Maßtheoretische Konzepte der Stochastik

im

Sommersemester 2014

am

Fachbereich Mathematik

der

Universität Hamburg

gelesen von

Holger Drees

Herzlichen Dank an Frau Julia Behnke für ihre Unterstützung bei der Erstellung einer früheren LaTeX-Version des Kurzskripts

Inhaltsverzeichnis

1 Maßräume

1.1 Definition *Sei $\Omega \neq \emptyset$ ein Grundraum. Ein Mengensystem $\mathcal{A} \subset 2^{\Omega}$ heißt* **σ-Algebra** *über Ω, falls*

(i) $\Omega \in \mathcal{A}$

(ii) $A \in \mathcal{A} \quad \Longrightarrow \quad A^c \in \mathcal{A}$

(iii) $A_n \in \mathcal{A} \; \forall \, n \in \mathbb{N} \quad \Longrightarrow \quad \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

Dann heißt $(\Omega, \mathcal{A})$ **Messraum** *und die Mengen $A \in \mathcal{A}$ heißen* **messbare Mengen**. □

1.2 Definition *Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine Abbildung $\mu : \mathcal{A} \to [0, \infty]$ heißt* **Maß**, *falls*

(i) $\mu(\emptyset) = 0$ *(Nulltreue)*

(ii) Für alle disjunkten Mengen $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, gilt $\mu\big(\bigcup_{n \in \mathbb{N}} A_n\big) = \sum_{n \in \mathbb{N}} \mu(A_n)$ (σ-Additivität)

Das Tupel $(\Omega, \mathcal{A}, \mu)$ heißt dann **Maßraum**.
Gilt $\mu(\Omega) = 1$, so wird μ ein **Wahrscheinlichkeitsmaß** *genannt und $(\Omega, \mathcal{A}, \mu)$ heißt* **Wahrscheinlichkeitsraum**. □

Im Folgenden verwenden wir stets die Konvention

$$x + \infty = \infty, \quad x - \infty = -\infty \quad \forall \, x \in \mathbb{R}$$

Analog zu den Sätzen 1.5 und 1.8 der Mathematischen Stochastik gilt

1.3 Satz *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Dann gilt*

(i) $\forall A_1, \ldots, A_n \in \mathcal{A}$ *disjunkt:* $\bigcup_{i=1}^{n} A_i \in \mathcal{A}, \quad \mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mu(A_i)$

(ii) $\forall A, B \in \mathcal{A}, A \subset B, \mu(A) < \infty : \quad B \setminus A \in \mathcal{A}, \quad \mu(B \setminus A) = \mu(B) - \mu(A)$

(iii) $\forall A, B \in \mathcal{A}, \mu(A \cap B) < \infty : \quad \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$

(iv) $\forall A_n \in \mathcal{A}, n \in \mathbb{N} : \quad \mu\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) \leq \sum_{n \in \mathbb{N}} \mu(A_n)$ *("σ-Subadditivität")*

(v) Ist $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, eine Mengenfolge mit $A_n \uparrow A$ (d.h. $A_n \subset A_{n+1}$ für alle $n \in \mathbb{N}$, $\bigcup_{n \in \mathbb{N}} A_n = A$), so folgt $\lim_{n \to \infty} \mu(A_n) = \mu(A)$. (Diese Eigenschaft wird σ-Stetigkeit von unten genannt.)

(vi) *Ist $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, eine Mengenfolge mit $A_n \downarrow A$ (d.h. $A_n \supset A_{n+1}$ für alle $n \in \mathbb{N}$, $\bigcap_{n\in\mathbb{N}} A_n = A$) und ist $\mu(A_m) < \infty$ für ein $m \in \mathbb{N}$, so folgt $A \in \mathcal{A}$ und $\lim_{n\to\infty} \mu(A_n) = \mu(A)$. (Diese Eigenschaft wird σ-Stetigkeit von oben genannt.)* $\square$

1.4 Satz und Definition (s. Mathematische Stochastik Def. 2.2, Lemma 2.3)
Sei $\Omega \neq \emptyset$ und $\mathcal{C} \subset 2^\Omega$. Das Mengensystem

$$\sigma(\mathcal{C}) := \bigcap_{\mathcal{A}\subset 2^\Omega\ \sigma\text{-Algebra},\ \mathcal{C}\subset\mathcal{A}} \mathcal{A}$$

ist die kleinste σ-Algebra, die $\mathcal{C}$ umfasst. Sie heißt die **von $\mathcal{C}$ erzeugte σ-Algebra**, *und $\mathcal{C}$ wird ein* **Erzeugendensystem von $\mathcal{A}$** *genannt.* $\square$

1.5 Beispiel (i) Die Borel-σ-Algebra $\mathbb{B}$ auf $\mathbb{R}$ wird u.a. von den folgenden Mengensystemen erzeugt:

- $\{(a, b] \mid a, b \in \mathbb{R}, a < b\}$
- $\{(-\infty, x] \mid x \in \mathbb{R}\}$
- $\{O \subset \mathbb{R} \mid O \text{ offen}\}$
- $\{A \subset \mathbb{R} \mid A \text{ abgeschlossen}\}$
- $\{K \subset \mathbb{R} \mid K \text{ kompakt}\}$ $\square$

Unter geeigneten Bedingungen existiert zu vorgegebenen Werten $\tilde{\mu}(C)$ für alle C aus einem Erzeugendensystem $\mathcal{C}$ von $\mathcal{A}$ genau ein Maß μ auf $(\Omega, \mathcal{A})$ mit $\mu(C) = \tilde{\mu}(C)$ für alle $C \in \mathcal{C}$.

1.6 Satz und Definition (Maßfortsetzungssatz) *Sei $(\Omega, \mathcal{A})$ ein Messraum.*

(i) **Eindeutigkeitsteil***:*
Sind μ_1, μ_2 Maße auf $(\Omega, \mathcal{A})$ und $\mathcal{C} \subset \mathcal{A}$ mit $\sigma(\mathcal{C}) = \mathcal{A}$ so, dass mit $C_1, C_2 \in \mathcal{C}$ auch $C_1 \cap C_2 \in \mathcal{C}$ (d.h. $\mathcal{C}$ ist **$\cap$-stabil***), und gilt*

- $\mu_1(C) = \mu_2(C) \quad \forall\, C \in \mathcal{C}$,
- $\exists\, C_n \in \mathcal{C} (n \in \mathbb{N})$ *mit* $C_n \uparrow \Omega$ *und* $\mu_1(C_n) < \infty$ *für alle* $n \in \mathbb{N}$,

dann folgt

$$\mu_1(A) = \mu_2(A) \quad \forall\, A \in \mathcal{A}.$$

(ii) **Existenzteil***:*
Sei $\mathcal{C} \subset \mathcal{A}$ ein **Ring** *über Ω, d.h.*

- $\emptyset \in \mathcal{C}$
- $A, B \in \mathcal{C} \implies A \setminus B \in \mathcal{C}, A \cup B \in \mathcal{C}$,

der $\mathcal{A}$ erzeugt, d.h. $\sigma(\mathcal{C}) = \mathcal{A}$. Sei ferner $\tilde{\mu} : \mathcal{C} \to [0, \infty]$ ein **Prämaß***, d.h. es gilt*

- $\tilde{\mu}(\emptyset) = 0$

- *für alle disjunkten Mengen $C_n \in \mathcal{C}$, $n \in \mathbb{N}$, für die $\bigcup_{n\in\mathbb{N}} C_n \in \mathcal{C}$, gilt $\tilde{\mu}\big(\bigcup_{n\in\mathbb{N}} C_n\big) = \sum_{n\in\mathbb{N}} \tilde{\mu}(C_n)$.*

Dann existiert (wenigstens) ein Maß μ auf $(\Omega, \mathcal{A})$ mit

$$\mu(C) = \tilde{\mu}(C) \quad \forall\, C \in \mathcal{C},$$

d.h. μ ist eine Fortsetzung von $\tilde{\mu}$ zu einem Maß auf $(\Omega, \mathcal{A})$. □

1.7 Beispiel (i) Da $\mathcal{C} := \{(a, b] \mid a, b \in \mathbb{R}, a \le b\}$ ein $\cap$-stabiles Erzeugendensystem von $\mathbb{B}$ ist, ist ein so genanntes **Radon-Maß** auf $(\mathbb{R}, \mathbb{B})$, also ein Maß μ mit $\mu(B) < \infty$ für alle beschränkten Mengen $B \in \mathbb{B}$, durch die Werte $\mu(a, b]$, $-\infty < a \le b < \infty$ eindeutig festgelegt.

Das Maß $\lambda\!\!\lambda$ auf $(\mathbb{R}, \mathbb{B})$ mit $\lambda\!\!\lambda(a, b] = b - a$ für alle $-\infty < a \le b < \infty$ heißt **Lebesgue-Maß** auf $\mathbb{R}$. Die Existenz kann man mit Hilfe von Satz 1.6(ii) angewendet auf den Ring

$$\mathcal{R} := \Big\{ \bigcup_{i=1}^{n} (a_i, b_i] \mid -\infty < a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n, n \in \mathbb{N}_0 \Big\}$$

(mit $\bigcup_{i=1}^{0} A_i := \emptyset$) und das darauf definierte Prämaß

$$\tilde{\mu}\Big(\bigcup_{i=1}^{n} (a_i, b_i]\Big) := \sum_{i=1}^{n} (b_i - a_i)$$

nachweisen.

(ii) Aus der Mathematischen Stochastik ist bekannt, dass sich jedes Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathbb{B})$ eindeutig durch seine Verteilungsfunktion $F : \mathbb{R} \to [0, 1]$, $F(x) = P(-\infty, x]$ beschreiben lässt. Dies folgt auch aus (i), da $P(a, b] = F(b) - F(a)$.

Umgekehrt kann man ähnlich wie beim Lebesgue-Maß zeigen, dass zu jeder rechtsseitig stetigen, monoton steigenden Funktion $F : \mathbb{R} \to [0, 1]$ mit $F(-\infty) := \lim_{x\downarrow-\infty} F(x) = 0$ und $F(\infty) := \lim_{x\uparrow\infty} F(x) = 1$ ein Wahrscheinlichkeitsmaß existiert, das F zur Verteilungsfunktion hat (s. Mathematische Stochastik, Korollar und Definition 2.7).

(iii) In Verallgemeinerung des Obigen lässt sich ebenso zeigen:
Zu jeder rechtsseitig stetigen, monoton steigenden Funktion $G : \mathbb{R} \to \mathbb{R}$ existiert genau ein Radon-Maß μ_G, so dass $\mu_G(a, b] = G(b) - G(a)$ für alle $-\infty < a \le b < \infty$. G heißt dann **maßdefinierende Funktion** zu μ_G (s. Behnen und Neuhaus (2003), Satz 39.3). Maßdefinierende Funktionen zu einem Radon-Maß μ sind nur bis auf eine additive Konstante bestimmt, d.h. ist G eine maßdefinierende Funktion von μ, so auch $\tilde{G} = G + c$ für eine beliebige Konstante c in $\mathbb{R}$. Eine maßdefinierende Funktion ist gegeben durch

$$G(x) := \begin{cases} \mu(0, x], & x \ge 0, \\ -\mu(x, 0], & x < 0. \end{cases}$$ □

Zum Beweis von Satz 1.6(i) benötigen wir

1.8 Definition *Sei $\Omega \neq \emptyset$. Ein Mengensystem $\mathcal{D} \subset 2^\Omega$ heißt* **Dynkin-System** *über Ω, falls*

(i) $\Omega \in \mathcal{D}$

(ii) $\forall\, A, B \in \mathcal{D}$ *mit* $A \subset B$: $\quad B \setminus A \in \mathcal{D}$

(iii) $\forall\, A_n \in \mathcal{D}\,(n \in \mathbb{N})$ *disjunkt:* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$ □

Offensichtlich ist jede σ-Algebra auch ein Dynkin-System. Umgekehrt gilt

1.9 Satz *Ist $\mathcal{D}$ ein $\cap$-stabiles Dynkin-System über Ω, so ist $\mathcal{D}$ auch eine σ-Algebra.* □

1.10 Lemma und Definition *Sei $\Omega \neq \emptyset$ und $\mathcal{C} \subset 2^\Omega$.*

$$d(\mathcal{C}) := \bigcap_{\mathcal{D} \text{ Dynkin-System über } \Omega,\, \mathcal{C} \subset \mathcal{D}} \mathcal{D}$$

ist das kleinste Dynkin-System, das $\mathcal{C}$ umfasst, und heißt das **von $\mathcal{C}$ erzeugte Dynkin-System.** □

1.11 Satz *Ist $\mathcal{C} \subset 2^\Omega$ $\cap$-stabil, so gilt $d(\mathcal{C}) = \sigma(\mathcal{C})$.* □

Das folgende Beweisprinzip erweist sich als sehr nützlich, um nachzuweisen, dass eine Eigenschaft E für alle Mengen einer σ-Algebra $\mathcal{A}$ (bzw. eines Dynkin-Systems $\mathcal{D}$) gilt:

- Für ein geeignetes Mengensystem $\mathcal{F} \subset 2^\Omega$ betrachte

$$\mathcal{G} := \{F \in \mathcal{F} \mid F \text{ besitzt Eigenschaft } E\} \qquad \text{(“gute Mengen”)}$$

- Man zeige:
  - $\mathcal{G}$ ist eine σ-Algebra (bzw. ein Dynkin-System)
  - $\mathcal{C} \subset \mathcal{G}$ für ein System $\mathcal{C} \subset 2^\Omega$ mit $\sigma(\mathcal{C}) = \mathcal{A}$ (bzw. $d(\mathcal{C}) = \mathcal{D}$)

Dann folgt $\mathcal{A} \subset \mathcal{G}$ (bzw. $\mathcal{D} \subset \mathcal{G}$), also die Behauptung.

Konstruktion einer Maßfortsetzung in der Situation von Satz 1.6(ii)

1. Schritt: Definiere zu *allen* $A \subset \Omega$

$$\mu^*(A) := \inf\Big\{\sum_{n\in\mathbb{N}} \tilde{\mu}(C_n) \mid \bigcup_{n\in\mathbb{N}} C_n \supset A, C_n \in \mathcal{C}\ \forall n \in \mathbb{N}\Big\}$$

mit der üblichen Konvention $\inf \emptyset = \infty$.
Insbesondere gilt dann

$$\mu^*(C) = \tilde{\mu}(C) \quad \forall\, C \in \mathcal{C}.$$

Man kann zeigen, dass $\mu^* : 2^\Omega \to [0,\infty]$ ein **äußeres Maß** ist, d.h.

(i) $\mu^*(\emptyset) = 0$

(ii) $\forall\, A \subset B \subset \Omega : \quad \mu^*(A) \leq \mu^*(B)$

(iii) $\forall\, A_n \subset \Omega\ (n \in \mathbb{N}) : \quad \mu^*\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leq \sum_{n\in\mathbb{N}} \mu^*(A_n)$

2. Schritt: Definiere das System der so genannten **additiven Zerleger** von μ^*:

$$\mathcal{A}^* := \big\{A \subset \Omega \mid \mu^*(A \cap M) + \mu^*(A^c \cap M) = \mu^*(M)\ \forall\, M \subset \Omega\big\}.$$

Man kann zeigen:

- $\mathcal{A}^*$ ist ein $\cap$-stabiles Dynkin-System, also eine σ-Algebra.
- $\mathcal{C} \subset \mathcal{A}^*$

Daraus folgt aber $\mathcal{A} \subset \mathcal{A}^*$.

3. Schritt: Man kann zeigen, dass $\mu^*|_{\mathcal{A}^*}$ ein Maß auf $(\Omega, \mathcal{A}^*)$ ist. Damit ist aber insbesondere

$$\mu = \mu^*|_{\mathcal{A}} : \quad \begin{aligned} &\mathcal{A} \to [0,\infty] \\ &\mu(A) = \mu^*(A) \end{aligned}$$

ein Maß, das gemäß den Ergebnissen aus dem 1. Schritt auf $\mathcal{C}$ mit $\tilde{\mu}$ übereinstimmt, also eine Fortsetzung von $\tilde{\mu}$ zu einem Maß auf $(\Omega, \mathcal{A})$ darstellt. $\square$

1.12 Bemerkung Der oben skizzierte Beweis von Satz 1.6(ii) zeigt, dass man $\tilde{\mu}$ sogar zu einem Maß auf $\mathcal{A}^*$ fortsetzen kann und i.allg. ist $\mathcal{A}^*$ echt größer als $\mathcal{A} = \sigma(\mathcal{C})$. Insbesondere ist $\mathcal{A}^*$ **vollständig** (bzgl. $\mu^*|_{\mathcal{A}^*}$), d.h.

$$\forall\, N \in \mathcal{A}^* \text{ mit } \mu^*(N) = 0,\ A \subset N : \quad A \in \mathcal{A}^*.$$

Solche Untermengen A von Mengen N vom Maß 0 heißen **Nullmengen**. Vollständige σ-Algebren sind also solche, die alle Nullmengen enthalten.

Ist $\mathcal{A}$ nicht vollständig, so ist also $\mathcal{A}$ eine echte Untermenge von $\mathcal{A}^*$. Beispielsweise ist die Borel-σ-Algebra $\mathbb{B}$ *nicht* vollständig und ein Maß auf $(\mathbb{R}, \mathbb{B})$ lässt sich daher mit obiger Konstruktion auf eine größere σ-Algebra fortsetzen.

Man kann eine σ-Algebra $\mathcal{A}$ allerdings auch direkt vervollständigen (bzgl. eines Maßes μ auf $\mathcal{A}$): Das Mengensystem

$$\bar{\mathcal{A}}_\mu := \left\{ A \cup N \mid A \in \mathcal{A}, N \text{ Nullmenge bzgl. } \mu \right\}$$

ist eine σ-Algebra und

$$\bar{\mu}(A \cup N) := \mu(A) \quad \forall\, A \in \mathcal{A}, N \text{ Nullmenge bzgl. } \mu$$

definiert die einzige Fortsetzung von μ zu einem Maß auf $(\Omega, \bar{\mathcal{A}}_\mu)$.

Die Fortsetzung des Lebesgue-Maßes $\lambda\!\!\lambda$ auf $\bar{\mathcal{B}}_{\lambda\!\!\lambda}$ wird ebenfalls als Lebesgue-Maß bezeichnet. $\square$

2 Integrationstheorie

In diesem Abschnitt soll ein das Lebesgue-Integral und Erwartungswerte verallgemeinernder Integralbegriff eingeführt werden. Dabei betrachten wir Funktionen $f : \Omega \to \bar{\mathbb{R}} := [-\infty, \infty]$, d.h. die Funktionswerte $\pm\infty$ sind zugelassen. Die Borel-σ-Algebra wird dann auf $\bar{\mathbb{R}}$ erweitert zu der σ-Algebra

$$\bar{\mathbb{B}} := \sigma\big(\mathbb{B} \cup \{\{\infty\}\} \cup \{\{-\infty\}\}\big) = \big\{B, B \cup \{\infty\}, B \cup \{-\infty\}, B \cup \{\infty, -\infty\} \mid B \in \mathbb{B}\big\}$$

auf $\bar{\mathbb{R}}$.

2.1 (Erinnerung) Sind $(\Omega, \mathcal{A}), (\tilde{\Omega}, \tilde{\mathcal{A}})$ Messräume, so heißt eine Abbildung $f : \Omega \to \tilde{\Omega}$ $\mathcal{A}, \tilde{\mathcal{A}}$**-messbar**, falls $f^{-1}(\tilde{A}) \in \mathcal{A}$ für alle $\tilde{A} \in \tilde{\mathcal{A}}$. In dem Fall schreibt man kurz $f : (\Omega, \mathcal{A}) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$. Ist $\tilde{\mathcal{C}}$ ein Erzeugendensystem von $\tilde{\mathcal{A}}$, so ist f genau dann $\mathcal{A}, \tilde{\mathcal{A}}$-messbar, wenn $f^{-1}(\tilde{C}) \in \mathcal{A}$ für alle $\tilde{C} \in \tilde{\mathcal{C}}$.

Da $\bar{\mathbb{B}} = \sigma\big(\{[-\infty, x] \mid x \in \mathbb{R}\}\big) = \sigma\big(\{[-\infty, x) \mid x \in \mathbb{R}\}\big) = \sigma\big(\{(x, \infty] \mid x \in \mathbb{R}\}\big) = \sigma\big(\{[x, \infty] \mid x \in \mathbb{R}\}\big)$, ist insbesondere $f : (\Omega, \mathcal{A}) \to (\bar{R}, \bar{\mathbb{B}})$ äquivalent zu jeder der folgenden Bedingungen

- $\{f \le x\} \in \mathcal{A} \quad \forall x \in \mathbb{R}$
- $\{f < x\} \in \mathcal{A} \quad \forall x \in \mathbb{R}$
- $\{f > x\} \in \mathcal{A} \quad \forall x \in \mathbb{R}$
- $\{f \ge x\} \in \mathcal{A} \quad \forall x \in \mathbb{R}$

Daraus folgt leicht, dass jede monotone Funktion $f : \mathbb{R} \to \bar{\mathbb{R}}$ $\mathbb{B}, \bar{\mathbb{B}}$-messbar ist. Ähnlich folgt durch Übertragung der Ergebnisse aus Beispiel 1.5 auf $\bar{\mathbb{B}}$ auch, dass jede stetige Funktion $\mathbb{B}, \bar{\mathbb{B}}$-messbar ist. Da für $\mathbb{B}, \bar{\mathbb{B}}$-messbare Funktionen f, g, f_n auch $f \pm g$, $f \cdot g$, f/g, $\sup_{n\in\mathbb{N}} f_n$, $\limsup_{n\in\mathbb{N}} f_n$, und $\lim_{n\in\mathbb{N}} f_n$ alle $\mathbb{B}, \bar{\mathbb{B}}$-messbar sind, falls die Ausdrücke wohldefiniert sind, lassen sich daraus sehr vielfältige $\mathbb{B}, \bar{\mathbb{B}}$-messbare Funktionen konstruieren.

Im folgenden wird außerdem verwendet, dass für $\mathcal{A}, \bar{\mathbb{B}}$-messbare Funktionen f, g und $a \in \bar{\mathbb{R}}$ alle Mengen des Typs $\{f = a\}$, $\{f \le a\}$, $\{f < g\}$ usw. zu $\mathcal{A}$ gehören. □

Ziel ist es zunächst, für einen beliebigen Maßraum $(\Omega, \mathcal{A}, \mu)$ und eine möglichst umfassende Klasse von messbaren Funktionen $f : (\Omega, \mathcal{A}) \to (\bar{R}, \bar{B})$ ein Integral $\int f\, d\mu = \int f(\omega)\, \mu(d\omega)$ zu definieren, das eine Mittelung von f gemäß den durch μ gegebenen Gewichten beschreiben soll. Insbesondere soll dieses Integral für Indikatorfunktionen von Mengen gleich dem Maß der Menge sein, linear in f sein und gewisse Stetigkeitseigenschaften besitzen. Diese Anforderungen lassen die nachfolgende Definition sinnvoll erscheinen.

2.2 Definition *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.*

(i) Für eine Funktion $f : \Omega \to [0, \infty]$ der Form

$$f = \sum_{i=1}^{n} a_i 1_{A_i} \quad \textit{mit } n \in \mathbb{N}, a_i \in [0, \infty) \textit{ und disjunkten Mengen } A_i \in \mathcal{A}\ \forall 1 \le i \le n$$

definiere

$$\int f\,d\mu := \sum_{i=1}^{n} a_i \mu(A_i).$$

(ii) Zu einer Funktion $f : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$ *mit* $f(\omega) \geq 0$ *für alle* $\omega \in \Omega$ *definiere eine Folge von Funktionen* $f^{(n)}$ *durch*

$$\begin{aligned} f^{(n)}(\omega) &= \min\left([f(\omega)2^n] \cdot 2^{-n}, n\right) \\ &= \sum_{i=0}^{n2^n-1} i2^{-n} 1_{[i2^{-n},(i+1)2^{-n})}(f(\omega)) + n 1_{[n,\infty]}(f(\omega)) \end{aligned}$$

und

$$\begin{aligned} \int f\,d\mu &:= \sup_{n\in\mathbb{N}} \int f^{(n)}\,d\mu \\ &= \sup_{n\in\mathbb{N}} \sum_{i=0}^{n2^n-1} i2^{-n}\mu\{i2^{-n} \leq f < (i+1)2^{-n}\} + n\mu\{f \geq n\}. \end{aligned}$$

(iii) Zu einer Funktion $f : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$ *definiere* $f^+ = \max(f, 0)$ *und* $f^- = -\min(f, 0)$ *sowie*

$$\int f\,d\mu := \int f^+ d\mu - \int f^- d\mu,$$

falls $\int f^+ d\mu < \infty$ *oder* $\int f^- d\mu < \infty$. *In diesem Fall heißt* f ***μ*-integrierbar**. *Ist* $|\int f\,d\mu| < \infty$, *so heißt* f **endlich *μ*-integrierbar**.

$$\mathcal{L}_1(\mu) := \{f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B}) \mid f \text{ endlich } \mu\text{-integrierbar}\}$$

Speziell: Ist μ *ein Wahrscheinlichkeitsmaß, so heißt* $E_\mu(f) := \int f\,d\mu$ **Erwartungswert von f unter *μ*.**
Ist $\mu = \lambda\!\!\lambda$, *so heißt* $\int f\,d\lambda\!\!\lambda$ **Lebesgue-Integral von f.** □

2.3 Bemerkung (i) Eine $\mathbb{R}$-wertige Funktion f besitzt genau dann eine Darstellung $f = \sum_{i=1}^n a_i 1_{A_i}$ für geeignetes $n \in \mathbb{N}$, $a_i \in \mathbb{R}$ und $A_i \in \mathcal{A}$, wenn sie $\mathcal{A}, \mathbb{B}$-messbar ist und nur endlich viele Werte annimmt. Werden die unterschiedlichen Werte mit $b_1, \ldots, b_k$ bezeichnet, so gilt dann nämlich $f = \sum_{j=1}^k b_j 1_{\{f=b_j\}}$. Ist zusätzlich wie in Definition 2.2(i) $f \geq 0$ gefordert, so heißt f **Elementarfunktion**.

Die in Definition 2.2(i) angegebene Definition des Integrals einer Elementarfunktion hängt nicht von der Wahl der Darstellung ab. Gilt also $f = \sum_{i=1}^n a_i 1_{A_i} = \sum_{j=1}^k b_j 1_{B_j}$ für geeignete $a_i, b_j \geq 0$ und disjunkte Mengen $A_i \in \mathcal{A}$ bzw. $B_j \in \mathcal{A}$, so auch $\sum_{i=1}^n a_i\mu(A_i) = \sum_{j=1}^k b_j\mu(B_j)$.

(ii) Die in Definition 2.2(ii) definierte Folge von Elementarfunktion $f^{(n)}$ konvergiert von unten monoton gegen f. Wir werden sehen, dass sich die Monotonie auch auf die Integrale von $f^{(n)}$ überträgt, so dass man in der Definition "sup" auch durch "lim" ersetzen kann.

Die in Definition 2.2(ii) gegebene Definition von $\int f\,d\mu$ stimmt für Elementarfunktionen mit der aus Definition 2.2(i) überein.

(iii) Soll in der Situation von Definition 2.2 die Abhängigkeit der Funktion f vom Argument explizit gemacht werden, so schreibt man statt $\int f\, d\mu$ auch $\int f(\omega)\, \mu(d\omega)$ (oder manchmal auch $\int f(\omega)\, d\mu(\omega)$).

In manchen Quellen werden auch nur die *endlich* μ-integrierbaren Funktionen als μ-integrierbar bezeichnet, während die in unserem Sinne μ-integrierbaren Funktionen (mit eventuell unendlichem Integral) als *quasi-μ-integrierbar* bezeichnet werden. □

2.4 Beispiel Für Maße mit endlichem Träger, also Maße der Form $\mu = \sum_{j=1}^{k} \alpha_j \varepsilon_{\omega_j}$ mit $\alpha_j > 0$ und $\omega_j \in \Omega$ für alle $1 \leq j \leq k$, (d.h. $\mu(A) = \sum_{j=1}^{k} \alpha_j 1_A(\omega_j)$) und $f : (\Omega, \mathcal{A}) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$ gilt

$$\int f\, d\mu = \sum_{j=1}^{k} \alpha_j f(\omega_j),$$

falls die rechte Seite wohldefiniert ist. Andernfalls ist f nicht μ-integrierbar.

Allgemeiner gilt für diskrete Maße $\mu = \sum_{j=1}^{\infty} \alpha_j \varepsilon_{\omega_j}$ mit $\alpha_j \geq 0$ und $\omega_j \in \Omega$ für alle $j \in \mathbb{N}$

$$\int f\, d\mu = \sum_{j=1}^{\infty} \alpha_j f(\omega_j),$$

falls $\sum_{j=1}^{\infty} \alpha_j f^+(\omega_j) < \infty$ oder $\sum_{j=1}^{\infty} \alpha_j f^-(\omega_j) < \infty$. □

2.5 Lemma *Sind f, g Elementarfunktionen und $\alpha \in \mathbb{R}$, so gilt*

(i) $\int \alpha f\, d\mu = \alpha \int f\, d\mu$

(ii) $\int f + g\, d\mu = \int f\, d\mu + \int g\, d\mu$

(iii) $f \leq g \quad \Longrightarrow \quad \int f\, d\mu \leq \int g\, d\mu$ □

Das folgende Lemma zeigt, dass in Definition 2.2(ii) die genaue Form der Elementarfunktionen $f^{(n)}$, die monoton von unten gegen f konvergieren, nicht wesentlich ist, d.h. jede andere solche Folge führt zum gleichen Integralbegriff.

2.6 Lemma *Sind f_n, g_n Elementarfunktionen mit $f_n, g_n \uparrow f$, so gilt*

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

□

2.7 Satz *Seien $f, g : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$ μ-integrierbar. Dann gilt*

(i) $\forall \alpha \in \mathbb{R}: \quad \alpha f$ *ist μ-integrierbar mit* $\int \alpha f\, d\mu = \alpha \int f\, d\mu.$

(ii) *Falls $f+g$ und $\int f d\mu + \int g d\mu$ wohldefiniert sind, so ist $f+g$ μ-integrierbar mit*

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

(iii) *Gilt $f \leq g$, so folgt $\int f\,d\mu \leq \int g d\mu$.*

(iv) *Ist $f \in \mathcal{L}_1(\mu)$ und gilt $|h| \leq f$ für $h : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$, so gilt auch $h, |h| \in \mathcal{L}_1(\mu)$ und*

$$\left|\int h\,d\mu\right| \leq \int |h|\,d\mu \leq \int f\,d\mu.$$

□

2.9 Satz (Hölder'sche Ungleichung) *Seien $f, g : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathbb{B})$ und $1 < p < \infty$. Dann gilt mit $q := p/(p-1)$ (d.h. $1/p + 1/q = 1$)*

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

(mit der Konvention $0 \cdot \infty = 0$). □

2.10 Korollar *Ist μ ein endliches Maß, so folgt aus $\int |f|^p\,d\mu < \infty$ auch $\int |f|^{\tilde{p}}\,d\mu < \infty$ für alle $1 \leq \tilde{p} \leq p$.* □

Ohne die Voraussetzung $\mu(\Omega) < \infty$ ist diese Folgerung i.allg. falsch.

Im folgenden bezeichnet für $p \geq 1$

$$\mathcal{L}_p(\mu) := \Big\{ f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B}) \;\Big|\; \int |f|^p\,d\mu < \infty \Big\}$$

den Raum der p-fach endlich μ-integrierbaren Funktionen. Für $f \in \mathcal{L}_p(\mu)$ definiert man

$$\|f\|_p := \Big(\int |f|^p\,d\mu\Big)^{1/p}.$$

Im Fall $\mu(\Omega) < \infty$ gilt also $\mathcal{L}_p(\mu) \subset \mathcal{L}_{\tilde{p}}(\mu)$ für alle $1 \leq \tilde{p} \leq p < \infty$. Allgemein gilt für $f \in \mathcal{L}_p(\mu)$ und $g \in \mathcal{L}_q(\mu)$ mit $p, q > 1$ und $1/p + 1/q = 1$ gemäß der Hölder'schen Ungleichung gerade $fg \in \mathcal{L}_1(\mu)$ und $\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q$.

2.11 Satz (Minkowski-Ungleichung) *Für $f, g \in \mathcal{L}_p(\mu)$ mit $p \geq 1$ gilt $f+g \in \mathcal{L}_p(\mu)$ und*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

□

2.12 Bemerkung Wegen der Sätze 2.7 und 2.11 ist $\mathcal{L}_p(\mu)$ ein linearer Vektorraum und es gilt $\|\alpha f\|_p = |\alpha| \|f\|_p$ für $f \in \mathcal{L}_p(\mu)$ und $\alpha \in \mathbb{R}$. Da gemäß Satz 2.11 $\|\cdot\|_p$ auch die Dreiecksungleichung erfüllt, ist dies eine *Halbnorm*, d.h. $\|\cdot\|_p$ besitzt alle definierenden Eigenschaften einer Norm außer der Implikation $\|f\| = 0 \Rightarrow f = 0$.

Es gilt allerdings $\|f\|_p = 0 \Rightarrow f = 0$ μ-f.ü. (d.h. $\mu\{f \neq 0\} = 0$). Identifiziert man nun alle Funktionen, die μ-f.ü. gleich sind, so kann daher $\|\cdot\|_p$ als Norm aufgefasst werden. Formal definiert man dazu die Äquivalenzrelation

$$f \sim g \ :\Longleftrightarrow\ f = g \ \mu\text{-f.ü.}$$

auf $\mathcal{L}_p(\mu)$ und betrachtet den zugehörigen Quotientenraum der Äquivalenzklassen $[f] := \{g \in \mathcal{L}_p(\mu) \mid f \sim g\}$, d.h.

$$L_p(\mu) := \big\{[f] \mid f \in \mathcal{L}_p(\mu)\big\}$$

versehen mit der Norm

$$\|[f]\|_p := \|f\|_p.$$

Meist schreibt man jedoch statt $[f]$ doch wieder f, und wir wollen diesem (praktischen, aber etwas unsauberen) Usus folgen.

$(L_p(\mu), \|\cdot\|_p)$ ist dann ein normierter Vektorraum und er ist sogar vollständig im topologischen Sinne (also ein Banach-Raum), d.h. jede Cauchy-Folge konvergiert:
Ist $f_n \in L_p(\mu)$, $n \in \mathbb{N}$, eine Folge, so dass

$$\forall\, \varepsilon > 0\ \exists n_\varepsilon \in \mathbb{N}\ \forall\, n, m \geq n_\varepsilon:\ \|f_n - f_m\|_p \leq \varepsilon,$$

dann existiert ein $f \in L_p(\mu)$ mit $\|f_n - f\|_p \to 0$.

Im Fall $p = 2$ besitzt dieser Vektorraum eine besonders schöne Struktur, da die Norm von dem Skalarprodukt (also der positiv definiten Bilinearform)

$$\langle f, g\rangle := \int f g \, d\mu, \quad f, g \in L_2(\mu),$$

erzeugt wird, d.h. $\|f\|_2 = \langle f, f\rangle$ für alle $f \in L_2(\mu)$. Damit ist $\big(L_2(\mu), \langle\cdot,\cdot\rangle\big)$ ein sog. *Hilbert-Raum*. In solchen gelten viele Gesetze, die aus der linearen Algebra für den $\mathbb{R}^n$ versehen mit dem üblichen Skalarprodukt bekannt sind. (Einige Unterschiede ergeben sich aber insbesondere daraus, dass $L_2(\mu)$ ein unendlichdimensionaler Vektorraum ist.)

Insbesondere gibt es zu jedem abgeschlossenen linearen Unterraum $M \subset L_2(\mu)$ und jedem $f \in L_2(\mu)$ genau ein $g \in M$, so dass

$$\|f - g\|_2 = \inf_{h \in M} \|f - h\|_2,$$

d.h. g ist das Element von M, das am nächsten an f liegt (oder anders ausgedrückt: g ist die beste Approximation von f durch eine Funktion in M). Die Funktion g ist eindeutig bestimmt durch die Bedingung

$$\langle f - g, h\rangle = 0 \quad \forall\, h \in M,$$

d.h. $f - g$ steht senkrecht auf dem Unterraum M, i.Z. $f - g \perp M$. g heißt daher auch die *Orthogonalprojektion* von f auf M. Es gilt dann das Gesetz von Pythagoras:

$$\|f - h\|_2^2 = \|f - g\|_2^2 + \|g - h\|_2^2 \quad \forall h \in M.$$

Mit Hilfe der Orthogonalprojektion lässt sich nun der folgende *Riesz'sche Darstellungssatz* beweisen:

Ist $\varphi : L_2(\mu) \to \mathbb{R}$ ein lineare, stetige Abbildung, dann gibt es ein $f \in L_2(\mu)$, so dass $\varphi(g) = \langle f, g \rangle = \int fg\, d\mu$ für alle $g \in L_2(\mu)$ gilt.

Die durch $g \mapsto \int fg\, d\mu$ definierten Funktionale sind also die einzigen stetigen, linearen Abbildungen von $L_2(\mu)$ nach $\mathbb{R}$. □

2.13 Definition *Man sagt "Eigenschaft E gilt μ-fast überall (kurz: μ-f.ü.)" oder "E gilt für μ-fast alle (kurz: μ-f.a.) $\omega \in \Omega$", falls $\{\omega \in \Omega \mid E$ gilt* nicht *für $\omega\}$ eine μ-Nullmenge ist, also eine Menge $N \in \mathcal{A}$ mit $\mu(N) = 0$ existiert, so dass die Eigenschaft E für alle $\omega \in N^c$ gilt.* □

Seien $f, f_n : (\Omega, \mathcal{A}, P) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$. Im allgemeinen folgt aus $\lim_{n\to\infty} f_n = f$ μ-f.ü. *nicht* die Konvergenz der Integrale $\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu$. Im folgenden sollen hinreichende Bedingungen angegeben werden, die diese Konvergenz sicherstellen.

2.15 Satz (von der monotonen Konvergenz)

Konvergieren die messbaren Funktionen $f_n \geq 0$, $n \in \mathbb{N}$, μ-f.ü. monoton von unten gegen f, so gilt $\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu$. □

2.16 Lemma (Lemma von Fatou) *Sind f_n, $n \in \mathbb{N}$, $[0, \infty]$-wertige messbare Funktionen, so gilt stets*

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

□

Beweis. Die Folge $g_n := \inf_{k\geq n} f_k \geq 0$ konvergiert für $n \to \infty$ von unten monoton gegen $\liminf_{n\to\infty} f_n$. Satz 2.15 liefert daher wegen $g_n \leq f_k$ für alle $k \geq n$

$$\int \liminf_{n\to\infty} f_n\, d\mu = \int \lim_{n\to\infty} g_n\, d\mu = \lim_{n\to\infty} \int g_n\, d\mu \leq \lim_{n\to\infty} \inf_{k\geq n} \int f_k\, d\mu = \liminf_{n\to\infty} \int f_n\, d\mu.$$

□

2.17 Satz (von Lebesgue, von der majorisierten/dominierten Konvergenz)

Gilt $\lim_{n\to\infty} f_n = f$ μ-f.ü. und $|f_n| \leq g$ μ-f.ü. für alle $n \in \mathbb{N}$ und eine endlich integrierbare Funktion $g \in \mathcal{L}_1(\mu)$, so gilt

$$\lim_{n\to\infty} \int |f_n - f|\, d\mu = 0 \tag{2.1}$$

(d.h. $f_n \to f$ in $(L_1, \|\cdot\|_1)$) und insbesondere

$$\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu. \tag{2.2}$$

□

Beweis. Wegen $|\int f_n\, d\mu - \int f\, d\mu| \leq \int |f_n - f|\, d\mu$ folgt (2.2) sofort aus (2.1).

Definiere

$$A := \bigcap_{n\in\mathbb{N}} \{|f_n| \le g\} \cap \{\lim_{n\to\infty} f_n = f\}.$$

Nach Voraussetzung gilt $\mu(A^c) = 0$ und wir können o.E. $A = \Omega$ annehmen. (Sonst ersetze in den nachfolgenden Beweisschritten f_n, f und g durch $\tilde{f_n} = f_n 1_A$, $\tilde{f} = f 1_A$ bzw. $\tilde{g} = g 1_A$.)

Definiere $h_n := 2g - |f_n - f| \ge 2g - (|f_n| + |f|) \ge 0$. Das Lemma 2.16 von Fatou liefert wegen $\lim_{n\to\infty} h_n = 2g$

$$\int 2g\, d\mu \le \liminf_{n\to\infty} \int h_n\, d\mu = \int 2g\, d\mu - \limsup_{n\to\infty} \int |f_n - f|\, d\mu,$$

also auch (2.1). □

2.18 Satz und Definition *Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : (\Omega, \mathcal{A}) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$, so wird durch*

$$\mu^f(\tilde{A}) = \mu\{f \in \tilde{A}\}, \quad \tilde{A} \in \tilde{\mathcal{A}},$$

ein Maß μ^f auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ definiert, das so genannte **Bildmaß von μ unter** *f.* □

2.19 Satz (Transformationssatz) *In der Situation von Satz und Definition 2.18 ist $g : (\tilde{\Omega}, \tilde{\mathcal{A}}) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$ genau dann μ^f-integrierbar, wenn $g \circ f$ μ-integrierbar ist. In dem Fall gilt*

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

□

3 Radon-Nikodym-Dichten

Aus der Mathematischen Stochastik sind Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathbb{B})$ mit Lebesgue-Dichte f bekannt, die definiert sind durch

$$P(A) = \int_A f \, d\lambda\!\!\lambda := \int f 1_A \, d\lambda\!\!\lambda \quad \forall A \in \mathbb{B}.$$

Analog kann man Maße mit Dichten bzgl. anderer integrierender Maße definieren.

3.1 Satz und Definition *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$ mit $f \geq 0$. Dann wird durch*

$$\nu(A) := \int_A f \, d\mu := \int f 1_A \, d\mu, \quad A \in \mathcal{A},$$

*ein Maß auf $(\Omega, \mathcal{A})$ definiert. Man sagt dann, dass ν die **μ-Dichte** f besitzt; i.Z. $f \in \frac{d\nu}{d\mu}$. f wird auch als **(Radon-Nikodym-)Dichte von ν bzgl. μ** bezeichnet.* □

3.2 Definition *Ein Maß μ auf $(\Omega, \mathcal{A})$ heißt σ-**endlich**, wenn es eine Folge von Mengen $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, gibt, so dass $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$ und $A_n \uparrow \Omega$.* □

3.3 Bemerkung μ-Dichten sind nur μ-f.ü. bestimmt. Ist also f eine μ-Dichte von ν und $f = g$ μ-f.ü., so ist auch g eine μ-Dichte von ν.

Umgekehrt gilt für jedes σ-endliche Maß μ: Sind f und g beides μ-Dichten von ν, so gilt $f = g$ μ-f.ü. (Dies gilt für Maße μ, die nicht σ-endlich sind, i.allg. nicht.) □

3.4 Beispiel Ist μ das Zählmaß auf $(\Omega, \mathcal{A})$, d.h. $\mu(A) = |A|$ für alle $A \in \mathcal{A}$, ν ein diskretes Maß auf $(\Omega, \mathcal{A})$ (d.h. $\nu(\Omega_0^c) = 0$ für eine höchstens abzählbare Menge Ω_0) und ist $\{\omega\} \in \mathcal{A}$ für alle $\omega \in \Omega$, so definiert $f(\omega) := \nu\{\omega\}$ für alle $\omega \in \Omega$ eine μ-Dichte von ν. Damit ist also die aus der Mathematischen Stochastik bekannte Zähldichte gerade gleich der Radon-Nikodym-Dichte bzgl. des Zählmaßes. □

3.5 Satz *Hat das Maß ν auf $(\Omega, \mathcal{A})$ eine μ-Dichte f, so ist eine Funktion $h : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$ genau dann ν-integrierbar, wenn $h \cdot f$ μ-integrierbar ist. In dem Fall gilt*

$$\int h \, d\nu = \int h \cdot f \, d\mu.$$

(Dabei setzen wir $\infty \cdot 0 = 0$.) □

3.6 Korollar *Seien μ, ν, τ Maße auf $(\Omega, \mathcal{A})$, wobei ν eine μ-Dichte f und τ eine ν-Dichte g besitzt. Dann ist $f \cdot g$ eine μ-Dichte von τ.* □

3.7 Satz und Definition *Besitzt ein Maß ν auf $(\Omega, \mathcal{A})$ eine μ-Dichte, so gilt*

$$\forall A \in \mathcal{A} : \quad \mu(A) = 0 \implies \nu(A) = 0 \tag{3.1}$$

*Ist Bedingung (3.1) erfüllt, so heißt ν **absolutstetig** bzgl. μ; i.Z. $\nu \ll \mu$.* □

3.8 Satz (von Radon-Nikodym) *Seien μ, ν Maße auf $(\Omega, \mathcal{A})$ und μ sei σ-endlich. Dann sind äquivalent:*

(i) ν besitzt eine μ-Dichte.

(ii) ν ist absolutstetig bzgl. μ. □

Das folgende Lemma liefert ein alternatives Kriterium für Absolutstetigkeit.

3.9 Lemma *Sei μ ein σ-endliches Maß auf $(\Omega, \mathcal{A})$.*

(i) Für alle endlichen Maße ν auf $(\Omega, \mathcal{A})$ gilt

$$\nu \ll \mu \iff \Big(\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall A \in \mathcal{A}: \quad \mu(A) < \delta \Rightarrow \nu(A) < \varepsilon\Big).$$

(ii) Insbesondere gilt für alle nicht-negativen Funktionen $f \in \mathcal{L}_1(\Omega, , \mathcal{A}, \mu)$:

$$\forall \varepsilon > 0 \; \exists \delta > 0 : \qquad \sup_{A \in \mathcal{A}, \, \mu(A) < \delta} \int_A f \, d\mu < \varepsilon.$$ □

Beliebige σ-endliche Maße lassen sich zerlegen in einen Anteil mit μ-Dichte und einem dazu "orthogonalen" Anteil.

3.10 Definition *Seien μ, ν Maße auf $(\Omega, \mathcal{A})$. Dann heißt μ **singulär** zu ν (i.Z. $\mu \perp \nu$), falls es ein $N \in \mathcal{A}$ gibt mit $\mu(N) = 0$ und $\nu(N^c) = 0$.* □

3.11 Satz *Sind μ und ν σ-endliche Maße auf $(\Omega, \mathcal{A})$, so existieren eindeutig bestimmte Maße ν_a und ν_s, so dass $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ und $\nu_s \perp \mu$.* □

3.12 Satz und Definition *Sei $F : \mathbb{R} \to \mathbb{R}$ eine maßdefinierende Funktion (s. Beispiel 1.7) und bezeichne μ_F das zugehörige Radon-Maß. F heißt absolutstetig bzw. singulär, wenn μ_F absolutstetig bzw. singulär bzgl. des Lebesgue-Maßes $\lambda\!\!\lambda$ ist.*

Die maßdefinierende Funktion F ist $\lambda\!\!\lambda$-f.ü. differenzierbar mit Ableitung F' und es gilt $\int_{(a,b]} F' \, d\lambda\!\!\lambda \leq F(b) - F(a)$ für alle $-\infty < a < b < \infty$. Die Funktion

$$F_a(x) := \begin{cases} F(0) + \int_{(0,x]} F' \, d\lambda\!\!\lambda, & x \geq 0, \\ F(0) - \int_{(x,0]} F' \, d\lambda\!\!\lambda, & x < 0, \end{cases}$$

ist eine maßdefinierende Funktion des bzgl. $\lambda\!\!\lambda$ absolutstetigen Anteils von μ_F und $F - F_a$ ist eine maßdefinierende Funktion des zu $\lambda\!\!\lambda$ singulären Anteils von μ_F. Daher ist F genau dann singulär, wenn $F' = 0$ $\lambda\!\!\lambda$-f.ü.

Der singuläre Anteil $F - F_a$ kann weiter zerlegt werden in einen diskreten Anteil F_d, der eine reine Sprungfunktion ist und durch

$$F_d(x) := \begin{cases} \sum_{0 < t \leq x} \Delta F(t), & x \geq 0, \\ \sum_{x < t \leq 0} \Delta F(t), & x < 0, \end{cases}$$

definiert wird, und einen singulären, aber stetigen Anteil $F_s := F - F_a - F_d$. (Hierbei bezeichnet $\Delta F(x) := F(x) - F(x-)$ die Sprunghöhe von F in x.)

4 Produkträume und Übergangsmaße

4.1 Definition *Eine Abbildung $K : \Omega_1 \times \mathcal{A}_2 \to \overline{\mathbb{R}}$ heißt **Übergangsmaß** oder **Übergangskern** von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, falls die beiden folgenden Bedingungen erfüllt sind:*

(i) Für alle $\omega_1 \in \Omega_1$ ist $K(\omega_1, \cdot) : \mathcal{A}_2 \to \overline{\mathbb{R}}$ ein Maß auf $(\Omega_2, \mathcal{A}_2)$.

(ii) Für alle $A_2 \in \mathcal{A}_2$ ist $K(\cdot, A_2) : \Omega_1 \to \overline{\mathbb{R}}$ $\mathcal{A}_1, \overline{\mathbb{B}}$-messbar.

*K heißt **σ-endlich**, falls es $C_n \in \mathcal{A}_2, n \in \mathbb{N}$, gibt mit*

$$C_n \uparrow \Omega_2 \quad \text{und} \quad \sup_{\omega_1 \in \Omega_1} K(\omega_1, C_n) < \infty.$$

*Ist $K(\omega_1, \cdot)$ für alle $\omega_1 \in \Omega_1$ ein Wahrscheinlichkeitsmaß, so heißt K **Markov-Kern** oder **Übergangswahrscheinlichkeitsmaß**.* □

4.2 Bemerkung Ist die Bedingung 4.1(ii) für alle Mengen A_2 aus einem $\cap$-stabilen Erzeugendensystem $\mathcal{C}_2$ von $\mathcal{A}_2$ erfüllt und gibt es Mengen $C_n \in \mathcal{C}_2$ mit $C_n \uparrow \Omega_2$, so dass in Satz und Definition 4.1(i) $K(\omega_1, \cdot)$ ein Maß ist mit $K(\omega_1, C_n) < \infty$ für alle $\omega_1 \in \Omega_1$, so ist K ein Übergangsmaß. □

4.4 Satz und Definition *(i) Für $A \subset \Omega_1 \times \Omega_2$ und $\omega_1 \in \Omega_1$ wird der **ω_1-Schnitt** von A definiert als $A_{\omega_1} := \{\omega_2 \in \Omega_2 \mid (\omega_1, \omega_2) \in A\}$.*

(ii) Ist μ_1 ein σ-endliches Maß auf $(\Omega_1, \mathcal{A}_1)$ und K ein σ-endliches Übergangsmaß von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, so definiert

$$\begin{aligned}(\mu_1 \otimes K)(A) &:= \int_{\Omega_1} K(\omega_1, A_{\omega_1})\, \mu_1(d\omega_1) \\ &= \int_{\Omega_1} \int_{\Omega_2} 1_A(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu_1(d\omega_1), \qquad A \in \mathcal{A}_1 \otimes \mathcal{A}_2,\end{aligned}$$

*ein σ-endliches Maß auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, die so genannte **Koppelung** von μ_1 und K.*
Insbesondere gilt für $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$

$$(\mu_1 \otimes K)(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2)\, \mu_1(d\omega_1).$$

□

4.5 Lemma *In der Situation von Satz und Definition 4.4 ist für alle $\omega_1 \in \Omega_1$ und alle $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ die Abbildung $1_A(\omega_1, \cdot) = 1_{A_{\omega_1}}$ $\mathcal{A}_2, \overline{\mathbb{B}}$-messbar und die Abbildung $g_A : \Omega_1 \to \overline{\mathbb{R}}$*

$$g_A(\omega_1) := \int_{\Omega_2} 1_A(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) = K(\omega_1, A_{\omega_1})$$

ist $\mathcal{A}_1, \overline{\mathbb{B}}$-messbar. □

4.6 Satz (von Fubini für Übergangsmaße) *In der Situation von Satz und Definition 4.4 sei*

$$f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$$

Dann gilt:

(i) Ist $f \geq 0$, so ist die Abbildung $g : \Omega_1 \to \overline{\mathbb{R}}$, $g(\omega_1) = \int f(\omega_1, \omega_2) K(\omega_1, d\omega_2)$ wohldefiniert und $\mathcal{A}_1, \overline{\mathbb{B}}$-messbar und es gilt

$$\int f \, d\mu_1 \otimes K = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \, K(\omega_1, d\omega_2) \, \mu_1(d\omega_1) \tag{4.1}$$

(ii) Ist f $(\mu_1 \otimes K)$-integrierbar, so ist $g(\omega_1)$ für μ_1-f.a. $\omega_1 \in \Omega_1$ wohldefiniert und es gilt (4.1) *(wobei das innere Integral $g(\omega_1)$ für die ω_1 gleich* 0 *gesetzt wird, für die es nicht wohldefiniert ist).* □

4.7 Bemerkung Man beachte, dass in der Situation von Satz 4.6(ii)

$$\left| \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \, K(\omega_1, d\omega_2) \, \mu_1(d\omega_1) \right| < \infty$$

nicht sicher stellt, dass f tatsächlich $\mu_1 \otimes K$-integrierbar ist. □

Einen wichtigen Spezialfall erhält man, wenn die Maße $K(\omega_1, \cdot)$ für alle $\omega_1 \in \Omega_1$ identisch sind.

4.9 Korollar und Definition *(i) Seien μ_i σ-endliche Maße auf $(\Omega_i, \mathcal{A}_i)$. Dann gibt es genau ein Maß $\mu =: \mu_1 \otimes \mu_2$ auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, so dass*

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \qquad \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \tag{1}$$

*(mit $0 \cdot \infty = 0$). $\mu_1 \otimes \mu_2$ heißt **Produkt** der Maße μ_1 und μ_2.*

(ii) (Satz von Fubini-Tonelli)
Für jede $\mathcal{A}_1 \otimes \mathcal{A}_2, \overline{\mathbb{B}}$-messbare Funktion f gilt

$$\begin{aligned} \int f d\mu_1 \otimes \mu_2 &= \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2) \, \mu_1(d\omega_1) \, \mu_2(d\omega_2) \\ &= \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \, \mu_2(d\omega_2) \, \mu_1(d\omega_1), \end{aligned}$$

falls ($f \geq 0$ oder) f $\mu_1 \otimes \mu_2$-integrierbar ist.

(iii) Insbesondere gilt für alle $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$

$$\mu_1 \otimes \mu_2(A) = \int_{\Omega_1} \mu_2(A_{\omega_1}) \, \mu_1(d\omega_1) = \int_{\Omega_2} \mu_1(A_{\omega_2}) \, \mu_2(d\omega_2)$$

. □

4.10 Bemerkung (i) Zwei Zufallsvariablen X_1, X_2 auf $(\Omega, \mathcal{A}, P)$ heißen genau dann stochastisch unabhängig, wenn $P^{(X_1,X_2)} = P^{X_1} \otimes P^{X_2}$ (s. Mathematische Stochastik, Definition 4.5). Produkte von Wahrscheinlichkeitsmaßen beschreiben also die *unabhängige* Hintereinanderausführung von Zufallsexperimenten.

(ii) Im Fall $\mu_1 = \mu_2 = \lambda\!\!\lambda$ ist $\mu_1 \otimes \mu_2 = \lambda\!\!\lambda^2$ das Flächenmaß in $\mathbb{R}^2$ und 4.9(iii) ist das so genannte Prinzip des Cavalieri, gemäß dem man einen Flächeninhalt berechnen kann, indem man die Längen aller Schnitte parallel zu einer Achse aufintegriert:

$$\lambda\!\!\lambda^2(A) = \int_{\mathbb{R}} \lambda\!\!\lambda(A_{\omega_1})\lambda\!\!\lambda(d\omega_1)$$

$\square$

4.12 Satz und Definition *Seien μ_Θ ein σ-endliches Maß auf einem Messraum $(\Theta, \mathcal{S})$ und $\mu_\vartheta, \vartheta \in \Theta$, Maße auf einem Messraum $(\Omega, \mathcal{A})$ so, dass $K(\vartheta, A) := \mu_\vartheta(A)$ ein σ-endliches Übergangsmaß von $(\Theta, \mathcal{S})$ nach $(\Omega, \mathcal{A})$ definiert. Dann definiert*

$$\nu(A) = \int_\Theta \mu_\vartheta(A)\,\mu_\Theta(d\vartheta), \qquad A \in \mathcal{A},$$

ein Maß auf $(\Omega, \mathcal{A})$, die **Mischung der μ_ϑ, $\vartheta \in \Theta$, mit Mischungsmaß** μ_Θ *oder kürzer die* **Mischung der μ_ϑ unter** μ_Θ. *Bezeichnet $pr_2 : \Theta \times \Omega \to \Omega$, $(\vartheta, \omega) \mapsto \omega$ die Projektion auf die 2. Koordinate, so ist ν gerade das Bildmaß der Koppelung $\mu_\Theta \otimes K$ unter pr_2: $\nu = (\mu_\Theta \otimes K)^{pr_2}$.* $\square$

Induktiv kann man nun aus einem σ-endlichen Maß μ_1 auf $(\Omega_1, \mathcal{A}_1)$ und Übergangsmaßen K_l von $\left(\times_{i=1}^{l-1} \Omega_i, \bigotimes_{i=1}^{l-1} \mathcal{A}_i\right)$ nach $(\Omega_l, \mathcal{A}_l)$, $2 \le l \le n$, ein Maß

$$((((\mu_1 \otimes K_2) \otimes K_3) \otimes K_4) \otimes \cdots) \otimes K_n =: \mu_1 \otimes \bigotimes_{l=2}^n K_l$$

konstruieren, bzw. speziell aus Maßen μ_i auf $(\Omega_i, \mathcal{A}_i), 1 \le i \le n$, ein Produktmaß

$$\begin{aligned} \bigotimes_{i=1}^n \mu_i &:= ((\mu_1 \otimes \mu_2) \otimes \mu_3) \cdots \otimes \mu_n \\ &= \mu_1 \otimes (\mu_2 \otimes (\cdots \otimes (\mu_{n-1} \otimes \mu_n))), \end{aligned}$$

das eindeutig charakterisiert ist durch die Bedingung

$$\bigotimes_{i=1}^n \mu_i \Big(\mathop{\times}_{i=1}^n A_i \Big) = \prod_{i=1}^n \mu_i(A_i) \qquad \forall A_i \in \mathcal{A}_i \ (1 \le i \le n).$$

Wir wollen abschließend noch kurz auf das Produkt unendlich vieler Messräume und darauf definierte Maße eingehen, wie sie z.B. zur Beschreibung stochastischer Prozesse benötigt werden.

4.14 Definition *Sei T eine beliebige Indexmenge und $(\Omega_t, \mathcal{A}_t)$ Messräume für alle $t \in T$. Dann wird die Produkt-σ-Algebra $\bigotimes_{t\in T} \mathcal{A}_t$ auf $\times_{t\in T} \Omega_t$ definiert als*

$$\bigotimes_{t\in T} \mathcal{A}_t = \sigma\left(\{pr_t^{-1}(A_t) | A_t \in \mathcal{A}_t, t \in T\}\right)$$

mit $pr_t : \times_{s\in T} \Omega_s \to \Omega_t,\ (\omega_s)_{s\in T} \mapsto \omega_t$. □

Im Fall $T = \mathbb{N}$ kann man die obige Konstruktion für das Produkt von n Messräumen für *Wahrscheinlichkeitsmaße* übertragen.

4.15 Satz und Definition (von Ionescu-Tulcea)

(i) Sei P_1 ein Wahrscheinlichkeitsmaß auf $(\Omega_1, \mathcal{A}_1)$ und für alle $l \geq 2$ sei K_l ein Markov-Kern von $\left(\times_{i=1}^{l-1} \Omega_i, \bigotimes_{i=1}^{l-1} \mathcal{A}_i\right)$ nach $(\Omega_l, \mathcal{A}_l)$. Dann gibt es genau ein Wahrscheinlichkeitsmaß Q auf $\left(\times_{n\in\mathbb{N}} \Omega_n, \bigotimes_{n\in\mathbb{N}} \mathcal{A}_n\right)$ mit

$$Q\Big(A_{1,n} \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big) = \Big(P_1 \otimes \bigotimes_{l=2}^{n} K_l\Big)(A_{1,n}) \qquad \forall A_{1,n} \in \bigotimes_{i=1}^{n} \mathcal{A}_i, n \in \mathbb{N}.$$

(ii) Ist speziell $K_l((\omega_1, \ldots, \omega_{l-1}), \cdot) = P_l$ für ein Wahrscheinlichkeitsmaß P_l für alle $l \geq 2$ und alle $(\omega_1, \ldots, \omega_{l-1}) \in \times_{i=1}^{l-1} \Omega_i$, so gilt

$$Q\left(\bigtimes_{i=1}^{n} A_i \times \bigtimes_{i=n+1}^{\infty} \Omega_i\right) = \prod_{i=1}^{n} P_i(A_i) \qquad \forall A_i \in \mathcal{A}_i, n \in \mathbb{N}$$

*und Q heißt **Produkt** der Wahrscheinlichkeitsmaße $P_i, i \in \mathbb{N}$; i.Z. $Q = \bigotimes_{i\in\mathbb{N}} P_i$.* □

Das Wahrscheinlichkeitsmaß auf dem Folgenraum $\times_{i\in\mathbb{N}} \Omega_i$ ist insbesondere durch alle endlich-dimensionalen Randverteilungen eindeutig festgelegt. Dies gilt sehr viel allgemeiner: Sind X_t Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in Ω_t für alle $t \in T$ (T beliebig!), so ist der daraus gebildete **stochastische Prozess** $(X_t)_{t\in T} : \Omega \to \times_{t\in T} \Omega_t$ $\mathcal{A}, \bigotimes_{t\in T} \mathcal{A}_t$-messbar. $P^{(X_t)_{t\in T}}$ ist dann durch alle Randverteilungen $P^{(X_{t_1},\ldots,X_{t_n})}$ für alle $t_1, \ldots, t_n \in T, n \in \mathbb{N}$ eindeutig festgelegt.

Umgekehrt kann man zu einem vorgegebenen "konsistenten" System von endlichdimensionalen Wahrscheinlichkeitsmaßen $\mu_{t_1,\ldots,t_n}$ auf $\times_{i=1}^{n} \Omega_{t_i}, (t_1, \ldots, t_n) \in T^n$, einen stochastischen Prozess finden, so dass $P^{(X_{t_1},\ldots,X_{t_n})} = \mu_{t_1,\ldots,t_n}$. Der Einfachheit halber beschränken wir uns hier auf den Fall, dass alle Bildräume identisch sind, d.h. es gelte $(\Omega_t, \mathcal{A}_t) = (\tilde{\Omega}, \tilde{\mathcal{A}})$ für alle $t \in T$. Insbesondere muss dann für alle Permutationen $\pi \in S_n$ auf $\{1, \ldots, n\}$ und die zugehörigen Transformationen $U_\pi : \tilde{\Omega}^n \to \tilde{\Omega}^n$, $U_\pi(x_1, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$ gelten

$$\mu_{t_1,\ldots,t_n}^{U_\pi} = \left(P^{(X_{t_1},\ldots,X_{t_n})}\right)^{U_\pi} = P^{U_\pi(X_{t_1},\ldots,X_{t_n})} = P^{(X_{t_{\pi(1)}},\ldots,X_{t_{\pi(n)}})} = \mu_{t_{\pi(1)},\ldots,t_{\pi(n)}}$$

und für alle Projektionen $\tilde{pr}_n : \tilde{\Omega}^n \to \tilde{\Omega}^{n-1}$, $\tilde{pr}_n(x_1, \ldots, x_n) = (x_1, \ldots, x_{n-1})$

$$\mu_{t_1,\ldots,t_n}^{\tilde{pr}_n} = \left(P^{(X_{t_1},\ldots,X_{t_n})}\right)^{\tilde{pr}_n} = P^{(X_{t_1},\ldots,X_{t_{n-1}})} = \mu_{t_1,\ldots,t_{n-1}}.$$

4.16 Satz (Konsistenzsatz/Existenzsatz von Kolmogorov) *Ist $\mu_{t_1,\ldots,t_n}$ für alle $n \in \mathbb{N}$, $t_1, \ldots, t_n \in T$ ein Wahrscheinlichkeitsmaß auf $(\tilde{\Omega}^n, \tilde{\mathcal{A}}^n)$, so dass für alle $\pi \in S_n, n \in \mathbb{N}$*

$$\mu_{t_1,\ldots,t_n}^{U_\pi} = \mu_{t_{\pi(1)},\ldots,t_{\pi(n)}}$$

und für alle $n \geq 2$

$$\mu_{t_1,\ldots,t_n}^{\tilde{pr}_n} = \mu_{t_1,\ldots,t_{n-1}}$$

(mit den oben definierten Abbildungen U_π und $\tilde{pr}_n$), dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und ein stochastischer Prozess $(X_t)_{t\in T}$ mit $X_t : (\Omega, \mathcal{A}) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$, so dass $P^{(X_{t_1},\ldots,X_{t_n})} = \mu_{t_1,\ldots,t_n}$ für alle $t_i \in T, n \in \mathbb{N}$.

$P^{(X_t)_{t\in T}}$ ist eindeutig bestimmt. □

5 Bedingte Erwartungen und bedingte Verteilungen

Sei im folgenden stets ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ gegeben. Zu zwei Zufallsvariablen

$$X : (\Omega, \mathcal{A}) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$$
$$Y : (\Omega, \mathcal{A}) \to (S, \mathcal{D})$$

soll ein mittlerer Wert oder eine beste Approximation von X bei bekanntem Wert von Y bestimmt werden.

Ist Y diskret, gibt es also eine abzählbare Menge $S_0 \subset S$ mit $P\{Y \in S_0\} = 1$ und gilt $\{y\} \in \mathcal{D}$ für alle $y \in S$, so definiert für alle $y \in S_0$ mit $P\{Y = y\} > 0$

$$P^{X|Y=y}(\cdot) = P(X \in \cdot \mid Y = y) : \quad \bar{\mathbb{B}} \to [0,1]$$
$$B \mapsto P(X \in B \mid Y = y) = \frac{P\{X \in B, Y = y\}}{P\{Y = y\}}$$

ein Wahrscheinlichkeitsmaß auf $(\bar{\mathbb{R}}, \bar{\mathbb{B}})$, das sich als bedingte Verteilung von X gegeben $Y = y$ auffassen lässt.

Bezeichnet Q_y das Wahrscheinlichkeitsmaß mit P-Dichte $1_{\{Y=y\}}/P\{Y = y\}$, so gilt für alle $B \in \bar{\mathbb{B}}$

$$P(X \in B \mid Y = y) = \int 1_{\{X \in B\}} \frac{1_{\{Y=y\}}}{P\{Y = y\}}\, dP = \int 1_B \circ X\, dQ_y = Q_y^X(B),$$

d.h. $P^{X|Y=y}$ ist zugleich die Verteilung von X unter Q_y. Damit lässt sich der Mittelwert von $P^{X|Y=y}$, der sich auch als bedingter Erwartungswert von X gegeben $Y = y$ interpretieren lässt, berechnen als

$$\begin{aligned} E(X|Y = y) &:= \int x\, P^{X|Y=y}(dx) \\ &= \int x\, Q_y^X(dx) \\ &= \int X\, dQ_y \\ &= \int X \frac{1_{\{Y=y\}}}{P\{Y = y\}}\, dP \\ &= \frac{\int_{\{Y=y\}} X\, dP}{P\{Y = y\}}, \end{aligned} \tag{5.1}$$

vorausgesetzt dass alle auftretenden Integrale wohldefiniert sind. Diese Größe kann man nun für alle möglichen Werte $y \in S$ zusammenfassend darstellen durch die Funktionen

$$E(X|Y = \cdot) : \quad (S, \mathcal{D}) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$$
$$y \mapsto E(X|Y = y)$$

bzw.

$$E(X|Y) = E(X|Y = \cdot) \circ Y : \quad \begin{aligned} &(\Omega, \mathcal{A}) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}}) \\ &\omega \mapsto E(X|Y = Y(\omega)). \end{aligned}$$

(Dabei wird $E(X|Y = y)$ für die P^Y-f.s. nicht auftretenden Werte $y \in S$ mit $P\{Y = y\} = 0$ beliebig so definiert, dass die Abbildung messbar ist.) Man beachte, dass $E(X|Y)$ eine *Zufallsvariable* ist, die vom Argument ω nur über den Wert $Y(\omega)$ der Zufallsvariablen Y abhängt, also eine messbare Funktion von Y ist.

Gemäß Bemerkung 5.17 (ii) der Mathematischen Stochastik ist der Erwartungswert einer quadratintegrierbaren Zufallsvariable X diejenige Konstante, die minimalen L_2-Abstand zu X besitzt. Daraus folgt sofort, dass $E(X|Y)$ unter allen Zufallsvariablen der Form $h \circ Y$ diejenige ist, die minimalen L_2-Abstand zu X besitzt, d.h. der bedingte Erwartungswert von X ist die beste L_2-Approximation von X unter allen Zufallsvariablen, die nur die von der durch Y gegebenen Information abhängt.

Es gilt dann in Verallgemeinerung von (5.1) für eine beliebige Menge $D \in \mathcal{D}$

$$\begin{aligned} \int_{\{Y \in D\}} X \, dP &= \sum_{y \in D \cap S_0} \int_{\{Y = y\}} X \, dP \\ &\overset{(5.1)}{=} \sum_{y \in D \cap S_0} E(X|Y = y) \cdot P\{Y = y\} \\ &= \sum_{y \in D \cap S_0} \int E(X|Y) 1_{\{Y = y\}} \, dP \\ &= \int_{\{Y \in D\}} E(X|Y) \, dP. \end{aligned} \tag{5.2}$$

Wird nun Y nicht mehr als diskret angenommen, so lässt sich (5.1) nicht mehr für die Definition eines bedingten Erwartungswerts verwenden, da z.B. $P\{Y = y\} = 0$ für *alle* $y \in S_0$, wenn P^Y keinen diskreten Anteil besitzt. Dagegen lässt sich der in (5.2) festgestellte Zusammenhang auf beliebige Zufallsvariablen Y verallgemeinern. Dabei fällt auf, dass die Zufallsvariable Y bloß noch über die von ihr erzeugte σ-Algebra

$$\sigma(Y) := Y^{-1}(\mathcal{D}) := \{Y^{-1}(D) \mid D \in \mathcal{D}\}$$

eingeht. Daher kann man entsprechend auch gleich einen bedingten Erwartungswert von X gegeben eine σ-Algebra definieren.

5.1 Definition *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : (\Omega, \mathcal{A}) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$ eine P-integrierbare Zufallsvariable.*

(i) Sei $\mathcal{C}$ eine Unter-σ-Algebra von $\mathcal{A}$. Eine Zv. $Z : (\Omega, \underline{\mathcal{C}}) \to (\bar{\mathbb{R}}, \bar{B})$ heißt **bedingter Erwartungswert von** X **gegeben** $\mathcal{C}$ *(i.Z. $\overline{\overline{E(X|\mathcal{C})}}$), falls die folgenden* **Radon-Nikodym-Gleichungen** *erfüllt sind:*

$$\int_C X \, dP = \int_C Z \, dP \quad \forall \, C \in \mathcal{C}.$$

(ii) Ist $Y : (\Omega, \mathcal{A}) \to (S, \mathcal{D})$ eine beliebige Zv. und

$$\mathcal{C} := \sigma(Y) := \{Y^{-1}(D) \mid D \in \mathcal{D}\}$$

die davon erzeugte σ-Algebra, so heißt eine Zv. $Z : (\Omega, \mathcal{C}) \to (\bar{\mathbb{R}}, \bar{B})$ **bedingter Erwartungswert von** X **gegeben** Y *(i.Z. $E(X|Y)$), falls sie ein bedingter Erwartungswert von X gegeben $\mathcal{C}$ ist, d.h. falls die folgenden* **Radon-Nikodym-Gleichungen** *erfüllt sind:*

$$\int_{\{Y \in D\}} X \, dP = \int_{\{Y \in D\}} Z \, dP \quad \forall D \in \mathcal{D}.$$

5.2 Bemerkung (i) Ist X endlich integrierbar, Z ein bedingter Erwartungswert von X gegeben $\mathcal{C}$ und $\tilde{Z}$ eine weitere $\mathcal{C}, \bar{\mathbb{B}}$-messbare Zv., so ist $\tilde{Z}$ genau dann eine bedingter Erwartungswert von X gegeben $\mathcal{C}$, wenn $Z = \tilde{Z}$ P-f.s. Man schreibt daher meist $Z = E(X|Y)$ P-f.s.

(ii) Ist X endlich integrierbar, so reicht es in der Situation von Definition 5.1(i) aus, die Radon-Nikodym-Gleichung für alle Mengen C eines $\cap$-stabilen Erzeugendensystems von $\mathcal{C}$ nachzuweisen. Für nicht-negative Zv. X und Z definieren nämlich beide Seiten der Radon-Nikodym-Gleichung (als Funktion von C betrachtet) endliche Maße auf $(\Omega, \mathcal{C})$, die gemäß dem Maßfortsetzungssatz durch die Werte auf einem $\cap$-stabilen Erzeugendensystem eindeutig bestimmt sind. Wenn die Maße also auf einem solchen Erzeugendensystem übereinstimmen, so auch auf der ganzen σ-Algebra $\mathcal{C}$. Für beliebige Zv. folgt das Ergebnis durch Zerlegung in den Positiv- und Negativteil.

Ebenso reicht es in der Situation von Definition 5.1(ii) aus, die Radon-Nikodym-Gleichung für alle Mengen D eines $\cap$-stabilen Erzeugendensystems von $\mathcal{D}$ nachzuweisen. Insbesondere reicht es für eine $\mathbb{R}$-wertige Zv. Y aus, Mengen der Form $D = (-\infty, t]$ für alle $t \in \mathbb{R}$ zu betrachten. □

5.3 Satz *Unter den Bedingungen von Definition 5.1(i) existiert stets ein bedingter Erwartungswert $E(X|\mathcal{C})$.* □

In der Situation von Definition 5.1(ii) ist die nachfolgende Charakterisierung der Messbarkeit bzgl. $\sigma(Y)$ nützlich.

5.4 Lemma (Faktorisierungslemma) *Sind $Y : (\Omega, \mathcal{A}) \to (S, \mathcal{D})$ und $Z : (\Omega, \mathcal{A}) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$ Zv., so ist Z genau dann $\sigma(Y), \tilde{\mathcal{A}}$-messbar, wenn es eine messbare Abbildung $h : (S, \mathcal{D}) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$ gibt, so dass $Z = h \circ Y$.* □

5.5 Definition *In der Situation von 5.1(ii) heißt eine Abbildung $h : (S, \mathcal{D}) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$* **faktorisierter bedingter Erwartungswert von** X **gegeben** Y*, falls $h \circ Y = E(X|Y)$ P-f.s.; i.Z. $E(X|Y = \cdot) = h$.*

Für alle $y \in S$ heißt dann $h(y) = E(X|Y = y)$ **(faktorisierter) bedingter Erwartungswert von** X **gegeben** $Y = y$*.* □

5.6 Bemerkung Es gilt also $h = E(X|Y = \cdot)$, falls h eine $\mathcal{D}, \bar{\mathbb{B}}$-messbare Abbildung ist, die die folgenden Radon-Nikodym-Gleichungen erfüllt:

$$\int_{\{Y\in D\}} X\,dP = \int_D h\,dP^Y.$$

$E(X|Y = \cdot)$ ist wegen $E(X|Y) = E(X|Y = \cdot) \circ Y$ nur P^Y-f.s. eindeutig bestimmt. □

Im folgenden betrachten wir bedingte Erwartungswerte i.W. nur noch für endlich integrierbare, reellwertige Zv., um die Formulierung der nachfolgenden Resultate möglichst einfach zu halten. In dem Fall können auch die bedingten Erwartungswerte stets reellwertig gewählt werden.

Im allgemeinen gibt es keinen konstruktiven Weg, bedingte Erwartungswerte zu berechnen. Ausnahmen sind insbesondere diskrete Zv. Y (s. die einführende Diskussion) und Paare (X, Y), die eine Dichte bzgl. eines Produktmaßes besitzen.

5.7 Satz und Definition *Sei in der Situation von Definition 5.1 (mit $E(|X|) < \infty$) μ ein σ-endliches Maß auf $(\bar{\mathbb{R}}, \bar{\mathbb{B}})$ und ν ein σ-endliches Maß auf $(S, \mathcal{D})$, so dass $P^{(X,Y)}$ eine $\mu\otimes\nu$-Dichte f besitzt, dann gilt für die* **bedingte μ-Dichte von X gegeben $Y = y$**

$$f(x|y) := \frac{f(x,y)}{\displaystyle\int_{-\infty}^{\infty} f(t,y)\,\mu(dt)}, \quad x \in \mathbb{R},$$

(mit der Konvention $0/0 := 0$) und P^Y-f.a. $y \in S$

$$E(X|Y = y) = \int x f(x|y)\,\mu(dx).$$

□

Die folgenden Rechenregeln erleichtern den Umgang mit bedingten Erwartungswerten

5.9 Satz *Seien $X, Y : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathbb{B})$ mit $E(|X|)+E(|Y|) < \infty$ und $Z : (\Omega, \mathcal{A}) \to (S, \mathcal{D})$. Dann gilt:*

(i) $E\big(E(X \mid Z)\big) = E(X)$

(ii) $E(aX + bY \mid Z) = aE(X \mid Z) + bE(Y \mid Z)$ *P-f.s. für alle* $a, b \in \mathbb{R}$

(iii) Gilt $X = g(Z)$ für ein $g : (S, \mathcal{D}) \to (\mathbb{R}, \mathbb{B})$, so folgt $E(X \mid Z) = X$ P-f.s.

(iv) Sind X und Z stochastisch unabhängig, so gilt $E(X \mid Z) = E(X)$ P-f.s.

(v) Gilt $X = Y$ P-f.s., so gilt $E(X \mid Z) = E(Y \mid Z)$ P-f.s.

(vi) Gilt $X \le Y$ P-f.s., so gilt $E(X \mid Z) \le E(Y \mid Z)$ P-f.s.

(vii) Konvergiert eine Folge von Zv. $X_n : (\Omega, \mathcal{A}) \to \big([0,\infty), \mathcal{B}([0,\infty))\big)$, $n \in \mathbb{N}$, von unten monoton P-f.s. gegen X, so gilt $\lim_{n\to\infty} E(X_n \mid Z) = E(X \mid Z)$ P-f.s.

(viii) *Konvergiert eine Folge von Zv.* $X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$, $n \in \mathbb{N}$, *gegen eine Zv.* X *P-f.s. und existiert eine majorisierende Zv.* $Y : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$ *(d.h.* $|X_n| \leq Y$ *P-f.s. für alle* $n \in \mathbb{N}$*) mit* $E(Y) < \infty$*, so gilt* $\lim_{n\to\infty} E(X_n \mid Z) = E(X \mid Z)$ *P-f.s.*

(ix) *Gilt* $X = g(Z)$ *für ein* $g : (S, \mathcal{D}) \to (\mathbb{R}, \mathbb{B})$ *und* $E(|XY|) < \infty$*, so folgt* $E(XY \mid Z) = XE(Y \mid Z)$ *P-f.s.*

(x) *Gilt* $E(|XY|) < \infty$*, so folgt* $E\big(E(X \mid Z) \cdot Y \mid Z\big) = E(X \mid Z) \cdot E(Y \mid Z)$ *P-f.s.*

(xi) *Ist* $Z = (Z_1, Z_2)$ *mit* $Z_i : (\Omega, \mathcal{A}) \to (S_i, \mathcal{D}_i)$ *(und daher* $S = S_1 \times S_2$ *und* $\mathcal{D} = \mathcal{D}_1 \otimes \mathcal{D}_2$*), so gilt*

$$E\big(E(X \mid Z_1) \mid Z\big) = E(X \mid Z_1) = E\big(E(X \mid Z) \mid Z_1\big) \quad P\text{-f.s.}$$

(xii) *Sind in der Situation von (xi)* (X, Z_1) *und* Z_2 *stochastisch unabhängig, dann gilt* $E(X \mid Z) = E(X \mid Z_1)$ *P-f.s.*

(xiii) *Sind in der Situation von (xi)* Z_2 *und* X **bedingt stochastisch unabhängig gegeben** Z_1*, d.h. gilt*

$$\begin{aligned} P^{(X,Z_2)|Z_1}(B \times D_2) &:= E\big(1_B(X) \cdot 1_{D_2}(Z_2) \mid Z_1\big) \\ &= E\big(1_B(X) \mid Z_1\big) \cdot E\big(1_{D_2}(Z_2) \mid Z_1\big) \\ &=: P^{X|Z_1}(B) \cdot P^{Z_2|Z_1}(D_2) \quad P\text{-f.s.} \quad \forall\, B \in \mathbb{B}, D_2 \in \mathcal{D}_2, \end{aligned}$$

so folgt $E(X \mid Z) = E(X \mid Z_1)$ *P-f.s.*

(xiv) *Jensen'sche Ungleichung: Ist* $I \subset \mathbb{R}$ *ein Intervall und* X *eine* I*-wertige endlich integrierbar Zufallsvariable, so gilt für jede konvexe Funktion* $\varphi : I \to \mathbb{R}$*:* $E(\varphi(X)|Z) \geq \varphi\big(E(X|Z)\big)$ *P-f.s.* □

Beweis. (i) Dies ist ein Spezialfall der Radon-Nikodym-Gleichung für $D = S$.

(ii) Die rechte Seite ist eine messbare Funktion von Z und die Radon-Nikodym-Gleichung verifiziert man leicht durch direktes Nachrechnen.

(iii) dito

(iv) Wegen der Unabhängigkeit von X und $1_D(Z)$ für alle $D \in \mathcal{D}$ gilt

$$\int_{\{Z\in D\}} X\, dP = E\big(1_D(Z) \cdot X\big) = E(1_D(Z)) \cdot E(X) = \int_{\{Z\in D\}} E(X)\, dP,$$

also die Radon-Nikodym-Gleichung.

(v) folgt aus (vi).

(vi) Angenommen $P(C) > 0$ für $C := \{E(X|Z) > E(Y|Z)\}$. Da $E(X|Z)$ und $E(Y|Z)$ messbare Funktionen von Z sind, kann man die Menge C darstellen als $C = \{Z \in D\}$ für ein $D \in \mathcal{D}$. Man erhält dann wegen der Radon-Nikodym-Gleichung einen Widerspruch:

$$0 \geq \int_{\{Z\in D\}} X - Y\, dP = \int_{\{Z\in D\}} E(X|Z) - E(Y|Z)\, dP > 0.$$

(vii) Gemäß (vi) sind die Zv. $E(X_n \mid Z)$ P-f.s. nicht-negativ und die Folge ist monoton steigend. Die Radon-Nikodym-Gleichung weist man daher leicht mit dem Satz von der monotonen Konvergenz nach, während die Messbarkeit von $\lim_{n\to\infty} E(X_n \mid Z)$ bzgl. $\sigma(Z)$ direkt aus der entsprechenden Messbarkeit von $E(X_n \mid Z)$ für alle $n \in \mathbb{N}$ folgt.

(viii) Wendet man (vii) auf die monoton steigende Folge $Y_n := Y + \inf_{k\geq n} X_k$ von nicht negativen Zv. an, die gegen $Y + X$ konvergiert, so erhält man $E(Y + \inf_{k\geq n} X_k \mid Z) \uparrow E(Y + X \mid Z)$ und folglich auch $E(\inf_{k\geq n} X_k \mid Z) \uparrow E(X \mid Z)$ P-f.s. Ebenso erhält man durch Anwendung von (vii) auf die Folge $\tilde{Y}_n := Y - \sup_{k\geq n} X_k \geq 0$ P-f.s. die Konvergenz von $E(\sup_{k\geq n} X_k \mid Z)$ von oben gegen $E(X \mid Z)$. Die Behauptung folgt daher wegen der Monotonieeigenschaft (vi) aus $\inf_{k\geq n} X_k \leq X_n \leq \sup_{k\geq n} X_k$.

(ix) Nach Voraussetzung und der Definition des bedingten Erwartungswerts ist $XE(Y|Z)$ eine messbare Funktion von Z. Die Radon-Nikodym-Gleichung weist man zunächst für einfache Zv. X nach, also für Zv. der Form $X = \sum_{i=1}^n a_i 1_{D_i}(Z)$: Für alle $D \in \mathcal{D}$ gilt dann

$$\begin{aligned}\int_{\{Z\in D\}} XE(Y \mid Z)\, dP &= \sum_{i=1}^n a_i \int_{\{Z\in D\cap D_i\}} E(Y \mid Z)\, dP \\ &= \sum_{i=1}^n a_i \int_{\{Z\in D\cap D_i\}} Y\, dP \\ &= \int_{\{Z\in D\}} XY\, dP.\end{aligned}$$

Ist X eine positive Zv., so lässt sie sich durch eine monoton steigende Folge von einfachen Zv. X_n approximieren und die Behauptung folgt mit (vii). Für eine beliebige Zv. X folgt die Behauptung schließlich durch Zerlegung in den Positiv- und Negativteil.

(x) Dies ist ein Spezialfall von (ix), da $E(X \mid Z)$ sich als $g(Z)$ mit $g = E(X \mid Z = \cdot)$ darstellen lässt.

(xi) Die erste Gleichung folgt direkt aus (iii), da $E(X \mid Z_1)$ messbar ist bzgl. $\sigma(Z_1)$ und damit erst recht messbar bzgl. $\sigma(Z) \supset \sigma(Z_1)$.

Für den Nachweis der zweiten Gleichung beachte man die Radon-Nikodym-Gleichungen

$$\begin{aligned}\int_{\{Z_1\in D_1\}} E(X \mid Z_1)\, dP &= \int_{\{Z_1\in D_1\}} X\, dP \qquad \forall\, D_1 \in \mathcal{D}_1 \\ \int_{\{Z\in D\}} E(X \mid Z)\, dP &= \int_{\{Z\in D\}} X\, dP \qquad \forall\, D \in \mathcal{D}\end{aligned}$$

Speziell für $D = D_1 \times S_2$ ergibt sich wegen $\{Z \in D\} = \{Z_1 \in D_1\}$

$$\int_{\{Z_1\in D_1\}} E(X \mid Z)\, dP = \int_{\{Z_1\in D_1\}} X\, dP = \int_{\{Z_1\in D_1\}} E(X \mid Z_1)\, dP,$$

d.h. $E(X|Z_1)$ erfüllt die Radon-Nikodym-Gleichung aus der Definition von $E\big(E(X|Z) \mid Z_1\big)$ und ist außerdem offensichtlich $\sigma(Z_1)$-messbar.

(xii) Die Behauptung folgt aus (xiii), wenn wir nachweisen, dass unter den genannten Voraussetzungen Z_2 und X bedingt unabhängig sind gegeben Z_1. Dies folgt aus (iv), (ix) und (xi):

$$\begin{aligned} E\big(1_B(X)\cdot 1_{D_2}(Z_2)\mid Z_1\big) &\overset{(xi)}{=} E\big(E\big(1_B(X)\cdot 1_{D_2}(Z_2)\mid (X,Z_1)\big)\mid Z_1\big)\\ &\overset{(ix)}{=} E\big(1_B(X)\cdot E(1_{D_2}(Z_2)\mid (X,Z_1))\mid Z_1\big)\\ &\overset{(iv)}{=} E\big(1_B(X)\cdot E(1_{D_2}(Z_2))\mid Z_1\big)\\ &\overset{(ii)}{=} E(1_B(X)\mid Z_1)\cdot E(1_{D_2}(Z_2))\\ &\overset{(iv)}{=} E(1_B(X)\mid Z_1)\cdot E(1_{D_2}(Z_2)\mid Z_1) \qquad P\text{-f.s.}\end{aligned}$$

(xiii) Offensichtlich ist $E(X|Z_1)$ messbar bzgl. $\sigma(Z)$, so dass nur die Radon-Nikodym-Gleichung zu zeigen bleibt, d.h. für alle $D\in\mathcal{D}$

$$\int_{\{Z\in D\}} E(X\mid Z_1)\,dP = \int_{\{Z\in D\}} X\,dP. \tag{5.3}$$

Dazu beschränken wir uns zunächst auf Mengen D der Form $D=D_1\times D_2$. Ähnlich wie im Beweis von (ix) kann man aus der Voraussetzung durch "algebraische Induktion" (für h) folgern, dass $E(h(X)\cdot 1_{D_2}(Z_2)\mid Z_1) = E(h(X)\mid Z_1)\cdot E(1_{D_2}(Z_2)\mid Z_1)$ P-f.s. gilt, wenn h eine messbare Funktion ist, so dass $E(|h(X)|)<\infty$ (Übungen). Man erhält damit insbesondere für $h(X)=X$

$$\begin{aligned}\int_{\{Z\in D\}} E(X\mid Z_1)\,dP &= E\big(1_{D_1}(Z_1)\cdot E(X\mid Z_1)\cdot 1_{D_2}(Z_2)\big)\\ &\overset{(i)}{=} E\Big(E\big(1_{D_1}(Z_1)\cdot E(X\mid Z_1)\cdot 1_{D_2}(Z_2)\mid Z_1\big)\Big)\\ &\overset{(ix),(x)}{=} E\big(1_{D_1}(Z_1)\cdot E(X\mid Z_1)\cdot E(1_{D_2}(Z_2)\mid Z_1)\big)\\ &\overset{s.o.}{=} E\big(1_{D_1}(Z_1)\cdot E(X\cdot 1_{D_2}(Z_2)\mid Z_1)\big)\\ &= \int_{\{Z_1\in D_1\}} X\cdot 1_{D_2}(Z_2)\,dP\\ &= \int_{\{Z\in D\}} X\,dP,\end{aligned}$$

wobei im vorletzten Schritt die Radon-Nikodym-Gleichung für $E(X1_{D_2}(Z_2)\mid Z_1)$ verwendet worden ist. Wir haben damit gezeigt, dass (5.3) für alle $D=D_1\times D_2$ gilt, woraus nach Bemerkung 5.2 (iii) die Behauptung folgt, da $\{D_1\times D_2\mid D_i\in\mathcal{D}_i\}$ ein $\cap$-stabiles Erzeugendensystem von $\mathcal{D}$ ist.

(xiv) Wegen (vi) nimmt $E(X|Z)$ P-f.s. auch nur Werte in I an. Da φ konvex ist, gibt es zu jedem $x_0\in I$ Konstanten $a,b\in\mathbb{R}$, so dass $\varphi(x_0)=ax_0+b$ und $\varphi(x)\ge ax+b$ für alle $x\in I$. Diese Konstanten kann man als messbare Funktionen angewendet auf x_0 darstellen. Speziell für $x_0=E(X|Z)$ gibt es daher Zufallsvariablen $a_Z=g_1(Z)$ und $b_Z=g_2(Z)$ für geeignete Abbildungen $g_1,g_2:(S,\mathcal{D})\to(\mathbb{R},\mathbb{B})$, so dass

$$\begin{aligned}\varphi\big(E(X|Z)\big) &= a_Z E(X|Z)+b_Z = E(a_Z X|Z)+E(b_Z|Z) = E(a_Z X+b_Z|Z)\\ &\le E(\varphi(X)|Z) \qquad P\text{-f.s.}\end{aligned}$$

Dabei haben wir beim 2. Gleichheitszeichen (iii) und (ix), beim letzten Gleichheitszeichen (ii) und für die Ungleichung (vi) angewendet. □

5.10 Bemerkung Natürlich überträgt sich Satz 5.9 sinngemäß auf bedingte Erwartungswerte gegeben eine Unter-σ-Algebra.

Seien dazu $X, Y : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathbb{B})$ mit $E(|X|) + E(|Y|) < \infty$ und $\mathcal{C} \subset \mathcal{A}$ eine Unter-σ-Algebra von $\mathcal{A}$. Dann gilt:

(i) $E\big(E(X \mid \mathcal{C})\big) = E(X)$

(ii) $E(aX + bY \mid \mathcal{C}) = aE(X \mid \mathcal{C}) + bE(Y \mid \mathcal{C})$ P-f.s. für alle $a, b \in \mathbb{R}$

(iii) Ist X $\mathcal{C}$-messbar, so folgt $E(X \mid \mathcal{C}) = X$ P-f.s.

(iv) Sind $\sigma(X)$ und $\mathcal{C}$ stochastisch unabhängig (d.h. sind A und C stochastisch unabhängig für alle $A \in \sigma(X)$ und alle $C \in \mathcal{C}$), so gilt $E(X \mid \mathcal{C}) = E(X)$ P-f.s.

(v) Gilt $X = Y$ P-f.s., so gilt $E(X \mid \mathcal{C}) = E(Y \mid \mathcal{C})$ P-f.s.

(vi) Gilt $X \leq Y$ P-f.s., so gilt $E(X \mid \mathcal{C}) \leq E(Y \mid \mathcal{C})$ P-f.s.

(vii) Konvergiert eine Folge von Zv. $X_n : (\Omega, \mathcal{A}) \to \big([0, \infty), \mathcal{B}([0, \infty))\big)$, $n \in \mathbb{N}$, von unten monoton P-f.s. gegen X, so gilt $\lim_{n\to\infty} E(X_n \mid \mathcal{C}) = E(X \mid \mathcal{C})$ P-f.s.

(viii) Konvergiert eine Folge von Zv. $X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$, $n \in \mathbb{N}$, gegen eine Zv. X und existiert eine majorisierende Zv. $Y : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$ mit $E(Y) < \infty$, so gilt $\lim_{n\to\infty} E(X_n \mid \mathcal{C}) = E(X \mid \mathcal{C})$ P-f.s.

(ix) Ist X $\mathcal{C}$-messbar und gilt $E(|XY|) < \infty$, so folgt $E(XY \mid \mathcal{C}) = XE(Y \mid \mathcal{C})$ P-f.s.

(x) Gilt $E(|XY|) < \infty$, so folgt $E\big(E(X \mid \mathcal{C}) \cdot Y \mid \mathcal{C}\big) = E(X \mid \mathcal{C}) \cdot E(Y \mid \mathcal{C})$ P-f.s.

(xi) Sind $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{A}$ Unter-σ-Algebren, so gilt

$$E\big(E(X \mid \mathcal{C}_1) \mid \mathcal{C}_2\big) = E(X \mid \mathcal{C}_1) = E\big(E(X \mid \mathcal{C}_2) \mid \mathcal{C}_1\big) \quad P\text{-f.s.}$$

(xii) Sind $\mathcal{C}_1, \mathcal{C}_2$ Unter-σ-Algebren von $\mathcal{A}$, so dass $\sigma(\sigma(X) \cup \mathcal{C}_1)$ und $\mathcal{C}_2$ stochastisch unabhängig sind, dann gilt $E(X \mid \mathcal{C}) = E(X \mid \mathcal{C}_1)$ P-f.s. mit $\mathcal{C} := \sigma(\mathcal{C}_1 \cup \mathcal{C}_2)$.

(xiii) Sind $\mathcal{C}_1, \mathcal{C}_2$ Unter-σ-Algebren von $\mathcal{A}$, so dass X und $\mathcal{C}_2$ **bedingt stochastisch unabhängig gegeben** $\mathcal{C}_1$ sind, d.h.

$$E\big(1_B(X) \cdot 1_{C_2} \mid \mathcal{C}_1\big) = E\big(1_B(X) \mid \mathcal{C}_1\big) \cdot E\big(1_{C_2} \mid \mathcal{C}_1\big) \quad P\text{-f.s.} \quad \forall\, B \in \mathbb{B}, C_2 \in \mathcal{C}_2,$$

so folgt $E(X \mid \mathcal{C}) = E(X \mid \mathcal{C}_1)$ P-f.s. mit $\mathcal{C} := \sigma(\mathcal{C}_1 \cup \mathcal{C}_2)$.

(xiv) Jensen'sche Ungleichung: Ist $I \subset \mathbb{R}$ ein Intervall und X eine I-wertige endlich integrierbar Zufallsvariable, so gilt für jede konvexe Funktion $\varphi : I \to \mathbb{R}$: $E(\varphi(X)|\mathcal{C}) \geq \varphi\big(E(X|\mathcal{C})\big)$ P-f.s. □

Falls $E(X^2) < \infty$, so kann man (wie schon im eingangs betrachteten Fall einer diskreten Zv. Z) $E(X \mid Z)$ als beste Approximation von X (im L_2-Sinn) auffassen, die nur die durch Z gegebene Information verwendet.

5.12 Satz *Seien $X, Y : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$ Zv. mit $E(X^2) < \infty$, und $Z : (\Omega, \mathcal{A}) \to (S, \mathcal{D})$ eine Zv., so dass $Y = h \circ Z$ für eine messbare Funktion $h : (S, \mathcal{D}) \to (\mathbb{R}, \mathbb{B})$ (d.h. Y ist $\sigma(Z), \mathbb{B}$-messbar). Dann gilt*

$$E\big((X - E(X \mid Z))^2\big) \leq E((X - Y)^2)$$

und "=" gilt genau dann, wenn $Y = E(X \mid Z)$ P-f.s. □

Beweis. Im Fall $E(Y^2) = \infty$ ist die rechte Seite der Ungleichung unendlich und die Behauptung daher trivial. Andernfalls gilt gemäß der Cauchy-Schwarzschen Ungleichung $E(|XY|) < \infty$. Für alle $\tilde{Y} = \tilde{h} \circ Z$ mit $\tilde{h} : (S, \mathcal{D}) \to (\mathbb{R}, \mathbb{B})$ und $E(\tilde{Y}^2) < \infty$ gilt

$$\begin{aligned} E\big((X - E(X \mid Z)) \cdot \tilde{Y} \mid Z\big) &\overset{5.9(ix)}{=} \tilde{Y} \cdot E(X - E(X \mid Z) \mid Z) \\ &\overset{5.9(ii),(iii)}{=} \tilde{Y} \cdot (E(X \mid Z) - E(X \mid Z)) \\ &= 0. \end{aligned} \tag{5.4}$$

Speziell mit $\tilde{Y} = E(X \mid Z) - Y = \big(E(X \mid Z = \cdot) - h\big) \circ Z$ folgt

$$\begin{aligned} E\big((X - Y)^2\big) &\overset{5.9(i)}{=} E\Big(E\big((X - Y)^2 \mid Z\big)\Big) \\ &= E\Big(E\big(((X - E(X \mid Z)) + \tilde{Y})^2 \mid Z\big)\Big) \\ &\overset{(5.4),5.9(ii)}{=} E\Big(E\big((X - E(X \mid Z))^2 \mid Z\big) + E(\tilde{Y}^2 \mid Z)\Big) \\ &\overset{5.9(i)}{=} E\big((X - E(X \mid Z))^2\big) + E(\tilde{Y}^2) \qquad (5.5) \\ &\geq E\big((X - E(X \mid Z))^2\big). \end{aligned}$$

Offensichtlich gilt beim letzten Schritt genau dann Gleichheit, wenn $\tilde{Y} = 0$ P-f.s. □

5.13 Bemerkung Die im Satz 5.12 beschriebene Ungleichung besagt gerade, dass $E(X|Z)$ (aufgefasst als Element im Hilbert-Raum $L_2(\Omega, \mathcal{A}, P)$) den kleinsten L_2-Abstand zu X unter allen Elementen des linearen Unterraums

$$\begin{aligned} M &:= \big\{Y \in L_2(\Omega, \mathcal{A}, P) \mid Y \text{ ist } \sigma(Z), \mathcal{B} - \text{messbar}\big\} \\ &= \big\{h \circ Z \mid h : (S, \mathcal{D}) \to (\mathbb{R}, \mathbb{B}) \text{ mit } E(h^2 \circ Z) < \infty\big\} \end{aligned}$$

besitzt, also gerade die Orthogonalprojektion von X auf M ist (s. Bemerkung 2.12). Wegen (5.4) steht nämlich $X - E(X|Z)$ orthogonal auf allen Elementen von M. Die Identität (5.5) ist gerade der Satz von Pythagoras in dieser Situation. □

Man kann wie in der eingangs geschilderten Situation, in der Y als diskret angenommen worden ist, auch den allgemeinen bedingten Erwartungswert $E(X \mid Y)$ als Mittelwert einer bedingten Verteilung $P^{X|Y}$ interpretieren. Eine solche bedingte Verteilung kann man aber auch für beliebige (d.h. nicht notwendig $\mathbb{R}$-wertige) Zv. X definieren.

5.14 Definition *Seien $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : (\Omega, \mathcal{A}) \to (\Omega_X, \mathcal{A}_X)$ und $Y : (\Omega, \mathcal{A}) \to (\Omega_Y, \mathcal{A}_Y)$ Zv. Dann heißt ein Markov-Kern K von $(\Omega, \sigma(Y))$ nach $(\Omega_X, \mathcal{A}_X)$ eine* **(reguläre) bedingte Verteilung von X gegeben Y**, *falls*

$$K(\cdot, A) = E(1_A(X) \mid Y)(\cdot) \quad P\text{-f.s.} \qquad \forall\, A \in \mathcal{A}_X;$$

i.Z. $K = P^{X|Y}$.

Ein Markov-Kern L von $(\Omega_Y, \mathcal{A}_Y)$ nach $(\Omega_X, \mathcal{A}_X)$ ist eine **faktorisierte (reguläre) bedingte Verteilung von X gegeben Y**, *falls*

$$L(\cdot, A) = E(1_A(X) \mid Y = \cdot) \quad P^Y\text{-f.s.} \qquad \forall\, A \in \mathcal{A}_X$$

(oder, äquivalent dazu, falls $K(\omega, A) = L(Y(\omega), A)$ für P-f.a. $\omega \in \Omega$ und alle $A \in \mathcal{A}_X$); i.Z. $L(y, \cdot) = P^{X|Y=y}(\cdot)$. □

5.15 Bemerkung (i) (Faktorisierte) bedingte Verteilungen existieren stets, wenn Ω_X ein sog. polnischer Raum ist (d.h. separabel und vollständig metrisierbar) und $\mathcal{A}_X$ die zugehörige Borel-σ-Algebra (also die von den offenen Mengen erzeugte σ-Algebra). Insbesondere gilt dies also für $\Omega_X \subset \mathbb{R}^k$ und $\mathcal{A}_X = \mathbb{B}^k \cap \Omega_X$.

(ii) Statt $E(1_A(X) \mid Y)$ schreibt man auch einfacher $P(X \in A \mid Y)$ und analog für die faktorisierten Versionen $P(X \in A \mid Y = y) := E(1_A(X) \mid Y = y)$. □

5.16 Satz *In der Situation von Definition 5.14 ist $P^{(Y,X)}$ gerade die Koppelung von P^Y und $P^{X|Y=\cdot}$:*

$$P^{(Y,X)} = P^Y \otimes P^{X|Y=\cdot}.$$

Insbesondere ist P^X die Mischung der Verteilungen $P^{X|Y=y}$, $y \in \Omega_Y$, unter der Mischungsverteilung P^Y:

$$P^X(A) = \int P^{X|Y=y}(A)\, P^Y(dy) \qquad \forall\, A \in \mathcal{A}_X.$$

□

Beweis. Für $A \in \mathcal{A}_X$ und $B \in \mathcal{A}_Y$ gilt

$$\begin{aligned} P^{(Y,X)}(B \times A) &= E\big(1_A(X) \cdot 1_B(Y)\big) \\ &\overset{5.9(i)}{=} E\big(E(1_A(X) \cdot 1_B(Y) \mid Y)\big) \\ &\overset{5.9(ix)}{=} E\big(E(1_A(X)|Y) \cdot 1_B(Y)\big) \\ &= E\big(P^{X|Y}(A) \cdot 1_B(Y)\big) \\ &= \int_B P^{X|Y=y}(A)\, P^Y(dy) \\ &= P^Y \otimes P^{X|Y=\cdot}(B \times A), \end{aligned}$$

wobei im vorletzten Schritt der Transformationssatz und die Definition von $P^{X|Y=y}$ verwendet worden sind und im letzten Schritt die Definition der Koppelung. Da die Rechteckmengen einen $\cap$-stabilen Erzeuger der Produkt-σ-Algebra bilden, ist die erste Behauptung gezeigt.

Die zweite Behauptung folgt durch Spezialisierung auf $B = \Omega_Y$. □

5.17 Korollar *Ist in der Situation von Definition 5.14* $(\Omega_X, \mathcal{A}_X) = (\mathbb{R}, \mathbb{B})$ *und* $E(|X|) < \infty$*, so gilt*

$$\begin{aligned} E(X \mid Y) &= \int x\, P^{X|Y}(dx) \quad P\text{-f.s.} \\ E(X \mid Y = y) &= \int x\, P^{X|Y=y}(dx) \quad \text{für } P^Y\text{-f.a. } y. \end{aligned}$$

Allgemeiner gilt für alle $h : (\Omega_X \times \Omega_Y, \mathcal{A}_X \otimes \mathcal{A}_Y) \to (\mathbb{R}, \mathbb{B})$ *mit* $E(|h(X,Y)|) < \infty$

$$\begin{aligned} E\big(h(X,Y)|Y\big) &= \int h(x,Y) P^{X|Y}(dx) \quad P\text{-f.s.} \\ E\big(h(X,Y)|Y = y\big) &= \int h(x,y) P^{X|Y=y}(dx) \quad \text{für } P^Y\text{-f.a. } y. \end{aligned}$$

□

Beweis. Es reicht, die letzte Identität nachzuweisen.
Gemäß dem Satz von Fubini für Übergangsmaße definiert

$$g(y) := \int h(x,y)\, P^{X|Y=y}(dx), \quad y \in \Omega_Y,$$

eine $\mathcal{A}_Y, \mathbb{B}$-messbare Abbildung und es gilt ferner für alle $B \in \mathcal{A}_Y$

$$\begin{aligned} \int_B g(y)\, P^Y(dy) &= \int \int h(x,y) 1_{\Omega_X \times B}(x,y)\, P^{X|Y=y}(dx)\, P^Y(dy) \\ &= \int \int h(x,y) 1_{\Omega_X \times B}(x,y)\, P^{(Y,X)}(d(y,x)) \\ &= \int h(X,Y) 1_{\Omega_X \times B}(X,Y)\, dP \\ &= \int_{\{Y \in B\}} h(X,Y)\, dP, \end{aligned}$$

d.h. g erfüllt auch die entsprechende Radon-Nikodym-Gleichung. □

6 Martingale

Martingale sind stochastische Modelle für Folgen von fairen Spielen: Gegeben die derzeitige Information ist der erwartete Spielstand nach der nächsten Spielrunde gleich dem jetzigen Spielstand.

6.1 Definition *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $T \subset \bar{\mathbb{R}}$ eine Indexmenge (z.B. $T = \mathbb{N}_0, T = [0, \infty)$ oder $T = \mathbb{Z}$). Eine Familie $(\mathcal{A}_t)_{t \in T}$ von Unter-σ-Algebren von $\mathcal{A}$ heißt* ***Filtration****, falls $\mathcal{A}_s \subset \mathcal{A}_t$ für alle $s \leq t$. Eine Familie $(X_t)_{t \in T}$ $\mathbb{R}$-wertiger Zufallsvariable (auf $(\Omega, \mathcal{A})$) wird auch als* ***stochastischer Prozess*** *bezeichnet. Der Prozess heißt* ***adaptiert*** *an $(\mathcal{A}_t)_{t \in T}$, falls X_t $\mathcal{A}_t, \mathbb{B}$-messbar ist für alle $t \in T$.*
Der Prozess heißt ***Martingal*** *(bzw.* ***Supermartingal****, bzw.* ***Submartingal****) bzgl. $(\mathcal{A}_t)_{t \in T}$, falls er an $(\mathcal{A}_t)_{t \in T}$ adaptiert ist, $E(|X_t|) < \infty$ für alle $t \in T$ gilt und $E(X_t | \mathcal{A}_s) = X_s$ (bzw. $E(X_t | \mathcal{A}_s) \leq X_s$ bzw. $E(X_t | \mathcal{A}_s) \geq X_s$) P-f.s. gilt für alle $s \leq t$, $s, t \in T$.* □

6.2 Bemerkung (i) Jeder Prozess ist adaptiert an seine **natürliche Filtration**

$$\mathcal{A}_t := \sigma\big((X_s)_{T \cap [-\infty, t]}\big) = \sigma\big\{X_s^{-1}(B) \mid s \in T, s \leq t, B \in \mathbb{B}\big\}, \quad t \in T.$$

Ist $(X_t)_{t \in T}$ ein Martingal (bzw. Supermartingal, bzw. Submartingal) bzgl. seiner natürlichen Filtration, so entfällt die Angabe der Filtration.

(ii) Im Fall $T = \mathbb{N}_0$ reicht es in der Definition eines Martingals, statt $E(X_t | \mathcal{A}_s) = X_s$ P-f.s. für alle $s, t \in T, s \leq t$ zu zeigen, nachzuweisen, dass $E(X_{n+1} | \mathcal{A}_n) = X_n$ P-f.s. für alle $n \in \mathbb{N}_0$ gilt. Analoges gilt für Super- und Submartingale. □

6.3 Beispiel (i) Sind X_i, $i \in \mathbb{N}$, unabhängige Zufallsvariablen mit $E(X_i) = 0$ (bzw. ≤ 0, bzw. ≥ 0) für alle $i \in \mathbb{N}$, so bilden die Partialsummen $S_n := \sum_{i=1}^n X_i$, $n \in \mathbb{N}_0$, ein Martingal (bzw. Supermartingal, bzw. Submartingal).

(ii) Ist Z eine endlich integrierbare Zufallsvariable, so definiert $X_t := E(Z | \mathcal{A}_t)$, $t \in T$, ein Martingal bzgl. $(\mathcal{A}_t)_{t \in T}$.

(iii) Ist $I \subset \mathbb{R}$ ein Intervall, $(X_t)_{t \in T}$ ein I-wertiges Martingal bzgl. $(\mathcal{A}_t)_{t \in T}$ und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion, so dass $E(|\varphi(X_t)|) < \infty$ für alle $t \in T$, so definiert $Y_t := \varphi(X_t)$, $t \in T$, wegen der Jensen'schen Ungleichung 5.10 (xiv) ein Submartingal bzgl. $(\mathcal{A}_t)_{t \in T}$. □

6.4 Lemma *Ein $\mathbb{R}$-wertiger Prozess $(X_t)_{t \in T}$ mit $E(|X_t|) < \infty$ für alle $t \in T$ ist genau dann ein Martingal (bzw. ein Supermartingal) bzgl. $(\mathcal{A}_t)_{t \in T}$, wenn*

- X_t *$\mathcal{A}_t, \mathbb{B}$-messbar ist für alle $t \in T$*
- $\int_A X_t dP \underset{(\leq)}{=} \int_A X_s dP \quad \forall s, t \in T, s \leq t, A \in \mathcal{A}_s$ □

6.5 Lemma *Seien $(X_t)_{t\in T}$ und $(Y_t)_{t\in T}$ Martingale (bzw. Supermartingale) bzgl. $(\mathcal{A}_t)_{t\in T}$. Dann gilt*

(i) $(-X_t)_{t\in T}$ *ist ein Martingal (bzw. ein Submartingal) bzgl.* $(\mathcal{A}_t)_{t\in T}$.

(ii) $(aX_t + Y_t)_{t\in T}$ *ist ein Martingal bzgl.* $(\mathcal{A}_t)_{t\in T}$ *für alle* $a \in \mathbb{R}$
(bzw. $(aX_t + Y_t)_{t\in T}$ *ist ein Supermartingal bzgl.* $(\mathcal{A}_t)_{t\in T}$ *für alle* $a \geq 0$*).*

(iii) $E(X_t) \underset{(\leq)}{=} E(X_s) \quad \forall s, t \in T, s \leq t.$ □

Es sollen nun zwei zentrale Resultate der Martingaltheorie für zeitdiskrete Martingale nachgewiesen werden. Angewendet auf Glücksspiele besagt das erste Resultat, dass es nicht möglich ist, durch geeignete Spielstrategien zu erreichen, dass man bei einem fairen Spiel im Mittel einen positiven Ertrag erzielt, wenn man nur einen endlichen Einsatz und nur endlich viel Zeit zur Verfügung hat. Dabei sind nur solche Strategien zulässig, die nicht in die Zukunft schauen. Insbesondere dürfen Zeitpunkte, zu denen das Spiel begonnen bzw. abgebrochen wird, im Sinne der folgenden Definition nur von der Vergangenheit abhängen und nicht davon, wie der zukünftige (noch unbekannte) Spielverlauf aussieht.

6.6 Satz und Definition *Sei $(\mathcal{A}_t)_{t\in T}$ eine Filtration auf $(\Omega, \mathcal{A})$. Eine Abbildung $S : \Omega \to T \cup \{\infty\}$ heißt* ***Stoppzeit*** *bzgl. $(\mathcal{A}_t)_{t\in T}$, falls*

$$\{S \leq t\} = \{\omega \in \Omega | S(\omega) \leq t\} \in \mathcal{A}_t \quad \forall t \in T.$$

Dann ist

$$\mathcal{A}_S := \{A \in \mathcal{A} \mid A \cap \{S \leq t\} \in \mathcal{A}_t \quad \forall t \in T\}$$

eine σ-Algebra, die sog. ***σ-Algebra der S-Vergangenheit*** *oder* ***σ-Algebra der Ereignisse bis S****.* □

6.7 Beispiel Ist $(X_n)_{n\in\mathbb{N}_0}$ ein an $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ adaptierter Prozess und $B \in \mathbb{B}$ beliebig, so definiert die sog. **Ersteintrittszeit** von $(X_n)_{n\in\mathbb{N}_0}$ in B

$$\begin{aligned} S_B : \quad & \Omega \to \mathbb{N}_0 \cup \{\infty\} \\ S_B(\omega) := & \begin{cases} \infty, & \text{falls } X_n(\omega) \notin B \text{ für alle } n \in \mathbb{N}_0 \\ \min\{n \in \mathbb{N}_0 \mid X_n(\omega) \in B\}, & \text{sonst} \end{cases} \end{aligned}$$

eine Stoppzeit bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$. □

6.8 Bemerkung (i) S ist genau dann eine Stoppzeit bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$, wenn $\{S = n\} \in \mathcal{A}_n$ für alle $n \in \mathbb{N}_0$.

(ii) Ist S eine Stoppzeit bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ und ist $(X_n)_{n\in\mathbb{N}_0}$ ein an $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ adaptierter Prozess, so ist die Zufallsvariable

$$X_S : \Omega \to \mathbb{R}, \; X_S(\omega) := X_{S(\omega)}(\omega)$$

$\mathcal{A}_S, \mathbb{B}$-messbar. □

6.9 Satz (optional sampling theorem) *Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal (bzw. Supermartingal) bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ und $0 \le T_1 \le T_2 \le \cdots \le T_k \le L$ eine monotone Folge durch $L \in \mathbb{N}_0$ beschränkter Stoppzeiten, so ist $(X_{T_i})_{1\le i\le k}$ ein Martingal (bzw. Supermartingal) bzgl. $(\mathcal{A}_{T_i})_{1\le i\le k}$.* □

6.10 Korollar (optional stopping theorem/Stoppsatz) *Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal (bzw. Supermartingal) bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ und T eine Stoppzeit bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$, so ist der in T gestoppte Prozess*

$$X_n^T := X_{\min(T,n)},\ n \in \mathbb{N}_0,$$

ein Martingal (bzw. Supermartingal) bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$. □

6.12 Korollar *Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ und $0 \le T \le L$ eine beschränkte Stoppzeit bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$, so gilt*

$$E(X_L) \le E(X_T) \le E(X_0) \tag{6.1}$$

und

$$E(|X_T|) \le E(X_0) + 2E(X_L^-) \tag{6.2}$$

Als zweites wichtiges Resultat wollen wir zeigen, dass unter geeigneten Integrabilitätsbedingungen *jedes* Martingal $(X_n)_{n\in\mathbb{N}_0}$ von der in Beispiel 6.3(ii) genannten Form ist und X_n für $n \to \infty$ f.s. gegen $X_\infty := E(Z|\mathcal{A}_\infty)$ konvergiert, wobei

$$\mathcal{A}_\infty := \sigma\Big(\bigcup_{n\in\mathbb{N}_0} \mathcal{A}_n\Big).$$

Dazu definieren wir zu einem stochastischen Prozess $X = (X_j)_{0\le j\le k}$ und einem beliebigen Intervall $[a,b]$ mit $a < b$ die sog. *Anzahl der aufsteigenden Überquerungen* von $[a,b]$ als

$$U_{[a,b]}(X,k) := \begin{cases} \max\big\{i \in \{1,\dots,k\} \mid T_i \le k\big\}, & \text{falls } T_1 \le k,\\ 0 & \text{sonst,} \end{cases}$$

wobei die Stoppzeiten $T_1,\dots,T_k$ wie folgt rekursiv definiert werden:

$$\begin{aligned} T_0 &:= 0\\ S_i &:= \min\{j \ge T_{i-1} \mid X_j \le a\}\\ T_i &:= \min\{j \ge S_i \mid X_j \ge b\} \end{aligned}$$

Die Idee des nachfolgenden Beweises ist es, dass ein Prozess $(X_n)_{n\in\mathbb{N}_0}$ genau dann in $\mathbb{R}$ konvergiert, wenn die Anzahl der aufsteigenden Überquerungen über jedes beliebige Intervall stets endlich ist. Um dies ausnutzen zu können, wird zunächst eine obere Schranke für den Erwartungswert der Anzahl der aufsteigenden Überquerungen hergeleitet.

6.13 Satz (Ungleichung von Doob) *Ist $(X_t)_{0\le t\le k}$ ein Supermartingal, so gilt*

$$E(U_{[a,b]}(X,k)) \le \frac{E((X_k-a)^-)}{b-a} \quad \forall -\infty < a < b < \infty.$$

□

6.14 Satz (Supermartingalkonvergenzsatz) *Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$.*

- *Gilt* $\sup_{n\in\mathbb{N}} E(X_n^-) < \infty$, *so konvergiert X_n P-f.s. gegen eine endlich integrierbare $\mathcal{A}_\infty, \mathbb{B}$-messbare Zufallsvariable X_∞.*
- *Ist $X_n \geq 0$ P-f.s. für alle $n \in \mathbb{N}_0$, oder sind die X_n, $n \in \mathbb{N}_0$, gleichgradig integrierbar, d.h. gilt*

$$\lim_{a\to\infty} \sup_{n\in\mathbb{N}} E(|X_n| 1_{\{|X_n|>a\}}) = 0,$$

so ist $(X_n)_{n\in\mathbb{N}_0\cup\{\infty\}}$ ein Supermartingal, d.h.

$$E(X_\infty|\mathcal{A}_n) \leq X_n \quad P\text{-f.s.} \qquad \forall n \in \mathbb{N}_0.$$

□

6.15 Satz *Ist $(X_n)_{n\in\mathbb{N}}$ eine gleichgradig integrierbare Folge von Zufallsvariablen, die P-stochastisch gegen eine Zufallsvariable X konvergiert, so konvergiert sie auch in L_1 gegen X, d.h.* $\lim_{n\to\infty} E(|X_n - X|) = 0$. □

6.16 Korollar *Ist $(X_n)_{n\in\mathbb{N}_0}$ ein gleichgradig integrierbares Martingal bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$, so konvergiert X_n P-f.s. und in L_1 gegen eine Zufallsvariable $X_\infty \in L_1$ und $(X_n)_{n\in\mathbb{N}_0\cup\{\infty\}}$ ist ein Martingal bzgl. $(\mathcal{A}_n)_{n\in\mathbb{N}_0\cup\{\infty\}}$.* □

6.17 Bemerkung Die gleichgradige Integrierbarkeit in Korollar 6.16 ist auch eine notwendige Voraussetzung für die Konvergenz. □

Der (Super-)Martingalkonvergenzsatz ist oft auch in solchen Situationen ein nützliches Hilfsmittel, die auf den ersten Blick wenig mit der Martingaltheorie zu tun haben.

6.18 Satz (0-1-Gesetz von Kolmogorov) *Sind X_n, $n \in \mathbb{N}$, unabhängige Zufallsvariablen und $\mathcal{T} := \bigcap_{n\in\mathbb{N}} \mathcal{T}_n$ die zugehörige* terminale σ-Algebra, *wobei $\mathcal{T}_n := \sigma((X_k)_{k\geq n})$, so gilt $P(A) \in \{0,1\}$ für alle $A \in \mathcal{T}$.* □

6.19 Beispiel In der Situation von Satz 6.18 gilt $P\{\bar{X}_n := \sum_{i=1}^n X_i/n \text{ konvergiert}\} \in \{0,1\}$. Mit Methoden der Martingaltheorie kann man ferner zeigen, dass sogar

$$P\{\bar{X}_n \to E(X_1)\} = 1$$

gilt, wenn die Zufallsvariablen identisch verteilt sind mit $E(|X_1|) < \infty$. Dies ist das sog. *Starke Gesetz der Großen Zahlen.* □