Skript zur Vorlesung

WAHRSCHEINLICHKEITS-THEORIE II

Jürgen Gärtner

unter Mitwirkung von
S. Ribbecke, N. Paetsch
und **S. Sturm**

Version Juli 2006

Inhaltsverzeichnis

Kapitel 1

Maßtheoretische Grundlagen

Bevor wir richtig in die Wahrscheinlichkeitstheorie II einsteigen, müssen wir dazu zuerst die dafür erforderlichen Grundlagen (das heißt die maßtheoretischen Grundlagen) betrachten. In diesem Kapitel werden hauptsächlich wichtige Definitionen und Sätze aus „Maß- und Integrationstheorie“ behandelt, welche zum Teil auch in der Analysis vorkommen. Zur weiterführenden Lektüre empfiehlt es sich, auch die Veranstaltung „Maß- und Integrationstheorie“ zu besuchen.

1.1. σ-Algebren und Maße

Definition 1.1.

a) Eine *σ-Algebra* ist ein System $\mathfrak{A}$ von Teilmengen von Ω mit

   (i) $\Omega \in \mathfrak{A}$;

   (ii) $A \in \mathfrak{A} \Rightarrow A^c := \Omega \setminus A \in \mathfrak{A}$;

   (iii) (A_n) Folge von Mengen aus $\mathfrak{A} \Rightarrow \bigcup\limits_n A_n \in \mathfrak{A}$.

b) Ein *Maß* auf einer σ-Algebra $\mathfrak{A}$ ist eine Abbildung $\mu : \mathfrak{A} \to [0, \infty]$ mit folgenden Eigenschaften:

   (i) $\mu(\emptyset) = 0$;

   (ii) *σ-Additivität*:
   (A_n) Folge *paarweiser disjunkter* Mengen aus $\mathfrak{A}$

$$\Rightarrow \mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

Bemerkung 1.2.

1. Die Mengen (Elemente) aus $\mathfrak{A}$ heißen $\mathfrak{A}$-*messbar* (kurz: *messbar*).

2. σ-Algebren sind „abgeschlossen“ bezüglich aller endlichen und abzählbar unendlichen Mengenoperationen (Vereinigung, Durchschnitt, Differenz, symmetrische Differenz).

3. Der Durchschnitt *beliebig vieler* σ-Algebren in Ω, ist wieder eine σ-Algebra in Ω.

4. Triviale σ-Algebren:

   $\mathfrak{A} = \{\emptyset, \Omega\}$

   $\mathfrak{P}(\Omega)$, das System aller Teilmengen von Ω (*Potenzmenge* von Ω)

5. $(\Omega, \mathfrak{A})$ und $(\Omega, \mathfrak{A}, \mu)$ heißen *messbarer Raum* bzw. *Maßraum* . μ heißt *endlich* (unendlich), falls $\mu(\Omega) < \infty$ $(= \infty)$. μ heißt σ-*endlich*, falls eine Folge (E_n) $\mathfrak{A}$-messbarer Mengen existiert mit $E_n \uparrow \Omega$ und $\mu(E_n) < \infty$ für alle n.

6. Ein Maß μ mit $\mu(\Omega) = 1$ heißt *Wahrscheinlichkeitsmaß*. $(\Omega, \mathfrak{A}, \mu)$ heißt dann *Wahrscheinlichkeitsraum*. Wahrscheinlichkeitsmaße werden üblicherweise mit P bzw. $\mathbb{P}$ bezeichnet.

Satz 1.3. *(Stetigkeit von Maßen bezüglich monotoner Mengenfolgen) Gegeben seien ein Maßraum $(\Omega, \mathfrak{A}, \mu)$ und eine Folge (A_n) $\mathfrak{A}$-messbarer Mengen. Dann gelten folgende Aussagen:*

(i) $A_n \uparrow A \Longrightarrow \mu(A_n) \uparrow \mu(A)$;

(ii) $A_n \downarrow A$ *und* $\mu(A_{n_0}) < \infty$ *für ein* $n_0 \Longrightarrow \mu(A_n) \downarrow \mu(A)$.

Definition 1.4. Sei $\mathfrak{E}$ ein System von Teilmengen von Ω. Dann wird mit

$$\sigma(\mathfrak{E}) := \bigcap_{\substack{\mathfrak{A} \supseteq \mathfrak{E} \\ \mathfrak{A} - \sigma - Algebra}} \mathfrak{A}$$

die kleinste $\mathfrak{E}$ umfassende σ-Algebra bezeichnet. Sie heißt die von $\mathfrak{E}$ *erzeugte σ-Algebra*; $\mathfrak{E}$ wird *Erzeuger* genannt.

Beispiel 1.5.

1. Sei $(\mathfrak{A}_j)_{j \in J}$ eine Familie von σ-Algebren in Ω. Ihre Vereinigung ist im Allgemeinen keine σ-Algebra.

   Bezeichung:

$$\bigvee_{j \in J} \mathfrak{A}_j := \sigma(\bigcup_{j \in J} \mathfrak{A}_j)$$

2. Sei $\mathfrak{E} := \{\{\omega\} : \ \omega \in \Omega\}$ das System aller einpunktigen Mengen

$$\Longrightarrow \sigma(\mathfrak{E}) \quad = \quad \{A \subseteq \Omega : \ A \text{ oder } A^c \text{ höchstens abzählbar } \}$$

   das heißt

$$\sigma(\mathfrak{E}) \qquad \begin{cases} = \mathfrak{P}(\Omega), & \text{falls } \Omega \text{ abzählbar} \\ \neq \mathfrak{P}(\Omega), & \text{falls } \Omega \text{ überabzählbar} \end{cases}$$

3. Seien E ein topologischer (oder metrischer) Raum
   $\mathfrak{O}_E$ das System aller offenen Teilmengen von E,
   $\mathfrak{C}_E$ das System aller abgeschlossenen Teilmengen von E.
   $\mathfrak{B}_E = \mathfrak{B}(E) := \sigma(\mathfrak{O}_E)$ heißt dann die *Borel-σ-Algebra* in E.
   Die Elemente von $\mathfrak{B}_E$ heißen *Borelmengen*. Es gilt

$$\sigma(\mathfrak{O}_E) = \sigma(\mathfrak{C}_E).$$

4. Spezialfall: Seien E = $\mathbb{R}^d$, $\mathfrak{B}^d := \mathfrak{B}(\mathbb{R}^d)$.
   Sei außerdem $\gamma := \{\bigtimes_{i=1}^{d} (a_i, b_i] : a_1, b_1, \ldots, a_d, b_d \in \mathbb{R}\}$ das System aller achsenparallelen Quader.
   Es gilt

$$\mathfrak{B}(\mathbb{R}^d) = \sigma(\gamma).$$

   $d = 1$: $\mathfrak{J}_0 := \{(-\infty, a] : a \in \mathbb{R}\}$ und $\mathfrak{J}_1 := \{(a, b] : a \leq b\}$.
   Dann gilt

$$\mathfrak{B}(\mathbb{R}) = \sigma(\mathfrak{J}_0) = \sigma(\mathfrak{J}_1).$$

Definition 1.6. Ein System ϑ von Teilmengen von Ω heißt *Dynkin-System*, falls gilt

a) $\Omega \in \vartheta$;

b) $A \in \vartheta \Rightarrow A^c \in \vartheta$;

c) Für eine Folge (A_n) *paarweise disjunkter* Mengen aus ϑ gilt $\bigcup_n A_n \in \vartheta$.

Beispiel 1.7.

1. Jede σ-Algebra ist ein Dynkin-System.

2. Der Durchschnitt beliebig vieler Dynkin-Systeme (in Ω) ist wieder ein Dynkin-System.

3. Seien μ und ν Maße auf $(\Omega, \mathfrak{A})$ und $\mu(\Omega) = \nu(\Omega) < \infty$, dann ist
   $\mathfrak{M} := \{A \in \mathfrak{A} : \mu(A) = \nu(A)\}$ ein Dynkin-System.

Beweis. $\Omega \in \mathfrak{M}$ ist klar. Sei nun $A \in \mathfrak{M}$, dann gilt

$$\mu(A^c) = \mu(\Omega) - \mu(A) = \nu(\Omega) - \nu(A) = \nu(A^c),$$

also ist auch $A^c \in \mathfrak{M}$. Für eine Folge (A_n) *paarweise disjunkter* Mengen aus $\mathfrak{M}$ gilt dann mit der σ-Additivität $(*)$

$$\mu(\bigcup_n A_n) \underset{(*)}{=} \sum_n \mu(A_n) \underset{A_n \in \mathfrak{M}}{=} \sum_n \nu(A_n) \underset{(*)}{=} \nu(\bigcup_n A_n),$$

also gilt $\bigcup_n A_n \in \mathfrak{M}$. □

Definition 1.8. Sei $\mathfrak{E}$ ein beliebiges System von Teilmengen von Ω. Dann wird mit

$$d(\mathfrak{E}) := \bigcap_{\substack{\vartheta \supseteq \mathfrak{E} \\ \vartheta \text{ Dynkin-System}}} \vartheta$$

das kleinste $\mathfrak{E}$ umfassende Dynkin-System bezeichnet. Es wird das *von $\mathfrak{E}$ erzeugte Dynkin-System* genannt. $\mathfrak{E}$ ist sein Erzeuger. $\mathfrak{E}$ heißt $\cap$-*stabil*, falls gilt:

$$A, B \in \mathfrak{E} \Longrightarrow A \cap B \in \mathfrak{E}.$$

Satz 1.9. *Sei ϑ ein Dynkin-System. Dann gilt:*

$$\vartheta \ \sigma - Algebra \Longleftrightarrow \vartheta \ \cap -stabil$$

Satz 1.10. *(Haupsatz über Dynkin-Systeme)*
Sei $\mathfrak{E}$ ein $\cap$-stabiles Mengensystem. Dann gilt:

$$\sigma(\mathfrak{E}) = d(\mathfrak{E}).$$

Beispiel 1.11. Maße sind durch ihre Werte auf $\cap$-stabilen Mengensystemen eindeutig festgelegt. Genauer:
Sei $(\Omega, \mathfrak{A})$ ein messbarer Raum, sowie μ und ϑ auf $\mathfrak{A}$ mit $\mu(A) = \vartheta(A) < \infty$ gegeben. Sei außerdem $\mathfrak{E}$ ein $\cap$-stabiler Erzeuger der σ-Algebra $\mathfrak{A}$.
Dann gilt:

$$\mu(A) = \vartheta(A) \quad \text{für alle } A \in \mathfrak{E} \Longrightarrow \mu = \vartheta.$$

Beweis. $\mathfrak{M} := \{A \in \mathfrak{A} : \mu(A) = \vartheta(A)\}$ ist ein Dynkin-System. Dann gilt mit dem Hauptsatz über Dynkin-Systeme $(*)$

$$\mathfrak{M} \supseteq \mathfrak{E} \Rightarrow \mathfrak{M} \supseteq d(\mathfrak{E}) \underset{(*)}{=} \sigma(\mathfrak{E}) = \mathfrak{A},$$

das heißt $\mathfrak{M} = \mathfrak{A}$, das bedeutet wiederum, dass $\sigma = \vartheta$. □

Im zweiten Teil dieses Abschnitts wollen wir uns nun mit der *Konstruktion von Maßen* befassen.

Definition 1.12.

a) Ein System γ von Teilmengen von Ω heißt *Semiring*, falls

(i) $\emptyset \in \gamma$

(ii) $A, B \in \gamma \Longrightarrow A \cap B \in \gamma$

(iii) $A, B \in \gamma$ und $A \supseteq B \Longrightarrow$ es existieren paarweise disjunkte $C_1, \dots, C_m \in \gamma$ mit $A \backslash B = \bigcup\limits_{i=1}^{m} C_i$.

b) Ein System $\mathfrak{A}$ von Teilmengen von Ω heißt *Algebra*, falls

(i) $\Omega \in \mathfrak{A}$;

(ii) $A \in \mathfrak{A} \Longrightarrow A^c \in \mathfrak{A}$;

(iii) $A_1, \dots, A_m \in \mathfrak{A} \Longrightarrow \bigcup\limits_{i=1}^{m} A_i \in \mathfrak{A}$.

Beispiel 1.13.

1. Jede σ-Algebra ist eine Algebra und jede Algebra ist ein Semiring.

2. Das System γ der achsenparallelen Quader aus Beispiel 1.5 bildet einen Semiring

Sei $\mathfrak{E}$ ein beliebiges Mengensystem und $\mu : \mathfrak{E} \to [0, \infty]$ eine beliebige nichtnegative Mengenfunktion. Dann heißt μ *σ-endlich*, falls eine Folge (E_n) von Mengen aus $\mathfrak{E}$ exisitiert mit $E_n \uparrow \Omega$ und $\mu(E_n) < \infty$ für alle n.
μ heißt *σ-additiv*, falls für eine beliebige Folge (A_n) paarweise disjunkter Mengen aus $\mathfrak{E}$ mit $\bigcup\limits_n A_n \in \mathfrak{E}$ gilt:

$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n).$$

Satz 1.14. *(Fortsetzungssatz von Carathéodory)*
Sei $\tilde{\mu}$ eine σ-endliche und σ-additive Mengenfunktion auf einem Semiring γ mit $\tilde{\mu}(\emptyset) = 0$. Dann besitzt $\tilde{\mu}$ genau eine Fortsetzung zu einem Maß μ auf der σ-Algebra $\sigma(\gamma)$.

Beweis. Existenzbeweis nichttrivial; zur Einzigkeit vergleiche Beispiel 1.11. □

Anwendung 1.14a). (Erste Beispiele)

1. Das *Lebesgue-Maß* λ auf $(\mathbb{R}^d, \mathfrak{B}^d)$ ist das einzige Maß auf $\mathfrak{B}^d$, welches achsenparallelen Quadern ihre Elementarvolumen zuordnet.
(Nichttrivial: σ-Additivität auf γ)

2. *Verteilungen* auf $(\mathbb{R}, \mathfrak{B})$.
Eine Funktion $F : \mathbb{R} \to [0, 1]$ heißt *Verteilungsfunktion*, falls

(i) F nichtfallend;
(ii) F rechtsstetig;
(iii) $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to+\infty} F(x) = 1$.

Jedem Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}, \mathfrak{B})$ ordnen wir die Funktion

$$F_\mu(x) := \mu((-\infty, x]), \qquad x \in \mathbb{R},$$

zu.

Satz 1.15. *Die Zuordnung $\mu \longmapsto F_\mu$ ist eine Bijektion zwischen der Menge der Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathfrak{B})$ und der Menge der Verteilungsfunktionen.*

Beweis. Übungsaufgabe; schwierig hierbei ist die Surjektivität zu zeigen, was mit Hilfe des Fortsetzungssatzes von Carathéodory geschieht. □

1.2. Meßbare Abbildungen

Definition 1.16.

a) Seien $(\Omega, \mathfrak{A})$ und $(\Omega', \mathfrak{A}')$ messbare Räume. Eine Abbildung $f : \Omega \longmapsto \Omega'$ heißt $\mathfrak{A}$-$\mathfrak{A}'$-*messbar* (kurz: *messbar*), falls

$$f^{-1}(A') \in \mathfrak{A} \quad \text{für alle } A' \in \mathfrak{A}'.$$

b) Sei wieder $(\Omega, \mathfrak{A})$ ein messbarer Raum. Eine Abbildung $f : \Omega \longmapsto \mathbb{R}$ heißt $\mathfrak{A}$-*messbare Funktion*, falls sie $\mathfrak{A}$-$\mathfrak{B}$-messbar ist.

c) Seien $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ die erweiterte Zahlengerade und die zugehörige Borel-σ-Algebra gegeben durch $\overline{\mathfrak{B}} := \sigma(\mathfrak{B} \cup \{-\infty\} \cup \{\infty\})$. Eine Abbildung $f : \Omega \longrightarrow \overline{\mathbb{R}}$ heißt *messbare numerische Funktion*, falls sie $\mathfrak{A}$-$\overline{\mathfrak{B}}$-messbar ist.

Für jedes $A \in \mathfrak{A}$ ist die Indikatorfunktion $\mathbb{1}_A$ messbar.

Satz 1.17. *(Messbarkeitskriterium)*
Gegeben seien messbare Räume $(\Omega, \mathfrak{A})$ und $(\Omega', \mathfrak{A}')$, ein Erzeuger $\mathfrak{E}'$ von $\mathfrak{A}'$ (das heißt $\sigma(\mathfrak{E}') = \mathfrak{A}'$) sowie eine Abbildung $f : \Omega \mapsto \Omega'$. Dann gilt

$$f \textit{ ist } \mathfrak{A} - \mathfrak{A}' - \textit{messbar} \Longleftrightarrow f^{-1}(A') \in \mathfrak{A} \quad \textit{für alle } A' \in \mathfrak{E}.$$

Beweisidee.

„$\Longrightarrow$“: trivial.
„$\Longleftarrow$“:
$\mathfrak{F}' := \{A' \in \mathfrak{A}' : f^{-1}(A') \in \mathfrak{A}$ ist eine σ-Algebra, was noch selbst nachzuprüfen ist. Dann gilt:

$$\mathfrak{E}' \subseteq \mathfrak{F}' \Rightarrow \mathfrak{A}' = \sigma(\mathfrak{E}') \subseteq \mathfrak{F}',$$

das heißt $\mathfrak{F}' = \mathfrak{A}'$. Damit ist f messbar. □

Bemerkung 1.18. Die Mengen $(-\infty, x)$, $x \in \mathbb{R}$, erzeugen die σ-Algebra $\mathfrak{B}$ (und die σ-Algebra $\overline{\mathfrak{B}}$). Deshalb ist eine Abbildung $f : \Omega \to \mathbb{R}$ $(f : \Omega \to \overline{\mathbb{R}})$ genau dann eine messbare Funktion (messbare numerische Funktion), wenn

$$\{f \leq x\} := \{\omega \in \Omega : f(\omega) \leq x\} \in \mathfrak{A} \qquad \text{für alle } x \in \mathbb{R}.$$

Satz 1.19. *(Messbarkeit und Grenzübergänge) Gegeben seien ein messbarer Raum $(\Omega, \mathfrak{A})$ und eine Folge (f_n) messbarer numerischer Funktionen auf Ω. Dann sind*

$$\inf_n f_n,\ \sup_n f_n,\ \liminf_{n\to\infty} f_n,\ \limsup_{n\to\infty} f_n$$

ebenfalls messbare numerische Funktionen. Existiert $f(\omega) := \lim_{n\to\infty} f_n(\omega)$ für alle $\omega \in \Omega$, so ist f ebenfalls eine messbare numerische Funktion.

Beweis. Z.B.:

- $\{\inf_n f_n < x\} = \bigcup_n \underbrace{\{f_n < x\}}_{\in \mathfrak{A}} \in \mathfrak{A} \qquad \text{für alle } x \in \mathbb{R}.$

- $\lim_{n\to\infty} \inf = \lim_{n\to\infty} \inf_{n\geq m} f_n$, das heißt $\inf_{n\geq m} f_n \uparrow \liminf_{n\to\infty} f_n$ für $m \uparrow \infty$
  $\Longrightarrow \{\liminf f_n \leq x\} = \bigcap_m \{\inf_{n\geq m} f_n \leq x\} \qquad \text{für alle } x \in \mathbb{R}.$

□

Satz 1.20. *(Messbarkeit und algebraische Operationen) Gegeben seien ein messbarer Raum $(\Omega, \mathfrak{A})$ sowie messbare Funktionen f und g auf Ω. Dann sind auch αf $(\alpha \in \mathbb{R})$, $f \pm g$, $f \cdot g$, f/g (falls $g(\omega) \neq 0$ für alle ω) messbar.*

Satz 1.21. *(Komposition messbarer Abbildungen) Gegeben seien messbare Räume $(\Omega_i, \mathfrak{A}_i)$ (mit i = 1,2,3). Sind $f : \Omega_1 \to \Omega_2$ und $g : \Omega_2 \to \Omega_3$ $\mathfrak{A}_1$-$\mathfrak{A}_2$-messbar, bzw. $\mathfrak{A}_2$-$\mathfrak{A}_3$-messbar, so ist $g \circ f : \Omega_1 \to \Omega_3$ eine $\mathfrak{A}_1$-$\mathfrak{A}_3$-messbare Abbildung.*

Satz 1.22. *(Messbarkeit und Stetigkeit) Gegeben seien topologische Räume $(E, \mathfrak{O})$ und $(E', \mathfrak{O}')$ sowie eine Abbildung $f : E \to E'$. Dann folgt aus f stetig, dass f $\mathfrak{B}(E)$-$\mathfrak{B}(E')$-messbar ist.*

Beweis. Anwendung des Messbarkeitskriteriums (Satz 1.17) für den Erzeuger $\mathfrak{O}'$ von $\mathfrak{B}(E')$:

$$\forall\, G' \in \mathfrak{O}' : \ \ f^{-1}(G') \in \mathfrak{O} \subseteq \mathfrak{B}(E), \ \text{da } f \text{ stetig ist.}$$

□

Ist f eine messbare (numerische) Funktion, so sind auch $f^+ = \max(f, 0)$, $f^- = \max(-f, 0)$ und $|f| = f^+ + f^-$ messbare (numerische) Funktionen.

Bezeichnungen. Sei Ω eine Menge, $(\Omega', \mathfrak{A}')$ ein messbarer Raum und f eine Abbildung mit $f : \Omega \to \Omega'$. Dann heißt $\sigma(f) := f^{-1}(\mathfrak{A}') := \{f^{-1}(A) : A \in \mathfrak{A}'\}$

die *von f erzeugte σ-Algebra.* Für eine Familie von Abbildungen $(f_i)_{i\in I}$, $f_i : \Omega \to \Omega'$ heißt $\sigma(f_i; i \in I) := \bigvee_{i\in I} f_i^{-1}(\mathfrak{A}')$ die *von den Abbildungen f_i, $i \in I$, erzeugte σ-Algebra.*

Anschaulich ist $\sigma(f_i; i \in I)$ die kleinste σ-Algebra in Ω, bezüglich der alle Abbildungen $f_i, i \in I$, messbar sind.

1.3. Lebesgue-Integral

Als Vorraussetzung für diesen Abschnitt sei $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum.

Eine Funktion $f : \Omega \to \mathbb{R}$ heiße *elementar*, wenn sie $\mathfrak{A}$-messbar ist und nur *endlich* viele Werte annimmt. Eine solche Funktion besitzt eine Darstellung

$$f = \sum_{i=1}^{n} \alpha_i \, \mathbb{1}_{A_i} \tag{1.1}$$

mit $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ und $A_1, \dots, A_n \in \mathfrak{A}$ paarweise disjunkt (*disjunkte Darstellung*).

Integral nichtnegativer elementarer Funktionen:

$$\int f \, d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i) \quad \in [0, \infty].$$

Vereinbarung: $0 \cdot \infty = 0$.

Satz 1.23. *Zu jeder nichtnegativen messbaren numerischen Funktion $f : \Omega \to \overline{\mathbb{R}}$ existiert eine Folge (f_n) nichtnegativer elementarer Funktionen mit $f_n \uparrow f$.*

Beweis. Man nehme

$$f_n := \sum_{k=1}^{n2^n} (k-1)2^{-n} \mathbb{1}_{\{(k-1)2^{-n} \le f < k2^{-n}\}} + n\mathbb{1}_{\{f \ge n\}}.$$

□

Integral nichtnegativer messbarer numerischer Funktionen:

Sei $f : \Omega \to \overline{\mathbb{R}}$ nichtnegativ und $\mathfrak{A}$-$\overline{\mathfrak{B}}$-messbar. Dann existiert eine Folge (f_n) elementarer Funktionen mit $0 \le f_n \uparrow f$. Man definiert

$$\int f \, d\mu := \lim_{n\to\infty} \int f_n \, d\mu \in [0, \infty].$$

Man beachte hier, dass der Limes existiert, da $0 \le \int f_n \, d\mu \le \int f_{n+1} \, d\mu$. Bezüglich der Korrektheit der Definition ist die Unabhängigkeit des Grenzwertes von der konkreten Approximationsfolge (f_n) noch nachzuweisen.

Integral messbarer numerischer Funktionen:

Sei $f : \Omega \to \overline{\mathbb{R}}$ eine messbare numerische Funktion, $f = f^+ - f^-$, $\int f^+ \, d\mu$, $\int f^- \, d\mu$ haben wir bereits definiert.
Ist eines der beiden Integrale endlich, so *existiert*

$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu \in [-\infty, \infty].$$

Die Funktion f heißt *μ-integrierbar*, falls beide Integrale endlich sind, das heißt falls

$$\int |f| \, d\mu = \int f^+ \, d\mu + \int f^- \, d\mu < \infty.$$

Für beliebige $A \in \mathfrak{A}$ setzt man

$$\int_A f \, d\mu := \int \mathbb{1}_A f \, d\mu.$$

Elementare Eigenschaften des Lebesgue-Integrals:

Seien $A \in \mathfrak{A}$, $\alpha \in \mathbb{R}$, f und g μ-integrierbare Funktionen. Dann gilt

(i) $\int \mathbb{1}_A \, d\mu = \mu(A)$;

(ii) αf μ-integrierbar und $\int \alpha f \, d\mu = \alpha \int f \, d\mu$ (Homogenität);

(iii) $f + g$ μ-intgerierbar und $\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu$ (Addidivität);

(iv) $f \leq g \Rightarrow \int f \, d\mu \leq \int g \, d\mu$ (Monotonie);

(v) $|\int f \, d\mu| \leq \int |f| \, d\mu$ (Dreiecksungleichung).

Beweis. Übungsaufgabe. □

Definition 1.24. Eine Menge $N \in \mathfrak{A}$ heißt *μ-Nullmenge*, falls $\mu(N) = 0$. Man sagt, eine Eigenschaft für die Elemente $\omega \in \Omega$ gelte *μ-fast überall* (*μ-f.ü.*), falls die Ausnahmemenge eine Teilmenge einer μ-Nullmenge ist.

Beispiel 1.25.

1. Meßbare Teilmengen und abzählbare Vereinigungen von μ-Nullmengen sind wieder μ-Nullmengen

2. Seien f, g und f_n messbare numerische Funktionen. Dann gelten folgende Aussagen:

$$\begin{aligned} f = g \ \mu\text{-fast überall} \quad &\Leftrightarrow \quad \mu(f \neq g) = 0 \\ f_n \uparrow f \ \mu\text{-fast überall} \quad &\Leftrightarrow \quad \{\omega \in \Omega \colon f_1(\omega) \leq f_2(\omega) \leq \ldots \text{ und } f(\omega) = \lim_n f_n(\omega)\} \\ &\qquad \text{ist eine } \mu\text{-Nullmenge} \\ &\Leftrightarrow \quad f_n \leq f_{n+1} \ \mu\text{-f.ü.} \, \forall n \text{ und } f = \lim_n f_n \ \mu\text{-f.ü.} \end{aligned}$$

Satz 1.26. *Seien f und g messbare numerische Funktionen auf $(\Omega, \mathfrak{A})$. Dann gelten die folgenden Aussagen:*

(i) Existiert $\int f\,d\mu$ und ist $f = g$ μ-fast überall, so existiert auch $\int g\,d\mu$ und

$$\int f\,d\mu = \int g\,d\mu.$$

(ii) Ist $f \geq 0$ und $\int f\,d\mu = 0$, so ist $f = 0$ μ-fast überall.

(iii) Ist $f \geq 0$ und $\int f\,d\mu < \infty$, so existiert eine messbare Funktion

$$g : \Omega \to [0, \infty) \text{ mit } \begin{cases} f = g \ \mu\text{-fast überall und} \\ \int f\,d\mu = \int g\,d\mu. \end{cases}$$

Bemerkung 1.27. Ist $f : [a, b] \to \mathbb{R}$ Borel-messbar und Riemann-integrierbar, so stimmt $\int_{[a,b]} f\,d\lambda$ (λ Lebesgue-Maß auf $[a, b]$) mit dem klassischen Riemann-Integral $\int_a^b f(x)dx$ überein.

Beweis. Übungsaufgabe. Zum Beweis benutzt man die Definition des Riemann-Integrals über Ober- und Untersummen sowie die Konvergenzsätze aus Abschnitt 1.5. □

1.4. Bildmaße

Für diesen Abschnitt seien die messbaren Räume $(\Omega, \mathfrak{A})$ und $(E, \mathfrak{E})$ sowie eine Abbildung $f : \Omega \to E$ gegeben.

Satz 1.28. *Für jedes Maß μ auf $(\Omega, \mathfrak{A})$ wird durch*

$$\nu(B) := \mu(f^{-1}(B)) = \mu(f \in B), \quad B \in \mathfrak{E},$$

ein Maß ν auf $(E, \mathfrak{E})$ definiert.

Definition 1.29. Das Maß ν heißt *Bild* des Maßes μ bezüglich f (kurz: *Bildmaß*) und wird mit $\mu \circ f^{-1}$ bezeichnet.

Satz 1.30. *(Transformationssatz für Bildmaße) Sei $g : E \to \overline{\mathbb{R}}$ eine messbare numerische Funktion. Dann gilt:*
Das Integral $\int g\,d\mu \circ f^{-1}$ existiert genau dann, wenn das Integral $\int g \circ f\,d\mu$ existiert. Ist dies erfüllt, so gilt

$$\int g\,d\mu \circ f^{-1} = \int g \circ f\,d\mu.$$

Beweisidee. Sei $g = \sum_{i=1}^{n} \beta_i \mathbb{1}_{B_i}$ nichtnegativ und elementar. Dann ist

$$
\begin{aligned}
g \circ f &= \sum_{i=1}^{n} \beta_i \mathbb{1}_{f^{-1}(B_i)} \\
\text{und} \quad \int g \circ f \, d\mu &= \sum_{i=1}^{n} \beta_i \mu(f^{-1}(B_i)) \\
&= \sum_{i=1}^{n} \beta_i (\mu \circ f^{-1})(B_i) \\
&= \int g \, d\mu \circ f^{-1}.
\end{aligned}
$$

□

1.5. Konvergenzsätze

Vertauschung von Integration und Grenzübergang

Für diesen Abschnitt sei $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum und f_n und f seien messbare numerische Funktionen auf $(\Omega, \mathfrak{A})$.

Satz 1.31. *(Satz über monotone Konvergenz) Sei f_n eine Folge nichtnegativer Funktionen mit $f_n \uparrow f$ μ-fast überall. Dann gilt*

$$0 \leq \int f_n \, d\mu \uparrow \int f \, d\mu.$$

Folgerung 1.32. *(Satz von Beppo Levi) Sei $f_n \geq 0$ für alle n. Dann folgt daraus*

$$\int \sum_n f_n \, d\mu = \sum_n \int f_n \, d\mu.$$

Man beachte: $0 \leq \sum_{k=1}^{n} f_k \uparrow \sum_{k=1}^{\infty} f_k$.

Satz 1.33. *(Lemma von Fatou) Sei f_n eine Folge nichtnegativer Funktionen. Dann gilt*

$$\int \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu.$$

Satz 1.34. *(Satz von Lebesgue über majorisierte Kovergenz) Es gelten die folgenden Vorraussetzungen:*

(i) $\lim_{n\to\infty} f_n = f$ *μ-fast überall;*

(ii) es existiere eine μ-integrierbare Funktion $g : \Omega \to \overline{\mathbb{R}}$ mit $|f| \leq g$ μ-f.ü. für alle n (also eine integrierbare Majorante).

Dann sind $f_n (n \in \mathbb{N})$ und f μ-integrierbar und

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Achtung: Ist μ endlich, so können konstante Funktionen g als integrierbare Majoranten dienen (Satz von Lebesgue über beschränkte Konvergenz).

1.6. Produktmaß und Mehrfachintegrale

Seien $(\Omega_1, \mathfrak{A}_1, \mu_1)$ und $(\Omega_2, \mathfrak{A}_2, \mu_2)$ *σ-endliche* Maßräume. Wir definieren das Produkt dieser Maßräume

$$(\Omega, \mathfrak{A}, \mu) := (\Omega_1, \mathfrak{A}_1, \mu_1) \otimes (\Omega_2, \mathfrak{A}_2, \mu_2)$$

wie folgt:

$$\begin{aligned}
\Omega &:= \Omega_1 \times \Omega_2; \\
\mathfrak{A} &:= \mathfrak{A}_1 \otimes \mathfrak{A}_2 \ (\textit{Produkt-}\sigma\textit{-Algebra}) \\
&\ \text{die vom Semiring } \gamma := \{A_1 \times A_2 : A_1 \in \mathfrak{A}_1, A_2 \in \mathfrak{A}_2\} \\
&\ \text{erzeugte } \sigma\text{-Algebra;} \\
\mu &:= \mu_1 \otimes \mu_2 \ (\textit{Produktmaß}) \\
&\ \textit{dasjenige Maß auf } \mathfrak{A}_1 \otimes \mathfrak{A}_2, \text{ für das} \\
&\ \mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \quad \forall A_1 \in \mathfrak{A}_1, A_2 \in \mathfrak{A}_2 \text{ gilt.}
\end{aligned}$$

Satz 1.35. *Sind μ_1 und μ_2 σ-endlich, so existiert genau ein Produktmaß $\mu_1 \otimes \mu_2$.*

Satz 1.36. *(Satz von Fubini) Seien die Maße μ_1 und μ_2 wieder σ-endlich und sei $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ eine $(\mathfrak{A}_1 \otimes \mathfrak{A}_2)$-messbare numerische Funktion.*

(i) Ist $f \geq 0$, so gilt

$$\begin{aligned}
&\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) \, \mu(d(\omega_1, \omega_2)) \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \, \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \, \mu_1(d\omega_1) \right) \mu_2(d\omega_2). \qquad (1.2)
\end{aligned}$$

(ii) Sei f $\mu_1 \otimes \mu_2$-integrierbar, dann gilt Formel (1.2).

Bemerkung 1.37.

1. Der Satz von Fubini (Satz 1.36) schließt implizit ein, dass die betrachteten Ausdrücke wohldefiniert sind. Insbesondere folgt aus der Nichtnegativität und aus der $\mathfrak{A}_1 \otimes \mathfrak{A}_2$-Messbarkeit von f, dass für jedes $\omega_1 \in \Omega_1$ die numerische Funktion $\omega_2 \longmapsto f(\omega_1, \omega_2)$ $\mathfrak{A}_2$-messbar ist und dass $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(d\omega_2)$ $\mathfrak{A}_1$-messbar ist.
   Im Teil b) kann es passieren, dass $\int f(\omega_1, \omega_2)\, \mu_2(d\omega_2)$ auf einer μ_1-Nullmenge von Punkten $\omega_1 \in \Omega_1$ nicht existiert. Dort ersetzt man dann das Integral durch 0.

2. Die $\mu_1 \otimes \mu_2$-Integrierbarkeit von f in b) kann man oft durch Anwendung von a) auf $|f|$ nachprüfen.

3. Für $f = \mathbb{1}_A$, $A \in \mathfrak{A}_1 \otimes \mathfrak{A}_2$, liefert die Aussage a) eine explizite Darstellung des Produktmaßes.

4. Das Produkt von Maßräumen und der Satz von Fubini lassen sich durch vollständige Induktion leicht auf endlich viele σ-endliche Maßräume ausdehnen. Etwas Mühe bereitet dabei der Nachweis des Assoziativgesetzes

$$(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3).$$

1.7. Absolutstetigkeit von Maßen

Sei $(\Omega, \mathfrak{A})$ ein messbarer Raum. Sind μ ein Maß auf $(\Omega, \mathfrak{A})$ und $f : \Omega \to \overline{\mathbb{R}}$ eine nichtnegative $\mathfrak{A}$-messbare numerische Funktion, so wird durch

$$\nu(A) := \int\limits_A f\, d\mu, \quad A \in \mathfrak{A} \tag{1.3}$$

ein Maß ν auf $(\Omega, \mathfrak{A})$ definiert.

Beweisskizze.
$\nu(\emptyset) = 0$ ist offensichtlich.
Zur σ-Additivität: Sei (A_n) eine Folge paarweise disjunkter $\mathfrak{A}$-messbarer Mengen, dann ist $\mathbb{1}_{\bigcup\limits_n A_n} = \sum\limits_n \mathbb{1}_{A_n}$. Und dann gilt mit dem Satz von Beppo Levi $(*)$

$$\nu(\bigcup A_n) = \int \sum_n (\mathbb{1}_{A_n} f)\, d\mu \underset{(*)}{=} \sum_n \int \mathbb{1}_{A_n} f\, d\mu = \sum_n \nu(A_n).$$

□

Definition 1.38. Seien μ,ν Maße auf $(\Omega, \mathfrak{A})$. Dann heißt das Maß ν *absolutstetig* bezüglich des Maßes μ (in Zeichen: $\nu \ll \mu$), falls für alle $A \in \mathfrak{A}$ gilt:

$$\mu(A) = 0 \Rightarrow \nu(A) = 0$$

(jede μ-Nullmenge ist eine ν-Nullmenge).

Das durch 1.3 definierte Maß ν ist absolutstetig bezüglich μ, denn aus

$$\mu(A) = 0 \text{ folgt } \mathbb{1}_A f = 0 \ \mu\text{-f.ü. und damit ist } \nu(A) = 0.$$

Satz 1.39. *(Satz von Radon-Nikodym)*
Seien μ und ν σ-endliche Maße auf $(\Omega, \mathfrak{A})$. Dann ist ν bezüglich μ genau dann absolutstetig, wenn eine nichtnegative $\mathfrak{A}$-messbare Funktion $f : \Omega \to \mathbb{R}$ existiert, mit

$$\nu(A) = \int\limits_A f \, d\mu, \quad A \in \mathfrak{A}.$$

Die Funktion f ist μ-fast überall eindeutig bestimmt.

Definition 1.40. Die μ-fast überall eindeutige Funktion f im Satz von Radon-Nikodym heißt *Radon-Nikodym-Ableitung* (*Dichte*) von ν bezüglich μ und wird mit $d\nu/d\mu$ bezeichnet.

Folgerung 1.41. *Seien μ und ν σ-endliche Maße auf $(\Omega, \mathfrak{A})$. Weiterhin sei ν absolutstetig bezüglich μ mit Dichte $f = d\nu/d\mu$. Dann gilt:*
Eine $\mathfrak{A}$-messbare Funktion $g : \Omega \to \mathbb{R}$ ist genau dann ν integrierbar, wnn $g \cdot f$ μ-integrierbar ist. Ist dies erfüllt, so gilt

$$\int g \, d\nu = \int gf \, d\nu. \tag{1.4}$$

Beweisidee. Nach dem Satz von Radon-Nikodym gilt (1.4) für $g = \mathbb{1}_A$, $A \in \mathfrak{A}$. □

1.8. $\mathcal{L}^p$-Räume

Im folgenden sei $(\Omega, \mathfrak{A}, \mu)$ wieder ein Maßraum und es gelte $p \geq 1$.

Eine $\mathfrak{A}$-messbare Funktion $f : \Omega \to \mathbb{R}$ heißt *p-fach μ-integrierbar*, falls $|f|^p$ μ-integrierbar ist (d.h. falls $\int |f|^p \, d\mu < \infty$).

$\mathcal{L}^p(\mu) = \mathcal{L}^p(\Omega, \mathfrak{A}, \mu)$ ist der Raum der p-fach μ-integrierbaren Funktionen.

$||f||_p := \left(\int |f|^p \, d\mu\right)^{1/p}$ ist eine Seminorm auf diesem Raum $\mathcal{L}^p$.

Diese Seminorm wird zu einer Norm, wenn man zu Äquivalenzklassen bezüglich $f \sim g :\Longleftrightarrow f = g$ μ-fast überall übergeht.

Der Raum der Äquivalenzklassen wird mit $L^p(\mu) = L^p(\Omega, \mathfrak{A}, \mu)$ bezeichnet und ist ein Banachraum. (Siehe dazu auch das Skript zur „Maß- und Integrationstheorie“.)

Wichtige Ungleichungen: Seien dazu f, g messbare Funktionen auf $(\Omega, \mathfrak{A})$, dann gelten die folgenden Ungleichungen:

Hölder-Ungleichung: Seien p und $q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$\int |fg|\, d\mu \leq \left(\int |f|^p \, d\mu \right)^{1/p} \left(\int |g|^q \, d\mu \right)^{1/q}.$$

Minkowski-Ungleichung: Für $p \geq 1$ gilt

$$\left(\int |f+g|^p \, d\mu \right)^{1/p} \leq \left(\int |f|^p \, d\mu \right)^{1/p} + \left(\int |g|^p \, d\mu \right)^{1/p}.$$

(Dies ist eine Dreiecksungleichung für die L^p-Norm.)

Kapitel 2

Zufallsvariable und Unabhängigkeit

2.1. Zufallsvariable und deren Verteilung

Sei für diesen Abschnitt $(\Omega, \mathfrak{A}, \mathbb{P})$ als ein Wahrscheinlichkeitsraum vorausgesetzt.

Definition 2.1. Sei $(E, \mathfrak{E})$ ein messbarer Raum.

a) Eine $\mathfrak{A}$-$\mathfrak{E}$-messbare Abbildung $X : \Omega \to E$ heißt *E-wertige Zufallsvariable.*

b) Das Bildmaß $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$ von $\mathbb{P}$ bezüglich der Zufallsvariablen X heißt *Verteilung von X.*

$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$, für $B \in \mathfrak{E}$, definiert ein Wahrscheinlichkeitsmaß auf dem messbaren Raum $(E, \mathfrak{E})$.
Ist X reellwertig (eine *Zufallsgröße*), das heißt $(E, \mathfrak{E}) = (\mathbb{R}, \mathfrak{B})$, so wird die Verteilung von X vollständig durch die zugehörige *Verteilungsfunktion* charakterisiert:

$$F_X(x) := \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R},$$

(siehe Satz 1.15).

Spezialfälle:

a) *Diskrete Verteilungen*
Sei $\mathbb{P}_X = \sum_{n=1}^{\infty} p_n \delta_{x_n}$ mit paarweise verschiedenen $x_n \in \mathbb{R}$, $p_n \geq 0$, $\sum_{n=1}^{\infty} p_n = 1$.
Charakterisierung durch Einzelwahrscheinlichkeiten:

$$p_n = \mathbb{P}_X(\{x_n\}) = \mathbb{P}(X = x_n), \qquad n \in \mathbb{N}.$$

Zugehörige Verteilungsfunktion:

$$F_X(x) = \sum_{n: x_n \geq x} p_n, \quad x \in \mathbb{R}.$$

b) *Absolutstetige Verteilungen*
Es gelte $\mathbb{P} \ll \lambda$ (λ Lebesgue-Maß auf $(\mathbb{R}, \mathfrak{B})$).
Nach dem Satz von Radon-Nikodym (Satz 1.39) besitzt $\mathbb{P}_X$ eine Dichte f bezüglich λ, mit $f = d\mathbb{P}_X/d\lambda$, und

$$\mathbb{P}(X \in B) = \mathbb{P}_X(B) = \int_B f \, d\lambda, \quad B \in \mathfrak{B}.$$

Insbesondere ist

$$F_X(x) = \int_{(-\infty,x]} f \, d\lambda \quad \left(= \int_{-\infty}^{x} f(y) \, dy \right), \quad x \in \mathbb{R}.$$

Ist X ein d-dimensionaler *Zufallsvektor* (das heißt es gilt $(E, \mathfrak{E}) = (\mathbb{R}^d, \mathfrak{B}^d)$) und bezeichnet außerdem

$$\begin{aligned} \pi_i : \mathbb{R}^d &\longrightarrow \mathbb{R} \\ (x_1, \dots, x_d) &\longmapsto x_i \end{aligned}$$

die Projektion auf die i-te Koordinate ($i = 1, \dots, d$), so läßt sich X in der Form

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

schreiben mit $X_i := \pi_i \circ X \ \ (i = 1, \dots, d)$.

Behauptung 2.2. *X ist genau dann ein Zufallsvektor, wenn $X_1, \dots, X_d$ Zufallsgrößen sind.*

Beweis.

„$\Longrightarrow$“:
π_i ist stetig und somit Borel-messbar, weshalb $X_i = \pi_i \circ X$ als Komposition messbarer Abbildungen wiederum messbar ist ($i = 1, 2, \dots, d$).

„$\Longleftarrow$“:
Die Quader Q der Gestalt $Q = \bigtimes_{i=1}^{d} (a_i, b_i]$ erzeugen $\mathfrak{B}^d$ und es gilt

$$X^{-1}(Q) = \left\{ X \in \bigtimes_{i=1}^{d} (a_i, b_i] \right\} = \bigcap_{i=1}^{d} \underbrace{\{X_i \in (a_i, b_i]\}}_{\in \mathfrak{A}, \text{ da} X_i \text{ messbar}} \in \mathfrak{A}.$$

Die Anwendung des Messbarkeitskriteriums (Satz 1.17) liefert nun die Behauptung. $\square$

2.2. Erwartungswerte

Als Vorraussetzung für diesen Abschnitt sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Nun sei X eine Zufallsgröße (das heißt eine *reellwertige* Zufallsvariable).
Der *Erwartungswert von* X ist definiert als

$$\mathbb{E}X := \int X \, d\mathbb{P},$$

falls das Integral auf der rechten Seite existiert. Ist $X \in \mathcal{L}^1(\mathbb{P})$, dass heißt $\mathbb{E}|X| < \infty$, so sagt man X besitze einen *endlichen* Erwartungswert.

Sei nun weiter $(E, \mathfrak{E})$ ein messbarer Raum, X eine E-wertige Zufallsvariable und $\varphi : E \to \mathbb{R}$ eine $\mathfrak{E}$-messbare Funktion. Dann ist $\varphi \circ X = \varphi(X)$ eine Zufallsgröße und (nach dem Integraltransformationssatz (Satz 1.28))

$$\mathbb{E}\,\varphi(X) = \int_\Omega \varphi \circ X \, d\mathbb{P} = \int_\mathbb{R} \varphi \, d\mathbb{P}_X,$$

falls eines der beiden Integrale auf der rechten Seite existiert. Insbesondere ist für reellwertige X (und $\varphi(x) = x$)

$$\mathbb{E}X = \int_\Omega X \, d\mathbb{P} = \int_\mathbb{R} x \, \mathbb{P}_X \, (dx) \; \underbrace{\left(= \int_{-\infty}^{\infty} x F_X \, (dx) \right)}_{\text{Lebesgue-Stieltjes-Integral}},$$

falls eines der Integrale existiert.

Bemerkung 2.3. Die betrachteten Erwartungswerte hängen von der Zufallsgröße X nur über deren Verteilung $\mathbb{P}_X$ ab.

Spezialfälle:

a) Sei X *diskret* mit Werten $x_n \in \mathbb{R}$ und Einzelwahrscheinlichkeiten

$$p_n = \mathbb{P}(X = x_n)(n \in \mathbb{N}), \text{ d.h. } \mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}.$$

Ist $\varphi : \mathbb{R} \to \mathbb{R}$ Borel-messbar, so gilt

$$\mathbb{E}|\varphi(X)| = \int |\varphi| \, d\mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n |\varphi(x_n)|.$$

Ist dieser Ausdruck endlich, so besitzt $\varphi(X)$ einen endlichen Erwartungswert und

$$\mathbb{E}\varphi(X) = \int \varphi \, d\mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n \varphi(x_n).$$

b) Sei X *absolutstetig* mit der Dichte $f := d\mathbb{P}_X/d\lambda$ und $\varphi : \mathbb{R} \to \mathbb{R}$ sei wieder Borel-messbar. Dann gilt mit Folgerung (1.41) $(*)$

$$\mathbb{E}|\varphi(X)| = \int |\varphi|\, d\mathbb{P}_X \underset{(*)}{=} \int |\varphi| \cdot f\, d\lambda$$

Ist dieser Ausdruck wiederum endlich, so besitzt $\varphi(X)$ wieder einen endlichen Erwartungswert und

$$\mathbb{E}\varphi(X) = \int \varphi\, d\mathbb{P}_X = \int \varphi \cdot f\, d\lambda.$$

Bemerkung 2.4. Viele Eigenschaften des Erwartungswertes (z. B. Linearität, Hölder-Ungleichung, ...) sind Eigenschaften der Integrale $\int \dots d\mathbb{P}$.

Beispiel 2.5. (Cauchy-Verteilung)
Die Cauchy-Verteilung zum Parameter $c > 0$ ist eine absolutstetige Verteilung mit Dichte

$$f(x) = \frac{1}{\pi} \frac{c}{c^2 + x^2} \quad \text{für } x \in \mathbb{R}.$$

f ist tatsächlich Dichte einer Verteilung, da f stetig und somit Borel-messbar ist, $f \geq 0$ gilt und da mit Hilfe des Satzes über monotone Konvergenz $(*)$

$$\int f\, d\lambda \underset{(*)}{=} \lim_{x\to\infty} \int_{-x}^{x} f(y)\, dy = \lim_{x\to\infty} \frac{1}{\pi} \left[\arctan(\frac{x}{c}) - \arctan(-\frac{x}{c})\right] = 1.$$

Der zugehörige Erwartungswert existiert nicht:

$$\varphi(x) = x,\ \varphi^+(x) = x\mathbb{1}_{[0,\infty)}(x),\ \varphi^-(x) = (-x)\mathbb{1}_{(-\infty,0]}(x)$$

$$\int x\, \mathbb{1}_{[0,\infty)}(x)\, f(x)\lambda\,(dx) = \lim_{R\to\infty} \int_0^R \frac{1}{\pi} \frac{cx}{c^2 + x^2} dx = \infty,$$

da $\int\limits_0^R \frac{1}{\pi} \frac{cx}{c^2+x^2} dx = \frac{c}{2\pi} \log(c^2 + x^2)\Big|_0^R$ und analog

$$\int (-x)\, \mathbb{1}_{(-\infty,0]}(x)\, f(x)\lambda\,(dx) = \infty.$$

2.3. Konvergenzarten für Folgen von Zufallsgrößen

Wir setzen X_n, X als Zufallsgrößen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ voraus.

Besteht eine Eigenschaft fast überall bezüglich eines Wahrscheinlichkeitsmaßes, so sagt man in der Wahrscheinlichkeitstheorie, dass diese Eigenschaft *fast sicher* (*f.s.*) gilt.

Definition 2.6.

a) (X_n) konvergiert *fast sicher* gegen X, falls $X_n \to X$ fast sicher, das heißt

$$\mathbb{P}(X_n \nrightarrow X) = 0.$$

b) (X_n) konvergiert *im p-ten Mittel* (*im $\mathcal{L}^p$*) gegen X, falls X_n und $X \in \mathcal{L}^p$ und $||X_n - X||_p \to 0$, das heißt

$$\mathbb{E}|X_n - X|^p \to 0 \quad \text{für } n \to \infty.$$

c) (X_n) konvergiert *in Wahrscheinlichkeit* (*stochastisch*) gegen X, falls für alle $\varepsilon > 0$ gilt

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

In Zeichen:

$$X_n \xrightarrow{f.s.} X, \quad X_n \xrightarrow{\mathcal{L}^p} X, \quad X_n \xrightarrow{\mathbb{P}} X.$$

Bemerkung 2.7. Die Grenzwerte sind nur $\mathbb{P}$-fast sicher eindeutig bestimmt.

Satz 2.8.

(i) Aus der fast sicheren Konvergenz folgt die Konvergenz in Wahrscheinlichkeit.

(ii) Aus der Konvergenz im p-ten Mittel folgt die Konvergenz in Wahrscheinlichkeit.

Beweis. Siehe Wahrscheinlichkeitstheorie I. □

Beispiel 2.9. Alle anderen Implikationen zwischen den drei Konvergenzarten sind im Allgemeinen (ohne Zusatzvorraussetzungen) falsch. Sei etwa

$$(\Omega, \mathfrak{A}, \mathbb{P}) = ([0,1]\mathfrak{B}_{[0,1]}, \lambda) \text{ mit } \lambda = \lambda_{[0,1]} \text{ Lebesgue-Maß auf } [0,1]$$

$$X_{2^m+k} := 2^{2m} \mathbb{1}_{[k2^{-m},(k+1)2^{-m}]} \qquad (m \in \mathbb{N}_0, 0 \le k < 2^m),$$

so konvergiert (X_n) in Wahrscheinlichkeit gegen 0, aber (X_n) konvergiert weder im p-ten Mittel (für $p \ge 1$) noch fast sicher:

$$n = 2^m + k \to \infty \quad \Longleftrightarrow \quad m \to \infty$$

$$0 < \varepsilon < 1, \;\; \mathbb{P}(|X_{2^m+k}| > \varepsilon) = \lambda([k2^{-m}, (k+1)2^{-m}]) = 2^{-m} \xrightarrow[m\to\infty]{} 0,$$

aber

(i) $||X_{2^m+k}||_p^p = \mathbb{E}|X_{2^m+k}|^p = 2^{2mp}\lambda([k2^{-m},(k+1)2^{-m}]) = 2^{(2p-1)m} \underset{m\to\infty}{\longrightarrow} \infty.$
(Würde $X_n \overset{\mathcal{L}^p}{\to} X$ gelten, so wäre

$$||X_n||_p \leq \underbrace{||X_n - X||_p}_{\to 0} + \underbrace{||X||_p}_{<\infty} \text{ beschränkt.)}$$

(ii) für jedes $\omega \in [0,1]$ findet man eine Teilfolge (X_{n_k}) mit $(X_{n_k}(\omega)) \to \infty$. Also konvergiert $(X_n(\omega))$ für kein ω.

Satz 2.10. *(Satz von Riesz) Konvergiert (X_n) in Wahrscheinlichkeit gegen X, so findet man eine Teilfolge (X_{n_k}), die fast sicher gegen X konvergiert.*

Beweis. Wie wählen die Teilfolge (X_{n_k}) so, dass

$$\mathbb{P}(|X_{n_k} - X| > 2^{-k}) < 2^{-k} \qquad (k = 1, 2, \dots)$$

gilt. Die Ereignisse $A_k := \{|X_{n_k} - X| > 2^{-k}\}$ erfüllen somit $\sum_k \mathbb{P}(A_k) < \infty$. Also ist

$$N := \limsup_{k\to\infty} A_k = \bigcup_n \bigcap_{l\geq k} A_l$$

eine $\mathbb{P}$-Nullmenge. N ist das Ereignis, dass unendlich viele der Ereignisse A_k eintreten. Für $\omega \in \Omega\setminus N$ folgt deshalb $X_{n_k}(\omega) \to X(\omega)$. Also gilt $X_{n_k} \to X$ $\mathbb{P}$-fast sicher. □

Alle drei Konvergenzarten sind *vollständig*, das heißt jede Cauchyfolge konvergiert. Wir formulieren dies nur für die Konvergenz in Wahrscheinlichkeit.

Satz 2.11. *Sei (X_n) eine Cauchyfolge in Wahrscheinlichkeit, das heißt für alle $\varepsilon > 0$ gilt:*

$$\lim_{m,n\to\infty} \mathbb{P}(|X_n - X_m| > \varepsilon) = 0. \tag{2.1}$$

Dann existiert eine Zufallsgröße X, so dass (X_n) in Wahrscheinlichkeit gegen X konvergiert.

Beweis. Wegen (2.1) findet man eine Teilfolge (X_{n_k}) mit

$$\mathbb{P}(|X_{n_{k+1}} - X_{n_k} > 2^{-k}) < 2^{-k} \quad (k = 1, 2, \dots).$$

Auf die Ereignisse $A_k := \{|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\}$ lässt sich somit das Lemma von Borel-Cantelli anwenden, weshalb $\limsup_{k\to\infty} A_k$ eine $\mathbb{P}$-Nullmenge ist. Für $\omega \in \limsup_{k\to\infty} A_k$ ist aber $(X_{n_k}(\omega))$ eine Cauchyfolge in $\mathbb{R}$. Da $\mathbb{R}$ vollständig ist, findet man eine Zufallsgröße X (Wie steht es um die Messbarkeit von X), so dass $X_{n_k} \to X$ fast sicher und folglich auch $X_{n_k} \to X$ in Wahrscheinlichkeit gilt.

Bleibt zu zeigen, dass die gesamte Folge (X_n) in Wahrscheinlichkeit gegen X konvergiert: Für beliebige $\varepsilon > 0$ gilt

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \underbrace{\mathbb{P}(|X_n - X_{n_k}| > \frac{\varepsilon}{2})}_{\to 0,\text{ für } n,k\to\infty \text{ wegen (2.1)}} + \underbrace{\mathbb{P}(|X_{n_k} - X| > \frac{\varepsilon}{2})}_{\to 0,\text{ für } k\to\infty \text{ (Konv. in Wkt.)}} .$$

□

2.4. Unabhängigkeit von Mengensystemen

Sei für diesen Abschnitt $(\Omega, \mathfrak{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Ereignisse $A_1, \dots, A_n \in \mathfrak{F}$ heißen *unabhängig*, wenn für beliebige $i_1, \dots, i_r$ mit $1 \leq i_1 < \dots < i_r \leq n$ gilt

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) = P(A_{i_1}) \cdots P(A_{i_r}).$$

Eine Familie $(A_i)_{i\in I}$ von Ereignissen heißt unabhängig, wenn je endlich viele dieser Ereignisse unabhängig sind.

Definition 2.12.

a) Teilmengen $\mathfrak{E}_1, \dots, \mathfrak{E}_n$ von $\mathfrak{A}$ mit $\Omega \in \mathfrak{E}_i$ $(i = 1, \dots, n)$ heißen unabhängig, wenn für beliebige $A_1 \in \mathfrak{E}_1, \dots, A_n \in \mathfrak{E}_n$ gilt

$$\mathbb{P}(A_1 \cap \dots A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n).$$

b) Eine Familie $(\mathfrak{E}_i)_{i\in I}$ von Teilmengen von $\mathfrak{A}$ mit $\Omega \in \mathfrak{E}_i$ für alle $i \in I$ heißt unabhängig, falls jeweils endliche viele dieser Mengensysteme unabhängig sind.

Die Unabhängigkeit von $\mathfrak{E}_1$ und $\mathfrak{E}_2$ drückt man symbolisch durch $\perp\!\!\!\perp$ aus.

Satz 2.13. *Sei $(\mathfrak{E}_i)_{i\in I}$ eine Familie von $\cap$-stabilen Teilmengen von $\mathfrak{A}$ mit $\Omega \in \mathfrak{E}_i$ für alle $i \in I$. Falls dann die Teilmengen $(\mathfrak{E}_i)_{i\in I}$ unabhängig sind, dann sind auch $(\sigma(\mathfrak{E}_i))_{i\in I}$ unabhängig.*

Beweis. Sei o.B.d.A. $I = \{1, \dots, n\}$ endlich. Wir betrachten die Aussagen

$$(H_k) \qquad \mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n) \text{ gilt für}$$
$$A_1 \in \sigma(\mathfrak{E}_1), \dots, A_k \in \sigma(\mathfrak{E}_k), A_{k+1} \in \mathfrak{E}_{k+1}, \dots, A_n \in \mathfrak{E}_n$$

mit $(k = 0, 1, \dots, n)$. Nach Vorraussetzung gilt (H_0), zu zeigen ist (H_n).
Wie beweisen dies mit nun Induktion über k.
Sei also (H_k) für ein k mit $0 \leq k < n$ erfüllt.

$$\begin{aligned} \text{Sei } \vartheta := \{B \in \sigma(\mathfrak{E}_{k+1}) \quad &: \quad \mathbb{P}(A_1 \cap \dots \cap A_k \cap B \cap A_{k+2} \cap \dots \cap A_n) \\ &= \quad \mathbb{P}(A_1) \cdots \mathbb{P}(A_k)\, \mathbb{P}(B)\, \mathbb{P}(A_{k+2}) \cdots \mathbb{P}(A_n) \\ &\qquad \text{für alle } A_1 \in \sigma(\mathfrak{E}_1), \dots, A_k \in \sigma(\mathfrak{E}_k), \\ &\qquad A_{k+2} \in \mathfrak{E}_{k+2}, \cdots, A_n \in \mathfrak{E}_n\} \end{aligned}$$

Es genügt zu zeigen, dass ϑ ein Dynkin-System ist. Da $\mathfrak{E}_{k+1} \subseteq \vartheta$ und wegen der $\cap$-Stabilität von $\mathfrak{E}_{k+1}$ folgt mit dem Hauptsatz über Dynkin-Systeme

$$\sigma(\mathfrak{E}_{k+1}) = d(\mathfrak{E}_{k+1}) \subseteq \vartheta, \quad \text{d.h. } (H_{k+1}) \text{ ist erfüllt.}$$

ϑ ist ein Dynkin-System, denn $\Omega \in \vartheta$, da $\Omega \in \mathfrak{E}_{k+1}$

$$\begin{aligned}
B \in \vartheta \quad \Rightarrow \quad & \mathbb{P}(A_1 \cap \cdots \cap A_k \cap B^c \cap A_{k+2} \cap \cdots \cap A_n) \\
= \quad & \mathbb{P}(A_1 \cap \cdots \cap A_k \cap \Omega \cap A_{k+2} \cap \ldots A_n) \\
& \quad - \mathbb{P}(A_1 \cap \cdots \cap A_k \cap B \cap A_{k+2} \cap \cdots \cap A_n) \\
= \quad & \mathbb{P}(A_1) \cdots \mathbb{P}(A_k)\, \mathbb{P}(\Omega)\, \mathbb{P}(A_{k+2}) \cdots \mathbb{P}(A_n) \\
& \quad - \mathbb{P}(A_1) \cdots \mathbb{P}(A_k)\, \mathbb{P}(B)\, \mathbb{P}(A_{k+2}) \cdots \mathbb{P}(A_n) \\
= \quad & \mathbb{P}(A_1) \cdots \mathbb{P}(A_k)\, \mathbb{P}(B^c)\, \mathbb{P}(A_{k+1}) \cdots (A_n), \text{ d.h. } B \in \vartheta.
\end{aligned}$$

Analog folgt für eine Folge (B_n) paarweise disjunkter Ereignisse aus ϑ mit der σ-Additivität von $\mathbb{P}$, dass $\bigcup\limits_n B_n \in \vartheta$. □

Folgerung 2.14. *Sei $(A_i)_{i \in I}$ eine Familie unabhängiger Ereignisse. Dann sind auch die σ-Algebren $\{\phi, A_i, A_i^c, \Omega\}$, $i \in I$, unabhängig.*

Beweis. Man wende den vorherigen Satz (Satz 2.13) auf die Mengensysteme $\mathfrak{E}_i := \{A_i, \Omega\}$, für $i \in I$, an. □

Satz 2.15. *Seien $\mathfrak{E}_i \subseteq \mathfrak{A}$, für $i \in I$, sowohl unabhängig als auch $\cap$-stabil und sei $\Omega \in \mathfrak{E}_i$ (für alle $i \in I$). Weiterhin sei $(I_k)_{k \in K}$ eine Familie von paarweise disjunkten Teilmengen der Indexmenge I. Dann sind auch die Mengensysteme*

$$\sigma\left(\bigcup_{i \in I_k} \mathfrak{E}_i\right) \quad \textit{für } k \in K, \ \textit{unabhängig.}$$

Beweis. Sei $\widehat{\mathfrak{E}_k}$ das Mengensystem, welches aus allen endlichen Durchschnitten von Ereignissen aus $\mathfrak{E}_i$, $i \in I_k$, besteht (für $k \in K$). Dann sind die Mengensysteme $\widehat{\mathfrak{E}_i}$, $k \in K$, sowohl unabhängig, als auch $\cap$-stabil und es gilt

$$\sigma\left(\widehat{\mathfrak{E}_k}\right) = \sigma\left(\bigcup_{i \in I_k} \mathfrak{E}_i\right).$$

Die Behauptung folgt nun unter der Verwendung des Satzes 2.13. □

Satz 2.16. (0-1-*Gesetz von Kolmogorov)*
Sei $(\mathfrak{A}_n)$ eine Folge unabhängiger Teil-σ-Algebren von $\mathfrak{A}$ und sei $\mathfrak{A}_{\geq m} := \bigvee\limits_{n \geq m} \mathfrak{A}_n$. Dann ist die Tail-$\sigma$-Algebra

$$\mathcal{T} := \bigcap_{m=1}^{\infty} \mathfrak{A}_{\geq m}$$

trivial, das heisst, dass für alle $A \in \mathcal{T}$ gilt:

$$\mathbb{P}(A) = 0 \text{ oder } \mathbb{P}(A) = 1.$$

Deutung: Eine Zufallsevolution mit unabhängigen Ausgängen zu verschiedenen Zeitpunkten $\mathfrak{A}_n$ enthält die Ereignisse, die nur vom Geschehen zum Zeitpunkt n abhängen.
$\mathfrak{A}_{\geq m}$ enthält die Ereignisse, die nur von der zukünftigen Entwicklung ab dem Zeitpunkt m abhängen.
$\mathcal{T}$ ist die σ-Algebra der Ereignisse, die nur vom Geschehen in der unendlich fernen Zukunft abhängen.

Beweis. Sei $\mathfrak{A}_{\leq m} := \bigvee_{i=1}^{m} \mathfrak{A}_i$. Satz 2.15 liefert uns

$$\begin{aligned}
\mathfrak{A}_{\leq m} \perp\!\!\!\perp \mathfrak{A}_{\geq m+1} \text{ für alle } m \quad &\Rightarrow \quad \mathfrak{A}_{\leq m} \perp\!\!\!\perp \mathcal{T} \qquad \text{für alle } m \text{ (da } \mathcal{T} \subseteq \mathfrak{A}_{\geq m})\\
&\underset{\text{Satz 2.15}}{\Rightarrow} \quad \bigvee_{n=1}^{\infty} = \bigvee_{m=1}^{\infty} \mathfrak{A}_{\leq m} \perp\!\!\!\perp \mathcal{T}\\
&\Rightarrow \quad \mathcal{T} \perp\!\!\!\perp \mathcal{T} \quad (\text{da } \mathcal{T} \subseteq \bigvee_{n=1}^{\infty} \mathfrak{A}_n),
\end{aligned}$$

das heißt für alle $A \in \mathcal{T}$ gilt $\mathbb{P}(A \cap A) = \mathbb{P}(A)$, also ist $\mathbb{P}(A)^2 = \mathbb{P}(A)$, weshalb $\mathbb{P}(A) = 0$ oder $\mathbb{P}(A) = 1$ gelten muss. □

Beispiel 2.17. (Illustrationsbeispiel)

1. Zum Lemma von Borel-Cantelli:
   Sei (A_n) eine Folge unabhängiger Ereignisse. Dann sind die von diesen Mengen erzeugten σ-Algebren $\mathfrak{A}_n := \{\emptyset, A_n, A_n^c, \Omega\}$ unabhängig (Folgerung 2.14). Also ist nach dem 0-1-Gesetz von Kolmogorov die Tail-σ-Algebra
   $$\mathcal{T} := \bigcap_{m=1}^{\infty} \bigvee_{n=m}^{\infty} \mathfrak{A}_n$$
   trivial.
   Da $A := \limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n \in \mathcal{T}$, ist somit $\mathbb{P}(A) = 0$ oder $\mathbb{P}(A) = 1$.

2. Zum starken Gesetz der großen Zahlen:
   Sei (A_n) eine Folge unabhängiger (und identisch verteilter) Zufallsgrößen. Angenommen, es gilt
   $$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = Y \quad \text{fast sicher.}$$
   Dann ist Y fast sicher konstant. Tatsächlich, die σ-Algebren $\sigma(X_n)$, mit $n \in \mathbb{N}$, sind unabhängig, weshalb nach dem 0-1-Gesetz von Kolmogorov

die Tail-σ-Algebra $\mathcal{T} := \bigcap_{m=1}^{\infty} \bigvee_{n=m}^{\infty} \sigma(A_n)$ trivial ist. Für beliebige $m \in \mathbb{N}$ ist

$$Y = \lim_{n\to\infty} \frac{1}{n} \sum_{i=m}^{n} X_i.$$

Also ist Y bezüglich der σ-Algebra $\bigvee_{n=m}^{\infty} \sigma(X_n)$ messbar. Da dies für beliebige m gilt, ist Y auch $\mathcal{T}$-messbar.
Wegen der Trivialität von $\mathcal{T}$ folgt die Behauptung. (Wäre Y nicht fast sicher konstant, so gäbe es ein $c \in \mathbb{R}$ mit $0 < \mathbb{P}(Y \leq c) < 1$; man betrachte zum Beispiel die Verteilungsfunktion von Y.)

2.5. Unabhängigkeit von Zufallsgrößen

Sei für diesen Abschnitt mit $(\Omega, \mathfrak{A}, \mathbb{P})$ wieder ein beliebiger Wahrscheinlichkeitsraum vorgegeben.

Aus der Definition der Unabhängigkeit von Zufallsvariablen ergibt sich sofort die nachstehende Behauptung.

Behauptung 2.18. *Sei I eine nichtleere Indexmenge. Für jedes $i \in I$ sei X_i eine $(E_, \mathfrak{E}_i)$-wertige Zufallsvariable auf dem Raum $(\Omega, \mathfrak{A}, \mathbb{P})$. Dann gilt:*
Die Famile $(X_i)_{i\in I}$ von Zufallsvariablen ist genau dann unabhängig, wenn die Familie $(\sigma(X_i))_{i\in I}$ der von ihnen erzeugenten σ-Algebren unabhängig ist.

Satz 2.19. *Seien X und Y zwei unabhängige Zufallsgrößen.*

(i) Sind X, $Y \geq 0$, so gilt

$$\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y. \tag{2.2}$$

(ii) Sind X, $Y \in \mathcal{L}^1(\mathbb{P})$, so ist auch $X \cdot Y \in \mathcal{L}^1(\mathbb{P})$ und die Behauptung (2.2) gilt ebenfalls.

Beweis.

(i) Da X und Y unabhängig sind, sind auch die σ-Algebren $\sigma(X)$ sowie $\sigma(Y)$ unabhängig (Behauptung (2.18)). Die Behauptung (2.2) gilt für nichtnegative elementare $\sigma(X)$- bzw. $\sigma(Y)$-messbare Zufallsgrößen $\tilde{X}$ und $\tilde{Y}$. Tatsächlich, ist

$$\begin{aligned} \tilde{X} &= \sum_{i=1}^{m} \alpha_i \mathbb{1}_{A_i} \quad \text{mit } A_1, \ldots, A_m \in \sigma(X) \\ \text{und } \tilde{Y} &= \sum_{j=1}^{n} \beta_j \mathbb{1}_{B_j} \quad \text{mit } B_1, \ldots, B_n \in \sigma(Y), \end{aligned}$$

so erhält man da $\mathbb{E}(\mathbb{1}_{A_i}\mathbb{1}_{B_j}) = \mathbb{P}(A_i \cap B_j) \underset{\text{Unabh.}}{=} \mathbb{P}(A_i)\,\mathbb{P}(B_j)$

$$\begin{aligned}
\mathbb{E}(\tilde{X}\tilde{Y}) &= \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_i\beta_j \cdot \mathbb{E}(\mathbb{1}_{A_i}\mathbb{1}_{B_j}) \\
&= \sum_{i=1}^{m} \alpha_i\beta_j \cdot \mathbb{E}\mathbb{1}_{A_i} \cdot \mathbb{E}\mathbb{1}_{B_j} \\
&= \left(\sum_{i=1}^{m} \alpha_i\, \mathbb{E}\mathbb{1}_{A_i}\right)\left(\sum_{j=1}^{n} \beta_j\, \mathbb{E}\mathbb{1}_{B_j}\right) \\
&= \mathbb{E}\tilde{X} \cdot \mathbb{E}\tilde{Y}.
\end{aligned}$$

Nach Satz (1.22) existieren elementare $\sigma(X)$- beziehungsweise $\sigma(Y)$-messbare Zufallsgrößen X_m und Y_n mit $0 \le X_m \uparrow X$ sowie $0 \le Y_n \uparrow Y$. Es gilt

$$\mathbb{E}(X_m \cdot Y_n) = \mathbb{E}X_m \cdot \mathbb{E}Y_n.$$

Die Behauptung (2.2) folgt nun mit dem Satz über monotone Konvergenz (Satz 1.29) für $m \uparrow \infty$ und $n \uparrow \infty$.

(ii) Da X und Y unabhängig sind, sind auch $|X|$ und $|Y|$ unabhängig (Da $\sigma(|X|) \subseteq \sigma(X)$, $\sigma(|Y|) \subseteq \sigma(Y)$). Also folgt mit (i)

$$\mathbb{E}(|X| \cdot |Y|) = \mathbb{E}|X| \cdot \mathbb{E}|Y| < \infty,$$

das heißt $X \cdot Y \in \mathcal{L}^1(\mathbb{P})$. Der Rest der Behauptung (ii) folgt aus (i) durch Zerlegung von X und Y in Positiv- und Negativteil.

□

Charakteristische Funktion einer Zufallsgröße X:

$$\chi_X(t) := \mathbb{E}e^{itX}, \quad \text{für } t \in \mathbb{R}.$$

Folgerung 2.20. *Sind X und Y unabhängige Zufallsgrößen, so gilt*

$$\chi_{X+Y} = \chi_X \cdot \chi_Y\,.$$

Beweis. Aus der Unabhängigkeit von X und Y folgt, dass e^{itX} und e^{itY} unabhängige komplexwertige Zufallsvariablen ($\sigma(e^{itX}) \subseteq \sigma(X)$, $\sigma(e^{itY}) \subseteq \sigma(Y)$) sind. Also folgt mit Satz 2.19 b), dass

$$\chi_{X+Y}(t) = \mathbb{E}(e^{itX} \cdot e^{itY}) = \mathbb{E}e^{itX} \cdot \mathbb{E}e^{itY} = \chi_X(t) \cdot \chi_Y(t).$$

Das Problem der Komplexwertigkeit von e^{itX} und e^{itY} lässt sich entweder durch eine Verallgemeinerung des Satzes 2.19 oder durch Zerlegung von e^{itX} und e^{itY} in Real- und Imaginärteil lösen. □

Satz 2.21. *Seien μ und ν zwei Wahrscheinlichkeitsmaße auf dem Raum $(\mathbb{R}, \mathfrak{B})$. Dann wird durch*

$$(\mu * \nu)(B) := \int \nu(B-x)\mu\,(dx) = \int \mu(B-y)\nu\,(dy), \quad \textit{für } B \in \mathfrak{B},$$

ein Wahrscheinlichkeitsmaß $\mu * \nu$ auf $(\mathbb{R}, \mathfrak{B})$ definiert.

Beweis. Sei also $B \in \mathfrak{B}$. Dann ist die Funktion

$$f(x,y) := \mathbb{1}_B(x+y) = \mathbb{1}_{B-x}(y) = \mathbb{1}_{B-y}(x), \quad \text{mit } (x,y) \in \mathbb{R}^2,$$

$\mathfrak{B} \otimes \mathfrak{B}(= \mathfrak{B}^2)$-messbar und der Satz von Fubini (Satz 1.34) liefert uns

$$\int \nu(B-x)\mu\,(dx) = \int \mu(B-y)\nu\,(dy) = \int \mathbb{1}_B(x+y)(\mu \otimes \nu)\,(d(x,y)). \tag{2.3}$$

Wendet man nun auf das letzte Integral den Satz von Beppo Levi (Folgerung 1.30) an, so sieht man, dass $\mu * \nu$ ein Wahrscheinlichkeitsmaß ist. □

Bemerkung 2.22.

1. $\mu * \nu$ heißt *Faltung* von μ und ν.
2. Es gilt $\mu * \nu = \nu * \mu$.
3. Sind μ und ν diskret (absolutstetig), so folgen aus Satz 2.21 die Faltungsformeln für Einzelwahrscheinlichkeiten (Dichten) aus Wahrscheinlichkeitstheorie I.

Satz 2.23. *Sind X und Y unabhängige Zufallsgrößen, so gilt für deren Verteilung*

(i) $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$ *sowie*

(ii) $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.

Beweis.

(i) Für beliebige B_1, $B_2 \in \mathfrak{B}$ gilt mit Hilfe der Unabhängigkeit $(*)$

$$\begin{aligned}
\mathbb{P}_{(X,Y)}(B_1 \times B_2) &= \mathbb{P}((X,Y) \in B_1 \times B_2) \\
&= \mathbb{P}(X \in B_1,\, Y \in B_2) \\
&\underset{(*)}{=} \mathbb{P}(X \in B_1) \cdot \mathbb{P}(Y \in B_2) \\
&= \mathbb{P}_X(B_1) \cdot \mathbb{P}_Y(B_2) \\
&= (\mathbb{P}_X \otimes \mathbb{P}_Y)(B_1 \times B_2).
\end{aligned}$$

Das bedeutet $\mathbb{P}_{(X,Y)}$ und $\mathbb{P}_X \otimes \mathbb{P}_Y$ stimmen auf einem $\cap$-stabilen Erzeuger von $\mathfrak{B} \otimes \mathfrak{B} = \mathfrak{B}^2$ überein. Also sind beide Maße identisch (siehe Beispiel 1.11).

(*ii*) Mit dem Integraltransformationssatz (Satz 1.28) folgt für beliebige $B \in \mathfrak{B}$:

$$\begin{aligned}
\mathbb{P}_{X+Y}(B) &= \mathbb{E}\mathbb{1}_B(X+Y) \\
&= \int\limits_{\mathbb{R}^2} \mathbb{1}_B(x+y)\mathbb{P}_{(X,Y)}\,(d(x,y)) \\
&\underset{a)}{=} \int\limits_{\mathbb{R}^2} \mathbb{1}_B(x+y)(\mathbb{P}_X \otimes \mathbb{P}_Y)\,(d(x,y)) \\
&\underset{2.3}{=} (\mathbb{P}_X * \mathbb{P}_Y)(B).
\end{aligned}$$

□

Kapitel 3

Bedingte Wahrscheinlichkeiten und bedingte Erwartungswerte

3.1. Klassischer Begriff

Sei für diesen Abschnitt $(\Omega, \mathfrak{A}, \mathbb{P})$ wieder ein Wahrscheinlichkeitsraum.

Für $B \in \mathfrak{A}$ mit $\mathbb{P}(B) > 0$ wird durch

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \qquad \text{für } A \in \mathfrak{A},$$

ein Wahrscheinlichkeitsmaß $\mathbb{P}(\,\cdot\,|B)$ auf $(\Omega, \mathfrak{A})$ definiert, die *bedingte Wahrscheinlichkeit* gegeben B. Der *bedingte Erwartungswert* gegeben B wird als Erwartungswert bezüglich $\mathbb{P}(\,\cdot\,|B)$ definiert:

$$\mathbb{E}[X|B] := \int X \, d\mathbb{P}(\,\cdot\,|B), \quad \text{mit } X \in \mathcal{L}^1(\mathbb{P}(\,\cdot\,|B)).$$

Behauptung 3.1. *Ist $X \in \mathcal{L}^1(\mathbb{P})$, so ist $X \in \mathcal{L}^1(\mathbb{P}(\,\cdot\,|B))$ und es gilt*

$$\mathbb{E}[X|B] = \frac{\mathbb{E}(X \mathbb{1}_B)}{\mathbb{P}(B)}.$$

Beweisidee. Die Behauptung gilt für $X = \mathbb{1}_A$, mit $A \in \mathfrak{A}$. □

Wegen

$$\mathbb{P}(A|B) = \mathbb{E}[\mathbb{1}_A|B], \qquad \text{mit } A \in \mathfrak{A},$$

kann die bedingte Wahrscheinlichkeit als „Spezialfall“ des bedingten Erwartungswertes angesehen werden.

Sei $B_1, \dots, B_r$ eine endliche messbare Zerlegung von Ω mit $\mathbb{P}(B_i) > 0$ für $i = 1, \dots, r$ (das heißt $B_1, \dots, B_r \in \mathfrak{A}$, $B_1 \cap B_j = \emptyset$ (für $i \neq j$), $\bigcup_{1=1}^{r} B_i = \Omega$).
Sei außerdem

$$\mathfrak{F} := \sigma(B_1, \dots, B_r) \subseteq \mathfrak{A}$$

die endliche σ-Algebra mit den „Atomen" $B_1, \dots, B_r$. Für $X \in \mathcal{L}^1(\mathbb{P})$ definieren wir dann

$$\begin{aligned} \mathbb{E}[X|\mathfrak{F}](\omega) &:= \mathbb{E}[X|B_i], \quad \text{falls } \omega \in B_i \\ &= \sum_{i=1}^{r} \mathbb{E}[X|B_i] \mathbb{1}_{B_i}(\omega), \end{aligned} \tag{3.1}$$

und für $A \in \mathfrak{A}$,

$$\mathbb{P}(A|\mathfrak{F}) := \mathbb{E}[\mathbb{1}_A|\mathfrak{F}] = \sum_{i=1}^{r} \mathbb{P}(A|B_i) \mathbb{1}_{B_i}.$$

Interpretation. Wir wissen bereits, welche der Ereignisse aus $\mathfrak{F}$ eingetreten sind.

a) Ist $\omega \in \Omega$ das bei unserer Beobachtung eingetretene Elementarereignis und $\omega \in B_i$ (das heißt B_i ist eingetreten), so ist $\mathbb{P}[A|\mathfrak{F}](\omega)$ die (bedingte) Wahrscheinlichkeit von A unter dieser Zusatzinformation.

b) Entsprechend ist $\mathbb{E}[X|\mathfrak{F}](\omega)$ der (bedingte) Mittelwert von X gegeben dieser Zusatzinformation.

Behauptung 3.2. *Unter den obigen Vorrausetzungen gilt*

(a) $\mathbb{E}[X|\mathfrak{F}] \in \mathcal{L}^1(\mathbb{P})$;

(b) $\mathbb{E}[X|\mathfrak{F}]$ *ist* $\mathfrak{F}$*-messbar;*

(c) $\mathbb{E}(\mathbb{1}_A \mathbb{E}[X|\mathfrak{F}]) = \mathbb{E}(\mathbb{1}_A X)$, *für* $A \in \mathfrak{F}$.

Beweis. Aus

$$\mathbb{E}[X|\mathfrak{F}] \underset{(3.1)}{=} \sum_{i=1}^{r} \mathbb{E}[X|B_i] \mathbb{1}_{B_i} \underset{\text{Beh. 3.1}}{=} \sum_{i=1}^{r} \frac{\mathbb{E}[X \mathbb{1}_{B_i}]}{\mathbb{P}(B_i)} \mathbb{1}_{B_i}$$

ergeben sich sofort die Aussagen (a) und (b). Zum Beweis von (c) sei $A \in \mathfrak{F}$, d.h.

$$A = \bigcup_{j \in J} B_j \text{ für eine Indexmenge } J \subseteq \{1, 2, \dots, r\}.$$

Dann ist

$$\begin{aligned} \mathbb{E}(\mathbb{1}_A \mathbb{E}[X|\mathfrak{F}]) &= \sum_{i=1}^{r} \sum_{j \in J} \underbrace{\frac{\mathbb{E}\mathbb{1}_{B_i \cap B_j}}{\mathbb{P}(B_i)}}_{=\delta_{ij}} \mathbb{E}(X \mathbb{1}_{B_i}) \\ &= \sum_{j \in J} \mathbb{E}(X \mathbb{1}_{B_j}) = \mathbb{E}(X \mathbb{1}_A). \end{aligned}$$

$\square$

3.2. Abstrakter Begriff

Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ wieder ein beliebiger Wahrscheinlichkeitsraum.

Satz 3.3. *Gegeben seien eine Zufallsgröße $X \in \mathcal{L}^1(\mathbb{P})$ und eine Teil-σ-Algebra $\mathfrak{F}$ von $\mathfrak{A}$. Dann existiert eine $\mathbb{P}$-fast sicher eindeutig bestimmte Zufallsgröße Y mit den folgenden Eigenschaften:*

(i) $Y \in \mathcal{L}^1(\mathbb{P})$;

(ii) Y *ist* $\mathfrak{F}$*-messbar;*

(iii) $\mathbb{E}(\mathbb{1}_A Y) = \mathbb{E}(\mathbb{1}_A X)$, *mit* $A \in \mathfrak{F}$.

Beweis.

a) *Existenz von Y*: Sei zunächst $X \geq 0$. Dann wird durch

$$Q(A) := \mathbb{E}(\mathbb{1}_A X) \text{ mit } A \in \mathfrak{F},$$

ein endliches Maß Q auf $(\Omega, \mathfrak{A})$ definiert. Wegen

$$Q \ll \mathbb{P}|_{\mathfrak{F}} \qquad \text{(Einschränkung von } \mathbb{P} \text{ auf } \mathfrak{F})$$

existiert nach dem Satz von Radon-Nikodym (Satz 1.39) eine nichtnegative $\mathfrak{F}$-messbare Zufallsgröße Y mit

$$Q(A) = \int_A Y \, d\mathbb{P}|_{\mathfrak{F}} = \int_A Y \, d\mathbb{P}, \text{ mit } A \in \mathfrak{F}.$$

Diese Zufallsgröße besitzt die gewünschten Eigenschaften:

$$\mathbb{E}Y = Q(\Omega) = \mathbb{E}X < \infty, \text{ das heißt } Y \in \mathcal{L}^1(\mathbb{P}).$$

Die $\mathfrak{F}$-Messbarkeit von Y gilt nach Definition von Y. Um (iii) zu zeigen, sei $A \in \mathfrak{F}$. Dann folgt

$$\mathbb{E}(\mathbb{1}_A Y) = \int_A Y \, d\mathbb{P} = Q(A) = \mathbb{E}(\mathbb{1}_A X).$$

Der allgemeine Fall $X \in \mathcal{L}^1(\mathbb{P})$ folgt hieraus durch die Zerlegung von X mit $X = X^+ - X^-$.

b) $\mathbb{P}$*-f.s. Eindeutigkeit von Y*: Sind Y_1 und Y_2 zwei Zufallsgrößen, welche die Eigenschaften (i)-(iii) erfüllen, so ist $Y := Y_1 - Y_2$ $\mathfrak{F}$-messbar und $Y \in \mathcal{L}^1(\mathbb{P})$ und wegen (iii) gilt für alle $A \in \mathfrak{F}$:

$$\mathbb{E}(\mathbb{1}_A Y) = 0.$$

Insbesondere gilt für jedes $n \in \mathbb{N}$:

$$0 = \mathbb{E}\mathbb{1}_{\{Y > \frac{1}{n}\}} Y \geq \frac{1}{n}\,\mathbb{P}(Y > \frac{1}{n}),$$

das bedeutet, dass $\{Y > \frac{1}{n}\}$ eine $\mathbb{P}$-Nullmenge ist. Dann sind auch $\{Y > 0\} = \bigcup_n \{Y > \frac{1}{n}\}$ sowie $\{Y < 0\}$ $\mathbb{P}$-Nullmengem, und das heißt $Y = 0$ fast sicher.

□

Definition 3.4. Die fast sicher eindeutig bestimmte Zufallsgröße Y aus dem Satz 3.3 heißt *bedingte Erwartung von X gegeben $\mathfrak{F}$* und wird mit $\mathbb{E}[X|\mathfrak{F}\,]$ bezeichnet. Für $A \in \mathfrak{A}$ heißt

$$\mathbb{P}[A|\mathfrak{F}\,] := \mathbb{E}[\mathbb{1}_A|\mathfrak{F}\,]$$

bedingte Wahrscheinlichkeit von A gegeben $\mathfrak{F}$.

Der bedingte Erwartungswert $\mathbb{E}[X|\mathfrak{F}\,]$ wird somit durch folgende Eigenschaften charakterisiert:

a) $\mathbb{E}[X|\mathfrak{F}\,] \in \mathcal{L}^1(\mathbb{P})$;

b) $\mathbb{E}[X|\mathfrak{F}\,]$ ist $\mathfrak{F}$-messbar;

c) für alle $A \in \mathfrak{F}$ gilt
$$\mathbb{E}(\mathbb{1}_A \mathbb{E}[X|\mathfrak{F}\,]) = \mathbb{E}(\mathbb{1}_A X).$$

Bemerkung 3.5. Mit der „üblichen Approximationsprozedur“ zeigt man, dass die folgende Eigenschaft äquivalent zu c) ist:

c) Für jede $\mathfrak{F}$-messbare Zufallsgröße Y mit $X \cdot Y \in \mathcal{L}^1(\mathbb{P})$ gilt $Y \cdot \mathbb{E}[X|\mathfrak{F}\,] \in \mathcal{L}^1(\mathbb{P})$ und
$$\mathbb{E}(Y \cdot \mathbb{E}[X|\mathfrak{F}\,]) = \mathbb{E}(Y \cdot X).$$

(Übungsaufgabe; man verwende zusätzlich die Eigenschaften (i) und (v) aus dem nachstehenden Satz.)

Satz 3.6. *(Fundamentaleigenschaften bedingter Erwartungen) Gegeben seien die Zufallsgrößen X, $Y \in \mathcal{L}^1(\mathbb{P})$ sowie die Teil-σ-Algebren $\mathfrak{F}$ und $\mathfrak{Y}$ von $\mathfrak{A}$. Dann gilt:*

(i) $\mathbb{E}[\alpha X + \beta Y|\mathfrak{F}\,] = \alpha\mathbb{E}[X|\mathfrak{F}\,] + \beta\mathbb{E}[Y|\mathfrak{F}\,]$ *fast sicher (für α, $\beta \in \mathbb{R}$) (Linearität)*

(ii) Aus $X \geq Y$ fast sicher folgt $\mathbb{E}[X|\mathfrak{F}\,] \geq \mathbb{E}[Y|\mathfrak{F}\,]$ *fast sicher (Monotonie)*

(*iii*) *Ist* Y $\mathfrak{F}$*-messbar, so gilt* $\mathbb{E}[Y|\mathfrak{F}] = Y$ *f.s. . Ist zusätzlich* $X \cdot Y \in \mathcal{L}^1(\mathbb{P})$*, so gilt weiter*

$$\mathbb{E}[X \cdot Y|\mathfrak{F}] = Y \cdot \mathbb{E}[X|\mathfrak{F}] \text{ fast sicher} \tag{3.2}$$

(Messbare Faktoren)

(*iv*) *Für* $\mathfrak{Y} \subseteq \mathfrak{F}$ *gilt*

$$\mathbb{E}[\,\mathbb{E}\,[X|\mathfrak{F}]\,|\mathfrak{Y}] = \mathbb{E}[[X|\mathfrak{Y}]|\mathfrak{F}] = \mathbb{E}[X|\mathfrak{Y}] \text{ fast sicher}$$

(Turmeigenschaft)

(*v*) $|\,\mathbb{E}[X|\mathfrak{F}]\,| \leq \mathbb{E}\,[\,|X|\,|\mathfrak{F}]$ *fast sicher*
(Dreiecksungleichung)

(*vi*) *Sind* $\sigma(X)$ *und* $\mathfrak{F}$ *unabhängig, so gilt* $\mathbb{E}[X|\mathfrak{F}] = \mathbb{E}X$ *fast sicher*

(*vii*) *Sind* X_n *und* X *Zufallsgrößen mit* $X_n \to X$ *fast sicher und* $|X_n| \leq Y$ *fast sicher für ein* $Y \in \mathcal{L}^1(\mathbb{P})$*, so gilt*

$$\mathbb{E}[X_n|\mathfrak{F}] \to \mathbb{E}[X|\mathfrak{F}] \text{ fast sicher und } \mathbb{E}[X|\mathfrak{F}] \in \mathcal{L}^1(\mathbb{P})$$

(Majorisierte Konvergenz)

Beweis.

(*i*) $Z := \alpha\mathbb{E}[X|\mathfrak{F}] + \beta\mathbb{E}[Y|\mathfrak{F}]$ gehört zum $\mathcal{L}^1(\mathbb{P})$ und ist $\mathfrak{F}$-messbar. Außerdem gilt für jedes $A \in \mathfrak{F}$:

$$\begin{aligned} \mathbb{E}(\mathbb{1}_A Z) &= \alpha\mathbb{E}[\mathbb{1}_A\,\mathbb{E}[X|\mathfrak{F}]\,] + \beta\mathbb{E}[\mathbb{1}_A\,\mathbb{E}[X|\mathfrak{F}]\,] \\ &= \alpha\mathbb{E}(\mathbb{1}_A X) + \beta(\mathbb{1}_A Y) = \mathbb{E}(\mathbb{1}_A\,(\alpha X + \beta Y)\,). \end{aligned}$$

Folglich ist $Z = \mathbb{E}[\alpha X + \beta Y|\mathfrak{F}]$ f.s. (Nach Satz 3.3 und Definition 3.4).

(*ii*) Nach Vorrausetzung ist $Z := \mathbb{E}[X|\mathfrak{F}] - \mathbb{E}[Y|\mathfrak{F}] \underset{(i)}{=} \mathbb{E}[X - Y|\mathfrak{F}]$ $\mathfrak{F}$-messbar und

$$\mathbb{E}(\mathbb{1}_A Z) = \mathbb{E}\mathbb{1}_A(X - Y) \geq 0, \text{ für alle } A \in \mathfrak{F}.$$

Hieraus folgt:

$$\begin{aligned} &Z \geq 0 \text{ fast sicher (Definiere } A_n := \{Z < -\frac{1}{n}\} \in \mathfrak{F}), \\ &0 \leq \mathbb{E}(\mathbb{1}_{A_n} Z) \leq -\frac{1}{n}\,\mathbb{P}(A_n) \Rightarrow \mathbb{P}(A_n) = 0, \text{ für alle } n \\ &\Rightarrow \{Z < 0\} = \bigcup_n A_n \text{ ist eine } \mathbb{P}\text{-Nullmenge.} \end{aligned}$$

(*iii*) Nach Definiton des bedingten Erwartungswertes gilt die Eigenschaft (3.2) für $Y = \mathbb{1}_B$, mit $B \in \mathfrak{F}$:
$Z := \mathbb{1}_B \mathbb{E}[X|\mathfrak{F}]$ ist $\mathfrak{F}$-messbar und $\in \mathcal{L}^1(\mathbb{P})$. Für beliebige $A \in \mathfrak{F}$ gilt:
$\mathbb{E}(\mathbb{1}_A Z) = \mathbb{E}(\mathbb{1}_{\underbrace{A \cap B}_{\in \mathfrak{F}}} \mathbb{E}[X|\mathfrak{F}]) = \mathbb{E}(\mathbb{1}_{A \cap B} \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot (\mathbb{1}_B X))$,
das heißt $Z = \mathbb{E}[\mathbb{1}_B \cdot X]$ fast sicher. Im allgemeinen Fall ist $Z := Y\,\mathbb{E}[X|\mathfrak{F}]$ $\mathfrak{F}$-messbar und nach Bemerkung 3.5 im $\mathcal{L}^1(\mathbb{P})$.
Für jedes $A \in \mathfrak{F}$ ist deshalb

$$\mathbb{E}(\mathbb{1}_A Z) \underset{\text{(3.2) für „}Y = \mathbb{1}_A\text{“}}{=} \mathbb{E}(Y \cdot \mathbb{E}[\mathbb{1}_A \cdot X|\mathfrak{F}] \underset{\text{Bem. 3.5}}{=} \mathbb{E}(\mathbb{1}_A(X \cdot Y)),$$

und das heißt, dass $Z = \mathbb{E}[XY|\mathfrak{F}]$ fast sicher

(*iv*) Da $\mathbb{E}[X|\mathfrak{Y}]$ $\mathfrak{Y}$-messbar ist, ist $\mathbb{E}[X|\mathfrak{F}]$ auch $\mathfrak{F}$-messbar (Man beachte dabei $\mathfrak{Y} \subseteq \mathfrak{F}$). Deshalb folgt aus (*iii*) sofort

$$\mathbb{E}[\,\mathbb{E}[X|\mathfrak{F}]|\mathfrak{F}] = \mathbb{E}[X|\mathfrak{F}] \text{ fast sicher}$$

Für jedes $A \in \mathfrak{Y}$ ist auch $A \in \mathfrak{F}$ und deshalb

$$\mathbb{E}(\mathbb{1}_A\, \mathbb{E}[X|\mathfrak{F}]) \underset{A \in \mathfrak{Y}}{=} \mathbb{E}(\mathbb{1}_A \cdot X) \underset{A \in \mathfrak{F}}{=} \mathbb{E}(\mathbb{1}_A\, \mathbb{E}[X|\mathfrak{F}])$$

Da $\mathbb{E}[X|\mathfrak{Y}]$ zudem $\mathfrak{Y}$-messbar und $\in \mathcal{L}^1(\mathbb{P})$ ist, fällt somit $\mathbb{E}[X|\mathfrak{Y}]$ fast sicher mit dem bedingten Erwartungswert der Zufallsgröße $\mathbb{E}[X|\mathfrak{F}]$ gegeben $\mathfrak{Y}$ zusammen.

(*v*) Mit $X = X^+ - X^-$ und mit $X^\pm \leq |X|$ folgt die Behauptung aus der Linearität und Monotonie des bedingten Erwartungswertes.

(*vi*) $Z := \mathbb{E}X$ ist trivialerweise $\mathfrak{F}$-messbar und $\in \mathcal{L}^1$. Für jedes $A \in \mathfrak{F}$ sind $\mathbb{1}_A$ und X unabhängig und folglich unkorreliert, weshalb

$$\mathbb{E}(\mathbb{1}_A \cdot Z) = \mathbb{E}\mathbb{1}_A \cdot \mathbb{E}X = \mathbb{E}(\mathbb{1}_A \cdot X)$$

gilt. Also ist $Z = \mathbb{E}[X|\mathfrak{F}]$ fast sicher

(*vii*) $\mathcal{L}^1$-Konvergenz:
$\mathbb{E}\,|\mathbb{E}[X_n|\mathfrak{F}] - \mathbb{E}[X|\mathfrak{F}]| \underset{(i),(ii)}{\leq} \mathbb{E}\,\mathbb{E}[|X_n - X|\,|\mathfrak{F}] = \mathbb{E}|X_n - X| \underset{n \to \infty}{\longrightarrow} 0$ nach dem Satz von Lebesgue über die majorisierte Konvergenz
(Man beachte $|X_n - X| \to 0$ fast sicher und $|X_n - X| \leq 2Y$ fast sicher).

fast sichere Konvergenz:
$Y_n := \sup_{k \geq n} |X_k - X| \to 0$ f.s. und $0 \leq Y_n \leq 2Y$.
$\mathbb{E}[Y_n|\mathfrak{F}] \downarrow \bar{Y}$ für ein $\bar{Y}$ und $|\,\mathbb{E}[Y_n|\mathfrak{F}]| \leq 2\mathbb{E}[|Y||\mathfrak{F}] \leq 2\mathbb{E}[|Y|] \leq \infty$
Nach dem Lemma von Fatou und dem Satz von Lebesgue gilt

$\mathbb{E}Y \leq \liminf\limits_{n\to\infty} \underbrace{\mathbb{E}\,[\mathbb{E}[Z_n|\mathfrak{F}\,]]}_{=\mathbb{E}Z_n} = 0$, das heißt $Y = 0$ fast sicher.
Hieraus folgt wegen

$$|\,\mathbb{E}[X_n|\mathfrak{F}\,] - \mathbb{E}[X|\mathfrak{F}\,]\,| \leq \mathbb{E}[Y_n|\mathfrak{F}\,]$$

die Behauptung.

□

Satz 3.7. *(Jensensche Ungleichung) Gegeben sei eine Zufallsgröße $X \in \mathcal{L}^1(\mathbb{P})$ sowie eine Teil-σ-Algebra $\mathfrak{F}$ von $\mathfrak{A}$. Dann gilt für jede konvexe Funktion φ mit $\varphi(X) \in \mathcal{L}^1(\mathbb{P})$ die Ungleichung*

$$\varphi(\,\mathbb{E}[X|\mathfrak{F}\,]\,) \leq \mathbb{E}[\varphi(X)|\mathfrak{F}\,] \quad \mathbb{P}\text{-fast sicher}$$

Beweis. Übungsaufgabe. □

Beispiel 3.8. (Optimale Approximation im quadratischen Mittel)
Seien $X \in \mathcal{L}^2(\mathbb{P})$ und $\mathfrak{F}$ eine Teil-σ-Algebra von $\mathfrak{A}$ (welche die uns zur Verfügung stehende „Information“ enthält).
Aufgabe: Man approximiere X möglichst genau durch eine $\mathfrak{F}$-messbare Zufallsgröse Y (deren Wert $Y(\omega)$ sich mit der vorhandenen „Information“ aus $\omega \in \Omega$ gewinnen lässt).
Genauer: Man löse das Variationsproblem

$$I := \inf\left\{\mathbb{E}(X-Y)^2 : Y\ \mathfrak{F}\text{-messbar und } Y \in \mathcal{L}^2(\mathbb{P})\right\}.$$

Lösung: Dieses Infimum wird für $Y = \mathbb{E}[X|\mathfrak{F}\,]$ angenommen und dieser Minimierer ist $\mathbb{P}$-f.s. eindeutig bestimmt. Außerdem ist $I = \mathbb{E}X^2 - \mathbb{E}\,(\mathbb{E}[X|\mathfrak{F}\,])^2$.

Beweis. Sei Y $\mathfrak{F}$-messbar und $Y \in \mathcal{L}^2(\mathbb{P})$. Dann ist

$$\mathbb{E}(X - \mathbb{E}[X|\mathfrak{F}\,])(Y - \mathbb{E}[X|\mathfrak{F}\,]) = 0. \tag{3.3}$$

Tatsächlich, mit den Rechenregeln für bedingte Erwartungen folgt

$$\begin{aligned}\mathbb{E}(X - \mathbb{E}[X|\mathfrak{F}\,])(Y - \mathbb{E}[X|\mathfrak{F}\,]) &= \mathbb{E}\,[\mathbb{E}(X - \mathbb{E}[X|\mathfrak{F}\,])(Y - \mathbb{E}[X|\mathfrak{F}\,])|\mathfrak{F}\,] \\ &\underset{\substack{Y-\mathbb{E}[X|\mathfrak{F}\,] \\ \mathfrak{F}\text{-messb.}}}{=} \mathbb{E}((Y - \mathbb{E}[X|\mathfrak{F}\,]) \cdot \underbrace{\mathbb{E}[X - \mathbb{E}[X|\mathfrak{F}\,]|\mathfrak{F}\,]}_{=\mathbb{E}[X|\mathfrak{F}\,]-\mathbb{E}[X|\mathfrak{F}\,]=0}) = 0.\end{aligned}$$

Hiermit erhält man

$$\begin{aligned}\mathbb{E}(X-Y)^2 &= \mathbb{E}\,(\,(X - \mathbb{E}[X|\mathfrak{F}\,]) - (Y - \mathbb{E}[X|\mathfrak{F}\,]\,)\,)^2 \\ &\underset{(3.2)}{=} \mathbb{E}(X - \mathbb{E}[X|\mathfrak{F}\,])^2 + \mathbb{E}(Y - \mathbb{E}[X|\mathfrak{F}\,])^2 \\ &\geq \mathbb{E}(X - \mathbb{E}[X|\mathfrak{F}\,])^2,\end{aligned}$$

wobei Gleichheit genau dann gilt, wenn $Y = \mathbb{E}[X|\mathfrak{F}\,]$ fast sicher gilt. Dies ist der gesuchte Minimierer, da $\mathbb{E}[X|\mathfrak{F}\,]$ $\mathfrak{F}$-messbar ist und im $\mathcal{L}^2(\mathbb{P})$ liegt. Letzeres folgt mit der Jensenschen Ungleichung (Satz 3.7) für die konvexe Funktion $\varphi(x) = x^2$:

$$\mathbb{E}(\mathbb{E}[X|\mathfrak{F}\,])^2 \leq \mathbb{E}\,\mathbb{E}[X^2|\mathfrak{F}\,] = \mathbb{E}X^2 < \infty.$$

Außerdem ist da $\mathbb{E}(X \cdot \mathbb{E}[X|\mathfrak{F}\,]) = \mathbb{E}\,\mathbb{E}[X \cdot \underbrace{\mathbb{E}[X|\mathfrak{F}\,]}_{\mathfrak{F}\text{-messb.}}|\mathfrak{F}\,] = \mathbb{E}(\mathbb{E}[X|\mathfrak{F}\,])^2$

$$\begin{aligned} I &= \mathbb{E}(X - \mathbb{E}[X|\mathfrak{F}\,])^2 = \mathbb{E}X^2 - 2\mathbb{E}(X \cdot \mathbb{E}[X|\mathfrak{F}\,]) + \mathbb{E}(\mathbb{E}[X|\mathfrak{F}\,])^2 \\ &= \mathbb{E}X^2 - \mathbb{E}(\mathbb{E}[X|\mathfrak{F}\,])^2. \end{aligned}$$

□

Definition 3.9. Seien $X \in \mathcal{L}^1(\mathbb{P})$ und Y eine Zufallsvariable mit Werten in einem beliebigen messbaren Raum. Dann definiert man den *bedingten Erwartungswert von X gegeben Y* durch

$$\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)].$$

Analog definiert man für $A \in \Omega$ die *bedingte Wahrscheinlichkeit von A gegeben Y* durch

$$\mathbb{P}(A|Y) := \mathbb{P}(A|\sigma(Y)).$$

Bemerkung 3.10. Für mehrere Zufallsvariablen $Y_1, \dots, Y_n$ definiert man entsprechend

$$\mathbb{E}[X|Y_1, \dots, Y_n] = \mathbb{E}[X|\sigma(Y_1, \dots, Y_n)].$$

Lemma 3.11. *(Faktorisierungslemma) Sei Y eine Zufallsvariable mit Werten in einem beliebigen messbaren Raum $(E, \mathfrak{E})$. Dann existiert zu jeder $\sigma(Y)$-messbaren Zufallgröße Z eine $\mathbb{P}_Y$-fast sicher eindeutig bestimmte messbare Funktion $\varphi : E \to \mathbb{R}$ mit*

$$Z = \varphi \circ Y.$$

Beweis.
1^o *Existenz von φ:*
Ist $Z = \sum\limits_{k=1}^{r} \alpha_k \mathbb{1}_{B_k}$ elementar, so sind $B_1, \dots, B_r \in \sigma(Y) = \{Y^{-1}(C) : C \in \mathfrak{E}\}$. Folglich existieren $C_1, \dots, C_r \in \mathfrak{E}$ mit $B_k = Y^1(C_k)$ mit $(k = 1, \dots, r)$, und die Behauptung ist für

$$\varphi = \sum_{k=1}^{r} \alpha_k \mathbb{1}_{C_k}$$

erfüllt. Tatsächlich, φ ist $\mathfrak{E}$-$\mathfrak{B}$-messbar und es gilt

$$\varphi \circ Y = \sum_{k=1}^{r} \alpha_k (\mathbb{1}_{C_k} \circ Y) = \sum_{k=1}^{r} \alpha_k \mathbb{1}_{Y^{-1}(C_k)} = Z.$$

Der allgemeine Fall folgt mit der „üblichen“ Approximationsprozedur.

2^o $\mathbb{P}_Y$*-fast sichere Eindeutigkeit von* φ (wobei $\mathbb{P}_Y = \mathbb{P} \circ Y^{-1}$):
Seien φ_1 und φ_2 zwei Funktionen mit den gewünschten Eigenschaften und sei $\psi := \varphi_1 - \varphi_2$. Dann ist $|\psi| \circ Y = |\psi \circ Y| = 0$ und somit

$$0 = \int_\Omega |\psi| \circ Y \, d\mathbb{P} \text{ und mit dem Integraltransformationssatz } = \int_E |\psi| \, d\mathbb{P}_Y.$$

Hieraus folgt $\psi = 0$ $\mathbb{P}_Y$-fast sicher und somit $\varphi_1 = \varphi_2$ $\mathbb{P}_Y$-fast sicher. □

Im Folgenden seien $X \in \mathcal{L}^1(\mathbb{P})$ und Y eine $(E, \mathfrak{E})$-wertige Zufallsvariable. Dann ist $\mathbb{E}[X|Y]$ eine $\sigma(Y)$-messbare Zufallsgröße. Nach dem Faktorisierungslemma existiert eine $\mathbb{P}_Y$-fast sicher eindeutig bestimmte messbare Funktion $\varphi : E \to \mathbb{R}$ mit

$$\mathbb{E}[X|Y] = \varphi \circ Y.$$

Definition 3.12. Für $y \in E$ wird $\varphi(y)$ mit $\mathbb{E}[X|Y = y]$ bezeichnet und *bedingter Erwartungswert von* X *gegeben* $Y = y$ *genannt*. Entsprechend definiert man $\mathbb{P}(A|Y = y)$.
$\mathbb{E}[X|Y = \cdot]$ ist somit eine auf E definierte messbare Funktion mit

$$\mathbb{E}[X|Y = \cdot] \circ Y = \mathbb{E}[X|Y] \ \ \mathbb{P}\text{-fast sicher}$$

Durch die Eigenschaften ist $\mathbb{E}[X|Y = \cdot]$ $\mathbb{P}_Y$-fast sicher eindeutig festgelegt.

3.3. Reguläre bedingte Wahrscheinlichkeiten

Sei $\mathfrak{F}$ eine beliebige Teil-σ-Algebra von $\mathfrak{A}$.

Frage: Kann man für jedes $A \in \mathfrak{A}$ eine Version $\mathbb{P}(A|\mathfrak{F})$ der bedingeten Wahrscheinlichkeit von A gegeben $\mathfrak{F}$ finden, so dass $\mathbb{P}(\,\cdot\,|\mathfrak{F})(\omega)$ ein Wahrscheinlichkeitsmaß ist für $\mathbb{P}$-fast alle $\omega \in \Omega$.

Zunächst kann man mit majorisierter Konvergenz (Satz 3.6 (vii)) zeigen, dass eine Art „fast sichere σ-Additivität“ gilt. Für jede Folge (A_n) paarweise disjunkter Mengen aus $\mathfrak{A}$ ist

$$\mathbb{P}\left(\bigcup_n A_n|\mathfrak{F}\right) = \sum_n \mathbb{P}\left(A_n|\mathfrak{F}\right) \ \ \mathbb{P}\text{-fast sicher} \tag{3.4}$$

Im Allgemeinen existieren jedoch *überabzählbar* viele solcher Folgen (A_n), deren Ausnahme-Nullmengen in (3.4) voneinander verschieden sein können. Die Vereinigung dieser Ausnahmemengen braucht deshalb im Allgemeinen keine Nullmenge mehr zu sein.
Die gestellte Frage lässt sich dennoch in großer Allgemeinheit positiv beantworten.

Definition 3.13. Gegeben seien messbare Räume $(E, \mathfrak{E})$ sowie $(F, \mathfrak{F})$. Eine Abbildung
$$K : E \times \mathfrak{F} \to \mathbb{R}$$
heißt *Markov-Kern von* $(E, \mathfrak{E})$ *nach* $(F, \mathfrak{F})$, wenn folgendes gilt:

a) $\forall A \in \mathfrak{F}: \quad K(\cdot, A)$ ist eine $\mathfrak{E}$-$\mathfrak{B}$-messbare Funktion auf E;

b) $\forall x \in E: \quad K(x, \cdot)$ ist ein Wahrscheinlichkeitsmaß auf $(F, \mathfrak{F})$.

Satz 3.14. *Sei Ω ein vollständiger separabler metrischer Raum („Polnischer Raum“) versehen mit der Borel-σ-Algebra $\mathfrak{B}_\Omega$. Weiterhin seien $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{B}_\Omega)$ und $\mathfrak{F}$ eine Teil-σ-Algebra von $\mathfrak{B}_\Omega$. Dann existiert ein Markov-Kern K von $(\Omega, \mathfrak{F})$ nach $(\Omega, \mathfrak{B}_\Omega)$ mit*

$$\int_B K(\omega, A)\, \mathbb{P}(d\omega) = \mathbb{P}(A \cap B) \quad \textit{(für } A \in \mathfrak{B}_\Omega \textit{ und } B \in \mathfrak{F}\textit{).} \tag{3.5}$$

Beweis. Siehe zum Beispiel Bauer, Wahrscheinlichkeitstheorie, Satz 44.3. □

Definition 3.15. Sei $\mathfrak{F}$ eine Teil-σ-Algebra von $\mathfrak{A}$.

a) $\mathbb{P}(\,\cdot\,|\mathfrak{F}\,)$ heißt *bedingte Wahrscheinlichkeit gegeben* $\mathfrak{F}$, wenn gilt:

   (i) $\forall A \in \mathfrak{A}: \quad \mathbb{P}(A|\mathfrak{F}\,)$ ist eine Version der bedingten Wahrscheinlichkeit von A gegeben $\mathfrak{F}$;

   (ii) $\forall \omega \in \Omega: \quad \mathbb{P}(\,\cdot\,|\mathfrak{F}\,)(\omega)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{A})$.

b) Sei weiter Y eine Zufallsvariable mit Werten in einem messbaren Raum $(E, \mathfrak{E})$.
   $\mathbb{P}(\,\cdot\,|Y = y)$ heißt *reguläre bedingte Wahrscheinlichkeit gegeben* $Y = y$ (mit $y \in E$), falls gilt:

   (i) $\forall A \in \mathfrak{A}: \quad \mathbb{P}(A|Y = y)$ ist eine Version der bedingten Wahrscheinlichkeit von A gegeben $Y = y$;

   (ii) $\forall y \in E: \quad \mathbb{P}(\,\cdot\,|Y = y)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{A})$.

Die zugehörigen *regulären bedingten Erwartungswerte* definiert man als Integrale bezüglich dieser regulären bedingten Wahrscheinlichkeiten: Für $X \in \mathcal{L}^1(\mathbb{P})$ setzt man

$$\begin{aligned} \mathbb{E}[Y|\mathfrak{F}\,](\omega) \quad &:= \quad \int_\Omega X(\tilde{\omega})\, \mathbb{P}(d\tilde{\omega}|\mathfrak{F}\,)(\omega), \quad \text{für } \omega \in \Omega, \\ \text{und} \quad \mathbb{E}[X|Y = y](\omega) \quad &:= \quad \int_\Omega X(\tilde{\omega})\, \mathbb{P}(d\tilde{\omega}|Y = y), \quad \text{für } y \in E. \end{aligned}$$

Man prüft leicht nach, dass dies tatsächlich (spezielle) Versionen bedingter Erwartungswerte sind.

Satz 3.16.

(i) *Seien Ω ein Polnischer Raum, $\mathfrak{B}_\Omega$ die Borel-σ-Algebra in Ω und $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{B}_\Omega)$. Dann existiert zu jeder Teil-σ-Algebra $\mathfrak{F}$ von $\mathfrak{B}_\Omega$ eine reguläre bedingte Wahrscheinlichkeit $\mathbb{P}(\,\cdot\,|\mathfrak{F})$.*

(ii) *Sei zusätzlich Y eine Zufallsvariable auf $(\Omega, \mathfrak{B}_\Omega, \mathbb{P})$ mit Werten in einem Polnischen Raum $(E, \mathfrak{B}_E)$. Dann existiert eine reguläre bedingte Wahrscheinlichkeit $\mathbb{P}(\,\cdot\,|Y = \,)$.*

Beweis.

(i) Der Markov-Kern K aus Satz 3.14 besitzt die gewünschten Eigenschaften: Für alle $A \in \mathfrak{A}$ ist $K(\cdot, A)$ $\mathfrak{F}$-messbar und im $\mathcal{L}^1$. (3.5) ist dann gleichbedeutend damit, dass für alle $B \in \mathfrak{F}$ gilt

$$\mathbb{E}\,\mathbb{1}_B K(\cdot, A) = \mathbb{E}\,\mathbb{1}_B \cdot \mathbb{1}_A.$$

Also ist $K(\,\cdot\,|A)$ eine Version der bedingten Wahrscheinlichkeit von A gegeben $\mathfrak{F}$. Ausserdem ist $K(\omega, \cdot)$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{B}_\Omega)$ für alle $\omega \in \Omega$.

(ii) *Beweisskizze:*
$\widehat{\Omega} := \Omega \times E$ ist polnisch mit $\mathfrak{B}_{\widehat{\Omega}} = \mathfrak{B}_\Omega \otimes \mathfrak{B}_E$.
Sei nun $\pi_2 : \widehat{\Omega} \to E$ mit $\pi_2(\omega, y) = y$ die Projektion auf die zweite Koordinate und sei $\mathbb{Q}$ das Bild des Maßes $\mathbb{P}$ bezüglich der messbaren Abbildung $\Omega \ni \omega \mapsto (\omega, Y(\omega)) \in \widehat{\Omega}$. Dies ist ein Wahrscheinlichkeitsmaß auf $(\widehat{\Omega}, \mathfrak{B}_{\widehat{\Omega}})$. Nach Behauptung a) existiert eine reguläre bedingte Wahrscheinlichkeit $\mathbb{Q}[\,\cdot\,|\pi_2]$. Für jedes $\widehat{A} \in \mathfrak{B}_{\widehat{\Omega}}$ ist $(\omega, y) \mapsto \mathbb{Q}[\widehat{A}|\pi_2](\omega, y)$ $\sigma(\pi_2) = \{\Omega \times B : B \in \mathfrak{B}_E\}$-messbar und hängt folglich nicht von ω ab (selbständig zu untersuchen ist hierbei, warum dies gilt).
Wir definieren

$$K(y, A) := \mathbb{Q}(A \times E|\pi_2)(\omega, y) \quad (\text{mit } y \in E \text{ und } A \in \mathfrak{B}_\Omega)$$

und zeigen, dass dies die gewünschte reguläre bedingte Wahrcheinlichkeit ist.
$K(y, A)$ ist eine $\sigma(\pi_2)$-messbare Funktion von (ω, y) und damit $\mathfrak{B}_E$-messbar als Funktion von y. Ausserdem ist $K(y, \cdot)$ ein Wahrscheinlichkeitsmaß für jedes $y \in E$. Also ist die Zufallsgröße $K(Y, A)$ $\sigma(Y)$-messbar und im $\mathcal{L}^1(\mathbb{P})$. Zu jedem $B \in \sigma(Y)$ existert ein $C \in \mathfrak{B}_\Omega$ mit $B = Y^{-1}(C)$ und mit Hilfe des Integraltransformationssatzes $(*)$ sowie der Definition der bedingten

Wahrscheinlichkeit $(**)$ gilt

$$
\begin{aligned}
\mathbb{E}\mathbb{1}_B K(Y,A) &= \int \mathbb{1}_{\Omega\times C}(\omega, Y(\omega))\,\mathbb{Q}[A\times E|\pi_2](\omega, Y(\omega))\,\mathbb{P}(d\omega) \\
&\underset{(*)}{=} \int_{\widehat{\Omega}} \mathbb{1}_{\Omega\times C}(\hat\omega)\,\mathbb{Q}[A\times E|\pi_2](\hat\omega)\,\mathbb{Q}(d\omega) \qquad (\Omega\times C\in\sigma(\pi_2)) \\
&\underset{(**)}{=} \mathbb{Q}((A\times E)\cap(\Omega\times C)) \\
&= \mathbb{Q}(A\times C) \underset{\text{Def. von }\mathbb{Q}}{=} \mathbb{P}(A\cap Y^{-1}(C)) \\
&= \mathbb{P}(A\cap B) \qquad (\text{für } A\in\mathfrak{B}_\Omega \text{ und } B\in\sigma(Y)).
\end{aligned}
$$

Also ist $K(\cdot,A)\circ Y$ eine Version der bedingten Wahrscheinlichkeit von A gegeben Y (mit $A\in\mathfrak{B}_\Omega$).

□

Satz 3.17. *Seien X und Y Zufallsvariablen mit Werten in einem messbaren Raum $(E,\mathfrak{E})$ bzw. $(F,\mathfrak{F})$. Weiterhin sei $H: E\times F\to\mathbb{R}$ eine $\mathfrak{E}\otimes\mathfrak{F}$-$\mathfrak{B}$-messbare und beschränkte Funktion. Dann gilt für den regulären bedingten Erwartungswert $\mathbb{E}[\,\cdot\,|Y=\,]$ bzw. $\mathbb{E}[\,\cdot\,|X=\,]$ (falls dieser existiert)*

$$\mathbb{E}[H(X,Y)|Y=y] = \mathbb{E}[H(X,y)|Y=y] \text{ für } \mathbb{P}_Y\text{-f.a. } y\in F \tag{3.6}$$

und

$$\mathbb{E}[H(X,Y)|Y] = (\mathbb{E}[H(X,y)|Y])_{y=Y} \;\; \mathbb{P}\text{-f.s.} \tag{3.7}$$

Beweisidee für (3.7).

1^o Die Aussage gilt für $H=\mathbb{1}_{A\times B}$ (für $A\in\mathfrak{E}$ und $B\in\mathfrak{F}$): $\mathbb{1}_B(Y)\,\mathbb{E}[\mathbb{1}_A(X)|Y] = \mathbb{1}_B(Y)\,\mathbb{P}[X^{-1}(A)|Y]$ ist eine Version des bedingten Erwartungswertes von $\mathbb{1}_{A\times B}(X,Y)=\mathbb{1}_A(X)\cdot\mathbb{1}_B(Y)$ gegeben Y. Da $\mathbb{P}[\,\cdot\,|Y]$ eine reguläre bedingte Wahrscheinlichkeit ist, gilt für alle y,ω

$$\mathbb{1}_B(y)\,\mathbb{P}[X^{-1}(A)|Y](\omega) = \mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(y)|Y](\omega)$$

und folglich

$$\mathbb{E}[\mathbb{1}_{A\times B}(A,Y)|Y] \overset{\text{f.s.}}{=} (\mathbb{E}[\mathbb{1}_{A\times B}(X,y)|Y])_{y=Y}\,.$$

2^o Die Menge $\vartheta := \{C\in\mathfrak{E}\otimes\mathfrak{F}:\ (3.7) \text{ gilt für } H=\mathbb{1}_C\}$ ist ein Dynkin-System und wegen 1^0 folgt mit dem Hauptsatz über Dynkin-Systeme, dass $\vartheta=\mathfrak{E}\otimes\mathfrak{F}$ ist.

3^o Der allgemeine Fall folgt mit der „üblichen“ Approximation.

□

Bemerkung 3.18. Die Beschränktheit von F kann durch $F(X,Y) \in \mathcal{L}^1(\mathbb{P})$ ersetzt werden. Dann existiert $\mathbb{E}H(X,y)$ im Allgemeinen nur für $\mathbb{P}_Y$-fast alle $y \in F$ und $\mathbb{E}[H(X,y)|Y]$ muss auf einer Nullmenge durch 0 ergänzt werden (wie beim Satz von Fubini).

Satz 3.19. *Seien X und Y unabhängige Zufallsvariablen mit Werten in einem messbaren Raum $(E,\mathfrak{E})$ bzw. $(F,\mathfrak{F})$. Weiterhin sei $H(X,Y) \in \mathcal{L}^1(\mathbb{P})$. Dann gilt*

$$\mathbb{E}[H(X,Y)|Y=y] = \mathbb{E}[H(X,y)] \text{ für } \mathbb{P}_Y\text{-fast alle } y \in F \tag{3.8}$$

und

$$\mathbb{E}[H(X,Y)|Y] = \left(\mathbb{E}[H(X,y)]\right)_{y=Y} \quad \mathbb{P}\text{-f.s} \tag{3.9}$$

Beweisidee für (3.9).
Für $H = \mathbb{1}_{A\times B}$ (mit $A \in \mathfrak{E}$ und $B \in \mathfrak{F}$) gilt

$$\begin{aligned}\mathbb{E}[H(X,Y)|Y] &\overset{\text{f.s.}}{=} \mathbb{1}_B(Y)\,\mathbb{E}[\mathbb{1}_A(X)|Y] \underset{\text{Unabh.}}{\overset{\text{f.s.}}{=}} \mathbb{1}_B(Y)\,\mathbb{E}\mathbb{1}_A(X) \\ &= \left(\mathbb{E}(\mathbb{1}_A(X)\mathbb{1}_B(y))\right)_{y=Y} = \left(\mathbb{E}H(X,y)\right)_{y=Y}.\end{aligned}$$

Die Verallgemeinerung auf „beliebige“ H geschieht analog zum Beweis des vorherigen Satzes. $\square$

Kapitel 4

Martingale in diskreter Zeit

4.1. Filtration und Stoppzeiten

Sei $(\Omega, \mathfrak{F}, \mathbb{P})$ für diesen Abschnitt ein Wahrscheinlichkeitsraum.

Definition 4.1. Eine Folge $\mathbb{F} = (\mathfrak{F}_n)_{n \in \mathbb{N}_0}$ von Teil-σ-Algebren von $\mathfrak{F}$ mit $\mathfrak{F}_0 \subseteq \mathfrak{F}_1 \subseteq \mathfrak{F}_2 \subseteq \dots$ heißt *Filtration*. Das Tupel $(\Omega, \mathfrak{F}, \mathbb{F}, \mathbb{P})$ heißt dann *filtrierter Wahrscheinlichkeitsraum*.

Interpretation: $\mathfrak{F}_n$ enthält die Ereignisse, deren Eintreten (oder Nichteintreten) durch Beobachtung bis zum Zeitpunkt n festgestellt werden kann (enthält die „Information“, die bis zum Zeitpunkt n bekannt ist).

Definition 4.2. Sei $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum. Eine Folge $X = (X_n)$ von Zufallsvariablen auf $(\Omega, \mathfrak{F}, \mathbb{P})$ heißt dann *adaptiert an die Filtration* $\mathbb{F} = (\mathfrak{F}_n)$, falls X_n $\mathfrak{F}_n$-messbar ist für alle $n \in \mathbb{N}_0$.

Ist (X_n) eine Folge von Zufallsvariablen auf $(\Omega, \mathfrak{F}, \mathbb{P})$, so wird durch

$$\mathfrak{F}_n^X := \sigma(X_0, X_1, \dots, X_n),\ n \in \mathbb{N}_0,$$

eine Filtration $\mathbb{F}^X = (\mathfrak{F}_n^X)$ definiert. Diese Filtration heißt die *von X erzeugte Filtration* (auch *kanonische Filtration* genannt). Dabei handelt es sich offenbar um die „kleinste“ Filtration, an die $X = (X_n)$ adaptiert ist.

Interpretation: Die Filtration F_n^X enthält genau die „Information“, die durch Beobachtung der Folge X bis zum Zeitpunkt n gewonnen werden kann.

Wir definieren $\overline{\mathbb{N}_0} := N_0 \cup \{\infty\}$.

Definition 4.3. Sei $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum. Eine Abbildung $\tau : \Omega \to \overline{\mathbb{N}_0}$ heißt *Stoppzeit*, falls

$$\{\tau \leq n\} \in \mathfrak{F}_n \quad \text{für alle } n \in \mathbb{N}_0$$

gilt.

Interpretation: Stoppzeiten sind nicht vorgreifende zufällige Zeiten. Das heißt, um zu entscheiden, ob τ bis zum Zeitpunkt n eingetreten ist, genügt es die zufällige zeitliche Entwicklung bis zum Zeitpunkt n zu kennen.

Beispiel 4.4.

1. Deterministische Zeiten sind Stoppzeiten: $\tau \equiv m$
   $\Rightarrow \{\tau \leq n\} = \emptyset$ für $n < m$ und $= \Omega$ für $n \geq m$.

2. Gegeben seien ein filtrierter Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$, eine adaptierte Folge (X_n) reeller Zufallsgrößen sowie $B \in \mathfrak{B}$. Dann ist

   $$\tau := \inf\{n \in \mathbb{N}_0 : X_n \in B\}$$

   (mit $\inf \emptyset := \infty$) eine Stoppzeit:

   $$\{\tau \leq n\} = \bigcup_{k=0}^{n} \underbrace{\{X_k \in B\}}_{\in \mathfrak{F}_k \subseteq \mathfrak{F}_n} \in \mathfrak{F}_n \qquad \forall n \in \mathbb{N}_0$$

   Man beachte: Kennt man $X_0(\omega), \ldots, X_n(\omega)$, so weiß man, ob $\tau(\omega) \leq n$ ist.
   Die Stoppzeit τ heißt *Zeit des ersten Erreichens der Menge* B (manchmal auch *Zeit des ersten Eintritts in die Menge* B).

3. Definiert man wie oben

   $$\tau_1 := \inf\{n : X_n \in B\} \text{ und } \sigma_1 := \inf\{n \geq \tau_1 : X_n \in B^c\}$$

   und rekursiv

   $$\tau_m := \inf\{n \geq \sigma_{m-1} : X_n \in B\} \text{ sowie } \sigma_m := \inf\{n \geq \tau_m : X_n \in B^c\}$$

   so erhält man eine Folge von Stoppzeiten

   $$0 \leq \tau_1 \leq \sigma_1 \leq \tau_2 \leq \sigma_2 \leq \ldots$$

   wobei

   $$\tau_m \text{ die Zeit des } m\text{-ten Eintritts in die Menge } B$$

   und

   $$\sigma_m \text{ die Zeit des } m\text{-ten Austritts aus der Menge } B$$

   ist (Übungsaufgabe).

4. Anschaulich ist klar, dass die Zeit τ des letzten Verlassens einer Menge B im Allgemeinen keine Stoppzeit ist: Durch die Beobachtung von $X_0, X_1, \ldots$ allein lässt sich nicht feststellen, ob B bis zum Zeitpunkt n das letze Mal verlassen wurde.

Lemma 4.5. *Sind τ_1 und τ_2 Stoppzeiten bezüglich einer Filtration $(\mathfrak{F}_n)$, so sind auch $\tau_1 \wedge \tau_2$, $\tau_1 \vee \tau_2$ und $\tau_1 + \tau_2$ $\mathfrak{F}_n$-Stoppzeiten.*

Beweis. Übungsaufgabe. □

Lemma 4.6. *Seien (X_n) eine adaptierte Folge (mit Werten in einem gemeinsamen messbaren Raum) und τ eine Stoppzeit auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$.*

a) Durch

$$\mathfrak{F}_\tau := \{A \in \mathfrak{F} : \ A \cap \{\tau \leq n\} \in \mathfrak{F}_n \ \ \forall n \in \mathbb{N}_0\}$$

wird eine Teil-σ-Algebra von $\mathfrak{F}$ definiert.

b) Ist $\tau < \infty$ (das heißt $\tau(\omega) < \infty$ für alle $\omega \in \Omega$), so wird durch

$$X_\tau(\omega) := X_{\tau(\omega)}(\omega), \quad \textit{mit } \omega \in \Omega,$$

eine $\mathfrak{F}_\tau$-messbare Zufallsvariable definiert.

Beweis.

a) Sei $\Omega \cap \{\tau \leq n\} \in \mathfrak{F}$ für alle n, dann ist $\Omega \in \mathfrak{F}_\tau$ und es folgt aus

$$\begin{aligned}
A \in \mathfrak{F}_\tau \quad &\Rightarrow \quad A^c \cap \{\tau \leq n\} \\
&= \quad (\underbrace{A \cap \{\tau \leq n\}}_{\in \mathfrak{F}_n})^c \cap \underbrace{\{\tau \leq n\}}_{\in \mathfrak{F}_n} \in \mathfrak{F}_n \qquad \text{für alle } n \\
&\Rightarrow \quad A^c \in \mathfrak{F}_\tau \quad \text{und aus} \\
A_m \in \mathfrak{F}_\tau \forall m \quad &\Rightarrow \quad (\bigcup_m A_m) \cap \{\tau \leq n\} \\
&= \quad \bigcup_m (\underbrace{A_m \cap \{\tau \leq n\}}_{\in \mathfrak{F}_n}) \in \mathfrak{F}_n \qquad \text{für alle } n \\
&\Rightarrow \quad \bigcup_m A_m \in \mathfrak{F}_\tau.
\end{aligned}$$

b) Ist B eine messbare Menge aus dem Bildraum der X_n, so folgt für beliebige $n \in \mathbb{N}_0$:

$$\{X_\tau \in B\} \cap \{\tau \leq n\} = \bigcup_{k=0}^{n} \underbrace{\{X_k \in B\}}_{\in \mathfrak{F}_k \subseteq \mathfrak{F}_n} \cap \underbrace{\{\tau = k\}}_{\in \mathfrak{F}_k \subseteq \mathfrak{F}_n} \in \mathfrak{F}_n.$$

□

Die σ-Algebra $\mathfrak{F}_\tau$ heißt *σ-Algebra der τ-Vergangenheit.* Anschaulich enthält $\mathfrak{F}_\tau$ alle Informationen über die zufällige zeitliche Entwicklung bis zum zufälligen Zeitpunkt τ.

4.2. Martingale

Sei nun $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$ ein beliebiger filtrierter Wahrscheinlichkeitsraum.

Definition 4.7. Eine Folge $M = (M_n)_{n\in\mathbb{N}_0}$ reeller Zufallsgrößen heißt *Martingal* (bezüglich der Filtration $(\mathfrak{F}_n)$), falls gilt:

a) (M_n) ist *integrierbar*, das heißt $M_n \in \mathcal{L}^1(\mathbb{P})$ für alle $n \in \mathbb{N}_0$;

b) (M_n) ist $(\mathfrak{F}_n)$-adaptiert;

c) $\mathbb{E}[M_{n+1}|\mathfrak{F}_n] = M_n$ $\mathbb{P}$-fast sicher für alle $n \in \mathbb{N}_0$.

Gilt in (iii) anstelle der Gleichheit die Ungleichung „$\geq$“ (bzw. „$\leq$“), so spricht man von einem *Submartingal* (bzw. einem *Supermartingal*).

Man kann ein Martingal als ein „gerechtes Spiel“ interpretieren, wobei M_n der Gewinn des ersten Spielers zum Zeitpunkt n und $\mathbb{E}[M_{n+1}|\mathfrak{F}_n]$ der *mittlere* Gewinn des ersten Spielers zum Zeitpunkt n *bei Kenntnis des Spielverlaufs* bis zum Zeitpunkt n ist.

Beispiel 4.8.

1. Sei (X_n) eine Folge unabhängiger identisch verteilter Zufallsgrößen mit endlichen Erwartungswerten. Als Filtration betrachten wir die kanonische Filtration $(\mathfrak{F}_n^X)$
$S_n := \sum_{k=0}^{n} X_n, \quad n \in \mathbb{N}_0.$ (Irrfahrt auf $\mathbb{R}$ mit Sprüngen).
$m := \mathbb{E}X_0 \ (= \mathbb{E}X_k$ für alle $k)$.
Dann gilt bezüglich $(\mathfrak{F}_n^X)$:
$$(S_n) \text{ ist ein } \begin{cases} \text{Martingal,} & \text{falls } m = 0 \\ \text{Submartingal,} & \text{falls } m > 0 \\ \text{Supermartingals,} & \text{falls } m < 0 \end{cases}$$
Beweis. Offensichtlich ist $S_n \in \mathcal{L}^1$ und $\sigma(X_0, \ldots, X_n) = \mathfrak{F}_n^X$-messbar (das heißt an $(\mathfrak{F}_n^X)$ adaptiert). Wegen $S_{n+1} = S_n + X_{n+1}$ gilt
$$\mathbb{E}[S_{n+1}|\mathfrak{F}_n^X] \overset{\text{f.s.}}{=} \mathbb{E}[S_n|\mathfrak{F}_n^X] + \mathbb{E}[X_{n+1}|\mathfrak{F}_n^X] \overset{\text{f.s.}}{=} S_n + \mathbb{E}X_{n+1} = S_n + m,$$
da S_n $\mathfrak{F}_n^X$-messbar und X_{n+1} unabhängig von $\mathfrak{F}_n^X$ ist.
Und daraus folgt die Behauptung. □

2. Sei $X = (X_n)_{n\in\mathbb{N}_0}$ eine Markovkette mit Zustandsraum I, einer Anfangsverteilung $p^0 = (p_i^0)_{i\in I}$ sowie einer Übergangsmatrix $P = (p_{ij})_{i,j\in I}$. Mit $G := P - E$ bezeichnen wir den *Generator* der Markovkette (E sei die

Einheitsmatrix).
Gegeben sei nun eine beschränkte Funktion $f : I \to \mathbb{R}$. Dann ist

$$M_n := f(X_n) - \sum_{k=0}^{n-1} Gf(X_k), \;\; n \in \mathbb{N}_0, \tag{4.1}$$

ein $(\mathfrak{F}_n^X)$-Martingal.

Beweis. Wir beweisen zunächst die folgende Version der *Markoveigenschaft*:

$$\mathbb{E}[f(X_{n+1})|\mathfrak{F}_n^X] = \sum_{j\in I} p_{X_n,j} f(j) = Pf(X_n). \qquad (*)$$

Hierzu ist in wesentlichen nur nachzuweisen, dass für alle $A \in \mathfrak{F}_n^X$

$$\mathbb{E}\mathbb{1}_A f(X_{n+1}) = \mathbb{E}\mathbb{1}_A \sum_{j\in I} p_{X_n,j} f(j)$$

gilt. Jede solche Menge A hat die Gestalt

$$A = \bigcup_{(i_0,\dots,i_n)\in B} \{X_0 = i_0, \dots, X_n = i_n\} \;\; \text{für ein } B \subseteq I^{n+1}.$$

Nun ist aber auf Grund der Markoveigenschaften

$$\begin{aligned}
& \mathbb{E}\, \mathbb{1}_{\{X_0=i_0,\dots,X_n=i_n\}} f(X_{n+1}) \\
= \;& \sum_{j\in I} \mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = j)\, f(j) \\
= \;& \sum_{j\in I} p^0_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} p_{i_n j}\, f(j) \\
= \;& \sum_{j\in I} \mathbb{E}\mathbb{1}_{\{X_0=i_0,\dots,X_n=i_n\}} p_{X_n,j}\, f(j)
\end{aligned}$$

und die Behauptung folgt hieraus durch Summation über $(i_0, \dots, i_n) \in B$.
Da f beschränkt ist, ist auch Gf beschränkt (Beweis als Übungsaufgabe) und daher (M_n) integrierbar. Offenbar ist (M_n) an $(\mathfrak{F}_n^X)$ adaptiert. Bleibt also zu zeigen, dass für jedes $n \in \mathbb{N}_0$

$$\mathbb{E}[M_{n+1} - M_n|\mathfrak{F}_n] = 0 \text{ f.s.}$$

gilt. Tatsächlich folgt mit der Markoveigenschaft, dass

$$\begin{aligned}
\mathbb{E}[M_{n+1} - M_n|\mathfrak{F}] &= \mathbb{E}[f(X_{n+1}) - \underbrace{f(X_n) - Gf(X_n)}_{=Pf(X_n)\;\; \mathfrak{F}_n\text{-mb.}}|\mathfrak{F}_n^X] \\
&= \mathbb{E}[f(X_{n+1})|\mathfrak{F}_n] - Pf(X_n) \\
&= 0.
\end{aligned}$$

□

Ist $Gf = 0$ (das heißt ist f *harmonisch*), so folgt aus (4.1), dass $(f(X_n))_{n\in\mathbb{N}_0}$ ein $\mathfrak{F}_n^X$-Martingal ist.
Ist die Markovkette irreduzibel, so gilt auch die Umkehrung (Übungsaufgabe).

Spezialfall: (X_n) sei eine einfache Irrfahrt auf $I := \mathbb{Z}$, das heißt $p_{ij} = \frac{1}{2}$ für $|j-i| = 1$ und $p_{ij} = 0$ sonst. Dann gilt $|X_n - X_{n-1}| = 1$ und deshalb (Achtung: Fallunterscheidung)

$$\begin{aligned} f(X_n) - f(X_{n-1}) &= \frac{f(X_{n-1}+1) - f(X_{n-1}-1)}{2}(X_n - X_{n-1}) \\ &\quad + \frac{f(X_{n-1}+1) + f(X_{n-1}-1)}{2} - f(X_{n-1}). \end{aligned}$$

Mit den *diskreten Ableitungen*

$$\begin{aligned} f'(i) &:= \frac{f(i+1) - f(i-1)}{2} \\ f''(i) &:= f(i+1) - 2f(i) + f(i-1), \quad i \in \mathbb{Z} \end{aligned}$$

lässt sich dies wie folgt schreiben:

$$f(X_n) - f(X_{n-1}) = f'(X_{n-1})(X_n - X_{n-1}) + \frac{1}{2}f''(X_{n-1}).$$

Durch Summation folgt hieraus die *diskrete Itô-Formel*

$$\begin{aligned} f(X_n) &= f(X_0) + \sum_{k=1}^{n} f'(X_{k-1})(X_k - X_{k-1}) + \frac{1}{2}\sum_{k=1}^{n} f''(X_{k-1}) \qquad (4.2) \\ &= f(X_0) + \underbrace{(F' \circ X)_n}_{\text{diskretes stoch. Integral (s.ÜA)}} + \frac{1}{2}\sum_{k=1}^{n} F''_k \end{aligned}$$

mit $F'_n := f'(X_{n-1})$ und $F''_n := f''(X_{n-1})$.
Andrerseits haben wir mit der Formel (4.1):

$$f(X_n) = M_n + \sum_{k=1}^{n} Gf(X_{k-1}). \qquad (4.3)$$

Wegen $Gf = \frac{1}{2}f''$ findet man durch Vergleich von (4.2) und (4.3)

$$M_n = f(X_0) + (F' \circ X)_n. \qquad (4.4)$$

Ist (X_n) ein beliebiges $\mathbb{Z}$-wertiges Martingal mit $|X_n - X_{n-1}| = 1$ (für alle n), so gilt die diskrete Itô-Formel (4.2) immer noch, wobei das diskrete stochastische Integral in (4.4) (und damit auch (M_n)) ein Martingal ist (siehe Übungsaufgabe).

Im folgenden führen wir einige einfache Eigenschaften von Martingalen an.

Satz 4.9. *Sei $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum.*

(i) X ist ein Submartingal $\Leftrightarrow -X$ ist ein Supermartingal;

(ii) Seien X und Y Martingale (Supermartingale) und $a, b \in \mathbb{R}$ $(a, b \geq 0)$, dann ist $aX + bY$ wieder ein Martingal (Supermartingal);

(iii) Sind X und Y Supermartingale, so ist $X \wedge Y$ auch ein Supermartingal;

(iv) Ist $Z \in \mathcal{L}^1$, so ist

$$X_n := \mathbb{E}[Z|\mathfrak{F}_n], \;\; n \in \mathbb{N}_0,$$

ein Martingal.

(v) Sei $X = (X_n)$ ein Supermartingal. Existiert eine Folge von Zeitpunkten $N_n \to \infty$ mit $\mathbb{E}X_{N_n} \geq \mathbb{E}X_0$ für alle n, so ist X ein Martingal.

Beweis. Übungsaufgabe. □

Satz 4.10. *Sei (X_n) ein Martingal und $\varphi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion mit $\varphi(X_n) \in \mathcal{L}^1$ für alle n. Dann ist $(\varphi(X_n))$ ein Submartingal.*

Beweis. $(\varphi(X_n))$ ist integrierbar und adaptiert. Die Jensensche Ungleichung liefert

$$\mathbb{E}[\varphi(X_{n+1})|\mathfrak{F}_n] \overset{\text{f.s.}}{\geq} \varphi(\mathbb{E}[X_{n+1}|\mathfrak{F}_n]) \overset{\text{f.s.}}{=} \varphi(X_n).$$

□

4.3. Doob-Zerlegung und quadratische Variation

Für diesen Abschnitt sei $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$ ein beliebiger filtrierter Wahrscheinlichkeitsraum.

Definition 4.11. Eine Folge $(X_n)_{n\in\mathbb{N}_0}$ von Zufallsvariablen heißt *previsibel* (vorhersehbar), falls X_0 konstant und

$$X_n \;\; \mathfrak{F}_{n-1}\text{-messbar ist für alle } n \in \mathbb{N}.$$

Satz 4.12. *(Doob-Zerlegung)*
Sei X eine adaptierte integrierbare Folge. Dann existiert eine fast sicher eindeutige Zerlegung

$$X = M + A,$$

wobei M ein Martingal und A eine previsible Folge mit $A_0 = 0$ ist. Ist X ein Submartingal, so ist die Folge A fast sicher nichtfallend.

Beweis.

1^o Existenz:

$$\begin{aligned} M_n &:= X_0 + \sum_{k=1}^{n} (X_k - \mathbb{E}[X_k|\mathfrak{F}_{k-1}]) \\ \text{und} \quad A_n &:= \sum_{k=1}^{n} (\mathbb{E}[X_k|\mathfrak{F}_{k-1}] - X_{k-1}) \end{aligned}$$

ist die gewünschte Zerlegung. Tatsächlich gilt

$$M_n + A_n = X_n,$$

mit M_n und $A_n \in \mathcal{L}^1$, wobei M_n $\mathfrak{F}_n$-messbar, A_n $\mathfrak{F}_{n-1}$-messbar sowie $A_0 = 0$ ist.
Bleibt zu zeigen, dass M ein Martingal ist:

$$\begin{aligned} \mathbb{E}[M_{n+1} - M_n|\mathfrak{F}_n] &= \mathbb{E}[X_{n+1} - \underbrace{\mathbb{E}[X_{n+1}|\mathfrak{F}_n]}_{\mathfrak{F}_n\text{-mb.}}|\mathfrak{F}_n] \\ &= \mathbb{E}[X_{n+1}|\mathfrak{F}_n] - \mathbb{E}[X_{n+1}|\mathfrak{F}_n] = 0 \quad \text{fast sicher.} \end{aligned}$$

2^o Eindeutigkeit: Seien $X = M + A = \widetilde{M} + \widetilde{A}$ zwei solche Zerlegungen. Dann ist $\widehat{M} := M - \widetilde{M} = \widetilde{A} - A$ gleichzeitig ein Martingal und previsibel, das heißt

$$\widehat{M}_n = \mathbb{E}[\widehat{M}_{n+1}|\mathfrak{F}_n] \underset{\text{previsibel}}{=} \widehat{M}_{n+1} \quad \text{f.s.}$$

Also ist $\widehat{M}_n = \widehat{M}_0 = 0$ fast sicher, woraus $M = \widetilde{M}$ fast sicher und $A = \widetilde{A}$ folgt.

3^o Ist X ein Submartingal, das heißt $\mathbb{E}[X_k|\mathfrak{F}_{k-1}] - X_{k-1} \geq 0$, so folgt aus der obigen Darstellung von A, dass die Folge A fast sicher nichtfallend ist.

□

Ist $M = (M_n)_{n\in\mathbb{N}_0}$ ein quadratisch integrierbares Martingal (das heißt $M_n \in \mathcal{L}^2(\mathbb{P})$, für alle n), so ist (M_n^2) ein Submartingal (da $\varphi(x) = x^2$ konvex ist, siehe Satz 4.10). Also besitzt (M_n^2) eine Doob-Zerlegung

$$M_n^2 = N_n + A_n \tag{4.5}$$

mit einem Martingal N und einer nichtfallenden previsiblen Folge A, mit $A_0 = 0$.

Definition 4.13. Die Folge A in (4.5) heißt *quadratische Variation* des quadratisch integrierbaren Martingals M und wird mit $\langle M\rangle$ bezeichnet (das heißt $\langle M\rangle_n := A_n$, mit $n \in \mathbb{N}_0$).

Beispiel 4.14.

1. Sei $X = (X_n)_{n\in\mathbb{N}}$ eine einfache symmetrische Irrfahrt auf $\mathbb{Z}$ mit $X_0 = 0$. Dann ist X ein Martingal bezüglich der kanonischen Filtration $(\mathfrak{F}_n^X)$ (als Summe unabhängiger Zufallsgrößen mit Erwartungswert 0). Also ist $|X|$ ein Submartingal (da $\varphi(x) = |x|$ konvex ist) und besitzt somit eine Doob-Zerlegung
$$|X| = M + A$$
mit einer nichtfallenden previsiblen folge $A = (A_n)$. Wegen
$$|X_{n+1}| - |X_n| = -\mathbb{1}_{X_n<0}(X_{n+1} - X_n) + \mathbb{1}_{X_n=0} + \mathbb{1}_{X_n>0}(X_{n+1} - X_n)$$
folgt
$$A_{n+1} - A_n = \mathbb{E}[|X_{n+1}| - |X|_n|\mathfrak{F}_n^X] = \mathbb{1}_{X_n=0},$$
das heißt
$$A_n = \sum_{k=0}^{n-1} \delta_0(X_k) \quad \textit{Lokalzeit} \text{ von } X \text{ in } 0.$$
Bemerkung: $|X|$ ist symmetrische Irrfahrt auf $\mathbb{N}_0$ mit Reflektion in 0.

2. Im Beispiel 4.8.2 wurden eine Markovkette (X_n) auf I, eine beschränkte Funktion $f : I \to \mathbb{R}$ und ein Generator $G := P - I$ betrachtet. Dort haben wir bereits die Doob-Zerlegung von $(f(X_n))$ (bezüglich $(\mathfrak{F}_n^X)$) erhalten:
$$f(X_n) = M_n + A_n$$
mit der previsiblen Folge
$$A_n = \sum_{k=0}^{n-1} Gf(X_k).$$
Da M quadratisch integrierbar ist (jedes M_n ist beschränkt), existiert die quadratische Variation $\langle M\rangle$. Diese ist gegeben durch
$$\langle M\rangle_n = \sum_{k=0}^{n-1} \underbrace{(Pf^2 - (Pf)^2)}_{\geq 0}(X_k).$$
Zum Beweis ist zu zeigen, dass
$$N_n := M_n^2 - \sum_{k=0}^{n-1} (Pf^2 - (Pf)^2(X_k)$$
ein Martingal ist, das heißt
$$\mathbb{E}[M_{n+1}^2 - M_n^2|\mathfrak{F}_n^X] = (Pf^2 - (Pf)^2)(X_k).$$

Nun gilt tatsächlich mit der Markoveigenschaft

$$\begin{aligned}
\mathbb{E}[M_{n+1}^2 - M_n^2|\mathfrak{F}_n^X] &= \mathbb{E}[(M_{n+1} - M_n)^2|\mathfrak{F}_n^X] \\
&= \mathbb{E}[(f(X_{n+1}) - Pf(X_n))^2|\mathfrak{F}_n^X] \\
&= \mathbb{E}[f^2(X_{n+1})|\mathfrak{F}_n^X] - 2Pf(X_n)\,\mathbb{E}[f(X_{n+1})|\mathfrak{F}_n^X] \\
&\quad +(Pf)^2(X_n) \\
&= (Pf^2 - (Pf)^2)(X_n).
\end{aligned}$$

4.4. Martingale und Stoppzeiten

Lemma 4.15. *Seien X ein Martingal und τ eine Stoppzeit mit $\tau \leq T$ für ein $T \in \mathbb{N}_0$. Dann gilt*

$$\mathbb{E}[X_T|\mathfrak{F}_\tau] = X_\tau \quad \textit{fast sicher}$$

und insbesondere $\mathbb{E}X_T = \mathbb{E}X_\tau$.

Beweis. Da X_τ $\mathfrak{F}_\tau$-messbar und integrierbar ($|X_\tau| \leq |X_0| + \ldots$) ist, bleibt nur zu zeigen

$$\mathbb{E}\mathbb{1}_A X_T = \mathbb{E}\mathbb{1}_A X_\tau \quad \text{für alle } A \in \mathfrak{F}_\tau.$$

Sei also $A \in \mathfrak{F}_\tau$. Dann ist

$$\mathbb{E}\mathbb{1}_A X_T = \sum_{n=0}^{T} \mathbb{E}\mathbb{1}_{\underbrace{A \cap \{\tau = n\}}_{\mathfrak{F}_n\text{-mb.}}} X_T.$$

Mit der Martingaleigenschaft von X ist dann

$$\begin{aligned}
\sum_{n=0}^{T} \mathbb{E}\mathbb{1}_{A\cap\{\tau=n\}} X_T &= \sum_{n=0}^{T} \mathbb{E}(\mathbb{E}[\mathbb{1}_{A\cap\{\tau=n\}} X_T|\mathfrak{F}_n]) \\
&= \sum_{n=0}^{T} \mathbb{E}\mathbb{1}_{A\cap\{\tau=n\}} X_n \\
&= \mathbb{E}\sum_{n=0}^{T} \mathbb{E}\mathbb{1}_{A\cap\{\tau=n\}} X_\tau \\
&= \mathbb{E}\mathbb{1}_A X_\tau.
\end{aligned}$$

□

Satz 4.16. *[Optional Sampling Theorem]*
Seien X ein Submartingal und σ sowie τ Stoppzeiten mit $\sigma \leq \tau$.

(i) Ist $\tau \leq T$ für ein $T \in \mathbb{N}_0$, so gilt

$$\mathbb{E}[X_\tau|\mathfrak{F}_\sigma] \leq X_\sigma \quad \textit{fast sicher}$$

und insbesondere $\mathbb{E}X_\tau \leq \mathbb{E}X_\sigma$. Ist X ein Martingal, so gilt jeweils Gleichheit.

(ii) *Ist X nichtnegativ und $\tau < \infty$, so gilt $\mathbb{E}X_\tau \leq \mathbb{E}X_\sigma \leq \mathbb{E}X_0 < \infty$ und*

$$\mathbb{E}[X_\tau | \mathfrak{F}_\sigma] \leq X_\sigma \quad \textit{fast sicher.}$$

Beweis.

(i) X besitzt eine Doob-Zerlegung $X = M + A$ mit einer nichtwachsenden previsiblen Folge A. Deshalb gilt

$$\mathbb{E}[X_\tau | \mathfrak{F}_\sigma] = \mathbb{E}[M_\tau | \mathfrak{F}_\sigma] + \mathbb{E}[A_\tau | \mathfrak{F}_\sigma]$$

Nun gilt nach Lemma 4.15 (∗) und mit der Turmeigenschaft

$$\mathbb{E}[M_\tau | \mathfrak{F}_\sigma] \underset{(*)}{=} \mathbb{E}[\mathbb{E}[M_T | \mathfrak{F}_\tau] | \mathfrak{F}_\sigma] = \mathbb{E}[M_T | \mathfrak{F}_\sigma] \underset{(*)}{=} M_\sigma$$

und

$$\mathbb{E}[A_\tau | \mathfrak{F}_\sigma] \leq \mathbb{E}[A_\sigma | \mathfrak{F}_\sigma] = A_\sigma,$$

da A nichtwachsend und previsibel (also auch adaptiert) ist.
Hieraus folgt $\mathbb{E}[X_\tau | \mathfrak{F}_\sigma] \leq X_\sigma$ und Gleichheit im Falle eines Martingals.

(ii) Folgt aus (i) für die beschränkten Stoppzeiten $\sigma_m := \sigma \wedge m$ und $\tau_n := \tau \wedge n$ nach Grenzübergang $n \to \infty$, $m \to \infty$ unter Verwendung von Konvergenzsätzen für bedingte Erwartungen (Lemma von Fatou, Satz von Lebesque). (Übungsaufgabe)

□

Folgerung 4.17.
Ist X ein Martingal (Submartingal) und ist $0 \leq \tau_0 \leq \tau_1 \leq \tau_2 \leq \dots$ eine Folge von beschränkten Stoppzeiten (das heißt $\tau_n \leq T_n$ für ein $T_n \in \mathbb{N}_0$ $(n \in \mathbb{N}_0)$), so ist $(X_{\tau_n})_{n \in \mathbb{N}_0}$ ein Martingal (Submartingal) bezüglich der Filtration $(\mathfrak{F}_{\tau_n})$.

Bemerkung 4.18. Ohne die Beschränktheit der Stoppzeiten sind die Aussagen in Satz 4.16a) und die Folgerung 4.17 im Allgemeinen falsch. (Ohne Beweis)

Satz 4.19. *(Optional Stopping) Seien X ein Martingal (Submartingal) bezüglich einer Filtration $\mathbb{F} = (\mathfrak{F}_n)$ und τ eine $\mathbb{F}$-Stoppzeit. Dann ist die gestoppte Folge X^τ ein Martingal (Submartingal) sowohl bezüglich der Filtration $\mathbb{F}$ als auch bezüglich der Filtration $\mathbb{F}^\tau = (\mathfrak{F}^\tau_n)$ mit $\mathfrak{F}^\tau_n := \mathfrak{F}_{n \wedge \tau}$.*

Beweis. (für Submartingale) Folgerung 4.17 liefert sofort, dass X^τ ein Submartingal bezüglich $\mathbb{F}$ ist. Insbesondere gilt

$$\mathbb{E}[X_{(n+1)\wedge\tau} | \mathfrak{F}_{n\wedge\tau}] \geq X_{n\wedge\tau} \quad \text{f.s.}$$

Außerdem ist $X_{n\wedge\tau} \in \mathcal{L}^1$ und $\mathfrak{F}_n$-messbar (da $\mathfrak{F}_{n\wedge\tau} \subseteq \mathfrak{F}_n$). Um zu zeigen, dass X^τ ein Submartingal bezüglich $\mathbb{F}$ ist, bleibt zu zeigen, dass

$$\mathbb{E}[X_{(n+1)\wedge\tau} | \mathfrak{F}_n] \geq X_{n\wedge\tau} \quad \text{f.s.}$$

Wegen

$$X_{(n+1)\wedge\tau} = \underbrace{\mathbb{1}_{\tau\le n} X_{\tau\wedge n}}_{\mathfrak{F}_n\text{-mb.}} + \underbrace{\mathbb{1}_{\tau>n}}_{\mathfrak{F}_n\text{-mb.}} X_{n+1}$$

erhalten wir

$$\begin{aligned}\mathbb{E}[X_{(n+1)\wedge\tau}|\mathfrak{F}_n] &= \mathbb{1}_{\tau\le n} X_\tau + \mathbb{1}_{\tau>n} \underbrace{\mathbb{E}[X_{n+1}|\mathfrak{F}_n]}_{\ge X_n \text{ (Submartingal)}} \\ &\ge \mathbb{1}_{\tau\le n} X_\tau + \mathbb{1}_{\tau>n} X_n = X_{n\wedge\tau} \quad \text{f.s.}\end{aligned}$$

□

Beispiel 4.20. (Ruin eines Spielers) Sei (X_n) eine einfache symmetrische Irrfahrt mit $X_0 = x$ (Startkapital des ersten Spielers) und $0 \le x \le N$ ($N - x$ ist das Startkapital des zweiten Spielers)

$$\text{Stoppzeiten} \begin{cases} \sigma := \inf\{n : X_n = 0\} & \text{Ruin des ersten Spielers} \\ \tau := \inf\{n : X_n = N\} & \text{Ruin des zweiten Spielers} \end{cases}$$

Es gilt $\sigma < \infty$ und $\tau < \infty$ fast sicher (ohne Beweis).
Man berechne $\mathbb{P}(\sigma < \tau)$ (der erste Spieler ist vor dem zweiten Spieler bankrott).
Da (X_n) ein Martingal ist, gilt für jedes $T \in \mathbb{N}$ mit Hilfe des Optional Samplings $(*)$

$$\begin{aligned} x &= \mathbb{E}X_0 \underset{(*)}{=} \mathbb{E}X_{(\sigma\wedge\tau)\wedge T} \\ &= \mathbb{E}\left(\mathbb{1}_{\sigma<\tau\le T} \underbrace{X_\sigma}_{=0} + \mathbb{1}_{\tau<\sigma\le T} \underbrace{X_\tau}_{=N} + \mathbb{1}_{(\sigma\wedge\tau)>T} X_T\right) \\ &= N\,\mathbb{P}(\tau < \sigma \le T) + \mathbb{E}\mathbb{1}_{(\sigma\wedge\tau)>T} X_T \\ &\underset{T\uparrow\infty}{\longrightarrow} N\,\mathbb{P}(\tau<\sigma) \qquad + \qquad 0 \\ &= N\,\mathbb{P}(\tau<\sigma), \end{aligned}$$

da $0 \le X_T \le N$ auf $\{\sigma \wedge \tau > T\}$, das heißt

$$\mathbb{P}(\tau < \sigma) = \frac{x}{N} \text{ und } \mathbb{P}(\sigma < \tau) = \frac{N-x}{N}.$$

Zur Abschwächung der Beschränktheit der Stoppzeiten in den obigen Aussagen benötigen wir die folgende Definition.

Definition 4.21. Eine Folge (X_n) von Zufallsgrößen heißt *gleichgradig integrierbar* (uniform integrierbar), falls zu jedem $\varepsilon > 0$ eine Zufallsgröße $Y \in \mathcal{L}^1(\mathbb{P})$ existiert mit

$$\sup_n \mathbb{E}\mathbb{1}_{|X_n|\ge Y} |X_n| < \varepsilon.$$

Ist

$$\sup_n \mathbb{E}|X_n|^p < \infty \text{ für ein } p > 1,$$

so ist (X_n) gleichgradig integrierbar, da

$$\sup_n \mathbb{E}\mathbb{1}_{|X_n|>c}|X_n| \leq \frac{1}{c^{p-1}} \sup_n \mathbb{E}|X_n|^p \to 0 \text{ für } c \uparrow \infty \text{ gilt.}$$

Satz 4.22. *(Ohne Beweis)*
Ist $X = (X_n)$ ein gleichgradig integrierbares Supermartingal und sind σ und τ endliche Stoppzeiten mit $\sigma \leq \tau$, so gilt X_σ und $X_\tau \in \mathcal{L}^1$ und

$$\mathbb{E}[X_\tau|\mathfrak{F}_\sigma] \leq X_\sigma \text{ f.s.}$$

Ist X ein Martingal, so gilt die Gleichheit.

4.5. Doobsche Ungleichungen

Für eine beliebige Folge (X_n) von Zufallsgrößen schreiben wir

$$X_n^* := \max_{k \leq n} X_k \quad (\text{mit } n \in \mathbb{N}_0).$$

Lemma 4.23. *Sei X ein Submartingal. Dann gilt für beliebige $\lambda > 0$*

$$\lambda \mathbb{P}(X_n^* \geq \lambda) \leq \mathbb{E}(\mathbb{1}_{X_n^* \geq \lambda} X_n) \leq \mathbb{E}(\mathbb{1}_{X_n^* \geq \lambda}|X_n|),$$

Beweis. Wir betrachten die Stoppzeit $\tau = \inf\{n : X_n \geq \lambda\}$. Dann ist

$$\{\tau \leq n\} = \{X_n^* \geq \lambda\}.$$

Das Optional Sampling Theorem liefert uns

$$\mathbb{E}X_n \geq X_{(\tau \wedge n)} = \mathbb{E}(\mathbb{1}_{\tau \leq n} X_\tau + \mathbb{1}_{\tau > n} X_n)$$

und folglich

$$\mathbb{E}(\mathbb{1}_{\tau \leq n} X_n) \geq \mathbb{E}(\mathbb{1}_{\tau \leq n} \underbrace{X_\tau}_{\geq \lambda}) \geq \lambda \mathbb{P}(\tau \leq n),$$

das heißt

$$\mathbb{E}(\mathbb{1}_{X_n^* \geq \lambda} X_n) \geq \lambda \mathbb{P}(X_n^* \geq \lambda).$$

□

Satz 4.24. *(Doobsche Ungleichungen)*
Sei X ein Martingal oder ein nicht- negatives Submartingal.

(i) *Für $p \geq 1$ und $\lambda > 0$ gilt*

$$\mathbb{P}(|X|_n^* \geq \lambda) \leq \frac{1}{\lambda^p} \mathbb{E}|X_n|^p. \tag{4.6}$$

(ii) Für $p > 1$ gilt

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

Beweis.

(*i*) $\varphi(x) = |x|^p$ (für $p \geq 1$) ist konvex. Für $c > 0$ setzen wir $\varphi_c(x) = \varphi(x)$ für $|x| \leq c$, $\varphi_c(x) = \varphi(c) + \varphi'(c)(x-c)$ für $x > c$ und $\varphi_c(x) = \varphi(-c) + \varphi'(-c)(x+c)$ für $x < -c$. Das heißt φ_c entsteht aus φ durch Linearisierung ausserhalb des Intervalls $[-c, c]$. Dann ist φ_c ebenfalls konvex und $(\varphi_c(X_n))$ ist ein Submartingal nach Satz 4.10. Der Beweis dieses Satzes funktioniert auch für nichtnegative Submartingale, da φ_c auf $[0, \infty)$ nichtfallend ist. Der Erwartungswert eines Submartingals ist aber nichtfallend, das heißt

$$\mathbb{E}\varphi_c(X_k) \leq \mathbb{E}\varphi_c(X_{k+1}) \quad (\text{für } k \in \mathbb{N}_0).$$

Durch einen monotonen Grenzübergang für $c \uparrow \infty$ folgt

$$\varphi_c(x) \uparrow \varphi(x) = |x|^p$$

und deshalb

$$\mathbb{E}|X_k|^p \leq \mathbb{E}|X_k|^{p+1}.$$

Zum Beweis der Ungleichung (4.6) sei $n \in \mathbb{N}$ beliebig fixiert und o.B.d.A. $\mathbb{E}|X_n|^p < \infty$. Dann ist nach obigem die (endliche) Folge $(|X_k|^p)_{k \in \{0,1,\ldots,n\}}$ integrierbar. Da $\varphi(x) = x^p$ konvex und auf $[0, \infty)$ nichtfallend ist, handelt es sich um ein Submartingal auf der Zeitmenge $\{0, 1, \ldots, n\}$. Deshalb erhalten wir

$$\mathbb{P}(|X|_n^* \geq \lambda) = \mathbb{P}((|X|^p)_n^* \geq \lambda^p) \leq \frac{1}{\lambda^p}\,\mathbb{E}|X_n|^p$$

nach Lemma 4.23 auf $\{0, 1, \ldots, n\}$.

(*ii*) Sei $K > 0$ beliebig. Dann ist $\mathbb{E}(|X|_n^* \wedge K)^p < \infty$ und wir erhalten mit dem Satz von Fubini $(*)$, dem Lemma 4.23 $(**)$ und der Hölderschen

Ungleichung $(***)$

$$\begin{aligned}
\mathbb{E}(|X|_n^* \wedge K)^p &= \mathbb{E} \int\limits_0^{|X|_n^* \wedge K} p\lambda^{p-1}\, d\lambda = \mathbb{E} \int\limits_0^K p\lambda^{p-1} \mathbb{1}_{|X|_n^* \geq \lambda}\, d\lambda \\
&\underset{(*)}{=} \int\limits_0^K p\lambda^{p-1} \mathbb{P}(|X|_n^* \geq \lambda)\, d\lambda \\
&\underset{(**)}{\leq} \int\limits_0^K p\lambda^{p-2} \mathbb{E}(\mathbb{1}_{|X|_n^* \geq \lambda} |X_n|)\, d\lambda \\
&\underset{(*)}{=} p\,\mathbb{E}\left(|X_n| \int_0^K \lambda^{p-2} \mathbb{1}_{|X|_n^* \geq \lambda}\, d\lambda \right) \\
&= p\,\mathbb{E}\left(|X_n| \int\limits_0^{|X|_n^* \wedge K} \lambda^{p-2}\, d\lambda \right) \\
&= \frac{p}{p-1}\, \mathbb{E}(|X_n|(|X|_n^* \wedge K)^{p-1}) \\
&\underset{(***)}{\leq} \frac{p}{p-1} \left(\mathbb{E}|X_n|^p\right)^{1/p} \left(\mathbb{E}(|X|_n^* \wedge K)^{(p-1)q}\right)^{1/q}
\end{aligned}$$

Wegen $\frac{1}{p} + \frac{1}{q} = 1$ ist dabei $(p-1)q = p$ und $\frac{1}{q} = 1 - \frac{1}{p}$, und wir erhalten

$$\mathbb{E}(|X|_n^* \wedge K)^p \leq \frac{p}{p-1} \left(\mathbb{E}|X_n|^p\right)^{1/p} \left(\mathbb{E}(|X|_n^* \wedge K)^p\right)^{1-1/p}.$$

Lösen wir diese Ungleichung nach $\mathbb{E}(|X|_n^* \wedge K)^p$ auf (was auf Grund der Endlichkeit dieses Ausdrucks möglich ist), so erhalten wir

$$\mathbb{E}(|X|_n^* \wedge K)^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_n|^p.$$

Durch den monotonen Grenzübergang für $K \uparrow \infty$ folgt nun die gewünschte Ungleichung.

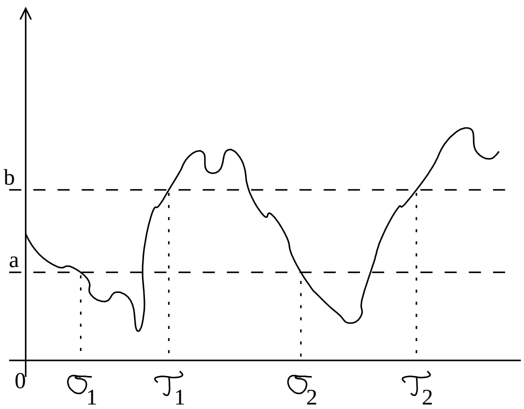
$\square$

4.6. Martingalkonvergenzsätze

Seien nun $(\Omega, \mathfrak{F}, (\mathfrak{F}_n), \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum und $\mathfrak{F}_\infty := \bigvee\limits_{n=1}^{\infty} \mathfrak{F}_n$. Außerdem sei $(X_n)_{n\in\mathbb{N}_0}$ eine adaptierte Folge von Zufallsgrößen.

Seien $a < b$ beliebig, aber fest.
Wir definieren Stoppzeiten $0 := \tau_0 \leq \sigma_1 \leq \tau_1 \leq \sigma_2 \leq \tau_2 \leq \ldots$ wie folgt:

$$\begin{aligned} \sigma_k &:= \inf\{n \geq \tau_{k-1} : X_n \leq a\}, \\ \tau_k &:= \inf\{n \geq \sigma_k : X_n \geq b\} \qquad (k \in \mathbb{N}). \end{aligned}$$



$(\tau_{k-1}, \sigma_k]$ „Abkreuzung"
$(\sigma_k, \tau_k]$ „Aufkreuzung"

Sei weiter definiert $U_n^{a,b} := \sup\{k \in \mathbb{N}_0 : \tau_k \leq n\}$ als die Anzahl der bis zum Zeitpunkt n vollendeten Aufkreuzungen.

Lemma 4.25. *(Upcrossing-Lemma) Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal, so gilt*

$$\mathbb{E}\, U_n^{a,b} \leq \frac{\mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+}{b - a}, \quad n \in \mathbb{N}_0.$$

Beweis. Wir beginnen mit einigen Vorbetrachtungen.

$$\text{Sei } H_n := \sum_{k=1}^{\infty} \mathbb{1}_{\sigma_k < n \leq \tau_k} = \begin{cases} 1 \\ 0 \end{cases} n \text{ befindet sich in } \begin{cases} \text{Aufkreuzungsschleife} \\ \text{Abkreuzungsschleife.} \end{cases}$$

H_n ist previsibel, denn $H_0 = 0$ und

$$\{\sigma_k < n \leq \tau_k\} = \{\sigma_k \leq n-1\} \cap \{\tau_k \leq n-1\}^c \in \mathfrak{F}_{n-1}.$$

$Y_n := X_n \vee a$ ist ein Submartingal als Maximum zweier Submartingale (vergleiche dazu Satz 4.9 (iii)). Dann ist wegen $1 - H_n \geq 0$ auch das stochastische Integral

$$((1 - H) \circ Y)_n = \sum_{k=1}^{n} (1 - H_k)(Y_k - Y_{k-1})$$

ein in 0 startendes Submartingal (Übungsaufgabe) und insbesondere

$$\mathbb{E}((1-H)\circ Y)_n \geq 0.$$

Andererseits ist

$$\begin{aligned}(H\circ Y)_n &= \sum_{k=1}^{n} H_k(Y_k - Y_{k-1}) \\ &\geq \sum_{l=1}^{U_n^{a,b}} \sum_{k=\sigma_l+1}^{\tau_l} (Y_k - Y_{k-1}) = \sum_{l=1}^{U_n^{a,b}} (Y_{\tau_l} - Y_{\sigma_l}) \\ &\geq (b-a)U_n^{a,b}.\end{aligned}$$

Dabei wurde verwendet, dass für $\sigma_l < j \leq \tau_l$ gilt

$$\sum_{k=\sigma_l+1}^{j} (Y_k - Y_{k-1}) = Y_j - Y_{\sigma_l} = (X_j \vee a) - a \geq 0$$

und

$$Y_{\tau_l} - Y_{\sigma_l} = (\underbrace{X_{\tau_l}}_{\geq b} \vee a) - (\underbrace{X_{\sigma_l}}_{\leq a} \vee a) \geq b-a.$$

Zusammenfassend erhalten wir

$$\begin{aligned}\mathbb{E}(X_n-a)^+ - \mathbb{E}(X_0-a)^+ &= \mathbb{E}(Y_n - Y_0) \\ &= \mathbb{E}(H\circ Y)_n + \mathbb{E}((1-H)\circ Y)_n \\ &\geq \mathbb{E}(H\circ Y)_n \\ &\geq (b-a)\mathbb{E}\,U_n^{a,b}.\end{aligned}$$

□

Satz 4.26. *(Martingalkonvergenzsatz)*
Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal mit $\sup_n \mathbb{E}X_n^+ < \infty$. Dann existiert eine $\mathfrak{F}_\infty$-messbare Zufallsgröße X_∞, für die gilt $X_\infty \in \mathcal{L}^1(\mathbb{P})$ und

$$X_n \to X_\infty \textit{ fast sicher}$$

Beweis. Für $a < b$ gilt

$$U_n^{a,b} \uparrow U^{a,b} \text{ Gesamtzahl der Aufkreuzungen des Intervalls } [a,b]$$

Das Upcrossing-Lemma liefert $((x-a)^+ \leq |a| + x^+)$

$$\mathbb{E}U_n^{a,b} \leq \frac{1}{b-a}(|a| + \mathbb{E}X_n^+),$$

weshalb durch monotonen Grenzübergang

$$\mathbb{E}U^{a,b} \leq \frac{1}{b-a}\left(|a| + \sup_n \mathbb{E}X_n^+\right) < \infty$$

folgt. Insbesondere ist $\mathbb{P}(U^{a,b} < \infty) = 1$.
Also ist

$$\{X_n \text{ konvergiert nicht}\} = \{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\} = \bigcup_{\substack{a,b\in\mathbb{Q} \\ a<b}} \{U^{a,b} = \infty\}$$

als abzählbare Vereinigung von Nullmengen wieder eine Nullmenge. Also existiert eine $\overline{\mathbb{R}}$-wertige Zufallsgröße X_∞ mit

$$X_n \to X_\infty \text{ fast sicher}$$

Da alle X_n $\mathfrak{F}_\infty$-messbar sind, kann X_∞ $\mathfrak{F}_\infty$-messbar gewählt werden. Bleibt $X_\infty \in \mathcal{L}^1(\mathbb{P})$ zu zeigen. Mit dem Lemma von Fatou erhalten wir

$$\mathbb{E}X_\infty^+ \leq \liminf_{n\to\infty} \mathbb{E}X_n^+ \leq \sup_n \mathbb{E}X_n^+ < \infty$$

und da X_n ein Submartingal ist

$$\begin{aligned} \mathbb{E}X_\infty^- &\leq \liminf_{n\to\infty} \mathbb{E}X_n^- = \liminf_{n\to\infty}\left(\mathbb{E}X_n^+ - \underbrace{\mathbb{E}X_n}_{\geq \mathbb{E}X_0}\right) \\ &\leq \sup_n \mathbb{E}X_n^+ - \mathbb{E}X_0 < \infty. \end{aligned}$$

□

Folgerung 4.27. *Ist (X_n) ein nichtnegatives Supermartingal, so existiert eine nichtnegative $\mathfrak{F}_\infty$-messbare Zufallsgröße X_∞ mit*

$$X_n \to X_\infty \textit{ fast sicher} \quad \textit{und} \quad \mathbb{E}X_\infty \leq \mathbb{E}X_0.$$

Satz 4.28. *(Ohne Beweis) Ist (X_n) ein gleichgradig integrierbares Martingal (Sub-,Supermartingal), so existiert eine $\mathfrak{F}_\infty$-messbare Zufallsgröße $X_\infty \in \mathcal{L}^1$ mit*

$$X_n \to X_\infty \textit{ fast sicher und in } \mathcal{L}^1(\mathbb{P}).$$

Außerdem gilt

$$X_n = \mathbb{E}[X_\infty|\mathfrak{F}_n] \textit{ fast sicher} \quad (n \in \mathbb{N}),$$

falls X ein Martingal ist („$\leq$“ für Submartingale, „$\geq$“ für Supermartingale).

Bemerkung 4.29.

1. Ist $Z \in \mathcal{L}^1(\mathbb{P})$, so ist die Folge $Z_n = \mathbb{E}[Z|\mathfrak{F}_n]$ gleichgradig integrierbar.

2. Für eine beliebige Zufallsgröße $Z \in \mathcal{L}^1(\mathbb{P})$ gilt
$$\mathbb{E}[Z|\mathfrak{F}_n] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[Z|\mathfrak{F}_\infty] \text{ fast sicher und in } \mathcal{L}^1.$$

Satz 4.30. *(Konvergenzsatz für Rückwärtsmartingale)(Ohne Beweis) Sei $(X_n)_{n\in-\mathbb{N}_0}$ ein Martingal bezüglich einer Filtration $(\mathfrak{F}_n)_{n\in-\mathbb{N}_0}$. Dann gilt*
$$X_n \to \mathbb{E}[X_0|\mathfrak{F}_{-\infty}] \textit{ fast sicher und in } \mathcal{L}^1$$
mit
$$\mathfrak{F}_{-\infty} := \bigcap_{n\in-\mathbb{N}_0} \mathfrak{F}_n.$$

Bemerkung 4.31. Aus der Martingaleigenschaft folgt sofort
$$X_n = \mathbb{E}[X_0|\mathfrak{F}_n] \text{ für } n < 0.$$
Zu zeigen bleibt, dass dieser bedingte Erwartungswert in geeignetem Sinne gegen $\mathbb{E}[X_0|\mathfrak{F}_{-\infty}]$ konvergiert.

Kapitel 5

Stochastische Prozesse

5.1. Definition und Verteilung

Ein weiteres wichtiges Thema sind stochastische Prozesse, worauf wir in diesem Kapitel näher eingehen wollen. Ein stochastischer Prozess ist ein stochastisches Modell für einen zufälligen Zeitablauf oder eine zufällige Bewegung. Jedem Zeitpunkt wird hierbei eine Zufallsvariable zugeordnet (z.B. Temperatur, Ort, Aktienkurs, ...).
Seien nun zur weiteren Betrachtung der Prozesstheorie ein Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, \mathbb{P})$, ein messbarer Raum $(E, \mathfrak{E})$ und eine nichtleere Indexmenge I gegeben.

Definition 5.1. Eine Familie $\mathbb{X} = (X_t)_{t \in I}$ von $(E, \mathfrak{E})$-wertigen Zufallsvariablen auf $(\Omega, \mathfrak{F}, P)$ heißt *stochastischer Prozess.*

Dabei bezeichnet $(E, \mathfrak{E})$ den *Zustandsraum* und I die *Zeitmenge* (Indexmenge) des stochastischen Prozesses $\mathbb{X}$.

Des weiteren können wir festhalten:

$\forall\ t \in I : \omega \longmapsto X_t(\omega)$	Die Abbildung ist eine $(E, \mathfrak{E})$-wertige Zufallsvariable
$\forall\ \omega \in \Omega : t \longmapsto X_t(\omega)$	Die Abbildung ist der *Pfad*, die *Trajektorie* des Prozesses im Zustandsraum E
I=$\{1, \ldots, n\}$ endliche Menge	$\mathbb{X} = (X_1, \ldots, X_n)$ ist ein n-dimensionaler Zufallsvektor
$I = \mathbb{N}$	$\mathbb{X} = (X_n)$ ist eine zufällige Folge
$I = \mathbb{N},\ \mathbb{N}_0,\ \mathbb{Z}, \ldots$	Prozess in *diskreter* Zeit
$I = \mathbb{R},\ \mathbb{R}_+,\ [0, T]$	Prozess in *stetiger* (kontinuierlicher) Zeit

$I \subseteq \mathbb{Z}^d,\ \mathbb{R}^d$	*zufälliges Feld* („mehrdimensionale Zeit“)
$(\mathbb{E}, \mathfrak{E}) = (\mathbb{R},\ \mathfrak{B})$	*reellwertiger* Prozess
$(\mathbb{E}, \mathfrak{E}) = (\mathbb{R}^d,\ \mathfrak{B}^d)$	*d-dimensionaler* Prozess

Ein stochastischer Prozess $\mathbb{X} = (X_t)_{t\in I}$ kann als *eine* Zufallsvariable $\mathbb{X}$ angesehen werden:

$$\Omega \ni \omega \longmapsto \mathbb{X}(\omega) = (X_t(\omega))_{t\in I} \in E^I$$

mit

$$\begin{aligned} E^I \quad &:= \quad \text{Menge aller Abbildungen } I \longrightarrow E \\ &= \quad \text{Menge aller Familien } (X_t)_{t\in I} \text{ mit } X_t \in E \text{ für alle } t \in I. \end{aligned}$$

Dazu definitiert man sich eine σ-Algebra $\mathfrak{E}^I$, so dass $\mathbb{X}$ eine $(E^I, \mathfrak{E}^I)$-wertige Zufallsvariable wird:

$$\begin{array}{rclll} \pi_t : E^I & \longrightarrow & E & & \text{Diese Abbildung ist die Projektion} \\ (x_s)_{s\in I} & \longmapsto & x_t & & \text{auf die t-te Koordinate} \end{array}$$

$\mathfrak{E}^I{:=}\sigma(\pi_t;\ t \in I)$ ist hierbei die kleinste σ-Algebra auf E^I, bezüglich der alle Projektionen $\pi_t,\ t \in I$, messbar sind.

Bemerkung 5.2.

1. Die σ-Algebra $\mathfrak{E}$ wird von Mengen der Gestalt

$$\pi_{t_1}^{-1}(A_{t_1}) \cap \cdots \cap \pi_{t_n}^{-1}(A_{t_n}) = \bigtimes_{t\in I} A_t$$

   mit $A_t := E$ für $t \notin \{t_1, \ldots, t_n\}$, $(t_1, \ldots, t_n) \in I$ und $A_{t_1}, \ldots, A_{t_n} \in \mathfrak{E}$ erzeugt.

2. Ist $I = \{1, \ldots, n\}$ endlich, so gilt insbesondere

$$\mathfrak{E}^I = \mathfrak{E} \otimes \cdots \otimes \mathfrak{E} \qquad \text{(n-mal)}$$

   Das bedeutet bei $\mathfrak{E}^I$ handelt es sich um eine Verallgemeinerung des Begriffs der Produkt-σ-Algebra auf unendlich viele Faktoren.

Bemerkung 5.3. Sei $\mathbb{X} = (X_t)_{t\in I}$ eine Familie von Abbildungen mit $X_t : \Omega \longrightarrow E$. Dann gilt:

> $\mathbb{X}$ ist ein stochastischer Prozess genau dann, wenn $\mathbb{X}$ eine $(E^I, \mathfrak{E}^I)$-wertige Zufallsvariable ist.

Beweis. Es gilt $X_t = \pi_t \circ \mathbb{X}$.

„$\Longrightarrow$“:
Die Mengen $\pi_t^{-1}(A)$ $(t \in I, A \in \mathfrak{E})$ bilden ein Erzeugendensystem von $\mathfrak{E}^I$.
Da X_t $\mathfrak{F}-\mathfrak{E}$-messbar ist, gilt dann $\mathbb{X}^{-1}(\pi_t^{-1}(A)) = (\pi_t \circ \mathbb{X})^{-1}(A) = X_t^{-1}(A) \in \mathfrak{F}$.
Also ist $\mathbb{X}$ $\mathfrak{F}$-$\mathfrak{E}^I$-messbar.

„$\Longleftarrow$“:
Für jedes t ist $X_t = \pi_t \circ \mathbb{X} : (\Omega, \mathfrak{F}) \xrightarrow{\mathbb{X}} (E^I, \mathfrak{E}^I) \xrightarrow{\pi_t} (E, \mathfrak{E})$ als Komposition messbarer Abbildungen wieder messbar, das heißt eine Zufallsvariable. $\square$

Definition 5.4. Das Bildmaß $\mathbb{P}_{\mathbb{X}} = \mathbb{P} \circ \mathbb{X}^{-1}$ auf $(E^I, \mathfrak{E}^I)$ heißt (unendlichdimensionale) *Verteilung des stochastischen Prozesses* $\mathbb{X}$.

Behauptung 5.5. *(ohne Beweis) (Kanonisches Modell eines stochastischen Prozesses* $\mathbb{X}$*).*
Sei $\mathbb{X} = (X_t)_{t \in I}$ *ein stochastischer Prozess mit Zustandsraum* $(E, \mathfrak{E})$*. Dann stimmt die Verteilung von* $\mathbb{X}$ *mit der Verteilung des auf dem Wahrscheinlichkeitsraum* $(E^I, \mathfrak{E}^I, \mathbb{P}_{\mathbb{X}})$ *gegebenen stochastischen Prozesses* $\pi = (\pi_t)_{t \in I}$ *überein.*

Bemerkung 5.6. Da man sich in der Regel nur für die Verteilungseigenschaften stochastischer Prozesse interessiert, nimmt man deshalb oft an, dass der Prozess in *„kanonischer Form“* $\pi = (\pi_t)_{t \in I}$ auf $(E^I, \mathfrak{E}^I, \mathbb{P}_{\mathbb{X}})$ gegeben ist.

Da die unendlich-dimensionale Verteilung eines stochastischen Prozesses meist nicht explizit angegeben werden kann, versucht man sie durch sogenannte endlich-dimensionale Verteilungen zu charakterisieren.
Wenn also $J \subseteq I$ eine (nichtleere) *endliche* Teilmenge der Indexmenge I ist, dann ist

$$\begin{array}{rcll} \pi_J : E^I & \longrightarrow & E^J & \quad \text{eine } \textit{endlich-dimensionale Projektion} \\ (x_t)_{t \in I} & \longmapsto & (x_t)_{t \in J} & \end{array}$$

Sind J und K zwei endliche Teilmengen von I mit $J \subseteq K$, so definieren wir entsprechend die Projektion

$$\pi_J^K : E^K \longrightarrow E^J.$$

Wir schreiben π_J^K für $J = \{t\}$, $t \in K$. Dann ist π_J $\mathfrak{E}^I$-$\mathfrak{E}^J$-messbar.
($\mathfrak{E}^J$ wird von Mengen der Gestalt $(\pi_t^J)^{-1}(A)$ $\quad (t \in J, A \in \mathfrak{E})$ erzeugt und es gilt $\pi_J^{-1}((\pi_t^J)^{-1}(A)) = (\pi_t^J \circ \pi_J)^{-1}(A) = \pi_t^{-1}(A) \in \mathfrak{E}^I$.) Entsprechend ist π_J^K auch $\mathfrak{E}^K$-$\mathfrak{E}^J$-messbar.
Mit $\mathbb{P}_{\mathbb{X}}$ bezeichnen wir die Verteilung von $\mathbb{X}_J = (X_t)_{t \in J}$ auf dem messbaren Raum $(E^J, \mathfrak{E}^J)$.

Definition 5.7. Die Wahrscheinlichkeitsmaße $\mathbb{P}_{\mathbb{X}_J}$, wobei J alle endlichen Teilmengen von I durchläuft, heißen *endlich-dimensionale* Verteilungen des Prozesses $\mathbb{X}$.

Behauptung 5.8. *Sei $\mathbb{X} = (X_t)_{t \in I}$ ein stochastischer Prozess mit Zustandsraum $(E, \mathfrak{E})$.*

(a) Für jede endliche Teilmenge J von I gilt

$$\mathbb{P}_{\mathbb{X}_J} = \mathbb{P}_{\mathbb{X}} \circ \pi_J^{-1}. \tag{5.1}$$

(b) Für je zwei endliche Teilmengen J und K von I mit $J \subseteq K$ gilt

$$\mathbb{P}_{\mathbb{X}_J} = \mathbb{P}_{\mathbb{X}_K} \circ (\pi_J^K)^{-1}. \tag{5.2}$$

Beweis.

(a) Wegen $\mathbb{X}_J = \pi_J \circ \mathbb{X}$ ist $\mathbb{P}_{\mathbb{X}_J} = \mathbb{P} \circ \mathbb{X}_J^{-1} = (\mathbb{P} \circ \mathbb{X}^{-1}) \circ \pi_J^{-1} = \mathbb{P}_{\mathbb{X}} \circ \pi_J^{-1}$

(b) Wegen $\mathbb{X}_J = \pi_J^K \circ \mathbb{X}_K$ folgt die Behauptung analog zu (i).

□

Die unendlich-dimensionale Verteilung eines stochastischen Prozesses bestimmt via (5.1) die endlich-dimensionalen Verteilungen eindeutig, und diese müssen die *Kolmogorovsche Verträglichkeitsbedingung (Kompatibilitätsbedingung)* (5.2) erfüllen.
Umgekehrt bestimmen aber auch die endlich-dimensionalen Verteilungen die unendlich-dimensionale Verteilung eindeutig.

Behauptung 5.9. *Für jedes endliche $J \subseteq I$ sei ein Wahrscheinlichkeitsmaß $\mathbb{P}_J$ auf $(E^I, \mathfrak{E}^I)$ gegeben. Dann existiert höchstens ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(E^I, \mathfrak{E}^I)$ mit*

$$\mathbb{P}_J = \mathbb{P} \circ \pi_J^{-1} \qquad \textit{für alle endlichen } J \subseteq I.$$

Beweis. Wir benutzen, dass das Wahrscheinlichkeitsmaß $\mathbb{P}$ durch seine Werte auf einem $\cap$-stabilen Erzeuger der σ-Algebra $\mathfrak{E}^I$ eindeutig festgelegt ist.
Wir zeigen zunächst, dass die Mengen der Gestalt

$$\pi_J^{-1}(A) \ (J \subseteq I \text{ endlich, } A \in \mathfrak{E}^I)$$

einen $\cap$-stabilen Erzeuger von $\mathfrak{E}^I$ bilden. Da diese Mengen auf einpunktigen J nach Definition $\mathfrak{E}^I$ erzeugen und diese Mengen (wegen der Messbarkeit von π_J) alle zu $\mathfrak{E}^I$ gehören, erzeugen sie diese σ-Algebra. Um die Durchschnittsstabilität zu zeigen, seien $J, K \subseteq I$ endlich und $A \in \mathfrak{E}^I, B \in \mathfrak{E}^K$. Es ist zu zeigen, dass ein endliches $L \subseteq I$ und ein $C \in \mathfrak{E}^L$ existiert mit

$$\pi_J^{-1}(A) \cap \pi_K^{-1}(B) = \pi_L^{-1}(C)$$

Setzen wir $L := J \cup K$, $\widetilde{A} := (\pi_J^L)^{-1}(A)$, $\widetilde{B} := (\pi_K^L)^{-1}(B)$ so ist die Beziehung für $C := \widetilde{A} \cap \widetilde{B}$ erfüllt. Aber die Werte von $\mathbb{P}$ auf dem obigen Erzeuger sind durch die Maße $\mathbb{P}_J$ eindeutig festgelegt:

$$\mathbb{P}(\pi_J^{-1}(A)) = (\mathbb{P} \circ \pi_J^{-1})(A) = \mathbb{P}_J(A).$$

□

5.2. Konstruktion stochastischer Prozesse

In diesem Abschnitt wollen wir uns mit der Frage beschäftigen, ob man zu vorgegebenen Verteilungen $\mathbb{P}_J$ auf $(E^J, \mathfrak{E}^J)$, mit $J \subseteq I$ endlich, einen stochastischen Prozess „konstruieren“ kann, dessen endlich-dimensionale Verteilungen mit den $\mathbb{P}_J$ übereinstimmen.

Satz 5.10. *(Kolmogorov)*
Sei E ein vollständiger separabler metrischer Raum („Polnischer Raum“) versehen mit der σ-Algebra $\mathfrak{E}$ der Borelmengen. Weiterhin sei I eine beliebige nichtleere Indexmenge. Für jede endliche Menge $J \subseteq I$ sei ein Wahrscheinlichkeitsmaß $\mathbb{P}_J$ auf $(E^J, \mathfrak{E}^J)$ derart gegeben, dass die Kolmogorovsche Verträglichkeitsbedingung (Kompatibilitätsbedingung)

$$\mathbb{P}_J = \mathbb{P}_K \circ (\pi_J^K)^{-1} \textit{ für } J \subseteq K \subseteq I,\ J, K \textit{ endlich},$$

erfüllt ist. Dann existiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(E^I, \mathfrak{E}^I)$ mit

$$\mathbb{P}_J = \mathbb{P} \circ \pi_J^{-1} \textit{ für endliches } J \subseteq I.$$

Folgerung 5.11. *Unter den Vorrausetzungen des Satzes von Kolmogorov ist der kanonische Prozess $\pi = (\pi_t)_{t \in I}$ auf dem Wahrscheinlichkeitsraum $(E^I, \mathfrak{E}^I, \mathbb{P})$, ein stochastischer Prozess, dessen endlich dimensionale Verteilungen mit den $\mathbb{P}_J$, $J \subseteq I$ endlich, übereinstimmen.*

Beweis. Die endlich-dimensionale Verteilung von $\pi = (\pi_t)_{t \in I}$ bezüglich der endlichen Indexmenge $J \subseteq I$ ist nach Definition die Verteilung der $(E^J, \mathfrak{E}^J)$-wertigen Zufallsvariablen $\pi_J = (\pi_t)_{t \in I}$ auf dem Wahrscheinlichkeitsraum $(E^I, \mathfrak{E}^I, \mathbb{P})$, das heißt $\mathbb{P} \circ \pi_J^{-1}$. □

Beweis. (Satz 5.10) Die Eindeutigkeit von $\mathbb{P}$ wurde in Behauptung 5.8 gezeigt. Zum Beweis der Existenz benutzen wir den Fortsetzungssatz von Carathéodory (Satz 1.14).

1^o Für jedes endliche $J \subseteq I$ ist $\mathfrak{Z}_J := \pi_J^{-1}(\mathfrak{E}^J)$ eine Teil-σ-Algebra von $\mathfrak{E}^I$ nämlich die *σ-Algebra der J-Zylinder*. Es gilt

$$\mathfrak{Z}_J \subseteq \mathfrak{Z}_K \textit{ für } J \subseteq K \textit{ endlich}. \tag{5.3}$$

(Für jedes $B \in \mathfrak{E}^J$ ist $\pi_J^{-1}(B) = \pi_K^{-1}((\pi_J^K)^{-1}(B))$.)
Deshalb ist

$$\mathfrak{Z} := \bigcup_{J \text{ endlich}} \mathfrak{Z}_J \text{ eine Algebra } \textit{(Algebra der Zylindermengen)}.$$

Diese Algebra erzeugt die σ-Algebra $\mathfrak{E}^I$:

$$\mathfrak{E}^I = \sigma(\mathfrak{Z}).$$

2^o Wir definieren $\mathbb{P}$ zunächst auf der Algebra $\mathfrak{Z}$ und zeigen, dass diese Mengenfunktion *endlich-additiv* ist. Wir setzen

$$\mathbb{P}(\pi_J^{-1}(A)) := \mathbb{P}_J(A) \qquad (J \subseteq I \text{ endlich, } A \in \mathfrak{E}^J).$$

Auf Grund der Verträglichkeitsbedingung $(*)$ ist $\mathbb{P}$ tatsächlich wohldefiniert. Seien nun $J, K \subseteq I$ endlich, $A \in \mathfrak{E}^J,\ B \in \mathfrak{E}^K$. Dann gilt

$$\begin{aligned} \pi_J^{-1}(A) = \pi_K^{-1}(B) \quad &\Rightarrow \quad (\pi_J^{J\cup K})^{-1}(A) = (\pi_K^{J\cup K})^{-1}(B) \\ &\Rightarrow \quad \mathbb{P}_J(A) \underset{(*)}{=} \mathbb{P}_{J\cup K}((\pi_J^{J\cup K})^{-1}(A)) \\ &= \quad \mathbb{P}_{J\cup K}((\pi_K^{J\cup K})^{-1}(B)) \underset{(*)}{=} \mathbb{P}_K(B). \end{aligned}$$

Offenbar ist $\mathbb{P}(Z) \geq 0$ für alle $Z \in \mathfrak{Z},\ \mathbb{P}(\emptyset)=0$ und $\mathbb{P}(E^I) = 1$. Die endliche Additivität ergibt sich wie folgt:
Seien $J, K \subseteq I$ endlich, $A \in \mathfrak{E}^I, B \in \mathfrak{E}^K$ und $\pi_J^{-1}(A) \cap \pi_K^{-1}(B) = \emptyset$
Dann gilt

$$\pi_J^{-1}(A) \cup \pi_K^{-1}(B) = \pi_{J\cup K}^{-1}(\widetilde{A}) \cup \pi_{J\cup K}^{-1}(\widetilde{B}) = \pi_{J\cup K}^{-1}(\widetilde{A} \cup \widetilde{B})$$

mit $\widetilde{A} := (\pi_J^{J\cup K})^{-1}(A)$ und $\widetilde{B} := (\pi_K^{J\cup K})^{-1}(B)$ und $\widetilde{A} \cap \widetilde{B} = \emptyset$.
Deshalb ist wegen der endlichen Additivität von $\mathbb{P}_{J\cup K}$ $(**)$ und der Verträglichkeitsbedingung $(*)$

$$\begin{aligned} \mathbb{P}(\pi_J^{-1}(A) \cup \pi_K^{-1}(B)) \quad &= \quad \mathbb{P}_{J\cup K}(\widetilde{A} \cup \widetilde{B}) \\ &\underset{(**)}{=} \quad \mathbb{P}_{J\cup K}(\widetilde{A}) + \mathbb{P}_{J\cup K}(\widetilde{B}) \\ &\underset{(*)}{=} \quad \mathbb{P}_J(A) + \mathbb{P}_K(B) \\ &= \quad \mathbb{P}(\pi_J^{-1}(A)) + \mathbb{P}(\pi_K^{-1}(B)). \end{aligned}$$

3^o Um den Satz von Carathéodory anwenden zu können, bleibt nur noch die σ-Additivität von $\mathbb{P}$ auf der Algebra $\mathfrak{Z}$ nachzuweisen. (Dies lässt sich nicht wie beim Beweis der endlichen Additivität machen, da die abzählbare Vereinigung endlicher Indexmengen im Allgemeinen nicht mehr endlich ist.) Da $\mathbb{P}$ endlich additiv ist, genügt es hierzu zu zeigen, dass folgendes gilt:

$$Z_n \in \mathfrak{Z},\ Z_n \downarrow \emptyset \Rightarrow \mathbb{P}(Z_n) \longrightarrow 0. \tag{5.4}$$

Hierzu werden wir ein *Kompaktheitsargument* benutzen.

Wir benutzen folgende Tatsachen über Polnische Räume:

(i) Sei J endlich und E ein vollständig separabler metrischer Raum mit der Borel-σ-Algebra $\mathfrak{E}$. Dann kann E^J wieder zu einem vollständigen separablen metrischen Raum gemacht werden, so dass dessen Borel-σ-Algebra mit der Produkt-σ-Algebra $\mathfrak{E}^J$ übereinstimmt und die Projektion $\pi_t^J : E^J \to E, t \in J$, stetig sind.

(ii) Jedes Wahrscheinlichkeitsmaß μ auf einem vollständigen separablen metrischen Raum E ist *regulär von innen*:
$\forall\, B \in \mathfrak{E}\ \forall\, \varepsilon > 0$ gibt es eine kompakte Menge $K \subseteq B$ sodass:

$$\mu(B) - \mu(K) = \mu(B \setminus K) < \varepsilon.$$

(siehe zum Beispiel H.Bauer, Maß und Integrationstheorie, Paragraph 26).

Zum Beweis von (5.4) gehen wir indirekt vor.
Hierzu sei $Z_n \in \mathfrak{Z},\ Z_n \downarrow$ und $\mathbb{P}(Z_n) \geq \alpha > 0\ \forall\, n$. Wir zeigen, dass dies

$$\bigcap_n Z_n \neq \emptyset$$

impliziert. Die Zylindermengen Z_n besitzen die Gestalt

$$Z_n = \pi_{J_n}^{-1}(A_n) \text{ mit } J_n \text{ endlich und } A_n \in \mathfrak{E}^{J_n}.$$

Wegen (5.3) können wir dann o.B.d.A

$$J_1 \subseteq J_2 \subseteq J_3 \subseteq \ldots$$

vorraussetzen. Auf Grund der inneren Regularität von $\mathbb{P}_{J_n}$ existieren kompakte Mengen $K_n \subseteq A_n$ mit

$$\mathbb{P}_{J_n}(A_n) - \mathbb{P}_{J_n}(K_n) < \alpha 2^{-n}.$$

Wir setzen

$$Z_n'' := \pi_{J_n}^{-1}(K_n) \text{ (im Allgemeinen nicht monoton)}$$

und

$$Z_n' := Z_1'' \cap \cdots \cap Z_n''.$$

Dann ist $Z_n' \downarrow$ und $Z_n' \subseteq Z_n$. Wir erhalten mit Hilfe der endlichen Subadditivität $(*)$

$$\begin{aligned}
\mathbb{P}(Z_n \setminus Z_n') &= \mathbb{P}(\bigcup_{k=1}^{n}(Z_n \setminus Z_k'')) \leq \mathbb{P}(\bigcup_{k=1}^{n}(Z_k \setminus Z_k'')) \\
&\underset{(*)}{\leq} \sum_{k=1}^{n} \mathbb{P}(Z_k \setminus Z_k'') = \sum_{k=1}^{n} \mathbb{P}_{J_k}(A_k \setminus K_k) \\
&< \sum_{k=1}^{n} \alpha 2^{-k} < \alpha.
\end{aligned}$$

Da aber nach Vorraussetzung $\mathbb{P}(Z_n) \geq \alpha$ ist, muss

$$Z_n' \neq \emptyset \qquad \text{für alle } n \text{ gelten.}$$

Also existiert zu jedem n ein

$$y^{(n)} \in Z'_n \subseteq Z''_m \qquad (m = 1, \ldots, n),$$

weshalb

$$y^{(n)}_{J_m} := \pi_{J_m}(y^{(n)}) \in K_m \qquad (m = 1, \ldots, n).$$

Deshalb existieren eine Teilfolge $(y^{(n_k)})$ und ein $y_J = (y_t)_{t \in J} \in E^J$ mit

$$y^{(n_k)}_t \longrightarrow y_t \ \forall\, t \in J := \bigcup_m J_m$$

und

$$y_{J_n} \in K_n \qquad \text{für alle } n.$$

(Man muss sukzessive für $m = 1, 2, \ldots$ Teilfolgen von Teilfolgen auswählen, deren Projektionen in K_m konvergieren und anschließend mit dem Diagonalisierungsverfahren eine Teilfolge $(y^{(n_k)})$ auswählen, die Teilfolge aller dieser Teilfolgen ist.)
Wir setzen $(y_t)_{t \in J} \in E^J$ beliebig zu einem $y = (y_t)_{t \in I} \in \mathfrak{E}^J$ fort. Dann gilt

$$y \in Z'_n \subseteq Z_n \qquad \text{für alle } n$$

und folglich $\bigcap_n Z_n \neq \emptyset$.

$\square$

Oft interessiert man sich auch für die *topologischen Eigenschaften der Pfade* stochastischer Prozesse (Stetigkeit, Rechtsstetigkeit, Differenzierbarkeit etc.).

Definition 5.12. Ein stochastischer Prozess $\widetilde{\mathbb{X}} = (\widetilde{X}_t)_{t \in I}$ heißt *Modifikation* des stochastischen Prozesses $\mathbb{X} = (X_t)_{t \in I}$, falls

$$\widetilde{X}_t = X_t \text{ fast sicher für jedes } t \in I$$

gilt.

Bemerkung 5.13. Ist $\widetilde{\mathbb{X}}$ eine Modifikation von $\mathbb{X}$, so besitzen $\mathbb{X}$ und $\widetilde{\mathbb{X}}$ die gleiche Verteilung. Tatsächlich ist für beliebige $t_1, \ldots, t_n \in I$

$$(\widetilde{X}_{t_1}, \ldots, \widetilde{X}_{t_n}) = (X_{t_1}, \ldots, X_{t_n}) \qquad \text{fast sicher}$$

und folglich

$$\mathbb{P}((\widetilde{X}_{t_1}, \ldots, \widetilde{X}_{t_n}) \in A) = \mathbb{P}((X_{t_1}, \ldots, X_{t_n}) \in A), \qquad \text{für alle } A \in \mathfrak{E}^n.$$

Also stimmen die endlich-dimensionalen Verteilungen (und damit auch die unendlich-dimensionalen Verteilungen) von $\mathbb{X}$ und $\widetilde{\mathbb{X}}$ überein.

Satz 5.14. *(Kolmogorov-Čentsov)*
Sei $\mathbb{X} = (X_t)_{t\in[0,T]}$ ein reellwertiger Prozess. Angenommen, es existieren positive Konstanten γ, c, ε mit

$$\mathbb{E}\,|X_s - X_t|^\gamma \leq c\,|s-t|^{1+\varepsilon}, \qquad s,t \in [0,T]. \tag{5.5}$$

Dann existiert eine Modifikation $\widetilde{\mathbb{X}} = (\widetilde{X}_t)_{t\in[0,T]}$ von X, deren Pfade stetig sind.

Beweis. Sei o.B.d.A. $T = 1$.
Sei $D_m := \{i2^{-m} := 0,1,\dots,2^m\}$. Dann ist $D := \bigcup\limits_m D_m$ die Menge aller dyadischen Zahlen in $[0,1]$.

$$\Delta_m := \left\{(s,t):\ s,t \in D_m, s<t, |s-t| = 2^{-m}\right\}$$
$$K_i := \sup_{(s,t)\in\Delta_i} |X_s - X_t|\,.$$

Wir erhalten

$$\mathbb{E}K_i^\gamma = \mathbb{E}\sup_{(s,t)\in\Delta_i} |X_s - X_t|^\gamma \leq \sum_{(s,t)\in\Delta_i} \mathbb{E}\,|X_s - X_t|^\gamma\,,$$

woraus mit der Vorraussetzung (5.5)

$$\mathbb{E}K_i^\gamma \leq c2^{-i\varepsilon}\ (i = 1,2,\dots) \tag{5.6}$$

folgt. Seien $s,t \in D$ mit $s < t$ beliebig gewählt. Für jedes m seien s_m und t_m die größten Zahlen aus D_m mit $s_m \leq s$ und $t_m \leq t$. Dann gilt $s_m \uparrow s$ und $t_m \uparrow t$ und $s_m = s,\ t_m = t$ für alle hinreichend großen m. Ist außerdem $|s-t| \leq 2^{-m}$, so ist entweder $s_m = t_m$ oder $(s_m, t_m) \in \Delta_m$.
In beiden Fällen gilt

$$X_s - X_t = \sum_{i=m}^{\infty} (X_{s_{i+1}} - X_{s_i}) + X_{s_m} - X_{t_m} + \sum_{i=m}^{\infty} (X_{t_i} - X_{t_{i+1}})$$

(Man beachte, dass dies nur für endliche Summen gilt.) und folglich

$$|X_s - X_t| \leq 2\sum_{i=m}^{\infty} K_i. \tag{5.7}$$

Für $\alpha > 0$ sei

$$M_\alpha := \sup\left\{\frac{|X_s - X_t|}{|s-t|^\alpha}:\ s,t\in D, s\neq t\right\}.$$

Wir erhalten

$$\begin{aligned}
M_\alpha \quad &:= \quad \sup_m \sup_{\substack{2^{-(m+1)} < |s-t| \leq 2^{-m} \\ s\neq t}} \frac{|X_s - X_t|}{|s-t|^\alpha} \\
&\leq \quad \sup_m \left(2^{(m+1)\alpha} \sup_{|s-t|\leq 2^{-m}} |X_s - X_t|\right) \\
&\underset{(5.7)}{\leq} \quad \sup_m \left((2\cdot 2^{(m+1)\alpha}\sum_{i=m}^{\infty} K_i\right)
\end{aligned}$$

und folglich

$$M_\alpha \leq 2^{\alpha+1} \sum_{i=0}^{\infty} 2^{i\alpha} K_i.$$

Ist $\gamma \geq 1$ so erhalten wir mit der $\mathcal{L}^\gamma(\mathbb{P})$-Norm $\|.\|_\gamma$, der Dreiecksungleichung $(*)$ und da $\|K_i\|_\gamma = (\mathbb{E}K_i^\gamma)^{1/\gamma} \underset{(5.6)}{\leq} c^{1/\gamma} 2^{-i\varepsilon/\gamma}$

$$\begin{aligned}
\|M_\alpha\|_\gamma &\underset{(*)}{\leq} 2^{\alpha+1} \sum_{i=0}^{\infty} 2^{i\alpha} \|K_i\|_\gamma \\
&\leq c^{1/\gamma} 2^{\alpha+1} \sum_{i=0}^{\infty} 2^{i(\alpha-\varepsilon\gamma)} < \infty \qquad \text{für } \alpha < \varepsilon/\gamma.
\end{aligned}$$

Ist $0 < \gamma < 1$ so gilt

$$\begin{aligned}
\mathbb{E}M_\alpha^\gamma &\leq 2^{\gamma(\alpha+1)} \mathbb{E}(\sum_{i=0}^{\infty} 2^{i\alpha} K_i)^\gamma \leq 2^{\gamma(\alpha+1)} \mathbb{E} \sum_{i=0}^{\infty} 2^{i\alpha\gamma} K_i^\gamma \\
&\underset{(5.6)}{\leq} c 2^{\gamma(\alpha+1)} \sum_{i=0}^{\infty} 2^{i(\alpha\gamma-\varepsilon)} < \infty \qquad \textit{für } \alpha < \varepsilon/\gamma.
\end{aligned}$$

In beiden Fällen ist $\mathbb{E}M_\alpha^\gamma < \infty$ und damit

$$M_\alpha < \infty \text{ fast sicher für } \alpha < \varepsilon/\gamma.$$

Also ist für fast alle ω der Pfad $X_.(\omega)$ gleichmäßig stetig auf D. Für diese ω kann man deshalb

$$\widetilde{X}_t(\omega) := \lim_{\substack{s \to t \\ s \in D}} X_s(\omega)$$

definieren, und die Abbildung $t \longrightarrow \widetilde{X}_t(\omega)$ ist stetig auf $[0,1]$.
Bleibt zu zeigen, dass $\widetilde{X}$ eine Modifikation von X ist. Wegen

$$\left|\widetilde{X}_t - X_t\right|^\gamma = \lim_{\substack{s \to t \\ s \in D}} |X_s - X_t|^\gamma \text{ fast sicher}$$

folgt nach dem Lemma von Fatou

$$\mathbb{E}\left|\widetilde{X}_t - X_t\right|^\gamma \leq \liminf_{\substack{s \to t \\ s \in D}} \mathbb{E}|X_s - X_t|^\gamma \underset{(5.5)}{=} 0,$$

das heißt $\widetilde{X}_t = X_t$ fast sicher (für jedes $t \in [0,1]$). $\square$

Bemerkung 5.15.

1. Aus dem Beweis des Satzes ist ersichtlich, dass eine Modifikation $(\widetilde{X}_t)_{t\in[0,T]}$ existiert, deren Pfade Hölder-stetig mit Hölderexponent $\alpha < \varepsilon/\gamma$ sind (das heißt für jedes $\alpha \in (0, \varepsilon/\gamma)$ und jedes $\omega \in \Omega$ existiert eine positive Konstante $c_{\alpha,\omega}$, so dass
$$\left|\widetilde{X}_s(\omega) - \widetilde{X}_t(\omega)\right| \le c_{\alpha,\omega} |s-t|^\alpha, \qquad s,t \in [0,T],$$
gilt).
2. Oft ist (5.5) für $\gamma = 4$ und $\varepsilon = 1$ erfüllt:
$$\mathbb{E}|X_s - X_t|^4 \le c|s-t|^2, \qquad s,t \in [0,T].$$
3. Der Satz gilt auch für Banachraum-wertige Prozesse (mit $\|X_s - X_t\|$ statt $|X_s - X_t|$ und für mehrdimensionale Zeitmengen $[0,T]^d$ (mit $|s-t|^{d+\varepsilon}$ statt $|s-t|^{1+\varepsilon}$).

Sei $\mathcal{C} = \mathcal{C}[0,T]$ der Banachraum der stetigen Funktionen ω:[0,T]$\longrightarrow \mathbb{R}$ mit der Supremumsnorm $\|\omega\| = \sup\limits_{t\in[0,T]} |\omega(t)|$. Mit $\mathfrak{B}_\mathcal{C}$ bezeichnen wir die zugehörige Borel-σ-Algebra. Für $t \in [0,T]$ bezeichne $\pi_t : \mathcal{C} \longrightarrow \mathbb{R}$ die durch $\pi_t(\omega) := \omega(t),\ \omega \in \mathcal{C}$, definierte eindimensionale Projektion.

Satz 5.16. *(Kanonisches Modell eines stochastischen Prozesses) (Übungsaufgabe)*
Sei $\mathbb{X} = (X_t)_{t\in[0,T]}$ ein reellwertiger stochastischer Prozess mit stetigen Pfaden.

(i) Dann ist $\mathbb{X} : \Omega \longrightarrow \mathcal{C}[0,T],\ \omega \longmapsto X_\cdot(\omega)$, eine $(\mathcal{C}, \mathfrak{B}_\mathcal{C})$-wertige Zufallsvariable.

(ii) Sei $\mathbb{P}_\mathbb{X} = \mathbb{P}\circ\mathbb{X}^{-1}$ die zugehörige Verteilung. Dann ist $(\pi_t)_{t\in[0,T]}$ ein Prozess mit stetigen Pfaden auf $(\mathcal{C}, \mathfrak{B}_\mathcal{C}, \mathbb{P}_\mathbb{X})$, dessen endlich-dimensionale Verteilungen auf denen von $\mathbb{X}$ übereinstimmen.

5.3. Beispiele

5.3.1. Folgen von unabhängigen Zufallsvariablen

Sei E ein polnischer Raum und $\mathfrak{E}$ die zugehörige Borel-σ-Algebra. Gegeben sei weiterhin eine Folge (X_n) *unabhängiger* $(E, \mathfrak{E})$-wertiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, \mathbb{P})$ derart, dass μ_n die Verteilung von X_n mit $(n \in \mathbb{N})$ ist, das heißt $\mu_n = \mathbb{P}\circ X_n^{-1}$. Die zugehörigen endlich-dimensionalen Verteilungen besitzen dann die Gestalt

$$\mathbb{P}_{n_1,\dots,n_r} = \mu_{n_1} \otimes \cdots \otimes \mu_{n_r}\ (1 \le n_1 < n_2 < \cdots < n_r) \tag{5.8}$$

(vgl. Satz 2.20). Da sich diese durch Projektion aus $\mathbb{P}_{1,2,\ldots,n_r}$ ergeben, genügt es im wesentlichen,

$$\mathbb{P}_{1,\ldots,n} := \mu_1 \otimes \cdots \otimes \mu_n \qquad (n \in \mathbb{N})$$

zu betrachten. Diese erfüllen die Kolmogorovsche Verträglichkeitsbedingung: Ist $\pi_n : E^{n+1} \longrightarrow E^n$ mit $\pi_n(x_1, \ldots, x_n, x_{n+1}) = (x_1, \ldots, x_n)$ die Projektion von E^{n+1} auf E^n, so gilt

$$\mathbb{P}_{1,\ldots,n} = \mathbb{P}_{1,\ldots,n,n+1} \circ \pi_n^{-1}.$$

Tatsächlich, für $A \in \mathfrak{E}^n$ ist $\pi_n^{-1}(A) = A \times E \in \mathbb{E}^{n+1}$ und

$$(\mu_1 \otimes \cdots \otimes \mu_n \otimes \mu_{n+1})(\pi_n(A)) = (\mu_1 \otimes ... \otimes \mu_n)(A) \cdot \underbrace{\mu_{n+1}(E)}_{=1}.$$

Also existiert nach dem Satz von Kolmogorov (Satz 5.9 und Folgerung 5.10) eine Folge (X_n) von $(E, \mathfrak{E})$-wertigen Zufallsvariablen, deren endlich-dimensionale Verteilungen die Gestalt (5.8) besitzen. Daraus folgt, dass diese Zufallsvariablen unabhängig sind (Übungsaufgabe) und X_n die Verteilung μ_n besitzt $(n \in \mathbb{N})$.

5.3.2. Markovketten

Gegeben seien eine höchstens abzählbare Menge I, eine Verteilung $(p_i^0)_{i \in I}$ und Übergangswahrscheinlichkeiten $(p_{ij})_{i,j \in I}$. Man „konstruiere“ eine homogene Markovkette $(X_n)_{n \in \mathbb{N}_0}$ mit Anfangsverteilung (p_i^0) und Übergangswahrscheinlichkeiten (p_{ij}). Die endlich-dimensionalen Verteilungen dieser Kette lassen sich dann durch deren Einzelwahrscheinlichkeiten beschreiben und besitzen die Gestalt

$$\mathbb{P}_{0,1,\ldots,n}(i_0, i_1, \ldots, i_n) = p_{i_0}^0 p_{i_0 i_1} \ldots p_{i_{n-1} i_n} \qquad (n \in \mathbb{N}_0; i_0, \ldots, i_n \in I).$$

Diese erfüllen die Kolmogorovsche Verträglichkeitsbedingung. Tatsächlich erhält man mit der Projektion $\pi_n : I^{n+2} \longrightarrow I^{n+1}$ für beliebige $A \subseteq I^{n+1}$:

$$\begin{aligned}
\mathbb{P}_{0,\ldots,n,n+1}(\pi_n^{-1}(A)) &= \sum_{\substack{(i_0,\ldots,i_n,i_{n+1}) \\ \in A \times I}} p_{i_0}^0 p_{i_0 i_1} \ldots p_{i_{n-1} i_n} p_{i_n i_{n+1}} \\
&= \sum_{(i_0,\ldots,i_n) \in A} p_{i_0}^0 p_{i_0 i_1} \ldots p_{i_{n-1} i_n} \underbrace{\sum_{i_{n+1} \in I} p_{i_n i_{n+1}}}_{=1} \\
&= \mathbb{P}_{0,1,\ldots,n}(A).
\end{aligned}$$

Also existiert nach dem Satz von Kolmogorov eine Folge $(X_n)_{n \in \mathbb{N}_0}$ I-wertiger Zufallsgrößen auf einem Wahrscheinlichkeitsraum, deren endlich-dimensionale Verteilungen durch (5.9) gegeben sind:

$$\begin{aligned}
&\mathbb{P}(X_0 = i_0, X_1 = i_1, ..., X_n = i_n) \\
= \;& p_{i_0}^0 p_{i_0 i_1} ... p_{i_{n-1} i_n} \qquad (n \in \mathbb{N}_0; i_0, ..., i_n \in I). \qquad (5.9)
\end{aligned}$$

Hieraus folgt leicht, dass (X_n) eine homogene Markovkette mit Zustandsraum I, Anfangswahrscheinlichkeiten (p_i^0) und Übergangswahrscheinlichkeiten (p_{ij}) ist.

5.3.3. Markovprozesse mit diskreter Zeit und allgemeinem Zustandsraum

Sei der Zustandsraum ein Polnischer Raum $(E, \mathfrak{E})$ mit durch ein Wahrscheinlichkeitsmaß μ_0 auf $(E, \mathfrak{E})$ gegebener Anfangsverteilung.
Die Übergangsfunktion sei $P(x, \Gamma)$ $(x \in E, \ \Gamma \in \mathfrak{E})$, ein Markovkern. Die endlich-dimensionalen Verteilungen sollen die Gestalt

$$\mathbb{P}_{0,1,\ldots,n}(A) = \int\int ... \int \mathbb{1}_A(x_0, x_1, ..., x_n) P(x_{n-1}, \, dx_n) ... P(x_0, \, dx_1) \mu^0(\, dx_0) \tag{5.10}$$

mit $n \in \mathbb{N}_0$ und $A \in \mathfrak{E}^{n+1}$ besitzen. Damit ist die Kolmogorovsche Verträglichkeitsbedingung erfüllt. Die Frage die bleibt ist die, wie man in diesem Falle die Markoveigenschaft und die Homogenität formuliert oder wie man diese aus (5.10) ableiten kann.

5.3.4. Gaußprozesse

Definition 5.17. Ein reellwertiger Prozess $(X_t)_{t \in I}$ heißt Gaußprozess, wenn seine endlich-dimensionalen Verteilungen Normalverteiungen sind.
Die Abbildungen $m : I \longrightarrow \mathbb{R}$ und $k : I \times I \longrightarrow \mathbb{R}$ definiert durch $m(t) := \mathbb{E}X_t$ und $k(s,t) := \mathbb{E}(X_s - m(s))(X_t - m(t)), \ s,t \in I$, heißen *Mittelwert-* bzw. *Kovarianzfunktion* des Gaußprozesses.

Die endlich-dimensionalen Verteilungen sind die Verteilungen der Gaußschen Vektoren $(X_{t_1}, ..., X_{t_n})^t$ $(t_1, ..., t_n \in I; t_k \neq t_l)$. Diese besitzen den Mittelwertvektor $(m(t_1), ..., m(t_n))^t$ und die Kovarianzmatrix $(k(t_i, t_j))_{i,j=1}^n$. Da die Kovarianzmatrix eines normalverteilten Vektors stets symmetrisch und nichtnegativ definit ist, ist die Kovarianzfunktion k ebenfalls symmetrisch und negativdefinit:

$$\begin{aligned} k(s,t) &= k(t,s) \qquad (s,t \in I) \\ \sum_{i,j=1}^{n} k(t_i, t_j)\lambda_i\lambda_j \geq 0 & \qquad (n \in \mathbb{N}; t_1, \ldots, t_n \in I; \lambda_1, \ldots \lambda_n \in \mathbb{R}). \end{aligned}$$

Satz 5.18. *Gegeben seien eine beliebige nichtleere Menge I und Funktionen $m : I \longrightarrow \mathbb{R}$, $k : I \times I \longrightarrow \mathbb{R}$. Ist k nichtnegativ definit, so existiert ein Gaußprozess $(X_t)_{t \in I}$ mit Mittelwertfunktion m und Kovarianzfunktion k.*

Beweis. Seien $n \in \mathbb{N}$ und $t_1, \ldots, t_n \in I$ mit $t_k \neq t_l$ beliebig gewählt. Dann existiert eine Normalverteilung $\mu_{t_1,\ldots,t_n}$ auf $(\mathbb{R}^n, \mathfrak{B}^n)$ mit dem Mittelwertvektor $m := (m(t_1), \ldots, m(t_n))^t$ und der Kovarianzmatrix $K = (k(t_i, t_j))_{i,j=1}^n$.
(Ist $\mu^{(n)}(d\,x) = (2\pi)^{-n/2} \exp\left\{-\frac{|x|^2}{2}\right\} dx$ die n-dimensionale Standard-Normalverteilung, so erhält man $\mu_{t_1,\ldots,t_n}$ als Bild von $\mu^{(n)}$ bezüglich der linearen Transformation

$$T(x) = K^{1/2}x + m, \ x \in \mathbb{R}^n,$$

das heißt $\mu_{t_1,\dots,t_n} = \mu^{(n)} \circ T^{-1}$, da die so definierte Verteilung die charakteristische Funktion

$$\varphi_{t_1,\dots,t_n}(\lambda) = \exp\left\{i(\lambda, m) - \frac{1}{2}(K\lambda, \lambda)\right\}, \ \lambda \in \mathbb{R}^n,$$

besitzt; siehe Wahrscheinlichkeitstheorie I).

Die Familie der Verteilungen $\mu_{t_1,\dots,t_n}$ erfüllt die Kolmogorovsche Verträglichkeitsbedingung: Mit der Projektion $\pi_n : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$ ist nur zu zeigen:

$$\mu_{t_1,\dots,t_n} = \mu_{t_1,\dots,t_{n+1}} \circ \pi_n^{-1}.$$

Da eine Verteilung auf $\mathbb{R}^n$ durch ihre charakteristische Funktion eindeutig bestimmt ist, kann man dies mit Hilfe des Integraltransformationssatzes $(*)$ und durch die Berechnung der charakteristischen Funktion von $\mu_{t_1,\dots,t_{n+1}}$ $(**)$ nachprüfen:

$$\begin{aligned}
& \int\limits_{\mathbb{R}^n} \exp\left\{i\sum_{k=1}^{n} \lambda_k x_k\right\} (\mu_{t_1,\dots,t_{n+1}} \circ \pi_n^{-1})(dx_1, \dots, dx_{n+1}) \\
\underset{(*)}{=} & \int\limits_{\mathbb{R}^{n+1}} \exp\left\{i\sum_{k=1}^{n} \lambda_k x_n\right\} \mu_{t_1,\dots,t_{n+1}}(dx_1, \dots, dx_{n+1}) \\
\underset{(**)}{=} & \exp\left\{i\sum_{k=1}^{n+1} m(t_k)\lambda_k - \frac{1}{2}\sum_{k,l=1}^{n+1} k(t_k, t_l)\lambda_k\lambda_l\right\}\Bigg|_{\lambda_{m+1}=0} \\
= & \exp\left\{i\sum_{k=1}^{n} m(t_k)\lambda_k - \frac{1}{2}\sum_{k,l=1}^{n} k(t_k, t_l)\lambda_k, \lambda_l\right\},
\end{aligned}$$

und dies ist die charakteristische Funktion von $\mu_{t_1,\dots,t_n}$. Also existiert nach dem Satz von Kolmogorov ein reellwertiger Prozess $(X_t)_{t\in I}$, dessen endlichdimensionale Verteilungen gerade $\mu_{t_1,\dots,t_n}$ sind. Dieser Prozess ist ein Gaußprozess (da $\mu_{t_1,\dots,t_n}$ Gaußsch ist), und seine Mittelwert- und Kovarianzfunktion stimmen mit m bzw. k überein:

$$\mathbb{E}X_t = \int\limits_{\mathbb{R}} x\mu_t(dx) = m(t)$$

$$\mathbb{E}(X_s - m(s))(X_t - m(t)) = \int\limits_{\mathbb{R}^2} (x - m(s))(y - m(t))\mu_{s,t}(dx, dy) = k(s, t)$$

(Man benutze die Darstellung von μ_t beziehungsweise $\mu_{s,t}$ als Bildmaß seiner Standard-Normalverteilung bezüglich einer linearen Transformation). □

5.3.5. Wienerprozess

R. Brown(1828), L. Bacheliér(1900), A. Einstein(1905), N. Wiener(1923)

Definition 5.19. Ein reellwertiger stochastischer Prozess $\mathbb{W} = (W_t)_{t\in[0,\infty)}$ heißt *Wienerprozess (Brownsche Bewegung)*, falls folgendes gilt:

a) $W_0(\omega) = 0$ für alle ω;

b) $\mathbb{W}$ besitzt unabhängige Zuwächse: das bedeutet für beliebige $n \in \mathbb{N}$ und $0 \le t_0 < \cdots < t_n$ sind die Zufallsgrößen $W_{t_0}, W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ unabhängig;

c) $\mathbb{W}$ besitzt stationäre Zuwächse: für beliebige $t, h > 0$ besitzen $W_{t+h} - W_h$ und W_t die gleiche Verteilung;

d) W_t ist $\mathcal{N}(0, t)$-verteilt;

e) $\mathbb{W}$ besitzt stetige Pfade.

Behauptung 5.20. *Ein reellwertiger Prozess $\mathbb{W} = (W_t)_{t\in[0,\infty)}$ ist genau dann ein Wienerprozess, wenn folgendes gilt:*

(a) $\mathbb{W}$ ist ein Gaußprozess mit Mittelwertfunktion $m(t) \equiv 0$ und Kovarianzfunktion $k(s,t) = s \wedge t$ $(s, t \ge 0)$.

(b) $\mathbb{W}$ besitzt stetige Pfade und $W_0 \equiv 0$.

Beweisskizze.

„$\Longrightarrow$"
Sei $n \in \mathbb{N}$ und $0 \le t_1 < \cdots < t_n$. Dann ist $(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}})^t$ ein Gaußscher Vektor mit Mittelwertvektor 0 und Kovarianzmatrix

$$\begin{pmatrix} t_1 & & & 0 \\ & t_2 - t_1 & & \\ & & \ddots & \\ 0 & & & t_n - t_{n-1} \end{pmatrix}. \tag{5.11}$$

Deshalb ist

$$\begin{pmatrix} W_{t_1} \\ W_{t_2} \\ \vdots \\ W_{t_n} \end{pmatrix} = \begin{pmatrix} 1 & & & 0 \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} W_{t_1} \\ W_{t_2} - W_{t_1} \\ \vdots \\ W_{t_n} - W_{t_{n-1}} \end{pmatrix} \tag{5.12}$$

ebenfalls Gaußsch (was man mittels linearer Transformation zeigt) mit Mittelwertvektor 0 und Kovarianzmatrix

$$\begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & & \vdots \\ t_1 & t_2 & \cdots & t_n \end{pmatrix} = (t_i \wedge t_j)_{i,j=1}^n \tag{5.13}$$

Das heißt, $\mathbb{W}$ ist ein Gaußprozess mit Mittelwertfunktion $m(t) \equiv 0$ und Kovarianzfunktion $k(s,t) = s \wedge t,\ s,t \geq 0$.

„$\Longleftarrow$“
Sei $\mathbb{W}$ Gaußsch mit der Mittelwertfunktion $m(t) \equiv 0$ und Kovarianzfunktion $k(s,t) = s \wedge t$. Sei weiterhin $n \in \mathbb{N}$ und $0 \leq t_1 < \cdots < t_n$. Dann ist $(W_{t_1}, W_{t_2}, \ldots, W_{t_n})^t$ ein Gaußvektor mit Mittelwertvektor 0 und Kovarianzmatrix (5.13). Dann erhält man mit (5.12) durch „Rücktransformation“, dass $(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}})^t$ ein Gaußvektor mit Mittelwertvektor 0 und Kovarianzmatrix (5.13) ist. Also sind $W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ unabhängig und normalverteilt mit Varianzen $t_1, t_2 - t_1, \ldots, t_n - t_{n-1}$ und Mittelwert 0. □

Bemerkung 5.21. Oft fordert man o.B.d.A $W_0 = 0$ und die Stetigkeit der Pfade nur fast sicher (statt für alle Realisierungen).

Behauptung 5.22. *Der Wienerprozess existiert.*

Beweis. Wir zeigen, dass ein Prozess $\mathbb{W}$ mit den Eigenschaften (i) und (ii) aus Behauptung 5.19 existiert.
Wegen Satz 5.17 genügt es zur Existenz eines Prozesses $\widetilde{\mathbb{W}} = (\widetilde{W}_t)_{t\in[0,\infty)}$ mit der Eigenschaft (i) nachzuweisen, dass die Funktion $k(s,t) = s \wedge t$ $(s,t \geq 0)$ nichtnegativ definit ist. Dies folgt daraus, dass für $n \in \mathbb{N}$ und $0 \leq t_1 < \cdots < t_n$ gilt

$$(k(t_i,t_j))_{i,j=1}^n = \begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_n \end{pmatrix} = CC^t$$

mit

$$C = \begin{pmatrix} \sqrt{t_1} & & & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & & \\ \vdots & \vdots & \ddots & \\ \sqrt{t_1} & \sqrt{t_1 - t_2} & \cdots & \sqrt{t_n - t_{n-1}} \end{pmatrix}$$

und eine Matrix der Gestalt CC^t stets nichtnegativ definit ist. Es bleibt zu zeigen, dass $\widetilde{\mathbb{W}}$ eine stetige Modifikation $\mathbb{W}$ besitzt. Hierzu überprüfen wir, dass die Vorraussetzung des Satzes von Kolmogorov-Čentsov mit $\gamma = 4$ und $\varepsilon = 1$ erfüllt ist:

$$\mathbb{E}\ \underbrace{\left|\widetilde{W}_s - \widetilde{W}_t\right|}_{\mathcal{N}(0,|s-t|)\text{-verteilt}}^4 = 3\,|s-t|^2 \ \ (s,t \geq 0).$$

Wegen $\mathbb{E}W_0^2 = 0$ ist außerdem $W_0 = 0$ fast sicher (und durch Abänderung von $\mathbb{W}$ auf einer Nullmenge kann man $W_0 \equiv 0$ erreichen). □

5.3.6. Poissonprozess

Definition 5.23. Ein $\mathbb{N}_0$-wertiger stochastischer Prozess $N = (N_t)_{t\in[0,\infty)}$ heißt Poissonprozess mit Parameter $\lambda > 0$, falls folgendes gilt:

a) $N_0(\omega) = 0$ für alle ω;

b) N besitzt unabhängige Zuwächse;

c) N besitzt stationäre Zuwächse;

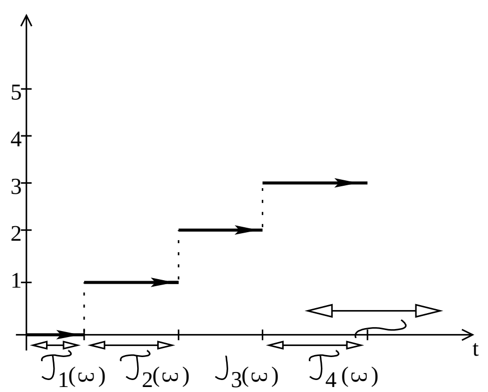
d) N_t ist Poisson-verteilt mit Parameter λt $(t \geq 0)$;

e) N besitzt stückweise konstante rechtsstetige Pfade.

Behauptung 5.24. *Der Poissonprozess existiert.*

Beweisidee. Der Poissonprozess lässt sich explizit konstruieren. Wir wissen, dass eine Folge (τ_n) unabhängig identisch exponential-verteilter Zufallsgrößen mit Parameter λ (auf einem gewissen Wahrscheinlichkeitsraum) existiert. Dann setzt man

$$N_t(\omega) = \max\left\{n \in \mathbb{N} : \sum_{i=1}^{n} \tau_i(\omega) \leq t\right\} \qquad (t \geq 0, \omega \in \Omega)$$



Dieser Prozess erfüllt (i)-(iv) und außerdem fast sicher Eigenschaft (v). □

Kapitel 6

Ergodentheorie

6.1. Grundlegende Begriffe

Als Vorrausetzung für diesen Abschnitt sei $(\Omega, \mathfrak{F}, \mathbb{P})$ ein beliebiger Wahrscheinlichkeitsraum.

Definition 6.1. Eine Abbildung $T : \Omega \longrightarrow \Omega$ heißt *maßerhaltende Transformation*, wenn sie $\mathfrak{F}$-$\mathfrak{F}$-messbar ist und $\mathbb{P}$ *invariant* läßt (das heißt wenn $\mathbb{P} \circ T^{-1} = \mathbb{P}$ gilt).

Definition 6.2. Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein stochastischer Prozess mit Werten in einem messbaren Raum $(E, \mathfrak{E})$. Der Prozess $\mathbb{X}$ heißt *stationär*, wenn er die gleiche Verteilung wie der Prozess $\widetilde{\mathbb{X}} = (X_{n+1})_{n\in\mathbb{N}_0}$ besitzt.

Ein stochastischer Prozess $\mathbb{X}$ ist also genau dann stationär, wenn für beliebige $n \in \mathbb{N}_0$ und beliebige $A \in \mathfrak{E}^{(n+1)}$

$$\mathbb{P}((X_0, \dots, X_n) \in A) = \mathbb{P}((X_1, \dots, X_{n+1}) \in A)$$

gilt.

Die nächste Behauptung gibt Aufschluss darüber, in welchem Sinne die Betrachtung maßerhaltender Transformationen äquivalent zur Betrachtung stationärer Prozesse ist.

Behauptung 6.3. *Sei $(E, \mathfrak{E})$ ein beliebiger messbarer Raum.*

(a) Ist T eine maßerhaltende Transformation und X_0 eine $(E, \mathfrak{E})$-wertige Zufallsvariable, so wird durch

$$X_n := X_0 \circ T^n \qquad (n \in \mathbb{N}_0)$$

ein stationärer Prozess $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ definiert $(T^0 := id,\ T^n := \underbrace{T \circ \ldots \circ T}_{(n-mal)})$.

(b) *Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein stationärer Prozess mit Zustandsraum $(E, \mathfrak{E})$ und $\mathbb{P}_{\mathbb{X}}$ seine Verteilung auf $(E^{\mathbb{N}_0}, \mathfrak{E}^{\mathbb{N}_0})$. Dann wird durch*

$$T((x_n)_{n\in\mathbb{N}_0}) := (x_{n+1})_{n\in\mathbb{N}_0}, \qquad (x_n)_{n\in\mathbb{N}_0} \in E^{\mathbb{N}_0},$$

eine maßerhaltende Transformation auf dem kanonischen Wahrscheinlichkeitsraum $(E^{\mathbb{N}_0}, \mathfrak{E}^{\mathbb{N}_0}, \mathbb{P}_{\mathbb{X}})$ definiert. Der kanonische Prozess $(\pi_n)_{n\in\mathbb{N}_0}$ stimmt dann mit dem in (a) zu T und π_0 konstruierten stationären Prozess überein. (Wie früher bezeichnet $\pi_n : E^{\mathbb{N}_0} \longrightarrow E$ dabei die Projektion auf die n-te Koordinate)

Beweis.

(a) Seien dazu $n \in \mathbb{N}_0$ und $B \in \mathfrak{E}^{n+1}$ beliebig gewählt.
Wegen $X_{n+1} = X_n \circ T$ $(n \in \mathbb{N}_0)$ gilt, da T maßerhaltend $(*)$ ist:

$$\begin{aligned}
\mathbb{P}((X_1, \dots, X_{n+1}) \in B) &= \mathbb{P}((X_0 \circ T, \dots, X_n \circ T) \in B) \\
&= \mathbb{P}(T^{-1}(\{(X_0, \dots, X_n) \in B\})) \\
&\underset{(*)}{=} \mathbb{P}((X_0, \dots, X_n) \in B)
\end{aligned}$$

(b) *T ist $\mathfrak{E}^{\mathbb{N}_0}$-$\mathfrak{E}^{\mathbb{N}_0}$-messbar*: Für beliebige $n \in \mathbb{N}_0$ und $A \in \mathfrak{E}$ ist

$$T^{-1}(\pi_n^{-1}(B)) = \pi_{n+1}^{-1}(B).$$

Da die eindimensionalen Projektionen π_n, $n \in \mathbb{N}_0$, die σ-Algebra $\mathfrak{E}^{\mathbb{N}_0}$ erzeugen, liefert dies die gewünschte Messbarkeit.

T läßt $\mathbb{P}_{\mathbb{X}}$ invariant: Sei $\widetilde{\mathbb{X}} = (X_{n+1})_{n\in\mathbb{N}_0}$. Dann gilt:

$$\widetilde{\mathbb{X}} = T \circ \mathbb{X} \text{ und } \mathbb{P}_{\widetilde{\mathbb{X}}} = \mathbb{P}_{\mathbb{X}} \qquad \text{(Stationarität von } \mathbb{X}).$$

Deshalb ist

$$\mathbb{P}_{\mathbb{X}} \circ T^{-1} = (\mathbb{P} \circ X^{-1}) \circ T^{-1} = \mathbb{P} \circ (T \circ \mathbb{X})^{-1} = \mathbb{P}_{\widetilde{\mathbb{X}}}.$$

Die letzte Aussage in (b) ist wegen

$$\pi_n = \pi_0 \circ T^n \qquad (n \in \mathbb{N}_0)$$

offensichtlich.

□

Definition 6.4. Sei $T : \Omega \longrightarrow \Omega$ und $A \subseteq \Omega$. Die Menge A heißt *T-invariant*, falls $T^{-1}(A) = A$ gilt.

Behauptung 6.5. *Sei T eine maßerhaltende Transformation. Dann ist*

$$\mathfrak{J} := \left\{ A \in \mathfrak{F} : \; T^{-1}(A) = A \right\}$$

eine Teil-σ-Algebra von $\mathfrak{F}$, die σ-Algebra der T-invariantan Ereignisse.

Beweis. Übungsaufgabe

Definition 6.6. Sei T eine maßerhaltende Transformation.

a) T heißt *ergodisch*, wenn die σ-Algebra $\mathfrak{J}$ der T-invarianten Ereignisse $\mathbb{P}$-trivial ist, das bedeutet, dass für jedes $A \in \mathfrak{J}$ endweder $\mathbb{P}(A) = 0$ oder $\mathbb{P}(A) = 1$ gilt.

b) T heißt *mischend*, falls für beliebige $A, B \in \mathfrak{F}$ gilt

$$\lim_{n\to\infty} \mathbb{P}(A \cap T^{-n}B) = \mathbb{P}(A)\mathbb{P}(B) \tag{6.1}$$

(das heißt die Ereignisse A und $T^{-n}B$ sind für $n \to \infty$ asymtotisch unabhängig).

Satz 6.7. *Jede mischende Transformation ist ergodisch.*

Beweis. Sei T mischend $(*)$ und $\mathfrak{J}$ die σ-Algebra der T-invarianten Ereignisse. Dann gilt für beliebige $A \in \mathfrak{J}$

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) \underset{A\in\mathfrak{J}}{=} \mathbb{P}(A \cap T^{-n}A) \underset{(*)}{\overset{n\to\infty}{\longrightarrow}} \mathbb{P}(A) \cdot \mathbb{P}(A).$$

Also ist $\mathbb{P}(A) = \mathbb{P}(A)^2$ und folglich $\mathbb{P}(A) \in \{0, 1\}$. □

Behauptung 6.8. *Gilt (6.1) für alle A, B aus einer Algebra, die $\mathfrak{F}$ erzeugt, so ist T mischend.*

Zum Beweis benutzt man das folgende Lemma aus der Maßtheorie (siehe z.B. Bauer, Maß- und Integrationstheorie, Satz 5.7)

Lemma 6.9. *Sei $\mathfrak{A}_0$ eine Algebra, welche die σ-Algebra $\mathfrak{F}$ erzeugt. Dann existiert zu jedem $A \in \mathfrak{F}$ und jedem $\varepsilon > 0$ eine Menge $A_0 \in \mathfrak{A}_0$ mit $\mathbb{P}(A \triangle A_0) < \varepsilon$.*

Beweis der Behauptung 6.8. Sei $\mathfrak{A}_0$ eine Algebra mit $\sigma(\mathfrak{A}_0) = \mathfrak{F}$, und (6.1) gelte für alle $A, B \in \mathfrak{A}_0$. Seien außerdem $A, B \in \mathfrak{F}$ und $\varepsilon > 0$ beliebig gewählt. Nach Lemma 6.9 existieren dann $A_0, B_0 \in \mathfrak{A}_0$ mit $\mathbb{P}(A \triangle A_0) < \varepsilon$ und $\mathbb{P}(B \triangle B_0) < \varepsilon$. Dann folgt, da T maßerhaltend $(*)$ ist,

$$\begin{aligned}
\left|\mathbb{P}(A \cap T^{-n}B) - \mathbb{P}(A_0 \cap T^{-n}B_0)\right| &\leq \mathbb{P}(A \triangle A_0) + \mathbb{P}(T^{-n}B \triangle T^{-n}B_0) \\
&\underset{(*)}{=} \mathbb{P}(A \triangle A_0) + \mathbb{P}(B \triangle B_0) \\
&< 2\varepsilon
\end{aligned}$$

und

$$|\mathbb{P}(A) - \mathbb{P}(A_0)| < \varepsilon, \ |\mathbb{P}(B) - \mathbb{P}(B_0)| < \varepsilon.$$

Da $\lim_{n\to\infty} \mathbb{P}(A_0 \cap T^{-n}B_0) = \mathbb{P}(A_0)\mathbb{P}(B_0)$ gilt und $\varepsilon > 0$ beliebig klein gewählt werden kann, folgt hiermit leicht die Behauptung

$$\lim_{n\to\infty} \mathbb{P}(A \cap T^{-n}B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

□

Bemerkung 6.10.

1. Die durch $T((x_n)_{n\in\mathbb{N}_0}) := (x_{n+1})_{n\in\mathbb{N}_0}$ definierte Abbildung $T : E^{\mathbb{N}_0} \longrightarrow E^{\mathbb{N}_0}$ wird oft *Verschiebungsoperator* oder auch *Gleitoperator* genannt.

2. Ein stationärer Prozess $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ mit Zustandsraum $(E, \mathfrak{E})$ heißt *ergodisch* (*mischend*), wenn der Gleitoperator T auf dem kanonischen Wahrscheinlichkeitsraum $(E^{\mathbb{N}_0}, \mathfrak{E}^{\mathbb{N}_0}, \mathbb{P}_{\mathbb{X}})$ ergodisch (mischend) ist.

6.2. Beispiele

1. Jede beliebige Folge $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ von *unabhängig identisch verteilten* $(E, \mathfrak{E})$-wertigen *Zufallsvariablen* ist ein ergodischer stationärer Prozess:
Sei μ die Verteilung von X_0 auf dem messbaren Raum $(E, \mathfrak{E})$. Dann besitzen $(X_0, \dots, X_n)$ und $(X_1, \dots, X_{n+1})$ die gleiche Verteilung, nämlich $\mu \otimes \cdots \otimes \mu$ ($(n+1)$-mal) ($n \in \mathbb{N}_0$). Also ist $\mathbb{X}$ stationär.
Die von den Zufallsvariablen X_n erzeugten σ-Algebren $\mathfrak{F}_n := \sigma(X_n)$, $n \in \mathbb{N}_0$, sind unabhängig. Seien nun T der Gleitoperator auf $E^{\mathbb{N}_0}$ und $A \in \mathfrak{E}^{\mathbb{N}_0}$ eine T-invariante Menge, das heißt $T^{-m}A = A$ für alle m. Mit $\mathbb{X}_m := (X_{m+n})_{n\in\mathbb{N}_0}$ bezeichnen wir den um m Zeiteinheiten verschobenen Prozess. Dann gilt:

$$\{\mathbb{X} \in A\} = \left\{\mathbb{X} \in T^{-m}A\right\} = \{\mathbb{X}_m \in A\} \in \bigvee_{k=m}^{\infty} \mathfrak{F}_k, \qquad \forall\, m \in \mathbb{N}_0.$$

Also ist $\{\mathbb{X} \in A\} \in \mathcal{T}_\infty := \bigcap\limits_{m=0}^{\infty} \bigvee\limits_{k=m}^{\infty} \mathfrak{F}_k$.
Nach dem 0-1-Gesetz von Kolmogorov (Satz 2.13) ist die Tail-σ-Algebra $\mathcal{T}_\infty$ $\mathbb{P}$-trivial, das bedeutet

$$\mathbb{P}_{\mathbb{X}}(A) = \mathbb{P}(\mathbb{X} \in A) \in \{0, 1\}\,.$$

Also ist T ergodisch auf $(E^{\mathbb{N}_0}, \mathfrak{E}^{\mathbb{N}_0}, \mathbb{P}_{\mathbb{X}})$. Der Gleitoperator ist sogar mischend, was leicht mit Hilfe der Behauptung 6.8 und der Algebra der Zylindermengen folgt (Übungsaufgabe).

2. *Drehoperator auf dem Einheitskreis.*
Sei $D := \{z \in \mathbb{C} : |z| = 1\}$ der Einheitskreis, $\mathfrak{B}_D$ die zugehörige Borel-σ-Algebra (die von den Kreisbögen erzeugt wird) und λ_D das normierte Lebesgue-Maß auf $\mathfrak{B}_D$ (das heißt dasjenige Maß, dass jedem Kreisbogen seine Länge$/2\pi$ zuordnet). Für jedes $c \in D$ definieren wir den Drehoperator

$$T_c(\omega) := c \cdot \omega \qquad (\omega \in D).$$

Dann ist T_c eine maßerhaltende Transformation auf dem Wahrscheinlichkeitsraum $(D, \mathfrak{B}_{\mathfrak{D}}, \lambda_D)$. (Die Urbilder von Kreisbögen bezüglich T_c sind Kreisbögen gleicher Länge)

Satz 6.11. *Der Drehoperator T_c ist genau dann ergodisch, wenn c keine Einheitswurzel ist (das heißt, wenn für kein $n \in \mathbb{N}$, ein $e \in D$ mit $e^n = 1$ existiert).*

Wir beweisen zunächst

Lemma 6.12. *Ist c keine Einheitswurzel, so liegt die Folge $(c^n)_{n\in\mathbb{N}_0}$ dicht in D.*

Beweis. Die Folge (c^n) besitzt mindestens einen Häufungspunkt $\omega_0 \in D$. Sei nun $\varepsilon > 0$ beliebig gewählt. Dann existieren natürliche Zahlen $m < n$ mit $|c^m - \omega_0| < \varepsilon$ und $|c^n - \omega_0| < \varepsilon$. Daraus folgt

$$0 < \left|c^{n-m} - 1\right| < 2\varepsilon.$$

Deshalb findet man zu jedem $\omega \in D$ ein $k \in \mathbb{N}_0$ mit $\left|c^{k(n-m)} - \omega\right| < 2\varepsilon$ □

Beweis des Satzes.

1^o Sei c eine Einheitswurzel,
das heißt es existiert ein $n \in \mathbb{N}$ mit $c^n = 1$. Für dieses n ist dann T_c^n die Identität, und für jedes $A \in \mathfrak{B}_D$ ist $A \cup T_c^{-1}A \cup \cdots \cup T_c^{-(n-1)}A$ eine T-invariante Menge. Wir wählen $A \in \mathfrak{B}_D$ so, dass $0 < \lambda_D(A) < 1/n$ gilt. Dann folgt, da T_c maßerhaltend $(*)$ ist,

$$0 < \lambda_D(A \cup T_c^{-1}A \cup \cdots \cup T_c^{-(n-1)}A) \leq \sum_{k=0}^{n-1} \lambda_D(T_c^{-k}A) \underset{(*)}{=} n\lambda_D(A) < 1.$$

Folglich ist T_c nicht ergodisch.

2^o Sei nun c keine Einheitswurzel.
Mit $\mathfrak{A}_0$ bezeichnen wir die Algebra aller endlichen disjunkten Vereinigungen von Kreisbögen. Dann ist $\sigma(\mathfrak{A}_0) = \mathfrak{B}_D$.
Sei nun $A \in \mathfrak{B}_D$ eine beliebige T-invariante Menge, das heißt $T^{-1}A = A$. Angenommen $\lambda_D(A) > 0$. Wir müssen zeigen, dass dann $\lambda_D(A) = 1$ gilt. Sei $\varepsilon \in (0, 1/2)$ beliebig gewählt. Nach Lemma 6.9 existieren endlich viele paarweise disjunkte Kreisbögen $I_1, \ldots, I_n$, so dass für $B := \bigcup_{k=1}^{n} I_k$ die Ungleichung

$$\lambda_D(A \triangle B) < \varepsilon\lambda_D(A)$$

gilt. O.B.d.A sei $\lambda_D(I_k) < \varepsilon$ für $k = 1, \ldots, n$. Wir erhalten

$$\lambda_D(A \triangle B) < \varepsilon\lambda_D(A) < 2\varepsilon(1-\varepsilon)\lambda_D(A) < 2\varepsilon(\lambda_D(A) - \lambda_D(B)) \leq 2\varepsilon\lambda_D(B).$$

Daraus folgt:

$$\begin{aligned}\sum_{k=1}^{n} \lambda_D(A \cap I_k) &= \lambda_D(A \cap B) \geq \lambda_D(B) - \lambda_D(A \triangle B) \\ &\geq (1 - 2\varepsilon)\lambda_D(B) = (1 - 2\varepsilon)\sum_{k=1}^{n} \lambda_D(I_k)\end{aligned}$$

Mindestens einer der Kreisbögen I_k, welchen wir mit I bezeichnen wollen, erfüllt deshalb

$$\lambda_D(A \cap I) \geq (1 - 2\varepsilon)\lambda_D(I).$$

Nach Lemma 6.12 existieren ein $r \in \mathbb{N}$ und $n_1, \ldots, n_r \in \mathbb{N}$ derart, dass die Kreisbögen $T_c^{-n_1}I, \ldots, T_c^{-n_r}I$ paarweise disjunkt sind und den Einheitskreis bis auf eine Menge vom Maß $< 2\varepsilon$ ausfüllen.
Für $j = 1, \ldots, r$ gilt da A T_c-invariant $(*)$ und maßerhaltend $(**)$ ist:

$$\begin{aligned} \lambda_D(A \cap T_c^{-n_j}I) &\underset{(*)}{=} \lambda_D(T_c^{-n_j}A \cap T_c^{-n_j}I) \\ &\underset{(**)}{=} \lambda_D(A \cap I) \\ &\geq (1 - 2\varepsilon)\lambda_D(I) \\ &= (1 - 2\varepsilon)\lambda_D(T^{-n_j}I) \end{aligned}$$

und folglich

$$\begin{aligned} \lambda_D(A) &\geq \sum_{j=1}^{r} \lambda_D(A \cap T_c^{-n_j}I) \\ &\geq (1 - 2\varepsilon)\sum_{j=1}^{r} \lambda_D(T_c^{-n_j}I) \\ &\geq (1 - 2\varepsilon)^2. \end{aligned}$$

Da $\varepsilon > 0$ beliebig klein gewählt werden kann, ist tatsächlich $\lambda_D(A) = 1$.

□

Bemerkung 6.13. Der Drehoperator T_c ist für kein $c \in D$ mischend

Beweis. Hierzu sei o.B.d.A. c keine Einheitswurzel. Da die Folge (c^n) dicht in D liegt, existiert eine Teilfolge (c^{n_k}) mit $c^{n_k} \longrightarrow 1$ für $k \longrightarrow \infty$. Ist I ein Kreisbogen mit $0 < \lambda_D(I) < 1$, so folgt:

$$\begin{aligned} \lim_{k\to\infty} \lambda_D(I \cap T_c^{-n_k}I) &= \lim_{k\to\infty} \lambda_D(I \cap T_{c^{n_k}}^{-1}I) \\ &= \lambda_D(I) \neq \lambda_D(I)^2. \end{aligned}$$

□

3. *Stationäre Markovketten*:
Jede irreduzible, aperiodische und stationäre Markovkette $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ mit einem endlichem Zustandsraum I ist mischend (und folglich ergodisch):
Dabei sei $(p_i^0)_{i\in I}$ die (stationäre) Anfangsverteilung und $(p_{ij})_{i,j\in I}$ die Übergangsmatrix.
Aus Wahrscheinlichkeitstheorie I wissen wir, dass die Stationarität der Markovkette $\mathbb{X}$ gleichbedeutend ist mit

$$\sum_{i\in I} p_i^0 p_{ij} = p_j^0, \qquad j \in I.$$

Unter den obigen Vorraussetzungen gilt die Aussage des Konvergenzsatzes für Markovketten:

$$\lim_{n\to\infty} p_{ij}^{(n)} = p_j^0, \qquad i,j \in I.$$

Wir zeigen, dass der Gleitoperator T auf $(I^{\mathbb{N}_0}, \mathfrak{P}(I)^{\mathbb{N}_0}, \mathbb{P}_{\mathbb{X}})$ mischend ist. Aufgrund der Behauptung 6.8 genügt es, Zylindermengen

$$A = \left\{(i_n)_{n\in\mathbb{N}_0} : (i_0,\dots,i_r) \in \widetilde{A}\right\} \text{ mit } \widetilde{A} \subseteq I^{r+1}$$

und

$$B = \left\{(i_n)_{n\in\mathbb{N}_0} : (i_0,\dots,i_s) \in \widetilde{B}\right\} \text{ mit } \widetilde{B} \subseteq I^{s+1}$$

zu betrachten. (Die Zylindermengen bilden eine Algebra, die die σ-Algebra $\mathfrak{P}(I)^{\mathbb{N}_0}$ erzeugt.) Für $n > r$ gilt dann:

$$\begin{aligned}
\mathbb{P}_{\mathbb{X}}(A \cap T^{-n}B) &= \mathbb{P}((X_0,\dots,X_r) \in \widetilde{A},\ (X_n,\dots,X_{n+s}) \in \widetilde{B}) \\
&= \sum_{(i_0,\dots,i_r)\in\widetilde{A}} \sum_{(j_0,\dots,j_s)\in\widetilde{B}} p_{i_0}^0 p_{i_0 i_1} \cdots p_{i_{r-1} i_r} p_{i_r j_0}^{(n-r)} p_{j_0 j_1} \cdots p_{j_{s-1} j_s}
\end{aligned}$$

Mit der Konvergenzaussage (6.2) folgt hieraus

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(A \cap T^{-n}B) &= \sum_{(i_0,\dots,i_r)\in\widetilde{A}} p_{i_0}^0 p_{i_0 i_1} \cdots p_{i_{r-1} i_r} \sum_{(j_0,\dots,j_s)\in\widetilde{B}} p_{j_0}^0 p_{j_0 j_1} \cdots p_{j_{s-1} j_s} \\
&= \mathbb{P}((X_0,\dots,X_r) \in \widetilde{A}) \cdot \mathbb{P}((X_0,\dots,X_s) \in \widetilde{B}) \\
&= \mathbb{P}_{\mathbb{X}}(A) \cdot \mathbb{P}_{\mathbb{X}}(B).
\end{aligned}$$

Also ist T mischend.

6.3. Ergodensätze

Sei $(\Omega, \mathfrak{F}, \mathbb{P})$ wieder ein beliebiger Wahrscheinlichkeitsraum

Satz 6.14. *(Individueller Ergodensatz von Birkhoff)*
Seien T eine maßerhaltende Transformation und $\mathfrak{J}$ die σ-Algebra der T-invarianten Ereignisse. Dann gilt für beliebige $f \in \mathcal{L}^1(\mathbb{P})$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k = \mathbb{E}\left[f|\mathfrak{J}\right] \qquad \mathbb{P}\text{-fast sicher.}$$

Bemerkung 6.15. T ist genau dann ergodisch, wenn $\mathbb{E}\left[f|\mathfrak{I}\right] = \mathbb{E}f$ fast sicher für alle $f \in \mathcal{L}^1(\mathbb{P})$ gilt. (Übungsaufgabe)

Folgerung 6.16. *Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein ergodischer stationärer Prozess mit Zustandsraum $(E, \mathfrak{E})$. Dann gilt für beliebige $n \in \mathbb{N}_0$ und jede messbare Funktion $f : E^{n+1} \longrightarrow \mathbb{R}$ mit $f(X_0,\dots,X_m) \in \mathcal{L}^1(\mathbb{P})$:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k,\dots,X_{k+m}) = \mathbb{E}f(X_0,\dots,X_m) \text{ fast sicher.}$$

Beweis. Nach den obigen Vorraussetzungen ist der Gleitoperator T auf dem Wahrscheinlichkeitsraum $(E^{\mathbb{N}_0}, \mathfrak{E}^{\mathbb{N}_0}, \mathbb{P}_{\mathbb{X}})$ ergodisch. Mit $\pi_m : E^{\mathbb{N}_0} \longrightarrow E^{m+1}$ bezeichnen wir die Projektion auf die ersten $m+1$ Koordinaten. Wegen

$$\mathbb{E}\,|f(X_0, \ldots, X_m)| = \int |f \circ \pi_m| \, d\,\mathbb{P}_{\mathbb{X}}$$

ist $f \circ \pi_m \in \mathcal{L}^1(\mathbb{P}_{\mathbb{X}})$, und der Birkhoffsche Ergodensatz liefert:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ \pi_m \circ T^k = \int f \circ \pi_m d\,\mathbb{P}_{\mathbb{X}} \qquad \mathbb{P}_{\mathbb{X}}\text{-fast sicher}$$

Die Behauptung folgt nun wegen

$$\begin{aligned} & \mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, \ldots, X_{k+m}\right) = \mathbb{E}f(X_0, \ldots, X_m)) \\ = \quad & \mathbb{P}_{\mathbb{X}}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ \pi_m \circ T^k\right) = \int f \circ \pi_m d\,\mathbb{P}_{\mathbb{X}}. \end{aligned}$$

□

Der Ergodensatz liefert also insbesondere eine Verallgemainerung des starken Gesetzes der großen Zahlen auf ergodische stationäre Prozesse. (Man beachte hier, dass die Zufallsvariablen X_n im Allgemeinen nicht unabhängig sind.)
Zum Beweis von Satz 6.14 setzen wir

$$X_n := f \circ T^n \text{ und } S_n := \sum_{k=0}^{n-1} X_k \qquad (n \in \mathbb{N}_0).$$

Insbesondere ist (X_n) ein reellwertiger stationärer Prozess mit $X_0 \in \mathcal{L}^1(\mathbb{P})$. Wir setzen

$$M_n := \max\{0, S_1, \ldots, S_n\} \qquad (n \in \mathbb{N}).$$

Lemma 6.17. *(Maximal-Ergodenlemma von Hopf)*

$$\mathbb{E}(X_0 \mathbb{1}_{M_n > 0}) \geq 0, \qquad n \in \mathbb{N}.$$

Beweis. Für $1 \leq k \leq n$ ist $M_n \circ T \geq S_k \circ T$. Also ist

$$X_0 + M_n \circ T \geq X_0 + S_k \circ T = S_{k+1},$$

das heißt

$$X_0 \geq S_{k+1} - M_n \circ T, \qquad 1 \leq k \leq n.$$

Dies gilt auch für $k = 0$. Folglich ist

$$X_0 \geq \max\{S_1, \ldots, S_n\} - M_n \circ T \tag{6.2}$$

Außerdem ist

$$\{M_n > 0\}^c = \{M_n = 0\} \cap \{M_n \circ T \geq 0\} \subseteq \{M_n - M_n \circ T \leq 0\}, \qquad (6.3)$$

deshalb erhalten wir, da T maßerhaltend $(*)$ ist,

$$\begin{aligned} \mathbb{E}X_0 \mathbb{1}_{M_n>0} &\underset{(6.2)}{\geq} \mathbb{E}(\max\{S_1, \ldots, S_n\} - M_n \circ T)\mathbb{1}_{M_n>0} \\ &= \mathbb{E}(M_n - M_n \circ T)\mathbb{1}_{M_n>0} \\ &\underset{(6.3)}{\geq} \mathbb{E}(M_n - M_n \circ T) \\ &\underset{(*)}{=} 0. \end{aligned}$$

□

Beweis von Satz 6.14. Wegen $\mathbb{E}[X_0|\mathfrak{J}] \circ T = \mathbb{E}[X_0|\mathfrak{J}]$ fast sicher (Übungsaufgabe) können wir o.B.d.A. $\mathbb{E}[X_0|\mathfrak{J}] = 0$ fast sicher annehmen (andernfalls betrachtet man $\widetilde{X}_n := X_n - \mathbb{E}[X_0|\mathfrak{J}]$ an Stelle von X_n, $n \in \mathbb{N}$).
Dann ist nur

$$Z := \limsup_{n\to\infty} \frac{1}{n} S_n \leq 0 \text{ fast sicher}$$

zu zeigen ($\liminf\limits_{n\to\infty} \frac{1}{n} S_n \geq 0$ fast sicher folgt hieraus durch den Übergang von X_n zu $-X_n$).
Sei $\varepsilon > 0$ beliebig und $F := \{Z > \varepsilon\}$. Dann ist also nur $\mathbb{P}(F) = 0$ zu zeigen. Wegen $Z \circ T = Z$ ist $F \in \mathfrak{J}$. Wir setzen

$$X_n^\varepsilon := (X_n - \varepsilon)\mathbb{1}_F, \qquad S_n^\varepsilon := \sum_{k=0}^{n-1} X_k^\varepsilon,$$

$$M_n^\varepsilon := \max\{0, S_1^\varepsilon, \ldots, S_n^\varepsilon\}, \qquad F_n := \{M_n^\varepsilon > 0\}$$

Offenbar ist $F_1 \subseteq F_2 \subseteq F_3 \subseteq \ldots$ und

$$\bigcup_{n=1}^{\infty} F_n = \left\{\sup_k \frac{1}{k} S_k^\varepsilon > 0\right\} = \left\{\sup_k \frac{1}{k} S_k > \varepsilon\right\} \cap F = F.$$

Also gilt $F_n \uparrow F$, und der Satz von Lebesgue über majorisierte Konvergenz liefert

$$\lim_{n\to\infty} \mathbb{E}(X_0^\varepsilon \mathbb{1}_{F_n}) = \mathbb{E}(X_0^\varepsilon \mathbb{1}_F) = \mathbb{E}X_0^\varepsilon.$$

Da das Maximal-Ergodenlemma von Hopf auf (X_n^ε) angewandt werden kann, gilt $\mathbb{E}(X_0^\varepsilon \mathbb{1}_{F_n}) \geq 0$ für alle n. Also ist

$$0 \leq \mathbb{E}X_0^\varepsilon = \mathbb{E}((X_0 - \varepsilon)\mathbb{1}_F) \underset{F\in\mathfrak{J}}{=} \mathbb{E}\left[\mathbb{E}[X_0 - \varepsilon|\mathfrak{J}]\, \mathbb{1}_F\right] = -\varepsilon\, \mathbb{P}(F)$$

und somit tatsächlich $\mathbb{P}(F) = 0$. □

Satz 6.18. *($\mathcal{L}^p$-Ergodensatz von Neumann) Seien T eine maßerhaltende Transformation, $p \geq 1$ und $f \in \mathcal{L}^p(\mathbb{P})$. Dann gilt.*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k = \mathbb{E}\left[f|\mathfrak{J}\right] \ \ \textit{in } \mathcal{L}^p(\mathbb{P}).$$

Beweis. Übungsaufgabe. Der Satz läßt sich nicht direkt mit majorisierter Konvergenz aus dem individuellen Ergodensatz ableiten, da im Allgemeinen keine p-fach integrierbare Majorante existiert.

Kapitel 7

Schwache Konvergenz von Wahrscheinlichkeitsmaßen

7.1. Definition und erste Eigenschaften

Motivation: Sei $(S_n)_{n\in\mathbb{N}_0}$ eine in 0 startende einfache symmetrische Irrfahrt auf $\mathbb{Z}$, das heißt $S_n = \sum_{k=1}^n X_k$, wobei (X_n) eine Folge von unabhängigen identisch verteilten Zufallsgrößen mit $\mathbb{P}(X_1 = -1) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$ ist. Der Zentrale Grenzwertsatz besagt, dass die Verteilungen der Zufallsgrößen $\frac{1}{\sqrt{n}} S_n$ für $n \to \infty$ schwach gegen die $\mathcal{N}(0,1)$-Verteilung konvegieren.
Betrachten wir den Prozess(S_n) auf dem Zeithorizont von 0 bis n, so ist dies eine asymptotische Aussage über die Verteilung des Prozesses zum Endzeitpunkt n. Dies sagt aber nichts über den zeitlichen Verlauf, die zufällige Dynamik, des Prozesses bis zum Endzeitpunkt aus. Um die Dynamik zu erfassen, reskalieren wir zusätzlich den Zeithorizont auf $[0,1]$:
Wir betrachten die Punkte

$$(0,0),\ \left(\frac{1}{n}, \frac{1}{\sqrt{n}}\ S_1(\omega)\right),\ \left(\frac{2}{n}, \frac{1}{\sqrt{n}}\ S_2(\omega)\right), dots,\ \left(\frac{n}{n}, \frac{1}{\sqrt{n}}\ S_n(\omega)\right)$$

und verbinden diese stückweise linear:
Damit erhalten wir für jedes n einen stochastischen Prozess

$$\mathbb{Y}_n = (Y_n(t))_{t\in[0,1]}$$

mit stetigen Pfaden.
Die Verteilung von $Y_n(1)$ konvergiert schwach gegen $\mathcal{N}(0,1)$.

Frage: Wie verhält sich die Verteilung des Prozesses $\mathbb{Y}_n$ auf $\mathcal{C}\,[0,1]$? Gilt hierfür eine unendlich-dimensionale Version des Zentralen Grenzwertsatzes?

Seien dazu

(E, ϱ) ein metrischer Raum und $\mathfrak{B}_E$ die zugehörige Borel-σ-Algebra,
$\mathcal{P}(E)$ die Menge aller Wahrscheinlichkeitsmaße auf $(E, \mathfrak{B}_E)$,
$\mathcal{C}(E)$ die Menge aller *beschränkten stetigen* Funktionen $f : E \longrightarrow \mathbb{R}$.

Definition 7.1.

a) Seien $\mu_n, \mu \in \mathcal{P}(E)$. Man sagt, die Folge (μ_n) von Wahrscheinlichkeitsmaßen *konvergiert schwach* gegen das Wahrscheinlichkeitsmaß μ, falls für alle $f \in \mathcal{C}(E)$ gilt:
$$\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu.$$
Schreibweisen: $\mu_n \xrightarrow{\text{w}} \mu, \ \mu = \text{w - }\lim_{n\to\infty} \mu_n$

b) Seien X_n, X $(E, \mathfrak{B}_E)$-wertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, \mathbb{P})$. Man sagt, die Folge (X_n) *konvergiert in Verteilung gegen* X, wenn die Verteilungen dieser Zufallsvariablen schwach konvergieren:
$$\mathbb{P} \circ X_n^{-1} \xrightarrow{\text{w}} \mathbb{P} \circ X^{-1}.$$
Schreibweisen: $X_n \xrightarrow{d} X, \ \mathcal{L}(X_n) \xrightarrow{\text{w}} \mathcal{L}(X).$

Das folgende Lemma zeigt, dass der Grenzwert μ in (i) eindeutig bestimmt ist.

Behauptung 7.2. *Für zwei Wahrscheinlichkeitsmaße $\mu, \nu \in \mathcal{P}(E)$ gelte für alle $f \in \mathcal{C}(E)$*
$$\int f \, d\mu = \int f \, d\nu \tag{7.1}$$
Dann ist $\mu = \nu$.

Beweis. Da die abgeschlossenen Mengen einen $\cap$-stabilen Erzeuger von $\mathfrak{B}_E$ bilden, genügt es $\mu(F) = \nu(F)$ für alle abgeschlossenen Mengen $F \subseteq E$ zu zeigen. Sei deshalb F abgeschlossen und
$$f_n(x) := e^{-n\varrho(x,F)}, \qquad x \in E, n \in \mathbb{N}.$$
Dann ist $f_n \in \mathcal{C}(E)$ und $f_n \downarrow \mathbb{1}_F$. Mit dem Satz von Lebesgue folgt dann
$$\mu(F) = \lim_{n\to\infty} \int f_n \, d\mu \underset{(7.1)}{=} \lim_{n\to\infty} \int f_n \, d\nu = \nu(F)$$
$\square$

Satz 7.3. *(Portmanteau-Theorem)*
Seien $\mu_n, \mu \in \mathcal{P}(E)$. Folgende Aussagen sind äquivalent:

(i) $\mu_n \xrightarrow{w} \mu$;

(ii) $\limsup\limits_{n\to\infty} \mu_n(F) \leq \mu(F)$ *für alle abgeschlossenen* $F \subseteq E$;

(iii) $\liminf\limits_{n\to\infty} \mu_n(G) \geq \mu(G)$ *für alle offenen* $G \subseteq E$;

(iv) $\lim\limits_{n\to\infty} \mu_n(B) = \mu(B)$ *für alle* $b \in \mathfrak{B}_E$ *mit* $\mu(\partial B) = 0$.

Beweis. Beweisschema: $(i) \iff (ii) \iff (iii) \implies (iv) \implies (ii)$

„$(i) \implies (ii)$":
Seien F abgeschlossen und $f_m(x) := e^{-m\varrho(x,F)}$. Dann gilt $f_m \in \mathcal{C}(E)$ und $f_m \downarrow \mathbb{1}_F$. Damit folgt

$$\mu_n(F) \leq \int f_m \, d\mu_n \underset{(i)}{\overset{n\to\infty}{\longrightarrow}} \int f_m \, d\mu,$$

das heißt

$$\limsup_{n\to\infty} \mu_n(F) \leq \int f_m \, d\mu \underset{m\to\infty}{\longrightarrow} \mu(F).$$

„$(ii) \iff (iii)$":
Offene und abgeschlossene Mengen gehen durch Komplementbildung auseinander hervor.

„$(ii) + (iii) \implies (iv)$":
Sei $B \in \mathfrak{B}_E$ mit $\mu(\partial B) = 0$. Dann ist $\mu(\mathring{B}) = \mu(B) = \mu(\bar{B})$, und wir erhalten

$$\limsup_{n\to\infty} \mu_n(B) \leq \limsup_{n\to\infty} \mu_n(\bar{B}) \underset{(ii)}{\leq} \mu(\bar{B}) = \mu(B),$$

$$\liminf_{n\to\infty} \mu_n(B) \geq \liminf_{n\to\infty} \mu_n(\mathring{B}) \underset{(iii)}{\geq} \mu(\mathring{B}) = \mu(B).$$

Zusammen ergibt dies

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

„$(iv) \implies (ii)$":
Sei $F \subseteq E$ abgeschlossen. Für bel. $\delta > 0$ bezeichne $F^\delta := \{x \in E : \varrho(x,F) \leq \delta\}$ die abgeschlossene δ-Umgebung von F. Wegen $\partial F^\delta \subseteq \{x \in E : \varrho(x,F) = \delta\}$ sind die Ränder ∂F^δ, $\delta > 0$, paarweise disjunkt. Deshalb ist die Menge

$$\left\{\delta > 0 : \mu(\partial F^\delta) > 0\right\} = \bigcup_{m=1}^{\infty} \left\{\delta > 0 : \mu(\partial F^\delta) > \frac{1}{m}\right\}$$

höchstens abzählbar, da $\left\{\delta > 0 : \mu(\partial F^\delta) > \frac{1}{m}\right\}$ endlich ist, und es existiert eine Folge (δ_m) mit $\delta_m \downarrow 0$ und $\mu(\partial F^{\delta_m}) = 0$ für alle m. Damit folgt

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F^{\delta_m}) \underset{(iv)}{=} \mu(F^{\delta_m}), \qquad m \in \mathbb{N}.$$

„$(ii) \Longrightarrow (i)$“:
Sei $f \in \mathcal{C}(E)$. Wir setzen o.B.d.A. $0 < f < 1$ vorraus. (Dies kann durch lineare Transformation $\widetilde{f} = af + b$ erreicht werden.) Die Mengen

$$F_k^{(m)} := \left\{ x \in E : \quad f(x) \geq \frac{k}{m} \right\} \qquad (k = 0, 1, \ldots, m; m \in \mathbb{N})$$

sind abgeschlossen und wegen

$$\sum_{k=1}^{m} \frac{k-1}{m} \mathbb{1}_{\left\{ \frac{k-1}{m} \leq f < \frac{k}{m} \right\}} \leq f \leq \sum_{k=1}^{m} \frac{k}{m} \mathbb{1}_{\left\{ \frac{k-1}{m} \leq f < \frac{k}{m} \right\}}$$

und $F_m^{(m)} = \emptyset$ gilt, da $\mathbb{1}_{\left\{ \frac{k-1}{m} \leq f < \frac{k}{m} \right\}} = \mathbb{1}_{F_{k-1}^{(m)}} - \mathbb{1}_{F_k^{(m)}}$, dass

$$\frac{1}{m} \sum_{k=1}^{m-1} \mathbb{1}_{F_k^{(m)}} \leq f \leq \frac{1}{m} \sum_{k=0}^{m-1} \mathbb{1}_{F_k^{(m)}}.$$

Damit erhalten wir

$$\begin{aligned} \limsup_{n\to\infty} \int f \, d\mu_n \quad & \leq \quad \frac{1}{m} \sum_{k=0}^{m-1} \limsup_{n\to\infty} \mu_n(F_k^{(m)}) \\ & \underset{(ii)}{=} \quad \frac{1}{m} \sum_{k=0}^{m-1} \mu(F_k^{(m)}) \\ & \leq \quad \frac{1}{m} + \int f \, d\mu. \end{aligned}$$

Für $m \to \infty$ folgt hieraus

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

Da dies auch für $1 - f$ statt f gilt, erhalten wir

$$\liminf_{n\to\infty} \int f \, d\mu_n \geq \int f \, d\mu,$$

und die Behauptung folgt. □

Lemma 7.4. *Sei $\mathfrak{E} \subseteq \mathfrak{B}_E$ ein $\cap$-stabiles Mengensystem mit der Eigenschaft, dass sich jede offene Menge als höchstens abzählbare Vereinigung von Mengen aus $\mathfrak{E}$ darstellen lässt. Dann gilt für beliebige $\mu_n, \mu \in \mathcal{P}(E)$:*

$$\lim_{n\to\infty} \mu_n(A) = \mu(A) \quad \forall A \in \mathfrak{E} \quad \Longrightarrow \quad \mu_n \xrightarrow{w} \mu.$$

Beweis. Wir benutzen die Aussage (iii) des Portmanteau-Theorems für offene Mengen. Sei also G offen und $\varepsilon > 0$ beliebig gewählt. Dann existieren $A_1, ..., A_m \in \mathfrak{E}$ mit

$$\bigcup_{k=1}^{m} A_k \subseteq G \text{ und } \mu\left(\bigcup_{k=1}^{m} A_k\right) \geq \mu(G) - \varepsilon.$$

Gemäß dem Einschluß-Ausschlußprinzip gilt

$$\mu_n\left(\bigcup_{k=1}^{m} A_k\right) = \sum_{k=1}^{m}(-1)^{k-1} \sum_{1\leq j_1<\ldots<j_k\leq m} \mu_n(\underbrace{A_{j_1} \cap \cdots \cap A_{j_k}}_{\in\mathfrak{E}})$$

und damit erhalten wir

$$\lim_{n\to\infty} \mu_n\left(\bigcup_{k=1}^{m} A_k\right) = \mu\left(\bigcup_{k=1}^{m} A_k\right).$$

Also gilt

$$\liminf_{n\to\infty} \mu_n(G) \geq \lim_{n\to\infty} \mu_n\left(\bigcup_{k=1}^{m} A_k\right) = \mu\left(\bigcup_{k=1}^{m} A_k\right) \geq \mu(G) - \varepsilon.$$

Da $\varepsilon > 0$ beliebig klein gewählt werden kann, folgt hieraus die gewünschte Ungleichung. □

Behauptung 7.5. *Seien $\mu_n, \mu \in \mathcal{P}(\mathbb{R})$ und F_n, F die zugehörigen Verteilungsfunktionen. Dann gilt*

$$\mu_n \xrightarrow{w} \mu \quad \Longleftrightarrow \quad F_n(t) \longrightarrow F(t) \textit{ für alle } t \in \mathbb{R}, \textit{ in denen } F \textit{ stetig ist.}$$

Beweis.
„$\Longrightarrow$“:
Sei F in t stetig. Dann ist

$$\mu(\partial(-\infty, t]) = \mu(\{t\}) = F(t) - F(t-0) = 0.$$

Deshalb können wir die Aussage (iv) des Portmanteau-Theorems benutzen und erhalten

$$\lim_{n\to\infty} F_n(t) = \lim_{n\to\infty} \mu_n((-\infty, t]) = \mu((-\infty, t]) = F(t).$$

„$\Longleftarrow$“:
Da höchstens abzählbar viele $t \in \mathbb{R}$ mit $\mu(\{t\}) \neq 0$ existieren ist

$$\begin{aligned} \mathfrak{E} \quad &:= \quad \{(a,b] : \quad \mu(\{a\}) = \mu(\{b\}) = 0\} \\ &= \quad \{(a,b] : \quad F \text{ stetig in } a \text{ und } b\} \end{aligned}$$

ein Mengensystem, das die Vorraussetzungen von Lemma 7.4 erfüllt. Insbesondere gilt für $(a,b] \in \mathfrak{E}$:

$$\lim_{n\to\infty} \mu_n((a,b]) = \lim_{n\to\infty} (F_n(b) - F_n(a)) = F(b) - F(a) = \mu((a,b]).$$

Also folgt $\mu_n \xrightarrow{\mathrm{w}} \mu$. □

Bemerkung 7.6. Für beliebige $x_n, x \in E$ gilt

$$x_n \longrightarrow x \text{ in } E \quad \Longleftrightarrow \quad \delta_{x_n} \xrightarrow{\mathrm{w}} \delta_x \text{ in } \mathcal{P}(E).$$

Tatsächlich, aus $x_n \longrightarrow x$ folgt für beliebige $f \in \mathcal{C}(E)$

$$\int f\, d\delta_{x_n} = f(x_n) \longrightarrow f(x) = \int f\, d\delta_x.$$

Gelte umgekehrt $\delta_{x_n} \xrightarrow{\mathrm{w}} \delta_x$. Da die Funktion $f(y) := e^{-\varrho(y,x)}$ stetig und beschränkt ist folgt

$$e^{-\varrho(x_n,x)} = \int f\, d\delta_{x_n} \longrightarrow \int f\, d\delta_x = e^{-\varrho(x,x)} = 1,$$

das heißt $\varrho(x_n, x) \longrightarrow 0$.

Behauptung 7.7. *Seien (E,ϱ) und (E',ϱ') metrische Räume und $h : E \longrightarrow E'$ eine stetige Abbildung. Dann gilt für beliebige $\mu_n, \mu \in \mathcal{P}(E)$:*

$$\mu_n \xrightarrow{w} \mu \text{ in } \mathcal{P}(E) \quad \Longrightarrow \quad \mu_n \circ h^{-1} \xrightarrow{w} \mu \circ h^{-1} \text{ in } \mathcal{P}(E').$$

Beweis. Angenommen $\mu_n \xrightarrow{\mathrm{w}} \mu$. Für beliebige $f \in \mathcal{C}(E')$ ist $f \circ h \in \mathcal{C}(E)$ und folglich gilt mit dem Integraltransformationssatz

$$\int_{E'} f\, d\mu_n \circ h^{-1} = \int_E f \circ h\, d\mu_n \longrightarrow \int_E f \circ h\, d\mu = \int_{E'} f\, d\mu \circ h^{-1}.$$

□

Für unstetige h gilt diese Aussage im Allgemeinen nicht (vergleiche Bemerkung 7.6). Ist $h : E \longrightarrow E'$ Borel-messbar, so ist die Menge der Unstetigkeitsstellen

$$D_h := \{x \in E : \quad h \text{ unstetig in } x\}$$

$\mathfrak{B}_E$-messbar und es gilt die folgende Verallgemeinerung der Behauptung 7.7.

Satz 7.8. *(ohne Beweis)*
Seien (E,ϱ) und (E',ϱ') metrische Räume, die Abbildung $h : E \longrightarrow E'$ Borel-messbar und D_h die Menge der Unstetigkeitsstellen von h. Dann gilt für beliebige $\mu_n, \mu \in \mathcal{P}(E)$ mit $\mu(D_h) = 0$:

$$\mu_n \xrightarrow{w} \mu \quad \Longrightarrow \quad \mu_n \circ h^{-1} \xrightarrow{w} \mu \circ h^{-1}.$$

Lemma 7.9. *Seien* (E_1, ϱ_1) *und* (E_2, ϱ_2) *separable metrische Räume. Dann ist* $E := E_1 \times E_2$ *versehen mit der Metrik*

$$\varrho((x_1, x_2), (y_1, y_2)) = \varrho_1(x_1, y_1) + \varrho_2(x_2, y_2), \qquad (x_1, x_2), (y_1, y_2) \in E,$$

ein separabler metrischer Raum, und es gilt

$$\mathfrak{B}_E = \mathfrak{B}_{E_1} \otimes \mathfrak{B}_{E_2}.$$

Beweisskizze. Man zeigt leicht, dass ϱ eine Metrik ist. Sind $\left\{x_m^{(1)} : \quad m \in \mathbb{N}\right\}$ und $\left\{x_n^{(2)} : \quad n \in \mathbb{N}\right\}$ Separabilitätsmengen in E_1 bzw. E_2 (das heißt abzählbare dichte Teilmengen), so ist $\left\{(x_m^{(1)}, x_n^{(2)}) : \quad m, n \in E\right\}$ eine Separabilitätsmenge in E. Also ist E separabel.
Die Mengen der Gestalt $G_1 \times G_2$ mit G_1 und G_2 offen in E_1 bzw. E_2 erzeugen die σ-Algebra $\mathfrak{B}_{E_1} \otimes \mathfrak{B}_{E_2}$ und sind offen. Deshalb ist $\mathfrak{B}_{E_1} \otimes \mathfrak{B}_{E_2} \subseteq \mathfrak{B}_E$.
Sind E_1 und E_2 separabel, so läßt sich jede offene Menge in E als abzählbare Vereinigung von Mengen der Gestalt $G_1 \times G_2$ mit G_1 und G_2 offen in E_1 bzw. E_2 darstellen. Folglich ist $\mathfrak{B}_E \subseteq \mathfrak{B}_{E_1} \otimes \mathfrak{B}_{E_2}$. $\square$

Definition 7.10. Seien X_n, X Zufallsvariablen auf einem gemeinsamen Wahrscheinlichleitsraum $(\Omega, \mathfrak{F}, \mathbb{P})$ mit Werten in einem *separablen* metrischen Raum (E, ϱ). Man sagt, die Folge (X_n) *konvergiert in Wahrscheinlichkeit gegen* X, falls für alle $\varepsilon > 0$ gilt:

$$\lim_{n \to \infty} \mathbb{P}(\varrho(X_n, X) > \varepsilon) = 0$$

Schreibweise: $\qquad X_n \xrightarrow{\mathbb{P}} X.$

Das Lemma 7.9 garantiert, dass für *separable* metrische Räume E die Abbildung $\varrho : E \times E \longrightarrow \mathbb{R}$ bezüglich $\mathfrak{B}_E \otimes \mathfrak{B}_E$ (und nicht nur bezüglich $\mathfrak{B}_{E \times E}$) messbar und folglich $\varrho(X_n, X)$ in obiger Definition eine Zufallsgröße ist.

Behauptung 7.11. *Seien* X_n, X *Zufallsvariablen mit Werten in einem separablen metrischen Raum* (E, ϱ). *Dann gilt*

$$X_n \xrightarrow{\mathbb{P}} X \quad \Longrightarrow \quad X_n \xrightarrow{d} X.$$

Beweis. Es gelte $X_n \xrightarrow{\mathbb{P}} X$. Wir wählen eine abgeschlossene Menge $F \subseteq E$ und $\varepsilon > 0$ beliebig. Wie früher bezeichnen wir mit $F^\varepsilon := \{x \in E : \quad \varrho(x, F) \leq \varepsilon\}$ die abgeschlossene ε-Umgebung von F.
Dann gilt die Abschätzung

$$\mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F^\epsilon + \mathbb{P}(\varrho(X_n, X) > \varepsilon)$$

und folglich da $\mathbb{P}(\varrho(X_n, X) > \varepsilon) \longrightarrow 0$ für $n \to \infty$

$$\limsup_{n \to \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F^\varepsilon).$$

Wegen $F^\varepsilon \downarrow F^0 = F$ für $\varepsilon \downarrow 0$ (und der Abgeschlossenheit von F) folgt hieraus

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F).$$

Deshalb liefert die Aussage (ii) des Portmanteau-Theorems (siehe Satz 7.3) die gewünschte Konvergenz $X_n \xrightarrow{d} X$. □

Behauptung 7.12. *Seien X, X_n, X_n' Zufallsvariablen mit Werten in einem separablen metrischen Raum (E, ϱ). Dann gilt*

$$X_n \xrightarrow{d} X \text{ und } \varrho(X_n, X_n') \xrightarrow{\mathbb{P}} 0 \quad \Longrightarrow \quad X_n' \xrightarrow{d} X.$$

Beweis. Seien F abgeschlossen, $\varepsilon > 0$ und F^ε die abgschlossene ε-Umgebung von F. Dann gilt

$$\mathbb{P}(X_n' \in F) \leq \mathbb{P}(X_n \in F^\varepsilon) + \mathbb{P}(\varrho(X_n', X_n) > \varepsilon).$$

Unter Benutzung der Aussage (ii) des Portmanteau-Theorems folgt

$$\limsup_{n\to\infty} \mathbb{P}(X_n' \in F) \leq \limsup_{n\to\infty} \mathbb{P}(X_n \in F^\varepsilon) \leq \mathbb{P}(X \in F^\varepsilon)$$

Wegen $F^\varepsilon \downarrow F$ für $\varepsilon \downarrow 0$ (und der Abgeschlossenheit von F) ist also

$$\limsup_{n\to\infty} \mathbb{P}(X_n' \in F) \leq \mathbb{P}(X \in F)$$

und eine nochmalige Anwendung des Portmanteau-Theorems liefert die gewünschte Konvergenz $X_n' \xrightarrow{d} X$. □

7.2. Schwache Kompaktheit und Straffheit

Behauptung 7.13. *Sei (E, ϱ) ein vollständiger separabler metrischer Raum und $\mu \in \mathcal{P}(E)$. Dann existiert zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subseteq E$ mit $\mu(K^\varepsilon) < \varepsilon$. (Das heißt μ ist auf einer σ-kompakten Menge konzentriert: Es existiert eine Folge (K_n) kompakter Mengen mit $\mu(\bigcup_n K_n) = 1$.)*

Unendlich-dimensionale Hilbert- und Banachräume sind im Allgemeinen nicht σ-kompakt. In solchen Räumen sind σ-kompakte Mengen „sehr dünn“.

Beweis. Seien $\mu \in \mathcal{P}(E)$ und $\varepsilon > 0$ beliebig gewählt.
Da E separabel ist, findet man zu jedem $n \in \mathbb{N}$ eine Folge $(B_{nk})_{k\in\mathbb{N}}$ offener Kugeln vom Radius $1/n$, die E überdecken. Insbesondere existiert zu jedem $n \in \mathbb{N}$ ein N_n mit

$$\mu\Big(\bigcup_{k\leq N_n} B_{nk}\Big) > 1 - \varepsilon 2^{-n}.$$

Die Menge

$$\bigcap_{n=1}^{\infty} \bigcup_{k \leq N_n} B_{nk}$$

ist relativkompakt, das heißt sie besitzt eine kompakte Abschließung K. Tatsächlich, für jedes beliebige $\delta > 0$ läßt sich diese Menge durch endlich viele offene δ-Kugeln überdecken („Existenz eines endlichen δ-Netzes“), und da E vollständig ist, impliziert dies die relative Kompaktheit dieser Menge. Außerdem gilt

$$\mu(K^c) \leq \mu(\bigcup_{n=1}^{\infty} \bigcap_{k \leq N_k} B_{nk}^c) < \sum_{n=1}^{\infty} \varepsilon 2^{-n} = \varepsilon.$$

□

Definition 7.14. Sei (E, ϱ) ein metrischer Raum.

a) Eine Menge $\Gamma \subseteq \mathcal{P}(E)$ heißt (sequentiell schwach) *relativkompakt*, wenn man aus jeder Folge von Wahrscheinlichkeitsmaßen in Γ eine schwach konvergente Teilfolge auswählen kann. Liegt der Grenzwert jeder solchen Teilfolge in Γ, so nennt man Γ (sequentiell schwach) *kompakt*.

b) Eine Menge $\Gamma \subseteq \mathcal{P}(E)$ heißt *straff* (englisch „tight“), wenn zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subseteq E$ existiert sodass für alle $\mu \in \Gamma$

$$\mu(K^c) < \varepsilon.$$

Bemerkung 7.15.

1. Ist (E, ϱ) vollständig und separabel, so ist jedes Wahrscheinlichkeitsmaß auf $(E, \mathfrak{B}_E)$ straff (Behauptung 7.13).
2. Ist E kompakt, so ist $\mathcal{P}(E)$ straff.
3. $\mathcal{P}(\mathbb{R})$ ist nicht straff, da $\{\delta_n : n \in \mathbb{N}\}$ nicht straff ist.

Satz 7.16. *(Prokhorov) (ohne Beweis)*
Sei (E, ϱ) ein vollständiger separabler metrischer Raum und $\Gamma \subseteq \mathcal{P}(E)$. Dann gilt

$$\Gamma \text{ relativkompakt} \quad \Longleftrightarrow \quad \Gamma \text{ straff}$$

Um den Satz von Prokhorov anzuwenden, benötigt man Straffheitkriterien für konkrete metrische Räume (E, ϱ). Für uns ist die Konvergenz in Verteilung von Folgen stochastischer Prozesse mit *stetigen* Pfaden besonders interessant.
Deshalb betrachten wir den vollständigen separablen metrischen Raum

$$\mathcal{C} = \mathcal{C}[0,1] \text{ der stetigen Funktionen } x : [0,1] \longrightarrow \mathbb{R}$$

mit der Supremumsmetrik

$$\varrho(x, y) := \sup_{t \in [0,1]} |x(t) - y(t)|, \qquad x, y \in \mathcal{C}.$$

Wir benötigen zunächst eine Charakterisierung der kompakten Teilmengen von $\mathcal{C}$. Hierzu führen wir das Stetigkeitsmodul w_x einer Funktion $x \in \mathcal{C}$ ein:

$$w_x(\delta) := \sup_{\substack{s,t\in[0,1] \\ |s-t|<\delta}} |x(s) - x(t)|\,, \qquad \delta > 0.$$

Da x gleichmäßig stetig ist, gilt $w_x(\delta) \downarrow 0$.

Satz 7.17. *(Arzelá-Ascoli)*
Eine Menge $K \subseteq \mathcal{C}$ ist genau dann relativkompakt, wenn die beiden folgenden Aussagen erfüllt sind:

(i) $\sup_{x\in K} |x(0)| < \infty$*;*

(ii) $\lim_{\delta\downarrow 0} \sup_{x\in K} w_x(\delta) = 0.$

(„Gleichmäßige Beschränktheit und gleichgradige Stetigkeit der Menge der Funktionen aus K“.)

Satz 7.18. *Eine Menge $\Gamma \subseteq \mathcal{P}(\mathcal{C})$ ist genau dann straff, wenn die beiden folgenden Bedingungen erfüllt sind:*

(i) $\lim_{a\to\infty} \sup_{\mu\in\Gamma} \mu(\{x:\ |x(0) > a|\}) = 0$*;*

(ii) $\lim_{\delta\downarrow 0} \sup_{\mu\in\Gamma} \mu(\{x:\ w_x(\delta) > \varepsilon\}) = 0, \qquad \forall \varepsilon > 0.$

Beweis.
„$\Longrightarrow$“:
Sei $\Gamma \subseteq \mathcal{P}(\mathcal{C})$ straff und $\varepsilon > 0$. Dann existiert eine Folge (K_n) kompakter Teilmengen in $\mathcal{C}$ mit

$$\lim_{n\to\infty} \sup_{\mu\in\Gamma} \mu(K_n^c) = 0.$$

Nach dem Satz von Arzelá-Ascoli existieren zu jedem n Zahlen $a_n, \delta_n > 0$ mit

$$K_n \subseteq \{x:\ |x(0)| \le a_n\}\,,\ K_n \subseteq \{x:\ w_x(\delta_n) \le \varepsilon\}\,.$$

Daraus folgt

$$\sup_{\mu\in\Gamma} \mu(\{x:\ |x(0)| > a_n\}) \le \mu(K_n^c) \underset{n\to\infty}{\longrightarrow} 0$$

und

$$\sup_{\mu\in\Gamma} \mu(\{x:\ w_x(\delta_n) > \varepsilon\}) \le \mu(K_n^c) \underset{n\to\infty}{\longrightarrow} 0.$$

Dies impliziert (i) und (ii).

„$\Longleftarrow$“:
Seien (i) und (ii) erfüllt und $\varepsilon > 0$ beliebig vorgegeben. Dann existiert ein $a > 0$, so dass

$$\sup_{\mu\in\Gamma} \mu(\{x: \ |x(0)| > a\}) < \frac{\varepsilon}{2} \text{ mit } \{x: \ |x(0)| > a\} =: A^c$$

ist. Außerdem findet man zu jedem $k \in \mathbb{N}$ ein $\delta_k > 0$ mit $\delta_k \downarrow 0$ und

$$\sup_{\mu\in\Gamma} \mu\left(\left\{x: \ w_x(\delta_k) > \frac{1}{k}\right\}\right) < \frac{\varepsilon}{2} 2^{-k} \text{ mit } \left\{x: \ w_x(\delta_k) > \frac{1}{k}\right\} =: A_k^c.$$

Die Menge $K := A \cap \bigcap_{k=1}^{\infty} A_k$ ist abgeschlossen (als Durchschnitt abgeschlossener Mengen) und erfüllt die Vorraussetzungen (i) und (ii) des Satzes von Arzelá-Ascoli. Also ist K kompakt. Außerdem gilt

$$\sup_{\mu\in\Gamma} \mu(K^c) \leq \sup_{\mu\in\Gamma}(\mu(A^c) + \sum_{k=1}^{\infty} \mu(A_k^c)) < \frac{\varepsilon}{2} + \sum_{k=1}^{\infty} \frac{\varepsilon}{2} 2^{-k} = \varepsilon.$$

Das heißt zu beliebigem $\varepsilon > 0$ findet man eine kompakte Menge $K \subseteq \mathcal{C}$ mit

$$\sup_{\mu\in\Gamma} \mu(K^c) < \varepsilon.$$

Mit anderen Worten Γ ist straff. □

Das folgende hinreichende Kriterium ist für Anwendungen oft komfortabler.

Satz 7.19. *Für die Straffheit einer Menge $\Gamma \subseteq \mathcal{P}(\mathcal{C})$ sind folgende Bedingungen hinreichend.*

(i) $\lim\limits_{a\to\infty} \sup\limits_{\mu\in\Gamma} \mu(\{x: \ |x(0)| > a\}) = 0,$

(ii) $\lim\limits_{\delta\downarrow 0} \sup\limits_{\mu\in\Gamma} \sup\limits_{s\in[0,1]} \frac{1}{\delta}\mu\left(\left\{x: \sup\limits_{0\leq t-s<\delta} |x(t) - x(s)| > \varepsilon\right\}\right) = 0,$ *für alle* $\varepsilon > 0$

Beweis. Es genügt zu zeigen, dass aus der Bedingung (ii) die Bedingung (ii) des Satzes 7.18 folgt. Zunächst gilt

$$\begin{aligned} w_x(\delta) &= \sup_{\substack{s\in[0,1] \\ 0\leq t-s\leq\delta}} |x(t) - x(s)| \\ &= \max_{1\leq i<\delta^{-1}} \sup_{\substack{s\in[(i-1)\delta, i\delta] \\ 0\leq t-s\leq\delta}} |x(t) - x(s)| \end{aligned}$$

Benutzen wir dabei die Abschätzung

$$|x(t) - x(s)| \leq |x(t) - x((i-1)\delta)| + |x(s) - x((i-1)\delta)|, \text{ falls } t \in [(i-1)\delta, i\delta],$$

und

$$|x(t) - x(s)| \leq |x(t) - x(i\delta)| + |x(i\delta) - x((i-1)\delta)| + |x(s) - x((i-1)\delta)|.$$

Falls $t \in [i\delta, (i+1)\delta]$, so folgt

$$w_x(\delta) \leq 3 \max_{1 \leq i < \delta^{-1}} \sup_{t \in [(i-1)\delta, i\delta]} |x(t) - x((i-1)\delta)| .$$

Deshalb ist, da

$$\begin{aligned} &\mu\left(\left\{x : \sup_{t \in [(i-1)\delta, i\delta]} |x(t) - x(s)| > \frac{\varepsilon}{3}\right\}\right) \\ \leq \quad &\delta \sup_{s \in [0,1]} \frac{1}{\delta}\mu\left(\left\{x : \sup_{0 \leq t-s \leq \delta} |x(t) - x(s)| > \frac{\varepsilon}{3}\right\}\right) \end{aligned}$$

$$\begin{aligned} \mu(\{x : w_\delta(x) > \varepsilon\}) \quad &\leq \quad \sum_{1 \leq i < \delta^{-1}} \mu\left(\left\{x : \sup_{t \in [(i-1)\delta, i\delta]} |x(t) - x(s)| > \frac{\varepsilon}{3}\right\}\right) \\ &\leq \quad \sup_{s \in [0,1]} \frac{1}{\delta}\mu\left(\left\{x : \sup_{0 \leq t-s \leq \delta} |x(t) - x(s)| > \frac{\varepsilon}{3}\right\}\right). \end{aligned}$$

Nach Vorraussetzung (ii) konvergiert der Ausdruck auf der rechten Seite für $\delta \downarrow 0$ gleichmäßig in $\mu \in \Gamma$ gegen 0, woraus die Vorraussetzung (ii) des Satzes 7.18 folgt. $\square$

Wie früher bezeichnen wir mit $\pi_{t_1,\ldots,t_r} : \mathcal{C}[0,1] \longrightarrow \mathbb{R}^r \quad (0 \leq t_1 < \cdots < t_r \leq 1)$ die endlich-dimensionale Projektion definiert durch

$$\pi_{t_1,\ldots,t_r}(x) = (x(t_1), \ldots, x(t_r)), \qquad x \in \mathcal{C}[0,1].$$

Der folgende Satz erhellt, wie man die Straffheit zum Studium der schwachen Konvergenz von Wahrscheinlichkeitsmaßen einsetzen kann.

Satz 7.20. *Seien $\mu_n, \mu \in \mathcal{P}(\mathcal{C})$. Dann gilt $\mu_n \xrightarrow{w} \mu$ denau dann, wenn folgende Bedingungen erfüllt sind:*

(i) $\mu_n \circ \pi^{-1}_{t_1,\ldots,t_r} \xrightarrow{w} \mu \circ \pi^{-1}_{t_1,\ldots,t_r}$ für alle $r \in \mathbb{N}$ und $0 \leq t_1 < \cdots < t_r \leq 1$. (Schwache Konvergenz der endlichen-dimensionalen Verteilungen);

(ii) $\{\mu_n : n \in \mathbb{N}\}$ ist straff.

Beweis.
„$\Longrightarrow$“:
Wegen der Stetigkeit der Projektionen folgt die Behauptung (i) aus Behauptung 7.7. Da jede konvergente Folge relativkompakt ist, liefert der Satz von Prokhorov (Satz 7.16) sofort die Straffheit von $\{\mu_n : n \in \mathbb{N}\}$

„$\Longleftarrow$“:
Aus (ii) folgt mit dem Satz von Prokhorov (Satz 7.16) die relative Kompaktheit von $\{\mu_n : n \in \mathbb{N}\}$. Würde (μ_n) nicht schwach konvergieren, gäbe es deshalb zwei

Teilfolgen (μ'_n) und (μ''_n) mit $\mu'_n \xrightarrow{w} \mu'$ und $\mu''_n \xrightarrow{w} \mu''$ fur gewisse $\mu', \mu'' \in \mathcal{P}(\mathcal{C})$ mit $\mu' \neq \mu''$. Da aber die endlich-dimensionalen Projektionen stetig sind, folgt aus (i), dass die endlich-dimensionalen Projektionen von μ' und μ'' mit denen von μ übereinstimmen. Hieraus folgt $\mu' = \mu'' = \mu$ (was ein Widerspruch zur Annahme ist). (Die Algebra der Zylindermengen erzeugt $\mathfrak{B}_{\mathcal{C}}$, Übungsaufgabe). Also konvergiert (μ_n) schwach gegen μ. □

Bemerkung 7.21. Im Beweis wurde folgende leicht zu beweisende Eigenschaft der schwachen Konvergenz benutzt: Kann man aus jeder Teilfolge einer Folge von Wahrscheinlichkeitsmaßen eine schwach konvergente Teilfolge auswählen und besitzen alle diese Teilfolgen den gleichen Grenzwert, so konvergiert die gesamte Folge. (Dies gilt für Konvergenz in metrischen Räumen aber z.B. nicht für die nicht metrisierbare Konvergenz fast sicher von Zufallsgrößen.)

7.3. Der Satz von Donsker und seine Anwendungen

Wir betrachten in leicht verallgemeinerter Form das bereits in der Motivation in Abschnitt 7.1 skizzierte Beispiel.

Gegeben sei eine Folge $(X_n)_{n\in\mathbb{N}}$ unabhängiger, identisch verteilter Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, \mathbb{P})$ mit

$$\mathbb{E}X_1 = 0, \qquad 0 < \sigma^2 := \mathbb{E}X_1^2 < \infty.$$

Mit

$$S_n := \sum_{k=1}^{n} X_k, \qquad n \in \mathbb{N}_0,$$

bezeichnen wir die zugehörigen Partialsummen. Dann wird für jedes $n \in \mathbb{N}$ durch

$$Y_n(t) := \frac{1}{\sigma\sqrt{n}} \left(S_{[nt]} + (nt - [nt])X_{[nt]+1} \right), \qquad t \in [0,1],$$

wobei $(nt - [nt])X_{[nt]+1}$ die lineare Interpolation zwischen $S_{[nt]}$ und $S_{[nt]+1}$ beschreibt, eine Zufallsvariable Y_n mit Werten in $(\mathcal{C}, \mathfrak{B}_{\mathcal{C}})$ definiert.

Mit $\mathcal{W}$ bezeichnen wir das *Wienermaß*, das heißt die Verteilung des Wienerprozesses auf $(\mathcal{C}, \mathfrak{B}_{\mathcal{C}})$.

Satz 7.22. *(Donsker)*
Unter den obigen Vorraussetzungen gilt

$$\mathbb{P} \circ Y_n^{-1} \xrightarrow{w} \mathcal{W} \qquad \textit{für } n \to \infty.$$

(Das heißt die Folge (Y_n) stochastischer Prozesse konvergiert auf $(\mathcal{C}, \mathfrak{B}_{\mathcal{C}})$ in Verteilung gegen einen aufdas Zeitintervall $[0,1]$ eingeschränkten Wienerprozess.)

Beweis. Der Beweis beruht auf einer Anwendung von Satz 7.20. Hierzu sei $\mathbb{W} = (W_t)_{t\in[0,1]}$ ein Wienerprozess (o.B.d.A. ebenfalls auf $(\Omega, \mathfrak{F}, \mathbb{P})$ definiert).

1^o *Konvergenz der endlich-dimensionalen Verteilungen*
Sei $0 \leq t_1 < \cdots < t_r \leq 1$ und $t_0 := 0$. Es ist

$$(Y_n(t_1), \ldots, Y_n(t_r)) \xrightarrow{d} (W_{t_1}, \ldots, W_{t_r}) \qquad \text{für } n \to \infty$$

zu zeigen. Wir beweisen

$$(Y_n(t_1) - Y_n(t_0), \ldots, Y_n(t_r) - Y_n(t_{r-1})) \xrightarrow{d} (W_{t_1} - W_{t_0}, \ldots, W_{t_r} - W_{t_{r-1}}), \tag{7.2}$$

woraus die obige Aussage durch lineare (und damit auch stetige) Transformation mit Hilfe der Matrix

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}$$

folgt (vergleiche Behauptung 7.7). Nun ist

$$Y_n(t) = Y_n'(t) + Y_n''(t)$$

mit

$$Y_n'(t) := \frac{1}{\sigma\sqrt{n}} S_{[nt]} \text{ und } Y_n''(t) := \frac{(nt - [nt])}{\sigma\sqrt{n}} X_{[nt]+1}.$$

Aufgrund des zentralen Grenzwertsatzes gilt

$$Y_n'(t_k) - Y_n'(t_{k-1}) \xrightarrow{d} W_{t_k} - W_{t_{k-1}} \tag{7.3}$$

für $k = 1, \ldots, n$. Tatsächlich,

$$\frac{S_{[nt_k]} - S_{[nt_{k-1}]}}{\sigma\sqrt{[nt_k] - [nt_{k-1}]}}$$

konvergiert für $n \to \infty$ in Verteilung gegen eine Standardnormalverteilung und $([nt_k]-[nt_{k-1}])/n \longrightarrow t_k - t_{k-1}$. Da $W_{t_k} - W_{t_{k-1}}$ $\mathcal{N}(0, t_k - t_{k-1})$-verteilt ist, folgt hieraus (zum Beispiel mit Verteilungsfunktionen) leicht die Behauptung (7.3). Da die Zuwächse auf der linken und der rechten Seite von (7.3) für verschiedene k unabhängig sind, erhält man hieraus (zum Beispiel mit charakteristischen Funktionen) die Konvergenz

$$(Y_n'(t_1) - Y_n'(t_0), \ldots, Y_n'(t_r) - Y_n'(t_{r-1})) \xrightarrow{d} (W_{t_1} - W_{t_0}, \ldots, W_{t_r} - W_{t_{r-1}}). \tag{7.4}$$

Mit der Chebyshev-Ungleichung erhält man

$$Y_n''(t_k) \xrightarrow{\mathbb{P}} 0 \qquad \text{für } n \to \infty$$

($k = 0, 1, \ldots, r$) und folglich

$$(Y_n''(t_1) - Y_n''(t_0), \ldots, Y_n''(t_r) - Y_n''(t_{r-1})) \xrightarrow{\mathbb{P}} 0 \qquad \text{für } n \to \infty. \tag{7.5}$$

2^o *Straffheit* der Folge (Y_n)
Wir benutzen das hinreichende Straffheitskriterium aus Satz 7.19:

(i) $\lim_{a\to\infty} \sup_n \mathbb{P}\left(|Y_n(0)| > a\right) = 0;$

(ii) $\lim_{\delta\downarrow\infty} \limsup_{n\to\infty} \sup_{s\in[0,1]} \frac{1}{\delta}\mathbb{P}\left(\sup_{0\le t-s\le\delta} |Y_n(t) - Y_n(s)| > \varepsilon\right) = 0, \qquad \forall\varepsilon > 0.$

(Dabei haben wir in der Formulierung von (ii) $\sup_n$ durch $\limsup_{n\to\infty}$ ersetzt. Dies kann man für Folgen von Wahrscheinlichkeitsmaßen im Satz 7.18 machen (da die enstsprechende Bedingung für jedes individuelle n erfüllt ist) und dies überträgt sich dann auf den Beweis von Satz 7.19.)
Wir merken an, dass

$$\sup_{0\le t-s\le\delta} |Y_n(t) - Y_n(s)|$$

durch ein entprechendes Maximum über zu s und t benachbarte Interpolationspunkte nach oben abgeschätzt werden kann. Genauer, für jeden Pfad $Y_n(.)$ ist entweder

$$\begin{aligned} \sup_{0\le t-s\le\delta} |Y_n(t) - Y_n(s)| \quad &\le \quad \max_{0\le l\le n\delta+2} \left|\frac{S_{[ns]+l} - S_{[ns]}}{\sigma\sqrt{n}}\right| \\ \text{oder} \quad &\le \quad \max_{0\le l\le n\delta+2} \left|\frac{S_{[ns]+1+l} - S_{[ns]+1}}{\sigma\sqrt{n}}\right|. \end{aligned}$$

Deshalb erhalten wir

$$\begin{aligned} &\sup_{s\in[0,1]} \mathbb{P}\left(\sup_{0\le t-s\le\delta} |Y_n(t) - Y_n(s)| > \varepsilon\right) \\ \le \quad & 2 \sup_{0\le k\le n} \mathbb{P}\left(\max_{0\le l\le[n\delta]+2} |S_{k+l} - S_k| > \varepsilon\delta\sqrt{n}\right) \\ = \quad & 2\mathbb{P}\left(\max_{0\le l\le[n\delta]+2} |S_l| > \varepsilon\delta\sqrt{n}\right) \\ = \quad & 2\mathbb{P}\left(\max_{0\le l\le[n\delta]+2} \left|\frac{S_l}{\sigma\sqrt{[n\delta]+2}}\right| > \varepsilon\sqrt{\frac{n}{[n\delta]+2}}\right) \end{aligned}$$

Wir benutzen nun die Ungleichung

$$\mathbb{P}\left(\max_{0\le l\le m} \left|\frac{S_l}{\sigma\sqrt{m}}\right| \ge \lambda\right) \le 2\mathbb{P}\left(\left|\frac{S_m}{\sigma\sqrt{m}}\right| \ge \lambda - \sqrt{2}\right) \tag{7.6}$$

($\lambda > 0,\ m \in \mathbb{N}$), die wir weiter unten beweisen. Damit folgt mit dem Zentralen Grenzwertsatz ($*$)

$$\begin{aligned}
&\mathbb{P}\left(\max_{0\le l\le[n\delta]+2}\left|\frac{S_l}{\delta\sqrt{[n\delta]+2}}\right| \ge \varepsilon\sqrt{\frac{n}{[n\delta]+2}}\right) \\
\le\quad & 2\mathbb{P}\left(\left|\frac{S_{[n\delta]+2}}{\delta\sqrt{[n\delta]+2}}\right| > \varepsilon\sqrt{\frac{n}{[n\delta]+2}}-\sqrt{2}\right) \\
\underset{(*)}{\xrightarrow{n\to\infty}}\quad & \frac{4}{\sqrt{2\pi}\int\limits_{\frac{\varepsilon}{\sqrt{\delta}}-\sqrt{2}}^{\infty} e^{-x^2/2}\,dx}
\end{aligned}$$

Setzen wir die obigen Teilergebnisse zusammen, so erhalten wir schließlich

$$\limsup_{n\to\infty}\sup_{s\in[0,1]}\frac{1}{\delta}\mathbb{P}\left(\sup_{0\le t-s\le\delta}|Y_n(t)-Y_n(s)|>\varepsilon\right) \le \frac{8}{\sqrt{2\pi}}\frac{1}{\delta}\int\limits_{\frac{\varepsilon}{\sqrt{\delta}}-\sqrt{2}}^{\infty} e^{-x^2/2}\,dx \quad (7.7)$$

Wegen

$$\int\limits_a^\infty e^{-x^2/2}\,dx \le \frac{1}{a}\int\limits_a^\infty xe^{-x^2/2}\,dx = \frac{1}{a}e^{-a^2/2}, \qquad a>0,$$

konvergiert der Ausdruck auf der rechten Seite von (7.7) für $\delta > 0$ gegen 0. Dies beweist die Bedingung (ii).

Es bleibt die Ungleichung (7.6) zu beweisen. Hierzu sei o.B.d.A. $\lambda > \sqrt{2}$. Wir betrachten die Ereignisse

$$A_i := \left\{\max_{j<i}\left|\frac{S_j}{\sigma\sqrt{m}}\right| < \lambda \le \left|\frac{S_i}{\sigma\sqrt{m}}\right|\right\} \qquad (i=1,...,m)$$

Dann gilt

$$\begin{aligned}
\mathbb{P}\left(\max_{i\le m}\left|\frac{S_i}{\sigma\sqrt{m}}\right|\ge\lambda\right) \quad &\le\quad \mathbb{P}\left(\left|\frac{S_m}{\sigma\sqrt{m}}\right|\ge\lambda-\sqrt{2}\right) \\
&+\quad \sum_{i=1}^{m-1}\mathbb{P}\left(A_i\cap\left\{\left|\frac{S_m}{\sigma\sqrt{m}}\right|<\lambda-\sqrt{2}\right\}\right)
\end{aligned}$$

Da aus $|S_i/(\sigma\sqrt{m})| \ge \lambda$ und $|S_m/\sigma\sqrt{m}| < \lambda-\sqrt{2}$ die Ungleichung

$$\left|(S_m-S_i)/(\sigma\sqrt{m})\right| \ge \sqrt{2}$$

folgt, erhalten wir wegen der Unabhängigkeit (*), der Chebyshevschen Ungleichung (**) und da $\mathbb{E}\left(\frac{S_m-S_i}{\sigma\sqrt{m}}\right)^2 \leq 1$

$$\begin{aligned}
\mathbb{P}\left(A_i \cap \left\{\left|\frac{S_m}{\sigma\sqrt{m}}\right| < \lambda - \sqrt{2}\right\}\right) &\leq \mathbb{P}\left(A_i \cap \left\{\left|\frac{S_m - S_i}{\sigma\sqrt{m}}\right| > \sqrt{2}\right\}\right) \\
&\underset{(*)}{=} \mathbb{P}(A_i)\mathbb{P}\left(\left|\frac{S_m - S_i}{\sigma\sqrt{m}}\right| > \sqrt{2}\right) \\
&\underset{(**)}{=} \frac{1}{2}\mathbb{P}(A_i)\mathbb{E}\left(\frac{S_m - S_i}{\sigma\sqrt{m}}\right)^2 \\
&\leq \frac{1}{2}\mathbb{P}(A_i)
\end{aligned}$$

Also gilt

$$\sum_{i=1}^{m-1} \mathbb{P}\left(A_i \cap \left\{\left|\frac{S_m}{\sigma\sqrt{m}}\right| < \lambda - \sqrt{2}\right\}\right) \leq \mathbb{P}\left(\max_{i\leq m}\left|\frac{S_i}{\sigma\sqrt{m}}\right| \geq \lambda\right).$$

Durch Einsetzen in (7.8) und „Auflösen“ der erhaltenen Ungleichung folgt die Abschätzung (7.6). □

Beispiel 7.23. Der Zentrale Grenzwertsatz als Spezialfall des Satzes von Donsker: Da die Projektion $\pi_1 : \mathcal{C}[0,1] \longrightarrow \mathbb{R}$ stetig ist, folgt aus dem Satz von Donsker

$$\mathbb{P} \circ Y_n^{-1} \circ \pi_1^{-1} \xrightarrow{\text{w}} \mathbb{P} \circ W^{-1} \circ \pi_1^{-1}$$

(Behauptung 7.7), das heißt

$$\pi_1 \circ Y_n \xrightarrow{d} \pi_1 \circ W.$$

Wegen $\pi_1 \circ Y_n = Y_n(1) = \frac{S_n}{\sigma\sqrt{n}}$, $\pi_1 \circ W = W_1$ und der Tatsache, dass W_1 standardnormalverteilt ist, ist dies nichts anderes als die Aussage des Zentralen Grenzwertsatzes:

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} W_1.$$

Wir zeigen nun, wie man aus Aussagen für einfache symmetrische Irrfahrten auf analoge Aussagen für den Wienerprozess schließen kann.

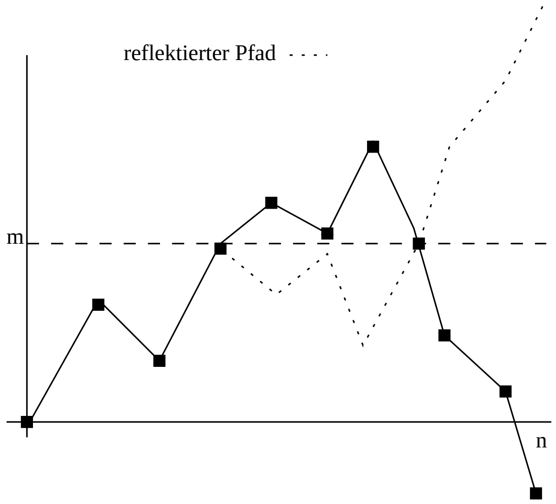
Beispiel 7.24. (*Reflektionspronzip*)
Wir betrachten den Spezialfall

$$\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = +1) = \frac{1}{2},$$

das heißt (S_n) ist eine einfache symmetrische Irrfahrt mit $S_0 := 0$. Für diese gilt das Reflektionsprinzip

$$\mathbb{P}\left(\max_{0\leq k\leq n} S_k \geq m\right) = 2\mathbb{P}(S_n > m) * \mathbb{P}(S_n = m), \qquad m \in \mathbb{N}.$$

Fallunterscheidung : $S_n < m$ (reflektierten Pfad benutzen),
$S_n > m$ und $S_n = m$



Mit Hilfe des Zentralen Grenzwertsatzes folgt hieraus leicht

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(\max_{0\le k\le n} \frac{S_k}{\sigma\sqrt{n}} \ge a\right) &= 2\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \ge a\right)\\ &= \max_{t\in[0,1]} Y_n(t)\\ &= 2\mathbb{P}(W_1 \ge a),\end{aligned}$$

das heißt

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{t\in[0,1]} Y_n(t) \underset{(=)}{>} a\right) = \mathbb{P}\left(W_1 \underset{(=)}{>} a\right), \qquad a > 0. \tag{7.8}$$

(Das Gleichheitszeichen kann weggelassen werden, da der Ausdruck auf der rechten Seite stetig von a abhängt.)
Die Abbildung $F : \mathcal{C}[0,1] \longrightarrow \mathbb{R}$ definiert durch

$$F(x) := \max_{t\in[0,1]} x(t), \qquad x \in \mathcal{C}[0,1],$$

ist stetig. Deshalb liefert eine Kombination des Satzes von Donsker mit der Behauptung 7.7 die Konvergenz

$$\mathbb{P}\circ Y_n^{-1} \circ F^{-1} \xrightarrow{\mathrm{w}} \mathbb{P}\circ W^{-1} \circ F,$$

das heißt

$$\max_{t\in[0,1]} Y_n(t) = F\circ Y_n \xrightarrow{d} F\circ W = \max_{t\in[0,1]} W_t.$$

In Termini der zugehörigen Verteilungsfunktionen ist dies gleichbedeutend mit

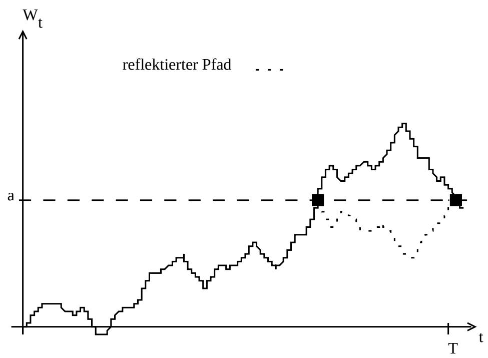
$$\lim_{n\to\infty} \mathbb{P}\left(\max_{t\in[0,1]} Y_n(t) > a\right) = \mathbb{P}\left(\max_{t\in[0,1]} W_t > a\right), \qquad a > 0. \tag{7.9}$$

Ein Vergleich von (7.8) und (7.9) liefert

$$\mathbb{P}\left(\max_{t\in[0,1]} W_t > a\right) = 2\mathbb{P}(W_1 > a), \qquad a > 0.$$

Eine Verallgemeinerung des Beweises auf beliebige Zeitintervalle $[0,T]$, $T > 0$, oder eine zusätzliche Anwendung der Skalierungsinvarianz des Wienerprozesses (Übungsaufgabe) liefert das *Reflektionsprinzip für den Wienerprozess*:

$$\mathbb{P}\left(\max_{t\in[0,T]} W_t \underset{(=)}{>} a\right) = 2\mathbb{P}(W_T \underset{(=)}{>} a) \qquad (T > 0,\ a > 0).$$



Fallunterscheidung $W_T < a$ (Reflektion des Pfades) und $W_T > a$
$(\mathbb{P}(W_T = a = 0))$
Die Heuristik dieses Bildes läßt sich nicht so einfach zu einem Beweis ausbauen, da der Zeitpunkt des ersten Erreichens von a zufällig (und nicht diskret) ist.

Beispiel 7.25. (*Arcsin-Gesetze*)
Die beiden Arcsin-Gesetze für einfache symmetrische Irrfahrten (S_n) lassen sich ebenfalls mit dem Satz von Donsker auf den Wienerprozess übertragen. So folgt etwa aus

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\#\{k \le n:\ S_k > 0\}}{n} \le a\right) = \frac{2}{\pi} \arcsin\sqrt{a}, \qquad 0 \le a \le 1.$$

die Aussage

$$\mathbb{P}(\lambda(\{t \in [0,1] : W_t > 0\}) \le a) = \frac{2}{\pi} \arcsin\sqrt{a}, \qquad 0 \le a \le 1,$$

das heißt das Lebesgue-Maß aller Punkte, in denen der Wienerprozess auf $[0,1]$ positiv ist, besitzt eine Arcsin-Verteilung. Hierzu benutzt man das Funktional

$$F(x) := \lambda(\{t \in [0,1] : \qquad x(t) > 0\}), \qquad x \in \mathcal{C}[0,1].$$

Dieses ist aber zum Beispiel in $x \equiv 0$ unstetig. Man kann aber zeigen, dass F messbar ist und die Menge der Unstetigkeitsstellen von F eine $\mathcal{W}$-Nullmenge ist (Satz von Fubini). Deshalb kann Satz 7.8 benutzt werden.

Stichwortverzeichnis