X_T) \leq \mathbb{E}(\mathbb{1}_{X_T>n}\mathbb{E}(X_m|\mathcal{F}_T)) = \mathbb{E}(\mathbb{1}_{X_T>n}X_m).$$

Aus Tschebyscheff und Optional Sampling folgt

$$\mathbb{P}(X_T>n) \leq \frac{1}{n}\mathbb{E}(X_T) \leq \frac{1}{n}\mathbb{E}(X_m) \longrightarrow 0, \quad n\to\infty.$$

Daraus lässt sich leicht die Behauptung folgern.
(b) Jede Stoppzeit T mit $T\leq m$ erfüllt insbesondere $\mathbb{P}(T<\infty)=1$.
(c) Übung.
(d) Es gilt mit der Hölder Ungleichung mit $\frac{1}{p}+\frac{1}{q}=1$

$$\mathbb{E}(|X_t|) \leq \mathbb{E}(|X_t|^p) \leq \sup_{t\in[0,T]} \mathbb{E}(X_t^p) < \infty$$

und damit $\sup\limits_{t\in[0,T]} \mathbb{E}(|X_t|) < \infty$. Ferner gilt

$$\mathbb{E}(\mathbb{1}_A|X_t|) \leq \left(\mathbb{E}(\mathbb{1}_A^q)\right)^{\frac{1}{q}} \left(\mathbb{E}(|X_t|^p)\right)^{\frac{1}{p}} \leq \left(\sup_{t\in[0,T]} \mathbb{E}(|X_t|^p)\right)^{\frac{1}{p}} \left(\mathbb{P}(A)\right)^{\frac{1}{q}}$$

welches die Behauptung zeigt. □

Wir werden später weitere hinreichende Bedingungen sehen.

Satz 7.49. *(Doob-Meyer Zerlegung, [KS91]) Sei $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ ein filtrierter Raum mit den üblichen Bedingungen und $(X_t)_{t\geq 0}$ ein cadlag Submartingal. Dann sind aquivalent:*

(a) $(X_t)_{t\geq 0}$ *ist von der Klasse (DL).*

(b) Es gibt ein rechtsstetiges Martingal $(M_t)_{t\geq 0}$ und einen rechtsstetigen vorhersehbaren Prozess $(A_t)_{t\geq 0}$ mit

$$A_0 = 0, \quad A_s \leq A_t, \quad s \leq t$$

derart, dass

$$X_t = X_0 + M_t + A_t, \quad t \geq 0.$$

Ist (M'_t, A'_t) eine weitere Zerlegung mit obigen Eigenschaften, so sind M_t, M'_t sowie A_t, A'_t ununterscheidbar. Ist zusätzlich $(X_t)_{t\geq 0}$ von der Klasse (D), so ist $(M_t)_{t\geq 0}$ gleichgradig integrierbar und $A_\infty := \lim_{t\to\infty} A_t$ ist integrierbar.

Diesen Satz zu beweisen erfordert einiges an Vorbereitung und soll hier nicht gemacht werden. Wir werden jedoch einen wichtigen Spezialfall beweisen.

Beispiel 7.50. *Dann ist $(M_t^2)_{t\geq 0}$ ein nicht-negatives Submartingal und somit von der Klasse (DL). Die Doob-Meyer Zerlegung liefert die Existenz und Eindeutigkeit eines vorhersehbaren, steigenden Prozesses $(A_t)_{t\geq 0}$ derart, dass $M_t^2 - M_0^2 - A_t$ ein Martingal ist.*

Definition 7.51. *Sei $(M_t)_{t\geq 0} \subset \mathcal{L}^2$ ein cadlag Martingal mit $M_0 = 0$. Dann wird der vorhersehbare, steigende Prozess $(A_t)_{t\geq 0}$ mit $\langle M \rangle_t$ bezeichnet. Dieser wird Kompensator bezeichnet.*

Beispiel 7.52.

- *Sei $(B_t)_{t\geq 0}$ die Brownsche Bewegung. Dann ist $\langle B \rangle_t = t$.*
- *Sei $(N_t)_{t\geq 0}$ der Poisson Prozess mit Intensität $\lambda > 0$. Dann ist $\langle N \rangle_t = \lambda t$.*

Im Folgenden wollen wir die Existenz und Eindeutigkeit des Kompensators $\langle M \rangle_t$ ohne Rückgriff auf die Doob-Meyer Zerlegung beweisen. Hierfür betrachten wir den Fall von Martingalen mit stetigen Pfaden. Der allgemeine Fall von cadlag Pfaden ist wesentlich schwieriger zu beweisen.

Satz 7.53. *Sei $p \in [1, \infty)$, $(M_t^{(n)}) \subset \mathcal{L}^p$ eine Familie von Martingalen mit stetigen Pfaden und $(M_t)_{t\geq 0} \subset \mathcal{L}^p$ ein stochastischer Prozess. Es gelte $M_t^{(n)} \longrightarrow M_t$ in $\mathcal{L}^p$ für $n \to \infty$ für jedes $t \geq 0$. Dann gelten die folgenden Aussagen.*

(a) $(M_t)_{t\geq 0}$ *hat eine Modifikation welche ein Martingal ist.*

(b) Sei $p > 1$. Es gibt eine Modifikation von $(M_t)_{t\geq 0}$, welche wir wieder mit $(M_t)_{t\geq 0}$ bezeichnen, derart dass M_t adaptiert ist, stetige Pfade hat und

$$\mathbb{E}\left(\sup_{t\in[0,T]} |M_t^{(n)} - M_t|^p\right) \longrightarrow 0, \quad \forall T > 0.$$

Beweis. (a) Sei $0 \le s < t$. Dann gilt $M_s^{(n)} = \mathbb{E}(M_t^{(n)}|\mathcal{F}_s)$ für alle $n \ge 1$ und

$$\left(\mathbb{E}(M_t^{(n)}|\mathcal{F}_s) - \mathbb{E}(M_t|\mathcal{F}_s)\right)^p = \left(\mathbb{E}(M_t^{(n)} - M_t|\mathcal{F}_s\right)^p \le \mathbb{E}((M_t^{(n)} - M_t)^p|\mathcal{F}_s).$$

Folglich gilt

$$\mathbb{E}\left(\left(\mathbb{E}(M_t^{(n)}|\mathcal{F}_s) - \mathbb{E}(M_t|\mathcal{F}_s)\right)^p\right) \le \mathbb{E}\left((M_t^{(n)} - M_t)^p\right)$$

und damit $\mathbb{E}(M_t^{(n)}|\mathcal{F}_s) \longrightarrow \mathbb{E}(M_t|\mathcal{F}_s)$ für $n \to \infty$ in $\mathcal{L}^p$. Daraus folgt durch Grenzübergang $n \to \infty$

$$M_s = \mathbb{E}(M_t|\mathcal{F}_s).$$

Also ist $(M_t)_{t \ge 0}$ ein Martingal.
(b) Sei $T > 0$ fest. Da $M^{(m)} - M^{(n)}$ ein Martingal ist, folgt aus der Doob Ungleichung

$$\begin{aligned}\mathbb{E}\left(\left(\sup_{t \in [0,T]} |M_t^{(m)} - M_t^{(n)}|\right)^p\right) &= \mathbb{E}\left(\sup_{t \in [0,T]} |M_t^{(m)} - M_t^{(n)}|^p\right) \\ &\le \left(\frac{p}{p-1}\right)^p \mathbb{E}(|M_T^{(m)} - M_T^{(n)}|^p)\end{aligned}$$

und somit

$$\|\sup_{t \in [0,T]} |M_t^{(m)} - M_t^{(n)}|\|_{\mathcal{L}^p} \le \frac{p}{p-1} \|M_T^{(m)} - M_T^{(n)}\|_{\mathcal{L}^p} \longrightarrow 0$$

für $n, m \to \infty$. Wir können daher eine Teilfolge $(n_k)_{k \in \mathbb{N}}$ finden derart, dass

$$\|M_T^{(n_{k+1})} - M_T^{(n_k)}\|_{\mathcal{L}^p} \le 2^{-k}, \quad k \ge 1.$$

Folglich gilt

$$\begin{aligned}\left\|\sum_{k=1}^{\infty} \sup_{t \in [0,T]} |M_t^{(n_{k+1})} - M_t^{(n_k)}|\right\|_{\mathcal{L}^p} &\le \sum_{k=1}^{\infty} \|\sup_{t \in [0,T]} |M_t^{(n_{k+1})} - M_t^{(n_k)}|\|_{\mathcal{L}^p} \\ &\le \frac{p}{p-1} \sum_{k=1}^{\infty} \|M_T^{(n_{k+1})} - M_T^{(n_k)}\|_{\mathcal{L}^p} \le \frac{p}{p-1} \sum_{k=1}^{\infty} \frac{1}{2^k} < \infty\end{aligned}$$

und somit

$$\mathbb{P}\left(\sum_{k=1}^{\infty} \sup_{t \in [0,T]} |M_t^{(n_{k+1})} - M_t^{(n_k)}| < \infty\right) = 1.$$

Also konvergiert $\sum_{k=1}^{\infty} |M_t^{(n_{k+1})} - M_t^{(n_k)}|$ fast sicher gleichmässig in $t \in [0, T]$. Insbesondere konvergiert $Y_t := \sum_{k=1}^{\infty} (M_t^{(n_{k+1})} - M_t^{(n_k)})$ fast sicher gleichmässig in $t \in [0, T]$. Wegen

$$M_t^{(n_N)} = M_t^{(n_1)} + \sum_{k=1}^{N-1} (M_t^{(n_{k+1})} - M_t^{(n_k)})$$

konvergiert $M_t^{(n_N)} \longrightarrow M_t^{(n_1)} + Y_t$ fast sicher gleichmässig in $t \in [0,T]$. Nach Übergang zu einer geeigneten Teilfolge sehen wir $M_t = M_t^{(n_1)} + Y_t$ fast sicher für alle $t \in [0,T]$. Die rechte Seite hat stetige Pfade und ist, da der Wahrscheinlichkeitsraum vollständig ist, adaptiert. Die letzte Behauptung folgt aus Doobs Ungleichung wegen

$$\mathbb{E}\left(\sup_{t\in[0,T]} |M_t^{(n)} - M_t|^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|M_T^{(n)} - M_T|^p).$$

□

Lemma 7.54. *Sei* $(M_t)_{t\geq 0} \subset \mathcal{L}^1$ *ein adaptierter stochastischer Prozess mit cadlag Pfaden. Genau dann ist* $(M_t)_{t\geq 0}$ *ein Martingal, wenn*

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) \tag{7.11}$$

für alle beschränkten Stoppzeiten gilt.

Beweis. Ist $(M_t)_{t\geq 0}$ ein Martingal, so gilt (7.11) nach Optional Sampling. Umgekehrt gelte (7.11). Für konstante Stoppzeiten $T = t \geq 0$ liefert dieses

$$\mathbb{E}(M_t) = \mathbb{E}(M_0), \quad t \geq 0.$$

Sei $s \leq t$ und $A \in \mathcal{F}_s$. Wir müssen

$$\mathbb{E}(M_t \mathbb{1}_A) = \mathbb{E}(M_s \mathbb{1}_A)$$

zeigen. Wähle $T = s\mathbb{1}_A + t\mathbb{1}_{A^c}$. Dann ist T eine beschränkte Stoppzeit und es gilt

$$\mathbb{E}(M_T) = \mathbb{E}(\mathbb{1}_A M_s + \mathbb{1}_{A^c} M_t) = \mathbb{E}(\mathbb{1}_A M_s) + \mathbb{E}(M_t) - \mathbb{E}(\mathbb{1}_A M_t).$$

Ferner gilt

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) = \mathbb{E}(M_t)$$

welches die Behauptung zeigt. □

Sei $(M_t)_{t\geq 0} \subset \mathcal{L}^2$ ein Martingal mit $M_0 = 0$ derart, dass $M_t(\omega)$ für jedes $\omega \in \Omega$ stetig in $t \geq 0$ ist. Sei $(\tau_n)_{n\in\mathbb{N}}$ eine Folge von Zerlegungungen

$$0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{N_n}^{(n)}$$

mit $t_{N_n}^{(n)} \longrightarrow \infty$, $\sup\limits_{k=0,\dots,n} (t_{k+1}^{(n)} - t_k^{(n)}) < \infty$ und $\tau_n \subset \tau_{n+1}$. Definiere einen Prozess

$$V_t^{(n)} := \sum_{s\in\tau_n} (M_{s'\wedge t} - M_{s\wedge t})^2,$$

wo s' den Nachfolger von s bezeichnet.

Satz 7.55. *Es gibt einen stochastischen Prozess* $\langle M\rangle_t$ *derart, dass die folgenden Eigenschaften gelten:*

(a) $\langle M\rangle_t \in \mathcal{L}^1$ *ist adaptiert, hat stetige Pfade und ist monoton steigend. (fast sicher)*

(b) Es gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{t\in[0,T]} |V_t^{(n)} - \langle M\rangle_t| \geq \varepsilon\right) = 0, \quad \forall \varepsilon > 0,\ T > 0.$$

(c) $M_t^2 - \langle M\rangle_t$ *ist ein Martingal.*

(d) Ist zusätzlich $(M_t)_{t\geq 0}$ *beschränkt, d.h.* $c := \sup_{(\omega,t)\in\Omega\times\mathbb{R}_+} |M_t(\omega)| < \infty$*, so gilt*

$$\mathbb{E}\left(\sup_{t\in[0,T]} |V_t^{(n)} - \langle M\rangle_t|^2\right) \longrightarrow 0, \quad \forall T > 0.$$

Beweis. Wir betrachten zuerst den Fall, dass M_t beschränkt ist. Wir zeigen zunächst, dass $(V_t^{(n)})_{n\in\mathbb{N}}$ für jedes $t \geq 0$ eine Cauchy Folge in $\mathcal{L}^2$.
Sei $N \in \mathbb{N}$ so gross dass für $m > n > N$: $t \leq t_{N_n}^{(n)} < t_{N_m}^{(m)}$. Wegen $(a-b)^2 = a^2 - 2(a-b)b - b^2$ folgt

$$\begin{aligned}
V_t^{(n)} &= \sum_{u\in\tau_n} \left(M_{u'\wedge t}^2 - M_{u\wedge t}^2\right) - 2\sum_{u\in\tau_n} M_{u\wedge t}(M_{u'\wedge t} - M_{u\wedge t}) \\
&= M_t^2 - 2\sum_{\substack{u\leq\tau_n \\ u\leq t}} M_u(M_{u'\wedge t} - M_{u\wedge t}) \\
&= M_t^2 - 2\sum_{\substack{u\leq\tau_n \\ u\leq t}} \sum_{\substack{u\leq s<u' \\ s\in\tau_m}} M_u(M_{s'\wedge t} - M_{s\wedge t})
\end{aligned}$$

wo die letzte Gleichung aus

$$M_{u'\wedge t} - M_{u\wedge t} = \sum_{\substack{u\leq s<u' \\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})$$

folgt. Analog erhalten wir

$$\begin{aligned}
V_t^{(m)} &= M_t^2 - 2\sum_{\substack{s\leq\tau_m \\ s\leq t}} M_s(M_{s'\wedge t} - M_{s\wedge t}) \\
&= M_t^2 - 2\sum_{\substack{u\leq t \\ u\in\tau_n}} \sum_{\substack{u\leq s<u' \\ s\in\tau_m}} M_s(M_{s'\wedge t} - M_{s\wedge t})
\end{aligned}$$

Daraus folgt

$$\begin{aligned}
V_t^{(m)} - V_t^{(n)} &= -2\sum_{u\in\tau_n} \sum_{\substack{u\leq s<u' \\ s\in\tau_m}} \left(M_s(M_{s'\wedge t} - M_s) - M_u(M_{s'\wedge t} - M_{s\wedge t})\right). \\
&= -2\sum_{\substack{u\leq t \\ u\in\tau_n}} \sum_{\substack{u\leq s<u' \\ s\in\tau_m}} (M_s - M_u)(M_{s'\wedge t} - M_{s\wedge t}).
\end{aligned}$$

Wir erhalten somit

$$\mathbb{E}\left((V_t^{(m)} - V_t^{(n)})^2\right) = 4\mathbb{E}\left(\sum_{\substack{u\le t\\ u\in\tau_n}} \sum_{\substack{u\le s<u'\\ s\in\tau_m}} (M_s - M_u)^2 (M_{s'\wedge t} - M_{s\wedge t})^2\right),$$

wo alle gemischten Terme wegen der Martingaleigenschaft verschwinden (nachrechnen!). Sei

$$\Delta_N := \sup_{n'\ge N} \sup_{m'>n'} \sup_{\omega\in\Omega} \left\{(M_s - M_u)^2 \mid u \in \tau_{n'},\ u \le t,\ s \in \tau_{m'},\ u \le s < u'\right\}.$$

Da $(M_t)_{t\ge 0}$ stetige Pfade hat, gilt $\Delta_N \longrightarrow 0$ fast sicher für $N \to \infty$. Ferner gilt $\Delta_N \le 4c^2$. Damit erhalten wir

$$\begin{aligned}
\mathbb{E}\left((V_t^{(m)} - V_t^{(n)})^2\right) &\le \mathbb{E}\left(\Delta_N \sum_{\substack{u\le t\\ u\in\tau_n}} \sum_{\substack{u\le s<u'\\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^2\right)\\
&\le \left(\mathbb{E}(\Delta_N^2)\right)^{\frac12} \left(\mathbb{E}\left(\sum_{\substack{u\le t\\ u\in\tau_n}} \sum_{\substack{u\le s<u'\\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^2\right)^2\right)^{\frac12}\\
&= \left(\mathbb{E}(\Delta_N^2)\right)^{\frac12} \left(\mathbb{E}\left(\sum_{\substack{s\le t\\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^2\right)^2\right)^{\frac12}
\end{aligned}$$

Aus dem Satz der dominierten Konvergenz folgt $\mathbb{E}(\Delta_N^2) \longrightarrow 0$. Ferner erhalten wir

$$\begin{aligned}
\mathbb{E}\left(\sum_{\substack{s\le t\\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^2\right)^2 &= \mathbb{E}\left(\sum_{\substack{s\le t\\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^4\right)\\
&\quad + 2\mathbb{E}\left(\sum_{\substack{s\le t\\ s\in\tau_m}} \sum_{\substack{u<s\\ u\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^2 (M_{u'\wedge t} - M_{u\wedge t})^2\right)\\
&\le 4c^2\mathbb{E}\left(\sum_{\substack{s\le t\\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^2\right)\\
&\quad + 2\sum_{\substack{s\le t\\ s\in\tau_m}} \sum_{\substack{s<u\\ u\in\tau_m}} \mathbb{E}\left((M_{s'\wedge t} - M_{s\wedge t})^2 (M_{u'\wedge t} - M_{u\wedge t})^2\right)\\
&= S_1 + S_2.
\end{aligned}$$

Es gilt

$$\begin{aligned}
S_1 &= 4c^2\mathbb{E}\left(\sum_{\substack{s\le t\\ s\in\tau_m}} (M_{s'\wedge t} - M_{s\wedge t})^2\right)\\
&= 4c^2 \sum_{\substack{s\le t\\ s\in\tau_m}} \mathbb{E}\left(M^2_{s'\wedge t} - 2M_{s'\wedge t}M_{s\wedge t} + M^2_{s\wedge t}\right)\\
&= 4c^2 \sum_{\substack{s\le t\\ s\in\tau_m}} \mathbb{E}\left(M^2_{s'\wedge t} - 2M_{s\wedge t}E(M_{s'\wedge t}|\mathcal{F}_s) + M^2_{s\wedge t}\right)\\
&= 4c^2 \sum_{\substack{s\le t\\ s\in\tau_m}} \mathbb{E}\left(M^2_{s'\wedge t} - M^2_{s\wedge t}\right) = 4c^2\mathbb{E}(M_t^2 - M_0^2) = 4c^2\mathbb{E}(M_t^2) \le 4c^4.
\end{aligned}$$

und analog

$$\begin{aligned}
\sum_{\substack{s<u\\ u\in\tau_m}} \mathbb{E}\left((M_{s'\wedge t} - M_{s\wedge t})^2(M_{u'\wedge t} - M_{u\wedge t})^2\right) &= \sum_{\substack{s<u\\ u\in\tau_m}} \mathbb{E}\left(\mathbb{E}((M_{u'\wedge t} - M_{u\wedge t})^2|\mathcal{F}_u)(M_{s'\wedge t} - M_{s\wedge t})^2\right)\\
&= \sum_{\substack{s<u\\ u\in\tau_m}} \mathbb{E}\left(\mathbb{E}(M^2_{u'\wedge t} - M^2_{u\wedge t}|\mathcal{F}_u)(M_{s'\wedge t} - M_{s\wedge t})^2\right)\\
&= \sum_{\substack{s<u\\ u\in\tau_m}} \mathbb{E}\left((M^2_{u'\wedge t} - M^2_{u\wedge t})(M_{s'\wedge t} - M_{s\wedge t})^2\right)\\
&= \mathbb{E}\left((M_t^2 - M^2_{s'\wedge t})(M_{s'\wedge t} - M_{s\wedge t})^2\right)\\
&\le 2c^2\mathbb{E}\left((M_{s'\wedge t} - M_{s\wedge t})^2\right) = 2c^2\mathbb{E}\left(M^2_{s'\wedge t} - M^2_{s\wedge t}\right).
\end{aligned}$$

Daraus folgt

$$S_2 \le 4c^2 \sum_{\substack{s\le t\\ s\in\tau_m}} \mathbb{E}\left(M^2_{s'\wedge t} - M^2_{s\wedge t}\right) = 4c^2\mathbb{E}(M_t^2) \le 4c^4.$$

Insgesamt folgt

$$\mathbb{E}\left((V_t^{(m)} - V_t^{(n)})^2\right) \le \left(\mathbb{E}(\Delta_N^2)\right)^{\frac{1}{2}} \sqrt{8}c^2 \longrightarrow 0.$$

Sei $V_t := \lim\limits_{n\to\infty} V_t^{(n)}$ der $\mathcal{L}^2$-Grenzwert. Setze

$$Y_t^{(n)} := 2\sum_{s\in\tau_n} M_{s\wedge t}(M_{s'\wedge t} - M_{s\wedge t}).$$

Dann ist $(Y_t^{(n)})_{t\ge 0}$ ein Martingal und es gilt

$$Y_t^{(n)} = M_t^2 - V_t^{(n)} \longrightarrow M_t^2 - V_t, \quad n\to\infty$$

in $\mathcal{L}^2$. Also gibt es eine Modifikation $(Y_t)_{t\geq 0}$ von $M_t^2 - V_t$ derart, dass Y ein Martingal mit stetigen Pfaden ist. Folglich hat V_t eine Modifikation mit stetigen Pfaden gegeben durch

$$\langle M\rangle_t = M_t^2 - Y_t. \tag{7.12}$$

Insbesondere ist $\langle M\rangle$ auch adaptiert und es gilt

$$\mathbb{E}\left(\sup_{t\in[0,T]} |V_t^{(n)} - \langle M\rangle_t|^2\right) \longrightarrow 0, \quad n\to\infty. \tag{7.13}$$

Betrachten wir den allgemeinen Fall. Sei

$$R_k(\omega) := \inf\left\{t>0 \mid |M_t(\omega)| > k\right\}.$$

Dann ist R_k eine Stoppzeit. Ferner gilt $|M_{t\wedge R_k}| \leq k$. Nach Optional Sampling ist

$$M_{t\wedge R_k} := M_t^k$$

ein beschränktes Martingal. Aus obigen Überlegungen gibt es einen stetigen, adaptierten Prozess $\langle M^k\rangle_t = V_t^k$. Es sei $(V_t^k)^{(n)}$ die Approximation von V_t^k über die Zerlegung τ_n.

Lemma 7.56. *(Diagonalfolgentrick) Es gibt es eine Teilfolge $(n_l)_{l\in\mathbb{N}}$ und $\Omega_0 \in \mathcal{F}_0$ mit $\mathbb{P}(\Omega_0) = 1$ derart, dass*

$$\sup_{s\in[0,t]} |(V_s^k)^{(n_l)} - \langle M^k\rangle_s| \longrightarrow 0, \quad l\to\infty, \quad \omega\in\Omega_0.$$

Beweis. Wegen (7.13) gibt es für $k=1$ eine Teilfolge $(n_1(l))_{l\in\mathbb{N}}$ und $\Omega_1 \in \mathcal{F}_0$ mit $\mathbb{P}(\Omega_1) = 1$ derart, dass

$$\sup_{s\in[0,t]} |(V_s^1)^{(n_1(l))} - \langle M^1\rangle_s| \longrightarrow 0, \quad l\to\infty, \quad \omega\in\Omega_1.$$

Wähle eine weitere Teilfolge $(n_2(l))_{l\in\mathbb{N}}$ von $(n_1(l))_{l\in\mathbb{N}}$ und $\Omega_2 \in \mathcal{F}_0$ mit $\mathbb{P}(\Omega_2) = 1$ derart, dass

$$\sup_{s\in[0,t]} |(V_s^2)^{(n_2(l))} - \langle M^2\rangle_s| \longrightarrow 0, \quad l\to\infty, \quad \omega\in\Omega_2.$$

Eine solche Wahl ist wegen der Konvergenz in (7.13) möglich. Iteration liefert eine Folge von Teilfolgen $(n_k(l))_{l\in\mathbb{N}}$, wo $(n_{k+1}(l))_{l\in\mathbb{N}}$ eine Teilfolge von $(n_k(l))_{l\in\mathbb{N}}$ ist, Mengen $\Omega_k \in \mathcal{F}_0$ mit $\mathbb{P}(\Omega_k) = 1$ derart, dass

$$\sup_{s\in[0,t]} |(V_s^k)^{(n_k(l))} - \langle M^k\rangle_s| \longrightarrow 0, \quad l\to\infty, \quad \omega\in\Omega_k.$$

Dann liefert $\Omega_0 := \bigcap\limits_{k\geq 1} \Omega_k$ sowie $n(l) := n_l(l)$ das gewünschte. □

Es gilt $R_k \leq R_{k+1}$ fast sicher. Für $R := \lim\limits_{k\to\infty} R_k \in [0,\infty]$ gilt

$$\mathbb{P}(R\geq m) = \mathbb{P}\left(\exists k: \ |M_m| \leq k\right) = 1, \quad \forall m\in\mathbb{N}.$$

Also ist $R = \infty$ fast sicher und somit $R_k \longrightarrow \infty$. Damit ist $\mathbb{P}(\Omega_0 \cap \{R = \infty\}) = 1$ Damit ist

$$\langle M \rangle_t(\omega) := \begin{cases} \langle M^k \rangle_t(\omega), & \text{für } k \in \mathbb{N} \text{ mit } \omega \in \{t \le R_k\} \cap \Omega_0 \\ 0, & \text{sonst} \end{cases},$$

auf $\Omega_0 \cap \{R = \infty\}$ definiert und auf dem Komplement 0. Ferner gilt

$$\mathbb{1}_{t \le R_k} V_t^{(n_l)} = \mathbb{1}_{t \le R_k} V_{t \wedge R_k}^{(n_l)} = \mathbb{1}_{t \le R_k} (V_t^k)^{(n_l)} \longrightarrow \mathbb{1}_{t \le R_k} \langle M^k \rangle_t \tag{7.14}$$

für $l \to \infty$ auf Ω_0. Dieses zeigt, dass $\langle M \rangle_t$ unabhängig von der Wahl von k ist.

Seien $\delta, \varepsilon > 0$ beliebig. Wähle $k(\delta) \in \mathbb{N}$ derart, dass

$$\mathbb{P}(R_{k(\delta)} < t) < \delta.$$

Dann folgt

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0 \le s \le t} |V_s^{(n)} - \langle M \rangle_s| \ge \varepsilon \right) \\
&= \mathbb{P}\left(\sup_{0 \le s \le t} |V_s^{(n)} - \langle M \rangle_s| \ge \varepsilon,\ R_k < t \right) + \mathbb{P}\left(\sup_{0 \le s \le t} |V_s^{(n)} - \langle M \rangle_s| \ge \varepsilon,\ R_k \ge t \right) \\
&\le \mathbb{P}\left(R_k < t\right) + \mathbb{P}\left(\sup_{0 \le s \le t} |(V_s^k)^{(n)} - \langle M^k \rangle_s| \ge \varepsilon, \quad R_k \ge t \right) \\
&\le \mathbb{P}\left(R_k < t\right) + \mathbb{P}\left(\sup_{0 \le s \le t} |(V_s^k)^{(n)} - \langle M^k \rangle_s| \ge \varepsilon \right)
\end{aligned}$$

Für den zweiten Summanden gilt nach Tschebyscheff

$$\mathbb{P}\left(\sup_{0 \le s \le t} |(V_s^k)^{(n)} - \langle M^k \rangle_s| \ge \varepsilon \right) \le \frac{1}{\varepsilon^2} \mathbb{E}\left(\sup_{0 \le s \le t} |(V_s^k)^{(n)} - \langle M^k \rangle_s|^2 \right) \longrightarrow 0, \quad n \to \infty.$$

Daraus folgt

$$\limsup_{n \to \infty} \mathbb{P}\left(\sup_{0 \le s \le t} |V_s^{(n)} - \langle M \rangle_s| \ge \varepsilon \right) \le \mathbb{P}\left(R_k < t\right) < \delta. \tag{7.15}$$

Dieses zeigt (b). Für Teil (a) bleibt es zu zeigen, dass $\langle M \rangle_t$ monoton steigend ist. Sei $t \ge 0$ fest, dann gilt

$$V_s^{(n)} = \sum_{u \in \tau_n} (M_{u' \wedge t} - M_{u \wedge t})^2 = \sum_{\substack{u \in \tau_n \\ u' \le t}} (M_{u'} - M_u)^2 + (M_s - M_{\delta_s(n)})^2$$

wo $\delta_s(n) = \sup\{u \in \tau_n \mid u \le s\}$. Daraus folgt

$$\begin{aligned}
&\mathbb{P}\left(\sup_{s \in [0,t]} \left| \sum_{\substack{u \in \tau_n \\ u' \le t}} (M_{u'} - M_u)^2 - \langle M \rangle_s \right| \ge \varepsilon \right) \\
&\le \mathbb{P}\left(\sup_{s \in [0,t]} (M_s - M_{\delta_s(n)})^2 + \sup_{s \in [0,t]} \left| V_s^{(n)} - \langle M \rangle_s \right| \ge \varepsilon \right) \\
&\le \mathbb{P}(\sup_{s \in [0,t]} (M_s - M_{\delta_s(n)})^2 \ge \varepsilon) + \mathbb{P}\left(\sup_{s \in [0,t]} \left| V_s^{(n)} - \langle M \rangle_s \right| \ge \varepsilon \right).
\end{aligned}$$

Der zweite Term konvergiert wegen (7.15) gegen 0, der erste wegen der Stetigkeit der Pfade. Da $\sum\limits_{\substack{s \in \tau_n \\ s' \leq t}} (M_{s'} - M_s)^2$ monoton wachsend in t ist, ist es auch $\langle M \rangle_t$.

Es bleibt (c) zu zeigen. Wir zeigen dass $M_T^2 - \langle M \rangle_T \in \mathcal{L}^1$ und

$$\mathbb{E}(M_T^2) = \mathbb{E}(\langle M \rangle_T)$$

für alle beschränkten Stoppzeiten gilt. Die Behauptung folgt dann aus Lemma 7.54. Wegen (7.12) ist

$$M_t^k - \langle M^k \rangle_t = M_{t \wedge R_k} - \langle M^k \rangle_t = M_{t \wedge R_k} - \langle M \rangle_{t \wedge R_k}$$

ein Martingal für alle $k \in \mathbb{N}$. Insbesondere gilt für alle beschränkten Stoppzeiten T

$$\mathbb{E}(M_{T \wedge R_k}^2) = \mathbb{E}(\langle M \rangle_{T \wedge R_k}).$$

Aus dem Satz der monotonen Konvergenz und der Submartingal Eigenschaft zusammen mit Optional Sampling folgt

$$\mathbb{E}(\langle M \rangle_T) = \lim_{k \to \infty} \mathbb{E}(\langle M \rangle_{T \wedge R_k}) = \lim_{k \to \infty} \mathbb{E}(M_{T \wedge R_k}^2) \leq \mathbb{E}(M_T^2).$$

Mit dem Lemma von Fatou gilt

$$\mathbb{E}(\langle M \rangle_T) = \lim_{k \to \infty} \mathbb{E}(M_{T \wedge R_k}^2) \geq \mathbb{E}(M_T^2).$$

Wir erhalten also

$$\mathbb{E}(\langle M \rangle_T) = \mathbb{E}(M_T^2).$$

Ist $m \in \mathbb{N}$ mit $T \leq m$ so folgt aus Optional Sampling

$$\mathbb{E}(M_T^2) \leq \mathbb{E}(M_m^2) < \infty$$

welches die Behauptung zeigt. $\square$

8 Appendix

Lemma 8.1. *Für jedes $N \in \mathbb{N}_0$ gibt es eine stetige Funktion $\Theta_{N+1}(x) \in \mathbb{C}$ mit $|\Theta_{N+1}(x)| \leq 1$ derart, dass*

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + \frac{(ix)^{N+1}}{(N+1)!} \Theta_{N+1}(x), \quad x \in \mathbb{R}.$$

Beweis. Wir zeigen über Induktion, dass für alle $N \in \mathbb{N}_0$ mit $t_0 = 1$ gilt

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + (ix)^{N+1} \int_0^{t_0} \cdots \int_0^{t_N} e^{ixt_{N+1}} dt_{N+1} \cdots dt_1$$

Dann hat

$$\Theta_{N+1}(x) := (N+1)! \int\limits_0^{t_0} \cdots \int\limits_0^{t_N} e^{ixt_{N+1}} dt_{N+1} \cdots dt_1$$

die gewünschten Eigenschaften. Für $N = 0$ gilt offensichtlich

$$e^{ix} = 1 + ix \int\limits_0^{t_0} e^{ixt} dt.$$

$N \longmapsto N+1$: Dann gilt nach Induktionsvoraussetzung

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + (ix)^{N+1} \int\limits_0^{t_0} \cdots \int\limits_0^{t_N} e^{ixt_{N+1}} dt_{N+1} \cdots dt_1.$$

Für das innere Integral gilt nach $N = 0$

$$\int\limits_0^{t_N} e^{ixt_{N+1}} dt_{N+1} = \int\limits_0^{t_N} 1 dt_N + ix \int\limits_0^{t_N} \int\limits_0^{t_{N+1}} e^{ixt_{N+2}} dt_{N+2} dt_{N+1}.$$

Daraus folgt

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + (ix)^{N+1} \int\limits_0^{t_0} \cdots \int\limits_0^{t_N} 1 dt_{N+1} \cdots dt_1 + (ix)^{N+2} \int\limits_0^{t_0} \cdots \int\limits_0^{t_{N+1}} e^{ixt_{N+2}} dt_{N+2} \cdots dt_1.$$

Die Behauptung folgt, da

$$\int\limits_0^{t_0} \cdots \int\limits_0^{t_N} 1 dt_{N+1} \cdots dt_1 = \frac{1}{(N+1)!}.$$

$\square$

Literatur

[Bil99] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[Chu74] Kai Lai Chung. *Elementary probability theory with stochastic processes.* Springer-Verlag, New York-Heidelberg, 1974. Undergraduate Texts in Mathematics.

[Dud02] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[Ete81] N. Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete*, 55(1):119–122, 1981.

[FL80] Wolfgang Fischer and Ingo Lieb. *Funktionentheorie*, volume 47 of *Vieweg Studium: Aufbaukurs Mathematik [Vieweg Studies: Mathematics Course].* Friedr. Vieweg & Sohn, Braunschweig, 1980. Aufbaukurs Mathematik.

[Jac01] N. Jacob. *Pseudo differential operators and Markov processes. Vol. I.* Imperial College Press, London, 2001. Fourier analysis and semigroups.

[KK06] Yu. G. Kondratiev and O. V. Kutoviy. On the metrical properties of the configuration space. *Math. Nachr.*, 279(7):774–783, 2006.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1991.

[KS07] Leonid B. Koralov and Yakov G. Sinai. *Theory of probability and random processes.* Universitext. Springer, Berlin, second edition, 2007.

[Wer00] Dirk Werner. *Funktionalanalysis.* Springer-Verlag, Berlin, extended edition, 2000.