Maßtheoretische Konzepte der Stochastik

Sommersemester 2016

Mathias Trabs*
Universität Hamburg

17. November 2016

Inhaltsverzeichnis

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

1 Wozu brauchen wir Maßtheorie? Ein Beispiel.[1]

Betrachten wir $n \in \mathbb{N}$ unabhängige Würfe einer fairen Münze, also den endlichen Wahrscheinlichkeitsraum $\Omega_n := \{0,1\}^n$ (mit der Interpretation, dass 0 "Kopf" entspricht und 1 "Zahl") versehen mit der Gleichverteilung $P_n(\{\omega\}) = 2^{-n}$ für jedes $\omega \in \Omega_n$. Der Anteil der Würfe mit dem Ergebnis "Zahl", ist dann gegeben durch

$$S_n(\omega) := \frac{1}{n}\sum_{i=1}^{n} \omega_i \quad \text{für alle} \quad \omega = (\omega_i)_{i=1,\dots,n} \in \Omega_n.$$

Ist die Münze fair, so sollte $S_n(\omega)$ etwa $\frac{1}{2}$ sein. Gleichheit tritt allerdings nur mit verschwindender Wahrscheinlichkeit auf, da für große n Stirlings Approximation ($n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$ für $n \to \infty$) ergibt, dass

$$\begin{aligned} P_{2n}\big(S_{2n} = \frac{1}{2}\big) =& 2^{-2n} \#\big\{\omega \in \Omega_{2n} : \omega \text{ enthält genau } n \text{ Einsen}\big\} \\ =& 2^{-2n}\binom{2n}{n} \approx \frac{1}{\sqrt{\pi n}}. \end{aligned}$$

Tatsächlich liefert uns das *schwache Gesetz der großen Zahlen*

$$\lim_{n\to\infty} P_n\big(\big|S_n - \frac{1}{2}\big| < \varepsilon\big) = 1 \qquad \text{für alle } \varepsilon > 0.$$

Intuitiv würde man sogar folgende stärkere Form dieses Konvergenzresultates erwarten: Bezeichne

$$\Omega := \{0,1\}^{\mathbb{N}} := \{\omega = (\omega_i)_{i\in\mathbb{N}} : \omega_i \in \{0,1\}, i \in \mathbb{N}\}$$

den Ergebnisraum eines unendlich langen Münzwurfes.

Bemerkung 1.1. $\Omega = \{0,1\}^{\mathbb{N}}$ ist überabzählbar (Übung ☑).

Mit der Notation $S_n(\omega) := \frac{1}{n}\sum_{i=1}^n \omega_i, \omega \in \Omega$, entspricht das Gesetz der großen Zahlen

$$\lim_{n\to\infty} S_n(\omega) = \frac{1}{2}.$$

Für alle $\omega \in \Omega$ gilt dies offensichtlich nicht (z.B. $\omega = (0,0,0\dots)$). Wir können also höchstens fragen, ob dieser Grenzwert für fast alle Wurffolgen $\omega \in \Omega$ gilt. Untersuchen wir also das Ereignis

$$\big\{\omega \in \Omega : \lim_{n\to\infty} S_n(\omega) = \frac{1}{2}\big\}.$$

Problem: Der Grenzwert hängt von dem asymptotischen Verhalten der Folge ω ab. Bisher können wir aber nur die Wahrscheinlichkeiten von Ereignissen, die von endlichen vielen Münzwürfen abhängen, berechnen, genauer

$$P(\{\omega : (\omega_1, \dots, \omega_n) \in A\}) = 2^{-n}\#A \quad \text{für alle } A \subseteq \Omega_n, n \in \mathbb{N}.$$

Die Frage ist also, wie wir Wahrscheinlichkeiten $P(A), A \in \mathcal{F}$, für zumindest ein Mengensystem $\mathcal{F} \subseteq \mathscr{P}(\Omega)$ von Ereignissen auf Ω definieren. Als Mindestanforderung sollte gelten

A1. *$\mathcal{F}$ enthält alle Ereignisse, die nur von endlichen vielen Münzwürfen abhängen, d.h. für alle $n \in \mathbb{N}$ und $A \subseteq \Omega_n$ gilt $\{\omega : (\omega_1, \dots, \omega_n) \in A\} \in \mathcal{F}$. Auf diesen Mengen ist P definiert durch*

$$P(\{\omega : (\omega_1, \dots, \omega_n) \in A\}) = P_n(A), \quad A \subseteq \Omega_n.$$

[1]Dieser Abschnitt folgt Kapitel 1 aus Breiman, L. (1992). Probability. Classics in Applied Mathematics. Siam, Philadelphia.

Versuchen wir die Wahrscheinlichkeit von $A := \{\overline{\lim}_{n\to\infty} \big|S_n(\omega) - \frac{1}{2}\big| > 0\}$ abzuschätzen. Bemerke

$$\lim_{n\to\infty} S_n(\omega) = \lim_{m\to\infty} S_{m^2}(\omega).$$

Für $\varepsilon > 0$ und $1 < m_0 < m_1$ definieren wir

$$A_\varepsilon := \big\{\omega : \overline{\lim}_{m\to\infty} \big|S_{m^2}(\omega) - \frac{1}{2}\big| > \varepsilon\big\},$$

$$A_{m_0,m_1} := \bigcup_{m=m_0}^{m_1} \big\{\omega : \big|S_{m^2}(\omega) - \frac{1}{2}\big| > \varepsilon\big\}.$$

Da A_{m_0,m_1} nur von den ersten m_1^2 Würfen abhängt, folgt aus (A1) und der Subadditivität von $P_{m_1^2}$

$$P(A_{m_0,m_1}) \leqslant \sum_{m=m_0}^{m_1} P_{m^2}\big(|S_{m^2} - \frac{1}{2}| > \varepsilon\big).$$

Aus der Chebyshev-Ungleichung folgt

$$P_{m^2}\big(\big|S_{m^2} - \frac{1}{2}\big| > \varepsilon\big) \leqslant \frac{\mathrm{Var}(S_{m^2})}{\varepsilon^2} = \frac{1}{4\varepsilon^2 m^2}$$

und damit $P(A_{m_0,m_1}) \leqslant \frac{1}{4\varepsilon^2}\sum_{m=m_0}^{m_1} \frac{1}{m^2}$. Für festes m_0 ist $(A_{m_0,m_1})_{m_1 \geqslant m_0}$ eine monoton wachsende Folge mit dem Grenzwert

$$A_{m_0} := \bigcup_{m_1 \geqslant m_0} A_{m_0,m_1} = \bigcup_{m \geqslant m_0} \big\{\omega : \big|S_{m^2}(\omega) - \frac{1}{2}\big| > \varepsilon\big\}.$$

Wir benötigen folgende Monotonieannahme.

A2. *Ist $(A_n)_{n\geqslant 1} \subseteq \mathcal{F}$ eine monotone Folge, d.h. $A_n \subseteq A_{n+1}$, dann gilt $\bigcup_{n\geqslant 1} A_n \in \mathcal{F}$ und*

$$\lim_{n\to\infty} P(A_n) = P\Big(\bigcup_{n\geqslant 1} A_n\Big).$$

Aus (A2) folgt

$$P(A_{m_0}) = \lim_{m_1\to\infty} P(A_{m_0,m_1}) \leqslant \frac{1}{4\varepsilon^2} \sum_{m=m_0}^{\infty} \frac{1}{m^2} \leqslant \frac{1}{4\varepsilon^2(m_0-1)}.$$

Da $\overline{\lim}_{m\to\infty} \big|S_{m^2}(\omega) - \frac{1}{2}\big| > \varepsilon$ genau dann, wenn für alle m_0 ein $m \geqslant m_0$ existiert, so dass $\big|S_{m^2}(\omega) - \frac{1}{2}\big| > \varepsilon$, folgt $A_\varepsilon = \bigcap_{m_0 \geqslant 1} A_{m_0}$. Da $(A_{m_0})_{m_0 \geqslant 2}$ eine fallende Folge von Mengen ist, folgt wieder aus (A2), dass

$$P(A_\varepsilon) = \lim_{m_0\to\infty} P(A_{m_0}) \leqslant \lim_{m_0\to\infty} \frac{1}{4\varepsilon^2 m_0} = 0.$$

Schließlich stellen wir fest, dass $A = \big\{\omega : \overline{\lim}_{m\to\infty} \big|S_{m^2}(\omega) - \frac{1}{2}\big| > 0\big\} = \bigcup_{k\geqslant 1} A_{1/k}$ mit wachsender Folge $(A_{1/k})_k$, so dass erneut (A2) ergibt

$$P(A) = \lim_{k\to\infty} P(A_{1/k}) = 0.$$

Wir haben somit folgendes *starkes Gesetz der großen Zahlen* für unser Münzwurfexperiment bewiesen:

Satz 1.2. *Erfüllen $\mathcal{F}$ und P die Annahmen (A1) und (A2), so gilt $P(\omega : \lim_{n\to\infty} S_n(\omega) = \frac{1}{2}) = 1$.*

Es bleibt die Frage, ob eine Wahrscheinlichkeitsverteilung P existiert, welche die Annahmen (A1) und (A2) erfüllt und ob P durch (A1) und (A2) eindeutig beschrieben wird (was man intuitiv erwarten würde). Die Antwort auf diese Frage liefert Caratheodorys Fortsetzungssatz, dem zentralen Satz der Maßtheorie.

2 σ-Algebren und Maße

2.1 Definitionen und Eigenschaften

Wie bereits bekannt, modellieren wir die Gesamtheit aller möglichen Ausgänge eines Zufallsexperimentes durch eine nichtleere Menge E. Die Potenzmenge $\mathscr{P}(E)$ ist definiert als die Menge aller Teilmengen von E. Alle uns interessierenden Ereignisse fassen wir durch ein Mengensystem $\mathcal{A}$ aus E zusammen.

Definition 2.1. Ein nichtleeres Mengensystem $\mathcal{A} \subseteq \mathscr{P}(E)$ heißt

(i) *Ring*, falls für alle $A, B \in \mathcal{A}$ gilt auch $A \cup B \in \mathcal{A}$ und $A \setminus B \in \mathcal{A}$.

(ii) *Algebra*, falls für alle $A, B \in \mathcal{A}$ gilt $A \cup B \in \mathcal{A}$ und $E \setminus A \in \mathcal{A}$.

(iii) *σ-Algebra,* falls $\mathcal{A}$ eine Algebra ist und für alle $(A_n)_{n\geqslant 1} \subseteq \mathcal{A}$ gilt $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$.

Satz 2.2. *Sei $\mathcal{A} \subseteq \mathscr{P}(E)$.*

(i) Ist $\mathcal{A}$ ein Ring mit $E \in \mathcal{A}$, so ist $\mathcal{A}$ auch eine Algebra.

(ii) Ist $\mathcal{A}$ eine Algebra, dann gilt für alle $A, B \in \mathcal{A}$:

$$\emptyset, E \in \mathcal{A}, \quad A \cap B \in \mathcal{A}, \quad A \setminus B \in \mathcal{A}.$$

Für $A_1, \dots, A_n \in \mathcal{A}$ gilt stets

$$\bigcup_{k=1}^{n} A_k \in \mathcal{A} \quad \textit{und} \quad \bigcap_{k=1}^{n} A_k \in \mathcal{A}.$$

(iii) Ist $\mathcal{A}$ eine σ-Algebra und $(A_n)_{n\geqslant 1} \subseteq \mathcal{A}$, so gilt

$$\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}.$$

(iv) Der Durchschnitt $\bigcap_{i\in I} \mathcal{A}_i$ beliebig vieler σ-Algebren $\mathcal{A}_i, i \in I$, ist wieder eine σ-Algebra.

Beweis. Übung ☑. □

Beispiel 2.3. Sei E eine beliebige nichtleere Menge.

(i) $\{\emptyset, E\}$ und $\mathscr{P}(E)$ sind σ-Algebren.

(ii) Sei $\mathcal{E} \subseteq \mathscr{P}(E)$ ein beliebiges Mengensystem. Dann ist

$$\sigma(\mathcal{E}) := \bigcap \big\{\mathcal{A} : \mathcal{A} \text{ ist } \sigma\text{-Algebra aus } E \text{ mit } \mathcal{E} \subseteq \mathcal{A}\big\} \tag{2.1}$$

die kleinste σ-Algebra, die $\mathcal{E}$ umfasst. $\sigma(\mathcal{E})$ heißt die *von $\mathcal{E}$ erzeugte σ-Algebra.* Es gilt (Übung ☑):

(a) Ist $\mathcal{A}$ eine σ-Algebra aus E, so ist $\mathcal{A} = \sigma(\mathcal{A})$.
(b) Sind $\mathcal{E}, \mathcal{E}'$ Mengensysteme aus E, so folgt aus $\mathcal{E} \subseteq \mathcal{E}'$, dass $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$ gilt.

(iii) Sei $\Omega_n = \{0,1\}^n$ bzw. $\Omega = \{0,1\}^{\mathbb{N}}$ der Ergebnisraum des (unendlichen) Münzwurfexperiments. Das Mengensystem

$$\begin{aligned} \mathcal{F} :&= \big\{A \subseteq \Omega : \exists n \in \mathbb{N}, B \subseteq \Omega_n, \text{ so dass } A = \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}\big\} \\ &= \{\pi_n^{-1}(B) : n \in \mathbb{N}, B \subseteq \Omega_n\} \end{aligned}$$

mit der Projektion auf die ersten n Koordinaten $\pi_n \colon \Omega \to \Omega_n$ ist eine Algebra, aber keine σ-Algebra.

(iv) Sei $f\colon X \to Y$ eine Abbildung und $\mathcal{A} \subseteq \mathscr{P}(Y)$. Dann definieren wir

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

Ist $\mathcal{A}$ eine σ-Algebra auf Y, so ist $f^{-1}(\mathcal{A})$ eine σ-Algebra auf X, welche die *durch f induzierte σ-Algebra* genannt wird. Für jedes Mengensystem $\mathcal{E} \subseteq \mathscr{P}(Y)$ gilt (Übung ☑)

$$\sigma(f^{-1}(\mathcal{E})) = f^{-1}(\sigma(\mathcal{E})).$$

(v) Für $n \geqslant 1$ bildet das folgende Mengensystem eine Algebra auf $\mathbb{R}^n$:

$$\mathcal{Q}_n := \Big\{ \bigcup_{k=1}^{K} (a_k, b_k] : (a_k, b_k] := \prod_{l=1}^{n} (a_{k,l}, b_{k,l}] \text{ mit } -\infty \leqslant a_{k,l} \leqslant b_{k,l} \leqslant \infty, k = 1, \dots, K, l = 1, \dots, n$$
$$\text{und } (a_k, b_k] \text{ sind paarweise disjunkt}, K \in \mathbb{N} \Big\},$$

wobei wir hier $(a, \infty] = (a, \infty)$ für $a \in \mathbb{R}$ setzen. Es ist aber keine σ-Algebra, denn $(n = 1)$

$$\bigcup_{k \geqslant 1} (0, 1 - \tfrac{1}{k}] = (0, 1) \notin \mathcal{A}.$$

(vi) Sei (E, d) ein metrischer (oder topologischer) Raum und bezeichne $\mathcal{O}(E)$ die Familie der offenen Teilmengen und $\mathcal{C}(E)$ die Familie der abgeschlossenen Teilmengen. Die von $\mathcal{O}(E)$ erzeugte σ-Algebra $\mathscr{B}_E := \sigma(\mathcal{O}(E))$ heißt *Borelsche σ-Algebra auf E*. Eine Teilmenge $A \in \mathscr{B}_E$ wird *Borelmenge des metrischen Raumes* (E, d) genannt. Es gilt (Übung ☑)

(a) $\mathscr{B}_E = \sigma(\mathcal{C}(E))$,

(b) $\mathscr{B}_{\mathbb{R}^n} = \sigma(\mathcal{Q}_n)$.

Definition 2.4. Ein Paar $(E, \mathcal{A})$, wobei E eine nichtleere Menge und $\mathcal{A} \subseteq \mathscr{P}(E)$ eine σ-Algebra ist, heißt *messbarer Raum*. Die Elemente von $\mathcal{A}$ heißen *messbare Mengen*.

Maße sind Abbildungen von einer σ-Algebra nach $[0, \infty]$, die intuitiv jeder Menge aus der σ-Algebra eine "Größe" zuordnen, bspw. Flächen oder Volumina. Es ist intuitiv, dass diese Abbildungen nichtnegativ und additiv sein sollten. In der Stochastik dienen Maße dazu, den Ereignissen Wahrscheinlichkeiten zuzuordnen, was den Wertebereich von Wahrscheinlichkeitsmaßen auf natürliche Weise auf $[0, 1]$ beschränkt.

Definition 2.5. Es sei $\mathcal{R}$ ein Ring auf $E \neq \emptyset$. Eine Abbildung $\mu\colon \mathcal{R} \to [0, \infty] := [0, \infty) \cup \{\infty\}$ heißt *Inhalt*, falls $\mu(\emptyset) = 0$ und falls μ *additiv* ist, d.h. für disjunkte Mengen $A, B \in \mathcal{R}$ gilt $\mu(A \cup B) = \mu(A) + \mu(B)$.

μ heißt *σ-additiv* oder *σ-Inhalt*, falls zusätzlich für jede abzählbare Folge $(A_k)_{k \geqslant 1} \subseteq \mathcal{R}$ paarweise disjunkter Mengen aus $\mathcal{R}$ mit $\bigcup_{k \geqslant 1} A_k \in \mathcal{R}$ gilt

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k).$$

Ein σ-additiver Inhalt $\mu\colon \mathcal{A} \to [0, \infty]$ auf einer σ-Algebra $\mathcal{A}$ heißt *Maß auf $\mathcal{A}$*.

Ein Tripel $(E, \mathcal{A}, \mu)$ aus einer nichtleeren Menge E, einer σ-Algebra $\mathcal{A}$ und einem Maß μ auf $\mathcal{A}$ heißt *Maßraum*. Die Mengen $A \in \mathcal{A}$ mit $\mu(A) = 0$ heißen *μ-Nullmengen*.

Satz 2.6 (Eigenschaften von Inhalten und Maßen)**.** *Für jeden Inhalt $\mu\colon \mathcal{A} \to [0, \infty]$ auf einer Algebra $\mathcal{A} \subseteq \mathscr{P}(E)$ gelten folgende Aussagen:*

(i) Für alle $A, B \in \mathcal{A}$ mit $A \subseteq B$ und $\mu(A) < \infty$ gilt $\mu(B \setminus A) = \mu(B) - \mu(A)$, insbesondere $\mu(A) \leqslant \mu(B)$ (Monotonie).

(ii) Für alle $A, B \in \mathcal{A}$ mit $\mu(A \cap B) < \infty$ gilt $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

(iii) Die beiden folgenden Eigenschaften sind äquivalent:

(a) Die additive Mengenfunktion μ ist σ-additiv.

(b) Für jede Folge $(A_n)_{n\geqslant 1} \subseteq \mathcal{A}$ mit $A_n \subseteq A_{n+1}, \mu(A_n) < \infty$ für $n \geqslant 1$ und $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$, gilt

$$\lim_{n\to\infty} \mu(A_n) = \mu\Big(\bigcup_{n\geqslant 1} A_n\Big) \quad \textit{(Stetigkeit von unten).}$$

(iv) Die beiden folgenden Eigenschaften sind äquivalent:

(a) Für jede Folge $(A_n)_{n\geqslant 1} \subseteq \mathcal{A}$ mit $A_n \supseteq A_{n+1}, n \geqslant 1$, und $\mu(A_1) < \infty$ und $\bigcap_{n\geqslant 1} A_n \in \mathcal{A}$, gilt

$$\lim_{n\to\infty} \mu(A_n) = \mu\Big(\bigcap_{n\geqslant 1} A_n\Big) \quad \textit{(Stetigkeit von oben).}$$

(b) Für jede Folge $(A_n)_{n\geqslant 1} \subseteq \mathcal{A}$ mit $A_n \supseteq A_{n+1}, n \geqslant 1$, und $\mu(A_1) < \infty$ und $\bigcap_{n\geqslant 1} A_n = \emptyset$ gilt $\lim_{n\to\infty} \mu(A_n) = 0$ (Stetigkeit in der leeren Menge).

(v) Die Eigenschaften in (iii) implizieren (iv). Im Fall $\mu(E) < \infty$ sind die Eigenschaften in (iii) und (iv) äquivalent.

(vi) Ist μ σ-additiv, so gilt für jede Folge $(A_n) \subseteq \mathcal{A}$ mit $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$ die Ungleichung

$$\mu\Big(\bigcup_{n\geqslant 1} A_n\Big) \leqslant \sum_{n=1}^{\infty} \mu(A_n) \quad \textit{(σ-Subadditivität).}$$

Beweis. (i) Wegen $B = A \cup (B \setminus A)$ gilt $\mu(B) = \mu(A) + \mu(B \setminus A)$.

(ii) Folgt aus $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$.

(iii) (a)⇒(b) Definiere $B_1 := A_1$ und $B_n := A_n \setminus A_{n-1}$ für $n \geqslant 1$. Dann gilt $\bigcup_{k=1}^n B_k = A_n$ und

$$\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \mu\big(\bigcup_{k=1}^{n} B_k\big) = \sum_{k=1}^{\infty} \mu(B_k) = \mu\Big(\bigcup_{k\geqslant 1} B_k\Big) = \mu\Big(\bigcup_{k\geqslant 1} A_k\Big).$$

(b) ⇒ (a) Für paarweise disjunkte $(B_n)_{n\geqslant 1} \subseteq \mathcal{A}$ folgt die σ-Additivität aus obiger Zeile mit $A_n := \bigcup_{k=1}^n B_k$.

(iv) Übung ☑.

(v) Wir zeigen *(iii)(a) ⇒ (iv)(b)*: Sei $(A_n)_{n\geqslant 1} \subseteq \mathcal{A}$ fallend mit $\mu(A_1) < \infty$ und $\bigcap_{n\geqslant 1} A_n = \emptyset$. Definiere die disjunkten Mengen $B_n := A_n \setminus A_{n+1}$. Es gilt $A_n = \bigcup_{k=n}^\infty B_k$ und $\sum_{k\geqslant 1} \mu(B_k) = \mu(A_1) < \infty$. Aus $\mu(A_n) = \sum_{k=n}^\infty \mu(B_k)$ folgt die Behauptung.

Falls $\mu(E) < \infty$, so gilt *(iv)(b) ⇒ (iii)(a):* Sei $(A_n) \subseteq \mathcal{A}$ eine Folge disjunkter Mengen mit $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$. Die Folge $B_n := \bigcup_{k=n}^\infty A_k \in \mathcal{A}$ ist monoton fallend mit $\bigcap_{n\geqslant 1} B_n = \emptyset$. Dann gilt

$$\mu\big(\bigcup_{k\geqslant 1} A_k\big) = \lim_{n\to\infty} \Big(\mu\big(\bigcup_{k=1}^{n} A_k\big) + \mu(B_{n+1})\Big) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k) + \lim_{n\to\infty} \mu(B_{n+1}).$$

Da letzterer Grenzwert gegen 0 konvergiert, folgt die σ-Additivität.

(vi) Für die Folge disjunkter Mengen $B_1 := A_1$ und $B_n := A_n \setminus \bigcup_{k=1}^{n-1} A_k$ gilt $\bigcup_{n\geqslant 1} A_n = \bigcup_{n\geqslant 1} B_n$ und es folgt aus der σ-Additivität und (i)

$$\mu\big(\bigcup_{n\geqslant 1} A_n\big) = \mu\big(\bigcup_{n\geqslant 1} B_n\big) = \sum_{n\geqslant 1} \mu(B_n) \leqslant \sum_{n\geqslant 1} \mu(A_n).$$

□

Definition 2.7. Ein Inhalt $\mu\colon \mathcal{A} \to [0,\infty]$ auf einer Algebra $\mathcal{A}$ über E heißt *σ-endlich*, falls eine Folge $(E_n)_{n\geqslant 1} \subseteq \mathcal{A}$ existiert mit $\bigcup_{n\geqslant 1} E_n = E$ und $\mu(E_n) < \infty$ für alle $n \geqslant 1$. Gilt $\mu(E) < \infty$, dann heißt μ *endlich*. Gilt $\mu(E) = 1$ für ein Maß μ auf einer σ-Algebra $\mathcal{A}$, so nennt man μ *Wahrscheinlichkeitsmaß* und $(E, \mathcal{A}, \mu)$ *Wahrscheinlichkeitsraum.*

Definition 2.8. Sei $(E, \mathcal{A}, \mu)$ ein Maßraum. Das Maß $\mu\colon \mathcal{A} \to [0,\infty]$ heißt *vollständig*, falls jede Teilmenge einer Nullmenge selbst messbar ist, d.h. ist $A \in \mathcal{A}$ eine Menge vom Maß $\mu(A) = 0$ und $B \subseteq A$, so gilt $B \in \mathcal{A}$.

Beispiel 2.9.

(i) Das *Zählmaß* zählt die Anzahl der Elemente einer Menge. Wir definieren $\mu\colon \mathscr{P}(E) \to [0,\infty]$ durch

$$\mu(A) := \begin{cases} \#A, & \text{falls } A \text{ endlich,} \\ +\infty, & \text{sonst.} \end{cases}$$

μ ist genau dann σ-endlich, wenn E abzählbar ist. μ ist vollständig (da $\mathcal{A} = \mathscr{P}(E)$). Ist E endlich, dann ist $P(A) := \frac{\#A}{\#E}, A \subseteq E$, ein Wahrscheinlichkeitsmaß, die *Gleichverteilung* (insbesondere ist P_n auf Ω_n aus Kaptiel 1 ein Wahrscheinlichkeitsmaß).

(ii) Sei $x \in E$ ein fixierter Punkt. Das *Dirac-Maß* entscheidet, ob x in einer Menge liegt oder nicht. Wir definieren $\delta_x\colon \mathscr{P}(E) \to [0,\infty]$ durch

$$\delta_x(A) = \begin{cases} 1, & \text{falls } x \in A, \\ 0, & \text{falls } x \notin A. \end{cases}$$

δ_x ist ein vollständiges Wahrscheinlichkeitsmaß.

(iii) Sei $(x_n)_{n=1}^\infty \subseteq \mathbb{R}^d$ eine Folge ohne Häufungspunkte und $f\colon \mathbb{N} \to [0,\infty)$ eine Gewichtsfunktion. Wir definieren $\mu_f\colon \mathscr{P}(\mathbb{R}^d) \to [0,\infty]$ durch

$$\mu_f(A) := \sum_{x_n \in A} f(n).$$

μ_f ist ein vollständiges, σ-endliches Maß. Die σ-Endlichkeit folgt, da man $\mathbb{R}^d$ durch kompakte Kugeln ausschöpfen kann, die jeweils nur endlich viele Folgeglieder x_n enthalten können, da $(x_n)_n$ keinen Häufungspunkt hat.

2.2 Konstruktion von Maßen

Da man im Allgemeinen nicht einmal alle Elemente der σ-Algebra eines messbaren Raumes $(E, \mathcal{A})$ kennt (z.B. Borelmengen), erscheint es hoffnungslos $\mu(A)$ für alle $A \in \mathcal{A}$ angeben zu können. Wie gelangt man dann aber zum konkreten Maß? Wir verwenden die Konstruktion von Caratheodory.

Einen σ-Inhalt μ auf einem Ring $\mathcal{R}$ über E nennen wir *Prämaß*. Unser Ziel ist nun, den zugehörigen *Prämaßraum* $(E, \mathcal{R}, \mu)$ zu einem (vollständigen) Maßraum fortzusetzen. Zunächst ein Reduktionsschritt.

Definition 2.10. Eine Abbildung $\mu^*\colon \mathscr{P}(E) \to [0,\infty]$ heißt *äußeres Maß auf E*, falls

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

(ii) aus $A \subseteq B$ für $A, B \subseteq E$ folgt $\mu^*(A) \leqslant \mu^*(B)$ (*Monotonie*),

(iii) für $A_n \subseteq E, n \geqslant 1$, gilt $\mu^*(\bigcup_{n=1}^\infty A_n) \leqslant \sum_{n=1}^\infty \mu^*(A_n)$ (*σ-Subadditivität*).

Definition 2.11. Sei μ^* ein äußeres Maß auf $\mathscr{P}(E)$. Eine Teilmenge $A \subseteq E$ heißt *μ^*-messbar*, falls für alle Teilmengen $B \subseteq E$ gilt

$$\mu^*(B) = \mu^*(A \cap B) + \mu^*(A^c \cap B). \tag{2.2}$$

Die Menge aller μ^*-messbaren Teilmengen von E bezeichnen wir mit $\mathcal{A}_{\mu^*}$.

μ^*-messbare Mengen A zerlegen also jedes $B \in \mathscr{P}(E)$ in zwei disjunkte Teilmengen, auf denen μ^* additiv ist. Aufgrund der Subadditivität von μ^* ist nur $\geqslant$ für die Messbarkeit zu zeigen.

Satz 2.12 (Caratheodory, 1914)**.** *Sei $\mu^*\colon \mathscr{P}(E) \to [0,\infty]$ ein äußeres Maß, dann ist $(E, \mathcal{A}_{\mu^*}, \mu^*|_{\mathcal{A}_{\mu^*}})$ ein vollständiger Maßraum.*

Beweis. Schritt 1: $\mathcal{A}_{\mu^*}$ ist eine Algebra.

Offenbar ist $E \in \mathcal{A}_{\mu^*}$ und da (2.2) symmetrisch in A und A^c ist, folgt aus $A \in \mathcal{A}_{\mu^*}$ auch $A^c \in \mathcal{A}_{\mu^*}$. Sind $A, B \in \mathcal{A}_{\mu^*}$, so gilt für alle $Q \subseteq E$:

$$\begin{aligned}
\mu^*(Q) \geqslant & \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \\
\geqslant & \mu^*(Q \cap A) + \mu^*(Q \cap A^c \cap B) + \mu^*(Q \cap A^c \cap B^c) \\
\geqslant & \mu^*\big((Q \cap A) \cup (Q \cap A^c \cap B)\big) + \mu^*(Q \cap (A \cup B)^c) \\
= & \mu^*(Q \cap (A \cup B)) + \mu^*(Q \cap (A \cup B)^c).
\end{aligned}$$

Somit ist $A \cup B \in \mathcal{A}_{\mu^*}$.

Schritt 2: Ist $(A_n)_{n\geqslant 1}$ eine Folge disjunkter Mengen aus $\mathcal{A}_{\mu^*}$, so ist $A := \bigcup_{n\geqslant 1} A_n \in \mathcal{A}_{\mu^*}$ und

$$\mu^*(A) = \sum_{n\geqslant 1} \mu^*(A_n). \tag{2.3}$$

Für disjunkte Mengen $M, N \in \mathcal{A}_{\mu^*}$ folgt aus (2.2) mit $B = Q \cap (M \cup N)$, dass $\mu^*(Q \cap (M \cup N)) = \mu^*(Q \cap M) + \mu^*(Q \cap N)$. Induktiv folgt für jedes $n \in \mathbb{N}$

$$\mu^*(Q \cap \bigcup_{j=1}^n A_j) = \sum_{j=1}^n \mu^*(Q \cap A_j).$$

Nach Schritt 1 ist $\bigcup_{j=1}^n A_j \in \mathcal{A}_{\mu^*}$ und wir erhalten für alle $Q \subseteq E$ und alle $n \in \mathbb{N}$:

$$\mu^*(Q) = \mu^*\Big(Q \cap \bigcup_{j=1}^n A_j\Big) + \mu^*\Big(Q \cap \Big(\bigcup_{j=1}^n A_j\Big)^c\Big) \geqslant \sum_{j=1}^n \mu^*(Q \cap A_j) + \mu^*(Q \cap A^c).$$

Also liefert die σ-Subadditivität

$$\mu^*(Q) \geqslant \sum_{j=1}^\infty \mu^*(Q \cap A_j) + \mu^*(Q \cap A^c) \geqslant \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \geqslant \mu^*(Q).$$

Diese Ungleichungskette zeigt $A \in \mathcal{A}_{\mu^*}$ sowie (2.3) im Fall $Q = A$.

Schritt 3: μ^* ist vollständig.

Ist $N \in \mathcal{A}_{\mu^*}$ eine μ^*-Nullmenge und $M \subseteq N$, dann gilt $\mu^*(M) = 0$ und für jedes $Q \subseteq E$ impliziert die Monotonie:

$$\mu^*(Q) \geqslant \mu^*(Q \cap M) + \mu^*(Q \cap M^c) \geqslant \mu^*(Q \cap N) + \mu^*(Q \cap N^c) = \mu^*(Q). \qquad \square$$

Satz 2.13 (Fortsetzungssatz)**.** *Sei $\mu\colon \mathcal{R} \to [0,\infty]$ ein Inhalt auf einem Ring $\mathcal{R}$ über E. Wir definieren $\mu^*\colon \mathscr{P}(E) \to [0,\infty]$ durch*

$$\mu^*(B) := \begin{cases} \inf\Big\{ \sum_{n=1}^\infty \mu(A_n) : B \subseteq \bigcup_{n=1}^\infty A_n \textit{ mit } A_n \in \mathcal{R} \textit{ für alle } n \geqslant 1\Big\} \\ +\infty, \textit{falls } B \textit{nicht in einer abzählbaren Vereinigung von Mengen aus } \mathcal{R} \textit{liegt.} \end{cases}$$

Dann gilt:

(i) μ^* ist ein äußeres Maß auf E (das von μ erzeugte äußere Maß*).*

(ii) $\mathcal{R} \subseteq \mathcal{A}_{\mu^*}$.

(iii) Ist μ ein σ-Inhalt, so gilt $\mu = \mu^*|_{\mathcal{R}}$.

Beweis. (i) Da $\emptyset \in \mathcal{R}$ und $\mu \geqslant 0$, gilt $\mu^*(\emptyset) = 0$. Ist $B_1 \subseteq B_2$, so ist jede Überdeckung von B_2 auch eine Überdeckung von B_1, woraus die Montonie $\mu^*(B_1) \leqslant \mu^*(B_2)$ folgt.

Sei nun $(A_n)_{n\geqslant 1} \subseteq \mathscr{P}(E)$. Ist $\mu^*(A_n) = \infty$ für ein n, so ist die Ungleichung

$$\mu^*(\bigcup_{n=1}^{\infty} A_n) \leqslant \sum_{n=1}^{\infty} \mu^*(A_n)$$

trivial. Es sei nun $\mu^*(A_n) < \infty$ für alle $n \geqslant 1$ und sei $\varepsilon > 0$. Für jedes $n \in \mathbb{N}$ gibt es eine Folge $(B_{nk})_{k\geqslant 1} \subseteq \mathcal{R}$, so dass $A_n \subseteq \bigcup_{k\geqslant 1} B_{nk}$ und

$$\sum_{k\geqslant 1} \mu(B_{nk}) \leqslant \mu^*(A_n) + \varepsilon \cdot 2^{-n}.$$

Nun ist $(B_{nk})_{(n,k)\in\mathbb{N}^2}$ eine abzählbare Familie von Mengen aus $\mathcal{R}$ mit $\bigcup_{n\geqslant 1} A_n \subseteq \bigcup_{n\geqslant 1} \bigcup_{k\geqslant 1} B_{nk}$ und es folgt

$$\mu^*\Big(\bigcup_{n\geqslant 1} A_n\Big) \leqslant \sum_{n\geqslant 1}\sum_{k\geqslant 1} \mu^*(B_{nk}) \leqslant \sum_{n\geqslant 1} \big(\mu^*(A_n) + \varepsilon \cdot 2^{-n}\big) = \sum_{n\geqslant 1} \mu^*(A_n) + \varepsilon.$$

Da ε beliebig war, folgt die σ-Subadditivität und μ^* ist ein äußeres Maß.

(ii) Seien nun $A \in \mathcal{R}$ und $B \subseteq E$ mit $\mu^*(B) < \infty$. Wir wählen $(A_n) \subseteq \mathcal{R}$ mit $B \subseteq \bigcup_{n\geqslant 1} A_n$ (existiert da $\mu^*(B) < \infty$). Da μ ein Inhalt auf $\mathcal{R}$ ist und $\mu^*|_{\mathcal{R}} \leqslant \mu$, folgt

$$\begin{aligned}
\sum_{n\geqslant 1} \mu(A_n) &= \sum_{n\geqslant 1} \mu(A_n \cap A) + \sum_{n\geqslant 1} \mu(A_n \cap A^c) \\
&\geqslant \sum_{n\geqslant 1} \mu^*(A_n \cap A) + \sum_{n\geqslant 1} \mu^*(A_n \cap A^c) \\
&\geqslant \mu^*\Big(A \cap \bigcup_{n\geqslant 1} A_n\Big) + \mu^*\Big(A^c \cap \bigcup_{n\geqslant 1} A_n\Big) \\
&\geqslant \mu^*(A \cap B) + \mu^*(A^c \cap B).
\end{aligned}$$

Bilden wir das Infimum über alle Überdeckungen $(A_n)_{n\geqslant 1}$ von B, erhalten wir $\mu^*(B) \geqslant \mu^*(A \cap B) + \mu^*(A^c \cap B)$.

(iii) Sei nun μ ein σ-Inhalt und $B \in \mathcal{R}$. Nach Definition gilt $\mu^*(B) \leqslant \mu(B)$. Für jede Überdeckung $(A_n)_{n\geqslant 1} \subseteq \mathcal{R}$ von B gilt $B = \bigcup_{n\geqslant 1} B_n$ mit der Folge $(B_k)_{k\geqslant 1} \subseteq \mathcal{R}$ disjunkter Mengen $B_1 := B \cap A_1$ und $B_n := B \cap \big(A_n \setminus \bigcup_{k=1}^{n-1} A_k\big), n \geqslant 2$. Dann folgt

$$\mu(B) = \mu\Big(\bigcup_{n\geqslant 1} B_n\Big) = \sum_{n\geqslant 1} \mu(B_n) \leqslant \sum_{n\geqslant 1} \mu(A_n).$$

Bilden wir das Infimum über alle Überdeckungen, folgt $\mu(B) \leqslant \mu^*(B)$. □

Diese beiden Sätze zeigen, dass jeder Prämaßraum $(E, \mathcal{R}, \mu)$ zu einem vollständigen Maßraum $(E, \mathcal{A}_{\mu^*}, \mu^*|_{\mathcal{A}_{\mu^*}})$ fortgesetzt werden kann. Beachte, dass die kleinste Fortsetzung zu einem Maßraum gegeben ist durch $(E, \sigma(\mathcal{R}), \widetilde{\mu})$ mit $\widetilde{\mu} = \mu^*|_{\sigma(\mathcal{R})}$.

Bemerkung 2.14. Man kann zeigen, dass $\mathcal{A}_{\mu^*}$ die Vervollständigung von $\sigma(\mathcal{R})$ ist, d.h.

$$\mathcal{A}_{\mu^*} = \big\{A \cup N : A \in \sigma(\mathcal{R}), N \subseteq N_0 \text{ für ein } N_0 \in \sigma(\mathcal{R}) \text{ mit } \widetilde{\mu}(N_0) = 0\big\}.$$

Satz 2.15 (Eindeutigkeitssatz)**.** *Es seien μ und ν zwei Maße auf einer σ-Algebra $\mathcal{A} = \sigma(\mathcal{R})$ über E erzeugt von einem Ring $\mathcal{R}$ mit*

(i) $\mu|_{\mathcal{R}} = \nu|_{\mathcal{R}}$ und

(ii) es existiert eine wachsende Folge $(E_n)_{n\geqslant 1} \subseteq \mathcal{R}$ mit $\mu(E_n) = \nu(E_n) < \infty$ für alle $n \in \mathbb{N}$ und $\bigcup_n E_n = E$,

dann gilt $\mu = \nu$.

Beweis. Für jedes $A \in \mathcal{A} = \sigma(\mathcal{R})$ haben wir eine monotone wachsende Ausschöpfung $A = \bigcup_{n\geqslant 1} A \cap E_n$. Aufgrund der Stetigkeit von unten gilt

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap E_n) \quad \text{und} \quad \nu(A) = \lim_{n\to\infty} \nu(A \cap E_n).$$

Es genügt also folgendes zu zeigen: *Sei $A \in \mathcal{R}$ und $B \in \mathcal{A}$ eine Teilmenge von A, dann gilt $\mu(B) = \mu^*(B) = \nu(B)$ mit*

$$\mu^*(B) = \inf\Big\{ \sum_{n=1}^{\infty} \mu(C_n) : B \subseteq \bigcup_{n\geqslant 1} C_n, C_n \in \mathcal{R} \Big\}.$$

Für jede Überdeckung $(C_n) \subseteq \mathcal{R}$ von B gilt:

$$\nu(B) \leqslant \nu\Big(\bigcup_{n\geqslant 1} C_n \Big) \leqslant \sum_{n\geqslant 1} \nu(C_n) = \sum_{n\geqslant 1} \mu(C_n).$$

Bilden wir das Infimum über alle Überdeckungen, erhalten wir $\nu(B) \leqslant \mu^*(B)$. Analog ergibt sich $\nu(A \setminus B) \leqslant \mu^*(A \setminus B)$. Es folgt

$$\mu^*(A) = \nu(A) = \nu(B) + \nu(A \setminus B) \leqslant \mu^*(B) + \mu^*(A \setminus B) \leqslant \mu^*(A),$$

so dass überall Gleichheit gilt. Insbesondere folgt $\nu(B) = \mu^*(B)$. □

Bemerkung 2.16. Diesen Satz kann man auf einen durchschnittsstabilen Erzeuger der σ-Algebra statt eines Ringes abschwächen.

Korollar 2.17. *Sei μ ein σ-endlicher σ-Inhalt auf einem Ring $\mathcal{R}$. Dann existiert genau ein Maß $\widetilde{\mu}$ auf $\sigma(\mathcal{R})$, welches μ fortsetzt, d.h. für das $\widetilde{\mu}|_{\mathcal{R}} = \mu$ gilt. Dieses Maß ist σ-endlich.*

Beispiel 2.18.

(i) Im unendlichen Münzwurfexperiment $\Omega = \{0,1\}^{\mathbb{N}}$ haben wir bereits gesehen, dass $\mathcal{F} = \{\pi_n^{-1}(B) : n \in \mathbb{N}, B \subseteq \Omega_n\}$ mit $\Omega_n = \{0,1\}^n$ und den Projektionen $\pi_n \colon \Omega \to \Omega_n$ eine Algebra ist. Auf $\mathcal{F}$ hatten wir $P(\pi_n^{-1}(B)) := P_n(B) := 2^{-n}\#B$ für alle $n \in \mathbb{N}, B \subseteq \Omega_n$ definiert. Da P_n Wahrscheinlichkeitsmaße sind, ist P ein σ-Inhalt auf $\mathcal{F}$ und wegen $\Omega = \pi_1^{-1}(\{0,1\})$ mit $P(\pi_1^{-1}(\{0,1\})) = P_1(\{0,1\}) = 1$ ist P auch normiert (insbesondere σ-endlich). Folglich können wir P zu einem eindeutigen Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\sigma(\mathcal{F})$ fortsetzen.

(ii) Ein *Radon-Maß* auf $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ ist ein Maß μ mit $\mu(B) < \infty$ für alle beschränkten Mengen $B \in \mathscr{B}_{\mathbb{R}}$. Da der Ring

$$\mathcal{Q}_1 := \Big\{ \bigcup_{k=1}^{n} (a_k, b_k] : -\infty \leqslant a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n \leqslant \infty, n \in \mathbb{N} \Big\} \cup \{\emptyset\}$$

auf $\mathbb{R}$ die Borel-σ-Algebra $\mathscr{B}_{\mathbb{R}}$ erzeugt, wird jedes Radon-Maß durch die Werte $\mu((a,b]), -\infty < a \leqslant b < \infty$ eindeutig festgelegt.
Das Maß λ auf $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ mit $\lambda((a,b]) = b - a$ für alle $-\infty < a \leqslant b < \infty$ heißt *Lebesgue-Maß* auf $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Die Existenz kann man mit Hilfe von Satz 2.13 angewendet auf den Ring $\mathcal{Q}_1$ und das darauf definierte Prämaß $\mu(\bigcup_{k=1}^n (a_k, b_k]) = \sum_{k=1}^n (b_k - a_k)$ nachweisen. Die Vervollständigung von $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ bzgl. λ wird mit $(\mathbb{R}, \mathscr{L}_{\mathbb{R}})$ bezeichnet mit der σ-Algebra der *Lebesguemessbaren Mengen* $\mathscr{L}_{\mathbb{R}} := \mathcal{A}_{\mu^*}$.

(iii) Aus der mathematischen Stochastik ist bekannt, dass sich jedes Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ eindeutig durch seine *Verteilungsfunktion* $F\colon \mathbb{R} \to [0,1], F(x) = \mathbb{P}((-\infty, x])$ beschreiben lässt. Dies folgt auch aus (ii), da $\mathbb{P}((a,b]) = F(b) - F(a)$. Umgekehrt kann man ähnlich wie beim Lebesgue-Maß zeigen, dass zu jeder rechtsseitig stetigen, monoton steigenden Funktion $F\colon \mathbb{R} \to [0,1]$ 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 als Verteilungsfunktion hat.

In Verallgemeinerung des letzten Beispiels lässt sich zeigen

Satz 2.19. *Zu jeder rechtsseitig stetigen, monoton steigenden Funktion $G\colon \mathbb{R} \to \mathbb{R}$ existiert genau ein Radon-Maß (oder Lebesgue-Stieltjes-Maß) μ_G , so dass $\mu_G((a,b]) = G(b) - G(a)$ für alle $-\infty < a \leqslant b < \infty$. G heißt maßdefinierende Funktion zu μ_G. Für jedes Radonmaß μ auf $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ ist die maßdefinierende Funktion bis auf eine additive Konstante eindeutig bestimmt. Eine maßdefinierende Funktion ist gegeben durch $G(a) := \mu((0,a])$ für $a \geqslant 0$ und $G(a) = -\mu((a,0])$ für $a < 0$.*

Beweis. Übung ☑. □

2.3 Dynkinsysteme*

Definition 2.20. Sei $E \neq \emptyset$. Ein Mengensystem $\mathcal{D} \subseteq \mathscr{P}(E)$ heißt *Dynkinsystem* über E, falls:

(i) $E \in \mathcal{D}$,

(ii) für alle $A, B \in \mathcal{D}$ mit $A \subseteq B$ gilt $B \setminus A \in \mathcal{D}$,

(iii) Für jede Folge diskjunkter Mengen $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$ gilt $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$.

Offensichtlich ist jede σ-Algebra ein Dynkinsystem. Umgekehrt gilt

Lemma 2.21. *Ist $\mathcal{D}$ ein $\cap$-stabiles Dynkinsystem, so ist $\mathcal{D}$ auch eine σ-Algebra.*

Beweis. Wir weisen die Eigenschaften einer σ-Algebra nach. Es gilt $E \in \mathcal{D}$ und $E \setminus A \in \mathcal{D}$ für jedes $A \in \mathcal{D}$ nach Definition. Sind $A, B \in \mathcal{D}$, so ist auch $A \setminus B = A \setminus (A \cap B) \in \mathcal{D}$ und daher auch $A \cup B = A \cup (B \setminus A) \in \mathcal{D}$. Für $(A_n)_{n\geqslant 1} \subseteq \mathcal{D}$ definieren wir $B_1 := A_1 \in \mathcal{D}$ und $B_n := A_n \setminus \bigcup_{k=1}^{n-1} A_k \in \mathcal{D}$, so dass $(B_n)_{n\geqslant 1}$ paarweise disjunkt sind mit $\bigcup_{n\geqslant 1} A_n = \bigcup_{n\geqslant 1} B_n \in \mathcal{D}$. □

Sei $E \neq \emptyset$ und $\mathcal{E} \subseteq \mathscr{P}(E)$. Ganz analog zur erzeugten σ-Algebra ist das kleinste Dynkin-System, welches $\mathcal{E}$ umfasst, gegeben durch

$$\delta(\mathcal{E}) := \bigcap \{\mathcal{D} : \mathcal{D} \text{ ist Dynkinsystem über } E \text{ mit } \mathcal{E} \subseteq \mathcal{D}\}.$$

$\delta(\mathcal{E})$ heißt das *von $\mathcal{E}$ erzeugte Dynkinsystem.*

Satz 2.22. *Ist $\mathcal{E} \subseteq \mathscr{P}(E)$ $\cap$-stabil, so ist auch $\delta(\mathcal{E})$ $\cap$-stabil. Insbesondere gilt in diesem Fall $\delta(\mathcal{E}) = \sigma(\mathcal{E})$.*

Beweis. Für jedes $C \in \delta(\mathcal{E})$ werden wir zeigen, dass

$$\delta(\mathcal{E}) \subseteq \mathcal{D}_C := \{A \in \delta(\mathcal{E}) : A \cap C \in \delta(\mathcal{E})\}.$$

Daraus folgt $\mathcal{D}_C = \delta(\mathcal{E})$ für jedes $C \in \delta(\mathcal{E})$, also ist $\delta(\mathcal{E})$ $\cap$-stabil und nach Lemma 2.21 eine σ-Algebra, d.h. $\sigma(\mathcal{E}) \subseteq \delta(\mathcal{E})$. Anderseits ist jede σ-Algebra ein Dynkinsystem, so dass $\delta(\mathcal{E}) \subseteq \sigma(\mathcal{E})$ und damit $\sigma(\mathcal{E}) = \delta(\mathcal{E})$ folgt.

Zunächst weisen wir nach, dass $\mathcal{D}_C$ ein Dynkinsystem ist:

(i) $E \in \mathcal{D}_C$, da $E \cap C = C \in \delta(\mathcal{E})$.

(ii) Sind $A, B \in \mathcal{D}_C$ mit $A \subseteq B$, dann ist $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C) \in \delta(\mathcal{E})$.

(iii) Sind $(A_n)_{n\geqslant 1} \subseteq \mathcal{D}_C$ paarweise disjunkt, so ist $(\bigcup_{n\geqslant 1} A_n) \cap C = \bigcup_{n\geqslant 1}(A_n \cap C) \in \delta(\mathcal{E})$.

Für $C \in \mathcal{E}$ gilt $\mathcal{E} \subseteq \mathcal{D}_C$ nach Annahme, damit auch $\delta(\mathcal{E}) \subseteq \mathcal{D}_C$. Für jedes $A \in \delta(\mathcal{E})$ und $C \in \mathcal{E}$ gilt also $A \cap C \in \delta(\mathcal{E})$ also $C \in \mathcal{D}_A$. Es folgt $\mathcal{E} \subseteq \mathcal{D}_A$, was $\delta(\mathcal{E}) \subseteq \mathcal{D}_A$ für jedes $A \in \delta(\mathcal{E})$ impliziert. $\square$

Damit lässt sich der Eindeutigkeitssatz wie folgt verschärfen.

Satz 2.23 (Eindeutigkeitssatz II). *Es seien μ und ν zwei Maße auf einer σ-Algebra $\mathcal{A} = \sigma(\mathcal{E})$ über E erzeugt von einem $\cap$-stabilen Erzeuger $\mathcal{E}$ mit*

(i) $\mu|_{\mathcal{E}} = \nu|_{\mathcal{E}}$ und

(ii) es existiert eine wachsende Folge $(E_n)_{n\geqslant 1} \subseteq \mathcal{E}$ mit $\mu(E_n) = \nu(E_n) < \infty$ für alle $n \in \mathbb{N}$ und $\bigcup_n E_n = E$,

dann gilt $\mu = \nu$.

Beweis. Für jedes $A \in \mathcal{A} = \sigma(\mathcal{E})$ haben wir eine monotone wachsende Ausschöpfung $A = \bigcup_{n\geqslant 1} A \cap E_n$. Aufgrund der Stetigkeit von unten gilt

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap E_n) \quad \text{und} \quad \nu(A) = \lim_{n\to\infty} \nu(A \cap E_n).$$

Es genügt also $\mu(A \cap C) = \nu(A \cap C)$ für jedes $A \in \mathcal{A}$ und $C \in \mathcal{E}$ mit $\mu(C) < \infty$ zu zeigen. Wir definieren

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

Dann ist $\mathcal{D}_C$ für ein beliebiges $C \in \mathcal{E}$ mit $\mu(C) < \infty$ ein Dynkinsystem:

(i) $E \in \mathcal{D}_C$, denn $E \cap C = C \in \mathcal{E}$.

(ii) Sind $A, B \in \mathcal{D}_C$ mit $A \subseteq B$, dann gilt

$$\begin{aligned}\mu\big((B \setminus A) \cap C\big) = \mu\big((B \cap C) \setminus (A \cap C)\big) &= \mu(B \cap C) - \mu(A \cap C) \\ &= \nu(B \cap C) - \nu(A \cap C) = \nu((B \setminus A) \cap C).\end{aligned}$$

(iii) Sind $(A_n)_{n\geqslant 1} \subseteq \mathcal{D}_c$ disjunkt, dann folgt

$$\begin{aligned}\mu\Big(\big(\bigcup_{n\geqslant 1} A_n\big) \cap C\Big) = \mu\Big(\bigcup_{n\geqslant 1}(A_n \cap C)\Big) &= \sum_{n\geqslant 1} \mu(A_n \cap C) \\ &= \sum_{n\geqslant 1} \nu(A_n \cap C) = \nu\Big(\big(\bigcup_{n\geqslant 1} A_n\big) \cap C\Big).\end{aligned}$$

Da $\mathcal{E}$ $\cap$-stabil, ist $\mathcal{E} \subseteq \mathcal{D}_c$ und damit $\delta(\mathcal{E}) \subseteq \mathcal{D}_C \subseteq \mathcal{A} = \sigma(\mathcal{E})$. Anderseits folgt $\delta(\mathcal{E}) = \sigma(\mathcal{E})$ aus Satz 2.22, also $\mathcal{D}_C = \mathcal{A}$ für jedes $C \in \mathcal{E}$. $\square$

3 Integrationstheorie

Um Erwartungswerte, Varianzen, Kovarianzen, etc. von Zufallsgrößen zu berechnen, benötigen wir einen Integralbegriff auf Wahrscheinlichkeitsräumen $(E, \mathcal{A}, \mu)$. Diesen einzuführen und dessen Eigenschaften kennen zu lernen, ist Ziel dieses Kapitels. Als Integranden werden wir Funktionen $f : E \to \overline{\mathbb{R}} := [-\infty, \infty]$ betrachten, d.h. die Funktionswerte $\pm\infty$ sind zugelassen. Bevor wir das (Lebesgue-)Integral konstruieren können, lernen wir im ersten Abschnitt die Klasse von Abbildungen kennen, die wir anschließend integrieren wollen.

3.1 Messbare Abbildungen

Es seien $(E, \mathcal{A})$ und $(F, \mathcal{B})$ zwei messbare Räume und $f\colon E \to F$ eine Abbildung.

Definition 3.1. Die Abbildung f heißt $(\mathcal{A}, \mathcal{B})$-*messbar* (wenn keine Verwechslungen möglich sind, auch kürzer $\mathcal{A}$-messbar, oder einfach messbar), falls $f^{-1}(\mathcal{B}) \subseteq \mathcal{A}$, d.h.

$$f^{-1}(B) \in \mathcal{A} \quad \text{für alle } B \in \mathcal{B}. \tag{3.1}$$

Ist (F, d) ein metrischer Raum, so betrachtet man die Borel-σ-Algebra $\mathscr{B}_F$ und jede $(\mathcal{A}, \mathscr{B}_F)$-messbare Abbildung $f\colon E \to F$ heißt auch *Borel-messbar*. Im Allgemeinen ist Messbarkeit schwer nachzuprüfen. Oft hilft folgendes Kriterium:

Lemma 3.2. *Es sei $\mathcal{E}$ ein Erzeuger von $\mathcal{B}$, d.h. $\mathcal{B} = \sigma(\mathcal{E})$. Dann ist f $(\mathcal{A}, \mathcal{B})$-messbar genau dann, wenn*

$$f^{-1}(\mathcal{E}) \subseteq \mathcal{A}. \tag{3.2}$$

Beweis. Aus der Messbarkeit von f folgt offensichtlich (3.2). Andererseits folgt aus Beispiel 2.3(iv), dass

$$f^{-1}(\mathcal{B}) = f^{-1}(\sigma(\mathcal{E})) = \sigma(f^{-1}(\mathcal{E})) \subseteq \sigma(\mathcal{A}) = \mathcal{A}.$$ □

Beispiel 3.3. Gegeben sei der messbare Raum $(E, \mathcal{A})$.

(i) Gilt $A_1, \dots, A_n \in \mathcal{A}$ und $a_1, \dots, a_n \in \mathbb{R} \setminus \{0\}$, so ist

$$f(y) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(y), \quad y \in E, \tag{3.3}$$

$\mathcal{A}$-messbar. Funktionen von der Gestalt (3.3) nennen wir *einfache Funktionen.* Man prüft leicht nach, dass für einfache Funktionen f und g auch $\alpha f + \beta g$ eine einfache Funktion für alle $\alpha, \beta \in \mathbb{R}$ ist(siehe unten).

(ii) Jede stetige Funktion $f\colon \mathbb{R} \to \mathbb{R}$ ist Borel-messbar. Die umgekehrte Implikation gilt nicht, da die *Dirichletsche Funktion* $f(y) = \mathbb{1}_{\mathbb{Q}}(y), y \in \mathbb{R}$, messbar, aber nirgendwo stetig ist.

(iii) Es seien E eine Menge, $(F, \mathcal{B})$ ein messbarer Raum und $f\colon E \to F$ eine Abbildung. Dann ist $\sigma(f) := f^{-1}(\mathcal{B})$ die kleinste σ-Algebra $\mathcal{A}$ auf E, so dass f eine $(\mathcal{A}, \mathcal{B})$-messbare Abbildung ist. Wir nennen $\sigma(f)$ die von f *erzeugte σ-Algebra.*

(iv) Es seien $f\colon (E, \mathcal{A}) \to (F, \mathcal{B})$ und $g\colon (F, \mathcal{B}) \to (G, \mathcal{C})$ messbare Abbildungen. Dann ist die zusammengesetzte Abbildung

$$h\colon E \to G, \quad y \mapsto g \circ f(y) := g(f(y))$$

eine $(\mathcal{A}, \mathcal{C})$-messbare Abbildung.

Der folgende Satz liefert Eigenschaften messbarer Funktionen.

Satz 3.4. *Es sei $(E, \mathcal{A})$ ein messbarer Raum.*

(i) $f\colon E \to \overline{\mathbb{R}}^d$ ist genau dann Borel-messbar, wenn für alle $x \in \overline{\mathbb{R}}^d$ gilt

$$\{y \in E : f(y) \leqslant x\} \in \mathcal{A}.$$

(ii) Sind $f_1, f_2, \dots, f_n\colon E \to \mathbb{R}$ reellwertige $(\mathcal{A}, \mathscr{B}_{\mathbb{R}})$-messbare Funktionen und ist $h\colon \mathbb{R}^n \to \mathbb{R}^d$ Borel-messbar, so ist $h(f_1, \dots, f_n)$ eine $\mathcal{A}$-messbare Funktion.

(iii) Sind $f, g\colon E \to \mathbb{R}$ Borel-messbar, so sind es auch $f + g, f \cdot g, \max(f, g), \min(f, g)$ und, falls $g \neq 0$, $\frac{f}{g}$.

(iv) Sind $f_n\colon E \to \overline{\mathbb{R}}, n \geqslant 1$*, Borel-messbar, so sind es auch*

$$\sup_{n\geqslant 1} f_n, \quad \inf_{n\geqslant 1} f_n, \quad \limsup_{n\to\infty} f_n, \quad \liminf_{n\to\infty} f_n$$

(hierbei folgt die Bildung von $\sup f_n, \inf f_n$ *etc. punktweise, d.h.* $(\sup_n f_n)(y) := \sup_n f_n(y)$*).*

(v) Sind $f_n\colon E \to \overline{\mathbb{R}}$ *Borel-messbar und konvergiert* f_n *punktweise gegen* f*, dann ist* f *Borel-messbar.*

Beweis. *(i)* folgt direkt aus $\mathscr{B}_{\overline{\mathbb{R}}^d} = \sigma(\{(-\infty, x], x \in \overline{\mathbb{R}}^d\})$ und Lemma 3.2.

(ii) Wegen Beispiel 3.3(iv) betrachten wir o.B.d.A. die Identitätsabbildung $h(x) = x, x \in \mathbb{R}^n$. Da $\sigma(\{(a,b] = \prod_{k=1}^n (a_k, b_k] : a, b \in \mathbb{R}^n\}) = \mathscr{B}_{\mathbb{R}^n}$, genügt es

$$(f_1^{-1}, \ldots, f_n^{-1})((a,b]) = \bigcap_{k=1}^n f_k^{-1}((a_k, b_k]) \in \mathcal{A}$$

für jeden Quader $(a,b]$ zu zeigen. Da $f_1, \ldots, f_n$ messbar sind, folgt dies aus der Durchschnittsstabilität von $\mathcal{A}$.

(iii) folgt aus (ii) und Beispiel 3.3(ii).

(iv) folgt aus

$$\big\{\sup_n f_n \leqslant x\big\} = \bigcap_n \{f_n \leqslant x\} \quad \text{und} \quad \big\{\inf_n f_n \leqslant x\big\} = \bigcap_m \big\{\inf_n f_n < x + \frac{1}{m}\big\} = \bigcap_m \bigcup_n \{f_n < x + \frac{1}{m}\}$$

sowie den Darstellungen $\limsup_{n\to\infty} f_n = \inf_{n\geqslant 1} \sup_{k\geqslant n} f_k$ und $\liminf_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} \inf_{k\geqslant n} f_n$.

(v) Es gilt $f = \limsup_{n\to\infty} f_n = \liminf_{n\to\infty} f_n$, so dass die Behauptung aus (iv) folgt. □

Die letzte Eigenschaft lässt sich auf vollständigen Maßräumen etwas abschwächen.

Definition 3.5. Sei $(E, \mathcal{A}, \mu)$ ein Maßraum. Man sagt: Eine Eigenschaft gilt auf E "*μ-fast überall*", falls sie auf $E \setminus N$ gilt, wobei $N \in \mathcal{A}$ eine μ-Nullmenge ist.

Lemma 3.6. *Es seien* $(E, \mathcal{A}, \mu)$ *ein vollständiger Maßraum,* $(f_n\colon E \to \mathbb{R})$ *eine Folge* $\mathcal{A}$*-messbarer Funktionen und* $f\colon E \to \overline{\mathbb{R}}$*. Die Folge* (f_n) *konvergiere* μ*-fast überall gegen* f*, d.h.*

$$\lim_{n\to\infty} f_n(x) = f(x), \quad \textit{für alle } x \in E \setminus N, \;\; \textit{wobei } N \in \mathcal{A}, \mu(N) = 0.$$

Dann ist die Grenzfunktion f *ebenfalls* $\mathcal{A}$*-messbar.*

Beweis. Übung ☑. □

Der folgende Satz zeigt, dass man jede messbare Funktion (zerlegbar in zwei nichtnegative Funktionen $f = f_+ - f_-$) durch einfache Funktionen beliebig genau approximiert werden kann. Dieses Resultat wird zentral für die Konstruktion von Integralen sein.

Satz 3.7 (Approximationssatz)**.** *Für jede nichtnegative Borel-messbare Funktion* $f\colon (E, \mathcal{A}) \to ([0,\infty], \mathscr{B}_{[0,\infty]})$ *gibt es eine nichtfallende Folge* $(f_n)_{n\geqslant 1}$ *einfacher, nichtnegativer, Borel-messbarer Funktionen, die punktweise von unten gegen* f *konvergiert:*

$$0 \leqslant f_n(y) \leqslant f_{n+1}(y) \leqslant f(y), \quad n \geqslant 1, y \in E \quad \textit{und} \quad \lim_{n\to\infty} f_n(y) = f(y), \quad y \in E.$$

Beweis. Sei $f\colon E \to \mathbb{R}$ messbar mit $f \geqslant 0$. Für $n \in \mathbb{N}$ definieren wir

$$f_n(y) := \begin{cases} k \cdot 2^{-n}, & \text{falls } f(y) \in [k2^{-n}, (k+1)2^{-n}) \text{ und } k = 0, \ldots, n2^n - 1, \\ n, & \text{falls } f(y) \geqslant n. \end{cases} \tag{3.4}$$

Schritt 1: $(f_n)_{n\geqslant 1}$ ist eine Folge einfacher Funktionen mit $f_n \leqslant f$.
Setzen wir

$$\begin{aligned} A_{n,k} :=& f^{-1}([k2^{-n},(k+1)2^{-n})) \in \mathcal{A}, \quad \text{für } k=0,\dots,n2^n-1,\\ A_n :=& f^{-1}([n,\infty]) \in \mathcal{A}, \end{aligned}$$

dann ist E die disjunkte Vereinigung von A_n und $A_{n,k}, k = 0,\dots,n2^n-1$. Weiter können wir f_n darstellen als

$$f_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n}\mathbb{1}_{A_{n,k}} + n\mathbb{1}_{A_n},$$

so dass f_n einfache Funktionen sind.

Schritt 2: $(f_n)_{n\geqslant 1}$ ist monoton wachsend.
Wegen

$$\Big[\frac{k}{2^n},\frac{k+1}{2^n}\Big) = \Big[\frac{2k}{2^{n+1}},\frac{2k+1}{2^{n+1}}\Big) \cup \Big[\frac{2k+1}{2^{n+1}},\frac{2(k+1)}{2^{n+1}}\Big)$$

gilt $A_{n,k} = A_{n+1,2k}\dot\cup A_{n+1,2k+1}$. Aus der Definition von f_n und f_{n+1} ergibt sich

$$\begin{aligned} f_{n+1}|_{A_{n+1,2k}} &= \frac{2k}{2^{n+1}} = \frac{k}{2^n} = f_n|A_{n,k},\\ f_{n+1}|_{A_{n+1,2k+1}} &= \frac{2k+1}{2^{n+1}} > \frac{k}{2^n} = f_n|A_{n,k}. \end{aligned}$$

Für die Mengen $A_n = A_{n+1} \cup f^{-1}([n,n+1))$ gilt

$$f_{n+1}|_{A_{n+1}} = n+1 > n = f_n|_{A_n} \quad \text{und} \quad f_{n+1}|_{f^{-1}([n,n+1))} \geqslant n = f_n|_{A_n}.$$

Schritt 3: $(f_n)_{n\geqslant 1}$ konvergiert gegen f.
Ist $f(x) = +\infty$, so ist $f_n(x) = n$ für alle $n \in \mathbb{N}$ und die Behauptung klar. Ist $f(x) < \infty$, so gilt für $n \in \mathbb{N}$ mit $f(x) < n$, dass $x \notin A_n$, also $x \in \bigcup_{k=0}^{n2^n-1} A_{n,k}$. Deshalb gibt es ein k_0, so dass $x \in A_{n,k_0}$ und somit

$$0 \leqslant f(x) - f_n(x) < \frac{k_0+1}{2^n} - \frac{k_0}{2^n} = 2^{-n}. \qquad \square$$

Nun noch eine Eigenschaft, die es Stochastikern erlaubt, sich auf Wahrscheinlichkeitsmaße auf dem Ergebnisraum eines Zufallexperiments zu beschränken, statt den gesamten zugrundeliegenden Wirkmechanismus beschreiben zu müssen.

Satz 3.8. *Es seien $(E,\mathcal{A},\mu)$ ein Maßraum, $(F,\mathcal{B})$ ein messbarer Raum und $f\colon (E,\mathcal{A}) \to (F,\mathcal{B})$ eine messbare Abbildung. Durch*

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

ist auf $\mathcal{B}$ ein Maß μ^f definiert, das als von f induziertes Maß *oder als* Bildmaß *von f bezeichnet wird. Ist μ endlich, so ist auch μ^f endlich.*

Beweis. Es gilt $\mu^f(\emptyset) = 0$ und für jede Folge $(B_k)_{k\geqslant 1} \subseteq \mathcal{F}$ paarweiser diskjunkter Mengen

$$\mu^f\Big(\bigcup_{k\geqslant 1} B_k\Big) = \mu\Big(f^{-1}\big(\bigcup_{k\geqslant 1} B_k\big)\Big) = \mu\Big(\bigcup_{k\geqslant 1} f^{-1}(B_k)\Big) = \sum_{k\geqslant 1}\mu(f^{-1}(B_k)) = \sum_{k\geqslant 1}\mu^f(B_k). \qquad \square$$

Damit $f^{-1}(B) \in \mathcal{A}$ und somit obiger Ausdruck für alle $B \in \mathcal{B}$ wohldefiniert ist, benötigen wir also genau die Messbarkeit von f. Messbare Abbildung $X\colon \Omega \to \mathcal{F}$ auf Wahrscheinlichkeitsräumen $(\Omega,\mathcal{F},P)$ heißen auch *Zufallsvariablen* und das Bildmaß P^X wird auch *Wahrscheinlichkeitsverteilung der Zufallsvariable X* genannt.

3.2 Integralkonstruktion

In diesem Abschnitt definieren wir das Integral über messbare Funktionen auf einem Maßraum $(E, \mathcal{A}, \mu)$ und studieren die Eigenschaften dieses Integrals. Um das Integral zu konstruieren, gehen wir in 3 Schritten vor:

(i) Integral für einfache Funktionen

(ii) Integral für nichtnegative Funktionen

(iii) Integral für beliebige messbare Funktionen

Definition 3.9. Sei $f\colon E \to \mathbb{R}$ eine einfache, nichtnegative Funktion mit der Darstellung $f = \sum_{i=1}^m a_i \mathbb{1}_{A_i}$ für $m \in \mathbb{N}, a_1, \dots, a_m > 0$ und $A_1, \dots, A_m \in \mathcal{A}$. Die Zahl

$$\int_E f \mathrm{d}\mu := \int_E f(y)\mu(\mathrm{d}y) := \sum_{i=1}^m a_i \mu(A_i)$$

heißt *Integral von f über E bezüglich des Maßes μ.*

Man prüft leicht nach, dass diese Definition korrekt ist, d.h. nicht von der Wahl der Darstellung von f als einfache Funktion abhängt (Übung ☑).

Satz 3.10. *Seien $f, g\colon E \to \mathbb{R}$ einfache Funktionen. Dann gilt:*

(i) Ist $f \leqslant g$, so gilt

$$\int_E f \mathrm{d}\mu \leqslant \int_E g \mathrm{d}\mu \quad \textit{(Montonie des Integrals).}$$

(ii) Die Funktion $\alpha f + \beta g$ für $\alpha, \beta \in \mathbb{R}$ ist einfach und es gilt

$$\int_E (\alpha f + \beta g) \mathrm{d}\mu = \alpha \int_E f \mathrm{d}\mu + \beta \int_E g \mathrm{d}\mu \quad \textit{(Linearität des Integrals).}$$

Beweis. Wir stellen f und g als einfache Funktionen für eine gemeinsame disjunkte Zerlegung dar. O.B.d.A sind f und g gegeben durch Normaldarstellungen

$$f = \sum_{i=1}^m a_i \mathbb{1}_{A_i} \quad \text{und} \quad g = \sum_{j=1}^n b_j \mathbb{1}_{B_j},$$

mit $a_1, \dots, a_n, b_1, \dots, b_m \geqslant 0$ und $(A_i)_{i=1,\dots,m}$ und $(B_j)_{j=1,\dots,n}$ sind disjunkte Zerlegungen von E (Übung ☑). Dann gilt

$$f = \sum_{i=1}^m \sum_{j=1}^n a_i \mathbb{1}_{A_i \cap B_j} \quad \text{und} \quad g = \sum_{i=1}^m \sum_{j=1}^n b_j \mathbb{1}_{A_i \cap B_j}.$$

Aus letzterer Darstellung folgen die Behauptungen aus der Definition. □

Definition 3.11. Sei $(E, \mathcal{A}, \mu)$ ein Maßraum. $\mathcal{M}_+$ bezeichne die Menge aller nichtnegativen, $\mathcal{A}$-messbaren Funktionen $f\colon E \to [0, \infty]$. Für $f \in \mathcal{M}_+$ heißt die Zahl

$$\int_E f \mathrm{d}\mu := \sup \left\{ \int_E \varphi \mathrm{d}\mu \middle| \varphi\colon E \to \mathbb{R} \text{ einfach}, \varphi \leqslant f \right\} \in [0, \infty]$$

Integral von f über E bzgl. des Maßes μ. Ist $A \in \mathcal{A}$, so definiert man das *Integral von f über A (bzgl. μ) durch*

$$\int_A f \mathrm{d}\mu := \int_E f \mathbb{1}_A \mathrm{d}\mu.$$

Satz 3.12. *Sei $(E, \mathcal{A}, \mu)$ ein Maßraum und seien $f, g \in \mathcal{M}_+$. Dann gilt*

(i) $f \leqslant g$ auf $A \in \mathcal{A}$ impliziert $\int_A f \mathrm{d}\mu \leqslant \int_A g \mathrm{d}\mu$.

(ii) Sind $A, B \in \mathcal{A}$ mit $A \subseteq B$, so gilt $\int_A f \mathrm{d}\mu \leqslant \int_B f \mathrm{d}\mu$.

(iii) Sind $A, B \in \mathcal{A}$ disjunkt, dann gilt

$$\int_{A \cup B} f \mathrm{d}\mu = \int_A f \mathrm{d}\mu + \int_B f \mathrm{d}\mu.$$

(iv) Es gilt $\int_E f \mathrm{d}\mu = 0$ genau dann, wenn $\mu(\{x \in E : f(x) > 0\}) = 0$ (d.h. $f = 0$ μ-fast überall).

Beweis. (i)-(iii) Übung ☑.

(iv) $\Rightarrow$: Wir setzen $A_n := \{x \in E : f(x) \geqslant \frac{1}{n}\}$ für $n \in \mathbb{N}$. Dann gilt $A_n \subseteq A_{n+1}$ und

$$A := \{x \in E : f(x) > 0\} = \bigcup_{n \geqslant 1} A_n.$$

Da f messbar ist, gilt $A_n \in \mathcal{A}$ und damit auch $A \in \mathcal{A}$. Aus den Monotonieeigenschaften (i) und (ii) folgt

$$0 = \int_E f \mathrm{d}\mu \geqslant \int_{A_n} f \mathrm{d}\mu \geqslant \int_{A_n} \frac{1}{n} \mathrm{d}\mu = \frac{\mu(A_n)}{n} \quad \text{für alle } n \in \mathbb{N}.$$

Also gilt $\mu(A_n) = 0$ für alle $n \in \mathbb{N}$. Aus der Stetigkeit des Maßes von unten folgt $\mu(A) = \lim_{n \to \infty} \mu(A_n) = 0$.

$\Leftarrow$: Sei $A := \{x \in E : f(x) > 0\}$ und $\mu(A) = 0$. Für jede einfache, nichtnegative Funktion $\varphi \leqslant f$ gilt dann $A_\varphi := \{x \in E : \varphi(x) > 0\} \subseteq A$, also $\mu(A_\varphi) = 0$ und folglich $\int_E \varphi \mathrm{d}\mu = 0$. Aus der Integraldefinition erhält man $\int_E f \mathrm{d}\mu = 0$. □

Als nächstes beweisen wir grundlegende Konvergenzeigenschaften für das Integral von Funktionenfolgen. Im Allgemeinen können wir nicht erwarten, dass das Integral der Grenzfunktion gleich dem Grenzwert der Integrale über die Funktionenfolge ist.

Beispiel 3.13. Auf dem Maßraum $([0,1], \mathscr{B}_{[0,1]}, \lambda)$ mit Lebesguemaß λ, betrachten wir die Folge einfacher Funktionen

$$f_n \colon \mathbb{R} \to \mathbb{R}_+, \quad f_n(y) = n \mathbb{1}_{[0,1/n]}(y), y \in \mathbb{R},$$

mit dem punktweisen Grenzwert $f \colon \mathbb{R} \to \overline{\mathbb{R}}_+$ mit $f(x) = 0$ für $x \neq 0$ und $f(0) = \infty$. Es gilt $\int_{[0,1]} f_n \mathrm{d}\lambda = n \cdot \frac{1}{n} = 1$, jedoch $\int_{[0,1]} f \mathrm{d}\lambda = 0$ nach Satz 3.12(iv).

Satz 3.14 (Satz von Beppo Levi über monotone Konvergenz, 1906)**.** *Sei $(E, \mathcal{A}, \mu)$ ein Maßraum, $(f_n)_{n \geqslant 1} \subseteq \mathcal{M}_+$ eine monoton wachsende Folge und $f \colon E \to [0, \infty]$ die Grenzfunktion $f(x) := \lim_{n \to \infty} f_n(x), x \in E$. Dann gilt*

$$\lim_{n \to \infty} \int_E f_n \mathrm{d}\mu = \int_E f \mathrm{d}\mu.$$

Beweis. Nach Satz 3.4 ist f messbar. Da $f_n \leqslant f_{n+1} \leqslant f$ folgt aus Satz 3.12 $\int_E f_n \mathrm{d}\mu \leqslant \int_E f_{n+1} \mathrm{d}\mu \leqslant \int_E f \mathrm{d}\mu$ und somit

$$\lim_{n \to \infty} \int_E f_n \mathrm{d}\mu \leqslant \int_E f \mathrm{d}\mu.$$

Es bleibt die Ungleichung in der anderen Richtung zu zeigen. Sei $\varphi \leqslant f$ eine einfache Funktion mit der Darstellung $\varphi = \sum_{k=1}^m \alpha_k \mathbb{1}_{A_k}$. Sei $\lambda \in (0,1)$ fixiert und

$$B_n := \{x \in E : \lambda \varphi(x) \leqslant f_n(x)\} \in \mathcal{A}.$$

Da $f_n \uparrow f$ und $\lambda\varphi(x) < f(x)$ für jedes $x \in E$ mit $f(x) \neq 0$, gilt $E = \bigcup_{n\geqslant 1} B_n$ und $B_n \subseteq B_{n+1}$. Die Mengen $A_k \in \mathcal{A}$ erfüllen $A_k = \bigcup_{n\geqslant 1}(A_k \cap B_n)$. Aus der Stetigkeit von μ von unten folgt $\mu(A_k) = \lim_{n\to\infty} \mu(A_k \cap B_n)$. Nach Definition ist

$$\begin{aligned}
\lambda \int_E \varphi \mathrm{d}\mu = \lambda \sum_{k=1}^{m} \alpha_k \mu(A_k) &= \lim_{n\to\infty} \lambda \sum_{k=1}^{m} \alpha_k \mu(A_k \cap B_n) \\
&= \lim_{n\to\infty} \int_E \Big(\sum_{k=1}^{m} \lambda\alpha_k \underbrace{\mathbb{1}_{A_k\cap B_n}}_{=\mathbb{1}_{A_k}\cdot\mathbb{1}_{B_n}} \Big) \mathrm{d}\mu \\
&= \lim_{n\to\infty} \int_{B_n} \lambda\varphi \mathrm{d}\mu.
\end{aligned}$$

Nach Definition gilt $\lambda\varphi \leqslant f_n$ auf B_n. Somit folgt

$$\lambda \int_E \varphi \mathrm{d}\mu \leqslant \lim_{n\to\infty} \int_{B_n} f_n \mathrm{d}\mu \leqslant \lim_{n\to\infty} \int_E f_n \mathrm{d}\mu.$$

Bilden wir den Grenzwert $\lambda \to 1$ und das Supremum über alle einfachen Funktionen $\varphi \leqslant f$, so erhalten wir

$$\int_E f \mathrm{d}\mu \leqslant \lim_{n\to\infty} \int_E f_n \mathrm{d}\mu. \qquad \square$$

Der Approximationssatz 3.7 liefert unmittelbar:

Korollar 3.15. *Sei $f \in \mathcal{M}_+$. Dann existiert eine monoton wachsende Folge nichtnegativer, einfacher Funktionen (φ_n) mit $\varphi_n \uparrow f$ und es gilt*

$$\int_E f \mathrm{d}\mu = \lim_{n\to\infty} \int_E \varphi_n \mathrm{d}\mu.$$

Diese äquivalente Darstellung des Integrals messbarer, nichtnegative Funktionen (als Grenzwert der Integrale über eine approximierende Folge einfacher Funktionen) wird häufig als Definition von $\int_E f \mathrm{d}\mu$ verwendet. Die Charakterisierung über das Supremum hat den Vorteil, dass die Unabhängigkeit des Integrals von der Auswahl der approximierenden Folge sofort ersichtlich ist.

Korollar 3.16. *Sind $f, g \in \mathcal{M}_+$ und $\alpha, \beta \in [0, \infty)$, so gilt*

$$\int_E (\alpha f + \beta g) \mathrm{d}\mu = \alpha \int_E f \mathrm{d}\mu + \beta \int_E g \mathrm{d}\mu.$$

Beweis. Seien $(\varphi_n)_{n\geqslant 1}$ und $(\psi_n)_{n\geqslant 1}$ monoton wachsende Folgen nichtnegativer, einfacher Funktionen mit $\varphi_n \uparrow f$ und $\psi_n \uparrow g$. Dann ist auch $(\alpha\varphi_n + \beta\psi_n)_{n\geqslant 1}$ eine wachsende Folge einfacher Funktionen mit $\lim_{n\to\infty} \alpha\varphi_n + \beta\psi_n = \alpha f + \beta g$. Somit liefert die Linearität des Integrals von einfachen Funktionen (Satz 3.10)

$$\begin{aligned}
\int_E (\alpha f + \beta g) \mathrm{d}\mu &= \lim_{n\to\infty} \int_E (\alpha\varphi_n + \beta\psi_n) \mathrm{d}\mu \\
&= \lim_{n\to\infty} \alpha \int_E \varphi_n \mathrm{d}\mu + \lim_{n\to\infty} \beta \int_E \psi_n \mathrm{d}\mu = \alpha \int_E f \mathrm{d}\mu + \beta \int_E g \mathrm{d}\mu. \qquad \square
\end{aligned}$$

Korollar 3.17. *Für jede Folge $(f_n)_{n\geqslant 1} \subseteq \mathcal{M}_+$ gilt*

$$\sum_{n=1}^{\infty} \int_E f_n \mathrm{d}\mu = \int_E \Big(\sum_{n=1}^{\infty} f_n \Big) \mathrm{d}\mu.$$

Beispiel 3.18.

(i) Betrachten wir den Maßraum $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \mu)$ mit dem Zählmaß μ. Dann gilt für jede Funktion $f\colon \mathbb{N} \to [0,\infty]$, dass

$$\int_{\mathbb{N}} f \mathrm{d}\mu = \sum_{n\in\mathbb{N}} f(n).$$

Dies sieht man wie folgt: Jede Funktion ist messbar, da wir als σ-Algebra die Potenzmenge gewählt haben. Mit $g_n := f(n)\mathbb{1}_{\{n\}} \geqslant 0$ gilt $f = \sum_{n\geqslant 1} g_n$. Dann folgt aus den vorangegangen zwei Korollaren

$$\int_{\mathbb{N}} f\mathrm{d}\mu = \sum_{n\geqslant 1} \int_{\mathbb{N}} g_n \mathrm{d}\mu = \sum_{n\geqslant 1} f(n) \int_{\mathbb{N}} \mathbb{1}_{\{n\}}\mathrm{d}\mu = \sum_{n\geqslant 1} f(n)\mu(\{n\}) = \sum_{n\geqslant 1} f(n).$$

(ii) Sei $(E, \mathcal{A}, \mu)$ ein Maßraum und $f \in \mathcal{M}_+$. Wir definieren die Abbildung $\nu\colon \mathcal{A} \to [0,\infty]$ durch

$$\nu(A) := \int_E f \cdot \mathbb{1}_A \mathrm{d}\mu, \quad \text{für alle } A \in \mathcal{A}.$$

Dann gilt (Übung ☑):

(a) ν ist ein Maß auf $(E, \mathcal{A})$, das sogenannte *Maß mit der Dichte f bezüglich μ.*

(b) ν ist stetig bezüglich μ im dem Sinn, dass für alle $A \in \mathcal{A}$ mit $\mu(A) = 0$ auch $\nu(A) = 0$ gilt.

(c) Ist $g \in \mathcal{M}_+$, so gilt $\int_E g\mathrm{d}\nu = \int_E (f \cdot g)\mathrm{d}\mu$.

Für eine beliebige Folge messbarer, nichtnegativer Funktionen gilt der folgenden Konvergenzsatz:

Satz 3.19 (Lemma von Fatou, 1906)**.** *Sei $(E, \mathcal{A}, \mu)$ ein Maßraum, $(f_n\colon E \to [0,\infty])$ eine Folge $\mathcal{A}$-messbarer Funktionen. Dann gilt*

$$\int_E \liminf_{n\to\infty} f_n \mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_E f_n \mathrm{d}\mu.$$

Beweis. Für den Limes Inferior einer Folge gilt

$$\liminf_{n\to\infty} f_n = \sup_n \inf_{k\geqslant n} f_k.$$

Sei nun $g_n := \inf_{k\geqslant n} f_k$. Nach Satz 3.4 ist die Funktion g_n $\mathcal{A}$-messbar. Weiterhin gilt $g_n \leqslant g_{n+1}$ und $g_n \leqslant f_k$ für alle $k \geqslant n$. Daraus folgt

$$\int_E g_n \mathrm{d}\mu \leqslant \int_E f_k \mathrm{d}\mu \quad \text{und damit} \quad \int_E g_n \mathrm{d}\mu \leqslant \inf_{k\geqslant n} \int_E f_k \mathrm{d}\mu, \quad n \in \mathbb{N}.$$

Wir wenden nun den Satz über monotone Konvergenz auf die Folge (g_n) an und erhalten

$$\begin{aligned}\int_E \liminf_{n\to\infty} f_n \mathrm{d}\mu &= \int_E \sup_n g_n \mathrm{d}\mu = \int_E \lim_{n\to\infty} g_n \mathrm{d}\mu = \lim_{n\to\infty} \int_E g_n \mathrm{d}\mu \\ &= \sup_n \int_E g_n \mathrm{d}\mu \leqslant \sup_n \inf_{k\geqslant n} \int_E f_k \mathrm{d}\mu = \liminf_{n\to\infty} \int_E f_n \mathrm{d}\mu.\end{aligned}$$ □

Nun werden wir das Integral für eine beliebige $\mathcal{A}$-messbare Funktion $f\colon E \to \overline{\mathbb{R}}$ definieren. Wir verwenden die Zerlegung

$$f = f_+ - f_-, \quad f_+ = \max\{0, f\}, f_-(y) = \max\{0, -f\},$$

wobei $f_+, f_- \in \mathcal{M}_+$. Es gilt außerdem $|f| = f_+ + f_-$.

Definition 3.20. Sei $(E, \mathcal{A}, \mu)$ ein Maßraum. Eine $\mathcal{A}$-messbare Funktion $f\colon E \to \overline{\mathbb{R}}$ heißt *μ-integrierbar*, wenn $\int_E f_+ \mathrm{d}\mu < \infty$ oder $\int_E f_- \mathrm{d}\mu < \infty$ gilt (oder beides endlich ist). In diesem Fall ist das *Integral von f über E bezüglich μ* definiert als

$$\int_E f(x)\mu(\mathrm{d}x) := \int_E f \mathrm{d}\mu := \begin{cases} \int_E f_+ \mathrm{d}\mu - \int_E f_- \mathrm{d}\mu, & \text{falls } \int_E f_+ \mathrm{d}\mu < \infty \text{ und } \int_E f_- \mathrm{d}\mu < \infty, \\ +\infty, & \text{falls } \int_E f_+ \mathrm{d}\mu = \infty \text{ und } \int_E f_- \mathrm{d}\mu < \infty, \\ -\infty, & \text{falls } \int_E f_+ \mathrm{d}\mu < \infty \text{ und } \int_E f_- \mathrm{d}\mu = \infty. \end{cases}$$

Wir bezeichnen mit $\mathcal{L}^1(\mu) = \mathcal{L}^1(E, \mathcal{A}, \mu)$ die Menge aller μ-integrierbaren $\mathcal{A}$-messbaren Funktionen $f\colon E \to \overline{\mathbb{R}}$ mit $|\int f \mathrm{d}\mu| < \infty$.

Ist $\mu = \lambda$ das Lebesguemaß auf $(\mathbb{R}, \mathscr{L}_{\mathbb{R}})$, so heißt $\int_{\mathbb{R}} f(x)\mathrm{d}x := \int_{\mathbb{R}} f \mathrm{d}\lambda$ *Lebesgueintegral* von f. Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, also ein Maßraum mit Wahrscheinlichkeitsmaß P, dann heißt $E[f] := \int_\Omega f \mathrm{d}P$ Erwartungswert von f unter P.

Bemerkung 3.21. Das Lebesgueintegral beruht auf der Annäherung von messbaren Abbildungen mittels einfacher Funktionen. Diese Approximation durch einfache Funktionen beruht auf einer Zerlegung des Wertebereichs, vgl. Satz 3.7. Im Gegensatz dazu wird das in der Schule gelehrte *Riemannintegral* von Funktionen $f\colon \mathbb{R} \to \mathbb{R}$ durch eine Zerlegung des Definitionsbereiches konstruiert.

Man kann zeigen, dass jede Riemannintegrierbare (z.B. stetige) Funktion auf einem Intervall $[a, b]$ auch Lebesgueintegrierbar bzgl. des Lebesguemaßes λ ist und beide Integrale übereinstimmen. Da das Lebesgueintegral auch die Integration bezüglich sehr beliebiger Maße zulässt, ist es sehr viel allgemeiner als das Riemannintegral.

Lemma 3.22. *Es seien $(E, \mathcal{A}, \mu)$ ein Maßraum und $f\colon E \to \overline{\mathbb{R}}$ eine $\mathcal{A}$-messbare Abbildung. Dann gilt $f \in \mathcal{L}^1(\mu)$ genau dann, wenn ein $g \in \mathcal{M}_+$ existiert mit $|f| \leqslant g$ und $\int_E g \mathrm{d}\mu < \infty$. g heißt* Majorante *von f.*

Beweis. $\Rightarrow$: Ist $f \in \mathcal{L}^1(\mu)$, so gilt $\int_E f_\pm \mathrm{d}\mu < \infty$, so dass $|f| = f_+ + f_-$ eine Majorante mit endlichem Integral ist.

$\Leftarrow$: Sei nun $g \in \mathcal{M}_+$ mit $|f| \leqslant g$ und $\int_E g \mathrm{d}\mu < \infty$, so liefert die Monotonie des Integral über nichtnegative, messbare Funktionen (Satz 3.12), dass

$$\int_E f_\pm \mathrm{d}\mu \leqslant \int_E |f| \mathrm{d}\mu \leqslant \int_E g \mathrm{d}\mu < \infty.$$

Damit gilt $f \in \mathcal{L}^1(\mu)$. □

Wir werden nun Rechenregeln für das Integral formulieren und beweisen.

Satz 3.23. *Für $f, g \in \mathcal{L}^1(\mu)$ gelten folgende Aussagen:*

(i) Für alle $\alpha, \beta \in \mathbb{R}$ ist $\alpha f + \beta g \in \mathcal{L}^1(\mu)$ und es gilt

$$\int_E (\alpha f + \beta g) \mathrm{d}\mu = \alpha \int_E f \mathrm{d}\mu + \beta \int_E g \mathrm{d}\mu.$$

(ii) Wenn $f \leqslant g$, so ist $\int_E f \mathrm{d}\mu \leqslant \int_E g \mathrm{d}\mu$.

(iii) Es gilt

$$\Big| \int_E f \mathrm{d}\mu \Big| \leqslant \int_E |f| \mathrm{d}\mu < \infty.$$

(iv) Wenn $f = g$ μ-fast überall, d.h. $\mu(\{x \in E : f(x) \neq g(x)\}) = 0$, so ist $\int_E f \mathrm{d}\mu = \int_E g \mathrm{d}\mu$.

(v) f ist μ-fast überall endlich, d.h. $\mu(\{x \in E : |f(x)| = \infty\}) = 0$.

Beweis. (i) *Schritt 1: Ist $f = p - q$ mit Funktionen $p, q \in \mathcal{M}_+$ mit endlichen Integralen, so gilt*

$$\int_E f \mathrm{d}\mu = \int_E p \mathrm{d}\mu - \int_E q \mathrm{d}\mu.$$

Aus $f = f_+ - f_- = p - q$ folgt $q + f_+ = p + f_-$ und die Additivität des Integrals über $\mathcal{M}_+$ liefert

$$\int_E q \mathrm{d}\mu + \int_E f_+ \mathrm{d}\mu = \int_E p \mathrm{d}\mu + \int_E f_- \mathrm{d}\mu.$$

Da alle Terme endlich sind, folgt hieraus

$$\int_E f \mathrm{d}\mu = \int_E f_+ \mathrm{d}\mu - \int_E f_- \mathrm{d}\mu = \int_E p \mathrm{d}\mu - \int_E q \mathrm{d}\mu.$$

Schritt 2: $f + g \in \mathcal{L}^1(\mu)$ und es gilt

$$\int_E (f+g) \mathrm{d}\mu = \int_E f \mathrm{d}\mu + \int_E g \mathrm{d}\mu.$$

Die Funktionen $p := f_+ + g_+$ und $q := f_- + g_-$ aus $\mathcal{M}_+$ haben ein endliches Integral und es gilt $f + g = p - q$. Da $|f + g| \leqslant p + q$, folgt aus Lemma 3.22, dass $f + g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$. Aus Schritt 1 und der Linearität des Integrals über $\mathcal{M}_+$ folgt

$$\begin{aligned}
\int_E (f+g) \mathrm{d}\mu &= \int_E p \mathrm{d}\mu - \int_E q \mathrm{d}\mu \\
&= \int_E f_+ \mathrm{d}\mu + \int_E g_+ \mathrm{d}\mu - \int_E f_- \mathrm{d}\mu - \int_E g_- \mathrm{d}\mu \\
&= \int_E f \mathrm{d}\mu + \int_E g \mathrm{d}\mu.
\end{aligned}$$

Schritt 3: $\alpha f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ und $\int_E \alpha f \mathrm{d}\mu = \alpha \int_E f \mathrm{d}\mu$.

Ist $\alpha \geqslant 0$ ist $(\alpha f)_+ = \alpha f_+$ und $(\alpha f)_- = \alpha f_-$ und die Behauptung folgt aus der Linearität des Integrals auf $\mathcal{M}_+$. Ist $\alpha < 0$, gilt $(\alpha f)_+ = |\alpha| f_-$ und $(\alpha f)_- = |\alpha| f_+$. Es folgt

$$\int_E \alpha f \mathrm{d}\mu = |\alpha| \Big(\int_E f_- \mathrm{d}\mu - \int_E f_+ \mathrm{d}\mu \Big) = \alpha \int_E f \mathrm{d}\mu.$$

(ii) Aufgrund der Linearität des Integrals genügt es $\int_E h \mathrm{d}\mu \geqslant 0$ für $h := g - f \geqslant 0$ zu zeigen. Letzteres folgt aus Satz 3.12.

(iii) Es gilt

$$\Big| \int_E f \mathrm{d}\mu \Big| = \Big| \int_E f_+ \mathrm{d}\mu - \int_E f_- \mathrm{d}\mu \Big| \leqslant \int_E f_+ \mathrm{d}\mu + \int_E f_- \mathrm{d}\mu = \int_E (f_+ + f_-) \mathrm{d}\mu = \int_E |f| \mathrm{d}\mu.$$

(iv) Es genügt $\int f \mathrm{d}\mu = 0$ für jedes messbare f mit $f = 0$ μ-fast überall zu zeigen. Für ein solches f sind auch f_+ und f_- μ-fast überall 0. Nach Satz 3.12 gilt dann $\int_E f_+ \mathrm{d}\mu = \int_E f_- \mathrm{d}\mu = 0$ und somit folgt die Behauptung.

(v) Da $|f| = f_+ + f_-$, gilt

$$\{x \in E : |f(x)| = \infty\} = \{x \in E : f_+(x) = \infty\} \cup \{x \in E : f_-(x) = \infty\}.$$

Wir können also o.B.d.A $f \geqslant 0$ annehmen. Sei $M = \{x \in E : f(x) = \infty\}$. Für $n \in \mathbb{N}$ gilt $f \geqslant f \cdot \mathbb{1}_M \geqslant n \cdot \mathbb{1}_M$. Folglich ist

$$+\infty > \int_E f \mathrm{d}\mu \geqslant n \int_E \mathbb{1}_M \mathrm{d}\mu = n \cdot \mu(M) \quad \text{für alle } n \in \mathbb{N}.$$

Dann muss aber $\mu(M) = 0$ gelten. $\square$

Als nächstes beweisen wir einen weiteren wichtigen Konvergenzsatz für das Integral von Funktionenfolgen.

Satz 3.24 (Satz von Lebesgue über die majorisierte Konvergenz, 1910)**.** *Sei* $(E, \mathcal{A}, \mu)$ *ein Maßraum,* $f\colon E \to \overline{\mathbb{R}}$ $\mathcal{A}$*-messbar und* $(f_n\colon E \to \overline{\mathbb{R}})_{n\geqslant 1}$ *eine Folge* $\mathcal{A}$*-messbarer Funktionen mit*

(i) $\lim_{n\to\infty} f_n(x) = f(x)$ *für* μ*-fast alle* $x \in E$*.*

(ii) Es existiert eine Funktion $F \in \mathcal{L}^1(\mu)$ *mit*

$$|f_n(x)| \leqslant F(x) \quad \textit{für } \mu\textit{-fast alle } x \in E.$$

Dann ist $f \in \mathcal{L}^1(\mu)$ *und es gilt*

$$\int_E f \mathrm{d}\mu = \lim_{n\to\infty} \int_E f_n \mathrm{d}\mu.$$

Beweis. Nach Satz 3.23 können wir, evtl. durch Abändern der Funktionen f, f_n bzw. F auf einer μ-Nullmenge, folgendes annehmen: f_n, f sind messbare Funktionen mit

(i) Für alle $n \in \mathbb{N}$ ist f_n über E integrierbar und endlich.

(ii) Die Folge (f_n) konvergiert auf dem ganzen Raum E punktweise gegen f.

(iii) $|f_n(x)| \leqslant F(x)$ für alle $x \in E$.

Da $|f| = \lim_{n\to\infty} |f_n| \leqslant F \in \mathcal{L}^1(\mu)$ folgt $f \in \mathcal{L}^1(\mu)$ aus Lemma 3.22. Wegen (iii) ist $0 \leqslant f_n + F$ und $0 \leqslant F - f_n$. Aus $\lim_{n\to\infty} f_n = f$ und dem Lemma von Fatou folgt

$$\begin{aligned}
0 \leqslant \int_E (f+F)\mathrm{d}\mu &= \int_E \liminf_{n\to\infty}(f_n + F)\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_E (f_n+F)\mathrm{d}\mu \\
&= \int_E F\mathrm{d}\mu + \liminf_{n\to\infty} \int_E f_n \mathrm{d}\mu, \\
0 \leqslant \int_E (F-f)\mathrm{d}\mu &= \int_E \liminf_{n\to\infty}(F - f_n)\mathrm{d}\mu \leqslant \int_E F\mathrm{d}\mu - \limsup_{n\to\infty} \int_E f_n \mathrm{d}\mu.
\end{aligned}$$

Aus $\int_E F\mathrm{d}\mu < \infty$ schließen wir

$$\int_E f\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_E f_n\mathrm{d}\mu \leqslant \limsup_{n\to\infty} \int_E f_n \mathrm{d}\mu \leqslant \int_E f\mathrm{d}\mu.$$

Somit existiert der Grenzwert $\lim_{n\to\infty} \int_E f_n \mathrm{d}\mu = \int_E f\mathrm{d}\mu$. □

Bemerkung 3.25. Für das Riemann-Integral benötigt man die gleichmäßige Konvergenz der Funktionenfolge (f_n) gegen f, um Integral und Grenzwert zu vertauschen. Das Lebesgue-Integral hat also bessere Konvergenzeigenschaften.

Die folgende wichtige Eigenschaft besagt, dass man zur Berechnung von $\int (h \circ f)\mathrm{d}\mu$ nur das Bildmaß μ^f kennen muss, nicht die Abbildung f noch das Maß μ selbst. Dies ist in der Statistik von fundamentaler Bedeutung, da wir häufig den zugrundeliegenden Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ nicht kennen, sondern nur abgeleitete Größen, die so genannten Zufallsvariablen, $X\colon \Omega \to F$ beobachten und daher (unter Verwendung von Stichproben) nur die Wahrscheinlichkeitsverteilung P^X zugänglich ist.

Satz 3.26 (Substitutionsregel)**.** *Es seien* $f\colon E \to F$ *eine* $(\mathcal{A}, \mathcal{B})$*-messbare Abbildung in einen messbaren Raum* $(F, \mathcal{B})$*,* $g\colon F \to \mathbb{R}$ *eine Borel-messbare Funktion und* μ^f *das Bildmaß von* μ *vermittels der Abbildung* f*. Es gilt:*

(i) $g \circ f = g(f) \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ *genau dann, wenn* $g \in \mathcal{L}^1(F, \mathcal{B}, \mu^f)$ *und*

(ii) falls $g \geqslant 0$ oder $g \in \mathcal{L}^1(F, \mathcal{B}, \mu^f)$, dann

$$\int_E g(f(y))\mu(\mathrm{d}y) = \int_F g(x)\mu^f(\mathrm{d}x). \tag{3.5}$$

Beweis. Schritt 1: Die Gleichung (3.5) gilt für $g = \mathbb{1}_B, B \in \mathcal{B}$, nach Definition des Bildmaßes μ^f:

$$\begin{aligned}\int_E g(x)\mu^f(\mathrm{d}x) &= \mu^f(B) = \mu(f^{-1}(B)) \\ &= \int_E \mathbb{1}_{f^{-1}(B)}(y)\mu(\mathrm{d}y) = \int_E \mathbb{1}_B(f(y))\mu(\mathrm{d}y) = \int_E g(f(y))\mu(\mathrm{d}y).\end{aligned}$$

Schritt 2: Aufgrund der Linearität des Integrals gilt (3.5) für alle einfachen Funktionen.

Schritt 3: Ist $(\varphi_n)_{n\geqslant 0}$ eine wachsende Folge nichtnegativer, einfacher Funktionen mit $g = \lim_{n\to\infty} \varphi_n$, und dank Schritt 2

$$\lim_{n\to\infty} \int_E (\varphi_n \circ f)\mathrm{d}\mu = \lim_{n\to\infty} \int_E \varphi_n \mathrm{d}\mu^f = \int_E g\mathrm{d}\mu^f.$$

Auch $\varphi_n \circ f$ ist eine wachsende Folge einfacher Funktionen und es folgt aus dem Satz über monotone Konvergenz

$$\int_E g\mathrm{d}\mu^f = \lim_{n\to\infty} \int_E (\varphi_n \circ f)\mathrm{d}\mu = \int_E \lim_{n\to\infty} (\varphi_n \circ f)\mathrm{d}\mu = \int_E (g \circ f)\mathrm{d}\mu.$$

Schritt 4: (i) und (ii) für den allgemeinen Fall folgen aus der Zerlegung $g = g_+ - g_-$ und Schritt 3. □

Die hier verwendete Beweistechnik kommt häufig zum Einsatz und wird mitunter "Lifting-Methode" oder "maßtheoretische Induktion" genannt.

4 L^p-Räume und Radon-Nikodym-Dichten

4.1 Räume integrierbarer Funktionen

Im Folgenden werden wir dem Maßraum $(E, \mathcal{A}, \mu)$ eine Serie von normierten Vektorräumen integrierbarer Funktionen studieren.

Definition 4.1. Für $p \geqslant 1$ bezeichnen wir mit $\mathcal{L}^p(\mu) = \mathcal{L}^p(E, \mathcal{A}, \mu)$ die Menge aller reellwertigen Borel-messbaren Funktionen $f\colon E \to \mathbb{R}$ mit

$$\int_E |f(y)|^p \mu(\mathrm{d}y) < \infty.$$

Für jedes $f \in \mathcal{L}^p(\mu)$ definieren wir

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

Ferner definieren wir $\mathcal{L}^\infty(\mu) := \mathcal{L}^\infty(E, \mathcal{A}, \mu)$ als Menge aller Borel-messbaren Funktionen $f\colon E \to \mathbb{R}$ mit

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

Die $\mathcal{L}^p(\mu)$-Räume erfüllen zwei wesentliche Ungleichungen.

Satz 4.2 (Hölder-Ungleichung). *Es seien $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$ oder $p = 1, q = \infty$. Dann gilt für $f \in \mathcal{L}^p(E, \mathcal{A}, \mu)$ und $g \in \mathcal{L}^q(E, \mathcal{A}, \mu)$, dass $f \cdot g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ und*

$$\int_E |f \cdot g| \mathrm{d}\mu \leqslant \|f\|_p \cdot \|g\|_q.$$

Beweis. Der Fall $p = 1, q = \infty$ ist trivial. Sei nun $p, q > 1$.

Wir zeigen zunächst die Ungleichung ("gewichtete Ungleichung von geometrischen und arithmetischen Mittel")

$$a^\rho b^{1-\rho} \leqslant \rho a + (1-\rho) b \quad \forall a, b \geqslant 0, \rho \in (0,1). \tag{4.1}$$

Die Ungleichung ist klar für $a = 0$ oder $b = 0$. Sind $a, b > 0$, ist (4.1) aufgrund der Monotonie des Logarithmus äquivalent zu

$$\rho \log a + (1-\rho) \log b = \log \left(a^\rho b^{1-\rho}\right) \leqslant \log \left(\rho a + (1-\rho) b\right),$$

was wiederum äquivalent zur Konkavität der Logarithmusfunktion ist. Eine C^2-Funktion f ist aber genau dann konkav, wenn $f'' \leqslant 0$ gilt. Für $f = \log$ ist dies erfüllt, da $f''(x) = -x^{-2}, x > 0$. Somit gilt (4.1).

Falls $\|f\|_p = 0$ oder $\|g\|_p = 0$, so ist f bzw. g μ-f.ü. 0 und die Behauptung gilt. Seien nun $\|f\|_p, \|g\|_q > 0$. Für beliebiges $x \in E$ wenden wir (4.1) auf

$$\rho = \frac{1}{p}, \quad a = \frac{|f(x)|^p}{\|f\|_p^p}, \quad b = \frac{|g(x)|^q}{\|g\|_q^q}$$

an. Also

$$\frac{|f(x)|}{\|f\|_p} \cdot \frac{|g(x)|}{\|g\|_q} \leqslant \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|_q^q}.$$

Integrieren wir über E, folgt

$$\begin{aligned}\frac{1}{\|f\|_p \|g\|_q} \int_E |f(x) g(x)| \mathrm{d}\mu \leqslant & \frac{1}{p\|f\|_p^p} \int |f(x)|^p \mathrm{d}\mu + \frac{1}{q\|g\|_q^q} \int |g(x)|^q \mathrm{d}\mu \\ = & \frac{1}{p} + \frac{1}{q} = 1\end{aligned}$$

und somit $\|f \cdot g\|_1 \leqslant \|f\|_p \cdot \|g\|_q$. □

Daraus ergeben sich zwei wichtige Spezialfälle:

Korollar 4.3.

*(i) Für $f, g \in \mathcal{L}^2(E, \mathcal{A}, \mu)$ gilt $\|f \cdot g\|_1 \leqslant \|f\|_2 \|g\|_2$ (*Cauchy-Schwarz-Ungleichung*).*

*(ii) Ist μ ein endliches Maß, so existiert für jedes $1 \leqslant p' \leqslant p$ eine Konstante $C_{p,p'} > 0$, so dass $\|f\|_{p'} \leqslant C_{p,p'} \|f\|_p$ (*stetige Einbettung von $\mathcal{L}^p(\mu)$ in $\mathcal{L}^{p'}(\mu)$*).*

Die zweite Ungleichung ist die Dreiecksungleichung für $\mathcal{L}^p(\mu)$-Räume.

Satz 4.4 (Minkowski-Ungleichung). *Falls $p \in [1, \infty]$ und $f, g \in \mathcal{L}^p(E, \mathcal{A}, \mu)$, so gilt $f + g \in \mathcal{L}^p(E, \mathcal{A}, \mu)$ und*

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

Beweis. Für $p = 1$ und $p = \infty$ ist die Behauptung trivial. Sei $p \in (1, \infty)$. Für beliebige $a, b \in \mathbb{R}$ gilt

$$\begin{aligned}|a + b|^p \leqslant (|a| + |b|)^p \leqslant & \big(2 \cdot \max\{|a|, |b|\}\big)^p \\ = & 2^p \cdot \max\{|a|^p, |b|^p\} \leqslant 2^p (|a|^p + |b|^p).\end{aligned}$$

Setzen wir nun $a = f(x)$ und $b = g(x)$ und integrieren beide Seiten über $x \in E$, erhalten wir

$$\int_E |f(x) + g(x)|^p \mathrm{d}\mu \leqslant 2^p \Big(\int_E |f(x)|^p \mathrm{d}\mu + \int_E |g(x)|^p \mathrm{d}\mu \Big) < \infty.$$

Somit ist $f + g \in \mathcal{L}^p(\mu)$ gezeigt. Um die Dreiecksungleichung zu zeigen, können wir o.B.d.A. $\|f + g\|_p > 0$ annehmen. Wir verwenden die Hölder-Ungleichung mit $q = \frac{p}{p-1}$:

$$\begin{aligned}
\int_E |f+g|^p \mathrm{d}\mu &= \int_E |f+g||f+g|^{p-1} \mathrm{d}\mu \\
&\leqslant \int_E |f| \cdot |f+g|^{p-1} \mathrm{d}\mu + \int_E |g| \cdot |f+g|^{p-1} \mathrm{d}\mu \\
&\leqslant \|f\|_p \Big(\int_E |f+g|^{q(p-1)} \mathrm{d}\mu \Big)^{1/q} + \|g\|_p \Big(\int_E |f+g|^{q(p-1)} \mathrm{d}\mu \Big)^{1/q} \\
&= \big(\|f\|_p + \|g\|_p \big) \Big(\int_E |f+g|^p \mathrm{d}\mu \Big)^{1/q}.
\end{aligned}$$

Somit

$$\|f+g\|_p = \Big(\int_E |f+g|^p \mathrm{d}\mu \Big)^{1-1/q} \leqslant \|f\|_p + \|g\|_p. \qquad \square$$

Aus der Minkowski-Ungleichung ergibt sich, dass für jedes $p \in [1, \infty]$ die Menge $\mathcal{L}^p(E, \mathcal{A}, \mu)$ einen *Vektorraumaum* bildet, d.h.

$$f, g \in \mathcal{L}^p(\mu), \alpha, \beta \in \mathbb{R} \quad \Rightarrow \quad \alpha f + \beta g \in \mathcal{L}^p(\mu),$$

und $\|\cdot\|_p$ eine Halbnorm auf $\mathcal{L}^p(E, \mathcal{A}, \mu)$ ist:

$$\|f\|_p \geqslant 0, \quad \|f+g\|_p \leqslant \|f\|_p + \|g\|_p, \quad \|\alpha f\|_p = |\alpha| \cdot \|f\|_p \quad \forall f, g \in \mathcal{L}^p(\mu), \lambda \in \mathbb{R}.$$

$\|\cdot\|_p$ ist jedoch keine Norm auf $\mathcal{L}^p(\mu)$, da aus $\|f\|_p = 0$ lediglich $f = 0$ μ-f.ü. folgt. Um einen normierten Vektorraum zu erhalten, betrachtet man die Mengen von Äquivalenzklassen bezüglich der Äquivalenzrelation

$$f \sim g :\Leftrightarrow f = g \ \mu\text{-fast überall}$$

auf $\mathcal{L}^p(\mu)$.

Definition 4.5. Sei $p \in [1, \infty)$. Der Quotientenraum der Äquivalenzklassen $[f] := \{g \in \mathcal{L}^p(\mu) : f \sim g\}$ wird bezeichnet mit

$$L^p(\mu) := L^p(E, \mathcal{A}, \mu) = \{[f] : f \in \mathcal{L}^p(E, \mathcal{A}, \mu)\} = {}^{\mathcal{L}^p(E,\mathcal{A},\mu)}\big/_{\{f\colon E \to \mathbb{R} | f \ \mathcal{A}\text{-messbar und } f=0 \ \mu\text{-f.ü.}\}}$$

und versehen mit der Norm

$$\|[f]\|_p := \Big(\int_E |g|^p \mathrm{d}\mu \Big)^{1/p} \quad \text{für einen beliebigen Repräsentanten } g \in [f].$$

Aus Gründen der Vereinfachung und weil Verwechslungen kaum möglich sind, nennt man $L^p(\mu), p \in [1, \infty)$, den Raum der p-integrierbaren Funktionen und schreibt dessen Elemente als f statt $[f]$, meint aber eigentlich den Raum der entsprechenden Äquivalenzklassen.

Der normierte Vektorraum $(L^p(E, \mathcal{A}, \mu), \|\cdot\|_p), 1 \leqslant p < \infty$, ist sogar vollständig, d.h. jede Cauchy-Folge konvergiert: Ist $(f_n)_{n \geqslant 1} \subseteq L^p(\mu)$ eine Folge, so dass

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

dann existiert ein $f \in L^p(\mu)$ mit $\lim_{n\to\infty} \|f - f_n\|_p = 0$ (Übung ☑). Wir haben also folgenden Satz bewiesen:

Satz 4.6 (Fischer-Riesz)**.** *Für $p \in [1,\infty)$ ist $(L^p(E,\mathcal{A},\mu), \|\cdot\|_p)$ ein Banachraum (d.h. ein vollständiger, normierter Vektorraum).*

Im Spezialfall $p = 2$ wird die Norm $\|\cdot\|_2$ vom Skalarprodukt (also der positiv definiten Bilinearform)

$$\langle f, g\rangle := \int_E f \cdot g \mathrm{d}\mu, \quad f, g \in L^2(\mu)$$

erzeugt, d.h. $\|f\|_2^2 = \langle f, f\rangle$ für alle $f \in L^2(\mu)$. Damit ist $(L^2(E,\mathcal{A},\mu), \langle\cdot,\cdot\rangle)$ ein (unendlichdimensionaler) *Hilbertraum*. Insbesondere gibt es zu jedem abgeschlossenen, linearen Unterraum $M \subseteq L^2(\mu)$ und jedem $f \in L^2(\mu)$ eine eindeutige Darstellung (*Orthogonalzerlegung*)

$$f = g + h \quad \text{mit} \quad g \in M, h \in L^2(\mu) \text{ mit } \langle h, z\rangle = \langle f - g, z\rangle = 0\,\forall z \in M.$$

Für alle $z \in M$ gilt dann der Satz von Pythagoras

$$\begin{aligned}\|f - z\|_2^2 &= \|f - g + g - z\|_2^2 \\ &= \|f - g\|_2^2 + 2\langle \underbrace{f - g}_{\perp M}, \underbrace{g - z}_{\in M}\rangle + \|g - z\|_2^2 = \|f - g\|_2^2 + \|g - z\|_2^2.\end{aligned}$$

Folglich kann die Funktion g auch durch $\|f - g\|_2 = \inf_{z\in M} \|f - z\|_2$ charakterisiert werden und wird *Orthogonalprojektion von f auf M* genannt. Mit Hilfe der Orthogonalprojektion lässt sich der folgende *Riesz'sche Darstellungssatz* beweisen (vgl. Funktionalanalysis): Eine Abbildung $\varphi\colon L^2(\mu) \to \mathbb{R}$ ist genau dann stetig und linear, falls ein $f \in L^2(\mu)$ existiert, so dass $\varphi(g) = \langle f, g\rangle$ für alle $g \in L^2(\mu)$ gilt.

4.2 Satz von Radon-Nikodym und Lebesguescher Zerlegungssatz

In diesem Abschnitt seien μ und ν stets zwei Maße auf dem messbaren Raum $(E, \mathcal{A})$. In Beispiel 3.18 ist uns bereits folgender Begriff untergekommen:

Definition 4.7. Eine messbare Funktion $f\colon E \to [0,\infty]$ heißt *Dichte von ν bezüglich μ* oder *μ-Dichte von ν*, falls

$$\nu(A) = \int_E f \cdot \mathbb{1}_A \mathrm{d}\mu = \int_A f \mathrm{d}\mu \quad \text{für jedes } A \in \mathcal{A}.$$

Existiert eine Dichte f von ν bzgl. μ, dann schreiben wir auch $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Lemma 4.8. *Sei ν σ-endlich. Sind f_1 und f_2 Dichten von ν bzgl. μ, so gilt $f_1 = f_2$ μ-fast überall.*

Beweis. Sei (E_n) eine Ausschöpfung von E, d.h. $E_n \subseteq E_{n+1}, n \geqslant 1$, mit $\bigcup_{n\geqslant 1} E_n = E$, und $\nu(E_n) < \infty$ für alle $n \geqslant 1$. Setzen wir $A_n := E_n \cap \{f_1 > f_2\}, n \geqslant 1$, dann ist $\nu(A_n) < \infty$, also

$$0 = \nu(A_n) - \nu(A_n) = \int_{A_n} (f_1 - f_2)\mathrm{d}\mu.$$

Folglich gilt $f_1\mathbb{1}_{A_n} = f_2\mathbb{1}_{A_n}$ μ-f.ü. (Satz 3.12(iv)). Da $f_1 > f_2$ auf A_n gilt, erhalten wir $\mu(A_n) = 0$ und damit

$$\mu(\{x \in E : f_1(x) > f_2(x)\}) = \mu\Big(\bigcup_{n\geqslant 1} A_n\Big) = 0.$$

Analog folgt $\mu(\{x \in E : f_1(x) > f_2(x)\}) = 0$, also $f_1 = f_2$ μ-fast überall. □

Satz 4.9.

(i) Hat das Maß ν auf $(E, \mathcal{A})$ eine μ-Dichte f, so ist eine messbare Funktion $h\colon E \to \overline{\mathbb{R}}$ genau dann ν-integrierbar, wenn $h \cdot f$ μ-integrierbar ist. In dem Fall gilt

$$\int_E h \mathrm{d}\nu = \int_E h \cdot f \mathrm{d}\mu.$$

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

(ii) Seien μ, ν, η Maße auf $(E, \mathcal{A})$, wobei ν eine μ-Dichte f besitzt und η eine ν-Dichte g besitzt. Dann ist $f \cdot g$ eine μ-Dichte von η also $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \cdot \frac{\mathrm{d}\eta}{\mathrm{d}\nu} = \frac{\mathrm{d}\eta}{\mathrm{d}\mu}$.

Beweis. Übung (☑). □

Definition 4.10. Seien μ und ν zwei Maße auf $(E, \mathcal{A})$. ν heißt *absolutstetig* bzgl. μ (kurz $\nu \ll \mu$), falls $\nu(A) = 0$ für jede μ-Nullmenge $A \in \mathcal{A}$ gilt. Die Maße μ und ν heißen *äquivalent* (kurz $\mu \sim \nu$), falls $\nu \ll \mu$ und $\mu \ll \nu$ gilt. μ heißt *singulär* zu ν (kurz $\mu \perp \nu$), falls es ein $A \in \mathcal{A}$ gibt mit $\mu(A) = 0$ und $\nu(A^c) = 0$.

Beispiel 4.11.

(i) Wir haben bereits gesehen, dass $\nu \ll \mu$, falls ν eine Dichte bzgl. μ besitzt. Der Satz von Radon-Nikodym wird sogar die Äquivalenz zeigen.

(ii) Auf $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ ist das Dirac-Maß δ_x singulär zum Lebesguemaß λ für jedes $x \in \mathbb{R}$, da $\lambda(\{x\}) = 0$ und $\delta(\mathbb{R} \setminus \{x\}) = 0$.

Satz 4.12 (Radon-Nikodym)**.** *Seien μ ein σ-endliches und ν ein beliebiges Maß auf $(E, \mathcal{A})$. Dann sind folgende Aussagen äquivalent*

(i) ν hat eine Dichte bzgl. μ.

(ii) $\nu \ll \mu$.

In diesem Fall heißt die Dichte Radon-Nikodym-Dichte $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ von ν bzgl. μ.

Beweis. Die Hinrichtung ist bereits bekannt. Wir zeigen nun (ii)⇒(i) im Fall, dass μ und ν endliche Maße sind. Für die Erweiterung auf σ-endliches μ und beliebige ν sei auf Elstrodt (2007, Satz VII.2.3) verwiesen.

Schritt 1: Gilt $\nu \leqslant \mu$, dann besitzt ν eine μ-Dichte $f\colon E \to [0, 1]$.

Wegen $\nu(E) \leqslant \mu(E) < \infty$ gilt $\mathcal{L}^2(\mu) \subseteq \mathcal{L}^2(\nu) \subseteq \mathcal{L}^1(\nu)$ (stetige Einbettungen, Korollar 4.3). Damit ist die lineare Abbildung

$$\mathcal{L}^2(\mu) \ni g \mapsto \int_E g \mathrm{d}\nu$$

wohldefiniert und stetig. Nach dem Darstellungssatz von Riesz existiert nun ein $f \in L^2(\mu)$, so dass

$$\int_E g \mathrm{d}\nu = \langle g, f \rangle = \int_E g \cdot f \mathrm{d}\mu \quad \text{für alle } g \in L^2(\mu).$$

Wählt man $g = \mathbb{1}_A$ für $A \in \mathcal{A}$, so folgt, dass nur noch $0 \leqslant f \leqslant 1$ μ-f.ü. zu zeigen ist, da wir dann eine Dichte $f\colon E \to [0, 1]$ wählen können.

Angenommen $\mu(B) > 0$ für $B := \{f > 1\}$. Dann ist $\nu(B) = \int_B f \mathrm{d}\mu > \mu(B)$: Widerspruch! Analog folgt aus $\mu(C) > 0$ mit $C := \{f < 0\}$, dass $\nu(C) = \int_C f \mathrm{d}\mu < 0$: Widerspruch!

Schritt 2: Seien nun μ, ν beliebige endliche Maße mit $\nu \ll \mu$, dann existiert eine μ-Dichte von ν.

Zum endlichen Maß $\tau := \mu + \nu$ existieren nach Schritt 1 die τ-Dichten $g, h\colon E \to [0, 1]$ von μ bzw. ν. Für $N = \{g = 0\}$ gilt $\mu(N) = \int_N g \mathrm{d}\tau = 0$ und wegen $\nu \ll \mu$ auch $\nu(N) = 0$. Die Funktion

$$f(x) := \begin{cases} h(x)/g(x), & \text{für } x \in N^c, \\ 0, & \text{für } x \in N \end{cases}$$

ist nichtnegativ, $\mathcal{A}$-messbar und für alle $A \in \mathcal{A}$ gilt

$$\nu(A) = \nu(A \cap N^c) = \int_{A\cap N^c} h \mathrm{d}\tau = \int_{A\cap N^c} f \cdot g \mathrm{d}\tau = \int_{A\cap N^c} f \mathrm{d}\mu = \int_A f \mathrm{d}\mu. \qquad \square$$

Satz 4.13 (Lebesguescher Zerlegungssatz). *Sind μ und ν σ-endliche Maße auf $(E, \mathcal{A})$, so gibt es genau eine Zerlegung*

$$\nu = \rho + \eta$$

von ν in zwei σ-endliche Maße ρ und η auf $\mathcal{A}$ mit $\rho \ll \mu$ und $\eta \perp \mu$.

Beweis. Existenz: Das Maß $\tau := \mu + \nu$ ist σ-endlich mit $\mu \ll \tau$. Nach dem Satz von Radon-Nikodym gibt es eine τ-Dichte $f \in \mathcal{M}^+$ von μ. Wir setzen $N := \{f = 0\}$ und definieren ρ und η vermöge

$$\rho(A) := \nu(A \cap N^c), \quad \eta(A) := \nu(A \cap N), \quad A \in \mathcal{A}.$$

Dann sind ρ, η σ-endliche Maße mit $\nu = \rho + \eta$. Wegen $\mu(N) = \int_N f \mathrm{d}\tau = 0$ und $\eta(N^c) = 0$, gilt $\mu \perp \eta$. Es bleibt $\rho \ll \mu$ zu zeigen. Sei dazu A eine μ-Nullmenge, also $0 = \mu(A) = \int_E f \mathbb{1}_A \mathrm{d}\tau$. Damit ist $f \mathbb{1}_A = 0$ τ-f.ü., d.h. $\tau(A \cap N^c) = 0$. Wegen $\rho(A) = \nu(A \cap N^c) \leqslant \tau(A \cap N^c) = 0$ ist also auch $\rho(A) = 0$.

Eindeutigkeit: Sei $\nu = \rho' + \eta'$ eine zweite Zerlegung mit $\rho' \ll \mu$ und $\eta' \perp \mu$. Dann existieren μ-Nullmengen $N, N' \in \mathcal{A}$ mit $\eta(N^c) = 0$ und $\eta'(N'^c) = 0$. Für alle $A \in \mathcal{A}$ gilt daher

$$\rho(A) = \rho(A \cap N^c \cap N'^c) = \nu(A \cap N^c \cap N'^c) = \rho'(A \cap N^c \cap N'^c) = \rho'(A).$$

Weiter ist

$$\eta(A) = \eta(A \cap N) = \nu(A \cap N) = \nu(A \cap N \cap N'),$$

denn $\nu(A \cap N \cap N'^c) = \rho'(A \cap N \cap N'^c) + \eta'(A \cap N \cap N'^c)$ und $\rho' = \rho$. Aus Symmetriegründen erhalten wir $\eta(A) = \nu(A \cap N \cap N') = \eta'(A)$. $\square$

Beispiel 4.14. Sei $F\colon \mathbb{R} \to [0,1]$ die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes P auf $(\mathbb{R}, \mathscr{B}_\mathbb{R})$. Nach dem Lebesgueschen Zerlegungungssatz können wir $P = P_a + P_s$ in einen bzgl. dem Lebesguemaß absolut stetigen Teil $P_a \ll \lambda$ und einen singulären Teil $P_s \perp \lambda$. zerlegen. Angenommen F ist stückweise (schwach) differenzierbar mit der Ableitung F' und höchstens abzählbar vielen Sprungstellen $(x_k)_{k\geqslant 1} \subseteq \mathbb{R}$. Setzen wir $F'(x_k) = 0, k \geqslant 1$, so ist F' die Radon-Nikodym-Dichte $\frac{\mathrm{d}P_a}{\mathrm{d}\lambda}$ von P_a bzgl. λ. Für den singulären Teil finden wir die Darstellung $P_s = \sum_{k\geqslant 1}(F(x_k) - \lim_{\varepsilon\downarrow 0} F(x_k - \varepsilon))\delta_{x_k}$.

5 Produkträume und Übergangskerne

Produktmengen und Produkt-σ-Algebren dienen in der Wahrscheinlichkeitstheorie zur Modellierung mehrerer gleichzeitig oder nacheinander ausgeführter zufälliger Versuche. In der Stochastik spielt der Begriff der Unabhängigkeit für Ereignisse bzw. Zufallsvariablen eine zentrale Rolle. Die gemeinsame Verteilung voneinander unabhängiger Zufallsgrößen ist von spezieller Gestalt, es ist eine sogenannte Produktverteilung oder, in der Sprache der Maßtheorie, ein Produktmaß. Letzteres wird im ersten Abschnitt eingeführt.

5.1 Endliche Produktmaße und der Satz von Fubini

Definition 5.1. Sei $(E_i)_{i\in I}$ eine beliebige Familie von Mengen. Der *Produktraum* $E = \times_{i\in I} E_i$ bezeichnet die Menge der Abbildungen $x\colon I \to \bigcup_{i\in I} E_i$ mit der Eigenschaft, dass $x(i) \in E_i$ für jedes $i \in I$ gilt. Sind alle E_i gleich einem E_0, so schreiben wir auch $E = E_0^I$, bzw. $E = E_0^n$ für den Spezialfall $I = \{1, \ldots, n\}$. Ist $i \in I$, so heißt $\pi_i : E \to E_i, x \mapsto x(i)$ die i-te *Koordinatenabbildung* oder *Projektion auf die* i *-te Koordinate.*

Definition 5.2. Seien $(E_i, \mathcal{A}_i), i \in I$, messbare Räume. Die *Produkt-σ-Algebra* $\mathcal{A} = \bigotimes_{i\in I} \mathcal{A}_i$ ist die kleinste σ-Algebra auf E, so dass für alle $i \in I$ die Projektion π_i $(\mathcal{A}, \mathcal{A}_i)$-messbar ist:

$$\mathcal{A} = \sigma(\pi_i, i \in I) = \sigma\big(\pi_i^{-1}(\mathcal{A}_i), i \in I\big).$$

Ist $(E_i, \mathcal{A}_i) = (E_0, \mathcal{A}_0)$ für jedes $i \in I$, schreiben wir $\mathcal{A} = \mathcal{A}_0^{\otimes I}$.

Beispiel 5.3.

(i) Für $i = 1, \dots n$ und $n \in \mathbb{N}$ seien $\mathcal{E}_i$ Erzeuger der σ-Algebren $\mathcal{A}_i = \sigma(\mathcal{E}_i)$ auf einer Menge E_i und es existiere eine Folge $(A_{i,n})_{n\geqslant 1} \subseteq \mathcal{E}_i$ mit $\bigcup_{n\geqslant 1} A_{i,n} = E_i$. Dann wird $\bigotimes_{i=1}^n \mathcal{A}_i$ vom System $\bigtimes_{i=1}^n \mathcal{E}_i$ erzeugt, d.h.

$$\bigotimes_{i=1}^n \mathcal{A}_i = \sigma\big(\{\bigtimes_{i=1}^n A_i : A_i \in \mathcal{E}_i, i = 1, \dots, n\}\big).$$

Insbesondere ist $\bigotimes_{i=1}^n \mathcal{A}_i = \sigma(\{\bigtimes_{i=1}^n A_i : A_i \in \mathcal{A}_i, i = 1, \dots, n\})$.

(ii) Ist $(E, \mathcal{A}) = (\mathbb{R}^n, \mathscr{B}_{\mathbb{R}^n})$ und $(F, \mathcal{B}) = (\mathbb{R}^m, \mathscr{B}_{\mathbb{R}^m})$ mit $n, m \in \mathbb{N}$, so ist $E \times F = \mathbb{R}^{n+m}$ und $\mathscr{B}_{\mathbb{R}^n} \otimes \mathscr{B}_{\mathbb{R}^m} = \mathscr{B}_{\mathbb{R}^{n+m}}$ (denn die Produkttopologie ist ganz analog zur Produkt-σ-Algebra definiert).

(iii) Für beliebige abzählbare Räume E und F gilt $\mathscr{P}(E) \otimes \mathscr{P}(F) = \mathscr{P}(E \times F)$.

Lemma 5.4. *Seien $(E_i, \mathcal{A}_i)$ für $i = 1, 2$ zwei messbare Räume, $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ und $f\colon E_1 \times E_2 \to \overline{\mathbb{R}}$ eine $\mathcal{A}_1 \otimes \mathcal{A}_2$-messbare Funktion. Dann gilt für jedes $y_1 \in E_1$ und $y_2 \in E_2$*

$$\begin{aligned}
&A_{y_1} := \{x_2 \in E_2 : (y_1, x_2) \in A\} \in \mathcal{A}_2,\\
&A_{y_2} := \{x_1 \in E_1 : (x_1, y_2) \in A\} \in \mathcal{A}_1,\\
&f_{y_1}\colon E_2 \to \overline{\mathbb{R}}, \quad x_2 \mapsto f(y_1, x_2) \quad \textit{ist } \mathcal{A}_2\textit{-messbar},\\
&f_{y_2}\colon E_1 \to \overline{\mathbb{R}}, \quad x_1 \mapsto f(x_1, y_2) \quad \textit{ist } \mathcal{A}_1\textit{-messbar.}
\end{aligned}$$

*A_{y_1} bzw. f_{y_1} heißt y_1-*Schnitt* der Menge A bzw. der Funktion f.*

Beweis. Für y_1 definiere die Einbettung $i\colon E_2 \to E_1 \times E_2, x_2 \mapsto (y_1, x_2)$. Da $\pi_1 \circ i$ konstant gleich y_1 ist (also $\mathcal{A}_1$-messbar) und $\pi_2 \circ i$ die Identität auf E_2 ist (also $\mathcal{A}_2$-messbar), ist die Abbildung i $(\mathcal{A}_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$-messbar. Folglich ist $A_{y_1} = i^{-1}(A) \in \mathcal{A}_2$ und $f_{y_1} = f \circ i$ $\mathcal{A}_2$-messbar. □

Nachdem wir nun den Produktraum mit einer σ-Algebra versehen haben, wollen nun auf diesem messbaren Raum das so genannte Produktmaß einführen. Hierzu benötigen wir zuerst noch ein Hilfslemma.

Lemma 5.5. *Seien $(E_i, \mathcal{A}_i, \mu_i), i = 1, 2$, σ-endliche Maßräume und $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Dann sind die Abbildungen $\varphi_1^A\colon E_1 \to [0, \infty], y_1 \mapsto \mu_2(A_{y_1})$ und $\varphi_2^A\colon E_2 \to [0, \infty], y_2 \mapsto \mu_1(A_{y_2})$ $(\mathcal{A}_1, \mathscr{B}_{\overline{\mathbb{R}}})$- bzw. $(\mathcal{A}_2, \mathscr{B}_{\overline{\mathbb{R}}})$-messbar.*

Beweis. Schritt 1: Nehmen wir zunächst an, dass $\mu_2(E_2) < \infty$. Dann ist das Mengensystem

$$\mathcal{D} := \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : \varphi_1^A \text{ ist } \mathcal{A}_1\text{-messbar}\}$$

ein Dynkin-System:

(i) $E_1 \times E_2 \in \mathcal{D}$, da $\varphi_1^{E_1\times E_2} = \mu_2(E_2) < \infty$ konstant ist.

(ii) Sind $A, B \in \mathcal{D}$ mit $A \subseteq B$, so folgt $A_{y_1} \subseteq B_{y_1}$ und

$$\varphi_1^B(y_1) = \int_{E_2} \mathbb{1}_{B_{y_1}} \mathrm{d}\mu_2 = \int_{E_2} \big(\mathbb{1}_{A_{y_1}} + \mathbb{1}_{(B\setminus A)_{y_1}}\big) \mathrm{d}\mu_2 = \varphi_1^A(y_1) + \varphi_1^{B\setminus A}(y_1).$$

Da μ_2 ein endliches Maß ist, folgt die Messbarkeit von $\varphi_1^{B\setminus A} = \varphi_1^B - \varphi_1^A$ und damit $B \setminus A \in \mathcal{D}$.

(iii) Sind $(A_n)_{n\geqslant 1} \subseteq \mathcal{D}$ disjunkte Mengen, so gilt $\mathbb{1}_{(\cup_n A_n)_{y_1}} = \mathbb{1}_{\cup_n (A_n)_{y_1}} = \sum_{n\geqslant 1} \mathbb{1}_{(A_n)_{y_1}}$, so dass $\varphi_1^{\cup_n A_n} = \sum_{n\geqslant 1} \varphi_1^{A_n}$.

Weiter ist $\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{A}_i, i = 1, 2\}$ $\cap$-stabil und $\mathcal{E} \subseteq \mathcal{D}$, da

$$\varphi_1^{A_1\times A_2}(y_1) = \mu_2\big((A_1 \times A_2)_{y_1}\big) = \mu_2(A_2)\mathbb{1}_{A_1}(y_1).$$

Nach Satz 2.22 folgt $\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{E}) = \delta(\mathcal{E}) \subseteq \mathcal{D} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$.

Schritt 2: Ist nun μ_2 σ-endlich, so existiert eine wachsende Folge $(B_n)_{n\geqslant 1} \subseteq \mathcal{A}_2$ mit $B_n \uparrow E_2$ und $\mu_2(B_n) < \infty$. Dann ist $\mu_{2,n} \colon \mathcal{A}_2 \to [0, \infty)$ mit $\mu_{2,n}(C) := \mu_2(C \cap B_n), C \in \mathcal{A}_2$, ein endliches Maß für jedes $n \geqslant 1$ und

$$\varphi_1^A(y_1) = \mu_2(A_{y_1}) = \lim_{n\to\infty} \mu_{2,n}(A_{y_1}).$$

Nach Schritt 1 ist $y_1 \mapsto \mu_{2,n}(A_{y_1})$ messbar für jedes $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ und somit auch φ_1^A. $\square$

Satz 5.6. *Es seien $(E_i, \mathcal{A}_i, \mu_i)$ für $i = 1, \ldots, n$ σ-endliche Maßräume. Dann existiert genau ein σ-endliches Maß μ auf $\mathcal{A} := \bigotimes_{i=1}^n \mathcal{A}_i$, so dass*

$$\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mu_i(A_i) \quad \textit{für alle } A_i \in \mathcal{A}_i, \quad i = 1, \ldots, n.$$

$\bigotimes_{i=1}^n \mu_i := \mu_1 \otimes \cdots \otimes \mu_n := \mu$ heißt Produktmaß *der μ_i. Falls alle Maßräume gleich einem Maßraum $(E_0, \mathcal{A}_0, \mu_0)$ sind, so schreiben wir kurz $\mu_0^{\otimes n} := \bigotimes_{i=1}^n \mu_i$.*

Beweis. Eindeutigkeit: μ ist für alle Mengen aus

$$\mathcal{E} := \big\{A_1 \times \cdots \times A_n : A_i \in \mathcal{A}_i \text{ für } i = 1, \ldots, n\big\}$$

festgelegt. Das Mengensystem $\mathcal{E}$ ist $\cap$-stabil und nach Beispiel 5.3(i) gilt $\mathcal{A} = \sigma(\mathcal{E})$. Für jedes $i = 1, \ldots, n$ seien $(B_{i,k})_{k\geqslant 1} \subseteq \mathcal{A}_i$ wachsende Überdeckungen von E_i mit $B_{i,k} \uparrow E_i$ und $\mu_i(B_{i,k}) < \infty$ für alle $k \geqslant 1$. Dann ist

$$\Big(\bigtimes_{i=1}^n B_{i,k}\Big)_{k\geqslant 1} \subseteq \mathcal{E}$$

eine wachsende Überdeckung von $E = \bigtimes_{i=1}^n E_i$ mit $\mu(\bigtimes_{i=1}^n B_{i,k}) = \prod_{i=1}^n \mu_i(B_{i,k}) < \infty$ für alle $k \geqslant 1$. Aus Satz 2.23 folgt die Eindeutigkeit.

Existenz: Die Existenz folgt induktiv aus dem Fall $n = 2$. Für $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ folgt aus dem vorherigen Lemma, dass $\varphi_1^A \colon E_1 \to [0, \infty]$ $\mathcal{A}_1$-messbar und nichtnegativ ist. Wir definieren

$$\mu(A) := \int_{E_1} \varphi_1^A \mathrm{d}\mu_1 = \int_{E_1} \mu_2(A_{y_1})\mathrm{d}\mu_1(y_1).$$

Dann ist μ ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ (Übung ☑) und für $A = A_1 \times A_2$ mit $A_i \in \mathcal{A}_i$ gilt

$$\begin{aligned}\mu(A_1 \times A_2) &= \int_{E_1} \mu_2\big((A_1 \times A_2)_{y_1}\big)\mathrm{d}\mu_1(y_1)\\ &= \int_{E_1} \mu_2(A_2)\mathbb{1}_{A_1}(y_1)\mathrm{d}\mu_1(y_1) = \mu_2(A_2)\mu_1(A_1).\end{aligned}$$ $\square$

Aus dem letzten Beweis lernen wir, dass das Produktmaß einer Menge $A \in \mathcal{A} \otimes \mathcal{B}$ berechnet werden kann, indem man über die Maße der x_1- bzw. x_2-Schnitte integriert:

$$\mu_1 \otimes \mu_2(A) = \int_{E_1} \mu_2(A_{x_1})\mathrm{d}\mu_1(x_1) = \int_{E_2} \mu_1(A_{x_2})\mathrm{d}\mu_2(x_2).$$

Dies nennt man das "*Prinzip von Cavalieri*". Da $\mu_2(A_{x_1}) = \int_{E_2} \mathbb{1}_A(x_1,\cdot)\,\mathrm{d}\mu_2$, können wir die letzte Zeile auch schreiben als

$$\begin{aligned}\int_{E_1\times E_2} \mathbb{1}_A \mathrm{d}(\mu_1\otimes\mu_2) &= \int_{E_1\times E_2} \mathbb{1}_A(x_1,x_2)\mathrm{d}(\mu_1\otimes\mu_2)(x_1,x_2)\\ &= \int_{E_1}\Big(\int_{E_2} \mathbb{1}_A(x_1,x_2)\mathrm{d}\mu_2(x_2)\Big)\mathrm{d}\mu_1(x_1)\\ &= \int_{E_2}\Big(\int_{E_1} \mathbb{1}_A(x_1,x_2)\mathrm{d}\mu_1(x_1)\Big)\mathrm{d}\mu_2(x_2). \end{aligned}\tag{5.1}$$

Man kann also die Reihenfolge der Integration von Indikatorfunktionen über Produktmaße vertauschen. Dieses Resultat wollen wir nun auf allgemeine messbare Funktionen verallgemeinern.

Satz 5.7 (Satz von Tonelli)**.** *Es seien $(E_i,\mathcal{A}_i,\mu_i)$ für $i=1,2$ σ-endliche Maßräume und $f\colon E_1\times E_2\to[0,\infty]$ eine nichtnegative $(\mathcal{A}_1\otimes\mathcal{A}_2,\mathscr{B}_{\overline{\mathbb{R}}})$-messbare Funktion. Dann sind die Funktionen*

$$\begin{aligned} E_1\ni x_1 &\mapsto \int_{E_2} f(x_1,\cdot)\mathrm{d}\mu_2\in[0,\infty] \quad \mathcal{A}_1\text{-}\textit{messbar},\\ E_2\ni x_2 &\mapsto \int_{E_1} f(\cdot,x_2)\mathrm{d}\mu_1\in[0,\infty] \quad \mathcal{A}_2\text{-}\textit{messbar}\end{aligned}$$

und es gilt

$$\begin{aligned}\int_{E_1\times E_2} f\,\mathrm{d}(\mu_1\otimes\mu_2) &= \int_{E_1}\Big(\int_{E_2} f(x_1,x_2)\mathrm{d}\mu_2(x_2)\Big)\mathrm{d}\mu_1(x_1)\\ &= \int_{E_2}\Big(\int_{E_1} f(x_1,x_2)\mathrm{d}\mu_1(x_1)\Big)\mathrm{d}\mu_2(x_2).\end{aligned}$$

Beweis. Beachte, dass für jedes $x_1\in E_1$ und $x_2\in E_2$ die Integrale

$$\int_{E_2} f(x_1,\cdot)\mathrm{d}\mu_2 \quad\text{und}\quad \int_{E_1} f(\cdot,x_2)\mathrm{d}\mu_1$$

wegen Lemma 5.4 existieren (und ggf. gleich ∞ sind). Wir verwenden maßtheoretische Induktion:

Schritt 1: Für Indikatorfunktionen $f=\mathbb{1}_A$ mit $A\in\mathcal{A}_1\otimes\mathcal{A}_2$ haben wir die Messbarkeit der Abbildungen in Lemma 5.5 und die Gleichheit bereits in (5.1) gezeigt.

Schritt 2: Sei nun $f=\sum_{k=1}^n a_k\mathbb{1}_{A_k}$ für $a_1,\dots,a_m\in\mathbb{R}, A_1,\dots,A_m\in\mathcal{A}_1\otimes\mathcal{A}_2$. Die Abbildung

$$x_1\mapsto\int_{E_2} f(x_1,\cdot)\mathrm{d}\mu_2=\sum_{k=1}^n a_k\int_{E_2}\mathbb{1}_{A_k}(x_1,\cdot)\mathrm{d}\mu_2$$

ist $\mathcal{A}_1$-messbar nach Lemma 5.5. Die Gleichheit der Integrale folgt, dann aus der Linearität des Integrals und (5.1).

Schritt 3: Sei $f\geqslant 0$ und $(\varphi_n)_{n\geqslant1}$ eine monoton wachsende Folge nichtnegativer, einfacher Funktionen mit $\varphi_n\uparrow f$. Für festes $x_1\in E_1$ gilt dann auch $\varphi_n(x_1,\cdot)\uparrow f(x_1,\cdot)$ und

$$\psi_n : E_1\to[0,\infty],\quad x_1\mapsto\int_{E_2}\varphi_n(x_1,\cdot)\mathrm{d}\mu_2$$

ist $\mathcal{A}_1$-messbar nach Schritt 2. Weiter gilt $\lim_{n\to\infty}\psi_n(x_1)=\int_{E_2} f(x_1,\cdot)\mathrm{d}\mu_2$ für jedes $x_1\in E_1$ nach dem Satz über monotone Konvergenz. Erneute Anwendung von Beppo Levi und in Kombination mit Schritt 2 liefert

$$\begin{aligned}\int_{E_1}\Big(\int_{E_2} f(x_1,x_2)\mathrm{d}\mu_2(x_2)\Big)\mathrm{d}\mu_1(x_1) &= \lim_{n\to\infty}\int_{E_1}\psi_n\mathrm{d}\mu_1\\ &= \lim_{n\to\infty}\int_{E_1}\Big(\int_{E_2}\varphi_n(x_1,\cdot)\mathrm{d}\mu_2\Big)\mathrm{d}\mu_1(x_1)\\ &= \lim_{n\to\infty}\int_{E_1\times E_2}\varphi_n\mathrm{d}(\mu_1\otimes\mu_2)\\ &= \int_{E_1\otimes E_2} f\,\mathrm{d}(\mu_1\otimes\mu_2).\end{aligned}$$ $\square$

Satz 5.8 (Satz von Fubini)**.** *Es seien $(E_i,\mathcal{A}_i,\mu_i)$ für $i=1,2$ σ-endliche Maßräume. Für jede Funktion $f\in\mathcal{L}^1(E_1\times E_2,\mathcal{A}_1\otimes\mathcal{A}_2,\mu_1\otimes\mu_2)$ sind die Funktionen*

$$E_1\ni x_1\mapsto\int_{E_2} f(x_1,\cdot)\mathrm{d}\mu_2\quad \textit{bzw.}\quad E_2\ni x_2\mapsto\int_{E_1} f(\cdot,x_2)\mathrm{d}\mu_1$$

μ_1-fast überall bzw. μ_2-fast überall endlich, $\mathcal{A}_1$-messbar bzw. $\mathcal{A}_2$-messbar und es gilt

$$\begin{aligned}\int_{E_1\times E_2} f(x_1,x_2)\,\mathrm{d}(\mu_1\otimes\mu_2)(x_1,x_2) &= \int_{E_1}\Big(\int_{E_2} f(x_1,x_2)\mathrm{d}\mu_2(x_2)\Big)\mathrm{d}\mu_1(x_1)\\ &= \int_{E_2}\Big(\int_{E_1} f(x_1,x_2)\mathrm{d}\mu_1(x_1)\Big)\mathrm{d}\mu_2(x_2).\end{aligned}\tag{5.2}$$

Beweis. Man wende den Satz von Tonelli auf die Zerlegung $f=f_+-f_-$ an (Übung ☑). $\square$

Bemerkung 5.9.

(i) Aus der Existenz der iterierten Integrale auf der rechten Seite von (5.2) folgt weder die Existenz des Integrals $\int_{E_1\times E_2} f\mathrm{d}(\mu_1\otimes\mu_2)$ noch ihre Gleichheit.

(ii) Wenn $f(x_1,x_2)=f_1(x_1)\cdot f_2(x_2), x_1\in E_1, x_2\in E_2$ für zwei Funktionen $f_i\in\mathcal{L}^1(E_i,\mathcal{A}_i,\mu_i), i=1,2$, dann gilt

$$\int_{E_1\times E_2} f_1\cdot f_2\mathrm{d}(\mu_1\otimes\mu_2)=\int_{E_1} f_1\mathrm{d}\mu_1\cdot\int_{E_2} f_2\mathrm{d}\mu_2.$$

5.2 Anwendung: Stochastische Unabhängigkeit

Wir betrachten einen Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$. Aus der mathematischen Stochastik ist bereits bekannt, dass zwei Ereignisse $A,B\in\mathcal{A}$ *unabhängig* sind, falls $P(A\cap B)=P(A)P(B)$. Allgemeiner gilt:

Definition 5.10. Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum. Für eine beliebige Indexmenge I und Mengensysteme $\mathcal{S}_i\subseteq\mathcal{A}, i\in I$, heißt die Familie $(\mathcal{S}_i)_{i\in I}$ *unabhängig*, falls für jede endliche Teilmenge $J\subseteq I$ und für jede Wahl von $E_j\in\mathcal{S}_j, j\in J$, gilt, dass

$$P\Big(\bigcap_{j\in J}E_j\Big)=\prod_{j\in J}P(E_j).$$

Sind $(\Omega_i,\mathcal{A}_i)$ messbare Räume und $X_i\colon(\Omega,\mathcal{A})\to(\Omega_i,\mathcal{A}_i)$ Zufallsvariablen für jedes $i\in I$, dann nennen wir $(X_i)_{i\in I}$ *unabhängig*, falls die Familie der erzeugten σ-Algebren $(\sigma(X_i))_{i\in I}$ unabhängig ist $(\sigma(X_i)=X_i^{-1}(\mathcal{A}_i))$.

Bemerkung 5.11. Ist $(\mathcal{E}_i\cup\{\emptyset\})$ $\cap$-stabil für jedes $i\in I$, dann gilt (Übung ☑)

$$(\mathcal{E}_i)_{i\in I}\text{ ist unabhängig}\quad\Longleftrightarrow\quad(\sigma(\mathcal{E}_i))_{i\in I}\text{ ist unabhängig.}$$

Sind $\mathcal{E}_i$ $\cap$-stabile Erzeuger von $\mathcal{A}_i$, so sind insbesondere die Zufallsvariablen $X_i\colon(\Omega,\mathcal{A})\to(\Omega_i,\mathcal{A}_i), i\in I$, genau dann unabhängig, wenn $(X_i^{-1}(\mathcal{E}_i))_{i\in I}$ unabhängig ist.

Satz 5.12. *Es seien* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum,* $(\Omega_i, \mathcal{A}_i)$ *messbare Räume und* $X_i \colon \Omega \to \Omega_i$ *Zufallsvariablen,* $i = 1, \ldots, n$ *für ein* $n \in \mathbb{N}$*. Dann sind* $X_1, \ldots, X_n$ *unabhängig genau dann, wenn die gemeinsame Verteilung das Produktmaß der Marginalverteilungen ist, d.h.* $P^{(X_1,\ldots,X_n)} = P^{X_1} \otimes \cdots \otimes P^{X_n}$*.*

Beweis. Nach Definition ist $(\sigma(X_i))_{i=1,\ldots,n}$ genau dann unabhängig, wenn für alle $A_i \in \mathcal{A}_i, i = 1, \ldots, n$, gilt

$$\underbrace{P\Big(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\Big)}_{P^{(X_1,\ldots,X_n)}(A_1\times\cdots\times A_n)} = \prod_{i=1}^{n} \underbrace{P(X_i^{-1}(A_i))}_{P^{X_i}(A_i)}$$

Nach Satz 5.6 ist dies aber gerade äquivalent zu $P^{(X_1,\ldots,X_n)} = P^{X_1} \otimes \cdots \otimes P^{X_n}$. □

Aus diesem Satz können wir insbesondere folgern, dass $X_1, \ldots, X_n$ genau dann unabhängig sind, wenn die Verteilungsfunktion des Zufallvektors $(X_1, \ldots, X_n)$ das Produkt der Verteilungsfunktionen der X_i sind und, falls Dichten existieren, die Dichte von $(X_1, \ldots, X_n)$ das Produkt der Dichten ist (Übung ☑).

Aus dem Satz von Fubini (Bemerkung 5.9) folgt außerdem:

Korollar 5.13. *Es seien* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum und* $X_i \colon \Omega \to \overline{\mathbb{R}}$ *reellwertige Zufallsvariablen für* $i = 1, 2$*. Sind* X_1 *und* X_2 *unabhängig mit endlichem Erwartungswert* $E[|X_i|] < \infty$*, so gilt* $E[X_1 \cdot X_2] = E[X_1] \cdot E[X_2]$*.*

5.3 Übergangskerne und abzählbare Produkte

Nachdem wir Produktmaße studiert haben, welche stochastische Unabhängigkeit modellieren, wollen wir nun unserer Resultate auf abhängige Experimente erweitern.

Definition 5.14. Seien $(E_i, \mathcal{A}_i), i = 1, 2$, zwei messbare Räume. Eine Abbildung $\kappa \colon E_1 \times \mathcal{A}_2 \to [0, \infty]$ heißt *Übergangskern* von $(E_1, \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$, falls

(i) $x_1 \mapsto \kappa(x_1, A_2)$ eine $\mathcal{A}_1$-messbare Abbildung ist für jedes $A_2 \in \mathcal{A}_2$ und

(ii) $A_2 \mapsto \kappa(x_1, A_2)$ ein Maß auf $\mathcal{A}_2$ ist für jedes $x_1 \in E_1$.

κ heißt *(σ-)endlich*, falls $\kappa(x_1, \cdot)$ für jedes $x_1 \in E_1$ ein (σ-)endliches Maß ist. Ist $\kappa(x_1, \cdot)$ für alle $x_1 \in E_1$ ein Wahrscheinlichkeitsmaß, so heißt κ *Markovkern*.

Lemma 5.15. *Seien* $(E_i, \mathcal{A}_i), i = 1, 2$, *messbare Räume,* κ *ein Übergangskern von* E_1 *nach* E_2 *und* $f \colon E_1 \times E_2 \to [0, \infty]$ $\mathcal{A}_1 \otimes \mathcal{A}_2$*-messbar. Dann ist die Funktion*

$$I_f \colon E_1 \to [0, \infty], \quad x_1 \mapsto \int_{E_2} f(x_1, x_2) \kappa(x_1, \mathrm{d}x_2)$$

wohldefiniert und $\mathcal{A}_1$*-messbar.*

Beweis. Ist $f = \mathbb{1}_A$ für ein $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, folgt die Messbarkeit von I_f exakt wie in Lemma 5.5 durch Betrachtung des Dynkin-Systems

$$\mathcal{D} := \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : I_{\mathbb{1}_A} \text{ ist } \mathcal{A}_1\text{-messbar}\}.$$

Folglich ist I_f messbar für jede einfache Funktion und über den Satz über monotone Konvergenz folgt die Messbarkeit für $f \in \mathcal{M}^+$, da $I_f(x_1) = \lim_{n\to\infty} I_{\varphi_n}(x_1)$ für eine approximierende Folge einfacher Funktionen $\varphi_n \uparrow f$ und jedes feste $x_1 \in E_1$. □

Satz 5.16. *Seien $(E_i, \mathcal{A}_i), i = 0, 1, 2$, messbare Räume und κ_1 ein Übergangskern von $(E_0, \mathcal{A}_0)$ nach $(E_1, \mathcal{A}_1)$ sowie κ_2 ein Übergangkern von $(E_0 \times E_1, \mathcal{A}_0 \otimes \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$. Dann ist die Abbildung*

$$\begin{aligned}\kappa_1 \otimes \kappa_2 \colon E_0 \times (\mathcal{A}_1 \otimes \mathcal{A}_2) &\to [0, \infty] \\ (x_0, A) &\mapsto \int_{E_1} \Big(\int_{E_2} \mathbb{1}_A(x_1, x_2) \kappa_2((x_0, x_1), \mathrm{d}x_2) \Big) \kappa_1(x_0, \mathrm{d}x_1)\end{aligned}$$

ein (wohldefinierter) Übergangskern von $(E_0, \mathcal{A}_0)$ nach $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$. Sind κ_1 und κ_2 endlich, so ist $\kappa_1 \otimes \kappa_2$ σ-endlich. Sind κ_1 und κ_2 Markovkerne, so ist auch $\kappa_1 \otimes \kappa_2$ ein Markovkern. $\kappa_1 \otimes \kappa_2$ heißt das Produkt *der Kerne κ_1 und κ_2 oder* Produktkern.

Beweis. Sei $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Die Abbildung

$$\varphi_A \colon E_0 \times E_1 \mapsto [0, \infty], \quad (x_0, x_1) \mapsto \int_{E_2} \mathbb{1}_A(x_1, x_2) \kappa_2((x_0, x_1), \mathrm{d}x_2)$$

ist nach Lemma 5.15 wohldefiniert und $\mathcal{A}_0 \otimes \mathcal{A}_1$-messbar. Daher ist wiederum nach dem selben Lemma die Abbildung

$$E_0 \ni x_0 \mapsto \kappa_1 \otimes \kappa_2(x_0, A) = \int_{E_1} \varphi_A(x_0, x_1) \kappa_1(x_0, \mathrm{d}x_1)$$

wohldefiniert und $\mathcal{A}_0$-messbar. Für jedes feste x_0 ist nach dem Satz über monotone Konvergenz die Abbildung $A \mapsto \kappa_1 \otimes \kappa_2(x_0, A)$ σ-additiv (vgl. Korollar 3.17), also ein Maß.

Seien nun κ_1, κ_2 endliche Kerne. Für $x_0 \in E_0$ und $n \in \mathbb{N}$ sei $A_{x_0,n} := \{x_1 \in E_1 : \kappa_2((x_0, x_1), E_2) < n\}$. Da κ_2 endlich ist, gilt $\bigcup_{n \geqslant 1} A_{x_0,n} = E_1$ und es gilt

$$\kappa_1 \otimes \kappa_2(x_0, A_{x_0,n} \times E_2) \leqslant n \cdot \kappa_1(x_0, A_n) < \infty.$$

Also ist $\kappa_1 \otimes \kappa_2(x_0, \cdot)$ für jedes x_0 σ-endlich. Sind κ_1, κ_2 Markovkerne, gilt für alle $x_0 \in E_0$

$$\kappa_1 \otimes \kappa_2(x_0, E_1 \times E_2) = \int_{E_1} \kappa_2((x_0, x_1), E_2) \kappa_1(x_0, \mathrm{d}x_1) = 1. \qquad \square$$

Bemerkung 5.17. Ist κ_2 ein Kern von $(E_1, \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$, so definieren wir das Produkt $\kappa_1 \otimes \kappa_2$ analog, wobei wir κ_2 als Kern von $(E_0 \times E_1, \mathcal{A}_0 \otimes \mathcal{A}_1)$ nach $(E_2, \mathcal{A}_2)$ auffassen, der nicht von der E_0-Koordinate abhängt. Ist $(E_i, \mathcal{A}_i) = (E_0, \mathcal{A}_0)$ für $i = 1, 2$ und $\kappa_1 = \kappa_2$ schreiben wir kurz $\kappa_1 \otimes \kappa_2 = \kappa_1^{\otimes 2}$.

Als Spezialfall, des vorherigen Satzes können wir die Existenz und Eindeutigkeit von Produktmaßen (Satz 5.6) auf Produkte mit Kernen erweitern.

Korollar 5.18. *Es seien $(E_1, \mathcal{A}_1, \mu)$ ein endlicher Maßraum, $(E_2, \mathcal{A}_2)$ ein messbarer Raum und κ ein endlicher Übergangskern von E_1 nach E_2. Dann existiert ein eindeutig bestimmtes, σ-endliches Maß $\mu \otimes \kappa$ auf $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ mit*

$$\mu \otimes \kappa(A_1 \times A_2) = \int_{A_1} \kappa(\omega_1, A_2) \mu(\mathrm{d}\omega_1) \quad \textit{für alle } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

$\mu \otimes \kappa$ heißt Kopplung *von μ und κ.* Ist *κ ein Markovkern und μ ein Wahrscheinlichkeitsmaß, so ist $\mu \otimes \kappa$ ein Wahrscheinlichkeitsmaß.*

Beweis. Man wende Satz 5.16 auf $\kappa_2 = \kappa$ und $\kappa_1(x_0, \cdot) = \mu$ an. Die Eindeutigkeit folgt wie in Satz 5.6. $\square$

Beispiel 5.19. Wir betrachten folgendes zweistufige Experiment. Zunächst ziehen wir eine gleichverteilte Zufallsvariable $p \sim \mathcal{U}([0, 1])$. Anschließend werfen wir eine Münze bei der mit Wahrscheinlichkeit p Zahl fällt und sonst Kopf. Das entsprechende Modell ist gegeben durch $(E_1, \mathcal{A}_1, \mu) = ([0, 1], \mathscr{B}_{[0,1]}, \lambda)$ sowie $(E_2, \mathcal{A}_2) = (\{0, 1\}, \mathscr{P}(\{0, 1\}))$ mit dem Markovkern $\kappa(q, \cdot) = q\delta_1 + (q - 1)\delta_0, q \in [0, 1]$. Wie hoch ist die Wahrscheinlichkeit, dass am Ende Zahl fällt, also $\mu \otimes \kappa([0, 1] \times \{1\})$?

Satz 5.20 (Fubini für Übergangskerne)**.** *Seien* $(E_1, \mathcal{A}_1, \mu)$ *ein Maßraum mit* $\mu(E_1) < \infty$, $(E_2, \mathcal{A}_2)$ *ein messbarer Raum und* κ *ein endlicher Übergangkern von* E_1 *nach* E_2 *sowie* $f\colon E_1 \times E_2 \to \overline{\mathbb{R}}$ $\mathcal{A}_1 \otimes \mathcal{A}_2$*-messbar. Ist* $f \geqslant 0$ *oder* $f \in \mathcal{L}^1(\mu \otimes \kappa)$*, dann gilt*

$$\int_{E_1 \times E_2} f \mathrm{d}(\mu \otimes \kappa) = \int_{E_1} \Big(\int_{E_2} f(x_1, x_2) \kappa(x_1, \mathrm{d}x_2) \Big) \mu(\mathrm{d}x_1).$$

Beweis. Für $f = \mathbb{1}_A$ mit $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ gilt die Aussage per Definition. Für allgemeine f verwende man maßtheoretische Induktion analog zum Satz von Fubini (Übung ☑). □

Abschließend wollen wir noch die Existenz und Eindeutigkeit von Produkten abzählbar vieler Markovkerne nachweisen. Dies ermöglicht es uns abzählbar viele sukzessive Experimente auf einem Wahrscheinlichkeitsraum realisieren zu können. Seien hierzu $(E_0, \mathcal{A}_0, P_0)$ ein Wahrscheinlichkeitsraum und $(E_i, \mathcal{A}_i), i \in \mathbb{N}$, messbare Räume. Wir setzen

$$E^k := \bigtimes_{i=0}^{k} E_i \quad \text{und} \quad \mathcal{A}^k := \bigotimes_{i=0}^{k} \mathcal{A}_i, \quad k \in \mathbb{N},$$

sowie

$$E := \bigtimes_{i=0}^{\infty} E_i \quad \text{und} \quad \mathcal{A} := \bigotimes_{i=0}^{\infty} \mathcal{A}_i.$$

Es sei κ_k ein Markovkern von $(E^{k-1}, \mathcal{A}^{k-1})$ nach $(E_k, \mathcal{A}_k)$ für jedes $k \in \mathbb{N}$. Dann definieren wir rekursiv

$$P_k := P_0 \otimes \bigotimes_{i=1}^{k} \kappa_i := P_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_k := \big(P_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_{k-1}\big) \otimes \kappa_k$$

für jedes $k \in \mathbb{N}$ ein Wahrscheinlichkeitsmaß auf $(E^k, \mathcal{A}^k)$ (induktive Anwendung von Korollar 5.18). Nach Konstruktion gilt

$$P_k\big(A \times E_{i+1} \times \cdots \times E_k\big) = P_i(A) \quad \text{für } k \geqslant i, A \in \mathcal{A}^i.$$

Satz 5.21 (Ionescu-Tulcea)**.** *Es seien* $(E_0, \mathcal{A}_0, P_0)$ *ein Wahrscheinlichkeitsraum,* $(E_i, \mathcal{A}_i), i \in \mathbb{N}$, *messbare Räume,* $E^k, \mathcal{A}^k, E, \mathcal{A}$ *wie oben definiert,* κ_k *ein Markovkern von* $(E^{k-1}, \mathcal{A}^{k-1})$ *nach* $(E_k, \mathcal{A}_k)$, $k \in \mathbb{N}$, *und* $P_k := P_0 \otimes \bigotimes_{i=1}^{k} \kappa_i$*. Dann gibt es ein eindeutig bestimmtes Wahrscheinlichkeitsmaß* P *auf* $(E, \mathcal{A})$ *mit*

$$P\Big(A \times \bigtimes_{k=i+1}^{\infty} E_k\Big) = P_i(A) \quad \textit{für alle } i \geqslant 0, A \in \mathcal{A}^i. \tag{5.3}$$

Beweis. Wir betrachten die Menge der Zylindermengen

$$\mathcal{Z} := \big\{A \times \bigtimes_{k=i+1}^{\infty} E_k : i \in \mathbb{N}, A \in \mathcal{A}^i\big\},$$

so dass die Mengenfunktion P durch (5.3) auf $\mathcal{Z}$ festgelegt ist. Offenbar ist $\mathcal{Z}$ eine Algebra mit $\sigma(\mathcal{Z}) = \mathcal{A}$. Weiterhin ist P additiv auf $\mathcal{Z}$, also ein Inhalt, mit $P(E) = P_0(E_0) = 1$. Ist nun P stetig in $\emptyset$, so ist P ein σ-Inhalt, also ein Prämaß auf $(E, \mathcal{Z})$ welches nach Caratheodory (Satz 2.13) zu einem Maß auf $\mathcal{A}$ fortgesetzt werden kann. Die Eindeutigkeit des Maßes folgt dann aus Satz 2.23, da die Menge der Rechteckzylinder

$$\big\{ \bigtimes_{k=0}^{i} A_k \times \bigtimes_{k=i+1}^{\infty} E_k : A_k \in \mathcal{A}_k, k = 0, \dots, i, i \in \mathbb{N}\big\} \subseteq \mathcal{Z}$$

ein $\cap$-stabiler Erzeuger von $\mathcal{A}$ ist.

Es bleibt also zu zeigen, dass P stetig in $\emptyset$ ist. Sei also $(A_n)_{n\geqslant 0} \subseteq \mathcal{Z}$ eine fallende Folge mit $\alpha := \inf_{n\geqslant 0} P(A_n) > 0$, so müssen wir zeigen, dass $\bigcap_{n\geqslant 0} A_n \neq \emptyset$. O.B.d.A. können wir annehmen, dass $A_n = A_n^{'} \times \bigtimes_{k=n+1}^{\infty} E_k$ für gewisse $A_n' \in \mathcal{A}^n$. Für $n \geqslant m$ setze

$$\begin{aligned} h_{m,n}(x_0,\dots,x_m) :&= \Big(\delta_{(x_0,\dots,x_m)} \otimes \bigotimes_{k=m+1}^{n} \kappa_k\Big)(A_n') \\ &= \int_{E_{m+1}\times\dots\times E_n} \mathbb{1}_{A_n'}(x_0,\dots,x_n)\Big(\bigotimes_{k=m+1}^{n} \kappa_k\Big)\big((x_0,\dots,x_m),\mathrm{d}(x_{m+1},\dots,x_n)\big) \end{aligned}$$

und $h_m := \inf_{n\geqslant m} h_{n,m}$. Wir zeigen jetzt induktiv, dass es $y_i \in E_i, i \geqslant 0$, gibt mit

$$h_m(y_0,\dots,y_m) \geqslant \alpha \quad \text{für alle } m \geqslant 0. \tag{5.4}$$

Wegen $A_{n+1}' \subseteq A_n' \times E_{n+1}$ gilt

$$\begin{aligned} h_{m,n+1}(x_0,\dots,x_m) &= \Big(\delta_{(x_0,\dots,x_m)} \otimes \bigotimes_{k=m+1}^{n+1} \kappa_k\Big)(A_{n+1}') \\ &\leqslant \Big(\delta_{(x_0,\dots,x_m)} \otimes \bigotimes_{k=m+1}^{n+1} \kappa_k\Big)(A_n' \times E_{n+1}) \\ &= \Big(\delta_{(x_0,\dots,x_m)} \otimes \bigotimes_{k=m+1}^{n} \kappa_k\Big)(A_n') = h_{m,n}(x_0,\dots,x_m). \end{aligned}$$

Also gilt $h_{m,n} \downarrow h_m$ für $n \to \infty$ und nach dem Satz über majorisierte Konvergenz und Fubini für Übergangskerne folgt

$$\int_{E_0} h_0 \mathrm{d}P_0 = \inf_{n\geqslant 0} \underbrace{\int_{E_0} h_{0,n}\,\mathrm{d}P_0}_{=P_n(A_n')} = \alpha.$$

Daher muss ein $y_0 \in E_0$ existieren mit $h_0(y_0) \geqslant \alpha$, also gilt (5.4) für $m = 0$. Gelte nun (5.4) für ein $m \geqslant 0$. Dann ist

$$\begin{aligned} &\int_{E_{m+1}} h_{m+1}(y_0,\dots,y_m,x_{m+1})\kappa_{m+1}((y_0,\dots,y_m),\mathrm{d}x_{m+1}) \\ &= \inf_{n\geqslant m+1} \int_{E_{m+1}} h_{m+1,n}(y_0,\dots,y_m,x_{m+1})\kappa_{m+1}((y_0,\dots,y_m),\mathrm{d}x_{m+1}) \\ &= \inf_{n\geqslant m+1} h_{m,n}(y_0,\dots,y_m) = h_m(y_0,\dots,y_m) \geqslant \alpha. \end{aligned}$$

Es folgt (5.4) für $m+1$.

Setzen wir nun $y = (y_0, y_1, \dots) \in E$, dann gilt nach Konstruktion $\alpha \leqslant h_{m,m}(y_0,\dots,y_m) = \mathbb{1}_{A_m'}(y_0,\dots,y_m)$ also $y \in A_m$ für jedes $m \geqslant 0$ und damit $\bigcap_{n\geqslant 0} A_n \neq \emptyset$. □

Beispiel 5.22 (Markovketten)**.** Wir betrachten den endlichen Ergebnisraum $(\Omega_n, \mathcal{A}_n)$ mit $\Omega_n := \{1,\dots,n\}$ und $\mathcal{A}_n := \mathscr{P}(\Omega_n)$. Auf $(\Omega_n, \mathcal{A}_n)$ ist eine Anfangsverteilung P_0 via $P_0(\{k\}) = p_k, k = 1,\dots,n$ eindeutig durch einen Vektor $p_0 \in [0,1]^n$ mit $\sum_{k=1}^n p_k = 1$ bestimmt. Auf $(\Omega_n, \mathcal{A}_n)$ heißt eine Familie von Zufallsvariablen (ein stochastischer Prozess) $(X_i)_{i\in\mathbb{N}}$ *Markovkette*, falls die Verteilung von X_i nur von X_{i-1} abhängt. Es gibt also einen Markovkern $\kappa\colon \Omega_n \times \mathcal{A}_n \to [0,1]$, eindeutig beschrieben durch eine *Übergangsmatrix*

$$\big(\kappa(k,\{l\})\big)_{k,l=1,\dots,n} = K \in [0,1]^{n\times n} \quad \text{mit} \quad \sum_{l=1}^{n} K_{k,l} = 1 \forall k \in \Omega_n,$$

so dass die Verteilung P_i von X_i gegeben ist durch $P_i = P_{i-1} \otimes \kappa = P_0 \otimes \kappa^{\otimes i}$, vgl. Bemerkung 5.17. Ist $p_i = (P_i(\{1\}), \dots, P_i(\{n\})) \in [0,1]^n$, so gilt

$$p_i = p_0 \cdot K^i = p_0 \cdot \underbrace{K \cdot \dots \cdot K}_{i \text{ mal}} \quad \text{für alle } i \geqslant 0.$$

Korollar 5.23. *Für jedes $n \geqslant 1$ sei $(E_n, \mathcal{A}_n, P_n)$ ein Wahrscheinlichkeitsraum. Dann existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P auf $(E, \mathcal{A})$ mit $E = \bigtimes_{n \geqslant 1} E_n, \mathcal{A} = \bigotimes_{n \geqslant 1} \mathcal{A}_n$ für das gilt*

$$P\Big(A_1 \times \dots \times A_n \times \bigtimes_{k \geqslant n+1} E_k\Big) = \prod_{k=1}^{n} P_k(A_k)$$

für jedes $n \in \mathbb{N}$ und alle $A_k \in \mathcal{A}_k, k = 1, \dots, n$. Wir nennen $\bigotimes_{k \geqslant 1} P_k := P$ das Produktmaß *der Maße $(P_k)_{k \geqslant 1}$.*

Unter dem Produktmaß P sind die Zufallsvariablen $X_i \colon E \to E_i, X_i(x) = \pi_i(x) = x_i$ gegeben durch die Koordinatenprojektionen π_i verteilt gemäß P_i, da

$$P^{X_i}(A) = P(X_i^{-1}(A)) = P\Big(\bigtimes_{k=0}^{i-1} E_k \times A \times \bigtimes_{k \geqslant i+1} E_k\Big) = P_i(A), \quad A \in \mathcal{A}_i.$$

und paarweise unabhängig, denn für $1 \leqslant i < j$ und $A \in \mathcal{A}_i, B \in \mathcal{A}_j$ gilt

$$\begin{aligned} P\big(X_i \in A, X_j \in B\big) &= P\big(X_i^{-1}(A) \cap X_j^{-1}(B)\big) \\ &= P\Big(\bigtimes_{k=1}^{i-1} E_k \times A \times \bigtimes_{k=i+1}^{j-1} E_k \times B \times \bigtimes_{k \geqslant j+1} E_k\Big) = P_i(A) P_j(B). \end{aligned}$$

Analog sind $(X_i)_{i \in J}$ für jede endliche Teilmenge $J \in \mathbb{N}$ unabhängig, also sind $(X_i)_{i \in \mathbb{N}}$ unabhängig. Wir haben also für abzählbar viele gegebene Wahrscheinlichkeiträume $(E_n, \mathcal{A}_n, P_n), n \geqslant 1$, einen Wahrscheinlichkeitsraum konstruiert auf dem unabhängige P_n-verteilte Zufallsvariablen $X_n \colon (E, \mathcal{A}) \to (E_n, \mathcal{A}_n)$ existieren. Dieses Resultat kann mit Hilfe von Kolmogorovs Erweiterungssatz auf überabzählbare Produkträume verallgemeinert werden (siehe Vorlesung "Stochastische Prozesse").

6 Bedingte Erwartung

Wenn über den Ausgang eines Zufallsexperiments eine Teilinformation vorhanden ist, ändern sich die Wahrscheinlichkeiten für die möglichen Ereignisse. Dieser Sachverhalt wird durch bedingte Wahrscheinlichkeiten bzw. Erwartungswerte formalisiert. Im folgenden sei $(\Omega, \mathcal{A}, P)$ stets ein Wahrscheinlichkeitsraum. Für "P-fast überall" schreiben wir auch "P-fast sicher" oder nur "fast sicher" (f.s.).

6.1 Bedingte Wahrscheinlichkeit

Beispiel 6.1. Wir werfen einen fairen Würfel. Der zugehörigen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ ist gegeben durch $\Omega = \{1, 2, \dots, 6\}, \mathcal{A} = \mathscr{P}(\Omega)$ und $P(\{\omega\}) = \frac{1}{6}$ für alle $\omega \in \Omega$. Betrachten wir die Ereignisse

$$\begin{aligned} A &:= \{\text{Augenzahl} \leqslant 3\} = \{1, 2, 3\} \quad \text{und} \\ B &:= \{\text{Augenzahl ungerade}\} = \{1, 3, 5\}. \end{aligned}$$

Offenbar gilt $P(A) = P(B) = \frac{1}{2}$. Wir groß ist die Wahrscheinlichkeit von B, wenn wir schon wissen, dass A eintritt?

Wenn wir wissen, dass A gilt, liegt es nahe auf A eine Gleichverteilung zu vermuten. Anderseits sollten Ergebnisse aus $\Omega \setminus A$ die Wahrscheinlichkeit 0 haben. Wir erhalten das Wahrscheinlichkeitsmaß P_A auf $(\Omega, \mathcal{A})$ mit

$$P_A(C) = \frac{\#(C \cap A)}{\# A} = \frac{P(C \cap A)}{P(A)} \quad \text{für alle } C \subseteq \Omega.$$

Damit gilt $P_A(B) = \frac{2}{3}$.

Dieses Beispiel führt uns auf folgende Definition:

Definition 6.2. Sei $A \in \mathcal{A}$. Dann ist die *bedingte Wahrscheinlichkeit gegeben A* definiert durch

$$P(B|A) := \begin{cases} \frac{P(A \cap B)}{P(A)}, & \text{falls } P(A) > 0, \\ 0, & \text{sonst.} \end{cases}$$

Offensichtlich ist $P(\cdot|A)$ tatsächlich ein Wahrscheinlichkeitsmaß, falls $P(A) > 0$. Zudem hat die bedingte Wahrscheinlichkeit folgende Eigenschaften:

Satz 6.3. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.*

(i) Für $A, B \in \mathcal{A}$ mit $P(A) > 0$ gilt

$$A, B \textit{ sind unabhängig} \quad \Longleftrightarrow \quad P(B|A) = P(B).$$

(ii) Sei I eine höchstens abzählbare Indexmenge und $(B_i)_{i \in I}$ paarweise disjunkte Mengen mit $P(\bigcup_{i \in I} B_i) = 1$. Dann gilt für jedes $A \in \mathcal{A}$:

$$P(A) = \sum_{i \in I} P(A|B_i)P(B_i) \quad \textit{(Formel von der totalen Wahrscheinlichkeit).}$$

(iii) Sei I eine höchstens abzählbare Indexmenge und $(B_i)_{i \in I}$ paarweise disjunkte Mengen mit $P(\bigcup_{i \in I} B_i) = 1$. Dann gilt für jedes $A \in \mathcal{A}$ mit $P(A) > 0$ und für jedes $k \in I$

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i \in I} P(A|B_i)P(B_i)} \quad \textit{(Bayes-Formel).}$$

Beweis. Übung ☑. □

Ist nun $X \colon \Omega \to \overline{\mathbb{R}}$ eine Zufallsvariable (also messbar) und $A \in \mathcal{A}$ mit $P(A) > 0$, so ist auch $\mathbb{1}_A \cdot X$ messbar. Ist X eine einfache Funktion, so folgt direkt für den Erwartungswert bzgl. $P(\cdot|A)$

$$\mathbb{E}[X|A] := \int_\Omega X(\omega) P(\mathrm{d}\omega|A) = \frac{\mathbb{E}[\mathbb{1}_A \cdot X]}{P(A)}.$$

Mit dem Satz über monotone Konvergenz setzt sich diese Gleichung auf $X \in \mathcal{M}_+$ und weiter auf $X \in \mathcal{L}^1(P)$ fort.

Definition 6.4. Sei $X \in \mathcal{M}_+$ oder $X \in \mathcal{L}^1(P)$ und $A \in \mathcal{A}$. Dann setzen wir

$$\mathbb{E}[X|A] := \int_\Omega X \mathrm{d}P_A = \begin{cases} \frac{\mathbb{E}[\mathbb{1}_A \cdot X]}{P(A)}, & P(A) > 0, \\ 0, & \text{sonst.} \end{cases}$$

Betrachten wir nun erneut den Fall, dass wir Ω in eine höchstens abzählbare Zerlegung von disjunkten Ereignissen $(B_i)_{i \in I}$ zerlegen können ($\bigcup_{i \in I} B_i = \Omega$). Wir setzen $\mathcal{F} = \sigma(B_i, i \in I)$ und für $X \in \mathcal{L}^1(P)$ definieren wir die Abbildung

$$\mathbb{E}[X|\mathcal{F}] \colon \Omega \to \mathbb{R}, \quad \omega \mapsto \sum_{i \in I} \mathbb{E}[X|B_i] \cdot \mathbb{1}_{B_i}(\omega).$$

Lemma 6.5. *Sei $X \in \mathcal{L}^1(P)$, I höchstens abzählbar und $(B_i)_{i\in I}$ paarweise disjunkte Mengen mit $\bigcup_{i\in I} B_i = \Omega$. Die Abbildung $\mathbb{E}[X|\mathcal{F}]$ hat folgende Eigenschaften:*

(i) $\mathbb{E}[X|\mathcal{F}]$ *ist $\mathcal{F}$-messbar,*

(ii) $\mathbb{E}[X|\mathcal{F}] \in \mathcal{L}^1(P)$ *und für jedes $A \in \mathcal{F}$ gilt* $\int_A \mathbb{E}[X|\mathcal{F}]\mathrm{d}P = \int_A X\mathrm{d}P$.

Beweis. (i) Die Messbarkeit folgt, da $\mathbb{E}[X|\mathcal{F}]$ eine (abzählbare) Linearkombination von Indikatorfunktionen über Mengen aus $\mathcal{F}$ ist

(ii) Da I höchstens abzählbar, existiert für jedes $A \in \mathcal{F}$ ein $J \subseteq I$ mit $A = \bigcup_{i\in J} B_i$. Für $J' := \{i \in J : P(B_i) > 0\}$ gilt dann

$$\int_A \mathbb{E}[X|\mathcal{F}]\mathrm{d}P = \sum_{i\in J'} \mathbb{E}[X|B_i]P(B_i) = \sum_{i\in J'} \mathbb{E}[X\mathbb{1}_{B_i}] = \mathbb{E}[X\mathbb{1}_A]. \qquad \square$$

6.2 Definition und Eigenschaften der bedingten Erwartung

Die obige Konstruktion erlaubt uns auf Ereignisse $B \in \mathcal{A}$ mit $P(B) > 0$ zu bedingen. In Anlehnung an Lemma 6.5, führen wir nun folgende wesentlich allgemeinere Definition ein:

Definition 6.6. Sei $\mathcal{F} \subseteq \mathcal{A}$ eine (Unter-)σ-Algebra und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Eine Zufallsvariable Y heißt *bedingte Erwartung* von X gegeben $\mathcal{F}$, falls gilt:

(i) Y ist $\mathcal{F}$-messbar und

(ii) für jedes $A \in \mathcal{F}$ gilt $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$.

Wir schreiben $\mathbb{E}[X|\mathcal{F}] := Y$. Für $B \in \mathcal{A}$ heißt $P(B|\mathcal{F}) := \mathbb{E}[\mathbb{1}_A|\mathcal{F}]$ die *bedingte Wahrscheinlichkeit* von B gegeben $\mathcal{F}$. Ist Z eine Zufallsvariable, schreiben wir kurz $\mathbb{E}[X|Z] := \mathbb{E}[X|\sigma(Z)]$.

Satz 6.7. *Für jede σ-Algebra $\mathcal{F} \subseteq \mathcal{A}$ und jedes $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ existiert die bedingte Erwartung $\mathbb{E}[X|\mathcal{F}]$ und ist eindeutig (bis auf Gleichheit P-f.s.).*

Beweis. Eindeutigkeit: Seien Y und Y' Zufallsvariablen, die (i) und (ii) aus Definition 6.6 erfüllen. Setze $A := \{Y > Y'\} \in \mathcal{F}$. Aus (ii) folgt

$$0 = \mathbb{E}[Y\mathbb{1}_A] - \mathbb{E}[Y'\mathbb{1}_A] = \mathbb{E}[\underbrace{(Y - Y')\mathbb{1}_A}_{\geqslant 0}].$$

Damit muss $P(A) = 0$ gelten, d.h. $Y \leqslant Y'$ P-f.s. Analog folgt $Y \geqslant Y'$ f.s.

Existenz: Mit der Zerlegung $X = X_+ - X_-$, $X_+ = \max\{X, 0\}, X_- = \max\{-X, 0\}$, definieren wir zwei endliche Maße Q_+ und Q_- auf $\mathcal{F}$ via

$$Q_\pm(A) := \mathbb{E}[X_\pm\mathbb{1}_A] \quad \text{für alle } A \in \mathcal{F}.$$

Offenbar ist $Q_\pm \ll P|_\mathcal{F}$, also existieren nach dem Satz von Radon-Nikodym $P|_\mathcal{F}$-Dichten $Y_\pm$, so dass

$$Q_\pm(A) = \int_\Omega \underbrace{Y_\pm\mathbb{1}_A}_{\mathcal{F}\text{-mb.}} \mathrm{d}P|_\mathcal{F} = \int_\Omega Y_\pm\mathbb{1}_A\mathrm{d}P = \mathbb{E}[Y_\pm\mathbb{1}_A]$$

(die zweite Gleichheit folgt leicht mit maßtheoretischer Induktion). Die Funktion $Y := Y_+ - Y_-$ erfüllt somit (i) und (ii). $\square$

Wir zeigen nun wichtige Eigenschaften, der bedingten Erwartung.

Satz 6.8. *Es seien $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ und $\mathcal{F} \subseteq \mathcal{A}$ eine Unter-σ-Algebra. Ferner sei $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Dann gilt:*

(i) $\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\big] = \mathbb{E}[X]$.

(ii) Ist X $\mathcal{F}$-messbar, dann ist $\mathbb{E}[X|\mathcal{F}] = X$ P-f.s.

(iii) $\mathbb{E}[\alpha X + \beta Y|\mathcal{F}] = \alpha\mathbb{E}[X|\mathcal{F}] + \beta\mathbb{E}[Y|\mathcal{F}]$ P-f.s. für alle $\alpha, \beta \in \mathbb{R}$ (Linearität).

(iv) Ist $X \geqslant Y$ P-f.s., so ist $\mathbb{E}[X|\mathcal{F}] \geqslant \mathbb{E}[Y|\mathcal{F}]$ P-f.s. (Monotonie).

(v) Für σ-Algebren $\mathcal{G} \subseteq \mathcal{F} \subseteq \mathcal{A}$ ist $\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\big|\mathcal{G}\big] = \mathbb{E}[X|\mathcal{G}]$ P-f.s. (Turmeigenschaft).

(vi) Sind $\sigma(X)$ und $\mathcal{F}$ unabhängig, so ist $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$ P-f.s.

(vii) Sind $(X_n)_{n\geqslant 1} \in \mathcal{M}_+$ mit $X_n \uparrow X$ P-f.s., dann gilt $\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}]$ P-f.s. (Monotone Konvergenz).

(viii) Ist $Y \in \mathcal{L}^1(P)$ und sind $(X_n)_{n\geqslant 1}$ Zufallsvariablen auf $(\Omega, \mathcal{A})$ mit $|X_n| \leqslant Y, n \geqslant 1$, und $\lim_{n\to\infty} X_n = X$ P-f.s., so gilt $\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}]$ P-f.s. (majorisierte Konvergenz).

(ix) Ist $\mathbb{E}[|XY|] < \infty$ und ist Y $\mathcal{F}$-messbar, dann ist $\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}]$ P-f.s.

Beweis. (i) Folgt aus der Definition: $\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\big] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\mathbb{1}_\Omega\big] = \mathbb{E}[X\mathbb{1}_\Omega] = \mathbb{E}[X]$.
(ii) Y selbst erfüllt die Bedingungen aus der Definition.
(iii) Übung ☑.
(iv) Übung ☑.
(v) Ist $A \in \mathcal{G} \subseteq \mathcal{F}$, so folgt

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_A\big] = \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\mathbb{1}_A\big] = \mathbb{E}\big[\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\big|\mathcal{G}\big]\mathbb{1}_A\big].$$

(vi) Da $\mathbb{E}[X]$ konstant ist, ist es auch $\mathcal{F}$-messbar. Für jedes $A \in \mathcal{F}$ sind nun $\mathbb{1}_A$ und X unabhängig, also ist $\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\mathbb{1}_A\big] = \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbb{1}_A]$.

(vii) Aus (iv) folgt $\mathbb{E}[X_n|\mathcal{F}] \leqslant \mathbb{E}[X_{n+1}|\mathcal{F}]$ f.s. Also ist $Y := \lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{F}]$ f.s. nicht-negativ und $\mathcal{F}$-messbar. Zweifache Anwendung des Satzes über monotone Konvergenz liefert für alle $A \in \mathcal{F}$

$$\mathbb{E}[Y\mathbb{1}_A] = \lim_{n\to\infty} \mathbb{E}\big[\mathbb{E}[X_n|\mathcal{F}]\mathbb{1}_A\big] = \lim_{n\to\infty} \mathbb{E}[X_n\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A].$$

(viii) Folgt wie die mayorisierte Konvergenz des Lebesgueintegrals (Satz 3.24). Übung ☑.
(ix) (maßtheoretische Induktion)
1. Für $Y = \mathbb{1}_A, A \in \mathcal{F}$, folgt für jedes $B \in \mathcal{F}$

$$\mathbb{E}\big[\mathbb{E}[\mathbb{1}_A X|\mathcal{F}]\mathbb{1}_B\big] = \mathbb{E}[X\underbrace{\mathbb{1}_A\mathbb{1}_B}_{\mathbb{1}_{A\cap B}}] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]\mathbb{1}_A\mathbb{1}_B\big].$$

2. Ist Y eine einfache $\mathcal{F}$-messbare Funktion verwenden wir (iii) und 1.

3. Für $Y \geqslant 0$ zerlegen wir $X = X_+ - X_-$. Sind nun Y_n einfache $\mathcal{F}$-messbare, nichtnegative Funktionen mit $Y_n \uparrow Y$, dann folgt aus (vii), dass

$$\begin{aligned}\mathbb{E}[X_\pm Y|\mathcal{F}] &= \lim_{n\to\infty} \mathbb{E}[X_\pm Y_n|\mathcal{F}] \\ &= \lim_{n\to\infty} Y_n\mathbb{E}[X_\pm|\mathcal{F}] = Y\mathbb{E}[X_\pm|\mathcal{F}].\end{aligned}$$

Mittels (iii) folgt $\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}]$.

4. Für $X \cdot Y \in \mathcal{L}^1(P)$ schreiben wir $X = X_+ - X_-$, $Y = Y_+ - Y_-$ und betrachten die nichtnegativen Funktionen X_+Y_+, X_+Y_-, X_-Y_+ sowie X_-Y_- separat. Die Behauptung folgt, dann wieder aus der Linearität. □

Satz 6.9 (Jensens Ungleichung)**.** *Sei $\varphi\colon \mathbb{R} \to \mathbb{R}$ konvex und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ mit $\varphi(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Für jede σ-Algebra $\mathcal{F} \subseteq \mathcal{A}$ gilt*

$$\varphi(\mathbb{E}[X|\mathcal{F}]) \leqslant \mathbb{E}[\varphi(X)|\mathcal{F}] \quad P\text{-f.s.}$$

Beweis. Wir verwenden folgende Eigenschaften konvexer Funktionen, vgl. Klenke (2008, Satz 7.7):

(i) φ ist stetig auf $\mathbb{R}$ und insbesondere Borel-messbar.

(ii) Für jedes $x \in \mathbb{R}$ ist der Differenzenquotient $\mathbb{R} \setminus \{x\} \ni y \mapsto \frac{\varphi(y)-\varphi(x)}{y-x}$ monoton wachsend.

(iii) Die Funktion
$$D^+\varphi(x) = \inf\left\{\frac{\varphi(y)-\varphi(x)}{y-x} : y > x\right\}$$
existiert, ist endlich und es gilt $\varphi(y) \geqslant (y-x)D^+\varphi(x) + \varphi(x)$ ($D^+\varphi(x)$ ist die maximale Tangentensteigung von φ an der Stelle x).

Angewendet an den Stellen $x = \mathbb{E}[X|\mathcal{F}]$ und $y = X$ erhalten wir
$$\varphi(X) \geqslant \big(X - \mathbb{E}[X|\mathcal{F}]\big)D^+\varphi(\mathbb{E}[X|\mathcal{F}]) + \varphi(\mathbb{E}[X|\mathcal{F}]).$$

Aufgrund der Stetigkeit von φ, gilt $D^+\varphi(x) = \inf\{\frac{\varphi(y)-\varphi(x)}{y-x} : y \in \mathbb{Q}, y > x\}$, so dass insbesondere $D^+\varphi(\mathbb{E}[X|\mathcal{F}])$ $\mathcal{F}$-messbar ist. Bedingen wir also obige Ungleichung auf $\mathcal{F}$ erhalten wir dank Satz 6.8 P-f.s.
$$\begin{aligned}\mathbb{E}[\varphi(X)|\mathcal{F}] \geqslant& \mathbb{E}\big[(X - \mathbb{E}[X|\mathcal{F}])D^+\varphi(\mathbb{E}[X|\mathcal{F}]) + \varphi(\mathbb{E}[X|\mathcal{F}])|\mathcal{F}\big]\\ =&\mathbb{E}\big[(X - \mathbb{E}[X|\mathcal{F}])|\mathcal{F}\big]D^+\varphi(\mathbb{E}[X|\mathcal{F}]) + \varphi(\mathbb{E}[X|\mathcal{F}])\\ =&\varphi(\mathbb{E}[X|\mathcal{F}]).\end{aligned}$$
□

Intuitiv ist $\mathbb{E}[X|\mathcal{F}]$ die beste Vorhersage, die wir für X machen können, wenn uns die Information aus $\mathcal{F}$ zur Verfügung steht. Ist $\sigma(X) \subseteq \mathcal{F}$, kennen wir X schon und $\mathbb{E}[X|\mathcal{F}] = X$ wie oben gezeigt. Im anderen Extrem sind X und $\mathcal{F}$ unabhängig. In diesem Fall wissen wir nichts zusätzlich über X und $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$. Formal können wir die bedingte Erwartung als Orthogonalprojektion von $X \in L^2(\Omega, \mathcal{A}, P)$ auf den Unterraum $L^2(\Omega, \mathcal{F}, P)$ auffassen:

Satz 6.10. *Sei $\mathcal{F} \subseteq \mathcal{A}$ eine σ-Algebra und $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Dann gilt für jedes $\mathcal{F}$-messbare Y mit $\mathbb{E}[Y^2] < \infty$:*
$$\mathbb{E}\big[(X-Y)^2\big] \geqslant \mathbb{E}\big[(X - \mathbb{E}[X|\mathcal{F}])^2\big]$$
mit Gleichheit genau dann, wenn $Y = \mathbb{E}[X|\mathcal{F}]$.

Beweis. Aus Jensens Ungleichung folgt $\mathbb{E}[\mathbb{E}[X|\mathcal{F}]^2] \leqslant \mathbb{E}[X^2] < \infty$, also $\mathbb{E}[X|\mathcal{F}] \in \mathcal{L}^2(P)$. Sei nun $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Die Ungleichung von Cauchy-Schwarz liefert $\mathbb{E}[|XY|] < \infty$, also $\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]Y]$. Entsprechend gilt
$$\mathbb{E}\big[X\mathbb{E}[X|\mathcal{F}]\big] = \mathbb{E}\Big[\mathbb{E}\big[X\mathbb{E}[X|\mathcal{F}]\big|\mathcal{F}\big]\Big] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{F}]^2\big].$$

Somit folgt
$$\begin{aligned}\mathbb{E}\big[(X-Y)^2\big] - \mathbb{E}\big[(X - \mathbb{E}[X|\mathcal{F}])^2\big] &= \mathbb{E}\big[Y^2 - 2XY + 2X\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[X|\mathcal{F}]^2\big]\\ &= \mathbb{E}\big[Y^2 - 2Y\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[X|\mathcal{F}]^2\big]\\ &= \mathbb{E}\big[(Y - \mathbb{E}[X|\mathcal{F}])^2\big] \geqslant 0.\end{aligned}$$
□

Als Charakterisierung von Messbarkeit ist folgendes Resultat hilfreich:

Lemma 6.11 (Faktorisierungslemma)**.** *Sei $(S, \mathcal{S})$ ein messbarer Raum, X eine S-wertige Zufallsvariable und Y eine reellwertige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Y ist genau dann $(\sigma(X), \mathscr{B}_{\mathbb{R}})$-messbar, wenn es eine $(\mathcal{S}, \mathscr{B}_{\mathbb{R}})$-messbare Abbildung $\varphi\colon S \to \mathbb{R}$ gibt, so dass $Y = \varphi \circ X$.*

Beweis. "$\Leftarrow$" $Y = \varphi(X)$ ist $\sigma(X)$-messbar, da X $\sigma(X)$-messbar ist und die Verknüpfung zweier messbarer Abbildungen wieder messbar ist.

"$\Rightarrow$" (Maßtheoretische Induktion)

Schritt 1: Sei $Y = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$ eine einfache Funktion (in Normaldarstellung) mit paarweise verschiedenen $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ und paarweise disjunkten $A_1, \dots, A_n \in \mathcal{A}$. Da Y $\sigma(X)$-messbar ist, gilt $A_k = Y^{-1}(\{\alpha_k\}) \in \sigma(X)$. Es existieren also $S_1, \dots, S_n \in \mathcal{S}$, so dass $A_k = X^{-1}(S_k)$ für $1 \leqslant k \leqslant n$. Dann gilt

$$Y = \varphi \circ X \quad \text{für} \quad \varphi = \sum_{k=1}^n \alpha_k \mathbb{1}_{S_k}.$$

Nach Konstruktion ist φ $\mathcal{S}$-messbar.

Schritt 2: Sei nun $Y \in \mathcal{M}_+$ $\sigma(X)$-messbar. Dann existiert eine Folge $(Y_n)_{n\geqslant 1} \subseteq \mathcal{M}_+$ einfacher, $\sigma(X)$-messbarer Funktionen mit $Y_n \uparrow Y$ für $n \to \infty$. Nach Schritt 1 gibt es messbare φ_n, so dass $Y_n = \varphi_n(X)$. Aus $Y_n \leqslant Y_{n+1}$ folgt $\varphi_n \leqslant \varphi_{n+1}$. Wir setzen $\varphi = \lim_{n\to\infty} \varphi_n \in [0, \infty]$. Wegen

$$Y(\omega) = \lim_{n\to\infty} Y_n(\omega) = \lim_{n\to\infty} \varphi_n(X(\omega)) = \varphi(X(\omega)) \quad \text{für alle } \omega \in \Omega,$$

gilt $\varphi(x) < \infty$ für alle x im Bild von X. Da φ $(\mathcal{S}, \mathscr{B}_{[0,\infty]})$-messbar ist, können wir φ auf $\varphi^{-1}(\{\infty\})$ zu 0 in messbarer Weise abändern.

Schritt 3: Für beliebige $\sigma(X)$-messbare Y zerlegen wir $Y = Y_+ - Y_-$ und erhalten nach Schritt 2 messbare Funktionen $\varphi_\pm$ mit $Y_\pm = \varphi_\pm \circ X$. Damit ist $\varphi := \varphi_+ - \varphi_-$ $\mathcal{S}$-messbar und erfüllt $Y = \varphi \circ X$. □

Sei Y eine $(S, \mathcal{S})$-wertige Zufallsvariable. Wenden wir das Faktorisierungslemma auf die bedingte Erwartung gegeben Y an, so existiert eine Funktion φ mit $\varphi(Y(\omega)) = \mathbb{E}[X|\sigma(Y)](\omega) = \mathbb{E}[X|Y](\omega)$. In diesem Fall wird $\mathbb{E}[X|Y = \cdot] := \varphi$ *bedingter Erwartungswert* von X gegeben Y genannt. Für diesen gilt also für alle $A \in \mathcal{S}$

$$\int_{\{Y\in A\}} X \mathrm{d}P = \int X \mathbb{1}_{Y^{-1}(A)} \mathrm{d}P = \int \varphi(Y) \mathbb{1}_A(Y) \mathrm{d}P = \int_A \varphi \mathrm{d}P^Y.$$

$\mathbb{E}[X|Y = \cdot]$ ist wegen $\mathbb{E}[X|Y] = \mathbb{E}[X|Y = \cdot] \circ Y$ nur P^Y-f.s. eindeutig bestimmt.

Es ist naheliegend mit Hilfe des bedingten Erwartungswertes auch die bedingte Verteilung $P(A|Y = y) := \mathbb{E}[\mathbb{1}_A|Y = y]$ für $y \in \mathcal{S}$ und $A \in \mathcal{A}$ zu definieren. Handelt es sich dabei aber wirklich um ein Wahrscheinlichkeitsmaß? Die Antwort ist nicht offensichtlich, da $P(A|Y = y)$ nur für P^Y-f.a. y eindeutig ist, wobei die Ausnahmemenge wieder von A abhängt.

Definition 6.12. Es seien $(S, \mathcal{S})$ ein messbarer Raum, X eine S-wertige Zufallsvariable und $\mathcal{F} \subseteq \mathcal{A}$ eine σ-Algebra. Ein Markovkern $\kappa_{X,\mathcal{F}}$ von $(\Omega, \mathcal{F})$ nach $(S, \mathcal{S})$ heißt *reguläre Version der bedingten Verteilung* von X gegeben $\mathcal{F}$, falls

$$\kappa_{X,\mathcal{F}}(\omega, A) = \mathbb{E}\big[\mathbb{1}_{X^{-1}(A)}|\mathcal{F}\big](\omega) \quad \text{für } P\text{-f.a. } \omega \in \Omega \text{ und jedes } A \in \mathcal{S}.$$

Ist speziell $\mathcal{F} = \sigma(Y)$ für eine Zufallsvariable Y (in einen beliebigen messbaren Raum), dann heißt ein Markovkern $\kappa_{X,Y}$ von $(\Omega, \mathcal{F})$ nach $(S, \mathcal{S})$ mit

$$\kappa_{X,Y}(y, A) = \mathbb{E}\big[\mathbb{1}_{X^{-1}(A)}|Y = y\big]$$

($= \varphi(y)$ für die Funktion aus dem Faktorierisierungslemma) *reguläre Version der bedingten Verteilung* von X gegeben Y, geschrieben $P^{X|Y=y}(A) := \kappa_{X,Y}(y, A)$.

Bemerkung 6.13. Man kann zeigen, dass für $\mathbb{R}^n$-wertige Zufallsvariablen X (allgemeiner mit Werten in polnischen Räumen) eine reguläre Version der bedingten Verteilung existiert.

Lemma 6.14. *Es seien $(S_X, \mathcal{S}_X)$ und $(S_Y, \mathcal{S}_Y)$ messbare Räume sowie X bzw. Y S_X- bzw. S_Y-wertige Zufallsvariablen. Ist $P^{X|Y}$ eine reguläre Version der bedingten Verteilung von X gegeben Y, dann ist $P^{(Y,X)}$ gerade die Kopplung von P^Y und $P^{X|Y}$, d.h. $P^{(Y,X)} = P^Y \otimes P^{X|Y}$.*

Beweis. Es genügt die Gleichheit auf Rechteckmengen $B \times A$ für $A \in \mathcal{S}_X$ und $B \in \mathcal{S}_Y$ nachzuweisen. Aufgrund der Eigenschaften der bedingten Erwartung und der Definition der Kopplung

$$\begin{aligned} P^{(Y,X)}(B \times A) &= \mathbb{E}\big[\mathbb{1}_A(X)\mathbb{1}_B(Y)\big] \\ &= \mathbb{E}\big[\mathbb{E}[\mathbb{1}_A(X)|Y] \cdot \mathbb{1}_B(Y)\big] \\ &= \mathbb{E}\big[P^{X|Y}(A)\mathbb{1}_B(Y)\big] \\ &= \int_B P^{X|Y=y}(A)P^Y(\mathrm{d}y) = P^Y \otimes P^{X|Y}(B \times A). \end{aligned}$$ $\square$

Als eine typische Anwendung, erlauben uns bedingte Erwartungen, die Berechnung von Erwartungswerten der Form $\mathbb{E}[g(X,Y)]$ für Zufallsvariablen X, Y und $g \in \mathcal{L}^1(P^{(X,Y)})$. Der Satz von Fubini für Übergangskerne liefert nämlich

$$\begin{aligned} \mathbb{E}[g(X,Y)] &= \mathbb{E}\big[\mathbb{E}[g(X,Y)|Y]\big] \\ &= \int_{S_Y} \mathbb{E}[g(X,y)|Y=y]P^Y(\mathrm{d}y) \\ &= \int_{S_Y}\int_{S_X} g(x,y)P^{X|Y=y}(\mathrm{d}x)P^Y(\mathrm{d}y). \end{aligned}$$

7 Martingale

Der wichtige Begriff des Martingals dient der Formalisierung von fairen Spielen. Er ist insbesondere in der Finanzmathematik zentral um den Handel mit Aktien an (vollständigen) Märkten zu modellieren.

7.1 Definition und Doob-Zerlegung

Definition 7.1. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $I \subseteq \mathbb{R}$. Eine Familie $(\mathcal{F}_t)_{t\in I}$ von σ-Algebren mit $\mathcal{F}_t \subseteq \mathcal{F}$ für jedes $t \in I$ heißt *Filtration*, falls $\mathcal{F}_s \subseteq \mathcal{F}_t$ für alle $s, t \in I$ mit $s \leqslant t$. Eine Familie von reellwertigen Zufallsvariablen $X = (X_t)_{t\in I}$ (auf $(\Omega, \mathcal{F}, P)$) heißt *stochastischer Prozess*. X heißt *adaptiert* an die Filtration $(\mathcal{F}_t)_{t\in I}$, falls X_t $\mathcal{F}_t$-messbar ist für jedes $t \in I$.

Definition 7.2. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $I \subseteq \mathbb{R}$ und $(\mathcal{F}_t)_{t\in I}$ eine Filtration. Ein stochastischer Prozess $X = (X_t)_{t\in I}$ heißt *Martingal* (bzw. *Submartingal* oder *Supermartingal*) bzgl. $(\mathcal{F}_t)_{t\in I}$, falls

(i) $X_t \in L^1(P), t \in I$,

(ii) X adaptiert an $(\mathcal{F}_t)_{t\in I}$ ist und

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

Bemerkung 7.3. (Übung ☑)

(i) Jeder stochastische Prozess ist adaptiert an seine *natürliche Filtration* $\mathcal{F}_t = \sigma(X_s, s \leqslant t)$ für jedes $t \in I$.

(ii) Ist $I = \mathbb{N} \cup \{0\}$, so folgt die Martingaleigenschaft $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ für alle $s, t \in I$ mit $s \leqslant t$ bereits aus der Bedingung $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ (analog für Sub- und Supermartingale).

(iii) Sind $(X_t)_{t\in I}$ und $(Y_t)_{t\in I}$ zwei Martingale (oder Supermartingale) bzgl. $(\mathcal{F}_t)_{t\in I}$, so gilt

  (a) $(-X_t)_{t\in I}$ ist ein Martingal (bzw. ein Submartingal) bzgl. $(\mathcal{F}_t)_{t\in I}$.

  (b) Für alle $\alpha \in \mathbb{R}$ ist $(\alpha X_t + Y_t)_{t\in I}$ ein Martingal (bzw. ein Supermartingal, falls $\alpha \geqslant 0$).

Beispiel 7.4.

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

(ii) Faires Spiel: Sei X_0 das Startkapital eines Spielers und X_n modelliere das Kapital nach n Runden. Die Filtration $\mathcal{F}_n$ interpretieren wir als die Information der Ausgänge der ersten $n \in \mathbb{N}$ Runden. Dann ist $\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n]$ die Vorhersage des Gewinns (bzw. Verlustes) in Runde $n+1$ gegeben aller Informationen bis zur "Zeit" n. Ist $(X_n)_{n\geqslant 0}$ ein Martingal, handelt es sich also um ein faires Spiel (das Kapital bleibt im Durchschnitt konstant). Ein Submartingal ist vorteilhaft für den Spieler, wohingegen ein Supermartingal im Mittel zu Verlusten führt.

(iii) Ist $Z \in L^1(P)$ und $(\mathcal{F}_t)_{t\in I}$, so definiert $X_t = \mathbb{E}[Z|\mathcal{F}_t], t \in I$, ein Martingal bzgl. $(\mathcal{F}_t)_{t\in I}$.

(iv) Ist $\varphi\colon \mathbb{R} \to \mathbb{R}$ konvex und $(X_t)_{t\in I}$ ein Martingal mit $\varphi(X_t) \in L^1(P)$, dann ist $(\varphi(X_t))_{t\in I}$ ein Submartingal (Jensens Ungleichung). Insbesondere ist $(X_t^2)_{t\in I}$ ein Submartingal für jedes Martingal aus $L^2(P)$.

Wir konzentrieren uns im folgenden auf (diskrete) Martingale indiziert durch $\mathbb{N} \cup \{0\}$ und betrachteten stets einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit einer Filtration $(\mathcal{F}_n)_{n\geqslant 0}$.

Definition 7.5. Ein stochastischer Prozess $(Y_n)_{n\geqslant 0}$ heißt *vorhersagbar* (oder *previsibel*) bzgl. der Filtration $(\mathcal{F}_n)_{n\geqslant 0}$, falls Y_0 konstant ist und Y_n $\mathcal{F}_{n-1}$-messbar ist für jedes $n \in \mathbb{N}$. Für einen vorhersagbaren Prozess $(Y_n)_{n\geqslant 0}$ und ein Martingal $(X_n)_{n\geqslant 0}$ (bzgl. der selben Filtration) ist die *Martingaltransformation* oder das *diskrete stochastische Integral* $(Y \cdot X)_{n\geqslant 0}$ definiert durch

$$(Y \cdot X)_0 = 0 \quad \text{und} \quad (Y \cdot X)_n = \sum_{k=1}^n Y_k(X_k - X_{k-1}), \quad n \geqslant 1.$$

Lemma 7.6. *Ist $(Y_n)_{n\geqslant 0}$ ein beschränkter, vorhersagbarer Prozess und $(X_n)_{n\geqslant 0}$ ein Martingal bzgl. einer Filtration $(\mathcal{F}_n)_{n\geqslant 0}$, so ist $(Y \cdot X)_{n\geqslant 0}$ ebenfalls ein Martingal.*

Beweis. Es gilt $Y_n \leqslant C$ für eine Konstante $C > 0$ und alle $n \geqslant 0$. Wir prüfen die Martingalbedingungen: Für $n \geqslant 1$ gilt

$$\mathbb{E}[|(Y \cdot X)_n|] \leqslant \sum_{k=1}^n \mathbb{E}[|Y_k| \cdot |X_k - X_{k-1}|] \leqslant \sum_{k=1}^n C\big(\mathbb{E}[|X_k|] + \mathbb{E}[|X_{k-1}|]\big) < \infty.$$

Weiterhin ist $(Y \cdot X)_n$ $\mathcal{F}_n$-messbar, da Y_k und X_k $\mathcal{F}_n$-messbar sind für alle $k \leqslant n$. Schließlich gilt

$$\begin{aligned}\mathbb{E}\big[(Y \cdot X)_{n+1} - (Y \cdot X)_n\big|\mathcal{F}_n\big] &= \mathbb{E}\big[Y_{n+1}(X_{n+1} - X_n)\big|\mathcal{F}_n\big] \\ &= Y_{n+1}\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] = 0.\end{aligned}$$

□

Die Martingaltransformation hat folgende Interpretation: X_n sei der Gewinn eines fairen Spiels bei einem Einsatz von 1€. Ein Spieler wählt nun vor dem $(n+1)$. Spiel seinen Einsatz Y_{n+1} für die nächste Runde. Dann ist $Y_{n+1}(X_{n+1} - X_n)$ der Gewinn im $(n+1)$. Spiel und folglich $(Y \cdot X)_n$ das Kapital nach n Runden. Für jede Strategie erhält sich also die "Fairness"-Eigenschaft.

Satz 7.7 (Doob-Zerlegung)**.** *Sei $X = (X_n)_{n\geqslant 0}$ ein bzgl. der Filtration $(\mathcal{F}_n)_{n\geqslant 0}$ adaptierter Prozess mit $X_n \in L^1(P), n \geqslant 0$. Dann existiert eine eindeutige Zerlegung*

$$X = M + A,$$

wobei $A = (A_n)_{n\geqslant 0}$ vorhersagbar ist mit $A_0 = 0$ und $M = (M_n)_{n\geqslant 0}$ ein Martingal. Diese Zerlegung heißt Doob-Zerlegung *und es gilt*

$$A_n = \sum_{k=1}^{n} \mathbb{E}[X_k - X_{k-1}|\mathcal{F}_{k-1}], \quad n \geqslant 1.$$

X ist ein Submartingal genau dann, wenn A monoton wachsend ist.

Beweis. Existenz: Per Konstruktion ist A vorhersagbar. Wir definieren den adaptierten Prozess

$$M_0 := X_0 \quad \text{und} \quad M_{n+1} := M_n + \big(X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n]\big).$$

Dann gilt

$$\begin{aligned} M_n + A_n &= X_0 + \sum_{k=1}^{n} \big(X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]\big) + \sum_{k=1}^{n} \mathbb{E}[X_k - X_{k-1}|\mathcal{F}_{k-1}] \\ &= X_0 + \sum_{k=1}^{n} \big(X_k - \mathbb{E}[X_{k-1}|\mathcal{F}_{k-1}]\big) = X_n. \end{aligned}$$

Weiterhin ist

$$\begin{aligned} \mathbb{E}[|M_n|] &\leqslant \mathbb{E}[|X_0|] + \sum_{k=1}^{n} \mathbb{E}\big[|X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]|\big] \\ &\leqslant \mathbb{E}[|X_0|] + \sum_{k=1}^{n} \big(\mathbb{E}[|X_k|] + \mathbb{E}\big[\mathbb{E}[|X_k||\mathcal{F}_{k-1}]|\big]\big) \leqslant 2\sum_{k=0}^{n} \mathbb{E}[|X_k|] < \infty. \end{aligned}$$

Schließlich gilt

$$\mathbb{E}\big[M_{n+1} - M_n\big|\mathcal{F}_n\big] = \mathbb{E}\big[X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n]\big|\mathcal{F}_n\big] = 0.$$

Eindeutigkeit: Seien $X = M + A = M' + A'$ zwei Zerlegungen mit den genannten Eigenschaften. Dann ist $M - M' = A' - A$ ein vorhersagbares Martingal. Somit gilt

$$M_n - M_n' = \mathbb{E}[M_{n+1} - M_{n+1}'|\mathcal{F}_n] = M_{n+1} - M_{n+1}' \quad P\text{-f.s.}$$

Aus $M_0 - M_0' = 0$ folgt also $M_n - M_n' = 0$ und somit auch $A_n = A_n'$ für alle $n \geqslant 0$. □

7.2 Stoppzeiten

Es sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum mit einer Filtration $(\mathcal{F}_n)_{n\geqslant 0}$.

Definition 7.8. Eine Zufallsvariable τ mit Werten $\mathbb{N} \cup \{0, \infty\}$ heißt *Stoppzeit (bzgl. $(\mathcal{F}_n)_{n\geqslant 0}$), falls für jedes $n \geqslant 0$ gilt, dass $\{\tau \leqslant n\} \in \mathcal{F}_n$.*

Bemerkung 7.9. τ ist genau dann eine Stoppzeit, wenn $\{\tau = n\} \in \mathcal{F}_n$. Dies folgt aus den Gleichheiten $\{\tau = n\} = \{\tau \leqslant n\} \setminus \{\tau \leqslant n-1\}$ und $\{\tau \leqslant n\} = \bigcup_{k=0}^{n} \{\tau = k\}$ für jedes $n \geqslant 0$.

Beispiel 7.10.

(i) Jede deterministische Abbildung $\tau(\omega) = n_0$ für ein festes $n_0 \in \mathbb{N} \cup \{0, \infty\}$ ist eine Stoppzeit.

(ii) Sei $(X_n)_{n\geqslant 0}$ ein adaptierter Prozess und $B \in \mathscr{B}_{\mathbb{R}}$ eine Borelmenge. Dann ist die *Eintrittszeit* von $(X_n)_{n\geqslant 0}$ in B

$$\tau_B := \inf\{n \geqslant 0 : X_n \in B\} \qquad (\inf \emptyset := \infty)$$

eine Stoppzeit, denn

$$\{\tau_B \leqslant n\} = \bigcup_{k=0}^{n} \{X_k \in B\} \in \mathcal{F}_n, \qquad \text{für alle } n \geqslant 0.$$

(iii) Sind σ und τ Stoppzeiten, dann sind auch $\sigma \wedge \tau := \min(\sigma, \tau), \sigma \vee \tau := \max(\sigma, \tau)$ und $(\sigma + \tau)$ Stoppzeiten, da

$$\begin{aligned}\{\sigma \wedge \tau \leqslant n\} &= \{\sigma \leqslant n\} \cup \{\tau \leqslant n\} \in \mathcal{F}_n,\\ \{\sigma \vee \tau \leqslant n\} &= \{\sigma \leqslant n\} \cap \{\tau \leqslant n\} \in \mathcal{F}_n,\\ \{\sigma + \tau \leqslant n\} &= \bigcup_{k=0}^{n} \{\sigma \leqslant k\} \cap \{\tau \leqslant n-k\} \in \mathcal{F}_n.\end{aligned}$$

Satz 7.11 (Optional Stopping)**.** *Ist $X = (X_n)_{n\geqslant 0}$ ein (Sub-, Super-)Martingal und τ eine Stoppzeit (beide bzgl. $(\mathcal{F}_n)_{n\geqslant 0}$), dann ist $X_n^\tau := X_{\tau\wedge n}, n \geqslant 0$, wieder ein (Sub-, Super-)Martingal.*

Beweis. Sei X zunächst ein Submartingal. Für jedes $n \geqslant 0$ ist $X_n^\tau \in L^1(P)$, da

$$\mathbb{E}[|X_{\tau\wedge n}|] \leqslant \sum_{k=1}^{n} \mathbb{E}[|X_k|] < \infty.$$

Setzen wir nun $Y_n^\tau := \mathbb{1}_{\{n\leqslant\tau\}}$, ist $(Y_n^\tau)_{n\geqslant 1}$ vorhersagbar, da $\{n \leqslant \tau\} = \{\tau \leqslant n-1\}^c \in \mathcal{F}_{n-1}$. Es gilt

$$(Y^\tau \cdot X)_n = \sum_{k=1}^{n} Y_k^\tau (X_k - X_{k-1}) = \sum_{k=1}^{\tau\wedge n} (X_k - X_{k-1}) = X_{\tau\wedge n} - X_0$$

und

$$\mathbb{E}\big[(Y^\tau \cdot X)_{n+1} - (Y^\tau \cdot X)_n \big| \mathcal{F}_n\big] = \mathbb{E}\big[Y_{n+1}^\tau (X_{n+1} - X_n) | \mathcal{F}_n\big] = \underbrace{Y_{n+1}^\tau}_{\in\{0,1\}} \underbrace{\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n]}_{\geqslant 0} \geqslant 0.$$

Damit ist $X_n^\tau = (Y^\tau \cdot X)_n + X_0$ ein Submartingal.

Ist X ein Supermartingal, betrachte das Submartingal $(-X_n)_{n\geqslant 0}$. Ist X ein Martingal, folgt die Behauptung, da X dann ein Sub- und ein Supermartingal ist. □

Definition 7.12. Für eine Stoppzeit τ ist die *σ-Algebra der τ-Vergangenheit* gegeben durch

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leqslant n\} \in \mathcal{F}_n \text{ für alle } n \geqslant 0\}.$$

Bemerkung 7.13.

(i) Durch nachprüfen der Eigenschaften von σ-Algebren, sieht man leicht, dass $\mathcal{F}_\tau$ tatsächlich eine σ-Algebra ist.

(ii) τ ist $\mathcal{F}_\tau$-messbar, da für jedes $n \in \mathbb{N} \cup \{0, \infty\}$ und jedes $m \geqslant 0$ gilt

$$\{\tau = n\} \cap \{\tau \leqslant m\} = \begin{cases} \{\tau = n\} \in \mathcal{F}_n \subseteq \mathcal{F}_m, & m \geqslant n,\\ \emptyset \in \mathcal{F}_m, & m < n.\end{cases}$$

(iii) Für Stoppzeiten σ, τ mit $\sigma \leqslant \tau$ gilt $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$: Ist $A \in \mathcal{F}_\sigma$ folgt für alle $n \geqslant 0$ wegen $\sigma \leqslant \tau$

$$A \cap \{\tau \leqslant n\} = \underbrace{A \cap \{\sigma \leqslant n\}}_{\in\mathcal{F}_n} \cap \{\tau \leqslant n\} \in \mathcal{F}_n.$$

Lemma 7.14. *Ist $(X_n)_{n\geqslant 0}$ ein adaptierter Prozess und τ eine endliche Stoppzeit, dann ist $X_\tau : \Omega \ni \omega \mapsto X_{\tau(\omega)}(\omega)$ $\mathcal{F}_\tau$-messbar.*

Beweis. Für $B \in \mathscr{B}_{\mathbb{R}}$ ist $\{X_\tau \in B\} \in \mathcal{F}_\tau$, da für alle $n \geqslant 0$ gilt

$$\begin{aligned}\{X_\tau \in B\} \cap \{\tau \leqslant n\} &= \bigcup_{k=0}^{n} \{X_\tau \in B\} \cap \{\tau = k\} \\ &= \bigcup_{k=0}^{n} \underbrace{\{X_k \in B\} \cap \{\tau = k\}}_{\in \mathcal{F}_k} \in \mathcal{F}_n.\end{aligned}$$

□

Satz 7.15 (Optional Sampling)**.** *Es seien $X = (X_n)_{n \geqslant 0}$ ein Martingal (bzw. Submartingal) und σ, τ Stoppzeiten mit $\sigma \leqslant \tau \leqslant C$ für eine Konstante $C > 0$. Dann gilt $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] = X_\sigma$ (bzw. $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \geqslant X_\sigma$).*

Beweis. Schritt 1: Sei X ein Martingal. Da $\tau \leqslant C$ ist $X_\tau \in L^1(P)$ und in obigem Lemma haben wir gezeigt, dass X_σ $\mathcal{F}_\sigma$-messbar ist. Es genügt also zu zeigen, dass für jedes $A \in \mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$

$$\mathbb{E}[X_\tau \mathbb{1}_A] = \mathbb{E}[X_\sigma \mathbb{1}_A]$$

gilt. Die Abbildung $\rho := \sigma \mathbb{1}_A + \tau \mathbb{1}_{A^c}$ ist ebenfalls eine Stoppzeit, da

$$\{\rho \leqslant n\} = (A \cap \{\sigma \leqslant n\}) \cup (A^c \cap \{\tau \leqslant n\}) \in \mathcal{F}_n.$$

Der Optional-Stopping-Satz impliziert

$$\mathbb{E}[X_\rho] = \mathbb{E}[X_{\rho \wedge C}] = \mathbb{E}[X_0] = \mathbb{E}[X_\tau]$$

und somit $0 = \mathbb{E}[X_\rho - X_\tau] = \mathbb{E}[(X_\sigma - X_\tau)\mathbb{1}_A]$.

Schritt 2: Ist X ein Submartingal mit Doob-Zerlegung $X = A + M$ für einen monoton wachsenden Prozess A, so folgt aus der Monotonie, $\tau \geqslant \sigma$ und Schritt 1, dass

$$\begin{aligned}\mathbb{E}[X_\tau|\mathcal{F}_\sigma] &= \mathbb{E}[A_\tau|\mathcal{F}_\sigma] + \mathbb{E}[M_\tau|\mathcal{F}_\sigma] \\ &\geqslant \mathbb{E}[A_\sigma|\mathcal{F}_\sigma] + M_\sigma = A_\sigma + M_\sigma = X_\sigma.\end{aligned}$$

□

Literaturemfehlung

- Bauer, H. (1992). *Maß-und Integrationstheorie.* de Gruyter.
- Elstrodt, J. (2007). *Maß- und Integrationstheorie.* Springer, Berlin Heidelberg.
- Georgii, H.-O. (2007). *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik.* de Gruyter, Berlin.
- Klenke, A. (2006). *Wahrscheinlichkeitstheorie.* Springer, Heidelberg.
- Küchler, U. (2015). *Maßtheorie für Statistiker: Grundlagen der Stochastik.* Springer, Berlin.