Wahrscheinlichkeitstheorie II

Prof. Dr. Wilhelm Stannat

Inhaltsverzeichnis

Der vorliegende Text ist eine Zusammenfassung von Teil I der Vorlesung **Wahrscheinlichkeitstheorie II** im Wintersemester 2015/16 an der TU Berlin.

Korrekturen und Kommentare bitte per Email an `stannat@math.tu-berlin.de`

1 Maß- und Integrationstheorie

1.1 σ-Algebren und Maße

Wir wiederholen kurz die Grundbegriffe der Wahrscheinlichkeitstheorie anhand von Beispielen, die für die Vorlesung Wahrscheinlichkeitstheorie II wegweisend sind. Ausgangspunkt für die mathematische Beschreibung von Zufallsexperimenten ist der **Ergebnisraum** Ω, der alle möglichen Ergebnisse des zu beschreibenden Zufallsexperiments enthält, und der endlich, abzählbar unendlich oder sogar überabzählbar sein kann.

Beispiel 1.1. Beispiele für Zufallsexperimente und zugehörigem Ergebnisraum Ω:

(i) **Münzwurf** Die möglichen Ausgänge sind `Kopf` oder `Zahl`, die wir mit 0 und 1 identifizieren, sodass $\Omega = \{0, 1\}$.

(ii) n-**maliger Münzwurf** In diesem Falle sind die möglichen Ergebnisse $0-1$-Folgen der Länge n, sodass

$$\Omega = \big\{(x_1, x_2, \ldots, x_n) \bigm| x_i \in \{0,1\}\big\} =: \{0,1\}^n .$$

(iii) **Unendlicher Münzwurf** In diesem Falle sind die Ergebnisse unendliche $0-1$-Folgen

$$\Omega = \big\{(x_i)_{i\in\mathbb{N}} \bigm| x_i \in \{0,1\}\big\} =: \{0,1\}^{\mathbb{N}} .$$

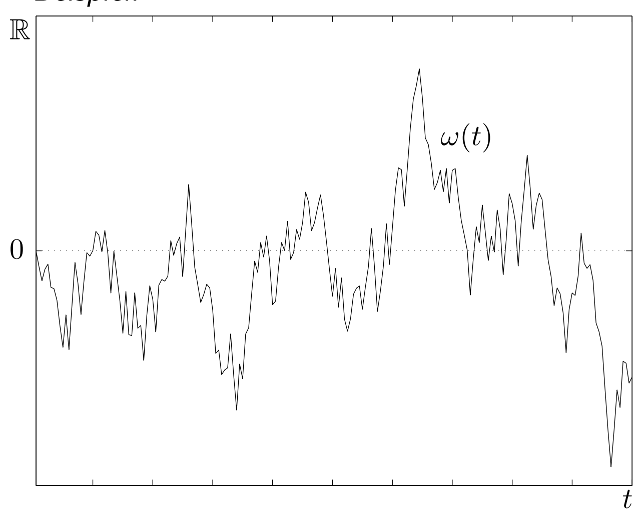
Der Ergebnisraum Ω ist überabzählbar und wir können ihn unter Ausnutzung der Binärentwicklung reeller Zahlen mit dem abgeschlossenen Intervall $[0,1] \subset \mathbb{R}$ identifizieren:

$$x = \sum_{i=1}^{\infty} x_i 2^{-i} .$$

(iv) **Zufallszahlen zwischen 0 und 1** $\Omega = [0,1]$.

(v) **Stochastische Prozesse, z.B. Brownsche Bewegung auf $\mathbb{R}$** In diesem Falle ist jede stetige Funktion auf $[0,1] \subset \mathbb{R}$ ein mögliches Ergebnis und damit $\Omega = \mathcal{C}\big([0,1]\big)$.

Beispiel:



Ereignisse

Teilmengen $A \subset \Omega$ heißen **Ereignisse**. Ist $\omega \in A$ so sagen wir, dass das Ereignis A **eingetreten ist**. Unter den Ereignissen besonders hervorzuheben sind

- das **sichere Ereignis** Ω,
- das **unmögliche Ereignis** $\emptyset$,
- die **Elementarereignisse** $\{\omega\}$ für $\omega \in \Omega$.

Kombination von Ereignissen

$A_1 \cup A_2, \quad \bigcup\limits_i A_i$	`mind. eines der Ereignisse` A_i `tritt ein`
$A_1 \cap A_2, \quad \bigcap\limits_i A_i$	`alle Ereignisse` A_i `treten ein`
$\limsup\limits_{n\to\infty} A_n := \bigcap\limits_n \bigcup\limits_{m \geqslant n} A_m$	`unendlich viele der` A_m `treten ein`
$\liminf\limits_{n\to\infty} A_n := \bigcup\limits_n \bigcap\limits_{m \geqslant n} A_m$	`alle bis auf endlich viele der` A_m `treten ein`

Beispiel 1.2. (i) **Münzwurf** 1 `geworfen`: $A = \{1\}$

(ii) n-**maliger Münzwurf** k `Einsen geworfen`:

$$A = \Big\{(x_1, \dots, x_n) \in \{0,1\}^n \;\Big|\; \sum_{i=1}^n x_i = k\Big\}.$$

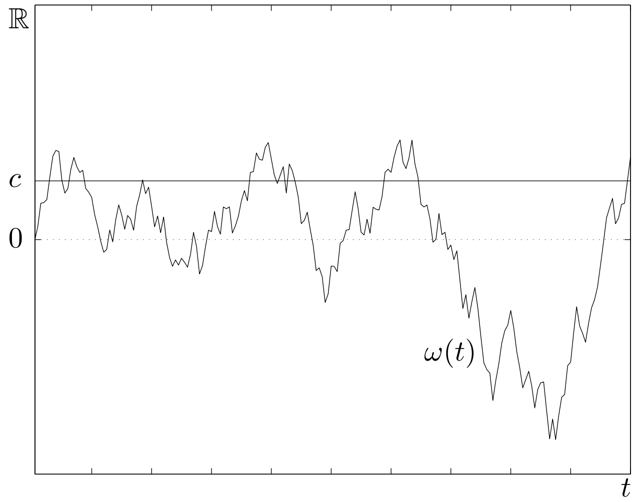
(iii) **Unendlicher Münzwurf** `rel. Häufigkeit der Einsen gleich` p:

$$A = \Big\{(x_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} \;\Big|\; \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i = p\Big\}.$$

(iv) **Zufallszahl zwischen** 0 **und** 1. `Zufallszahl` $\in [a,b]$: $A = [a,b] \subset \Omega = [0,1]$.

(v) **Stochastische Prozesse** `Überschreiten des Niveaus` c:

$$A = \big\{\omega \in \mathcal{C}\big([0,1]\big) \;\big|\; \max_{0\leqslant t\leqslant 1} \omega(t) > c\big\}.$$



Im Falle **diskreter** (d.h. endlicher oder abzählbar unendlicher) Ergebnisräume Ω können wir mithilfe einer Zähldichte (bzw. Wahrscheinlichkeitsfunktion)

$$p : \Omega \to [0,1] \text{ mit } \sum_{\omega\in\Omega} p(\omega) = 1$$

die Wahrscheinlichkeiten jedes einzelnen Ergebnisses festlegen.
Für jedes Ereignis $A \subset \Omega$ kann dann seine Wahrscheinlichkeit $P(A)$ einfach durch Addition der Einzelwahrscheinlichkeit ausgerechnet werden

$$P(A) = \sum_{\omega\in A} p(\omega)$$

(siehe VL Wahrscheinlichkeitstheorie I).
Im überabzählbarem Fall gibt es jedoch keine sinnvolle Weise eine überabzählbare Summe von Wahrscheinlichkeiten zu bilden. Es gibt keine konsistente mathematische Theorie, die aufbauend auf Zähldichten für Ereignisse mit überabzählbaren Ergebnissen (nichttriviale) Wahrscheinlichkeiten zuordnen kann. Als Ausweg bietet sich daher an, den Ereignissen direkt Wahrscheinlichkeiten zuzuordnen. Dies ist im Falle überabzählbarer Ereignisräume allerdings im Allgemeinen nicht für beliebige Teilmengen von Ω, d.h. für alle Elemente der Potenzmenge $\mathcal{P}(\Omega)$, möglich, sondern nur für bestimmte Teilsysteme $\mathcal{A}$. Um mit diesen insbesondere obige Kombinationen von Ereignissen bilden zu können, sollten sie unter Komplementbildung und Vereinigung abgeschlossen sein im Sinne folgender Definition:

Definition 1.3. Ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt σ-**Algebra** falls

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

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

(c) $(A_n)_{n\in\mathbb{N}} \in \mathcal{A} \quad \Longrightarrow \quad \bigcup_{n\geq 1} A_n \in \mathcal{A}$

Das Paar $(\Omega, \mathcal{A})$ heißt **messbarer Raum** oder **Raum der messbaren Ereignisse**.
Eine Teilmenge $A \subseteq \Omega$ mit $A \in \mathcal{A}$ heißt $\mathcal{A}$-**messbar** (oder auch einfach nur messbar, wenn der Bezug zur zugrundeliegenden σ-Algebra $\mathcal{A}$ klar ist).

Bemerkung 1.4. *(i) Es sei $\mathcal{A}$ eine σ-Algebra. Dann gilt:*

- $\emptyset = \Omega^c \in \mathcal{A}$.
- $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *impliziert*

$$\bigcap_{i\in\mathbb{N}} A_i = \Big(\bigcup_{i\in\mathbb{N}} A_i^c\Big)^c \in \mathcal{A}.$$

- $A_1, \ldots, A_n \in \mathcal{A}$ *impliziert*

$$\bigcup_{i=1}^n A_i \in \mathcal{A} \quad \textit{und} \quad \bigcap_{i=1}^n A_i \in \mathcal{A}$$

(betrachte $A_m = \emptyset$ für alle $m > n$, bzw. $A_m = \Omega$ für alle $m > n$).

- $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *impliziert*

$$\bigcap_n \bigcup_{m\geqslant n} A_m \in \mathcal{A} \quad \textit{und} \quad \bigcup_n \bigcap_{m\geqslant n} A_m \in \mathcal{A}.$$

(ii) Das System $\{\emptyset, \Omega\}$ ist eine σ-Algebra und offensichtlich die kleinste σ-Algebra auf Ω. Da in diesem Falle nur das sichere und das unmögliche Ereignisse messbar sind, spricht man auch von der **trivialen σ-Algebra**.

(iii) Die Potenzmenge $\mathcal{P}(\Omega)$ ist eine σ-Algebra und offensichtlich die größte σ-Algebra auf Ω. Jede andere σ-Algebra ist also eine Teil-σ-Algebra von $\mathcal{P}(\Omega)$.

(iv) Es sei I eine (beliebige) Indexmenge (möglicherweise überabzählbar) und für $i \in I$ sei $\mathcal{A}_i$ eine σ-Algebra. Dann ist auch $\bigcap_{i\in I} \mathcal{A}_i$ wieder eine σ-Algebra.

(v) **Typische Konstruktion einer σ-Algebra** *Es sei $\mathcal{A}_0 \neq \emptyset \subset \mathcal{P}(\Omega)$ eine Klasse von Ereignissen. Dann ist das folgende Mengensystem*

$$\sigma(\mathcal{A}_0) := \bigcap_{\substack{\mathcal{B} \text{ ist } \sigma\text{-Algebra},\\ \mathcal{A}_0 \subset \mathcal{B}}} \mathcal{B}$$

wieder eine σ-Algebra. $\sigma(\mathcal{A}_0)$ ist die kleinste σ-Algebra die $\mathcal{A}_0$ enthält und heißt **die von $\mathcal{A}_0$ erzeugte σ-Algebra**.

Beispiel 1.5. Es sei Ω ein topologischer Raum und $\mathcal{A}_o$ das Mengensystem der offenen Teilmengen von Ω. Dann heißt $\mathcal{B}(\Omega) := \sigma(\mathcal{A}_o)$ die **Borel-σ-Algebra** auf Ω (oder auch σ-Algebra der **Borelmengen**).

Beispiele für Borelmengen: abgeschlossene Mengen, abzählbare Vereinigungen von abgeschlossenen Mengen, usw. ...

Man beachte: nicht jede Teilmenge eines topologischen Raumes ist eine Borelmenge, d.h. im Allgemeinen gilt $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.

Das wichtigste Beispiel sind die Borelmengen $\mathcal{B}(\mathbb{R})$ auf $\mathbb{R}$. Mann kann zeigen, dass

$$\begin{aligned}\mathcal{B}(\mathbb{R}) &= \sigma(\{(a,b) : -\infty \leq a < b \leq +\infty\}) \\ &= \sigma(\{(\infty, t] : t \in \mathbb{R}\}) \\ &= \sigma(\{(\infty, t] : t \in \mathbb{Q}\}),\end{aligned}$$

d.h. $\mathcal{B}(\mathbb{R})$ wird auch von dem System der offenen Intervalle, dem der links unbeschränkten und rechts beschränkten, abgeschlossenen Intervalle (mit rationalem Randpunkt) erzeugt (Übungsaufgaben).

Wahrscheinlichkeiten

Definition 1.6. Es sei $(\Omega, \mathcal{A})$ ein messbarer Raum und

$$\mu : \mathcal{A} \to [0, \infty)$$

eine Abbildung.

(i) μ heißt ein **Maß**, falls

  (a) $\mu(\emptyset) = 0$

  (b) μ ist σ-**additiv**, d.h., für jede Folge $(A_n)_{n\in\mathbb{N}}$ paarweiser disjunkter Mengen in $\mathcal{A}$ gilt

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n)\,.$$

  In diesem Fall heißt das Tripel $(\Omega, \mathcal{A}, \mu)$ ein **Maßraum**.

(ii) μ heißt **endlich**, falls $\mu(\Omega) < \infty$.

(iii) μ heißt ein **Wahrscheinlichkeitsmaß**, falls μ ein Maß ist mit $\mu(\Omega) = 1$. In diesem Falle heißt das Tripel $(\Omega, \mathcal{A}, \mu)$ ein **Wahrscheinlichkeitsraum** und wir schreiben meist P statt μ.

Beispiel 1.7. (i) **Münzwurf** Es sei $\mathcal{A} := \mathcal{P}(\Omega) = \big\{\emptyset, \{0\}, \{1\}, \{0,1\}\big\}$.
Werfen einer **fairen** Münze bedeutet, dass `Kopf` und `Zahl` gleiche Wahrscheinlichkeit 0.5 haben. Folglich

$$P(\{0\}) := P(\{1\}) := \frac{1}{2}, \quad P(\emptyset) := 0, \quad P(\underbrace{\{0,1\}}_{=\Omega}) := 1.$$

(iii) **Unendlicher Münzwurf** $\Omega = \{0,1\}^{\mathbb{N}}$.
Es sei $\mathcal{A} := \sigma(\mathcal{A}_0)$ mit

$$\begin{aligned}\mathcal{A}_0 := \big\{B \subset \Omega \;\big|\; \exists\, n \in \mathbb{N} \text{ und } B_0 \in \mathcal{P}\big(\{0,1\}^n\big), \\ \text{so dass } B = B_0 \times \{0,1\} \times \{0,1\} \times \ldots\big\}.\end{aligned}$$

Ein Ereignis B ist in $\mathcal{A}_0$ enthalten, falls es nur von **endlich vielen** Münzwürfen abhängt!
Für $\bar{x}_1, \dots, \bar{x}_n \in \{0,1\}$ definiere man nun

$$P\big(\underbrace{\{(x_1, x_2, \dots) \in \{0,1\}^{\mathbb{N}} \mid x_1 = \bar{x}_1, \dots, x_n = \bar{x}_n\}}_{\in \mathcal{A}_0}\big) := 2^{-n}.$$

Wir werden später sehen, dass P (eindeutig) zu einem Wahrscheinlichkeitsmaß auf $\mathcal{A} = \sigma(\mathcal{A}_0)$ fortgesetzt werden kann. Für dieses Wahrscheinlichkeitsmaß gilt

$$P\left(\left\{(x_1, x_2, \dots) \in \{0,1\}^{\mathbb{N}} \;\middle|\; \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{2}\right\}\right) = 1.$$

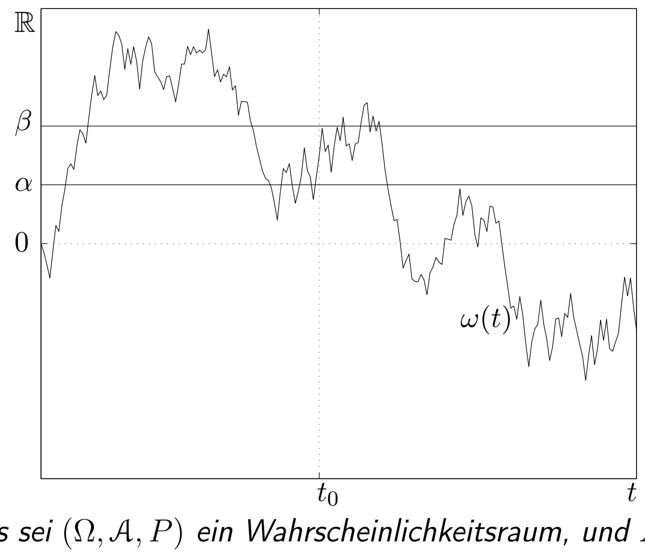
(Beweis: Später!)

(v) **Stochastische Prozesse** $\Omega = \mathcal{C}([0,1])$, $P =$ Wienermaß ("Brownsche Bewegung") Für festes $t_0 \in \mathbb{R}_+$ und $\alpha, \beta \in \mathbb{R}$ gilt

$$P\big(\{\omega \mid \omega(t_0) \in [\alpha, \beta]\}\big) := \frac{1}{\sqrt{2\pi t_0}} \int_\alpha^\beta e^{-\frac{x^2}{2t_0}}\, \mathrm{d}x$$

(Normalverteilung $N(0, t_0)$).
Wie groß ist nun die Wahrscheinlichkeit $P\big(\{\omega \mid \max_{0\leqslant t\leqslant 1} \omega(t) > c\}\big)$?

Antwort später!



Bemerkung 1.8. *Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, und $A_1, \dots, A_n \in \mathcal{A}$ paarweise disjunkt. Dann gilt:*

$$P\Big(\dot{\bigcup_{i\leqslant n}} A_i\Big) = \sum_{i=1}^{n} P(A_i)$$

(setze einfach $A_m = \emptyset$ für alle $m > n$). Insbesondere

$$\begin{aligned} A, B \in \mathcal{A},\ A \subset B &\Rightarrow P(B) = P(A) + P(B \setminus A) \\ &\Rightarrow P(B \setminus A) = P(B) - P(A), \end{aligned}$$

und $P(A^c) = P(\Omega \setminus A) = P(\Omega) - P(A) = 1 - P(A)$.
P *ist* **subadditiv**, *d.h. für* $A, B \in \mathcal{A}$ *gilt*

$$\begin{aligned} P(A \cup B) &= P\big(A \,\dot\cup\, [B \setminus (A \cap B)]\big) \\ &= P(A) + P(B) - P(A \cap B) \\ &\leqslant P(A) + P(B), \end{aligned} \tag{1.1}$$

und eine Induktion nach n *ergibt die* **Formel von Sylvester** *(siehe Satz 1.7, VL Wahrscheinlichkeitstheorie I):*
Es sei I *eine endliche Indexmenge,* A_i, $i \in I$, *eine Familie von Ereignissen in* $\mathcal{A}$ *(nicht notwendigerweise disjunkt). Dann gilt*

$$\begin{aligned} P\Big(\bigcup_{i\in I} A_i\Big) \quad &= \quad \sum_{\substack{J\subset I,\\ J\neq\emptyset}} (-1)^{|J|-1} \cdot P\Big(\bigcap_{j\in J} A_j\Big) \qquad (1.2) \\ &\overset{I=\{1,\dots,n\}}{=} \sum_{k=1}^{n} (-1)^{k-1} \cdot \sum_{1\leqslant i_1<\dots<i_k\leqslant n} P(A_{i_1} \cap \dots \cap A_{i_k}). \end{aligned}$$

Satz 1.9. *Es sei* $\mathcal{A}$ *eine* σ*-Algebra, und* $P : \mathcal{A} \to \mathbb{R}_+$ *eine Abbildung mit* $P(\Omega) = 1$. *Dann sind äquivalent:*

(i) P *ist ein Wahrscheinlichkeitsmaß.*

(ii) P *ist* **endlich additiv**, *d.h.* $A, B \in \mathcal{A}$, $A \cap B = \emptyset$ *impliziert* $P(A \,\dot\cup\, B) = P(A) + P(B)$,

und

stetig von unten, *d.h.* $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *mit* $A_i \subset A_{i+1}$ *für alle* $i \in \mathbb{N}$, *impliziert*

$$P\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \lim_{i\to\infty} P(A_i).$$

(iii) P *ist* **endlich additiv** *und* **stetig von oben**, *d.h.* $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *mit* $A_i \supset A_{i+1}$ *für alle* $i \in \mathbb{N}$, *impliziert*

$$P\Big(\bigcap_{i\in\mathbb{N}} A_i\Big) = \lim_{i\to\infty} P(A_i).$$

Korollar 1.10 (σ-Subadditivität). *Es sei* A_i, $i \in \mathbb{N}$, *eine Folge von Ereignissen in* $\mathcal{A}$ *(nicht notwendigerweise paarweise disjunkt). Dann gilt:*

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \leqslant \sum_{i=1}^{\infty} P(A_i).$$

Beweis.

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \overset{1.9}{=} \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} A_i\Big) \overset{(1.1)}{\leqslant} \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i). \qquad \square$$

Lemma 1.11 (Borel-Cantelli, Teil 1). *Es sei* $A_i \in \mathcal{A}$, $i \in \mathbb{N}$. *Dann gilt*

$$\sum_{i=1}^{\infty} P(A_i) < \infty \quad \Rightarrow \quad P\Big(\underbrace{\bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} A_m}_{=: \limsup\limits_{n\to\infty} A_n}\Big) = 0.$$

Beweis. Da

$$\bigcup_{m\geqslant n} A_m \overset{n\to\infty}{\searrow} \bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} A_m$$

folgt aus der Stetigkeit von oben von P dass

$$P\big(\limsup_{n\to\infty} A_n\big) \overset{1.9}{=} \lim_{n\to\infty} P\Big(\bigcup_{m\geqslant n} A_m\Big) \overset{1.10}{\leqslant} \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) = 0,$$

wegen $\sum_{m=1}^{\infty} P(A_m) < \infty$. □

Beispiel 1.12. (i) **Gleichverteilung auf** $[0,1]$ Es sei $\Omega = [0,1]$ und $\mathcal{A}$ die Borel-σ-Algebra auf Ω $(= \sigma\big(\{[a,b] \mid 0 \leqslant a \leqslant b \leqslant 1\}\big))$. Weiterhin sei P die Einschränkung des Lebesguemaßes auf die Borelmengen in $[0,1]$. Dann ist $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Das Wahrscheinlichkeitsmaß P heißt Gleichverteilung auf $[0,1]$, da $P([a,b]) = b - a$ für $0 \leq a \leq b \leq 1$ (**translationsinvariant**).

(ii) **Dirac-Maß, Einpunktverteilung** Es sei $\Omega \neq \emptyset$ und $\omega_0 \in \Omega$. Weiter sei $\mathcal{A}$ eine beliebige σ-Algebra auf Ω (z.B. $\mathcal{A} = \mathcal{P}(\Omega)$). Dann definiert

$$P(A) := 1_A(\omega_0) := \begin{cases} 1 & \text{für } \omega_0 \in A \\ 0 & \text{für } \omega_0 \notin A. \end{cases}$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$. P heißt **Diracmaß** (bzw. **Einpunktverteilung in** ω_0) und wird mit δ_{ω_0} bezeichnet.

(iii) **Konvexkombinationen von Wahrscheinlichkeitsmaßen** Es sei $\Omega \neq \emptyset$ und $\mathcal{A}$ eine σ-Algebra auf Ω. Weiter sei I eine abzählbare Indexmenge. Für $i \in I$ sei P_i ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. Weiter sei $\alpha_i \in [0,1]$, $i \in I$, mit $\sum_{i\in I} \alpha_i = 1$ gegeben. Dann ist $P := \sum_{i\in I} \alpha_i \cdot P_i$ wieder ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$.

Dies gilt insbesondere für

$$P := \sum_{i\in I} \alpha_i \cdot \delta_{\omega_i}$$

mit $\omega_i \in \Omega$, $i \in I$.

Das letzte Beispiel klärt die allgemeine Struktur von Wahrscheinlichkeitsmaßen für diskrete Stichprobenräume aufgrund des folgenden Satzes:

Satz 1.13. *Es sei* Ω *abzählbar.*

(i) Ist $p : \Omega \to [0,1]$ eine Zähldichte, so definiert

$$P(A) := \sum_{\omega \in A} p(\omega) \qquad \forall A \subset \Omega$$

ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{P})$.

(ii) Umgekehrt kann jedes Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{P})$ in dieser Form dargestellt werden mit $p(\omega) := P(\{\omega\})$ für alle $\omega \in \Omega$.

Beweis. (i)

$$P = \sum_{\omega \in \Omega} p(\omega) \cdot \delta_\omega \,.$$

(ii) Übung. □

1.2 Konstruktion von Maßen

Für überabzählbare Stichprobenräume sind (Wahrscheinlichkeits-) Maße schwieriger zu definieren. Denkt man etwa an die Gleichverteilung λ auf $([0,1], \mathcal{B}([0,1]))$, so ist $\lambda(A)$ zunächst einmal nur für Intervalle $A = [a,b]$ explizit gegeben, $\lambda([a,b]) = b - a$, und dann natürlich auch für disjunkte Vereinigungen von Intervallen, aber eben nicht für alle Borelmengen. Es ist noch nicht einmal von vorneherein klar, ob es überhaupt auf $([0,1], \mathcal{B}([0,1]))$ ein solches Maß gibt, das auf Intervallen mit λ übereinstimmt. In der Tat gibt es zum Beispiel keine Fortsetzung von λ zu einem Maß auf der Potenzmenge $\mathcal{P}([0,1])$ von $[0,1]$. Einen Beweis hierfür findet man in [Ba91, Satz 8.1].

Es ist von daher also wichtig, entscheiden zu können, ob auf Teilsystemen einer σ-Algebra $\mathcal{A}$ gegebene Mengenfunktionen sich (eindeutig) zu einem Maß auf ganz $\mathcal{A}$ fortsetzen lassen. Für die Existenz einer Fortsetzung liefert der Forsetzungssatz von Caratheodory (s.u.) ein hinreichendes Kriterium. Doch zunächst einmal klären wir die Eindeutigkeit.

Dynkin-Systeme und Eindeutigkeit von (Wahrscheinlichkeits-) Maßen

Es sei $\Omega \neq \emptyset$.

Definition 1.14. Ein Mengensystem $\mathcal{D} \subset \mathcal{P}(\Omega)$ heißt **Dynkin-System**, falls gilt :

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

(ii) $A \in \mathcal{D} \quad \Rightarrow \quad A^c \in \mathcal{D}$.

(iii) Für $A_i \in \mathcal{D}$, $i \in \mathbb{N}$, paarweise disjunkt, folgt

$$\dot{\bigcup_{i \in \mathbb{N}}} A_i \in \mathcal{D}.$$

Beispiel 1.15. (i) Jede σ-Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ ist ein Dynkin-System.

(ii) Es seien μ_1, μ_2 zwei endliche Maße auf $(\Omega, \mathcal{A})$ mit $\mu_1(\Omega) = \mu_2(\Omega)$. Dann ist das Mengensystem

$$\mathcal{D} := \{A \in \mathcal{A} \mid \mu_1(A) = \mu_2(A)\}$$

ein Dynkin-System.

Bemerkung 1.16. *(i) Es sei $\mathcal{D}$ ein Dynkin-System. Dann gilt*

$$A, B \in \mathcal{D}\,, A \subset B \qquad \Rightarrow \qquad B \setminus A = (B^c \dot{\cup} A)^c \in \mathcal{D}$$

(ii) Jedes Dynkin-System, das abgeschlossen ist unter endlichen Durchschnitten (Notation: $\cap$ stabil), ist eine σ-Algebra, denn:

(a) $A, B \in \mathcal{D} \quad \Rightarrow \quad A \cup B = A \dot{\cup} \underbrace{\big(B \setminus \underbrace{(A \cap B)}_{\in \mathcal{D}}\big)}_{\overset{(i)}{\in} \mathcal{D}} \in \mathcal{D}.$

(b) $A_i \in \mathcal{D},\ i \in \mathbb{N} \quad \Rightarrow \quad \bigcup_{i \in \mathbb{N}} A_i = \dot{\bigcup}_{i \in \mathbb{N}} \underbrace{\Big[A_i \cap \underbrace{\Big(\bigcup_{n=1}^{i-1} A_n\Big)^c}_{\overset{(a)}{\in} \mathcal{D}}\Big]}_{\substack{\in \mathcal{D} \text{ nach Ann.,} \\ \text{paarweise disjunkt}}} \in \mathcal{D}.$

Satz 1.17. *Es sei $\mathcal{B} \subset \mathcal{P}(\Omega)$ ein $\cap$-stabiles Mengensystem. Dann gilt*

$$\sigma(\mathcal{B}) = \delta(\mathcal{B})\,,$$

wobei

$$\delta(\mathcal{B}) := \bigcap_{\substack{\mathcal{D} \text{ Dynkin-System} \\ \mathcal{B} \subset \mathcal{D}}} \mathcal{D}$$

das **von $\mathcal{B}$ erzeugte Dynkin-System** *ist.*

Beweis. Da jede σ-Algebra ein Dynkin-System ist, ist klar, dass $\delta(\mathcal{B}) \subset \sigma(\mathcal{B})$. Wenn wir nun zeigen, dass $\delta(\mathcal{B})$ eine σ-Algebra ist, folgt also Gleichheit. Nach Bemerkung 1.16(ii) reicht es aus zu zeigen, dass $\delta(\mathcal{B})$ $\cap$-stabil ist. Dazu betrachten wir für beliebiges $D \in \delta(\mathcal{B})$ das Mengensystem

$$\mathcal{D}_D = \{A \subset \Omega : A \cap D \in \delta(\mathcal{B})\}\,.$$

Es ist einfach zu sehen, dass $\mathcal{D}_D$ ein Dynkin-System ist. Für jedes $B \in \mathcal{B}$ gilt nun wegen der Annahme an $\mathcal{B}$, dass $\mathcal{B} \subset \mathcal{D}_B$ und damit dann auch $\delta(\mathcal{B}) \subset \mathcal{D}_B$. Dies bedeutet nun aber, dass für alle $D \in \delta(\mathcal{B})$ und jedes $B \in \mathcal{B}$ $D \cap B \in \delta(\mathcal{B})$ gilt. Damit gilt dann aber umgekehrt auch $B \in \mathcal{D}_D$ für alle $B \in \mathcal{B}$ und damit also auch $\delta(\mathcal{B}) \subset \mathcal{D}_D$ für alle $D \in \delta(\mathcal{B})$. $\square$

Satz 1.18. *Es seien μ, ν endliche Maße auf $(\Omega, \mathcal{A})$ mit $\mu(\Omega) = \nu(\Omega)$, und $\mathcal{B} \subseteq \mathcal{A}$ ein $\cap$ - stabiles Mengensystem. Dann gilt*

$$\mu(A) = \nu(A) \text{ für alle } A \in \mathcal{B} \quad \Rightarrow \quad \mu = \nu \text{ auf } \sigma(\mathcal{B}).$$

Beweis. Das Mengensystem

$$\mathcal{D} := \left\{ A \in \mathcal{A} \mid \mu(A) = \nu(A) \right\}$$

ist ein Dynkin-System, das $\mathcal{B}$ enthält. Folglich,

$$\sigma(\mathcal{B}) \overset{1.17}{=} \delta(\mathcal{B}) \subset \mathcal{D}.$$

□

Beispiel 1.19. (i) Für $p \in [0, 1]$ ist das Wahrscheinlichkeitsmaß P_p auf $(\Omega := \{0, 1\}^{\mathbb{N}}, \mathcal{A})$ eindeutig bestimmt durch

$$P_p\left(X_1 = x_1, \ldots, X_n = x_n\right) = p^k (1-p)^{n-k}, \quad \text{mit } k := \sum_{i=1}^n x_i$$

für alle $x_1, \ldots, x_n \in \{0, 1\}$, $n \in \mathbb{N}$, da das Mengensystem aller Zylindermengen

$$\{X_1 = x_1, \ldots, X_n = x_n\}, \quad n \in \mathbb{N}_0, \; x_1, \ldots, x_n \in \{0, 1\}$$

ein $\cap$-stabiler Erzeuger von $\mathcal{A}$ ist (siehe Beispiel 1.7 (iii)).

(ii) Ein Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist eindeutig durch seine Verteilungsfunktion $F(t)$ $(:= \mu\big(]-\infty, t\,]\big))$ bestimmt, denn

$$\mu\big(]a, b]\big) = F(b) - F(a),$$

und das Mengensystem der Intervalle $]a, b], a, b \in \mathbb{R}$, ist ein $\cap$-stabiler Erzeuger der Borelschen σ-Algebra $\mathcal{B}(\mathbb{R})$.

Wir können obigen Eindeutigkeitssatz noch ausweiten auf σ-endliche Maße. Ein Maß μ auf $(\Omega, \mathcal{A})$ heißt σ-**endlich**, falls eine aufsteigende Folge $\Omega_n \in \mathcal{A}$, $n \geq 1$, existiert mit $\Omega = \bigcup_{n\geq 1}^{\infty} \Omega_n$ und $\mu(\Omega_n) < \infty$ für alle $n \geq 1$. In diesem Falle schreiben wir einfach $\Omega_n \uparrow \Omega$.

Satz 1.20. *Es seien μ, ν zwei σ-endliche Maße auf $(\Omega, \mathcal{A})$, $\mathcal{B} \subseteq \mathcal{A}$ ein $\cap$-stabiles Erzeugendensystem von $\mathcal{A}$. Stimmen μ und ν auf $\mathcal{B}$ überein und gibt es weiterhin eine Folge $\Omega_n \uparrow \Omega \in \mathcal{B}$ mit $\mu(\Omega_n) = \nu(\Omega_n) < \infty$ für alle $n \in \mathbb{N}$, so gilt $\mu = \nu$.*

Beweis. Für $n \in \mathbb{N}$ definieren wir die Einschränkungen $\mu_n(A) := \mu(A \cap \Omega_n)$, $A \in \mathcal{A}$, von μ (bzw. $\nu_n(A) := \nu(A \cap \Omega_n)$, $A \in \mathcal{A}$) von μ (bzw. von ν) auf Ω_n. Dann sind μ_n und ν_n endliche Maße auf $\mathcal{A}$, die auf $\mathcal{B}$ übereinstimmen. Nach Satz 1.18 folgt $\mu_n = \nu_n$ für alle n und damit für beliebiges $A \in \mathcal{A}$

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

□

Fortsetzung von Prämaßen

Wir wollen Mengenfunktionen auf geeigneten Mengensystemen zu Maßen auf σ-Algebren erweitern. Als geeignete Mengensysteme bieten sich an:

Definition 1.21. Ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt **Algebra**, falls

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

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

(c) $A, B \in \mathcal{A} \quad \Longrightarrow \quad A \cup B \in \mathcal{A}$

Bemerkung 1.22. *(i) Ist $\mathcal{A}$ eine Algebra so folgt*

$$A, B \in \mathcal{A} \quad \Longrightarrow \quad A \cap B \in \mathcal{A}\,,$$

denn

$$A \cap B = \Omega \setminus (A^c \cup B^c)\ .$$

(ii) Eine Algebra ist insbesondere abgeschlossen unter **endlichen** *Vereinigungen oder* **endlichen** *Durchschnitten, im allgemeinen aber eben nicht unter* **abzählbar unendlichen** *Vereinigungen oder* **abzählbar unendlichen** *Durchschnitten. Dies ist der fundamentale Unterschied zu σ-Algebren.*

(iii) Das System

$$\left\{ A = \bigcup_{i=1}^{n}]a_i, b_i\rangle : -\infty \le a_i \le b_i \le +\infty \right\}$$

mit der Konvention $]a, b\rangle =]a, b]$ für $b < \infty$ und $]a, \infty\rangle =]a, \infty[$ aller **endlichen** *Vereinigungen von (möglicherweise unbeschränkten) nach links halboffenen Intervallen auf $\mathbb{R}$ ist eine Algebra, aber keine σ-Algebra.*

Als geeignete Mengenfunktion betrachten wir Inhalte und Prämaße.

Definition 1.23. Es sei $\mathcal{A}$ eine Algebra.

(i) Eine Abbildung

$$\mu : \mathcal{A} \to [0, \infty]$$

heißt **Inhalt**, falls gilt:

(a) $\mu(\emptyset) = 0$

(b) μ ist **endlich additiv**, d.h., für $A, B \in \mathcal{A}$ disjunkt gilt $\mu(A \cup B) = \mu(A) + \mu(B)$.

(ii) Ein Inhalt

$$\mu : \mathcal{A} \to [0, \infty]$$

heißt **Prämaß**, falls μ **zusätzlich auch σ-additiv** ist, d.h., für jede Folge $(A_n)_{n\in\mathbb{N}}$ paarweiser disjunkter Mengen in $\mathcal{A}$ mit $\bigcup_{n\ge1} A_n \in \mathcal{A}$ gilt

$$\mu\left(\bigcup_{n\ge1} A_n\right) = \sum_{n\ge1} \mu(A_n)\,.$$

(iii) Ein Prämaß μ heißt σ-**endlich**, falls eine aufsteigende Folge Ω_n, $n \in \mathbb{N}$, von Mengen $\Omega_n \in \mathcal{A}$ existiert mit $\Omega = \bigcup_{n\geq 1} \Omega_n$ und $\mu(\Omega_n) < \infty$ für alle $n \geq 1$.

Man beachte in Teil (ii) der obigen Definition, dass die Bedingung $\bigcup_{n\geq 1} A_n \in \mathcal{A}$ für Algebren $\mathcal{A}$ nicht aus der Definition folgt.

Lemma 1.24. *Es sei μ ein endlicher Inhalt auf einer Algebra $\mathcal{A}$. Dann sind äquivalent:*

(i) μ ist σ-additiv (d.h. ein Prämaß),

(ii) μ ist $\emptyset$-stetig, d.h., für alle Mengen $A_1, A_2, \ldots \in \mathcal{A}$ mit $A_n \downarrow \emptyset$ gilt $\lim_{n\to\infty} \mu(A_n) = 0$.

Beweis. $(i) \implies (ii)$ Es gelte $A_n \downarrow \emptyset$. Wir definieren $B_n = A_n \setminus A_{n+1}$ für $n \in \mathbb{N}$. Dann sind die Mengen B_n paarweise disjunkt, und es gilt $A_n = \bigcup_{m=n}^{\infty} B_m$. Folglich ist $\mu(A_n) = \sum_{m=n}^{\infty} \mu(B_m)$. Da diese Summe konvergiert, folgt $\lim_{n\to\infty} \mu(A_n) = 0$.
$(ii) \implies (i)$ Es sei $(B_n)_{n\in\mathbb{N}}$ eine Folge von paarweise disjunkten Mengen in $\mathcal{A}$, mit $B = \bigcup_{n\in\mathbb{N}} B_n \in \mathcal{A}$. Für $n \in \mathbb{N}$ betrachte $A_{n+1} = \bigcup_{m=n+1}^{\infty} B_m = B \setminus (B_1 \cup \ldots \cup B_n)$ ($\in \mathcal{A}$). Dann sind die Mengen $B_1, \ldots, B_n, A_{n+1}$ paarweise disjunkt und in $\mathcal{A}$. Wegen der endlichen Additivität gilt

$$\mu(B) = \sum_{m=1}^{n} \mu(B_m) + \mu(A_{n+1})$$

und wegen $A_{n+1} \downarrow \emptyset$ gilt $\lim_{n\to\infty} \mu(A_{n+1}) = 0$, also

$$\mu(B) = \sum_{m=1}^{\infty} \mu(B_m)\,,$$

d.h. μ ist σ-additiv. □

Satz 1.25. *(Carathéodory) Es sei μ_0 ein σ-endliches Prämaß auf einer Algebra $\mathcal{A}$. Dann gibt es genau ein Maß μ auf $\sigma(\mathcal{A})$, das μ_0 erweitert, d.h. das auf $\mathcal{A}$ mit μ_0 übereinstimmt.*

Der Existenzteil dieses Satzes wird z.B. in [Ba91, I.5] bewiesen. Die Eindeutigkeit folgt direkt aus Satz 1.20.

Verteilungsfunktionen

Ist μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ so ist die zugehörige Verteilungsfunktion $F(t) := \mu(]-\infty, t])$ monoton wachsend, rechtsstetig mit $0 \leq F \leq 1$, sowie $\lim_{t\to-\infty} F(t) = 0$ und $\lim_{t\to\infty} F(t) = 1$ (siehe Satz 3.6, VL Wahrscheinlichkeitstheorie I). Es ist leicht einzusehen, dass auch für allgemeine Maße μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $\mu(]s, t]) < \infty$ für alle $s, t \in R$ mit $s < t$ die Verteilungsfunktion F monoton wachsend und rechtsstetig ist (mit Werten in $[0, \infty]$). Mithilfe des Fortsetzungssatzes von Carathéodory können wir nun umgekehrt zeigen (siehe Bemerkung 3.8 aus Wahrscheinlichkeitstheorie I):

Satz 1.26. *Es sei $F : \mathbb{R} \to \mathbb{R}$ monoton wachsend und rechtsstetig. Dann gibt es genau ein Maß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $\mu(]s,t]) = F(t) - F(s)$ für alle $s, t \in \mathbb{R}$, $s < t$.*

Beweis. Es sei $\mathcal{J}$ die Menge aller Intervalle der Form $]a,b\rangle$ mit $]a,b\rangle =]a,b]$ für $-\infty \leq a \leq b < \infty$ und und $]a,\infty\rangle =]a,\infty[$ für $-\infty \leq a < \infty$. Desweiteren sei $\mathcal{F}$ die Vereinigung aller **endlichen, disjunkten** Vereinigungen von Intervallen aus $\mathcal{J}$. Wie bereits in Teil (iii) von Bemerkung 1.22 festgestellt, ist $\mathcal{F}$ eine Algebra. Setze

$$\mu(]s,t\rangle) := F(t) - F(s) \quad \text{für} \quad -\infty \leq s \leq t \leq \infty$$

mit

$$F(-\infty) := \lim_{s \downarrow -\infty} F(s) \quad \text{und} \quad F(\infty) := \lim_{t \uparrow -\infty} F(t)\,.$$

Wir können μ auf natürliche Weise auf endliche disjunkte Vereinigungen von Intervallen $]a,b\rangle$ fortsetzen, und damit zu einer endlich additiven Mengenabbildung, d.h. einem Inhalt, auf $\mathcal{F}$. Offensichtlich ist μ σ-endlich, denn für $\Omega_n :=]-n,n] \uparrow \mathbb{R}$ ist $\mu(]-n,n]) = F(n) - F(-n) < \infty$. Um μ zu einem Maß auf die ganze Borelsche σ-Algebra fortsetzen zu können, reicht es nun zu zeigen, dass μ σ-additiv ist. Dazu reicht es nach Lemma 1.24 aus zu zeigen, dass für jede absteigende Folge $(A_n)_n$ von Mengen aus $\mathcal{F}$ mit $A_n \downarrow \emptyset$ folgt $\lim_{n\to\infty} \mu(A_n) = 0$. Durch Einschränkung auf $\Omega_m =]-m,m]$ mit anschließendem Grenzwertübergang $m \uparrow \infty$ können wir annehmen, dass $\mu(A_1) < \infty$. Wir nehmen nun umgekehrt an, (A_n) sei eine absteigende Folge von Mengen in $\mathcal{F}$ mit $m := \lim_{n\to\infty} \mu(A_n) > 0$ (und $\mu(A_1) < \infty$) und haben zu zeigen, dass $\bigcap_n A_n \neq \emptyset$. Wir konstruieren uns dazu iterativ eine absteigende Folge von kompakten Mengen K_n und Mengen $B_n \in \mathcal{F}$ mit $B_n \subset K_n \subset A_n$ und $\mu(A_n \setminus B_n) \leq m2^{-(n+1)}$. In der Tat, da jedes A_n eine endliche disjunkte Vereingung von Teilintervallen der Form $]a,b\rangle$ ist, finden wir hierzu eine endliche Vereinigung $K_n = \bigcup_{i=1}^m [a_i, b_i]$ von abgeschlossenen disjunkten Teilintervallen mit $\mu(B_n) \geq \mu(A_n) - m2^{-(n+1)}$ für $B_n := \bigcup_{i=1}^m]a_i, b_i]$.
Es gilt nun bei dieser Wahl wegen

$$A_n \setminus (B_1 \cap \ldots \cap B_n) \subset \bigcup_{i=1}^n A_i \setminus B_i\,,$$

dass

$$\begin{aligned}\mu(B_1 \cap \ldots \cap B_n) &\geq \mu(A_n) - \mu\left(\bigcup_{i=1}^n A_i \setminus B_i\right) \geq \mu(A_n) - \sum_{i=1}^n \mu(A_i \setminus B_i) \\ &\geq \mu(A_n) - \sum_{i=1}^n m2^{-(i+1)} \geq \frac{m}{2}\,.\end{aligned}$$

Daher ist $B_1 \cap \ldots \cap B_n \neq \emptyset$ für alle n und damit auch $\bar{K}_n := K_1 \cap \ldots \cap K_n \neq \emptyset$ für alle n. Damit ist $\bar{K}_n$ eine absteigende Folge **nichtleerer, kompakter** Mengen, und damit ist ihr Schnitt $\bigcap_{n=1}^\infty \bar{K}_n \neq \emptyset$ woraus die Behauptung folgt, dass auch $\bigcap_{n\geq 1}^\infty A_n \neq \emptyset$. □

Beispiel 1.27. (i) Im Falle $F(\infty) - F(-\infty) = 1$ ist das im letzten Satz konstruierte Maß μ ein Wahrscheinlichkeitsmaß und wir können ohne Beschränkung der Allgemeinheit $F(-\infty) = 0$ annehmen. Damit ist dann F die zugehörige Verteilungsfunktion, insbesondere gilt also $F(t) = \mu(]-\infty, t])$ für $t \in \mathbb{R}$.

(ii) Für $F(t) = t$ erhalten wir aus dem letzten Satz das **Lebesguemaß** λ (mit $\lambda(]s,t]) = t - s$ für alle $s \leq t$).

(iii) Ist $f : \mathbb{R} \to [0, \infty[$ eine (Riemann-) integrierbare Dichte (nicht notwendigerweise normiert), so erfüllt $F(t) := \int_0^t f(x)\,dx$, $t \in \mathbb{R}$, die Voraussetzungen des letzten Satzes und damit gibt es (genau) ein Maß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit

$$\mu(]s,t]) = \int_s^t f(x)\,dx \quad \text{für} \quad -\infty \leq s \leq t \leq +\infty\,.$$

1.3 Transformation von Wahrscheinlichkeitsräumen

Im ganzen Abschnitt seien $(\Omega, \mathcal{A})$ und $(\tilde{\Omega}, \tilde{\mathcal{A}})$ messbare Räume.

Definition 1.28. Eine Abbildung $T : \Omega \to \tilde{\Omega}$ heißt $\mathcal{A}/\tilde{\mathcal{A}}$-**messbar** (oder einfach **messbar**), falls $T^{-1}(\tilde{A}) \in \mathcal{A}$ für alle $\tilde{A} \in \tilde{\mathcal{A}}$.
Bezeichnung:

$$\{T \in \tilde{A}\} := T^{-1}(\tilde{A}) = \big\{\omega \in \Omega \;\big|\; T(\omega) \in \tilde{A}\big\}.$$

Bemerkung 1.29. *(i) Im Falle $\mathcal{A} := \mathcal{P}(\Omega)$ ist offensichtlich* **jede** *Abbildung $T : \Omega \to \tilde{\Omega}$ messbar.*

(ii) Hinreichende Bedingung für Messbarkeit *Es sei $\tilde{\mathcal{A}}_0 \subset \mathcal{P}(\tilde{\Omega})$ Erzeugendensystem für $\tilde{\mathcal{A}}$, d.h. $\tilde{\mathcal{A}} := \sigma(\tilde{\mathcal{A}}_0)$. Dann ist T $\mathcal{A}/\tilde{\mathcal{A}}$-messbar genau dann wenn $T^{-1}(\tilde{A}) \in \mathcal{A}$ für alle $\tilde{A} \in \tilde{\mathcal{A}}_0$.*

(iii) Insbesondere: Sind $\Omega, \tilde{\Omega}$ topologische Räume und $\mathcal{A}, \tilde{\mathcal{A}}$ die zugehörigen Borel σ-Algebren. Dann gilt:

$$T : \Omega \to \tilde{\Omega} \text{ ist stetig} \quad \Rightarrow \quad T \text{ ist } \mathcal{A}/\tilde{\mathcal{A}}\text{-messbar.}$$

(iv) Es seien $(\Omega_i, \mathcal{A}_i)$, $i = 1,2,3$, messbare Räume, und $T_i : \mathcal{A}_i \to \mathcal{A}_{i+1}$, $i = 1,2$, messbare Abbildungen. Dann gilt:

$$T_2 \circ T_1 \text{ ist } \mathcal{A}_1/\mathcal{A}_3\text{-messbar.}$$

Beweis. (ii) $\{\tilde{A} \in \mathcal{P}(\tilde{\Omega}) \,|\, T^{-1}(\tilde{A}) \in \mathcal{A}\}$ ist eine σ-Algebra die $\mathcal{A}_0$ enthält. Folglich gilt

$$\sigma(\tilde{\mathcal{A}}_0) \subset \big\{\tilde{A} \in \mathcal{P}(\tilde{\Omega}) \;\big|\; T^{-1}(\tilde{A}) \in \mathcal{A}\big\}.$$

(iii) Unmittelbare Folgerung aus (ii).

(iv) Übung. □

Definition 1.30. Es sei $T : \bar{\Omega} \to \Omega$ eine Abbildung und $\mathcal{A}$ eine σ-Algebra auf Ω. Dann ist auch

$$\sigma(T) := \big\{T^{-1}(A) \;\big|\; A \in \mathcal{A}\big\}$$

eine σ-Algebra (auf $\bar{\Omega}$); $\sigma(T)$ heißt **die von** T **erzeugte** σ-**Algebra**. Weiterhin gilt: $\sigma(T)$ ist **die kleinste** σ-Algebra $\bar{\mathcal{A}}$ auf $\bar{\Omega}$, für die T $\bar{\mathcal{A}}/\mathcal{A}$-messbar ist.

Satz 1.31. *Es sei* $T : \Omega \to \tilde{\Omega}$ $\mathcal{A}/\tilde{\mathcal{A}}$*-messbar und* P *ein Wahrscheinlichkeitsmaß auf* $(\Omega, \mathcal{A})$*. Dann gilt:*

$$\tilde{P}(\tilde{A}) := T(P)(\tilde{A}) := P\big(T^{-1}(\tilde{A})\big) =: P(T \in \tilde{A}), \quad \tilde{A} \in \tilde{\mathcal{A}},$$

definiert ein Wahrscheinlichkeitsmaß auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$*.* $\tilde{P}$ *heißt das von* T **auf** $(\tilde{\Omega}, \tilde{\mathcal{A}})$ **induzierte Bildmaß** *oder einfach nur* **die Verteilung von** T **(unter** P**)***.*
Bezeichnung: $\tilde{P} = P \circ T^{-1}$ *oder* $\tilde{P} = T(P)$*.*

Beweis. Offensichtlich gilt $\tilde{P}(\tilde{A}) \geqslant 0$ für alle $\tilde{A} \in \tilde{\mathcal{A}}$, $\tilde{P}(\emptyset) = 0$ und $\tilde{P}(\tilde{\Omega}) = 1$. Für paarweise disjunkte Ereignisse $\tilde{A}_i \in \tilde{\mathcal{A}}$, $i \in \mathbb{N}$, sind auch ihre Urbilder $T^{-1}(\tilde{A}_i)$ paarweise disjunkt, und damit folgt

$$\tilde{P}\Big(\dot{\bigcup_{i\in\mathbb{N}}} \tilde{A}_i\Big) = P\Bigg(\underbrace{T^{-1}\Big(\dot{\bigcup_{i\in\mathbb{N}}} \tilde{A}_i\Big)}_{\dot{\bigcup}_{i\in\mathbb{N}} T^{-1}(\tilde{A}_i)}\Bigg) \overset{\substack{P \text{ ist} \\ \sigma\text{-additiv}}}{=} \sum_{i=1}^{\infty} P\big(T^{-1}(\tilde{A}_i)\big) = \sum_{i=1}^{\infty} \tilde{P}(\tilde{A}_i). \qquad \square$$

Bemerkung 1.32. *Es sei* $T(\Omega)$ *abzählbar, also* $T(\Omega) = \{\tilde{\omega}_i \,|\, i \in \mathbb{N}\}$*. Dann gilt*

$$\tilde{P} = \sum_{i=1}^{\infty} P(T = \tilde{\omega}_i) \cdot \delta_{\tilde{\omega}_i}.$$

Beweis. Für jedes Ereignis $\tilde{A} \in \tilde{\mathcal{A}}$ gilt offensichtlich

$$\Big\{T \in \tilde{A}\Big\} = \dot{\bigcup_{\{i\in\mathbb{N} \,|\, \tilde{\omega}_i \in \tilde{A}\}}} \{T = \tilde{\omega}_i\},$$

so dass

$$\tilde{P}(\tilde{A}) = P(T \in \tilde{A}) = \sum_{i=1}^{\infty} P(T = \tilde{\omega}_i) \cdot \underbrace{1_{\tilde{A}}(\tilde{\omega}_i)}_{=\delta_{\tilde{\omega}_i}(\tilde{A})} = \Big(\sum_{i=1}^{\infty} P(T = \tilde{\omega}_i) \cdot \delta_{\tilde{\omega}_i}\Big)(\tilde{A}). \qquad \square$$

Beispiel 1.33. Unendlicher Münzwurf: **Existenz** eines Wahrscheinlichkeitsmaßes.
Es sei $\Omega := [0,1]$ und $\mathcal{A}$ die Borelsche σ-Algebra auf $[0,1]$. Weiterhin sei P die Gleichverteilung auf $[0,1]$ und

$$\tilde{\Omega} := \Big\{\tilde{\omega} = (x_n)_{n\in\mathbb{N}} \;\Big|\; x_i \in \{0,1\} \;\forall\, i \in \mathbb{N}\Big\} = \{0,1\}^{\mathbb{N}}.$$

Definiere $X_i : \tilde{\Omega} \to \{0,1\}$ durch

$$X_i\big((x_n)_{n\in\mathbb{N}}\big) := x_i, \quad i \in \mathbb{N},$$

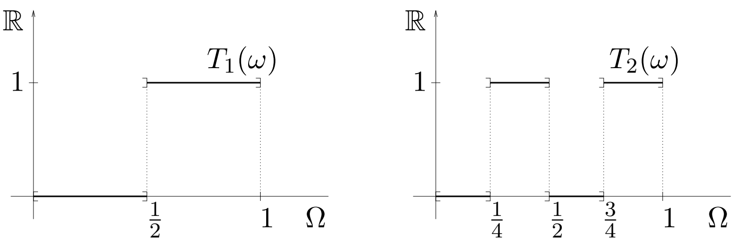
sowie

$$\tilde{\mathcal{A}} := \sigma\big(\big\{\{X_i = 1\} \;\big|\; i \in \mathbb{N}\big\}\big).$$

Man beachte, dass $\tilde{\mathcal{A}} = \sigma(\mathcal{A}_0)$, wobei $\mathcal{A}_0$ die Algebra der Zylindermengen aus Beispiel 1.7 (iii) bezeichnet. Die Binärentwicklung für $\omega \in [0,1]$ definiert eine Abbildung

$$\begin{aligned} T &: \Omega \to \tilde{\Omega} \\ &\omega \mapsto T(\omega) = (T_1(\omega), T_2(\omega), \ldots), \end{aligned}$$

mit



(und analog für T_3, T_4, ...). Man beachte, dass $T_i = X_i \circ T$ für alle $i \in \mathbb{N}$. T ist $\mathcal{A}/\tilde{\mathcal{A}}$-messbar, denn

$$T^{-1}\big(\{X_i = 1\}\big) = \{T_i = 1\} = \text{endl. Vereinigung von Intervallen} \in \mathcal{A}\,.$$

Definiere $\tilde{P} := P \circ T^{-1}$. Für fest gewählte $(x_1, \dots, x_n) \in \{0,1\}^n$ gilt nun offenbar, dass

$$\begin{aligned}\tilde{P}(X_1 = x_1, \dots, X_n = x_n) &= \tilde{P}\Big(\bigcap_{i=1}^{n} \{X_i = x_i\}\Big) \\ &= P(\text{Intervall der Länge } 2^{-n}) = 2^{-n}\,.\end{aligned}$$

Es folgt also, dass $\tilde{P}$ für festes n mit der Verteilung von n Münzwürfen übereinstimmt (letzteres entspricht der Gleichverteilung auf den $0-1$-Folgen der Länge n). Wir haben somit die Existenz eines Wahrscheinlichkeitsmaßes $\tilde{P}$ auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ gezeigt und damit das Problem aus 1.7 gelöst.

1.4 Zufallsvariablen

Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Wir erinnern zunächst an die Definition 3.1 aus der VL Wahrscheinlichkeitstheorie I einer Zufallsvariablen.

Definition 1.34. Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(E, \mathcal{E})$ ein messbarer Raum, (zB $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$). Eine Abbildung $X : \Omega \to E$ heißt **(E-wertige) Zufallsvariable** (kurz: ZV), falls

$$\{X \in B\} \in \mathcal{A} \quad \text{für alle} \quad B \in \mathcal{E}\,.$$

Der Wert $X(\omega)$ heißt **Realisierung** der Zufallsvariablen X (zum Ergebnis ω).

M.a.W. Zufallsvariablen sind also nichts anderes als auf Wahrscheinlichkeitsräumen definierte messbare Abbildungen. Als wichtigen Spezialfall wollen wir im folgenden insbesondere auch den Raum $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ der **erweiterten reellen Zahlen** zulassen mit (Borelscher-) σ-Algebra $\mathcal{B}(\bar{\mathbb{R}}) = \big\{B \subset \bar{\mathbb{R}} \;\big|\; B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\big\}$.

Bemerkung 1.35. *(i) $X : \Omega \to \bar{\mathbb{R}}$ ist eine Zufallsvariable falls für alle $c \in \mathbb{R}$ $\{X \leqslant c\} \in \mathcal{A}$.*

(ii) Im Spezialfall $\mathcal{A} = \mathcal{P}(\Omega)$ ist jede Funktion von Ω in die erweiterten reellen Zahlen $\bar{\mathbb{R}}$ eine Zufallsvariable auf $(\Omega, \mathcal{A})$.

(iii) Es sei X eine Zufallsvariable auf $(\Omega, \mathcal{A})$ und $h : \mathbb{R} \to \mathbb{R}$ (bzw. $\bar{\mathbb{R}} \to \bar{\mathbb{R}}$) eine messbare Transformation. Dann ist auch $h(X)$ wieder eine Zufallsvariable.

Beispiele: $|X|$, X^2, $|X|^p$, e^X, $\ldots$

(iv) Die Familie der Zufallsvariablen auf $(\Omega, \mathcal{A})$ ist abgeschlossen unter abzählbaren Operationen.

Insbesondere: Sind $X_1, X_2, \ldots$ Zufallsvariablen, so sind auch

- $\sum_i \alpha_i \cdot X_i$
- $\sup_i X_i$, $\inf_i X_i$
- $\limsup_{i\to\infty} X_i$, $\liminf_{i\to\infty} X_i$

wieder Zufallsvariablen.

Beweis. (i) Offensichtlich, denn $\sigma(\{]-\infty, c] : c \in \mathbb{R}\}) = \mathcal{B}(\bar{\mathbb{R}})$. Es reicht, $\{X \leq c\} \in \mathcal{A}$ für alle $c \in \mathbb{Q}$ oder jeder anderen dichten Teilmenge von $\mathbb{R}$ zu betrachten (vgl. Beispiel 1.5).

(ii) und (iii) Offensichtlich.

(iv) zum Beispiel:

- Supremum

$$\big\{(\sup_i X_i) \leqslant c\big\} = \bigcap_{i\in\mathbb{N}} \underbrace{\{X_i \leqslant c\}}_{\in\mathcal{A}} \in \mathcal{A}.$$

- Summe

$$\big\{X + Y < c\big\} = \bigcup_{\substack{r,s\in\mathbb{Q}\\ r+s<c}} \underbrace{\{X < r\} \cap \{Y < s\}}_{\in\mathcal{A}} \in \mathcal{A}$$

□

Beispiel 1.36. (i) **Indikatorfunktion** (zum Ereignis $A \in \mathcal{A}$):

$$\omega \mapsto 1_A(\omega) := \begin{cases} 1 & \text{für } \omega \in A \\ 0 & \text{für } \omega \notin A \end{cases} \qquad \text{alternative Bezeichnung } \mathbb{I}_A$$

ist eine Zufallsvariable, denn

$$\{1_A \leqslant c\} = \begin{cases} \emptyset & \text{für } c < 0 \\ A^c & \text{für } 0 \leqslant c < 1 \\ \Omega & \text{für } c \geqslant 1. \end{cases}$$

(ii) **Einfache Zufallsvariablen**

$$X = \sum_{i=1}^{n} c_i \cdot 1_{A_i}, \quad c_i \in \mathbb{R},\ A_i \in \mathcal{A},$$

Man beachte, dass jede Zufallsvariable, die nur endlich viele Werte annimmt, von diesem Typ ist, denn ist $X(\Omega) = \{c_1, \ldots, c_n\}$, so folgt

$$X = \sum_i c_i 1_{A_i} \quad \text{für} \quad X^{-1}\big(\{c_i\}\big) =: A_i.$$

Einfache Zufallsvariablen sind also in diesem Zusammenhang nichts anderes als die Elementarfunktionen in der allgemeinen Maßtheorie.

Satz 1.37 (Struktur von Zufallsvariablen auf $(\Omega, \mathcal{A})$)**.** *Es sei X eine Zufallsvariable auf $(\Omega, \mathcal{A})$. Dann gilt:*

(i) $X = X^+ - X^-$, *mit*

$$X^+ := \max(X, 0), \quad X^- := -\min(X, 0) \quad \textit{(Zufallsvariablen!)}.$$

(ii) Es sei $X \geqslant 0$. Dann gibt es eine Folge X_n, $n \in \mathbb{N}$ von einfachen Zufallsvariablen mit $X_n \leqslant X_{n+1}$ und $X = \lim_{n\to\infty} X_n$.

Beweis. (von (ii))

$$X_n := \sum_{i=0}^{n2^n-1} \frac{i}{2^n} 1_{\{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\}} + n 1_{\{X \geq n\}}$$

□

Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

Definition 1.38. Es sei X eine Zufallsvariable auf $(\Omega, \mathcal{A})$ mit

$$\min\Big(\int X^+ \,\mathrm{d}P, \int X^- \,\mathrm{d}P\Big) < \infty. \tag{1.3}$$

Dann ist

$$E(X) := \int X \,\mathrm{d}P \quad \Big(= \int_\Omega X \,\mathrm{d}P\Big)$$

wohldefiniert und heißt **Erwartungswert** von X (bzgl. P).

Definition/Konstruktion des Integrals bzgl. P

Es sei X eine Zufallsvariable.

1. Ist $X = 1_A$, $A \in \mathcal{A}$, so setzt man

$$\int X \,\mathrm{d}P := P(A)\,.$$

2. Ist $X = \sum_{i=1}^n c_i \cdot 1_{A_i}$, $c_i \in \mathbb{R}$, $A_i \in \mathcal{A}$, einfach, so setzt man

$$\int X \,\mathrm{d}P := \sum_{i=1}^n c_i \cdot P(A_i)$$

(unabhängig von der speziellen Darstellung von X!)

3. Ist $X \geq 0$, so gibt es nach Satz 1.37 eine aufsteigende Folge X_n von einfachen nichtnegativen Zufallsvariablen mit $X_n \uparrow X$. Dann setzt man

$$\int X \, \mathrm{d}P := \lim_{n\to\infty} \int X_n \, \mathrm{d}P \quad \big(\in [0,\infty]\big).$$

(unabhängig von der speziellen Wahl der approximierenden Folge X_n!)

4. Für allgemeine X zerlege $X = X^+ - X^-$ und definiere

$$\big(E(X) =\big) \quad \int X \, \mathrm{d}P := \int X^+ \, \mathrm{d}P - \int X^- \, \mathrm{d}P.$$

(wohldefiniert, falls (1.3) gilt.)

Definition 1.39. Der **(Vektor-) Raum aller P-integrierbaren Zufallsvariablen** ist definiert durch

$$\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{A}, P) := \big\{X \text{ ZV } \bigm| E\big(|X|\big) < \infty\big\}.$$

Wir definieren: Eine Eigenschaft E von Punkten $\omega \in \Omega$ **gilt P-f.s.**, falls eine ($\mathcal{A}$-) messbare Nullmenge N existiert, d.h. eine Menge $N \in \mathcal{A}$ mit $P(N) = 0$, sodass jedes $\omega \in \Omega \setminus N$ die Eigenschaft E besitzt.

Im folgenden sei

$$\mathcal{N} := \{X \text{ Zufallsvariable } \mid X = 0 \ P\text{-f.s.}\}$$

dann ist der Quotientenraum

$$L^1 := {}^{\mathcal{L}^1}\big/_{\mathcal{N}}$$

ein Banachraum bzgl. der Norm $E\big(|X|\big)$.

Bemerkung 1.40. *Spezialfall: Es sei X eine Zufallsvariable, $X \geq 0$, $X(\Omega)$ abzählbar. Dann gilt:*

$$E(X) = E\Big(\sum_{x\in X(\Omega)} x \cdot 1_{\{X=x\}}\Big) \overset{\substack{\text{“3.” und} \\ \text{“2.” oben}}}{=} \sum_{x\in X(\Omega)} x \cdot P(X = x)\,. \tag{1.4}$$

Für allgemeines X nicht notwendigerweise endlich, jedoch mit wohldefiniertem Erwartungswert $E(X)$:

$$E(X) = \sum_{\substack{x\in X(\Omega),\\ x\geqslant 0}} x \cdot P(X = x) - \sum_{\substack{x\in X(\Omega),\\ x<0}} (-x) \cdot P(X = x)\,.$$

Falls zusätzlich Ω abzählbar und $X \geqslant 0$ so gilt

$$X = \sum_{\omega\in\Omega} X(\omega) \cdot 1_{\{\omega\}}\,, \qquad \textit{und}$$

$$E(X) = \sum_{\omega\in\Omega} X(\omega) \cdot E(1_{\{\omega\}}) = \sum_{\omega\in\Omega} X(\omega) \cdot \underbrace{P(\{\omega\})}_{=:p(\omega)} = \sum_{\omega\in\Omega} p(\omega) \cdot X(\omega)\,.$$

Beispiel 1.41. Unendliche Münzwürfe Es sei $\Omega = \{0,1\}^{\mathbb{N}}$. $\mathcal{A}$ und P wie in 1.33

(i) **Erwartung des i-ten Münzwurfs** $X_i\big((x_n)_{n\in\mathbb{N}}\big) := x_i$

$$E(X_i) \overset{(1.4)}{=} 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = \frac{1}{2}.$$

(ii) **Erwartungswert der Anzahl der** `Erfolge`

$S_n := X_1 + \cdots + X_n =$ Anzahl der `Erfolge` (= Einsen) in n Münzwürfen

Dann gilt:

$$P(S_n = k) = \sum_{\substack{(x_1,\dots,x_n)\in\{0,1\}^n \\ \text{mit} \\ x_1+\cdots+x_n=k}} P(X_1 = x_1, \dots, X_n = x_n) = \binom{n}{k} \cdot 2^{-n}\,, k = 0,1,\dots,n$$

Es folgt

$$E(S_n) \overset{(1.4)}{=} \sum_{k=0}^{n} k \cdot P(S_n = k) = \sum_{k=1}^{n} k \cdot \binom{n}{k} \cdot 2^{-n} = \frac{n}{2}.$$

Einfacher, unter Ausnutzung der Linearität von $E(\,\cdot\,)$ (siehe folgenden Satz):

$$E(S_n) = \sum_{i=1}^{n} E(X_i) \overset{\text{(i)}}{=} \frac{n}{2}.$$

(iii) **Wartezeit auf den ersten Erfolg** Es sei

$$\begin{aligned} T(\omega) &:= \min\{n \in \mathbb{N} \,|\, X_n(\omega) = 1\} \\ &= \text{ Wartezeit auf den ersten Erfolg }. \end{aligned}$$

Dann gilt

$$P(T = k) = P(X_1 = \cdots = X_{k=1} = 0,\ X_k = 1) = 2^{-k},$$

und damit

$$E(T) \overset{(1.4)}{=} \sum_{k=1}^{\infty} k \cdot P(T = k) = \sum_{k=1}^{\infty} k \cdot 2^{-k} = \frac{1}{2} \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^{k-1} = 2.$$

(Zur Erinnerung: $\frac{\mathrm{d}}{\mathrm{d}q}\big(\frac{q}{1-q}\big) = \frac{\mathrm{d}}{\mathrm{d}q} \sum_{k=1}^{\infty} q^k = \sum_{k=1}^{\infty} k q^{k-1}$)

Bemerkung 1.42. $X = Y$ P-f.s., *d.h.* $P(X = Y) = 1$, *impliziert* $E(X) = E(Y)$.

Satz 1.43. $X \mapsto E(X)$ *ist ein* ***positives lineares Funktional*** *auf* $\mathcal{L}^1$, *d.h.*

(i) $X \geqslant 0$ *P-f.s. impliziert* $E(X) \geqslant 0$.

(ii) $E\Big(\sum_{i=1}^{n} c_i \cdot X_i\Big) = \sum_{i=1}^{n} c_i \cdot E(X_i)$.

((i) und (ii) implizieren Monotonie von $E(\,\cdot\,)$: $X \leqslant Y \;\Rightarrow\; E(X) \leqslant E(Y)$.)

Beweis. Siehe Lehrbücher zur Maß- und Integrationstheorie (zB [Ba91]). □

Zusätzlich ist $X \mapsto E(X)$ stetig bzgl. monoton wachsender Folgen von Zufallsvariablen aufgrund des folgenden Satzes:

Satz 1.44 (Monotone Integration, B. Levi). *Es sei X_n Zufallsvariablen mit $0 \leqslant X_1 \leqslant X_2 \leqslant \dots$. Dann gilt:*

$$\lim_{n\to\infty} E(X_n) = E\big(\lim_{n\to\infty} X_n\big)\,.$$

Beweis. Siehe Lehrbücher zur Maß- und Integrationstheorie (zB [Ba91]). □

Korollar 1.45. *Es seien $X_n \geqslant 0$, $n \in \mathbb{N}$, Zufallsvariablen. Dann gilt*

$$E\Big(\sum_{n=1}^{\infty} X_n\Big) = \sum_{n=1}^{\infty} E(X_n)\,.$$

Lemma 1.46 (Lemma von Fatou). *Es sei $X_n \geq 0$, $n \in \mathbb{N}$, eine Folge von Zufallsvariablen (bzw. allgemeiner $X_n \geqslant Y \in \mathcal{L}^1$). Dann gilt:*

$$E\big(\liminf_{n\to\infty} X_n\big) \leqslant \liminf_{n\to\infty} E(X_n)\,.$$

Beweis.

$$\begin{aligned} E\big(\liminf_{n\to\infty} X_n\big) &= E\Big(\lim_{n\to\infty}\big(\inf_{k\geqslant n} X_k\big)\Big) \overset{\text{Levi}}{=} \lim_{n\to\infty} E\big(\inf_{k\geqslant n} X_k\big) \\ &\leqslant \lim_{n\to\infty} \inf_{k\geqslant n} E(X_k) = \liminf_{n\to\infty} E(X_n)\,. \end{aligned}$$

□

Satz 1.47 (Lebesguescher Konvergenzsatz). *Es sei X_n, $n \in \mathbb{N}$, eine Folge von Zufallsvariablen und $Y \in \mathcal{L}^1$ mit $|X_n| \leqslant Y$ P-f.s. Angenommen, der punktweise Grenzwert $X := \lim_{n\to\infty} X_n$ existiert P-f.s., so gilt*

$$E\big(X\big) = E\big(\lim_{n\to\infty} X_n\big) = \lim_{n\to\infty} E(X_n)\,.$$

Beweis.

$$\begin{aligned} E\big(\underbrace{Y - \limsup_{n\to\infty} X_n}_{\geqslant 0}\big) &= E\big(\liminf_{n\to\infty} \underbrace{(Y - X_n)}_{\geqslant 0}\big) \overset{\text{Fatou}}{\leqslant} \liminf_{n\to\infty} E(Y - X_n) \\ &= E(Y) - \limsup_{n\to\infty} E(X_n)\,. \end{aligned}$$

Aus $\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n = \lim_{n\to\infty} X_n$ P-f.s. folgt nun

$$\begin{aligned} E\big(X\big) = E\big(\lim_{n\to\infty} X_n\big) = E\big(\liminf_{n\to\infty} X_n\big) &\overset{\text{Fatou}}{\leqslant} \liminf_{n\to\infty} E(X_n) \leqslant \limsup_{n\to\infty} E(X_n) \\ &\leqslant E\big(\limsup_{n\to\infty} X_n\big) = E\big(\lim_{n\to\infty} X_n\big) = E\big(X\big) \end{aligned}$$

□

Beispiel 1.48. Fairer Münzwurf (Petersburg Paradox)
Man betrachte folgendes Spiel: Eine faire Münze werde geworfen und der Spieler kann einen beliebigen Betrag (in Euro) auf `Kopf` oder `Zahl` setzen. Zeigt die richtige Seite nach oben, so erhält der Spieler das Doppelte seines Einsatzes zurück, andernfalls verfällt sein Einsatz.

Wir betrachten nun folgende Strategie: der Spieler verdoppelt bei jedem Misserfolg seinen Einsatz, bis er zum ersten Mal gewinnt. Ist sein ursprünglicher Einsatz 1 Euro, so beträgt sein Einsatz nach der n-ten Runde

$$X_n = 2^{n-1} \cdot 1_{\{T>n-1\}},$$

wobei T = Wartezeit auf den ersten Erfolg bezeichnet. Dann gilt

$$E(X_n) = 2^{n-1} \cdot \underbrace{P(T > n-1)}_{=\left(\frac{1}{2}\right)^{n-1}} = 1\,.$$

Andererseits jedoch ist $\lim_{n\to\infty} X_n = 0$ P-f.s. (genauer: für alle $\omega \neq (0,0,0,\dots)$).

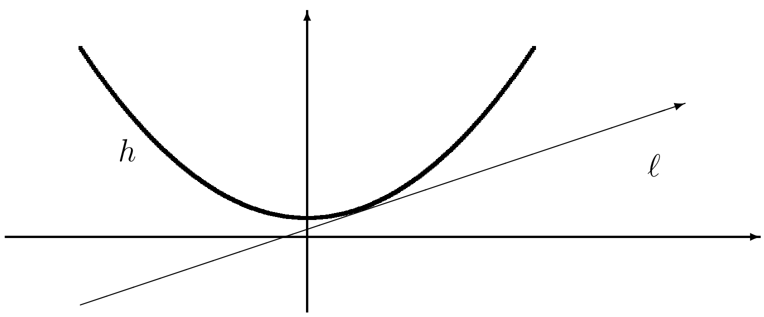
1.5 Ungleichungen

Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

Satz 1.49 (Jensensche Ungleichung). *Es sei h eine auf einem Intervall $I \subseteq \mathbb{R}$ definierte konvexe Funktion, X aus $\mathcal{L}^1$, also integrierbar, und $X(\Omega) \subset I$. Dann folgt $\mathbb{E}(X) \in I$ und*

$$h\big(E(X)\big) \leqslant E\big(h(X)\big)\,.$$

Beweis. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass X nicht P-f.s. konstant ist. Dann ist $x_0 := E(X) \in \mathring{I}$, d.h. im Inneren des Intervalls I. Da h konvex ist, gibt es eine affin lineare Funktion ℓ mit $\ell(x_0) = h(x_0)$ und $\ell \leqslant h$ (Stützgerade).



Folglich

$$h\big(\mathbb{E}[X]\big) = \ell\big(E(X)\big) \overset{\text{Linearität}}{=} E\big(\ell(X)\big) \overset{\text{Monotonie}}{\leqslant} E\big(h(X)\big)\,. \qquad \square$$

Beispiel 1.50.

$$E(X)^2 \leqslant E(X^2)\,.$$

Weiterhin gilt auch für $0 < p \leqslant q$:

$$\underbrace{E\big(|X|^p\big)^{\frac{1}{p}}}_{=:\|X\|_p} \leqslant \underbrace{E\big(|X|^q\big)^{\frac{1}{q}}}_{=:\|X\|_q}\,.$$

Beweis. $h(x) := |x|^{\frac{q}{p}}$ ist konvex. Da $\big(|X| \wedge n\big)^p \in \mathcal{L}^1$ für $n \in \mathbb{N}$, folgt

$$\Big(\mathbb{E}\Big(\big(|X| \wedge n\big)^p\Big)\Big)^{\frac{q}{p}} \leqslant \mathbb{E}\Big(\big(|X| \wedge n\big)^q\Big)\,,$$

woraus im Grenzwert $n \to \infty$ die Behauptung folgt. □

Definition 1.51. Für $1 \leqslant p < \infty$ sei

$$\mathcal{L}^p := \big\{X \;\big|\; X \text{ ZV und } \mathbb{E}\big(|X|^p\big) < \infty\big\}.$$

$\mathcal{L}^p$ heiß der **Raum der p-fach integrierbaren Zufallsvariablen**.

Bemerkung 1.52. *(i) Für $1 \leqslant p \leqslant q$ folgt $\mathcal{L}^q \subset \mathcal{L}^p$.*

(ii) Es sei $\mathcal{N} = \{X \mid X \text{ ZV und } X = 0 \ P\text{-f.s.}\}$, und $p \geqslant 1$. Dann ist $\mathcal{N} \subset \mathcal{L}^p$ ein linearer Unterraum und der Quotientenraum

$$L^p := {}^{\mathcal{L}^p}\!\big/_{\mathcal{N}}$$

ein Banachraum bzgl. der Norm $\|\cdot\|_p$ (d.h. ein vollständiger normierter Vektorraum)

Der folgende Satz wurde bereits als Satz 5.1 in der VL Wahrscheinlichkeitstheorie I bewiesen.

Satz 1.53 (Markovsche Ungleichung). *Es sei X eine Zufallsvariable, $h \geqslant 0$ monoton wachsend. Dann gilt*

$$P(X \geqslant c) \leqslant \frac{E\big(h(X)\big)}{h(c)} \qquad \text{für alle } c \in \mathbb{R} \text{ mit } h(c) > 0\,.$$

Beweis.

$$\begin{aligned} h(c)\,P(X \geqslant c) &\leqslant h(c)\,P\big(h(X) \geqslant h(c)\big) = E\big(h(c)\,1_{\{h(X)\geqslant h(c)\}}\big) \\ &\leqslant E\big(h(X)\big)\,. \end{aligned}$$

□

Beispiel 1.54. Spezialfälle

(i)

$$h(t)=\begin{cases} t^p & t>0 \\ 0 & t\leqslant 0 \end{cases} \qquad \text{für } 1\le p<+\infty$$

ist monoton wachsend, also:

$$P\left(|X-E(X)|\ge a\right)\le \frac{E\left(|X-E(X)|^p\right)}{a^p} \quad \forall a>0\,.$$

Im Spezialfall $p=2$ erhält man die **Chebychevsche Ungleichung**

$$P\left(|X-E(X)|\ge a\right)\le \frac{E\left(|X-E(X)|^2\right)}{a^2}=\frac{1}{a^2}\mathsf{V}(X) \quad \forall a>0\,.$$

(ii) $h(t)=e^{\lambda t}$, $\lambda>0$, ergibt

$$P(X-E(X)\ge a)\le e^{-\lambda a}\,E\left(e^{\lambda(X-E(X))}\right) \quad \forall a\in\mathbb{R}\,.$$

1.6 Konvergenz von Zufallsvariablen und Gleichgradige Integrierbarkeit

Definition 1.55. Es seien $X, X_1, X_2, \dots$ Zufallsvariablen $(\Omega, \mathcal{A}, P)$.

(i) $\mathcal{L}^p$-**Konvergenz** $(p\geqslant 1)$

$$\lim_{n\to\infty} E\big(|X_n-X|^p\big)=0$$

(alternativ $\lim_{n\to\infty}\|X_n-X\|_p=0$)

(ii) **Konvergenz in Wahrscheinlichkeit (Stochastische Konvergenz)**

$$\forall\,\varepsilon>0:\lim_{n\to\infty} P\big(|X_n-X|\geqslant\varepsilon\big)=0.$$

(iii) P-**f.s. Konvergenz**

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

Bemerkung. *Wir haben bereits in der Vorlesung Wahrscheinlichkeitstheorie I im Zusammenhang mit dem starken Gesetz der großen Zahlen gesehen, dass zwischen Konvergenz in Wahrscheinlichkeit und P-f.s.-Konvergenz folgender Zusammenhang besteht (siehe VL-Skript Wahrscheinlichkeitstheorie I, Abschnitt 5.3):*

$$X_n \overset{\text{P-f.s.}}{\longrightarrow} X \Leftrightarrow \quad \sup_{k\geqslant n}|X_k-X|\longrightarrow 0 \;\; \textit{in Wahrscheinlichkeit für } n\to\infty\,.$$

Satz 1.56 (Vergleich der drei Konvergenzarten).

$$\begin{tikzcd}[column sep=large, row sep=large]
\textit{(i)} \arrow[rr, Rightarrow] & & \textit{(ii)} \arrow[dl, Rightarrow, shift left, "\substack{\textit{entlang einer}\\ \textit{Teilfolge}}"] \\
& \textit{(iii)} \arrow[ul, Rightarrow, "\substack{\textit{falls } \sup_{n\in\mathbb{N}}|X_n| \in \mathcal{L}^p \\ \textit{(bzw. } |X_n|^p \textit{ gleichgr. int.)}}"] \arrow[ur, Rightarrow, shift left] &
\end{tikzcd}$$

Beweis.

(i)⇒(ii): Die Markovsche Ungleichung impliziert:

$$P\big(|X_n - X| \geqslant \varepsilon\big) \leqslant \frac{E\big(|X_n - X|^p\big)}{\varepsilon^p}.$$

(iii)⇒(ii): Es sei $\varepsilon > 0$. Für die Zufallsvariablen $Y_n := 1_{\{|X_n - X| \geq \varepsilon\}}$ gilt $\lim_{n\to\infty} Y_n = 0$ P-f.s. und $|Y_n| \leq 1$. Aus dem Lebesgueschen Konvergenzsatz 1.47 folgt nun

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

(iii)⇒(i): $Y := \sup_{n\in\mathbb{N}}|X_n| \in \mathcal{L}^p$, $\lim_{n\to\infty} X_n = X$ P-f.s. impliziert $|X| \leqslant Y$
Insbesondere, $|X_n - X|^p \leqslant 2^p Y^p \in \mathcal{L}^1$.
Aus $\lim_{n\to\infty}|X_n - X|^p = 0$ P-f.s. und dem Lebesgueschen Konvergenzsatz folgt nun

$$\lim_{n\to\infty} E\big(|X_n - X|^p\big) = 0.$$

(ii)⇒(iii): Durch Iteration finden wir für alle $l \in \mathbb{N}$ eine Teilfolge $\exists\big(n_k(l)\big)_k$ von $\big(n_k(l-1)\big)_k$ so dass

$$\sum_{k=1}^{\infty} P\left(\big|X_{n_k(l)} - X\big| \geqslant \frac{1}{l}\right) \leq 1\,.$$

Hierbei setzen wir $n_k(0) = k$. Für die Diagonalfolge $\big(n_k(k)\big)_k$ ergibt sich für alle l

$$\sum_{k=1}^{\infty} P\left(\big|X_{n_k(k)} - X\big| \geqslant \frac{1}{l}\right) \leq \sum_{k=1}^{l-1} P\left(\big|X_{n_k(k)} - X\big| \geqslant \frac{1}{l}\right) + 1 \leq l + 1 < \infty\,.$$

Das folgende Lemma 1.58 impliziert

$$\lim_{k\to\infty} X_{n_k(k)} = X \quad P\text{-f.s.}$$

□

Bemerkung 1.57. *Das Diagramm kann wie folgt vervollständigt werden:*

- *(ii)⇒ (i) gilt, falls $\sup_{n\in\mathbb{N}}|X_n| \in \mathcal{L}^p$ (bzw. $|X_n|^p$ gleichgradig integrierbar)(vgl. Satz 1.59 und Bemerkung 1.63 unten)*

- *im allgemeinen (i)$\not\Rightarrow$(iii) und (iii)$\not\Rightarrow$(i) (also auch (ii)$\not\Rightarrow$(i)). Für Beispiele siehe die Übungen.*

Es bleibt noch der folgende Zusammenhang zwischen stochastischer und fast sicherer Konvergenz nachzutragen:

Lemma 1.58. *Es seien $Z_1, Z_2, \dots$ Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Falls für alle $\varepsilon > 0$*

$$\sum_{n=1}^{\infty} P\big(|Z_n| \geqslant \varepsilon\big) < \infty$$

(= "schnelle stochastische Konvergenz gegen 0"*), so folgt*

$$P\big(\{\omega \mid \lim_{n\to\infty} Z_n(\omega) = 0\}\big) = 1.$$

(= "f.s. Konvergenz gegen 0"*).*

Beweis. Aus dem Lemma von Borel-Cantelli (Lemma 1.11) folgt für alle $\varepsilon > 0$

$$P\Big(\limsup_{n\to\infty}\{|Z_n| > \varepsilon\}\Big) = 0.$$

Insbesondere ergibt sich für alle $k \in \mathbb{N}$ die Existenz einer Nullmenge $N_k \in \mathcal{A}$ mit

$$\limsup_{n\to\infty} \big|Z_n(\omega)\big| \leqslant \frac{1}{k} \qquad \forall\, \omega \in \Omega \setminus N_k\,.$$

Es folgt für $\omega \notin N := \bigcup_{k=1}^{\infty} N_k$ (beachte dass $P(N) = 0!$)

$$\lim_{n\to\infty} \big|Z_n(\omega)\big| = 0\,.$$

□

Der nächste Satz ist die stärkste Version des Lebesgueschen Konvergenzsatzes.

Satz 1.59. *Es seien $X_n \in \mathcal{L}^1$ und X Zufallsvariablen. Dann sind äquivalent:*

(i) $\lim\limits_{n\to\infty} X_n = X$ *in* $\mathcal{L}^1$.

(ii) $\lim\limits_{n\to\infty} X_n = X$ *in Wahrscheinlichkeit und* $(X_n)_{n\in\mathbb{N}}$ *gleichgradig integrierbar.*

Korollar 1.60. $\lim_{n\to\infty} X_n = X$ *P-f.s. und* $(X_n)_{n\in\mathbb{N}}$ *gleichgradig integrierbar impliziert*

$$\lim_{n\to\infty} E(X_n) = E(X).$$

Definition 1.61. Eine Familie $(X_i)_{i\in I} \subseteq \mathcal{L}^1$ von Zufallsvariablen heißt **gleichgradig integrierbar**, falls

$$\lim_{c\to\infty} \sup_{i\in I} \underbrace{\int_{\{|X_i|\geqslant c\}} |X_i|\, dP}_{=:E(1_{\{|X_i|\geqslant c\}}\cdot|X_i|)} = 0.$$

Lemma 1.62 (ε-δ Kriterium). *Es sei $(X_i)_{i\in I} \subset \mathcal{L}^1$. Dann sind äquivalent:*

(i) $(X_i)_{i\in I}$ *ist gleichgradig integrierbar.*

(ii) $\sup_{i\in I} E\big(|X_i|\big) < \infty$ *und* $\forall\, \varepsilon > 0\ \exists \delta > 0$ *sodass*

$$P(A) < \delta \quad \Rightarrow \quad \int_A |X_i|\, dP < \varepsilon \quad \forall\, i \in I.$$

Beweis.

(i) $\Rightarrow$(ii): $\exists c$ so dass $\sup_{i\in I} \int_{\{|X_i|\geqslant c\}} |X_i|\, dP \leqslant 1$. Folglich

$$\begin{aligned}\sup_{i\in I} \int |X_i|\, dP &= \sup_{i\in I} \bigg\{ \int_{\{|X_i|<c\}} |X_i|\, dP + \int_{\{|X_i|\geqslant c\}} |X_i|\, dP \bigg\} \\ &\leqslant c + 1 < \infty.\end{aligned}$$

Es sei $\varepsilon > 0$. Dann gibt es ein $c \geqslant 0$ so dass

$$\sup_{i\in I} \int_{\{|X_i|\geqslant c\}} |X_i|\, dP < \frac{\varepsilon}{2}.$$

Für $\delta := \frac{\varepsilon}{2c}$ und $A \in \mathcal{A}$ mit $P(A) < \delta$ ergibt sich

$$\begin{aligned}\int_A |X_i|\, dP &= \int_{A\cap\{|X_i|<c\}} |X_i|\, dP + \int_{A\cap\{|X_i|\geqslant c\}} |X_i|\, dP \\ &\leqslant c \int_A dP + \int_{\{|X_i|\geqslant c\}} |X_i|\, dP < c \cdot P(A) + \frac{\varepsilon}{2} < \varepsilon.\end{aligned}$$

(ii)$\Rightarrow$(i): Es sei $\varepsilon > 0$ und δ wie in (ii). Weiter sei c so groß, dass

$$\frac{1}{c} \cdot \sup_{i\in I} E\big(|X_i|\big) < \delta\,.$$

Dann folgt mit der Markovschen Ungleichung

$$P\big(|X_i| \geqslant c\big) \leqslant \frac{1}{c} \cdot E\big(|X_i|\big) < \delta\,.$$

Insbesondere ergibt sich

$$\int_{\{|X_i|\geqslant c\}} |X_i|\, dP < \varepsilon \quad \forall\, i \in I \quad \Rightarrow \quad \sup_{i\in I} \int_{\{|X_i|\geqslant c\}} |X_i|\, dP \leq \varepsilon\,.$$

□

Bemerkung 1.63. *(i) Existenz einer Majorante impliziert gleichgradige Integrierbarkeit :* $|X_i| \leq Y \in \mathcal{L}^1\ \forall i \in I$

$$\Rightarrow \qquad \int_{\{|X_i|\geqslant c\}} |X_i|\, dP \leqslant \int_{\{Y\geqslant c\}} Y\, dP = E\big(1_{\{Y\geqslant c\}} \cdot Y\big) \xrightarrow{\text{Lebesgue} \quad c\nearrow\infty} 0,$$

denn $1_{\{Y\geqslant c\}} \cdot Y \xrightarrow{c\to\infty} 0$ *P-f.s. (Markovsche Ungleichung)*
Insbesondere: I *endlich* $\Rightarrow (X_i)_{i\in I} \subset \mathcal{L}^1$ *gleichgradig integrierbar.*

(ii) Es seien $(X_i)_{i\in I}$, $(Y_i)_{i\in I}$ gleichgradig integrierbar, $\alpha,\beta\in\mathbb{R}$

$$\Rightarrow \qquad (\alpha X_i+\beta Y_i)_{i\in I} \qquad \textit{gleichgradig integrierbar}$$

(siehe Übungen)

Beweis von Satz 1.59.

(i)⇒(ii): siehe Übungen (Hinweis: Benutze Lemma 1.62).

(ii)⇒(i): a) $X\in\mathcal{L}^1$, denn entlang einer Teilfolge (n_k) gilt $\lim_{k\to\infty} X_{n_k}=X$ P-f.s., und somit

$$E\big(|X|\big)=E\big(\liminf_{k\to\infty}|X_{n_k}|\big)\overset{\text{Fatou}}{\leqslant}\liminf_{k\to\infty}E\big(|X_{n_k}|\big)\leqslant\sup_{n\in\mathbb{N}}E\big(|X_n|\big)<\infty.$$

b) O.B.d.A. $X=0$ (denn $(X_n)_{n\in\mathbb{N}}$ gleichgradig integrierbar impliziert $(X_n-X)_{n\in\mathbb{N}}$ ebenfalls gleichgradig integrierbar nach Bemerkung 1.63)

Es sei $\varepsilon>0$. Dann gibt es ein $\delta>0$ sodass für alle $A\in\mathcal{A}$ mit $P(A)<\delta$ folgt $\int_A|X_n|\,dP<\frac{\varepsilon}{2}$.
Da $X_n\to 0$ in Wahrscheinlichkeit, existiert ein $n_0\in\mathbb{N}$, sodass $P\big(|X_n|\geqslant\frac{\varepsilon}{2}\big)<\delta$ $\forall\, n\geq n_0$. Es folgt für $n\geq n_0$

$$E\big(|X_n|\big)=\underbrace{\int_{\{|X_n|<\frac{\varepsilon}{2}\}}|X_n|\,dP}_{<\frac{\varepsilon}{2}}+\underbrace{\int_{\{|X_n|\geqslant\frac{\varepsilon}{2}\}}|X_n|\,dP}_{<\frac{\varepsilon}{2}}<\varepsilon,$$

und somit $\lim_{n\to\infty}E(|X_n|)=0$. □

Korollar 1.64. *$\lim_{n\to\infty}X_n=X$ in Wahrscheinlichkeit und $\big(|X_n|^p\big)_{n\in\mathbb{N}}$ gleichgradig integrierbar, $p>0$*

$$\Rightarrow \qquad \lim_{n\to\infty}X_n=X \quad \textit{in } \mathcal{L}^p.$$

Beweis. $\lim_{n\to\infty}|X_n-X|^p\to 0$ in Wahrscheinlichkeit und wegen

$$|X_n-X|^p\leqslant 2^p\cdot\underbrace{\big(|X_n|^p+|X|^p\big)}_{\text{glgr. integrierbar}},$$

ist auch $(|X_n-X|^p)_{n\in\mathbb{N}}$ gleichgradig integrierbar. Satz 1.59 impliziert nun

$$\lim_{n\to\infty}E\big(|X_n-X|^p\big)=0.$$ □

Satz 1.65. *(De la Vallée Poussin) Es sei $g:[0,\infty[\,\to[0,\infty[$ messbar und es gelte $\lim_{x\to\infty}\frac{g(x)}{x}=\infty$. Dann gilt:*

$$\sup_{i\in I}E\left(g\big(|X_i|\big)\right)<\infty \qquad\Rightarrow\qquad (X_i)_{i\in I}\ \textit{gleichgradig integrierbar}$$

Beweis. Es sei $\varepsilon > 0$. Wähle $c > 0$, sodass $\frac{g(x)}{x} \geqslant \frac{1}{\varepsilon} \sup_{i\in I} E\big(g\big(|X_i|\big)\big) + 1$ für $x \geqslant c$. Dann folgt

$$\begin{aligned}\int_{\{|X_i|\geqslant c\}} |X_i|\, dP &= \int_{\{|X_i|\geqslant c\}} g\big(|X_i|\big) \cdot \frac{|X_i|}{g\big(|X_i|\big)}\, \mathrm{d}P \\ &\leqslant \frac{\varepsilon}{\sup\limits_{j\in I} E\big(g\big(|X_j|\big)\big) + 1} \cdot \int_{\{|X_i\geqslant c\}} g\big(|X_i|\big)\, dP \leqslant \varepsilon \qquad \forall i \in I \qquad \square\end{aligned}$$

Beispiel 1.66. (i) $p > 1$, $\sup_i E(|X_i|^p) < \infty \Rightarrow (X_i)_{i\in I}$ gleichgradig integrierbar

(ii) (endliche Entropie Bedingung)

$$\sup_{i\in I} E\Big(|X_i|\cdot\log^+\big(|X_i|\big)\Big) < \infty \qquad \Rightarrow \qquad (X_i)_{i\in I} \text{ gleichgradig integrierbar} \quad (1.5)$$

Beispiel 1.67. Anwendung auf das starke Gesetz der großen Zahlen. Es seien Zufallsvariablen $X_1, X_2, \ldots$ in $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ mit $E(X_i) = m$ für alle $i \in \mathbb{N}$.
Angenommen es gelte

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} m \quad P-f.s. \tag{1.6}$$

Frage: Unter welchen Bedingungen gilt $\mathcal{L}^1$-Konvergenz?
Wir haben in Satz 5.8 (Starkes Gesetz der großen Zahlen, VL Wahrscheinlichkeitstheorie I) gesehen, dass (1.6) gilt, falls $(X_i)_{i\in\mathbb{N}} \subset \mathcal{L}^1$ unabhängig und identisch verteilt.
Antwort:

$$\sup_{i\in\mathbb{N}} E\Big(|X_i| \cdot \log^+\big(|X_i|\big)\Big) < \infty \quad \text{impliziert} \quad \lim_{n\to\infty} \frac{S_n}{n} = m \quad \text{in } \mathcal{L}^1.$$

Beweis: $g(x) := x \cdot \log^+(x)$ ist monoton wachsend und konvex. Folglich

$$E\left(g\left(\frac{|S_n|}{n}\right)\right) \overset{\text{Monotonie}}{\leqslant} E\left(g\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|\right)\right) \overset{\text{Konvexität}}{\leqslant} E\left(\frac{1}{n}\sum_{i=1}^{n} g\big(|X_i|\big)\right) \leqslant \text{const.} \qquad \forall n\,.$$

Folglich ist $\big(\frac{S_n}{n}\big)_{n\in\mathbb{N}}$ gleichgradig integrierbar und somit

$$\lim_{n\to\infty} \frac{S_n}{n} = m \quad \text{in } \mathcal{L}^1.$$

Eine abschließende Bemerkung zum Lebesgueschen Konvergenzsatz.

Satz 1.68. *Es seien $X_n \geq 0$, $\lim\limits_{n\to\infty} X_n = X$ P-f.s. (oder in Wahrscheinlichkeit). Dann gilt*

$$\begin{aligned} &\lim_{n\to\infty} X_n = X \quad \textit{in } \mathcal{L}^1 \\ \Leftrightarrow \quad &\lim_{n\to\infty} E(X_n) = E(X) \quad \textit{und} \quad E(X) < \infty. \end{aligned}$$

Beweis.
„$\Rightarrow$“: Offensichtlich
„$\Leftarrow$“:

$$X + X_n = \underbrace{X \vee X_n}_{:=\sup\{X,X_n\}} + \underbrace{X \wedge X_n}_{:=\inf\{X,X_n\}}$$

Somit

$$\lim_{n\to\infty} E(X \wedge X_n) \overset{\text{Lebesgue}}{=} E(X)$$

und

$$\lim_{n\to\infty} E(X \vee X_n) = E(X)\,.$$

Aus $|X_n - X| = (X \vee X_n) - (X \wedge X_n)$ folgt nunmehr

$$\lim_{n\to\infty} E\big(|X_n - X|\big) = E(X) - E(X) = 0.$$ □

$\mathcal{L}^p$-Vollständigkeit

Satz 1.69 (Riesz-Fischer). *Es sei $1 \leqslant p < \infty$ und $X_n \in \mathcal{L}^p$ mit*

$$\lim_{n,m\to\infty} \int |X_n - X_m|^p \, dP = 0.$$

Dann gibt es eine Zufallsvariable $X \in \mathcal{L}^p$ so dass

(i) $\lim_{k\to\infty} X_{n_k} = X$ *P-f.s. entlang einer Teilfolge,*

(ii) $\lim_{n\to\infty} X_n = X$ *in $\mathcal{L}^p$.*

Beweis. Siehe Literatur zur Maß- und Integrationstheorie (zB [Ba91]). □

1.7 Schwache Konvergenz von Wahrscheinlichkeitsmaßen

Es sei S ein topologischer Raum und $\mathcal{S}$ die Borelsche σ-Algebra auf S. Weiter seien μ, μ_n, $n \in \mathbb{N}$, Wahrscheinlichkeitsmaße auf $(S, \mathcal{S})$.
Was ist eine vernünftige Definition für die Konvergenz der Folge μ_n gegen μ?
Offenbar ist punktweise Konvergenz im Sinne von $\mu_n(A) \xrightarrow{n\to\infty} \mu(A)$ für alle $A \in \mathcal{S}$ zu stark für viele Anwendungen. Im Zusammenhang mit dem Zentralen Grenzwertsatz haben wir in der VL Wahrscheinlichkeitstheorie I, Definition 6.1, folgenden Konvergenzbegriff eingeführt:

Definition 1.70. Es seien μ und μ_n, $n \in \mathbb{N}$, Wahrscheinlichkeitsmaße auf $(S, \mathcal{S})$. Die Folge (μ_n) konvergiert **schwach** gegen μ, falls für alle $f \in C_b(S)$ (= Raum der beschränkten stetigen Funktionen auf S) gilt

$$\int f \, d\mu_n \xrightarrow{n\to\infty} \int f \, d\mu.$$

Beispiel 1.71. (i) $x_n \xrightarrow{n\to\infty} x$ in S impliziert $\delta_{x_n} \xrightarrow{n\to\infty} \delta_x$ schwach.

(ii) Es sei $S := \mathbb{R}^1$ und $\mu_n := N\big(0, \frac{1}{n}\big)$. Dann gilt $\mu_n \to \delta_0$ schwach, denn für alle $f \in C_b(\mathbb{R})$ gilt

$$\begin{aligned}\int f\, d\mu_n &= \int f(x) \cdot \frac{1}{\sqrt{2\pi \frac{1}{n}}} \cdot e^{-\frac{x^2}{2\cdot\frac{1}{n}}}\, dx \\ &\overset{x=\frac{y}{\sqrt{n}}}{=} \int f\Big(\frac{y}{\sqrt{n}}\Big) \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}\, dy \\ &\overset{\text{Lebesgue}}{\xrightarrow{n\to\infty}} f(0) \quad = \quad \int f\, d\delta_0.\end{aligned}$$

Satz 1.72 (Portemanteau-Theorem). *Es sei S ein metrischer Raum mit Metrik d. Dann sind folgende Aussagen äquivalent:*

(i) $\mu_n \to \mu$ *schwach*

(ii) $\int f\, d\mu_n \xrightarrow{n\to\infty} \int f\, d\mu$ *für alle f beschränkt und gleichmäßig stetig (bzgl. d)*

(iii) $\limsup_{n\to\infty} \mu_n(F) \leqslant \mu(F)$ *für alle $F \subseteq S$ abgeschlossen*

(iv) $\liminf_{n\to\infty} \mu_n(G) \geqslant \mu(G)$ *für alle $G \subseteq S$ offen*

(v) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *für alle $A \in \mathcal{S}$ mit $\mu(\bar{A} \setminus \mathring{A}) = 0$.*

Beweis. **(iii)⇔(iv):** Offensichtlich bei Übergang zum Komplement.

(i)⇒(ii): Offensichtlich.

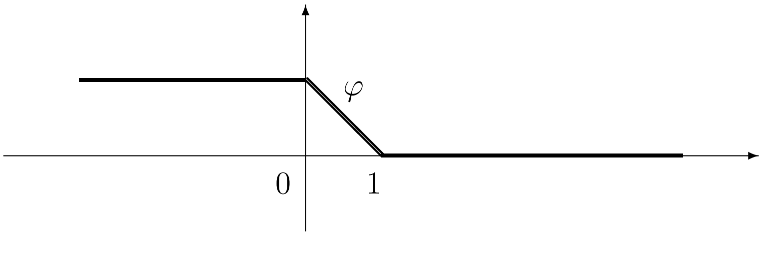
(ii)⇒(iii): Es sei $F \subset S$ abgeschlossen und

$$G_m := \left\{x \in S \;\middle|\; d(x,F) < \frac{1}{m}\right\}, \quad m \in \mathbb{N} \qquad \text{offen!}$$

Dann gilt $G_m \searrow F$, und somit $\mu(G_m) \searrow \mu(F)$.

Für $\varepsilon > 0$ existiert ein $m \in \mathbb{N}$ mit $\mu(G_m) < \mu(F) + \varepsilon$. Definiere

$$\varphi(x) := \begin{cases} 1 & \text{if } x \leqslant 0 \\ 1 - x & \text{if } x \in [0,1] \\ 0 & \text{if } x \geqslant 1. \end{cases}$$



sowie $f := \varphi\big(m \cdot d(\,\cdot\,, F)\big)$.

f ist Lipschitz, insbesondere gleichmäßig stetig. Weiter gilt $f = 0$ auf G_m^c und $f = 1$ auf F. Somit gilt

$$\begin{aligned}\limsup_{n\to\infty} \mu_n(F) &\leqslant \limsup_{n\to\infty} \int f \, d\mu_n \overset{\text{(ii)}}{=} \int f \, d\mu \\ &\leqslant \mu(G_m) < \mu(F) + \varepsilon..\end{aligned}$$

(iii)⇒(v): Es sei A mit $\mu(\bar{A} \setminus \mathring{A}) = 0$ gegeben. Dann folgt

$$\begin{aligned}\mu(A) = \mu(\mathring{A}) &\overset{\text{(iv)}}{\leqslant} \liminf_{n\to\infty} \mu_n(\mathring{A}) \leqslant \liminf_{n\to\infty} \mu_n(A) \leqslant \limsup_{n\to\infty} \mu_n(A) \\ &\leqslant \limsup_{n\to\infty} \mu_n(\bar{A}) \overset{\text{(iii)}}{\leqslant} \mu(\bar{A}) = \mu(A).\end{aligned}$$

(v)⇒(iii): Es sei $F \subseteq S$ abgeschlossen. Für alle $\delta > 0$ gilt

$$\partial\big\{d(\,\cdot\,, F) \geqslant \delta\big\} \subseteq \big\{d(\,\cdot\,, F) = \delta\big\}.$$

Beachte: Die Menge

$$D := \Big\{\delta > 0 \;\Big|\; \mu\big(\{d(\,\cdot\,, F) = \delta\}\big) > 0\Big\}$$

ist abzählbar, denn für alle $n \in \mathbb{N}$ ist die Menge

$$D_n := \left\{\delta > 0 \;\middle|\; \mu\big(\underbrace{\{d(\,\cdot\,, F) = \delta\}}_{\text{disjunkt!}}\big) > \frac{1}{n}\right\}$$

endlich. Insbesondere existiert eine Folge $\delta_k \in \,]0, \infty[\, \setminus D$, $\delta_k \downarrow 0$ sodass für die Menge

$$F_k := \big\{d(\,\cdot\,, F) \leqslant \delta_k\big\}$$

$\mu(\bar{F}_k \setminus \mathring{F}_k) = 0$ gilt. Wegen $F_k \searrow F$ folgt nun

$$\limsup_{n\to\infty} \mu_n(F) \leqslant \limsup_{n\to\infty} \mu_n(F_k) \overset{\text{(v)}}{=} \mu(F_k) \xrightarrow{k\to\infty} \mu(F).$$

(iii)⇒(i): Es sei $f \in C_b(S)$. Es reicht aus zu zeigen, dass

$$\limsup_{n\to\infty} \int f \, d\mu_n \leqslant \int f \, d\mu,$$

(denn dann folgt

$$-\liminf_{n\to\infty} \int f \, d\mu_n \leqslant \int (-f) \, d\mu,$$

und somit $\liminf_{n\to\infty} \int f\, d\mu_n \geqslant \int f\ d\mu$).

O.B.d.A. $0 \leqslant f \leqslant 1$

Zu $k \in \mathbb{N}$ fest gewählt sei $F_j := \left\{ f \geqslant \frac{j}{k} \right\}$, $j \in \mathbb{N}$ (F_j abgeschlossen!)

Dann gilt

$$\frac{1}{k} \sum_{i=1}^{k} 1_{F_i} \leqslant f \leqslant \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} 1_{F_i}$$

Somit folgt für ein beliebiges Wahrscheinlichkeitsmaß ν auf $(S, \mathcal{S})$:

$$\frac{1}{k} \sum_{i=1}^{k} \nu(F_i) \leqslant \int f\, d\nu \leqslant \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \nu(F_i).$$

sowie

$$\begin{aligned}\limsup_{n\to\infty} \int f\ d\mu_n - \frac{1}{k} &\leqslant \frac{1}{k} \cdot \limsup_{n\to\infty} \sum_{i=1}^{k} \mu_n(F_i) \\ &\leqslant \frac{1}{k} \sum_{i=1}^{k} \limsup_{n\to\infty} \mu_n(F_i) \overset{\text{(iii)}}{\leqslant} \frac{1}{k} \sum_{i=1}^{k} \mu(F_i) \leqslant \int f\, d\mu\end{aligned}$$

□

Korollar 1.73. *Es seien X, X_n, $n \in \mathbb{N}$, messbare Abbildungen von $(\Omega, \mathcal{A}, P)$ nach $(S, \mathcal{S})$ mit Verteilungen μ, μ_n, $n \in \mathbb{N}$. Dann gilt:*

$$X_n \xrightarrow{n\to\infty} X \quad \textit{in Wahrscheinlichkeit} \quad \Rightarrow \quad \mu_n \xrightarrow{n\to\infty} \mu \quad \textit{schwach}$$

Hierbei sagen wir $\lim_{n\to\infty} X_n = X$ in Wahrscheinlichkeit, falls $\lim_{n\to\infty} P(d(X, X_n) > \delta) = 0$ für alle $\delta > 0$.

Beweis. Es sei $f \in C_b(S)$ gleichmäßig stetig und $\varepsilon > 0$. Dann gibt es ein $\delta = \delta(\varepsilon) > 0$ sodass gilt: $x, y \in S$ mit $d(x, y) \leqslant \delta$ impliziert $|f(x) - f(y)| < \varepsilon$.
Somit folgt

$$\begin{aligned}&\left| \int f\ d\mu - \int f\ d\mu_n \right| = \big| E\big(f(X)\big) - E\big(f(X_n)\big) \big| \\ &\leqslant \int_{\{d(X,X_n)\leqslant\delta\}} \big|f(X) - f(X_n)\big|\, dP + \int_{\{d(X,X_n)>\delta\}} \big|f(X) - f(X_n)\big|\, dP \\ &\leqslant \varepsilon + 2\|f\|_\infty \cdot \underbrace{P\big(d(X_n, X) > \delta\big)}_{\xrightarrow{n\to\infty} 0}.\end{aligned}$$

□

Korollar 1.74. *Es sei $S = \mathbb{R}^1$ und μ, μ_n, $n \in \mathbb{N}$, seien Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit zugehörigen Verteilungsfunktionen F, F_n. Dann sind folgende Aussagen äquivalent:*

(i) $\mu_n \xrightarrow{n\to\infty} \mu$ *vage, d.h.* $\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu$ *für alle* $f \in C_0(\mathbb{R}^1)$ *(= Raum der stetigen Funktionen mit kompakten Träger)*

(ii) $\mu_n \xrightarrow{n\to\infty} \mu$ *schwach*

(iii) $F_n(x) \xrightarrow{n\to\infty} F(x)$ *für alle Stetigkeitspunkte* x *von* F

(iv) $\mu_n\big(]a,b]\big) \xrightarrow{n\to\infty} \mu\big(]a,b]\big)$ *für alle* $]a,b]$ *mit* $\mu(\{a\}) = \mu(\{b\}) = 0.$

Beweis. **(i)⇒(ii):** Übung.

(ii)⇒(iii): Es sei F stetig in x. Dann gilt $\mu(\{x\}) = 0$, und somit nach dem Portmanteau Theorem:

$$F_n(x) = \mu_n\big(]-\infty, x]\big) \xrightarrow{n\to\infty} \mu\big(]-\infty, x]\big) = F(x).$$

(iii)⇒(iv): Es sei $]a,b]$ mit $\mu(\{a\}) = \mu(\{b\}) = 0$ gegeben. Dann ist F stetig in a und b und folglich

$$\begin{aligned} \mu\big(]a,b]\big) &= F(b) - F(a) \overset{\text{(iii)}}{=} \lim_{n\to\infty} F_n(b) - \lim_{n\to\infty} F_n(a) \\ &= \lim_{n\to\infty} \mu_n\big(]a,b]\big). \end{aligned}$$

(iv)⇒(i): Es sei $D := \big\{x \in \mathbb{R} \;\big|\; \mu(\{x\}) = 0\big\}$. Dann ist $\mathbb{R} \setminus D$ abzählbar, also $D \subseteq \mathbb{R}$ dicht. Es sei $f \in C_0(\mathbb{R})$, also f gleichmäßig stetig. Damit finden wir zu $\varepsilon > 0$ reelle Zahlen $c_0 < \cdots < c_m \in D$ mit

$$\Big\| f - \underbrace{\sum_{k=1}^m f(c_{k-1}) \cdot 1_{]c_{k-1},c_k]}}_{=:g} \Big\|_\infty \leqslant \sup_k \sup_{x\in[c_{k-1},c_k]} \big| f(x) - f(c_{k-1}) \big| < \varepsilon.$$

Es folgt

$$\begin{aligned} &\left| \int f\, d\mu - \int f\, d\mu_n \right| \\ &\quad\leqslant \underbrace{\int |f-g|\, d\mu}_{<\varepsilon} + \left| \int g\, d\mu - \int g\, d\mu_n \right| + \underbrace{\int |f-g|\, d\mu_n}_{<\varepsilon} \\ &\quad\leqslant 2\varepsilon + \sum_{k=1}^m f(c_{k-1}) \cdot \big| \mu\big(]c_{k-1}, c_k]\big) - \mu_n\big(]c_{k-1}, c_k]\big) \big| \xrightarrow[n\to\infty]{\text{(iv)}} 2\varepsilon. \end{aligned}$$

□

Zum Abschluss dieses Abschnittes betrachten wir noch folgende Charakterisierung der schwachen Konvergenz von Wahrscheinlichkeitsmaßen mithilfe ihrer charakteristischen Funktionen. Wie in Abschnitt 4.2, VL Wahrscheinlichkeitstheorie I, für (reellwertige)

Zufallsvariablen, definieren wir für ein Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ seine **charakteristische Funktion**

$$\chi_\mu(t) := \int e^{itx}\,\mu(dx) \left(:= \int \cos(tx)\,\mu(dx) + \int \sin(tx)\,\mu(dx)\right)\,, t \in \mathbb{R}\,.$$

Analog zu Bemerkung 4.11, VL Wahrscheinlichkeitstheorie I, gilt dann folgender

Satz 1.75. *Es seien μ, μ_n, $n \in \mathbb{N}$, Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann sind folgende Aussagen äquivalent:*

(i) $\mu_n \xrightarrow{n\to\infty} \mu$ *schwach,*

(ii) $\chi_{\mu_n}(t) = \chi_\mu(t)$ *für alle $t \in \mathbb{R}$.*

Beispiel 1.76. $\mu_n = N\left(0, \frac{1}{n}\right)$, also $\chi_{\mu_n}(t) = e^{-\frac{t^2}{2n}}$. Dann folgt offensichtlich $\lim_{n\to\infty} e^{-\frac{t^2}{2n}} = 1$ für alle $t \in \mathbb{R}$, also $\lim_{n\to\infty} \mu_n = \delta_0$ schwach.

2 (Ergänzungen zur) Unabhängigkeit

2.1 Unabhängige Ereignisse

Wir haben (stochastische) Unabhängigkeit von Ereignissen in Abschnitt 2.2 der VL Wahrscheinlichkeitstheorie II eingeführt. In diesem Abschnitt wollen wir allgemeiner **Unabhängigkeit von Ereignissystemen** betrachten.
Im ganzen Abschnitt bezeichne $(\Omega, \mathcal{A}, P)$ einen Wahrscheinlichkeitsraum.

Definition 2.1. (i) Eine Familie von Ereignissen $A_i \in \mathcal{A}$, $i \in I$, heißt **unabhängig** (bzgl. P), falls für jede endliche Teilmenge $J \subseteq I$ die Produktformel

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

gilt.

(ii) Eine Familie von Ereignissystemen $\mathcal{B}_i \subseteq \mathcal{A}$, $i \in I$, heißt **unabhängig**, falls für jede endliche Teilmenge $J \subseteq I$ und für alle Ereignisse $A_j \in \mathcal{B}_j$, $j \in J$, die Produktformel

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

gilt.

Satz 2.2. *Es sei $\mathcal{B}_i$, $i \in I$, eine Familie von unabhängigen $\cap$-stabilen Mengensystemen. Dann gilt:*

(i) $\sigma(\mathcal{B}_i)$, $i \in I$, sind unabhängig.

(ii) Es sei J_k, $k \in K$, eine Partition der Indexmenge I. Dann sind auch die σ-Algebren

$$\sigma\Big(\bigcup_{i\in J_k} \mathcal{B}_i\Big), \quad k \in K,$$

unabhängig.

Beweis. (i) Es sei $J \subseteq I$, J endlich, von der Form $J = \{j_1, \dots, j_n\}$. Es seien $A_{j_1} \in \sigma(\mathcal{B}_{j_1}), \dots, A_{j_n} \in \sigma(\mathcal{B}_{j_n})$. Wir haben zu zeigen, dass

$$P(A_{j_1} \cap \dots \cap A_{j_n}) = P(A_{j_1}) \cdots P(A_{j_n}). \tag{2.1}$$

Dazu nehme zuerst an, dass $A_{j_2} \in \mathcal{B}_{j_2}, \dots, A_{j_n} \in \mathcal{B}_{j_n}$, und betrachte

$$\begin{aligned}\mathcal{D}_{j_1} := \big\{A \in \sigma(\mathcal{B}_{j_1}) \;\big|\; & P(A \cap A_{j_2} \cap \dots \cap A_{j_n}) \\ & = P(A) \cdot P(A_{j_2}) \cdots P(A_{j_n})\big\}.\end{aligned}$$

Beachte, dass $\mathcal{D}_{j_1}$ ein Dynkin-System ist (!) welches $\mathcal{B}_{j_1}$ enthält. Aus Satz 1.17 folgt nun, dass

$$\sigma(\mathcal{B}_{j_1}) = \delta(\mathcal{B}_{j_1}) \subseteq \mathcal{D}_{j_1} \ ,$$

und somit $\sigma(\mathcal{B}_{j_1}) = \mathcal{D}_{j_1}$. Durch Iteration desselben Argumentes folgt in gleicher Weise $\mathcal{D}_{j_2} = \sigma(\mathcal{B}_{j_2})$, $\mathcal{D}_{j_3} = \sigma(\mathcal{B}_{j_3})$, ... und damit (2.1).

(ii) Für $k \in K$ definiere das Mengensystem

$$\mathcal{C}_k := \Big\{ \bigcap_{j \in J} A_j \;\Big|\; J \subseteq J_k,\ J \text{ endlich},\ A_j \in \mathcal{B}_j \Big\} .$$

Dann ist $\mathcal{C}_k$ $\cap$-stabil und die Familie der Ereignissysteme $\mathcal{C}_k$, $k \in K$, unabhängig, denn für $k_1, \dots, k_n \in K$ und jede endliche Teilmenge $J^1 \subseteq J_{k_1}, \dots, J^n \subseteq J_{k_n}$, folgt

$$P\Big(\underbrace{\Big(\bigcap_{i \in J^1} A_i\Big)}_{\in \mathcal{C}_{k_1}} \cap \dots \cap \underbrace{\Big(\bigcap_{i \in J^n} A_i\Big)}_{\in \mathcal{C}_{k_n}} \Big) \overset{\substack{\mathcal{B}_i, i \in I \\ \text{unabh.}}}{=} \prod_{j=1}^{n} P\Big(\bigcap_{i \in J^j} A_i \Big).$$

Teil (i) impliziert nun, dass auch die Ereignissysteme

$$\sigma(\mathcal{C}_k) = \sigma\Big(\bigcup_{i \in J_k} \mathcal{B}_i \Big), \quad k \in K,$$

unabhängig sind. □

Beispiel 2.3. Es seien $A_i \in \mathcal{A}$, $i \in I$, unabhängig. Dann sind auch A_i, A_i^c, $i \in I$, unabhängig.

Bemerkung 2.4. *Wir erinnern an dieser Stelle, dass paarweise Unabhängigkeit von Ereignissen im Allgemeinen nicht die (vollständige) Unabhängigkeit von Ereignissen impliziert (vgl. Teil (ii) der Bemerkung 2.10 aus der Wahrscheinlichkeitstheorie I): Betrachte als Beispiel das zweimalige Werfen einer fairen Münze. Dann sind die Ereignisse*

$$A = \texttt{1.Wurf Zahl}$$

$$B = \texttt{2.Wurf Zahl}$$

$$C = \texttt{1.und 2.Wurf gleich}.$$

paarweise unabhängig aber nicht unabhängig, denn $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$, *aber*

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

Satz 2.5 (Kolmogorovsches Null-Eins-Gesetz). *Es seien* $\mathcal{B}_n$, $n \in \mathbb{N}$, *unabhängige* σ*-Algebren, und*

$$\mathcal{B}_\infty := \bigcap_{n=1}^{\infty} \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{B}_m \Big)$$

*die **terminale** σ-**Algebra** (bzw. das **tail-field**). Dann folgt*

$$P(A) \in \{0,1\} \qquad \forall\, A \in \mathcal{B}_\infty$$

d.h., P ist deterministisch auf $\mathcal{B}_\infty$.

Illustration: Unabhängige 0-1-Experimente
Es sei $\Omega := \{0,1\}^{\mathbb{N}}$, $X_i(\omega) = x_i$ für $\omega = (x_i)_{i\in\mathbb{N}}$ der Ausgang des i-ten Experiments.
Für $p \in [0,1]$ sei P_p das (eindeutig bestimmte) Wahrscheinlichkeitsmaß auf
$\mathcal{A} := \sigma\left(\{X_i = 1\} : i = 1,2,3,\ldots\right) = \sigma\left(\mathcal{A}_0\right)$ mit

(i) $P_p\left(X_i = 1\right) = p$

(ii) $\{X_i = 1\}$, $i \in \mathbb{N}$, unabhängig

Es ist klar, dass aus (i) und (ii) folgt, dass

$$P_p\left(X_1 = x_1, \ldots, X_n = x_n\right) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

und damit stimmt das Wahrscheinlichkeitsmaß P_p mit dem in Beispiel 1.19 betrachteten Wahrscheinlichkeitsmaß überein.
Es sei nun $\mathcal{B}_i = \sigma\big(\{X_i = 1\}\big)$. Dann enthält die zugehörige terminale σ-Algebra

$$\mathcal{B}_\infty = \bigcap_{n\in\mathbb{N}} \sigma\Big(\bigcup_{m\geqslant n} \mathcal{B}_m\Big)$$

Ereignisse, die in der unendlich fernen Zukunft liegen, wie zum Beispiel:

$$\limsup_{i\to\infty}\{X_i = 1\} = \texttt{ unendlich viele Einsen}$$

$$\left\{\omega \in \{0,1\}^{\mathbb{N}} \;\middle|\; \lim_{n\to\infty} \underbrace{\frac{1}{n}\sum_{i=1}^n X_i(\omega)}_{=:\frac{S_n(\omega)}{n}} \quad \texttt{existiert}\right\}$$

Beweis des Null-Eins-Gesetzes. Satz 2.2 impliziert für alle n, dass die σ-Algebren

$$\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{n-1}, \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{B}_m\Big)$$

unabhängig sind. Wegen $\mathcal{B}_\infty \subseteq \sigma\Big(\bigcup_{m\geqslant n} \mathcal{B}_m\Big)$ folgt hieraus auch, dass für alle n die σ-Algebren

$$\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{n-1}, \mathcal{B}_\infty$$

unabhängig sind. Nach Definition ergibt sich hieraus, dass die σ-Algebren

$$\mathcal{B}_\infty\,, \mathcal{B}_n\,, n \in \mathbb{N}$$

unabhängig sind, und damit aufgrund von Satz 2.2 (ii) dann auch, dass

$$\sigma\Big(\bigcup_{n\in\mathbb{N}}\mathcal{B}_n\Big) \quad \text{und} \quad \mathcal{B}_\infty$$

unabhängig sind. Wegen $\mathcal{B}_\infty \subseteq \sigma\Big(\bigcup_{n\geqslant 1}\mathcal{B}_n\Big)$ sind damit schließlich die σ-Algebren $\mathcal{B}_\infty$ und $\mathcal{B}_\infty$ unabhängig (!). Damit folgt das Null-Eins-Gesetz aus dem folgenden Lemma. □

Lemma 2.6. *Es sei $\mathcal{B} \subseteq \mathcal{A}$ eine σ-Algebra für die $\mathcal{B}$ unabhängig ist von sich selber. Dann gilt*

$$P(A) \in \{0,1\} \qquad \forall\, A \in \mathcal{B}\,.$$

Beweis. Für alle $A \in \mathcal{B}$

$$P(A) = P(A \cap A) = P(A)\cdot P(A) = P(A)^2.$$

Somit folgt entweder $P(A) = 0$ oder $P(A) = 1$. □

Korollar 2.7. *Es sei A_n, $n \in \mathbb{N}$, eine Folge unabhängiger Ereignisse. Dann folgt*

$$P\left(\limsup_{n\to\infty} A_n\right) = P\left(\bigcap_{n\in\mathbb{N}}\bigcup_{m\geqslant n} A_m\right) \in \{0,1\}\,.$$

Beweis. Die σ-Algebren $\mathcal{B}_n := \sigma\{A_n\} = \{\emptyset, \Omega, A_n, A_n^c\}$, $n \in \mathbb{N}$, sind unabhängig nach Satz 2.2. Wegen $\limsup_{n\to\infty} A_n \in \mathcal{B}_\infty$ folgt die Behauptung somit aus dem Null-Eins-Gesetz von Kolmogorov. □

Lemma 2.8 (Borel-Cantelli). *(i) Es sei $A_i \in \mathcal{A}$, $i \in \mathbb{N}$. Dann gilt:*

$$\sum_{i=1}^{\infty} P(A_i) < \infty \quad \Rightarrow \quad P\big(\limsup_{i\to\infty} A_i\big) = 0.$$

(ii) Sind zusätzlich die Ereignisse $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, unabhängig, so folgt auch:

$$\sum_{i=1}^{\infty} P(A_i) = \infty \quad \Rightarrow \quad P\big(\limsup_{i\to\infty} A_i\big) = 1.$$

Beweis. (i) Siehe Lemma 1.11.

(ii) Es reicht zu zeigen, dass

$$P\Big(\bigcup_{m=n}^{\infty} A_m\Big) = 1 \quad \text{bzw.} \quad P\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) = 0 \qquad \forall\, n\,.$$

Die letzte Gleichheit folgt aus der Abschätzung

$$P\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) = \lim_{k\to\infty} \underbrace{P\Big(\bigcap_{m=n}^{n+k} A_m^c\Big)}_{=\prod_{m=n}^{n+k} P(A_m^c) \quad \text{unabh.}}$$
$$= \lim_{k\to\infty} \prod_{m=n}^{n+k} (1 - P(A_m)) \leq \lim_{k\to\infty} \exp\left(-\sum_{m=n}^{n+k} P(A_m)\right) = 0$$

wobei wir verwendet haben, dass $1-\alpha \leqslant e^{-\alpha}$ für alle $\alpha \in \mathbb{R}$.

$\square$

Beispiel 2.9. Unabhängige 0-1-Experimente mit Erfolgswahrscheinlichkeit $p \in\,]0,1[$. Es sei $(x_1, \dots, x_N) \in \{0,1\}^N$ ("Binärtext der Länge N").

$$P_p\,(\texttt{Text erscheint}) \quad ?$$

Zur Berechnung dieser Wahrscheinlichkeit teilen wir die unendliche Folge $\omega = (y_n) \in \{0,1\}^{\mathbb{N}}$ in Blöcke

$$(\underbrace{y_1, y_2, \dots}_{\substack{\text{1. Block}\\ \text{Länge} = N}} \; \underbrace{\dots\dots}_{\substack{\text{2. Block}\\ \text{Länge} = N}} \; \dots) \in \Omega := \{0,1\}^{\mathbb{N}}.$$

der Länge N auf und betrachten die Ereignisse $A_i = \texttt{Text erscheint in Block } i$. Dann sind die Ereignisse A_i, $i \in \mathbb{N}$, unabhängig (!) nach Satz 2.2 (ii) mit identischer Wahrscheinlichkeit

$$P_p(A_i) = p^K(1-p)^{N-K} =: \alpha > 0$$

wobei $K := \sum_{i=1}^N x_i$ Anzahl der Einsen. Insbesondere gilt $\sum_{i=1}^\infty P_p(A_i) = \sum_{i=1}^\infty \alpha = \infty$, und das Lemma von Borel-Cantelli impliziert nun, dass

$$P_p\left(\limsup_{i\to\infty} A_i\right) := P_p\,(\texttt{Text erscheint unendlich oft}) = 1\,.$$

Es gilt sogar noch viel mehr: Die Indikatorfunktionen $1_{A_1}, 1_{A_2}, \dots$ sind unabhängig und damit folgt nach dem starken Gesetz der großen Zahlen, dass

$$\frac{1}{n}\sum_{i=1}^n 1_{A_i} \xrightarrow{P_p\text{-f.s.}} E\,(1_{A_i}) = \alpha\,,$$

d.h. der vorgegebene Text $x_1.\dots, x_N$ erscheint fast sicher mit strikt positiver relativer Häufigkeit in einer beliebigen $0-1$-Folge.

2.2 Unabhängige Zufallsvariablen und das Kolmogorovsche Gesetz der großen Zahlen

Es sei wieder im ganzen Abschnitt $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Wir erinnern zunächst an die Definition der Unabhängigkeit von Zufallsvariablen (Definition 3.18, VL Wahrscheinlichkeitstheorie I).

Definition 2.10. Eine Familie X_i, $i \in I$, von Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ heißt **(stochastisch) unabhängig**, genau dann wenn die von $(X_i)_{i\in I}$ erzeugten σ-Algebren

$$\sigma(X_i) := X_i^{-1}\big(\mathcal{B}(\bar{\mathbb{R}})\big) \quad \Big(= \big\{\{X_i \in A\} \mid A \in \mathcal{B}(\bar{\mathbb{R}})\big\}\Big), \quad i \in I,$$

unabhängig sind, d.h. also wenn für beliebige endliche Teilmengen $J \subseteq I$ und beliebige Borelmengen $A_j \in \mathcal{B}(\bar{\mathbb{R}})$ die Produktformel

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

gilt.

Bemerkung 2.11. *Aus der Definition ergibt sich unmittelbar: Sind X_i, $i \in I$, unabhängig und $h_i : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$, $i \in I$, $\mathcal{B}(\bar{\mathbb{R}})/\mathcal{B}(\bar{\mathbb{R}})$-messbar. Dann sind auch die transformierten Zufallsvariablen $Y_i := h_i(X_i)$, $i \in I$, unabhängig, denn $\sigma(Y_i) \subset \sigma(X_i)$ für alle $i \in I$.*

Satz 2.12. *Es seien $X_1, \ldots, X_n$ unabhängige, nichtnegative Zufallsvariablen. Dann gilt:*

$$E\left(X_1 \cdot \ldots \cdot X_n\right) = E\left(X_1\right) \cdot E\left(X_2\right) \cdot \ldots \cdot E\left(X_n\right) .$$

Beweis. O.B.d.A. können wir annehmen, dass $n = 2$. Der allgemeine Fall ergibt sich dann mithilfe einer Induktion nach n, da $X_1 \cdot \ldots \cdot X_{n-1}$ messbar bzgl. der σ-Algebra $\sigma\big(X_1, \ldots, X_{n-1}\big)$ und die beiden σ-Algebren $\sigma\big(X_1, \ldots, X_{n-1}\big)$ und $\sigma(X_n)$ unabhängig aufgrund von Satz 2.2.)
Wir können uns daher auf den Fall zweier nichtnegativer Zufallsvariablen X und Y zurückziehen und haben zu zeigen, dass

$$E\left(XY\right) = E\left(X\right) \cdot E\left(Y\right) . \tag{2.2}$$

Mit maßtheoretischer Induktion können wir uns auf den Fall einfacher Zufallsvariablen zurückziehen, denn für allgemeine nichtnegative X und Y finden wir aufsteigende Folgen von einfachen Zufallsvariablen X_n (bzw. Y_n), die $\sigma(X)$-messbar (resp. $\sigma(Y)$-messbar) sind, und punktweise gegen X (bzw. Y) konvergieren. Dann folgt aus $E\left(X_nY_n\right) = E\left(X_n\right) \cdot E\left(Y_n\right)$ für alle n auch (2.2) mithilfe des Satzes von der monotonen Integration.
Sind X und Y also einfache Zufallsvariablen, also von der Form

$$X = \sum_{i=1}^{m} \alpha_i 1_{A_i} \quad \text{und} \quad Y = \sum_{j=1}^{n} \beta_j 1_{B_j},$$

mit $\alpha_i, \beta_j \geqslant 0$ und $A_i \in \sigma(X)$ bzw. $B_j \in \sigma(Y)$ so folgt

$$E(XY) = \sum_{i,j} \alpha_i\beta_j \cdot P(A_i \cap B_j) = \sum_{i,j} \alpha_i\beta_j \cdot P(A_i) \cdot P(B_j) = E(X) \cdot E(Y). \square$$

Korollar 2.13. *Es seien $X_1, \ldots, X_n$ in $\mathcal{L}^1$ unabhängige Zufallsvariablen. Dann gilt*

$$X_1 \cdot \ldots \cdot X_n \in \mathcal{L}^1 \qquad \textit{und} \qquad E(X_1 \cdot \ldots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \ldots \cdot E(X_n).$$

Beweis. Wie im Beweis von Satz 2.12 können wir uns mithilfe einer Induktion nach n auf den Fall $n = 2$ zurückziehen. Es seinen also zwei unabhängige Zufallsvariablen X und Y in $\mathcal{L}^1$ gegeben. Weiterhin sei $\varepsilon_1, \varepsilon_2 \in \{+, -\}$. Dann sind X^{ε_1} und Y^{ε_2} unabhängig aufgrund von Bemerkung 2.11 und nichtnegativ. Satz 2.12 impliziert nun, dass

$$E(X^{\varepsilon_1} \cdot Y^{\varepsilon_2}) = E(X^{\varepsilon_1}) \cdot E(Y^{\varepsilon_2}).$$

Innsbesondere liegt $X^{\varepsilon_1} \cdot Y^{\varepsilon_2}$ in $\mathcal{L}^1$, denn $E(X^{\varepsilon_1}) \cdot E(Y^{\varepsilon_2}) < \infty$. Folglich liegt auch

$$X \cdot Y = X^+ \cdot Y^+ + X^- \cdot Y^- - (X^+ \cdot Y^- + X^- \cdot Y^+) \quad \in \mathcal{L}^1,$$

und $E(XY) = E(X) \cdot E(Y)$. □

Bemerkung 2.14. *(i) Im allgemeinen ist die Umkehrung zu obigem Korrolar falsch: Betrachte dazu als Beispiel eine $N(0,1)$-Zufallsvariable X und $Y = X^2$. Dann sind X und Y nicht unabhängig, jedoch gilt*

$$E(XY) = E\left(X^3\right) = 0 = E(X) \cdot E(Y).$$

(ii)

$$X, Y \in \mathcal{L}^2 \quad \textit{unabhängig} \quad \Rightarrow \quad X, Y \quad \textit{unkorreliert}$$

denn

$$\mathit{Cov}(X,Y) = E(XY) - E(X) \cdot E(Y) = 0.$$

Das Beispiel aus Teil (i) zeigt auch, dass die Umkehrung dieser Implikation im Allgemeinen falsch ist.

Wir kommen nun zum Beweis des starken Gesetzes der großen Zahlen im allgemeinen Fall unabhängig und identisch verteilter Zufallsvariablen $(X_i)_{i\in\mathbb{N}} \in \mathcal{L}^1$ (Satz 5.8 aus der VL Wahrscheinlichkeitstheorie I).

Satz 2.15 (Kolmogorov, 1930). *Es sei $(X_n) \subseteq \mathcal{L}^1$ eine Folge unabhängig, identisch verteilter Zerfallsvariablen. Dann gilt:*

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{\textit{f.s.}} E(X_1) \quad \textit{für } n \to \infty.$$

In der VL Wahrscheinlichkeitstheorie I wurde der Satz nur im Spezialfall $E(|X_i|^4) < \infty$ bewiesen. Wir werden im folgenden sogar noch eine etwas stärkere Version des Satzes beweisen:

Satz 2.16 (Etemadi, 1981). *Es sei* $(X_n) \subseteq \mathcal{L}^1$ *eine Folge paarweise unabhängiger und identisch verteilter Zufallsvariablen. Dann gilt*

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{\text{f.s.}} E(X_1) \quad \text{für } n \to \infty\,.$$

Beweis. O.B.d.A. $X_i \geqslant 0$
(andernfalls betrachte $X_1^+, X_2^+, \dots$ (paarweise unabhängig und identisch verteilt)
und $X_1^-, X_2^-, \dots$ (paarweise unabhängig und identisch verteilt))

1. Ersetze X_i durch $\tilde{X}_i := 1_{\{X_i < i\}} X_i$.

   Offensichtlich gilt

   $$\tilde{X}_i = h_i(X_i) \quad \text{mit} \quad h_i(x) := \begin{cases} x & \text{falls } x < i \\ 0 & \text{falls } x \geqslant i \end{cases}$$

   Dann sind $\tilde{X}_1, \tilde{X}_2, \dots$ paarwise unabhängig aufgrund von Bemerkung 2.11. Für den Beweis des Satzes reicht es nun aus zu zeigen, dass für $\tilde{S}_n := \sum_{i=1}^n \tilde{X}_i$ gilt:

   $$\frac{\tilde{S}_n}{n} \xrightarrow{n\to\infty} m \quad P\text{-f.s.}$$

   wobei wir $m := E\left(X_i\right)$ gesetzt haben.

   In der Tat gilt

   $$\begin{aligned} \sum_{n=1}^{\infty} P(X_n \neq \tilde{X}_n) &= \sum_{n=1}^{\infty} P(X_n \geqslant n) = \sum_{n=1}^{\infty} P(X_1 \geqslant n) \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\big(X_1 \in [k, k+1[\big) = \sum_{k=1}^{\infty} k \cdot P\big(X_1 \in [k, k+1[\big) \\ &= \sum_{k=1}^{\infty} E\big(\underbrace{k \cdot 1_{\{X_1 \in [k,k+1[\}}}_{\leqslant X_1 \cdot 1_{\{X_1 \in [k,k+1[\}}}\big) \leqslant E(X_1) < \infty \end{aligned}$$

   und aus dem Lemma von Borel-Cantelli (Teil (i)) folgt

   $$P(X_n \neq \tilde{X}_n \text{ unendlich oft}) = 0\,.$$

2. In einem weiteren Schritt reduzieren wir den Beweis der Konvergenz auf die Konvergenz entlang der Teilfolge $k_n = \lfloor \alpha^n \rfloor$ für $\alpha > 1$

   (wobei ($\lfloor x \rfloor$ = größte natürliche Zahl $\leq x$).

   Wir zeigen im Schritt 3, dass

   $$\frac{\tilde{S}_{k_n} - E(\tilde{S}_{k_n})}{k_n} \xrightarrow{n\to\infty} 0 \qquad \text{P-f.s.} \tag{2.3}$$

Hieraus folgt dann die Behauptung des Satzes, denn

$$E(\tilde{X}_i) = E\big(1_{\{X_i<i\}} \cdot X_i\big) = E\big(1_{\{X_1<i\}} \cdot X_1\big) \overset{i\to\infty}{\nearrow} E(X_1)\,(=m)\,,$$

damit

$$\frac{1}{k_n} \cdot E(\tilde{S}_{k_n}) = \frac{1}{k_n} \sum_{i=1}^{k_n} E(\tilde{X}_i) \xrightarrow{n\to\infty} m,$$

und somit schließlich

$$\frac{1}{k_n} \cdot \tilde{S}_{k_n} \xrightarrow{n\to\infty} m \quad P\text{-f.s.}$$

Ist $l \in \mathbb{N} \cap [k_n, k_{n+1}[$, so folgt

$$\underbrace{\frac{k_n}{k_{n+1}}}_{\xrightarrow{n\to\infty} \frac{1}{\alpha}} \cdot \underbrace{\frac{\tilde{S}_{k_n}}{k_n}}_{\substack{\xrightarrow{n\to\infty} m \\ P\text{-f.s.}}} \leqslant \frac{\tilde{S}_l}{l} \leqslant \underbrace{\frac{\tilde{S}_{k_{n+1}}}{k_{n+1}}}_{\substack{\xrightarrow{n\to\infty} m \\ P\text{-f.s.}}} \cdot \underbrace{\frac{k_{n+1}}{k_n}}_{\xrightarrow{n\to\infty} \alpha} .$$

Also gibt es eine P-Nullmenge $N_\alpha \in \mathcal{A}$, sodass für alle $\omega \notin N_\alpha$

$$\frac{1}{\alpha} \cdot m \leqslant \liminf_{l\to\infty} \frac{\tilde{S}_l(\omega)}{l} \leqslant \limsup_{l\to\infty} \frac{\tilde{S}_l(\omega)}{l} \leqslant \alpha \cdot m.$$

Schließlich wähle man eine Nullfolge $\alpha_n \searrow 1$. Dann folgt für alle $\omega \notin N := \bigcup_{n\geqslant 1} N_{\alpha_n}$

$$\lim_{l\to\infty} \frac{\tilde{S}_l(\omega)}{l} = m.$$

3. Aufgrund von Lemma 1.58 reicht es für den Beweis von (2.3) aus zu zeigen, dass

$$\forall \varepsilon > 0 : \qquad \sum_{n=1}^{\infty} P\left(\left|\frac{\tilde{S}_{k_n} - E(\tilde{S}_{k_n})}{k_n}\right| \geqslant \varepsilon\right) < \infty$$

(schnelle stochastische Konvergenz gegen 0)
Aus der paarweisen Unabhängigkeit der $\tilde{X}_i$ folgt, dass $\tilde{X}_i$ unkorreliert sind, und damit

$$\begin{aligned} P\left(\left|\frac{\tilde{S}_{k_n} - E(\tilde{S}_{k_n})}{k_n}\right| \geqslant \varepsilon\right) \leqslant \frac{1}{k_n^2\varepsilon^2} \cdot \operatorname{var}(\tilde{S}_{k_n}) &= \frac{1}{k_n^2\varepsilon^2} \sum_{i=1}^{k_n} \operatorname{var}(\tilde{X}_i) \\ &\leqslant \frac{1}{k_n^2\varepsilon^2} \sum_{i=1}^{k_n} E\big(\tilde{X}_i^2\big). \end{aligned}$$

Es reicht daher aus zu zeigen, dass

$$s := \sum_{n=1}^{\infty} \left(\frac{1}{k_n^2} \sum_{i=1}^{k_n} E\big(\tilde{X}_i^2\big) \right) = \sum_{\substack{(i,n)\in\mathbb{N}^2, \\ i\leqslant k_n}} \frac{1}{k_n^2} \cdot E\big(\tilde{X}_i^2\big) < \infty\,.$$

Zum Beweis der Konvergenz beachte, dass

$$s = \sum_{i=1}^{\infty} \Big(\sum_{n\,:\,k_n \geqslant i} \frac{1}{k_n^2} \Big) \cdot E\big(\tilde{X}_i^2\big).$$

Wir zeigen im folgenden, dass eine Konstante c existiert mit

$$\sum_{n\,:\,k_n \geqslant i} \frac{1}{k_n^2} \leqslant \frac{c}{i^2}. \tag{2.4}$$

Hieraus ergibt sich dann

$$\begin{aligned}
s &\overset{(2.4)}{\leqslant} c \sum_{i=1}^{\infty} \frac{1}{i^2} \cdot E\big(\tilde{X}_i^2\big) = c \sum_{i=1}^{\infty} \frac{1}{i^2} \cdot E\big(1_{\{X_1 < i\}} \cdot X_1^2\big) \\
&\leqslant \; c \sum_{i=1}^{\infty} \Big(\frac{1}{i^2} \sum_{l=1}^{i} l^2 \cdot P\big(X_1 \in [l-1, l[\big) \Big) \\
&= \; c \sum_{l=1}^{\infty} \Bigg(l^2 \cdot \underbrace{\Big(\sum_{i=l}^{\infty} \frac{1}{i^2} \Big)}_{\leqslant 2l^{-1}} \cdot P\big(X_1 \in [l-1, l[\big) \Bigg) \\
&\leqslant \; 2c \sum_{l=1}^{\infty} l \cdot P\big(X_1 \in [l-1, l[\big) = 2c \sum_{l=1}^{\infty} E\big(\underbrace{l \cdot 1_{\{X_1 \in [l-1, l[\}}}_{\leqslant (X_1+1)\cdot 1_{\{X_1 \in [l-1,l[\}}} \big) \\
&\leqslant \; 2c \cdot \big(E(X_1) + 1\big) < \infty,
\end{aligned}$$

wobei wir verwendet haben, dass

$$\sum_{i=l}^{\infty} \frac{1}{i^2} \leqslant \frac{1}{l^2} + \sum_{i=l+1}^{\infty} \frac{1}{(i-1)i} = \frac{1}{l^2} + \sum_{i=l+1}^{\infty} \Big(\frac{1}{i-1} - \frac{1}{i} \Big) = \frac{1}{l^2} + \frac{1}{l} \leqslant \frac{2}{l}\,.$$

Es bleibt schließlich noch (2.4) zu zeigen. Hierzu beachte, dass

$$\begin{aligned}
& \lfloor \alpha^n \rfloor = k_n \leqslant \alpha^n < k_n + 1 \\
\Rightarrow \quad & k_n > \alpha^n - 1 \overset{\alpha > 1}{\geqslant} \alpha^n - \alpha^{n-1} = \underbrace{\Big(\frac{\alpha - 1}{\alpha} \Big)}_{=: c_\alpha} \alpha^n.
\end{aligned}$$

Es sei n_i die kleinste natürliche Zahl mit $k_{n_i} = \lfloor \alpha^{n_i} \rfloor \geqslant i$, also $\alpha^{n_i} \geqslant i$, so folgt

$$\sum_{n\,:\,k_n \geqslant i} \frac{1}{k_n^2} \leqslant c_\alpha^{-2} \sum_{n \geqslant n_i} \frac{1}{\alpha^{2n}} = c_\alpha^{-2} \cdot \frac{1}{1-\alpha^{-2}} \cdot \alpha^{-2n_i} \leqslant \frac{c_\alpha^{-2}}{1-\alpha^{-2}} \cdot \frac{1}{i^2}. \qquad \square$$

Korollar 2.17. *Es seien $X_1, X_2, \ldots$ paarweise unabhängig und identisch verteilt (im Folgenden kurz: u.i.d.) mit $X_i \geqslant 0$. Dann gilt*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = E(X_1) \quad \big(\in [0, \infty]\big) \qquad \textit{P-f.s.}$$

Beweis. O.B.d.A. $E(X_1)=\infty$. Dann folgt $\frac{1}{n}\sum_{i=1}^n\big(X_i(\omega)\wedge N\big)\xrightarrow{n\to\infty}E(X_1\wedge N)$, P-f.s. für alle N, und somit

$$\frac{1}{n}\sum_{i=1}^n X_i(\omega)\geqslant\frac{1}{n}\sum_{i=1}^n\big(X_i(\omega)\wedge N\big)\xrightarrow{n\to\infty}E(X_1\wedge N)\overset{N\to\infty}{\nearrow}E(X_1)\quad P\text{-f.s.}\qquad\square$$

Beispiel 2.18. Wachstum in zufälligen Medien
Es seien $Y_1,Y_2,\dots$ u.i.d., $Y_i>0$, mit $m:=E(Y_i)$.
Definiere $X_0=1$ und induktiv $X_n:=X_{n-1}\cdot Y_n$. Dann gilt offenbar $X_n=Y_1\cdots Y_n$ und $E(X_n)=E(Y_1)\cdots E(Y_n)=m^n$, und somit

$$E(X_n)\to\begin{cases}+\infty & \text{falls } m>1 \qquad \text{exponentielles Wachstum (superkritisch)}\\ 1 & \text{falls } m=1 \qquad \text{kritisch}\\ 0 & \text{falls } m<1 \qquad \text{exponentieller Abfall (subkritisch)}\end{cases}$$

Wie sieht das Langzeitverhalten von $X_n(\omega)$ aus?
Überraschenderweise ist es auch im superkritischen Fall $m>1$ möglich, dass $\lim_{n\to\infty}X_n=0$ mit positiver Wahrscheinlichkeit.
Erklärung: Wir nehmen an, dass $\log Y_i\in\mathcal{L}^1$. Dann folgt

$$\frac{1}{n}\log X_n=\frac{1}{n}\sum_{i=1}^n\log Y_i\xrightarrow{n\to\infty}E\left(\log Y_1\right)=:\alpha\quad P\text{-f.s.}$$

und es folgt:

$\alpha<0$: $\exists\,\varepsilon>0$ mit $\alpha+\varepsilon<0$, sodass $X_n(\omega)\leqslant e^{n(\alpha+\varepsilon)}\ \forall\,n\geqslant n_0(\omega)$, und damit P-f.s. exponentieller Abfall

$\alpha>0$: $\exists\,\varepsilon>0$ mit $\alpha-\varepsilon>0$, sodass $X_n(\omega)\geqslant e^{n(\alpha-\varepsilon)}\ \forall\,n\geqslant n_0(\omega)$, und damit P-f.s. exponentielles Wachstum

Beachte, dass aufgrund der Jensenschen Ungleichung

$$\alpha=E\left(\log Y_1\right)\quad\leqslant\quad\log\underbrace{E(Y_1)}_{=m},$$

und im Allgemeinen ist diese Ungleichung strikt, d.h. $\alpha<\log m$, sodass es passieren kann, dass $\alpha<0$ obwohl $m>1$ (!)
Illustration Als spezielles Beispiel betrachte

$$Y_i:=\begin{cases}\frac{1}{2}(1+c) & \text{mit Wahrscheinlichkeit } \frac{1}{2}\\ \frac{1}{2} & \text{mit Wahrscheinlichkeit } \frac{1}{2}\end{cases},$$

sodass $E(Y_i)=\frac{1}{4}(1+c)+\frac{1}{4}=\frac{1}{2}+\frac{1}{4}c$ (superkritisch für $c>2$)
Andererseits

$$E\left(\log Y_1\right)=\frac{1}{2}\cdot\left(\log\left(\frac{1}{2}(1+c)\right)+\log\frac{1}{2}\right)=\frac{1}{2}\cdot\log\frac{1+c}{4}\overset{c<3}{<}0.$$

Folglich $X_n\xrightarrow{n\to\infty}0$ P-f.s. mit exponentieller Rate für $c<3$, während zugleich für $c>2$ $E(X_n)=m^n\nearrow\infty$ mit exponentieller Rate.

Zurück zum Kolmogorovschen Gesetz der großen Zahlen:
Es seien $X_1, X_2, \ldots \in \mathcal{L}^1$ u.i.d. mit $m := \mathbb{E}(X_i)$. Dann folgt

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \xrightarrow{n\to\infty} E(X_1) \quad P\text{-f.s.}$$

Definiere die **empirische Verteilung**

$$\begin{aligned}\varrho_n(\omega, A) &:= \frac{1}{n}\sum_{i=1}^{n} 1_A\big(X_i(\omega)\big)\\ &= \texttt{relative Häufigkeit des Ereignisses } X_i \in A\end{aligned}$$

der ersten n **Beobachtungen** $X_1(\omega), \ldots, X_n(\omega)$. Beachte, dass

$$\varrho_n(\omega, \cdot\,) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$$

ein Wahrscheinlichkeitsmaß ist auf $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ für jedes ω.

Satz 2.19. *Für P-fast alle $\omega \in \Omega$ gilt:*

$$\varrho_n(\omega, \cdot\,) \xrightarrow{n\to\infty} \mu := P \circ X_1^{-1} \quad \textit{schwach.}$$

Beweis. Offensichtlich folgt aus dem Kolmogorovschen Gesetz der großen Zahlen für jedes $x \in \mathbb{R}$

$$\begin{aligned}F_n(\omega, x) := \varrho_n\big(\omega,]-\infty, x]\big) &= \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,x]}\big(X_i(\omega)\big)\\ &\to E(1_{]-\infty,x]}(X_1)) = P(X_1 \le x) = \mu\big(]-\infty, x]\big) =: F(x)\end{aligned}$$

P-f.s., also für alle $\omega \notin N(x)$ für eine P-Nullmenge $N(x)$. Dann ist auch

$$N := \bigcup_{r \in \mathbb{Q}} N(r).$$

eine P-Nullmenge, and für alle $x \in \mathbb{R}$ and alle $s, r \in \mathbb{Q}$ mit $s < x < r$ und $\omega \notin N$:

$$\begin{aligned}F(s) = \lim_{n\to\infty} F_n(\omega, s) &\leqslant \liminf_{n\to\infty} F_n(\omega, x)\\ &\leqslant \limsup_{n\to\infty} F_n(\omega, x) \leqslant \lim_{n\to\infty} F_n(\omega, r) = F(r).\end{aligned}$$

Ist also F stetig in x, so folgt für $\omega \notin N$

$$\lim_{n\to\infty} F_n(\omega, x) = F(x)\,,$$

und die schwache Konvergenz folgt aus dem Portmanteau Theorem. $\square$

3 Stochastische Prozesse in diskreter Zeit

Wir erinnern kurz an die Definition der (elementaren) bedingten Wahrscheinlichkeit (siehe Kapitel 2 der VL Wahrscheinlichkeitstheorie I).

Definition 3.1. Es sei $(\Omega, \mathcal{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum und A, $B \in \mathcal{A}$ Ereignisse mit $P(B) > 0$. Dann heißt

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

die **bedingte Wahrscheinlichkeit von A unter der Bedingung B** (oder auch: **die bedingte Wahrscheinlichkeit von A gegeben B)**. Im Falle $P(B) = 0$ setzen wir einfach $P(A \mid B) := 0$.

Wir wollen (elementare) bedingte Wahrscheinlichkeiten im folgenden dazu benutzen, die Verteilungen eines stochastischen Prozesses $X_0, X_1, \ldots, X_n, X_{n+1}, \ldots$, d.h. einer endlichen oder unendlichen Folge von Zufallsvariablen, aus den (elementaren) bedingten Wahrscheinlichkeiten

$$P\left(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0\right)$$

zu berechnen, die wir als Übergangswahrscheinlichkeiten des stochastischen Prozesses auffassen können.
Dazu betrachten wir im folgenden Beispiel die kanonische Situation eines stochastischen Prozesses mit einem diskreten Zustandsraum:

Beispiel 3.2 (Berechnung totaler Wahrscheinlichkeiten aus bedingten Wahrscheinlichkeiten). Es sei S eine endliche Menge, $\Omega := S^{n+1}$, $n \in \mathbb{N}$, und P ein Wahrscheinlichkeitsmaß auf Ω. Weiter sei $X_i : \Omega \to S$, $i = 0, \ldots, n$, die kanonische Projektion $X_i(\omega) := x_i$ für $\omega = (x_0, \ldots, x_n)$.
Wir fassen $0, 1, \ldots, n$ als Zeitpunkte auf und somit die kanonischen Projektionen $(X_i)_{0 \leqslant i \leqslant n}$ als **stochastischen Prozess**. Dementsprechend heisst $\big(X_0(\omega), \ldots, X_n(\omega)\big)$ **Trajektorie** des Prozesses.
Für alle $\omega \in \Omega$ gilt im diskreten Fall entweder $P(\{\omega\}) = 0$ oder

$$\begin{aligned}
P(\{\omega\}) &= P(X_0 = x_0, \ldots, X_n = x_n) \\
&= P(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) \\
&\qquad \cdot P(X_n = x_n \,|\, X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) \\
&\vdots \\
&= P(X_0 = x_0) \\
&\qquad \cdot P(X_1 = x_1 \,|\, X_0 = x_0) \\
&\qquad \cdot P(X_2 = x_2 \,|\, X_0 = x_0, X_1 = x_1) \\
&\qquad \cdots \\
&\qquad \cdot P(X_n = x_n \,|\, X_0 = x_0, \ldots, X_{n-1} = x_{n-1})\,.
\end{aligned}$$

Beachte: $P(\{\omega\}) \neq 0$ impliziert $P(X_0 = x_0, \ldots, X_k = x_k) \neq 0$ für alle $k \in \{0, \ldots, n\}$.
Folgerung: Ein Wahrscheinlichkeitsmaß P auf Ω ist eindeutig durch folgende beiden Daten bestimmt:

Anfangsverteilung: $\mu := P \circ X_0^{-1}$

Übergangswahrscheinlichkeiten: die bedingten Wahrscheinlichkeiten

$$P\left(X_k = x_k \,|\, X_0 = x_0, \ldots, X_{k-1} = x_{k-1}\right)$$

für alle $k \in \{1, \ldots, n\}$ und $(x_0, \ldots, x_k) \in S^{(k+1)}$.

Die **Existenz** eines solchen Wahrscheinlichkeitsmaßes P für gegebene Anfangsverteilung und gegebene Übergangswahrscheinlichkeiten zeigen wir im folgenden Abschnitt. Ein wichtiger Spezialfall sind Markovketten (siehe Kapitel 8, VL Wahrscheinlichkeitstheorie I):

$$P\left(X_k = x_k \,|\, X_0 = x_0, \ldots, X_{k-1} = x_{k-1}\right) = P\left(X_k = x_k \,|\, X_{k-1} = x_{k-1}\right),$$

d.h., die Übergangswahrscheinlichkeiten für X_k hängen nur von X_{k-1} ab.

3.1 Übergangswahrscheinlichkeiten und Satz von Fubini

Es seien $(S_1, \mathcal{S}_1)$ und $(S_2, \mathcal{S}_2)$ messbare Räume.

Definition 3.3. Eine Abbildung

$$\begin{array}{rrcl} K: & S_1 \times \mathcal{S}_2 & \to & [0,1] \\ & (x_1, A_2) & \mapsto & K(x_1, A_2) \end{array}$$

heisst **stochastischer Kern** (von $(S_1, \mathcal{S}_1)$ nach $(S_2, \mathcal{S}_2)$), falls gilt:

(a) $\forall x_1 \in S_1$: $K(x_1, \cdot\,)$ ist Wahrscheinlichkeitsmaß auf $(S_2, \mathcal{S}_2)$.

(b) $\forall A_2 \in \mathcal{S}_2$: $K(\,\cdot\,, A_2)$ ist $\mathcal{S}_1$-messbar.

Beispiel 3.4. (i) Für jedes Wahrscheinlichkeitsmaß μ auf $(S_2, \mathcal{S}_2)$ definiert

$$K(x_1, \cdot\,) := \mu \qquad \forall\, x_1 \in S_1 \text{ keine Kopplung!}$$

einen **stochastischen Kern**.

(ii) Es sei $T : S_1 \to S_2$ eine $\mathcal{S}_1/\mathcal{S}_2$ messbare Abbildung. Dann definiert

$$K(x_1, \cdot\,) := \delta_{T(x_1)} \qquad \forall x_1 \in S_1 .$$

einen **stochastischen Kern**.

(iii) **Stochastische Matrizen** Es sei S_1, S_2 höchstens abzählbar und $\mathcal{S}_i = \mathcal{P}(S_i)$, $i = 1, 2$. In diesem Fall ist jeder stochastische Kern von $(S_1, \mathcal{S}_1)$ nach $(S_2, \mathcal{S}_2)$ bestimmt durch

$$K(x_1, x_2) := K\big(x_1, \{x_2\}\big), \quad x_1 \in S_1,\ x_2 \in S_2,$$

wobei $K : S_1 \times \mathcal{S}_2 \to [0,1]$ eine Abbildung ist mit $\sum_{x_2 \in S_2} K(x_1, \{x_2\}) = 1$ für alle $x_1 \in S_1$. Folglich können wir K mit einer stochastischen Matrix identifizieren, also mit einer Matrix mit nichtnegativen Einträgen und Zeilensummen 1.

Beispiel 3.5. (i) **Übergangswahrscheinlichkeiten der sym. Irrfahrt auf $\mathbb{Z}^d$**

$S_1 = S_2 = S := \mathbb{Z}^d$ mit $\mathcal{S} := \mathcal{P}(\mathbb{Z}^d)$

$$K(x, \cdot\,) := \frac{1}{2d} \sum_{y \in N(x)} \delta_y \,, x \in \mathbb{Z}^d \,,$$

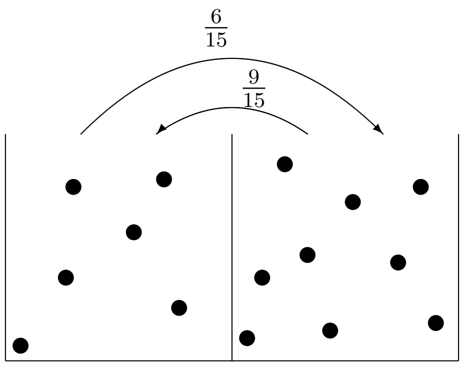
wobei

$$N(x) := \big\{ y \in \mathbb{Z}^d \;\big|\; \|x - y\| = 1 \big\}$$

die Menge der nächsten Nachbarn von x bezeichnet.

(ii) **Ehrenfest Modell**

Ein Behälter enthalte N Kugeln und wird in zwei Teile (*links* und *rechts*) unterteilt.
Nun wird zufällig eine Kugel gewählt und in die jeweils andere Hälfte gelegt.



"Mikroskopische Beschreibung" Zustandsraum $S := \{0,1\}^N$ mit $x = (x_1, \dots, x_N) \in S$ wobei

$$x_i := \begin{cases} 1 & \text{falls die i-te Kugel in der } \textit{linken} \text{ Hälfte} \\ 0 & \text{falls die i-te Kugel in der } \textit{rechten} \text{ Hälfte} \end{cases}$$

zugehöriger stochastischer Kern.

$$K(x, \cdot\,) := \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_1, \dots, x_{i-1}, 1 - x_i, x_{i+1}, \dots, x_N)}.$$

"Makroskopische Beschreibung" Zustandsraum $S := \{0, \dots, N\}$, wobei $j \in S$ die Anzahl der Kugeln in der linken Hälfte bezeichnet. Zugehöriger stochastischer Kern:

$$K(j, \cdot\,) := \frac{N - j}{N} \cdot \delta_{j+1} + \frac{j}{N} \cdot \delta_{j-1}.$$

(iii) **Übergangswahrscheinlichkeiten des Ornstein-Uhlenbeck Prozesses**
$S = S_1 = S_2 = \mathbb{R}$, $K(x, \cdot\,) := N(\alpha x, \sigma^2)$ mit $\alpha \in \mathbb{R}$, $\sigma^2 > 0$.

Wir kommen nun zum Satz von Fubini. Es sei μ_1 in Wahrscheinlichkeitsmaß auf $(S_1, \mathcal{S}_1)$ und $K(\,\cdot\,,\,\cdot\,)$ ein stochastischer Kern von $(S_1, \mathcal{S}_1)$ nach $(S_2, \mathcal{S}_2)$.
Unser Ziel ist die Konstruktion eines Wahrscheinlichkeitsmaßes P $(:= \mu_1 \otimes K)$ auf dem Produktraum $(\Omega, \mathcal{A})$, wobei

$$\begin{aligned}\Omega &:= S_1 \times S_2 \\ \mathcal{A} &:= \mathcal{S}_1 \otimes \mathcal{S}_2 = \sigma\big(\{A_1 \times A_2 \,|\, A_1 \in \mathcal{S}_1,\ A_2 \in \mathcal{S}_2\}\big) \stackrel{!}{=} \sigma(X_1, X_2)\,,\end{aligned}$$

und

$$\begin{array}{rrcll} X_i: & \Omega = S_1 \times S_2 & \to & S_i, & i = 1,2, \\ & (x_1, x_2) & \mapsto & x_i, & \end{array}$$

mit der Eigenschaft

$$P(A_1 \times A_2) = \int_{A_1} K(x_1, A_2)\, \mu_1(dx_1)$$

für alle $A_1 \in \mathcal{S}_1$ und $A_2 \in \mathcal{S}_2$.

Satz 3.6 (Fubini). *Es sei μ_1 ein Wahrscheinlichkeitsmaß auf $(S_1, \mathcal{S}_1)$, K ein stochastischer Kern von $(S_1, \mathcal{S}_1)$ nach $(S_2, \mathcal{S}_2)$, und*

$$\begin{aligned}\Omega &:= S_1 \times S_2, \\ \mathcal{A} &:= \sigma\big(\{A_1 \times A_2 \,|\, A_i \in \mathcal{S}_i\}\big) =: \mathcal{S}_1 \otimes \mathcal{S}_2\,.\end{aligned}$$

Dann gibt es ein Wahrscheinlichkeitsmaß P $(=: \mu_1 \otimes K)$ auf $(\Omega, \mathcal{A})$, so dass für alle $\mathcal{A}$-messbaren Funktionen $f \geqslant 0$ gilt:

$$\int_\Omega f\, dP = \int \left(\int f(x_1, x_2)\, K(x_1, dx_2) \right) \mu_1(dx_1), \tag{3.1}$$

Insbesondere gilt für alle $A \in \mathcal{A}$

$$P(A) = \int K(x_1, A_{x_1})\, \mu_1(dx_1)\,. \tag{3.2}$$

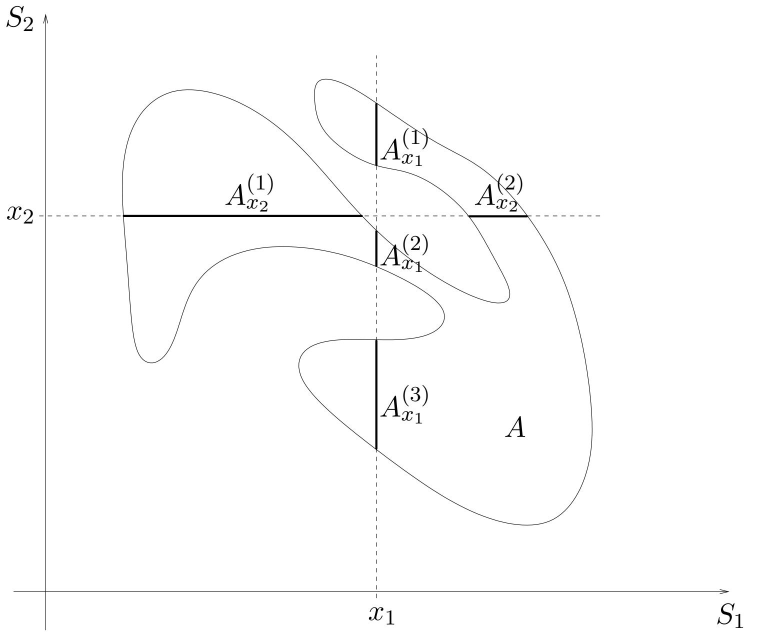
Hierbei ist

$$A_{x_1} = \{x_2 \in S_2 \,|\, (x_1, x_2) \in A\}$$

der x_1 Schnitt der Menge A. Insbesondere folgt für $A_1 \in \mathcal{S}_1, A_2 \in \mathcal{S}_2$:

$$P(A_1 \times A_2) = \int_{A_1} K(x_1, A_2)\, \mu_1(dx_1). \tag{3.3}$$

P ist durch (3.3) *eindeutig bestimmt.*



Beachte

$$A_{x_1} = A_{x_1}^{(1)} \cup A_{x_1}^{(2)} \cup A_{x_1}^{(3)} \quad \text{und} \quad A_{x_2} = A_{x_2}^{(1)} \cup A_{x_2}^{(2)}.$$

Beweis. Eindeutigkeit: Die Menge aller Zylindermengen $A_1 \times A_2$ mit $A_i \in \mathcal{S}_i$ ist $\cap$-stabiler Erzeuger der Produkt-σ-Algebra $\mathcal{A}$. Die Eindeutigkeit folgt nun aus Satz 1.18
Existenz: Für $x_1 \in S_1$ sei

$$\varphi_{x_1}(x_2) := (x_1, x_2)\,.$$

$\varphi_{x_1} : S_2 \to \Omega$ ist messbar, denn für $A_1 \in \mathcal{S}_1, A_2 \in \mathcal{S}_2$ gilt

$$\varphi_{x_1}^{-1}(A_1 \times A_2) = \begin{cases} \emptyset & \text{für } x_1 \notin A_1 \\ A_2 & \text{für } x_1 \in A_1 \end{cases}$$

Somit ist für jede $\mathcal{A}$-messbare Funktion $f : \Omega \to \mathbb{R}$ und jedes $x_1 \in S_1$ die Abbildung

$$f_{x_1} := f \circ \varphi_{x_1} : S_2 \to \mathbb{R}\,, x_2 \mapsto f(x_1, x_2)$$

$\mathcal{S}_2/\mathcal{B}(\mathbb{R})$-messbar. Es sei nun $f \geqslant 0$ oder beschränkt. Dann ist die Abbildung

$$x_1 \mapsto \int f(x_1, x_2)\; K(x_1, dx_2) \left(= \int f_{x_1}(x_2)\; K(x_1, dx_2) \right) \tag{3.4}$$

wohldefiniert.

Wir zeigen im folgenden, dass diese Abbildung $\mathcal{S}_1$-messbar ist. Dazu sei zunächst $f = 1_A$ für $A \in \mathcal{A}$ gegeben. Für allgemeine f folgt die Messbarkeit dann mit Hilfe einer maßtheoretischen Induktion.
Für $f = 1_A$ gilt nun aber

$$\int \underbrace{1_A(x_1, x_2)}_{=1_{A_{x_1}}(x_2)} \, K(x_1, dx_2) = K(x_1, A_{x_1}).$$

und wir betrachten entsprechend im folgenden das Mengensystem

$$\mathcal{D} := \big\{ A \in \mathcal{A} \;\big|\; x_1 \mapsto K(x_1, A_{x_1}) \ \mathcal{S}_1\text{-messbar} \big\}.$$

Dann ist $\mathcal{D}$ ein Dynkin-System (!) und enthält alle Zylindermengen $A = A_1 \times A_2$ mit $A_i \in \mathcal{S}_i$, denn

$$K\big(x_1, (A_1 \times A_2)_{x_1}\big) = 1_{A_1}(x_1) \cdot K(x_1, A_2)\,.$$

Da das Mengensystem der Zylindermengen ein $\cap$-stabiler Erzeuger der Produkt-σ-Algebra $\mathcal{A}$ ist, folgt $\mathcal{D} = \mathcal{A}$. Insbesondere folgt, dass für alle nichtnegativen oder beschränkten $\mathcal{A}$-messbaren Funktionen $f : \Omega \to \mathbb{R}$ das Integral

$$\int \left(\int f(x_1, x_2) \, K(x_1, dx_2) \right) \mu(dx_1)$$

wohldefiniert ist.
Für alle $A \in \mathcal{A}$ können wir nun definieren

$$P(A) := \int \left(\int \underbrace{1_A(x_1, x_2)}_{=1_{A_{x_1}}(x_2)} \, K(x_1, dx_2) \right) \mu(dx_1) = \int K(x_1, A_{x_1}) \, \mu(dx_1).$$

P ist ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$, denn

$$P(\Omega) = \int K(x_1, S_2) \, \mu(dx_1) = \int 1 \, \mu(dx_1) = 1.$$

Zum Beweis der σ-Additivität seien $A_1, A_2, \ldots$ paarweise disjunkte Mengen in $\mathcal{A}$. Dann sind für alle $x_1 \in S_1$ die Mengen $(A_1)_{x_1}, (A_2)_{x_1}, \ldots$ ebenfalls paarweise disjunkt und somit

$$\begin{aligned} P\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) &= \int K\bigg(x_1, \Big(\bigcup_{n \in \mathbb{N}} A_n\Big)_{x_1}\bigg) \, \mu(dx_1) \\ &= \int \sum_{n=1}^{\infty} K\big(x_1, (A_n)_{x_1}\big) \, \mu(dx_1) \\ &= \sum_{n=1}^{\infty} \int K\big(x_1, (A_n)_{x_1}\big) \, \mu(dx_1) = \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

In der zweiten Gleichheit haben wir verwendet, dass $K(x_1, \cdot)$ für alle x_1 ein Wahrscheinlichkeitsmaß ist und in der dritten Gleichheit wurde der Satz von der monotonen Integration verwendet. Schließlich folgt (3.1) mit maßtheoretischer Induktion. $\square$

Beispiele und Anwendungen

Bemerkung 3.7. *Der klassische Satz von Fubini ist ein Spezialfall von Satz 3.6:* $K(x_1, \cdot\,) = \mu_2$. *In diesem Fall heißt das im Satz von Fubini konstruierte Maß* $\mu_1 \otimes K$, *das* ***Produktmaß*** *von* μ_1 *und* μ_2 *und wird mit* $\mu_1 \otimes \mu_2$ *bezeichnet. Insbesondere gilt in diesem Fall*

$$\int f\,dP = \int \left(\int f(x_1, x_2)\, \mu_2(dx_2) \right) \mu_1(dx_1).$$

Bemerkung 3.8 (Randverteilungen). *Es seien* $X_i : \Omega \to S_i$, $i = 1, 2$, *die natürlichen Projektionen* $X_i\big((x_1, x_2)\big) := x_i$. *Die Verteilungen von* X_i *unter* $\mu_1 \otimes K$ *heißen* ***Randverteilungen*** *und sind gegeben durch*

$$\begin{aligned}(P \circ X_1^{-1})(A_1) &= P(X_1 \in A_1) = P(A_1 \times S_2) \\ &= \int_{A_1} \underbrace{K(x_1, S_2)}_{=1}\, \mu_1(dx_1) = \mu_1(A_1)\end{aligned}$$

sowie

$$\begin{aligned}(P \circ X_2^{-1})(A_2) &= P(X_2 \in A_2) = P(S_1 \times A_2) \\ &= \int K(x_1, A_2)\, \mu_1(dx_1) =: (\mu_1 K)(A_2).\end{aligned}$$

Insbesondere gilt also für die Randverteilungen

$$P \circ X_1^{-1} = \mu_1 \qquad P \circ X_2^{-1} = \mu_1 K\,.$$

Definition 3.9. Es sei $S_1 = S_2 = S$ und $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$. Ein Wahrscheinlichkeitsmaß μ auf $(S, \mathcal{S})$ heißt **invariantes Maß** (bzw. **Gleichgewichtsverteilung**) **für** K falls $\mu = \mu K$.

Beispiel 3.10. (i) **Ehrenfest Modell** (makroskopisch) Es sei $S = \{0, 1, \dots, N\}$ und

$$K(y, \cdot\,) = \frac{y}{N} \cdot \delta_{y-1} + \frac{N-y}{N} \cdot \delta_{y+1}.$$

In diesem Fall ist die Binomialverteilung $Bin(N, \frac{1}{2})$ ein invariantes Maß, denn

$$\begin{aligned}(\mu K)(\{x\}) &= \sum_{y \in S} \mu(\{y\}) \cdot K(y, x) \\ &= \mu(\{x+1\}) \cdot \frac{x+1}{N} + \mu(\{x-1\}) \cdot \frac{N-(x-1)}{N} \\ &= 2^{-N} \binom{N}{x+1} \cdot \frac{x+1}{N} + 2^{-N} \binom{N}{x-1} \cdot \frac{N-(x-1)}{N} \\ &= 2^{-N} \left[\binom{N-1}{x} + \binom{N-1}{x-1} \right] = 2^{-N} \cdot \binom{N}{x} = \mu(\{x\}).\end{aligned}$$

(ii) **Ornstein-Uhlenbeck Prozess** Es sei $S = \mathbb{R}$ und $K(x, \cdot\,) = N(\alpha x, \sigma^2)$ mit $|\alpha| < 1$ gegeben. Dann ist

$$\mu = N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$$

ein invariantes Maß (Übung.)

Wir betrachten nun das umgekehrte Problem: Gegeben sei ein Wahrscheinlichkeitsmaß P auf dem Produktraum $(\Omega, \mathcal{A})$. Können wir dann P **disintegrieren**, d.h., gibt es ein Wahrscheinlichkeitsmaß μ_1 auf $(S_1, \mathcal{S}_1)$ und einen stochastischen Kern K von S_1 to S_2 so dass

$$P = \mu_1 \otimes K\,?$$

In den meisten Fällen ja, z.B. falls S_1 und S_2 polnische Räume sind (d.h., topologische Räume mit abzählbarer Basis, deren Topologie durch eine vollständige Metrik induziert ist) ist dies mithilfe bedingter Erwartungen möglich (siehe unten).

Beispiel 3.11. Im Spezialfall, wenn S_1 abzählbar ist (und $\mathcal{S}_1 = \mathcal{P}(S_1)$), können wir P folgendermaßen disintegrieren: Notwendigerweise ist μ_1 die Verteilung der Projektion X_1 auf die erste Koordinate. Für die Definition des Kerns K sei ν ein beliebiges Wahrscheinlichkeitsmaß auf $(S_2, \mathcal{S}_2)$. Setze dann

$$K(x_1, A_2) := \begin{cases} P(X_2 \in A_2 \,|\, X_1 = x_1) & \text{falls } \underbrace{\mu_1(\{x_1\})}_{=P(X_1=x_1)} > 0 \\ \\ \nu(A_2) & \text{falls } \mu_1(\{x_1\}) = 0. \end{cases}$$

Dann gilt

$$\begin{aligned} P(A_1 \times A_2) = P(X_1 \in A_1,\ X_2 \in A_2) &= \sum_{x_1 \in A_1} P(X_1 = x_1,\ X_2 \in A_2) \\ &= \sum_{\substack{x_1 \in A_1, \\ \mu_1(\{x_1\})>0}} P(X_1 = x_1) \cdot P(X_2 \in A_2 \,|\, X_1 = x_1) \\ &= \sum_{x_1 \in A_1} \mu_1(\{x_1\}) \cdot K(x_1, A_2) = \int_{A_1} K(x_1, A_2)\, \mu_1(dx_1) \\ &= (\mu_1 \otimes K)(A_1 \times A_2), \end{aligned}$$

und somit $P = \mu_1 \otimes K$.

Im nächsten Satz geben wir für die Disintegration absolutstetiger Wahrscheinlichkeitsmaße eine explizite Formel an.
Beachte: Ist P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ und $\varphi : \Omega \to \mathbb{R}_+$ $\mathcal{A}$-messbar mit $\int \varphi\, dP = 1$, so definiert

$$(\varphi P)(A) := \int_A \varphi\, dP$$

ein weiteres Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$.
Für einen stochastischen Kern K von S_1 nach S_2 und eine $\mathcal{A}$-messbare Funktion $\varphi : \Omega \to \mathbb{R}_+$, sei

$$K\varphi(x) := \int K(x, dy)\ \varphi(x, y)\,, x \in S_1\,.$$

Satz 3.12. *Es sei $P = \mu \otimes K$ und $\tilde{P} := \varphi P$. Dann gilt $\tilde{P} = \tilde{\mu} \otimes \tilde{K}$ mit*

$$\tilde{\mu} = (K\varphi)\mu \quad \text{und} \quad \tilde{K}(x, dy) := \frac{\varphi(x, y)}{K\varphi(x)} \cdot K(x, dy)$$

für alle $x \in S_1$ mit $K\varphi(x) > 0$ (und $\tilde{K}(x, \cdot\,) = \nu, \nu$ beliebiges Wahrscheinlichkeitsmaß auf $(S_2, \mathcal{S}_2)$ falls $x \in S_2$ so dass $K\varphi(x) = 0$).

Beweis. (i) Es sei $\tilde{\mu}$ die Verteilung von X_1 unter $\tilde{P}$. Dann gilt für alle $A \in \mathcal{S}_1$

$$\begin{aligned}\tilde{\mu}(A) = \tilde{P}(A \times S_2) &= \int_{A \times S_2} \varphi(x, y)\ dP \\ &= \int 1_A(x) \left(\int \varphi(x, y)\ K(x, dy) \right) \mu(dx) = \int_A (K\varphi)(x)\ \mu(dx),\end{aligned}$$

also $\tilde{\mu} = (K\varphi)\mu$. Insbesondere ist $K\varphi > 0\ \tilde{\mu}$-f.-ü., denn

$$\tilde{\mu}(K\varphi = 0) = \int_{\{K\varphi = 0\}} (K\varphi)(x)\, \mu(dx) = 0\,.$$

(ii) Es sei $\tilde{K}$ wie oben. Damit ist $\tilde{K}$ ein stochastischer Kern, denn

$$\int \varphi(x, y)\ K(x, dy) = K\varphi(x)\,,\ \text{so dass } \tilde{K}(x, S_2) = 1 \quad \text{für alle } x \in S_1\,.$$

Für alle $A \in \mathcal{S}_1$ und $B \in \mathcal{S}_2$ folgt

$$\begin{aligned}\tilde{P}(A \times B) &= \int \Bigg(\underbrace{\int 1_{A \times B}(x, y) \cdot \varphi(x, y)\ K(x, dy)}_{\leqslant K\varphi(x)} \Bigg) \mu(dx) \\ &= \int_{\{K\varphi > 0\}} \left(\int 1_{A \times B}(x, y) \cdot \varphi(x, y)\ K(x, dy) \right) \mu(dx) \\ &= \int_A K\varphi(x)\ \tilde{K}(x, B)\ \mu(dx) = \int_A \tilde{K}(x, B)\ \tilde{\mu}(dx) \\ &= (\tilde{\mu} \otimes \tilde{K})(A \times B).\end{aligned}$$

$\square$

3.2 Das kanonische Modell für die zeitliche Entwicklung stochastischer Prozesse in diskreter Zeit

Wir betrachten im gesamten Abschnitt folgende Situation: Gegeben seien

- messbare Räume $(S_i, \mathcal{S}_1)$, $i = 0, 1, 2, \ldots$ und wir definieren

$$S^n := S_0 \times S_1 \times \cdots \times S_n$$
$$\mathcal{S}^n := \mathcal{S}_0 \otimes \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n = \sigma\big(\{A_0 \times \cdots \times A_n \,|\, A_i \in \mathcal{S}_i\}\big).$$

- - eine Anfangsverteilung μ_0 auf $(S_0, \mathcal{S}_0)$
  - stochastische Kerne

$$K_n\big((x_0, \ldots, x_{n-1}), dx_n\big)$$

  von $(S^{n-1}, \mathcal{S}^{n-1})$ nach $(S_n, \mathcal{S}_n)$, $n = 1, 2, \ldots$.

Mithilfe des Satzes von Fubini können wir dann Wahrscheinlichkeitsmaße P^n auf S^n, $n = 0, 1, 2, \ldots$ folgendermaßen definieren:

$$\begin{aligned} P^0 &:= \mu_0 && \text{auf } S_0, \\ P^n &:= P^{n-1} \otimes K_n && \text{auf } S^n = S^{n-1} \times S_n,\ n = 0, 1, 2, \ldots \end{aligned}$$

Nach dem Satz von Fubini (Satz 3.6) gilt dann für jede $\mathcal{S}^n$-messbare Funktion $f : S^n \to \mathbb{R}_+$:

$$\begin{aligned} &\int f \, dP^n \\ &= \int P^{n-1}\big(d(x_0, \ldots, x_{n-1})\big) \int K_n\big((x_0, \ldots, x_{n-1}), dx_n\big)\ f(x_0, \ldots, x_{n-1}, x_n) \\ &= \cdots \\ &= \int \mu_0(dx_0) \int K_1(x_0, dx_1) \cdots \int K_n\big((x_0, \ldots, x_{n-1}), dx_n\big)\ f(x_0, \ldots, x_n). \end{aligned}$$

Es sei $\Omega := S_0 \times S_1 \times \ldots$ die Menge aller Pfade (oder Trajektorien) $\omega = (x_0, x_1, \ldots)$ mit $x_i \in S_i$, und

$$\begin{aligned} &X_n(\omega) := x_n \qquad \text{(Projektion auf } n\text{-te Koordinate)}, \\ &\mathcal{A}_n := \sigma(X_0, \ldots, X_n) \quad (\subseteq \mathcal{A}), \\ &\mathcal{A} := \sigma(X_0, X_1, \ldots) = \sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{A}_n\Big). \end{aligned}$$

Unser Hauptziel in diesem Abschnitt ist dann die Konstruktion eines Wahrscheinlichkeitsmaßes P auf $(\Omega, \mathcal{A})$ für das gilt

$$\int f(X_0, \ldots, X_n) \, dP = \int f \, dP^n \qquad \forall\, n = 1, 2, \ldots$$

Mit anderen Worte: Die **endlichdimensionalen Verteilungen** unter P, d.h., die gemeinsamen Verteilungen von $(X_0, \ldots, X_n)$ unter P sind gerade die Wahrscheinlichkeitsmaße P^n.

Satz 3.13 (Ionescu-Tulcea). *Es gibt ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{A})$, so dass für alle $n \geqslant 0$ und alle $\mathcal{S}^n$-messbaren Funktionen $f : S^n \to \mathbb{R}_+$:*

$$\int_{S^n} f \, dP_{(X_0,\ldots,X_n)} = \int_\Omega f(X_0, \ldots, X_n) \, dP = \int_{S^n} f \, dP^n. \tag{3.5}$$

Mit anderen Worte: es gibt genau ein P mit $P^n = P \circ (X_0, \ldots, X_n)^{-1}$ für alle n.

Beweis. Eindeutigkeit: Offensichtlich, denn die Familie der endlichdimensionalen Zylindermengen

$$\mathcal{E} := \left\{ \bigcap_{i=0}^{n} \{X_i \in A_i\} \,\middle|\, n \geqslant 0,\ A_i \in \mathcal{S}_i \right\}$$

ist $\bigcap$ - stabiler Erzeuger von $\mathcal{A}$.
Existenz: Es sei $A \in \mathcal{A}_n$, also $A = (X_0, \ldots, X_n)^{-1}(A^n)$ für ein $A^n \in \mathcal{S}^n$. Dann gilt insbesondere $1_A = 1_{A^n}(X_0, \ldots, X_n)$. Damit (3.5) erfüllt ist, müssen wir notwendigerweise definieren:

$$P(A) := P^n(A^n)\,. \tag{3.6}$$

P ist durch (3.6) in der Tat wohldefiniert, denn für $A \in \mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ mit

$$A = A^n \times S_{n+1} \times S_{n+2} \times \cdots \quad = A^{n+1} \times S_{n+2} \times \cdots\,,$$

gilt $A^{n+1} = A^n \times S_{n+1}$ und somit,

$$\begin{aligned} P^{n+1}(A^{n+1}) &= P^{n+1}(A^n \times S_{n+1}) \\ &= \int_{A^n} \underbrace{K_{n+1}\big((x_0, \ldots, x_n), S_{n+1}\big)}_{=1} \, dP^n = P^n(A^n)\,. \end{aligned}$$

Wir erhalten also durch (3.6) einen wohldefinierten Inhalt P auf $\mathcal{B} = \bigcup_{n=0}^\infty \mathcal{A}_n$.
$\mathcal{B}$ ist eine Algebra (d.h., ein Mengensystem, das Ω enthält und abgeschlossen ist unter Komplementen und endlichen (!) Vereinigungen (vgl. Definition 1.21)) und P ist offensichtlich endlich additiv auf $\mathcal{B}$, denn die Einschränkungen von P auf $\mathcal{A}_n$ sind für alle n (σ-) additiv. Um P nun zu einem σ-additiven Wahrscheinlichkeitsmaß auf $\mathcal{A} = \sigma(\mathcal{B})$ fortzusetzen, genügt es nach dem Fortsetzungssatz von Caratheodory (Satz 1.25) zu zeigen, dass P ein Prämaß, also σ-additiv auf $\mathcal{B}$, ist. Hierzu ist es nach Lemma 1.24 ausreichend zu zeigen, dass P $\emptyset$-stetig ist, d.h., dass folgende Bedingung erfüllt ist:

$$B_n \in \mathcal{B}\,, B_n \downarrow \emptyset \quad \Rightarrow \quad P(B_n) \xrightarrow{n\to\infty} 0\,.$$

O.B.d.A. $B_0 = \Omega$ und $B_n \in \mathcal{A}_n$ (für $B_n \in \mathcal{A}_m$, wiederhole man B_{n-1} m-fach !). Dann gilt

$$B_n = A^n \times S_{n+1} \times S_{n+2} \times \ldots$$

mit

$$A^{n+1} \subseteq A^n \times S_{n+1}$$

und wir müssen zeigen, dass

$$\big(P(B_n) =\big) \quad P^n(A^n) \xrightarrow{n\to\infty} 0 \quad \text{(d.h., } \inf_n P^n(A^n) = 0\text{).}$$

Nehmen wir an, es würde

$$\inf_{n\in\mathbb{N}} P^n(A^n) > 0 \text{ gelten.}$$

Dann haben wir zu zeigen, dass $\bigcap_{n=0}^{\infty} B_n \neq \emptyset$. Beachte nun, dass

$$P^n(A^n) = \int \mu_0(dx_0)\ f_{0,n}(x_0)$$

mit

$$f_{0,n}(x_0) := \int K_1(x_0, dx_1) \cdots \int K_n\big((x_0, \dots, x_{n-1}), dx_n\big)\ 1_{A^n}(x_0, \dots, x_n).$$

Es ist klar, dass die Folge $(f_{0,n})_{n\in\mathbb{N}}$ monoton fällt, denn

$$\begin{aligned}
&\int K_{n+1}\big((x_0, \dots, x_n), dx_{n+1}\big)\ 1_{A^{n+1}}(x_0, \dots, x_{n+1}) \\
&\qquad \leqslant \int K_{n+1}\big((x_0, \dots, x_n), dx_{n+1}\big)\ 1_{A^n\times S_{n+1}}(x_0, \dots, x_{n+1}) \\
&\qquad = 1_{A^n}(x_0, \dots, x_n)\,,
\end{aligned}$$

und somit gilt

$$\begin{aligned}
f_{0,n+1}(x_0) &= \int K_1(x_0, dx_1) \cdots \int K_{n+1}\big((x_0, \dots, x_n), dx_{n+1}\big)\ 1_{A^{n+1}}(x_0, \dots, x_{n+1}) \\
&\leq \int K_1(x_0, dx_1) \cdots \int K_n\big((x_0, \dots, x_{n-1}), dx_n\big)\ 1_{A^n}(x_0, \dots, x_n) = f_{0,n}(x_0)\ .
\end{aligned}$$

Insbesondere ergibt sich

$$\int \inf_{n\in\mathbb{N}} f_{0,n}\ d\mu_0 = \inf_{n\in\mathbb{N}} \int f_{0,n}\ d\mu_0 = \inf_{n\in\mathbb{N}} P^n(A^n) > 0\,.$$

Es gibt also ein $\bar{x}_0 \in S_0$ mit

$$\inf_{n\in\mathbb{N}} f_{0,n}(\bar{x}_0) > 0\,.$$

Andererseits können wir schreiben

$$f_{0,n}(\bar{x}_0) = \int K_1(\bar{x}_0, dx_1)\ f_{1,n}(x_1)$$

mit

$$f_{1,n}(x_1) := \int K_2\big((\bar{x}_0, x_1), dx_2\big) \\ \cdots \int K_n\big((\bar{x}_0, x_1, \dots, x_{n-1}),\ dx_n\big)\ 1_{A^n}(\bar{x}_0, x_1, \dots, x_n)\,.$$

Mithilfe derselben Argumentation wie oben (mit $\mu_1 = K_1(\bar{x}_0, \cdot)$) finden wir ein $\bar{x}_1 \in S_1$ mit

$$\inf_{n\in\mathbb{N}} f_{1,n}(\bar{x}_1) > 0\,.$$

Iterieren wir dieses Verfahren, so finden wir für jedes $i = 0, 1, \dots$ ein $\bar{x}_i \in S_i$ so dass für alle $m \geqslant 0$

$$\inf_{n\in\mathbb{N}} \int K_m\big((\bar{x}_0, \dots, \bar{x}_{m-1}), dx_m\big) \\ \cdots \int K_n\big((\bar{x}_0, \dots, \bar{x}_{m-1}, x_m, \dots, x_{n-1}), dx_n\big) 1_{A^n}(\bar{x}_0, \dots, \bar{x}_{m-1}, x_m, \dots, x_n) > 0\,.$$

Insbesondere gilt für $m = n$

$$0 < \int K_m\big((\bar{x}_0, \dots, \bar{x}_{m-1}), dx_m\big)\ 1_{A^m}(\bar{x}_0, \dots, \bar{x}_{m-1}, x_m) \\ \leqslant 1_{A^{m-1}}(\bar{x}_0, \dots \bar{x}_{m-1})\,,$$

so dass

$$(\bar{x}_0, \dots, \bar{x}_{m-1}) \in A^{m-1} \quad \text{und} \quad \bar{\omega} := (\bar{x}_0, \bar{x}_1, \dots) \ \in \ \underbrace{B_{m-1}}_{=A^{m-1}\times S_m \times S_{m+1}\times\cdots}$$

für alle $m \geq 1$, d.h.,

$$\bar{\omega} \in \bigcap_{m=0}^{\infty} B_m\,.$$

Hieraus folgt die Behauptung. □

Definition 3.14. Es gelte $(S_i, \mathcal{S}_i) = (S, \mathcal{S})$ für alle $i = 0, 1, 2, \dots$. Dann heißt die Folge der Projektionen $(X_n)_{n\geqslant 0}$ auf $(\Omega, \mathcal{A}, P)$ (mit P wie im vorherigen Satz) **stochastischer Prozess** (in diskreter Zeit) mit **Zustandsraum** $(S, \mathcal{S})$, **Anfangsverteilung** μ_0 und **Übergangswahrscheinlichkeiten** $\big(K_n(\,\cdot\,,\,\cdot\,)\big)_{n\in\mathbb{N}}$.

Beispiele

1) Unendliche Produktmaße

Es sei

$$K_n\big((x_0, \dots, x_{n-1}),\ \cdot\,\big) = \mu_n\,,$$

unabhängig von $(x_0, \dots, x_{n-1})$: Dann heißt

$$P =: \bigotimes_{n=0}^{\infty} \mu_n$$

das **Produktmaß von** $\mu_0, \mu_1, \dots$.
Für alle $n \geqslant 0$ und $A_0, \dots, A_n \in \mathcal{S}$ gilt

$$\begin{aligned}
P(X_0 \in A_0, \dots, X_n \in A_n) &\overset{\text{I.-T.}}{=} P^n(A_0 \times \cdots \times A_n) \\
&= \int \mu_0(dx_0) \int \mu_1(dx_1) \cdots \int \mu_n(dx_n)\, 1_{A_0 \times \cdots \times A_n}(x_0, \dots, x_n) \\
&= \mu_0(A_0) \cdot \mu_1(A_1) \cdots \mu_n(A_n).
\end{aligned}$$

Insbesondere also $P \circ X_n^{-1} = \mu_n$ für alle n, und die natürlichen Projektionen $X_0, X_1, \dots$ sind unabhängig. Wir haben also

Satz 3.15. *Es sei (μ_n) eine Folge von Wahrscheinlichkeitsmaßen auf einen messbaren Raum $(S, \mathcal{S})$. Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine Folge (X_n) von unabhängigen Zufallsvariablen mit $P \circ X_n^{-1} = \mu_n$ für alle n.*

Insbesondere haben wir damit die Existenz eines Wahrscheinlichkeitsraumes für unendlich viele unabhängige $0-1$- Experimente gezeigt!

2) Markovketten

$$K_n\big((x_0, \dots, x_{n-1}), \,\cdot\,\big) = \tilde{K}_n(x_{n-1}, \,\cdot\,)$$

zeitlich homogen, falls $\tilde{K}_n = K$ für alle n.
Für gegebene Anfangsverteilung μ_0 und Übergangswahrscheinlichkeiten K gibt es ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{A})$. P heißt **kanonisches Modell** für die zeitliche Entwicklung der **Markovkette** mit $P \circ X_0^{-1} = \mu_0$ und

$$P(X_{n+1} \in A_{n+1} \mid\; X_0 \in A_0, \dots, X_n \in A_n) = P(X_{n+1} \in A_{n+1} \mid\; X_n \in A_n)\,.$$

Beispiel 3.16. Es sei $S = \mathbb{R}$, $\beta > 0$, $x_0 \in \mathbb{R} \setminus \{0\}$, $\mu_0 = \delta_{x_0}$ und $K(x, \,\cdot\,) = N(0, \beta x^2)$ $(K(0, \,\cdot\,) = \delta_0)$
Für welche β konvergiert die Folge (X_n) und was ist ihr Grenzwert?
Für $n \geqslant 1$

$$\begin{aligned}
E(X_n^2) &\overset{\text{I.-T.}}{=} \int x_n^2\, P^n\big(d(x_0, \dots, x_n)\big) \\
&= \int \Bigg(\underbrace{\int x_n^2\, K(x_{n-1}, dx_n)}_{\substack{= \beta x_{n-1}^2, \\ K(x_{n-1}, dx_n) = N(0, \beta x_{n-1}^2)}} \Bigg)\, P^{n-1}(dx_0, \dots, dx_{n-1}) \\
&= \beta \cdot E(X_{n-1}^2) = \cdots = \beta^n x_0^2\,.
\end{aligned}$$

Ist $\beta < 1$ so folgt

$$E\Big(\sum_{n=1}^{\infty} X_n^2\Big) = \sum_{n=1}^{\infty} E(X_n^2) = \sum_{n=1}^{\infty} \beta^n x_0^2 < \infty,$$

also $\sum\limits_{n=1}^{\infty} X_n^2 < \infty$ P-f.s., und damit

$$\lim_{n\to\infty} X_n = 0 \quad P\text{-f.s.}$$

Eine ähnliche Rechnung für das erste absolute Moment zeigt

$$E\big(|X_n|\big) = \cdots = \sqrt{\frac{2}{\pi}\beta} \cdot E\big(|X_{n-1}|\big) = \cdots = \left(\frac{2}{\pi}\beta\right)^{\frac{n}{2}} \cdot \underbrace{E\big(|X_0|\big)}_{=|x_0|},$$

denn

$$\int |X_n| \; K(x_{n-1}, dx_n) = \sqrt{\frac{2}{\pi}}\sigma = \sqrt{\frac{2}{\pi}\beta\,|x_{n-1}|}\,.$$

Folglich,

$$E\Big(\sum_{n=1}^{\infty} |X_n|\Big) = \sum_{n=1}^{\infty} \left(\frac{2}{\pi} \cdot \beta\right)^{\frac{n}{2}} \cdot |x_0|\,,$$

so dass für $\beta < \frac{\pi}{2}$:

$$\lim_{n\to\infty} X_n = 0 \quad P\text{-f.s.}$$

In der Tat, gilt sogar

$$\forall\ \beta < \beta_0 : X_n \xrightarrow{n\to\infty} 0 \quad P\text{-f.s. mit exp. Rate}$$
$$\forall\ \beta > \beta_0 : |X_n| \xrightarrow{n\to\infty} \infty \quad P\text{-f.s. mit exp. Rate.}$$

wobei

$$\begin{aligned} \beta_0 &:= \exp\Big(-\frac{4}{\sqrt{2\pi}} \int_0^{\infty} \log x \cdot e^{-\frac{x^2}{2}} \; dx\Big) \\ &= 2e^C \approx 3.56, \end{aligned}$$

Hierbei ist

$$C := \lim_{n\to\infty} \Big(\sum_{k=1}^{n} \frac{1}{k} - \log n\Big) \approx 0.577$$

die Euler-Mascheroni Konstante.

Beweis. Es ist einfach zu sehen dass für alle n: $X_n \neq 0$ P-f.s. Für $n \in \mathbb{N}$ definiere

$$Y_n := \begin{cases} \frac{X_n}{X_{n-1}} & \text{auf } \{X_{n-1} \neq 0\} \\ 0 & \text{auf } \{X_{n-1} = 0\}. \end{cases}$$

Dann sind $Y_1, Y_2, \dots$ unabhängige Zufallsvariablen mit Verteilung $N(0, \beta)$, denn für alle messbaren Funktionen $f : \mathbb{R}^n \to \mathbb{R}_+$ gilt

$$\begin{aligned}\int f(Y_1, \dots, Y_n)\, dP &\overset{\text{I.-T.}}{=} \int f\Big(\frac{x_1}{x_0}, \dots, \frac{x_n}{x_{n-1}}\Big) \cdot \Big(\frac{1}{2\pi\beta}\Big)^{\frac{n}{2}} \cdot \Big(\frac{1}{x_0^2 \cdots x_{n-1}^2}\Big)^{\frac{1}{2}} \\ &\qquad \cdot \exp\Big(-\frac{x_1^2}{2\beta x_0^2} - \dots - \frac{x_n^2}{2\beta x_{n-1}^2}\Big)\, dx_1 \dots dx_n \\ &= \int f(y_1, \dots, y_n) \cdot \Big(\frac{1}{2\pi\beta}\Big)^{\frac{n}{2}} \cdot \exp\Big(-\frac{y_1^2 + \dots + y_n^2}{2\beta}\Big)\, dy_1 \dots dy_n\,.\end{aligned}$$

Beachte nun dass

$$|X_n| = |x_0| \cdot |Y_1| \cdots |Y_n|$$

und somit

$$\frac{1}{n} \cdot \log|X_n| = \frac{1}{n} \cdot \log|x_0| + \frac{1}{n} \sum_{i=1}^n \log|Y_i|\,.$$

Nun folgt die Aussage aus dem starken Gesetz der großen Zahlen, denn $(\log|Y_i|)_{i \in \mathbb{N}}$ sind unabhängig und identisch verteilt mit

$$E\big(\log|Y_i|\big) = 2 \cdot \frac{1}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}}\, dx\,,$$

und somit

$$\lim_{n\to\infty} \frac{1}{n} \cdot \log|X_n| = \frac{2}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}}\, dx \quad P\text{-f.s.}$$

Folglich,

$$|X_n| \xrightarrow{n\to\infty} 0 \quad \text{mit exp. Rate, falls } \int \dots < 0,$$

$$|X_n| \xrightarrow{n\to\infty} \infty \quad \text{mit exp. Rate, falls } \int \dots > 0.$$

Beachte, dass

$$\begin{aligned}\frac{2}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}}\, dx &\overset{y = \frac{x}{\sqrt{\beta}}}{=} \frac{2}{\sqrt{2\pi}} \int_0^\infty \log(\sqrt{\beta} y) \cdot e^{-\frac{y^2}{2}}\, dy \\ &= \frac{1}{2} \cdot \log\beta + \frac{2}{\sqrt{2\pi}} \int_0^\infty \log y \cdot e^{-\frac{y^2}{2}}\, dy \\ &< 0 \quad \Leftrightarrow \quad \beta < \beta_0.\end{aligned}$$

□

Bleibt noch zu zeigen, dass

$$-\frac{4}{\sqrt{2\pi}}\int_0^\infty \log x \cdot e^{-\frac{x^2}{2}}\,dx = \log 2 + C$$

wobei C die Euler-Mascheroni Konstante bezeichnet (Übung!).

Beispiel 3.17. Betrachte unabhängige 0-1- Experimente mit unbekannter Erfolgswahrscheinlichkeit $p \in [0,1]$. Zugehöriges kanonisches Modell:

$$S_i := \{0,1\},\ i \in \mathbb{N};\quad \Omega := \{0,1\}^{\mathbb{N}},$$
$$X_i : \Omega \to \{0,1\},\ i \in \mathbb{N},\quad \text{Projektionen},$$
$$\mu_i := p\varepsilon_1 + (1-p)\varepsilon_0,\ i \in \mathbb{N};\quad P_p := \bigotimes_{i=1}^\infty \mu_i$$

$\mathcal{A}_n$ und $\mathcal{A}$ wie oben definiert.
Für die unbekannte Wahrscheinlichkeit p wählen wir eine a priori-Verteilung μ auf $\big([0,1],\mathcal{B}([0,1])\big)$.
Behauptung: $K(p,\,\cdot\,) := P_p(\,\cdot\,)$ ist ein stochastischer Kern von $\big([0,1],\mathcal{B}([0,1])\big)$ nach $(\Omega,\mathcal{A})$.

Beweis. Es bleibt zu zeigen, dass für $A \in \mathcal{A}$ die Abbildung $p \mapsto P_p(A)$ messbar auf $[0,1]$ ist. Betrachte dazu das Mengensystem

$$\mathcal{D} := \big\{A \in \mathcal{A} \;\big|\; p \mapsto P_p(A) \text{ ist } \mathcal{B}([0,1])\text{-messbar}\big\}$$

$\mathcal{D}$ ist ein Dynkin-System und enthält alle (endlichen) Zylindermengen

$$\{X_1 = x_1, \ldots, X_n = x_n\},\quad n \in \mathbb{N},\ x_1,\ldots,x_n \in \{0,1\}\,,$$

denn

$$P_p(X_1 = x_1, \ldots, X_n = x_n) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

ist messbar (sogar stetig!) in p.
Die Behauptung folgt nun aus der Tatsache, dass das Mengensystem der endlichen Zylindermengen ein $\bigcap$-stabiler Erzeuger der σ-Algebra $\mathcal{A}$ ist. $\square$

Es sei nun $\bar{P} := \mu \otimes K$ auf $\bar{\Omega} := [0,1] \times \Omega$, versehen mit der Produkt-σ-Algebra $\mathcal{B}([0,1]) \otimes \mathcal{A}$. Aus Bemerkung 3.8 folgt, dass $\bar{P}$ als Randverteilungen zum einen die a priori-Verteilung μ und zum anderen das Wahrscheinlichkeitsmaß

$$P(\,\cdot\,) := \int P_p(\,\cdot\,)\ \mu(dp) \tag{3.7}$$

auf $(\Omega,\mathcal{A})$ besitzt. Aufgrund von (3.7) kann das Maß P als Mischung der P_p bzgl. der a priori-Verteilung μ des Erfolgsparameters p aufgefasst werden.

Beachte: Unter P sind die kanonischen Projektionen X_i nicht mehr unabhängig !

Für den Spezialfall, dass μ Gleichverteilung auf $[0,1]$, können wir nun die Anfangsverteilung $P \circ X_1^{-1}$ und die Übergangswahrscheinlichkeiten explizit angeben:

$$\begin{aligned} P \circ X_1^{-1} &= \int \big(p\varepsilon_1 + (1-p)\varepsilon_0\big)(\,\cdot\,)\ \mu(dp) \\ &= \int p\ \mu(dp) \cdot \varepsilon_1 + \int (1-p)\ \mu(dp) \cdot \varepsilon_0 = \frac{1}{2} \cdot \varepsilon_1 + \frac{1}{2} \cdot \varepsilon_0. \end{aligned}$$

Für $n \in \mathbb{N}$ und $x_1, \dots, x_n \in \{0,1\}$ mit $s_n := \sum_{i=1}^n x_i$ folgt

$$\begin{aligned} &P(X_{n+1} = 1 \,|\, X_1 = x_1, \dots, X_n = x_n) \\ &\qquad = \frac{P(X_{n+1} = 1, X_n = x_n, \dots, X_1 = x_1)}{P(X_n = x_n, \dots, X_1 = x_1)} \\ &\qquad \overset{(3.7)}{=} \frac{\displaystyle\int p^{s_n+1}(1-p)^{n-s_n}\ \mu(dp)}{\displaystyle\int p^{s_n}(1-p)^{n-s_n}\ \mu(dp)} \\ &\qquad = \frac{\Gamma(s_n+2)\Gamma(n-s_n+1)}{\Gamma(n+3)} \frac{\Gamma(n+2)}{\Gamma(s_n+1)\Gamma(n-s_n+1)} = \frac{s_n+1}{n+2} \\ &\qquad = \underbrace{\left(1 - \frac{n}{n+2}\right) \cdot \frac{1}{2} + \frac{n}{n+2} \cdot \frac{s_n}{n}}_{\text{Konvexkombination}}. \end{aligned}$$

Satz 3.18. *Es sei P ein Wahrscheinlichkeitsmaß auf dem kanonischen Modell $(\Omega, \mathcal{A})$ und*

$$\mu_n := P \circ X_n^{-1}, \quad n \in \mathbb{N}_0.$$

Dann gilt:

$$X_n,\ n \in \mathbb{N},\ \textit{unabhängig} \quad \Leftrightarrow \quad P = \bigotimes_{n=0}^{\infty} \mu_n.$$

Beweis. Es sei $\tilde{P} := \bigotimes_{n=0}^{\infty} \mu_n$. Dann gilt

$$P = \tilde{P}$$

genau dann wenn für alle $n \in \mathbb{N}_0$ und alle $A_0 \in \mathcal{S}_0, \dots, A_n \in \mathcal{S}_n$ gilt

$$\begin{aligned} &P(X_0 \in A_0, \dots, X_n \in A_n) = \tilde{P}(X_0 \in A_0, \dots, X_n \in A_n) \\ &\quad = \prod_{i=0}^n \mu_i(A_i) = \prod_{i=0}^n P(X_i \in A_i). \end{aligned}$$

Dies ist offensichtlich genau dann der Fall, wenn die kanonischen Projektionen X_n, $n \in \mathbb{N}_0$, unabhängig sind. $\square$

Definition 3.19. Es sei $S_i := S$, $i \in \mathbb{N}_0$, $(\Omega, \mathcal{A})$ das zugehörige kanonische Modell und P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. Insbesondere ist $(X_n)_{n \geqslant 0}$ ein stochastischer Prozess im Sinne von Definition 3.14. Es sei $J \subseteq \mathbb{N}_0$, $|J| < \infty$. Dann heißt die Verteilung von $(X_j)_{j \in J}$ unter P

$$\mu_J := P \circ (X_i)_{i \in J}^{-1}$$

die **endlichdimensionale Randverteilung zur Koordinatenmenge** J (auf $(S^J, \mathcal{S}^J)$).

Bemerkung 3.20. *P ist durch die Familie der endlichdimensionalen Randverteilungen*

$$\mu_{\{0,\dots,n\}}, \quad n \in \mathbb{N} \text{ eindeutig festgelegt.}$$

3.3 Bedingte Erwartungen

Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{A}_0 \subseteq \mathcal{A}$ eine Unter-σ-Algebra.

Definition 3.21. Es sei $X \geqslant 0$ eine Zufallsvariable. Eine Zufallsvariable $X_0 \geqslant 0$ heißt **(Version der) bedingten Erwartung von** X **gegeben** $\mathcal{A}_0$ falls gilt:

(a) X_0 ist $\mathcal{A}_0$-messbar.

(b) $E(Y_0 X) = E(Y_0 X_0)$ für alle $\mathcal{A}_0$-messbaren Zufallsvariablen $Y_0 \geqslant 0$.

Satz 3.22. *Es sei X und $\mathcal{A}_0$ wie oben gegeben. Dann gilt:*

(i) Eine Zufallsvariable X_0 mit den Eigenschaften (a) und (b) der vorherigen Definition existiert (siehe Unterabschnitt 3.3.2 unten).

(ii) Je zwei Zufallsvariablen mit den Eigenschaften (a) und (b) stimmen P-f.s. überein.

Bezeichnung:

$$X_0 =: E[X \,|\, \mathcal{A}_0].$$

Bemerkung. *Um zu überprüfen, ob eine nichtnegative $\mathcal{A}_0$-messbare Zufallsvariable X_0 eine (Version der) bedingten Erwartung $E[X \,|\, \mathcal{A}_0]$ ist, reicht es aus zu zeigen, dass*

$$E\left(1_{A_0} X\right) = E\left(1_{A_0} X_0\right) \qquad \forall A_0 \in \mathcal{A}_0 \,. \tag{3.8}$$

In der Tat folgt mithilfe einer maßtheoretischen Induktion unmittelbar, dass (3.8) *äquivalent ist zur Eigenschaft (b) der vorherigen Definition, d.h. $E(Y_0 X) = E(Y_0 X_0)$ für alle $\mathcal{A}_0$-messbaren Zufallsvariablen $Y_0 \geq 0$.*

Bemerkung 3.23. *(i)* ***Extremfälle***

$$E\big(X \;\big|\; \{\emptyset, \Omega\}\big) = E(X) \qquad \text{und} \qquad E(X \,|\, \mathcal{A}) = X$$

(ii) Es sei X nicht notwendigerweise nichtnegativ. Betrachte dann die Zerlegung $X = X^+ - X^-$ in Positiv- und Negativteil.

Gilt dann

$$\min\bigl(E(X^+ \,|\, \mathcal{A}_0)\,,\ E(X^- \,|\, \mathcal{A}_0)\bigr) < \infty \quad P-\textit{f.s.}$$

so können wir definieren:

$$E(X \,|\, \mathcal{A}_0) := E(X^+ \,|\, \mathcal{A}_0) - E(X^- \,|\, \mathcal{A}_0)\,.$$

Beachte, dass

$$\begin{aligned} X \in \mathcal{L}^1 \quad &\Leftrightarrow \quad E(X \,|\, \mathcal{A}_0) \in \mathcal{L}^1 \\ &\Leftrightarrow \quad \begin{cases} E(X^+) = E\bigl(E(X^+ \mid \mathcal{A}_0)\bigr) < \infty \\ E(X^-) = E\bigl(E(X^- \mid \mathcal{A}_0)\bigr) < \infty\,. \end{cases} \end{aligned}$$

(iii) Für beliebiges $A \in \mathcal{A}$ sei

$$P(A \,|\, \mathcal{A}_0) := E(1_A \,|\, \mathcal{A}_0)\,.$$

*$P(A \,|\, \mathcal{A}_0)$ heißt **bedingte Wahrscheinlichkeit (des Ereignisses A) gegeben** $\mathcal{A}_0$.*

*(iv) **diskreter Fall**: Es seien $B_i \in \mathcal{A}$, $i \in \mathbb{N}$ paarweise disjunkt mit $\Omega = \bigcup_{i\in\mathbb{N}} B_i$, so dass $\mathcal{A}_0 = \sigma\{B_i \,|\, i \in \mathbb{N}\}$. Dann gilt für eine beliebige nichtnegative Zufallsvariable $X \geqslant 0$:*

$$E(X \,|\, \mathcal{A}_0) = \sum_{i\in\mathbb{N}:P(B_i)>0} E(X \,|\, B_i) \cdot 1_{B_i},$$

wobei

$$E(X \,|\, B_i) = \frac{1}{P(B_i)} \int_{B_i} X\, dP$$

***die elementare bedingte Erwartung von X bzgl. des Ereignisses B_i** bezeichnet.*

(v) Es sei $(\Omega', \mathcal{A}')$ ein messbarer Raum und $Y : \Omega \to \Omega'$ $\mathcal{A}/\mathcal{A}'$-messbar. Es sei $\mathcal{A}_0 := \sigma(Y)$ und $X \geqslant 0$ Zufallsvariable auf Ω. Das Faktorisierungslemma impliziert die Existenz einer Funktion $f_X : \Omega' \to \bar{\mathbb{R}}_+$, so dass $P - f.s.$

$$E(X \,|\, Y) := E\bigl(X \bigm| \sigma(Y)\bigr) = f_X \circ Y$$

Notation:

$$E(X \,|\, Y = \omega') := f_X(\omega') \qquad \omega' \in \Omega'\,.$$

$$
\begin{tikzcd}
& \begin{array}{l}(\Omega,\mathcal{A})\\ \mathcal{A}_0 := \sigma(Y)\subset\mathcal{A}\end{array} \arrow[r, "Y"] \arrow[d, "{E[X|Y]}"] & (\Omega',\mathcal{A}') \arrow[dl, "f_X"] \\
(\Omega,\mathcal{A}) \arrow[r, "X"] & \mathbb{R}_+ &
\end{tikzcd}
$$

Insbesondere folgt, $Y^{-1}(A') \in \mathcal{A}_0 = \sigma(Y)$ *für alle* $A' \in \mathcal{A}'$ *und*

$$
\begin{aligned}
\int_{Y^{-1}(A')} X\,dP &\overset{3.21\,(ii)}{=} \int_{Y^{-1}(A')} f_X(Y)\,dP = \int_\Omega 1_{A'}(Y)\,f_X(Y)\,dP \\
&= \int_{A'} f_X\,d(P\circ Y^{-1}).
\end{aligned}
$$

Somit ist, f_X $P\circ Y^{-1}$*-f.s. eindeutig bestimmt.*

3.3.1 Eigenschaften der bedingten Erwartung

(a) Linearität und Monotonie

$$E(c_1X_1 + c_2X_2 \,|\, \mathcal{A}_0) = c_1E(X_1 \,|\, \mathcal{A}_0) + c_2E(X_2 \,|\, \mathcal{A}_0)$$

$$X \leqslant Y \quad P-f.s. \quad \Rightarrow \quad E(X \,|\, \mathcal{A}_0) \leqslant E[Y \,|\, \mathcal{A}_0]\,.$$

(b) Konvergenzsätze B. Levi, monotone Konvergenz

$$0 \leqslant X_1 \leqslant X_2 \leqslant \dots \quad P-f.s. \quad \Rightarrow \quad E\big(\lim_{n\to\infty} X_n \;\big|\; \mathcal{A}_0\big) = \lim_{n\to\infty} E(X_n \,|\, \mathcal{A}_0)\,.$$

Fatou

$$X_n \geq 0 \quad \forall\, n \in \mathbb{N} \quad \Rightarrow \quad E\big(\liminf_{n\to\infty} X_n \;\big|\; \mathcal{A}_0\big) \leqslant \liminf_{n\to\infty} E(X_n \,|\, \mathcal{A}_0)\,.$$

Lebesgue, majorisierte Konvergenz

$|X_n| \leqslant Y \in \mathcal{L}^1$ für alle $n \in \mathbb{N}$ und $X_n \to X$ P-f.s.
(oder (X_n) gleichgradig integrierbar und $X_n \to X$ P-stochastisch).
Dann gilt

$$E\big(\lim_{n\to\infty} X_n \;\big|\; \mathcal{A}_0\big) = \lim_{n\to\infty} E(X_n \,|\, \mathcal{A}_0)\,.$$

(c) Kontraktionseigenschaften **Jensensche Ungleichung** $X \in \mathcal{L}^1$ und u konkave (!) Funktion auf $\mathbb{R}$. Dann gilt

$$E\big(u(X) \bigm| \mathcal{A}_0\big) \leqslant u\big(E(X \,|\, \mathcal{A}_0)\big) \quad P-\text{f.s.}$$

Kontraktion auf $\mathcal{L}^p$
Für $p \geqslant 1$ und $X \in \mathcal{L}^p$ gilt

$$\|E(X \,|\, \mathcal{A}_0)\|_p \leqslant \|X\|_p.$$

Insbesondere ist die Abbildung

$$X \mapsto E(X \,|\, \mathcal{A}_0)$$

stetig auf $\big(L^p, \|\cdot\|_p\big)$

(d) Glättungseigenschaften Es sei $X \geqslant 0$ und $Y_0 \geqslant 0$ $\mathcal{A}_0$-messbar. Dann folgt

$$E(Y_0 X \,|\, \mathcal{A}_0) = Y_0 \cdot E(X \,|\, \mathcal{A}_0) \quad P-\text{f.s.}$$

Projektivität (engl. tower property)
Insbesondere folgt für zwei Unter-σ-Algebren und $\mathcal{A}_0, \mathcal{A}_1$ mit $\mathcal{A}_0 \subseteq \mathcal{A}_1 \subseteq \mathcal{A}$, dass

$$E\big(E(X \,|\, \mathcal{A}_1) \bigm| \mathcal{A}_0\big) = E\big(E(X \,|\, \mathcal{A}_0) \bigm| \mathcal{A}_1\big) = E(X \,|\, \mathcal{A}_0) \quad P-\text{f.s.}$$

(e) Bedingte Erwartung und Unabhängigkeit

Satz 3.24. *Es seien $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{A}$ Unter-σ-Algebren und $X \in \mathcal{L}^1$. Es sei $\sigma(\mathcal{A}_1, \mathcal{A}_2)$ (bzw. $\sigma(\mathcal{A}_1, X)$) die von $\mathcal{A}_1 \cup \mathcal{A}_2$ (bzw. $\mathcal{A}_1 \cup \sigma(X)$) erzeugte σ - Algebra. Dann gilt: Ist $\sigma(\mathcal{A}_1, X)$ unabhängig von $\mathcal{A}_2$, so folgt*

$$E\big(X \bigm| \sigma(\mathcal{A}_1, \mathcal{A}_2)\big) = E(X \,|\, \mathcal{A}_1) \quad P\text{-f.s.}$$

Insbesondere gilt

$$X \text{ unabhängig von } \mathcal{A}_0 \quad \Rightarrow \quad E(X \,|\, \mathcal{A}_0) = E(X)\,.$$

Der Beweis des Satzes folgt aus dem folgenden Satz:

Satz 3.25. *Es seien $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{A}$ Unter-σ-Algebren und $X \in \mathcal{L}^1$, $X \geq 0$. Dann sind die folgenden Aussagen äquivalent:*

(i) $E\big(X \bigm| \sigma(\mathcal{A}_1, \mathcal{A}_2)\big) = E(X \,|\, \mathcal{A}_1)$.

(ii) $E(XY \,|\, \mathcal{A}_1) = E(X \,|\, \mathcal{A}_1) E(Y \,|\, \mathcal{A}_1)$ *für alle* $Y \geq 0$ $\sigma(\mathcal{A}_1, \mathcal{A}_2)$*-messbar.*

(iii) $E(X X_2 \,|\, \mathcal{A}_1) = E(X \,|\, \mathcal{A}_1) E(X_2 \,|\, \mathcal{A}_1)$ *für alle* $X_2 \geq 0$ $\mathcal{A}_2$*-messbar.*

Beweis. Übung □

Beispiel 3.26 (Markovkette). Es sei P_μ die Verteilung einer Markovkette $X_0, X_1, \dots$ auf $\Omega = S^{\{0,1,\dots\}}$ mit Anfangsverteilung μ und Übergangswahrscheinlichkeiten $p(x, dy)$ auf $(S, \mathcal{S})$. Es sei $\mathcal{A}_n := \sigma(X_0, \dots, X_n)$, $\hat{\mathcal{A}}_n := \sigma(X_i \,|\, i \geqslant n)$ und ϑ^n der **Shift** um n Schritte, d.h.

$$\vartheta^n((x_0, x_1, x_2, \dots)) = (x_n, x_{n+1}, \dots)$$

also insbesondere $X_k \circ \vartheta^n = X_{n+k}$. Dann folgt aus der Markoveigenschaft, angewandt zur Zeit n, insbesondere

$$E_\mu(\psi \circ \vartheta^n \,|\, \mathcal{A}_n) = E_{X_n}(\psi) \tag{3.9}$$

für alle $\psi \geq 0$ $\mathcal{A}$-messbar. Somit gilt für alle $\hat{\mathcal{A}}_n$-messbaren $X \geqslant 0$, dass

$$E_\mu(X \,|\, \mathcal{A}_n) = E_\mu(X \,|\, X_n)\,.$$

Nach Satz 3.25 ist dies äquivalent dazu, dass für alle $\mathcal{A}_n$-messbaren Zufallsvariablen $Y \geqslant 0$ gilt.

$$E_\mu(YX \,|\, X_n) = E_\mu(Y \,|\, X_n) E_\mu(X \,|\, X_n)\,,$$

d.h. also, gegeben die Gegenwart $\sigma(X_n)$, ist die Zukunft $\hat{\mathcal{A}}_n$ unabhängig von der Vergangenheit $\mathcal{A}_n$ (Elementare Markoveigenschaft!)

(f) Beste $\mathcal{L}^2$ Approximation Es sei $X \in \mathcal{L}^2$. Dann gilt

$$E\Big(\big(X - E(X \,|\, \mathcal{A}_0)\big)^2\Big) \leqslant E\big((X - Y_0)^2\big)$$

für alle $Y_0 \in \mathcal{L}^2$, $\mathcal{A}_0$-messbar.

3.3.2 Existenz

(a) Hilbertraummethode. Es sei $L^2 := \mathcal{L}^2/\sim$, mit der Äquivalenzrelation $X \sim Y$ genau dann wenn $X = Y$ P-f.s.. Für gegebenes $X \in \mathcal{L}^2$, bezeichne $\bar{X}$ die zugehörige Äquivalenzklasse d.h. $Y \in \bar{X}$ genau dann wenn $X = Y$ P-f.s. Jedes $Y \in \bar{X}$ heißt Vertreter der Äquivalenzklasse $\bar{X}$. Für zwei Äquivalenzklassen $\bar{X}, \bar{Y} \in L^2$ definieren wir ihr Skalarprodukt durch

$$(\bar{X}, \bar{Y}) := E(XY)\,,$$

wobei X (bzw. Y) Vertreter von $\bar{X}$ (bzw. $\bar{Y}$) sind. Dann ist $\big(L^2, (\ ,\ )\big)$ ein **Hilbertraum**.

Es sei nun $\mathcal{L}_0^2 := \mathcal{L}^2(\Omega, \mathcal{A}_0, P) (\subseteq \mathcal{L}^2)$ der lineare Unterraum der quadratintegrierbaren Zufallsvariablen, die messbar bzgl. der kleineren Unter-σ-Algebra $\mathcal{A}_0$ sind, und es sei $L_0^2 := \mathcal{L}_0^2/\sim$. Dann ist L_0^2 ein abgeschlossener (!) Unterraum von L^2 (folgt aus dem Satz von Riesz-Fischer (Satz 1.69), denn jede L^2-Cauchy-Folge X_1 besitzt eine $P-$f.s. -konvergente Teilfolge, so dass aus X_n $\mathcal{A}_0$-messbar für alle n folgt, dass auch der L^2-Grenzwert $\mathcal{A}_0$-messbar ist).

Aus der Eigenschaft (f) der bedingten Erwartung folgt, dass $E(X \mid \mathcal{A}_0)$ für $X \in \mathcal{L}^2$ ein Vertreter der orthogonalen Projektion $\bar{\pi}(\bar{X})$ von $\bar{X} \in L^2$ auf L_0^2 ist. Wenn wir also die Existenz der orthogonalen Projektion in $E(X \,|\, \mathcal{A}_0)$ ausnutzen, können wir folgendermaßen die bedingte Erwartung $E(X \,|\, \mathcal{A}_0)$ definieren:

Schritt 1: Für $X \in \mathcal{L}^2$ definiere

$$X_0 := \pi(\bar{X}) \quad \big(:= \mathcal{A}_0\text{-messbarer Repräsentant von } \bar{\pi}(\bar{X})\big)$$

Es folgt für alle $Y_0 \in \mathcal{L}_0^2$, dass

$$\begin{aligned} E([Y_0 X) &= E(Y_0 X_0) + E\big(Y_0(X - X_0)\big) \\ &= E(Y_0 X_0) + \underbrace{(\bar{Y}_0, \overline{X - X_0})}_{=0} . \end{aligned} \tag{3.10}$$

Also

$$E(Y_0 X) = E(Y_0 X_0)$$

für alle $Y_0 \geq 0$ $\mathcal{A}_0$-messbar, so dass

$$X_0 = E(X \,|\, \mathcal{A}_0) .$$

Wie im letzten Abschnitt, zeigt man wieder die Monotonie:

$$X \leqslant Y \quad P\text{-f.s.} \quad \Rightarrow \quad \pi(\bar{X}) \leqslant \pi(\bar{Y}) \quad P\text{-f.s.}$$

Schritt 2: Für allgemeine $X \geqslant 0$, nicht notwendigerweise in $\mathcal{L}^2$, betrachte man $X \wedge n \in \mathcal{L}^2$. Aus der Monotonie folgt, dass

$$Z_0 := \lim_{n\to\infty} \pi(\overline{X \wedge n})$$

P-f.s. konvergiert (im uneigentlichen Sinne) ($\mathcal{A}_0$-messbar!). Aus dem Satz von der monotonen Konvergenz ergibt sich für $Y_0 \geqslant 0$

$$E(Y_0 Z_0) = \lim_{n\to\infty} E\big(Y_0 \pi(\overline{X \wedge n})\big) = \lim_{n\to\infty} E(Y_0 X \wedge n) = E(Y_0 \cdot X) .$$

Folglich ist $Z_0 = E(X \,|\, \mathcal{A}_0)$ eine Version der bedingten Erwartung.

(b) Radon-Nikodym Theorem Im ganzen Unterabschnitt sei $(\Omega, \mathcal{A})$ ein messbarer Raum.

Definition 3.27. Es seien μ, ν zwei σ-endliche Maße auf $(\Omega, \mathcal{A})$. Dann heißt ν **absolutstetig bzgl.** μ ($\nu \ll \mu$), falls

$$\mu(N) = 0 \,, N \in \mathcal{A} \quad \Rightarrow \quad \nu(N) = 0 .$$

M.a.W.: Jede μ-Nullmenge ist auch eine ν-Nullmenge (aber nicht notwendigerweise umgekehrt!).

Beispiel 3.28. Es sei μ ein σ-endliches Maß auf Ω und $f \in \mathcal{L}^1$, $f \geq 0$. Definiere hierzu das endliche Maß

$$\nu(A) := \int_A f\, d\mu := \int 1_A f\, d\mu\,, \quad A \in \mathcal{A}\,. \tag{3.11}$$

Dann gilt $\nu \ll \mu$.
Der Satz von Radon-Nikodym (siehe Satz 3.30) besagt nun umgekehrt, dass zu $\nu \ll \mu$ eine $\mathcal{A}$-messbare nichtnegative Funktion $f : \Omega \to \mathbb{R}_+$ existiert, so dass (3.11) gilt. f ist μ-f.s. eindeutig bestimmt und heißt **Dichte** von ν bzgl. μ (Bezeichnung: $\frac{d\nu}{d\mu}$).

Mithilfe des Satzes von Radon-Nikodym erhalten wir insbesondere einen zweiten, unabhängigen Beweis der Existenz der bedingten Erwartung. Wir werden diesen Beweis für endliche Maße beweisen.

Lemma 3.29. *Es seien σ und τ zwei endliche (positive) Maße auf einem messbaren Raum $(\Omega, \mathcal{A})$ mit $\sigma(\Omega) < \tau(\Omega)$. Dann gibt es eine messbare Menge $\Omega' \in \mathcal{A}$ mit folgenden Eigenschaften:*

(i) $\sigma(\Omega') < \tau(\Omega')$.

(ii) $\sigma(A) \leqslant \tau(A)$ *für alle* $A \in \Omega' \cap \mathcal{A} := \{A \subset \Omega' \mid A \in \mathcal{A}\}$.

Beweis. (i) Es sei $\delta := \tau - \sigma$ (d.h., $\delta(A) := \tau(A) - \sigma(A)$ für alle $A \in \mathcal{A}$). δ ist beschränkt auf $\mathcal{A}$, denn

$$-\sigma(\Omega) \leqslant \delta(A) \leqslant \tau(\Omega)\,.$$

Wir definieren induktiv zwei Folgen

$$(A_n)_{n \in \mathbb{N} \cup \{0\}}\,, \quad (\Omega_n)_{n \in \mathbb{N} \cup \{0\}}$$

durch:

$A_0 := \emptyset$ und $\Omega_0 := \Omega \setminus A_0$ $(= \Omega)$. Im Induktionsschritt seien $A_0, \ldots, A_n$ und $\Omega_0, \ldots, \Omega_n$ bereits konstruiert. Dann gilt

$$\alpha_n := \inf_{A \in \Omega_n \cap \mathcal{A}} \delta(A) \leqslant 0$$

denn $\emptyset \in \Omega_n \cap \mathcal{A}$ und somit $\alpha_n \leqslant \delta(\emptyset) = 0$.
Falls $\alpha_n = 0$, so setze $A_{n+1} := \emptyset$ und $\Omega_{n+1} := \Omega_n \setminus A_{n+1}$ $(= \Omega_n)$.

Falls $\alpha_n < 0$, so wähle $A_{n+1} \in \Omega_n \cap \mathcal{A}$ mit $\delta(A_{n+1}) \leqslant \frac{\alpha_n}{2}$ und setze $\Omega_{n+1} := \Omega_n \setminus A_{n+1}$.

Dann sind A_n, $n \in \mathbb{N}$, paarweise disjunkt und somit ist

$$\sum_{n=0}^{\infty} \delta(A_n) \quad \Big(= \tau\Big(\bigcup_{n \geq 0} A_n\Big) - \sigma\Big(\bigcup_{n \geq 0} A_n\Big)\Big)$$

konvergent. Insbesondere ergibt sich $\lim_{n\to\infty} \delta(A_n) = 0$ und somit $\lim_{n\to\infty} \alpha_n = 0$. Setze nun

$$\Omega' := \bigcap_{n\geqslant 0} \Omega_n .$$

Da (Ω_n) absteigend ist, folgt

$$\delta(\Omega') = \lim_{n\to\infty} \tau(\Omega_n) - \lim_{n\to\infty} \sigma(\Omega_n) = \lim_{n\to\infty} \delta(\Omega_n) > \delta(\Omega)\,,$$

denn

$$\delta(\Omega_{n+1}) \geq \delta(\Omega_n) - \delta(A_{n+1}) \geq \delta(\Omega_n) \geq \delta(\Omega_0) = \delta(\Omega)\,.$$

Hieraus folgt (i).

(ii) Es sei $A \in \Omega' \cap \mathcal{A}$. Dann gilt $A \in \Omega_n \cap \mathcal{A}$ für alle n, somit $\delta(A) \geqslant \alpha_n$ für alle n, und damit $\delta(A) \geqslant \lim_{n\to\infty} \alpha_n = 0$. □

Satz 3.30 (Radon-Nikodym). *Es seien μ und ν endliche (positive) Maße auf dem messbaren Raum $(\Omega, \mathcal{A})$. Dann sind folgende Aussagen äquivalent:*

(i) Es existiert eine $\mathcal{A}$-messbare Funktion $f \geq 0$ (μ-f.s. eindeutig bestimmt) mit $\nu = f \cdot \mu$ (d.h., $\nu(A) = \int_A f\, d\mu$ für alle $A \in \mathcal{A}$).

(ii) $\nu \ll \mu$ (d.h., $\mu(N) = 0$ für $N \in \mathcal{A}$ impliziert $\nu(N) = 0$).

Beweis. (i)⇒(ii) ist offensichtlich.
(ii)⇒(i)
Es sei G die Familie aller $\mathcal{A}$-messbaren $\overline{\mathbb{R}}$-wertigen Funktionen $g \geqslant 0$ auf Ω mit

$$g\mu \leqslant \nu,$$

d.h., $\nu(A) \geqslant \int_A g\, d\mu$ für alle $A \in \mathcal{A}$. Beachte, dass $g \equiv 0 \in G$, und dass G abgeschlossen ist unter Suprema, denn für $g, h \in G$ folgt

$$\begin{aligned}\int_A \sup(g,h)\, d\mu &= \int_{A\cap\{g\geqslant h\}} g\, d\mu + \int_{A\cap\{g<h\}} h\, d\mu \\ &\leqslant \nu\big(A \cap \{g \geqslant h\}\big) + \nu\big(A \cap \{g < h\}\big) = \nu(A) \qquad \forall\, A \in \mathcal{A}\,,\end{aligned}$$

also $\sup(g,k) \in G$. Es sei

$$\gamma := \sup_{g\in G} \int g\, d\mu \quad \big(\leqslant \nu(\Omega) < \infty\big).$$

Da G $\sup$-stabil ist, gibt es eine monoton wachsende Folge (g_n) von Funktionen aus G mit

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

Hierbei haben wir den Satz von der monotonen Integration verwendet.
Es sei $f := \lim\limits_{n\to\infty} g_n$. Dann gilt

$$\int_A f \, d\mu = \lim_{n\to\infty} \int_A g_n \, d\mu \leqslant \nu(A) \quad \forall\, A \in \mathcal{A}\,.$$

Folglich ist, $f \in G$ und somit ist f ein Maximum von

$$g \mapsto \int g \, d\mu \text{ auf } G\,.$$

Wir zeigen als nächstes, dass $f\mu = \nu$. Offensichtlich gilt $f\mu \leqslant \nu$, da $f \in G$. Definiere nun

$$\tau := \nu - f\mu\,. \tag{3.12}$$

τ ist ein endliches positives Maß auf $\mathcal{A}$ und zu zeigen ist dass $\tau \equiv 0$.
Nehme umgekehrt an, es gelte $\tau(\Omega) > 0$ (also insbesondere $\mu(\Omega) > 0$) und setze

$$\beta := \frac{1}{2} \cdot \frac{\tau(\Omega)}{\mu(\Omega)} > 0.$$

Dann gilt

$$\tau(\Omega) = 2\beta \cdot \mu(\Omega) > \beta \cdot \mu(\Omega)\,.$$

Aufgrund des vorangegangenen Lemmas existiert eine Menge $\Omega' \in \mathcal{A}$ mit

$$\tau(\Omega') > \beta \cdot \mu(\Omega') \quad \text{und} \quad \tau(A) \geqslant \beta \cdot \mu(A) \tag{3.13}$$

für alle $A \in \Omega' \cap \mathcal{A}$. Definiere $f_0 := f + \beta \cdot 1_{\Omega'}$. Dann ist f_0 $\mathcal{A}$-messbar und für alle $A \in \mathcal{A}$ gilt

$$\int_A f_0 \, d\mu = \int_A f \, d\mu + \beta \cdot \mu(A \cap \Omega') \leqslant \int_A f \, d\mu + \tau(A) = \nu(A).$$

Insbesondere folgt $f_0 \in G$. Andererseits ist

$$\int f_0 \, d\mu = \int f \, d\mu + \beta \cdot \mu(\Omega') = \gamma + \beta \cdot \mu(\Omega') > \gamma\,,$$

wobei wir verwendet haben, dass aus $\nu \ll \mu$ insbesondere $\mu(\Omega') > 0$ folgt. $\int f_0 d\mu > \gamma$ steht nun aber im Widerspruch dazu, dass f ein Maximum von

$$g \mapsto \int g \, d\mu$$

auf G ist. Es muss also notwendigerweise $\tau \equiv 0$ gelten. $\square$

Bemerkung 3.31. *(i) Es seien μ, ν endliche Maße, $\nu \ll \mu$ und $\frac{d\nu}{d\mu}$ sei die zugehörige Dichte. Dann gilt*

$$\nu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$$

aber im allgemeinen nicht $\mu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$.
(ii) Es seien μ und ν endliche Maße. μ und ν heißen **äquivalent** *(Notation: $\mu \sim \nu$), falls $\mu \ll \nu$ und $\nu \ll \mu$. Es ist leicht einzusehen, dass $\mu \sim \nu$ genau dann wenn $\nu \ll \mu$, also $\nu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$, und zusätzlich $\mu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$. In diesem Fall gilt*

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$

(iii) Es seien ν, μ und λ endliche Maße, $\nu \ll \mu$ und $\mu \ll \lambda$. Dann folgt

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda} \quad \lambda - \textit{f.ü.}$$

Anwendung auf die Konstruktion der bedingten Erwartung

Es sei $\mathcal{A}_0 \subseteq \mathcal{A}$ Unter-σ-Algebra, $X \geq 0$, $X \in \mathcal{L}^1(P)$. Dann ist

$$Q(A) := \int_A X\,dP = \int 1_A X\,dP\,, \quad A \in \mathcal{A}_0$$

ein endliches Maß auf $(\Omega, \mathcal{A}_0)$. Offensichtlich ist $Q \ll P_{|\mathcal{A}_0}$, und somit existiert nach dem Satz von Radon-Nikodym die Dichte

$$\exists \quad X_0 := \frac{dQ}{dP_{|\mathcal{A}_0}} \qquad \mathcal{A}_0 \text{ -messbar}\,.$$

Für $A \in \mathcal{A}_0$ gilt nach Definition der Dichte,

$$\int 1_A X\,dP = Q(A) = \int 1_A \frac{dQ}{dP}\,dP = \int 1_A X_0\,dP\,. \tag{3.14}$$

Mithilfe einer maßtheoretischen Induktion zeigt man nun dass (3.14)

(a) aufgrund der Linearität für alle einfachen Zufallsvariablen $Y_0 = \sum_{k=1}^n a_k 1_{A_k}$ gilt,

(b) mithilfe der Approximation durch monoton wachsende, punktweise konvergente Folgen einfacher Zufallsvariablen, auch für alle $\mathcal{A}_0$-messbaren $Y_0 \geqslant 0$ gilt.

Folglich ist $X_0 = \frac{dQ}{dP_{|\mathcal{A}_0}}$ eine Version der bedingten Erwartung $E(X|\mathcal{A}_0)$. Für allgemeine $X \geq 0$ betrachte nun wieder die monotone Approximation $X \wedge n \uparrow X$.

3.3.3 Reguläre bedingte Wahrscheinlichkeiten

Betrachte die Abbildung

$$A \mapsto P(A \,|\, \mathcal{A}_0) \quad \big(:= E(1_A \,|\, \mathcal{A}_0)\big).$$

Die folgenden Eigenschaften ergeben sich aus den allgemeinen Eigenschaften der bedingten Erwartung:

- $0 \leqslant P(A \,|\, \mathcal{A}_0) \leqslant 1$ P-f.s.
- $P(\emptyset \,|\, \mathcal{A}_0) = 0$ und $P(\Omega \,|\, \mathcal{A}_0) = 1$ P-f.s.
- $A_1 \subseteq A_2$ impliziert $P(A_1 \,|\, \mathcal{A}_0) \leqslant P(A_2 \,|\, \mathcal{A}_0)$ P-f.s.
- A_n, $n \in \mathbb{N}$, paarweise disjunkt
$$\Rightarrow \quad P\Big(\bigcup_{n=1}^{\infty} A_n \,\Big|\, \mathcal{A}_0\Big) = \sum_{n=1}^{\infty} P(A_n \,|\, \mathcal{A}_0) \ P\text{-f.s.}$$

Dies reicht im allgemeinen jedoch noch nicht aus zu folgern, dass

$$A \mapsto P(A \,|\, \mathcal{A}_0)(\omega), \quad A \in \mathcal{A}, \tag{3.15}$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$ für P-f.a. $\omega \in \Omega$.
Ist Ω jedoch ein diskreter Raum, so ist dies richtig, und wir wollen uns im folgenden fragen, unter welchen Bedingungen dies auch im allgemeinen Fall gilt, d.h. unter welchen Annahmen man gute Versionen von $P(A \,|\, \mathcal{A}_0)$, $A \in \mathcal{A}$, derart auswählen kann, so dass (3.15) in der Tat ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ definiert, zumindest für P-f.a. $\omega \in \Omega$.

Definition 3.32. Ein messbarer Raum $(\Omega, \mathcal{A})$ heißt **Borel-Raum**, falls eine Borelmenge $U \in \mathcal{B}(\mathbb{R})$ und eine (mengentheoretische) Bijektion $\varphi : \Omega \to U$ existieren, so dass φ und φ^{-1} beide messbar sind.

Satz 3.33. *Es sei $(\Omega, \mathcal{A})$ ein* ***Borel-Raum*** *und P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. Weiter sei $\mathcal{A}_0 \subseteq \mathcal{A}$ eine Unter-σ-Algebra. Dann existiert ein stochastischer Kern $K_{\mathcal{A}_0}$ von $(\Omega, \mathcal{A}_0)$ nach $(\Omega, \mathcal{A})$, so dass für alle $A \in \mathcal{A}$ gilt:*

$$K_{\mathcal{A}_0}(\omega, A) = P(A \,|\, \mathcal{A}_0)(\omega) \quad P - f.s.$$

M.a.W. : $K_{\mathcal{A}_0}(\cdot, A)$ ist eine Version der bedingten Wahrscheinlichkeit $P(A \,|\, \mathcal{A}_0)$ für alle $A \in \mathcal{A}$. $K_{\mathcal{A}_0}$ heißt ***(eine) reguläre bedingte Wahrscheinlichkeit gegeben*** *$\mathcal{A}_0$.*

$K_{\mathcal{A}_0}$ ist im folgenden Sinne eindeutig bestimmt: Ist $\tilde{K}_{\mathcal{A}_0}$ ein zweiter stochastischer Kern von $(\Omega, \mathcal{A}_0)$ nach $(\Omega, \mathcal{A})$ mit obigen Eigenschaften, so gibt es eine P-Nullmenge $N \in \mathcal{A}$, so dass für alle $\omega \in \Omega \setminus N$ und alle $A \in \mathcal{A}$

$$K_{\mathcal{A}_0}(\omega, A) = \tilde{K}_{\mathcal{A}_0}(\omega, A).$$

Einen Beweis des Satzes findet man in [Kl08], Satz 8.36.

4 Martingaltheorie

Eine wichtige Struktur in der Analysis stochastischer Prozesse sind Martingale, die als Verallgemeinerungen von Summen unabhängiger Zufallsvariablen angesehen werden können. Im folgenden sei

- $(\Omega, \mathcal{A}, P)$ - Wahrscheinlichkeitsraum
- I eine diskrete Indexmenge mit (partieller) Ordnung $\leq$, z.B. $I \subseteq \mathbb{N} \cup \{0\}$ mit der gewöhnlichen Ordnung
- $(\mathcal{A}_t)_{t\in I}$ - **Filtration auf** $(\Omega, \mathcal{A})$, d.h., $\mathcal{A}_t \subseteq \mathcal{A}$ σ-Algebra und $\mathcal{A}_s \subseteq \mathcal{A}_t$ für $s \leq t$

Wir interpretieren $\mathcal{A}_t$ als verfügbare Information zur Zeit t.

Definition 4.1. Ein Familie integrierbarer, $(\mathcal{A}_t)$-adaptierter Zufallsvariablen $(X_t)_{t\in I}$, d.h. $(X_t) \subseteq \mathcal{L}^1(P)$ und X_t $\mathcal{A}_t$-messbar für alle t, heißt

(i) **Martingal**, falls

$$X_s = E(X_t|\mathcal{A}_s) \quad \forall s,t \in I \text{ mit } s \leqslant t$$

(ii) **Submartingal**, falls

$$X_s \leqslant E(X_t|\mathcal{A}_s) \quad \forall s,t \in I \text{ mit } s \leqslant t \quad \text{("im Mittel aufsteigend")}$$

(iii) **Supermartigal**, falls

$$X_s \geqslant E(X_t|\mathcal{A}_s) \quad \forall s,t \in I \text{ mit } s \leqslant t \quad \text{("im Mittel absteigend")}$$

Man beachte, dass zur Martingaleigenschaft immer auch die Filtration $(\mathcal{A}_t)_{t\in I}$ gehört. Deshalb spricht man mitunter auch von $(\mathcal{A}_t)$-Martingal (bzw. $(\mathcal{A}_t)$-Submartingal bzw. $(\mathcal{A}_t)$-Supermartingal). Ist die Martingaleigenschaft für die Filtration $(\mathcal{A}_t)$ erfüllt, so gilt sie auch für **jede gröbere** Filtration $(\mathcal{B}_t)_{t\in I}$, bzgl. der (X_t) adaptiert ist, denn für $s, t \in I$ mit $s \leq t$ gilt wegen $\mathcal{B}_s \subseteq \mathcal{A}_s$ und der $(\mathcal{A}_t)$-Martingaleigenschaft

$$E(X_t \mid \mathcal{B}_s) = E(E(X_t \mid \mathcal{A}_s) \mid \mathcal{B}_s) = E(X_s \mid \mathcal{B}_s) = X_s\,.$$

Die entsprechenden Aussagen für Sub- bzw. Supermartingale gelten ebenfalls.

Beispiel 4.2. (1) **"sukzessive Prognosen "** Sei $X \in \mathcal{L}^1(P)$. Dann definiert

$$X_t := E(X|\mathcal{A}_t),\ t \in I,$$

ein $(\mathcal{A}_t)$-Martingal.

Beweis: Für $s, t \in I$ mit $s \leqslant t$ gilt:
$E(X_t|\mathcal{A}_s) = E(E(X|\mathcal{A}_t)|\mathcal{A}_s) = E(X|\mathcal{A}_s) = X_s$ aufgrund der Projektivität der bedingten Erwartung. □

Wir werden später im Zusammenhang mit den Martingalkonvergenzsätzen sehen, dass jedes gleichgradig integrierbare, $\mathcal{L}^1$-beschränkte $(\mathcal{A}_t)$-Martingal von dieser Form ist.

(2) **"Zentrierte Summen"** Es seien $Y_n \in \mathcal{L}^1(P)$ unabhängig und $\mathcal{A}_n = \sigma(Y_1, \ldots, Y_n)$. Dann definiert

$$X_n := \sum_{k=1}^{n} (Y_k - E(Y_k)),\ n \geqslant 0,$$

ein $(\mathcal{A}_n)$-Martingal.

Beweis: $E(X_{n+1}|\mathcal{A}_n) = E(Y_{n+1} - E(Y_{n+1})|\mathcal{A}_n) + E(X_n|\mathcal{A}_n) = X_n$, da Y_{n+1} unabhängig von $\mathcal{A}_n$ und damit $E(Y_{n+1} \mid \mathcal{A}_n) = E(Y_{n+1})$.

(3) **"Martingale einer Markovkette"** Es sei $(S, \mathcal{S})$ ein messbarer Raum, P_μ Verteilung einer Markovkette auf $\Omega := S^{\{0,1,2,\ldots\}}$ mit Startverteilung μ und Übergangswahrscheinlichkeiten $p(x, dy)$. Es sei $f : S \to \mathbb{R}_+$, $\mathcal{S}$-messbar. Setze

$$pf(x) := \int p(x, dy) f(y)\,, x \in S\,.$$

Weiterhin sei X_n die Projektion auf die n-te Koordinate, also die Position der Markovkette zur Zeit n, und $\mathcal{A}_n := \sigma\{X_0, X_1, X_2, \ldots, X_n\}$. Dann folgt aus der Markoveigenschaft

$$E_\mu(f(X_{n+1} \mid \mathcal{A}_n) = E_\mu(f(X_{n+1}) \mid X_n) = pf(X_n)\,.$$

Es sei nun f beschränkt. Dann folgt hieraus, dass

$$M_n(f) := f(X_n) - \sum_{k=0}^{n-1} (pf(X_k) - f(X_k))$$

ein $(\mathcal{A}_n)$-Martingal ist, denn

$$\begin{aligned} E(M_{n+1}(f) \mid \mathcal{A}_n) &= E(f(X_{n+1}) \mid \mathcal{A}_n) - \sum_{k=0}^{n} (pf(X_k) - f(X_k)) \\ &= pf(X_n) - \sum_{k=0}^{n} (pf(X_k) - f(X_k)) \\ &= f(X_n) - \sum_{k=0}^{n-1} (pf(X_k) - f(X_k))\ . \end{aligned}$$

Sei nun h beschränkt und **harmonisch (bzgl. p)**, d.h. $ph = h$, dann ist insbesondere

$$M_n(h) = h(X_n)\,, n \in \mathbb{N}_0\,,$$

bereits ein Martingal. Diese Verbindung zwischen Markovketten und Martingalen bildet den Grundstein zur Analyse von Markovketten mithilfe der Martingaltheorie.

(4) **"Martingaltransformierte durch previsible Prozesse"** Es sei $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ eine Filtration, $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal und $(V_n)_{n\in\mathbb{N}}$ previsibel bzgl. $(\mathcal{A}_n)$, d.h., V_n $\mathcal{A}_{n-1}$-messbar $\forall n \geq 1$. Ist dann $V_k \cdot (X_k - X_{k-1}) \in \mathcal{L}^1(P)$, so ist

$$(V \cdot X)_n := X_0 + \sum_{k=1}^{n} V_k(X_k - X_{k-1}),\, n \in \mathbb{N}_0$$

(ebenfalls) ein $(\mathcal{A}_n)$-Martingal.

Beweis: $E((V \cdot X)_{n+1}|\mathcal{A}_n) = E((V \cdot X)_n|\mathcal{A}_n) + E(V_{n+1}(X_{n+1} - X_n)|\mathcal{A}_n) = (V \cdot X)_n + V_{n+1} \cdot E(X_{n+1} - X_n|\mathcal{A}_n) = (V \cdot X)_n$. □

(5) **"Faire Spiele"** Es seien Y_n, $n \in \mathbb{N}$ unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A})$ mit $P(Y_n = 1) = p(> 0)$ und $P(Y_n = -1) = 1 - p$. Man interpretiere Y_n als Ausgangs des n-ten Spiels, das ein Spieler in einer Spielbank spielt. $Y_n = +1$ bedeutet dabei "Gewinn" und $Y_n = -1$ "Verlust". Der Spieler legt eine Einsatzstrategie wie folgt fest: er wählt für jedes n eine Funktion $e_n : \{-1, +1\}^n \to \mathbb{R}_+$, die seinen Einsatz $e_n(Y_1, \ldots, Y_n)$ (abhängig von den Spielausgängen bis zum n-ten Spiel) festlegt. Z.B. könnte er

$$e_n(-1, \ldots, -1) = 2^{n-1}\,, = 0 \text{ sonst}$$

wählen ("Verdoppelungsstrategie"). Setze

$$S_1 := x \in \mathbb{R}_+\,, S_n := x + \sum_{k=2}^{n} e_{k-1}(Y_1, \ldots, Y_{k-1})Y_k\,, n \geq 2\,.$$

dann ist S_n gerade der Gewinn des Spielers nach Spiel n. Bezgl. $\mathcal{A}_n := \sigma(Y_1, \ldots, Y_n)$ ist dann (S_n) ein Martingal, genau dann wenn $p = \frac{1}{2}$, also genau im Fall eines fairen Spiels. Falls $p > \frac{1}{2}$, also der Spieler bevorzugt wird, liegt ein Submartingal vor, im Falle $p < \frac{1}{2}$ ein Supermartingal.

(6) **"Lokale Dichten"** Es sei Q ein Wahrscheinlichkeitsmass auf $(\Omega, \mathcal{A})$ mit

$$Q_{|\mathcal{A}_t} \ll P_{|\mathcal{A}_t} \,\forall t \in I\,.$$

Nach dem Satz von Radon-Nikodym 3.30 gibt es für $t \in I$ ein $X_t \geq 0$, $\mathcal{A}_t$-messbar mit

$$Q_{|\mathcal{A}_t} = X_t P_{|\mathcal{A}_t}\,.$$

Dann ist $(X_t)_{t\in I}$ ein $(\mathcal{A}_t)$-Martingal.

Beweis: $(\mathcal{A}_t)$-Adaptiertheit und Integrierbarkeit sind offensichtlich. Sei nun $s, t \in I$, $s \leq t$. Dann folgt für beliebige $A \in \mathcal{A}_s$:

$$\int_A X_t\, dP = \int_A dQ = \int_A X_s\, dP\,,$$

und damit $E(X_t \mid \mathcal{A}_s) = X_s$. □

Satz 4.3. *Es seien $(X_t)_{t\in I}$, $(Y_t)_{t\in I}$ zwei $(\mathcal{A}_t)_{t\in I}$-Martingale (bzw. Submartingale, bzw. Supermartingale). Weiter seien α, $\beta \geq 0$, und $q : \mathbb{R} \to \mathbb{R}$ konvex (bzw. konvex und monoton wachsend bzw. konkav und monoton wachsend). Dann gilt:*

(i) $(\alpha X_t + \beta Y_t)_{t\in I}$ *ist* $(\mathcal{A}_t)_{t\in I}$*-Martingal (bzw. Submartingal bzw. Supermartingal).*

(ii) $(q(X_t))_{t\in I}$ *ist ein* $(\mathcal{A}_t)_{t\in I}$*-Submartingal (bzw. Submartingal bzw. Supermartingal) falls* $q(X_t) \in \mathcal{L}(P)$ *für alle* $t \in I$.

Beweis. Es sei $s, t \in I$, $s \leq t$.
(i) $E(\alpha X_t + \beta Y_t \mid \mathcal{A}_s) = \alpha E(X_t \mid \mathcal{A}_s) + \beta E(Y_t \mid \mathcal{A}_s) = \alpha X_s + \beta Y_t$.
(ii) Nach Jensen gilt $E(q(X_t) \mid \mathcal{A}_s) \geq q\left(E(X_t \mid \mathcal{A}_s)\right) = q(X_s)$ (bzw. $\geq$ falls (X_t) Submartingal und q konvex und monoton wachsend). □

Satz 4.4. *(Doob-Meyer-Zerlegung - diskreter Fall)*
Es sei $I := \mathbb{N}_0$ und $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal. Dann gibt es genau ein Martingal $(M_n)_{n\in\mathbb{N}_0}$ und eine aufsteigende previsible Folge $(A_n)_{n\in\mathbb{N}_0}$ von Zufallsvariablen auf $(\Omega, \mathcal{A})$ mit $A_0 = 0$, so dass

$$X_n = M_n + A_n \quad \forall n \in \mathbb{N}_0\,.$$

Beweis. Eindeutigkeit: Gilt $X_n = M_n + A_n = M'_n + A'_n$, so folgt:

$$M_n - M'_n = A'_n - A_n \text{ ist } \mathcal{A}_{n-1} - \text{messbar}$$

und somit

$$M_n - M'_n = E(M_n - M'_n \mid \mathcal{A}_{n-1}) = M_{n-1} - M'_{n-1} \quad \forall n\,.$$

Folglich ist $M_n - M'_n = M_0 - M'_0 = A'_0 - A_0 = 0$.
Existenz: Definiere $A_0 := 0$ und

$$A_n := \sum_{k=1}^{n} E(X_k - X_{k-1} \mid \mathcal{A}_{k-1})\,, n \geq 1\,, \mathcal{A}_{n-1} - \text{messbar}\,.$$

Dann ist $M_n := X_n - A_n$, $\mathcal{A}_n$-messbar und integrierbar, sowie

$$\begin{aligned} E(M_n \mid \mathcal{A}_{n-1}) &= E(X_n - A_n \mid \mathcal{A}_{n-1}) = E(X_n \mid \mathcal{A}_{n-1}) - A_n \\ &= E(X_n \mid \mathcal{A}_{n-1}) - E(X_n - X_{n-1} \mid \mathcal{A}_{n-1}) - A_{n-1} \\ &= X_{n-1} - A_{n-1} = M_{n-1}\,. \end{aligned}$$

□

4.1 Maximalungleichungen

Die Martingaleigenschaft eines stochastischen Prozesses $(X_n)_{n\geqslant 0}$ impliziert, dass die Zufallsvariable X_n alle Informationen des Prozesses bis zur Zeit n besitzt, denn

$$X_m = E(X_n \mid \mathcal{A}_m) \quad \forall\, m \leq n\,.$$

Eine wichtige Aussage, die diese Eigenschaft ausnutzt, ist die folgende Maximalungleichung:

Theorem 4.5 (Maximalungleichung/Doobsche Ungleichung). *Es sei $(X_n)_{n\geqslant 0}$ ein Martingal oder ein nichtnegatives Submartingal und es sei*

$$X_n^* := \max_{0\leqslant k\leqslant n} |X_k| \,, \quad n \geqslant 0\,.$$

Dann gilt:

(i)

$$P(X_n^* \geqslant R) \leqslant \frac{1}{R}\, E(|X_n|\,; X_n^* \geqslant R) \leqslant \frac{1}{R}\, E(|X_n|) \quad \forall R > 0.$$

(ii) Falls $X_n \in \mathcal{L}^p(P)$ für ein $p > 1$ so folgt

$$E((X_n^*)^p)^{\frac{1}{p}} \leqslant \frac{p}{p-1}\; E(|X_n|^p)^{\frac{1}{p}}\,.$$

Beweis. (i) Offensichtlich gilt

$$\begin{aligned}
P\left(\max_{0\leqslant k\leqslant n} |X_k| \geqslant\ R\right) &= \sum_{l=0}^{n} P\left(|X_l| \geqslant R\,;\, \max_{0\leqslant k<l} |X_k| < R\right) \\
&\underset{\substack{\uparrow \\ \text{Markovsche Ungleichung}}}{\leqslant} \frac{1}{R}\sum_{l=0}^{n} E\Big(\underbrace{\mid X_l \mid}_{\leq E(\,|X_n|\,|\,\mathcal{A}_l)} \;;\, |X_l| \geqslant R,\ \max_{0\leqslant k<l} \mid X_k \mid < R\Big) \\
&\underset{\substack{\uparrow \\ |X_n|\text{ Sub-Martingal}}}{\leqslant} \frac{1}{R}\sum_{l=0}^{n} E\Big(E(|X_n| \mid \mathcal{A}_l);\, \underbrace{|X_l| \geqslant R,\ \max_{0\leqslant k<l} \mid X_k \mid < R}_{\in \mathcal{A}_l}\Big) \\
&\underset{\substack{\uparrow \\ \text{Projektivität}}}{=} \frac{1}{R}\sum_{l=0}^{n}\ E\left(|X_n|;\, |X_l| \geqslant R,\ \max_{0\leqslant k<l} |X_k| < R\right) \\
&= \frac{1}{R}E\left(|X_n|\,;\, \max_{0\geqslant k\geqslant n} \mid X_k \mid\geqslant R\right) \leqslant \frac{1}{R}E\left(|X_n|\,;\, X_n^* \geqslant R\right)\,.
\end{aligned}$$

Wegen $\{X_n^* \geq R\} = \{\max_{0\leq k\leq n} X_k \geq R\}$ folgt hieraus die Behauptung.

(ii)

$$\begin{aligned}
E((X_n^*)^p) &= E\left(p\int_0^{X_n^*} u^{p-1}\, du\right) = E\left(p\int_0^{\infty} 1_{\{u\leqslant X_n^*\}} u^{p-1}\, du\right) \\
&\overset{\text{Fubini}}{=} p\int_0^{\infty} \underbrace{E(1_{\{u\leqslant X_n^*\}})}_{\leqslant \frac{1}{u}E(|X_n|;\, X_n^*\geqslant u)\text{ nach (i)}} \quad u^{p-1}\, du \leqslant p\int_0^{\infty} E\left(\mid X_n \mid\,;\, X_n^* \geqslant u\right)\ u^{p-2}\, du \\
&\overset{\text{Fubini}}{=} p\ E\Big(\underbrace{\int_0^{X_n^*} u^{p-2}\, du}_{\frac{1}{p-1}(X_n^*)^{p-1}} \cdot \mid X_n \mid\Big) \leqslant \frac{p}{p-1}\ E\left(|X_n|^p\right)^{\frac{1}{p}}\ E\left((X_n^*)^p\right)^{\frac{p-1}{p}}\,.
\end{aligned}$$

Im letzten Schritt haben wir die Hölder Ungleichung mit p und $q = \frac{p}{p-1}$ angewandt.

□

4.2 Stoppzeiten und Stoppsatz

Im ganzen Unterabschnitt sei

- $(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum
- $I \subseteq \mathbb{N}_0$ Teilmenge
- $(\mathcal{A})_{t \in I}$ - Filtration

Das Ziel dieses Unterabschnittes ist die Verallgemeinerung der Martingaleigenschaft

$$E(X_t \mid \mathcal{A}_s) = X_s \quad \forall\, s \leq t$$

auf "zufällige" Zeiten S und T mit $S \leq T$. Die Klasse der Zeiten, für die eine solche Aussage richtig ist, wird in der folgenden Definition gegeben:

Definition 4.6. Eine Abbildung $T : \Omega \to I$ mit

$$\{T \leqslant n\} \in \mathcal{A}_n \quad \forall n \in I \tag{4.1}$$

heisst $(\mathcal{A}_n)$-**Stoppzeit**.

(4.1) bedeutet, dass T "nicht vorgreifend" ist, d.h. das Eintreten der Stoppregel bis zur Zeit n hängt nur von der bis zu diesem Zeitpunkt zur Verfügung stehenden Information $\mathcal{A}_n$ ab, aber nicht etwa von weiter in der Zukunft liegenden Informationen.

Beispiel 4.7. (1) **Eintrittszeiten:** $(X_n)_{n \in \mathbb{N}_0}$ $(\mathcal{A}_n)$-adaptiert mit Werten in einem messbaren Raum $(S, \mathcal{S})$. Sei $A \in \mathcal{S}$ und

$$\begin{aligned} T_A(\omega) &:= \inf\{n \geqslant 0 :\ X_n(\omega) \in A\},\ \omega \in \Omega \\ &= \text{erste Eintrittszeit in } A \quad (\inf \emptyset = +\infty) \end{aligned}$$

Dann ist T_A eine $(\mathcal{A}_n)$-Stoppzeit, denn

$$\{T_A \leqslant m\} = \bigcup_{n=0}^{m} \{X_n \in A\}\,.$$

(2) Es sei $I = \mathbb{N}_0$. Sind S und T $(\mathcal{A}_t)$-Stoppzeiten, so auch $S + T$ und $S \wedge T$, denn

$$\{S + T \leq n\} = \bigcup_{k,l \in \mathbb{N}_0 : k+l \leq n} \underbrace{\{S \leq k\}}_{\in \mathcal{A}_k} \cup \underbrace{\{T \leq l\}}_{\in \mathcal{A}_l} \in \mathcal{A}_n$$

und

$$\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}\ \in \mathcal{A}_n\,.$$

(3) Es sei $T : \Omega \to I$ eine Abbildung. Dann sind äquivalent:

(i) T ist eine $(\mathcal{A}_n)$-Stoppzeit.

(ii) $\{T = n\} \in \mathcal{A}_n$ für alle $n \in I$.

Beweis: $(i) \Rightarrow (ii)$ Für $n \in I$ bezeichne n' den Vorgänger von n in I. Dann folgt:

$$\{T = n\} = \underbrace{\{T \leq n\}}_{\in \mathcal{A}_n} \setminus \underbrace{\{T \leq n'\}}_{\in \mathcal{A}_{n'} \subseteq \mathcal{A}_n} \in \mathcal{A}_n \,.$$

Gibt es keinen solchen Vorgänger, ist also n das kleinste Element in I, so gilt offensichtlich $\{T \leq n\} = \{T = n\}$.

$(ii) \Rightarrow (i)$

$$\{T \leq n\} = \bigcup_{k \in I : k \leq n} \underbrace{\{T = k\}}_{\in \mathcal{A}_k \subseteq \mathcal{A}_n} \in \mathcal{A}_n \,.$$

□

Für eine Stoppzeit T definieren wir weiterhin

$$\mathcal{A}_T = \{A \subseteq \Omega \mid A \cap \{T \leqslant t\} \in \mathcal{A}_t \quad \forall t \in I\} \,.$$

Man kann zeigen, dass $\mathcal{A}_T$ eine σ-Algebra ist (Übung!). Sie heisst **σ-Algebra der T-Vergangenheit**.
Ist $I \subseteq \mathbb{N}_0$ und $(X_t)_{t \in I}$ ein $(\mathcal{A}_t)_{t \in I}$-adaptierter Prozess mit Werten in einem messbaren Raum $(S, \mathcal{S})$, so ist der durch T gestoppte Prozess

$$X_T(\omega) := X_{T(\omega)}(\omega)$$

$\mathcal{A}_T$-messbar, denn für $A \in \mathcal{S}$ gilt

$$\{X_T \in A\} \cap \{T \leq n\} = \bigcup_{m \in I : m \leq n} \underbrace{\{X_m \in A\}}_{\in \mathcal{A}_m} \cap \underbrace{\{T = m\}}_{\in \mathcal{A}_m} \in \mathcal{A}_n \quad \forall n \in I \,.$$

Die wichtigste Aussage im Zusammenhang mit Stoppzeiten:

Theorem 4.8. *(Diskreter Stoppsatz)*
Es sei $I = \mathbb{N}_0$ und $(X_n)_{n \in \mathbb{N}_0}$ ein $(\mathcal{A}_n)$-Martingal, S, T beschränkte $(\mathcal{A}_n)$-Stoppzeiten, $S \leqslant T$. Dann gilt

$$E(X_T | \, \mathcal{A}_S) = X_S \,.$$

Insbesondere gilt für jede $(\mathcal{A}_t)$-Stoppzeit T:

(i) $(X_{T \wedge t})$ ist ein $(\mathcal{A}_{T \wedge t})$-Martingal,

(ii) $E(X_{T \wedge t}) = E(X_0)$ ist konstant in der Zeit t.

Zum Beweis des Stoppsatzes benötigen wir das folgende Lemma:

Lemma 4.9. *Es sei (X_n), T, S wie in Theorem 4.8. Dann gilt*

$$E(X_T) = E(X_S) \,.$$

Beweis. Es sei $S \leqslant T \leqslant K$. Dann gilt

$$X_T = X_S + \sum_{k=S+1}^{T} X_k - X_{k-1} = X_S + \sum_{k=1}^{K} \underbrace{1_{\{S<k\leqslant T\}}}_{\in \mathcal{A}_{k-1}} (X_k - X_{k-1})$$

denn $\{S < k \leqslant T\} = \{S \leqslant k-1\} \cap \{T > k-1\}$. Somit

$$\begin{aligned} E(X_T) &= E(X_S) + \sum_{k=1}^{K} \underbrace{E\left(1_{\{S<k\leqslant T\}}(X_k - X_{k-1})\right)}_{=E\left(1_{\{S<k\leqslant T\}} \underbrace{E(X_k - X_{k-1} \mid \mathcal{A}_{k-1})}_{=0}\right)} \\ &= E(X_S)\,. \end{aligned}$$

□

Beweis von Theorem 4.8. Es gelte wieder $T \leqslant K$. Für $B \in \mathcal{A}_S$ sei

$$S^B := S1_B + K1_{B^c}$$

$$T^B := T1_B + K1_{B^c}$$

Dann sind $S^B, T^B (\mathcal{A}_n)$-Stoppzeiten, denn

$$\{S^B \leqslant n\} = \Big(\underbrace{\{S \leqslant n\} \cap B}_{\in \mathcal{A}_n}\Big) \cup \Big(\underbrace{\{K \leqslant n\} \cap B^c}_{=\begin{cases} \emptyset & , \ K > n \\ B^c & , \ K \leqslant n \end{cases} \in \mathcal{A}_n}\Big)$$

und

$$\{T^B \leqslant n\} = \Big(\underbrace{\{T \leqslant n\} \cap B}_{\in \mathcal{A}_n}\Big) \cup (\{K \leqslant n\} \cap B^c)$$

Das vorherige Lemma impliziert

$$\begin{array}{ccc} E(X_{S^B}) & = & E(X_{T^B}) \\ \shortparallel & & \shortparallel \\ E(X_S 1_B) + E(X_K 1_{B^c}) & & E(X_T 1_B) + E(X_K 1_{B^c}) \end{array}$$

und somit $E(X_S 1_B) = E(X_T 1_B)$. Da $B \in \mathcal{A}_S$ beliebig und X_S $\mathcal{A}_S$-messbar, erhalten wir $X_S = E(X_T|\mathcal{A}_S)$. □

Bemerkung 4.10. *Die Aussage von Theorem 4.8 ist für unbeschränkte Stoppzeiten T im allgemeinen falsch. Betrachte dazu als Beispiel die symmetrische Irrfahrt*

$$S_n := Y_1 + \ldots + Y_n \,, \mathcal{A}_n := \sigma(Y_1, \ldots, Y_n)$$

mit $(Y_k)_{k\in\mathbb{N}}$ unabhängig und identisch verteilt mit $P(Y_k = \pm 1) = \frac{1}{2}$. Betrachte dann folgende Stoppzeit:

$$T := \min\{n \geqslant 1 : S_n = +1\}$$

also die erste Trefferzeit von "1". Dann ist $T < \infty$ P-f.s., also $E(S_T) = 1$, jedoch $E(S_0) = 0$.

Korollar 4.11. *Es sei $(X_t)_{t\geqslant 0}$ ein $(\mathcal{A}_t)$-Martingal, T eine $(\mathcal{A}_t)$-Stoppzeit, so dass $(X_{T\wedge k})_{k\geqslant 1}$ gleichgradig integrierbar ist (z.B. beschränkt in k). Dann gilt*

$$E(X_T) = E(X_0)\,.$$

Beweis. $T \wedge k \uparrow T$ somit $X_{T\wedge k} \to X_T$ (P-f.s.) und folglich

$$E(X_T) = \lim_{n\to\infty} E\left(X_{T\wedge k}\right) = E(X_0).$$

□

Bemerkung. *(zur Definition von Stoppzeiten) Wir haben bislang nur Stoppzeiten betrachtet, die Werte in I annehmen, also endlich sind. Im weiteren Verlauf der Vorlesung wollen wir jedoch auch Stoppzeiten T zulassen, die möglicherweise den Wert $+\infty$ annehmen, was bereits für die in Beispiel 4.7 (1) definierten Eintrittszeiten*

$$T_A := \inf\{n \geq 0 : X_n \in A\} \quad (\inf \emptyset = +\infty)$$

der Fall sein kann, falls der Prozess $(X_n)_{n\in\mathbb{N}_0}$ die Menge A niemals erreicht und damit das Infimum über die leere Menge gebildet wird.

Alle bisherigen Aussagen bleiben für Stoppzeiten gültig, die P-f.s. endlich sind. Insbesondere ist für eine P-f.s. endliche Stoppzeit auch die Zufallsvariable X_T P-f.s. wohldefiniert.

Anwendungen des Stoppsatzes 4.8
(a) Lösung des Dirichletproblems und Berechnung von Austrittswahrscheinlichkeiten
Betrachte die Situation aus Beispiel 4.2 (3).
Dirichletproblem für eine Menge $A \in \mathcal{S}$
Gegeben: Beschränkte Funktion f auf A^c
Gesucht: h beschränkt auf S, $\mathcal{S}$-messbar mit

(a) $ph = h$ auf A, d.h. h p-harmonisch auf A,

(b) $h = f$ auf A^c

Satz 4.12. *Es gelte $P_x(T_{A^c} < \infty) = 1\ \forall\, x \in A$. Dann ist*

$$h(x) := E\left(f(X_{T_{A^c}}) \mid X_0 = x\right), x \in S\,,$$

die eindeutige minimale Lösung des Dirichletproblems für die Menge A mit "Randwerten" f.

Beweis. Der Einfachheit halber bezeichne P_x im folgenden die bedingte Verteilung $P(\cdot \mid X_0 = x)$.

Eindeutigkeit: Es sei h Lösung und $T := T_{A^c}$. Dann ist $M_n := h(X_{T\wedge n})$, $n \geq 0$, ein Martingal bezüglich $\mathcal{A}_n := \sigma(\{X_0, \ldots, X_n\})$, denn für $x \in A$ gilt

$$\begin{aligned} E_x(M_{n+1} \mid \mathcal{A}_n) &= E_x(h(X_{T\wedge(n+1)})1_{\{T\leq n\}} \mid \mathcal{A}_n) + E_x(h(X_{T\wedge(n+1)})1_{\{T>n\}} \mid \mathcal{A}_n) \\ &= E_x(\underbrace{h(X_{T\wedge n})1_{\{T\leq n\}}}_{\mathcal{A}_n-\text{messbar}} \mid \mathcal{A}_n) + E_x(h(X_{n+1}) \underbrace{1_{\{T>n\}}}_{\mathcal{A}_n-\text{messbar}} \mid \mathcal{A}_n) \\ &= h(X_{T\wedge n})1_{\{T\leq n\}} + 1_{\{T>n\}} \underbrace{E_x(h(X_{n+1}) \mid \mathcal{A}_n)}_{=ph(X_n)} \\ &= h(X_{T\wedge n})1_{\{T\leq n\}} + 1_{\{T>n\}} \underbrace{h(X_n)}_{=h(X_{T\wedge n})} = h(X_{T\wedge n})\,. \end{aligned}$$

Die Anwendung des Stoppsatzes ergibt

$$\begin{aligned} h(x) &= E(h(X_0) \mid X_0 = x) = E_x(h(X_0)) \\ &= E_x(M_0) = E_x(M_T) = E_x(h(X_T)) = E_x(f(X_T)) \end{aligned}$$

für alle $x \in A$. (Haben hierbei benutzt, dass $M_{T\wedge n} = h(X_{T\wedge n})$, $n \geq 0$, gleichgradig integrierbar, da h beschränkt.)
Existenz: Definiere $h(x) := E_x(f(X_T))$, $x \in S$. Dann gilt für $x \in A^c$ wegen $T_{A^c} = 0$ P_x-f.s. offensichtlich $h(x) = E_x(f(X_0)) = f(x)$. Bleibt noch zu zeigen: h ist p-harmonisch auf A. Dazu beachte, dass

$$\begin{aligned} ph(x) &= E_x(h(X_1)) = E_x(E_x(f(X_T) \mid X_1)) && \text{(Def. von } h\text{)} \\ &= E_x(E_x(f(X_T) \circ \theta \mid \mathcal{A}_1)) && \text{(Markoveigenschaft)} \\ &= E_x(f(X_T) \circ \Theta) \underbrace{=}_{x\in A} E_x(f(X_T)) = h(x)\,. \end{aligned}$$

□

(b) Waldsche Identität
Es seien $Y_1, Y_2, \ldots \in \mathcal{L}^1(P)$ unabhängig und identisch verteilt (unter P) und $m := E(Y_k)$. Aus Beispiel 4.2 (2) wissen wir, dass

$$M_n := \underbrace{\sum_{k=1}^{n} Y_k}_{=:S_n} - m \cdot n\,, n \geq 0$$

ein Martingal ist bezüglich der Filtration $\mathcal{A}_n := \sigma\{Y_1, \ldots, Y_n\}$, $n \geq 0$. Es sei nun T eine Stoppzeit mit $E(T) < \infty$ (nicht notwendigerweise unabhängig von $Y_1, Y_2, \ldots$). Dann gilt die Waldsche Identität

$$E(S_T) = m \cdot E(T)\,. \tag{4.2}$$

Beweis. Zunächst folgt aus dem Stoppsatz 4.8, dass
$0 = E(M_0) = E(M_{T\wedge n}) = E(S_{T\wedge n}) - m \cdot E(T \wedge n)$ für alle $n \geq 0$. Folglich ist

$$E(S_{T\wedge n}) = m \cdot E(T \wedge n) \quad \forall\, n \geq 0\,.$$

Wir wollen nun auf beiden Seiten dieser Gleichung den Grenzwert $n \to \infty$ nehmen. Dazu beachten wir, dass wegen $E(T) < \infty$ folgt, dass $T < \infty$ P-f.s., also $\lim_{n\to\infty} S_{T\wedge n} \to S_T$ P-f.s. Entsprechend gilt für $\tilde{S}_n := |Y_1| + \ldots + |Y_n|$, $n \in \mathbb{N}$, und $\tilde{m} := E(|Y_k|)$, dass

$$E(\tilde{S}_{T\wedge n}) = \tilde{m} \cdot E(T \wedge n) \quad \forall\, n \quad \text{und} \quad \tilde{S}_{T\wedge n} \to \tilde{S}_T \quad P-\text{f.s.}$$

Aus dem Satz über monotone Integration folgt nun für diesen Fall die Waldsche Identität, $E(\tilde{S}_T) = \tilde{m} \cdot E(T)$, also insbesondere $\tilde{S}_T \in \mathcal{L}^1(P)$. Aber wegen $|S_T| \le \tilde{S}_T$ impliziert der Lebesguesche Konvergenzsatz, dass

$$E(S_T) = \lim_{n\to\infty} E(S_{T\wedge n}) = \lim_{n\to\infty} m \cdot E(T \wedge n) = m \cdot E(T)\,.$$

□

4.3 Konvergenzsätze

Im ganzen Abschnitt sei $I = \mathbb{N}_0$ und eine Filtration $(\mathcal{A}_n)_{n\in I}$ gegeben.
Es sei $(X_n)_{n\in\mathbb{N}_0}$ ein $(\mathcal{A}_n)$-adaptierter Prozess. Für $a, b \in \mathbb{R}$ mit $a < b$ und $N \in \mathbb{N}$ definieren wir:

(i) $X_n(\omega)$ überquert das Intervall $[a, b]$ aufwärts (mindestens) n-mal auf $[0, N]$, falls s_i, t_i, $1 \le i \le n$ existieren mit $0 \le s_1 < t_1 < s_2 < t_2 < \ldots < s_n < t_n \le N$ mit $X_{s_i}(\omega) \le a$ und $X_{t_i}(\omega) \ge b$, $\forall 1 \le i \le n$.

(ii) $U(a, b; N)(\omega) := \inf\{n \ge 0 \mid X_n(\omega)$ überquert $[a, b]$ höchstens n-mal auf $[0, N]\}$

Theorem 4.13. *(Doob's Upcrossing Lemma)*
Es sei $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal und $N \in \mathbb{N}$. Dann gilt

$$E(U(a, b; N)) \le \frac{E\left((X_N - a)^+\right)}{(b - a)}\,.$$

Beweis. Definiere Stoppzeiten $T_0 := 0$, und S_n, T_n induktiv für $n \ge 1$ durch

$$S_n := \inf\{k \ge T_{n-1} \mid X_k \le a\} \wedge N\,,$$
$$T_n := \inf\{k \ge S_n \mid X_k \ge b\} \wedge N\,.$$

S_n, T_n sind offensichtlich $(\mathcal{A}_n)$-Stoppzeiten und es gilt

$$U(a, b; N) = \sup\{n \ge 0 \mid T_n > S_n, X_{T_n} \ge b\}\,.$$

Beachte dass $Y_n := (X_n - a)^+$, $n \in \mathbb{N}_0$, ein Submartingal ist nach Satz 4.3 und nichtnegativ. Weiterhin gilt

$$\begin{aligned}(b - a)U(a, b; N) &\le \sum_{n=1}^{N} (Y_{T_n} - Y_{S_n}) \\ &= Y_{\underbrace{T_N}_{=N}} - \sum_{n=2}^{N} (Y_{S_n} - Y_{T_{n-1}}) - \underbrace{Y_{S_1}}_{\ge 0}\,.\end{aligned}$$

Aus dem Stoppsatz 4.8 folgt nun, dass $E(Y_{S_n} - Y_{T_{n-1}}) \geq 0$, und damit

$$(b-a)E(U(a,b;N)) \leq E(Y_N) = E\left((X_N - a)^+\right) .$$

□

Korollar 4.14. *($f.s.$-Martingalkonvergenzsatz) Es sei $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal mit*

$$\sup_{n\geq 0} E(X_n^+) < \infty .$$

Dann existiert eine Zufallsvariable $X_\infty \in \mathcal{L}^1(P)$ mit :

$$X_\infty = \lim_{n\to\infty} X_n \quad P - f.s.$$

Beweis. Für $a < b$ gilt offenbar

$$U(a,b;N) \uparrow U(a,b;\infty) \qquad \text{für } N \uparrow \infty .$$

Hierbei bezeichnet $U(a,b;\infty)(\omega)$ die Anzahl aller Aufwärtsüberquerungen von $[a,b]$ auf $[0,\infty[$. Aus Doobs Upcrossing Lemma 4.13 folgt mithilfe einer montonen Integration

$$\begin{aligned} E(U(a,b;\infty)) = \lim_{N\uparrow\infty} E(U(a,b;N)) &\leq \lim_{N\uparrow\infty} \frac{E\left((X_N - a)^+\right)}{(b-a)^+} \\ &\leq \sup_{N\geq 0} \frac{E\left((X_N)^+\right) + |a|}{(b-a)^+} < \infty \end{aligned}$$

nach Annahme. Folglich gilt $P(N_{a,b}) = 0$ für

$$N_{a,b} := \{U(a,b;\infty) = \infty\} \subseteq \{\liminf_{n\to\infty} X_n \leq a\} \cap \{\limsup_{n\to\infty} X_n \geq b\} .$$

Dann ist auch die Menge

$$N := \bigcup_{a,b\in\mathbb{Q}:a<b} N_{a,b}$$

wieder eine P-Nullmenge. Für $\omega \in N^c$ ist $\liminf_{n\to\infty} X_n(\omega) = \limsup_{n\to\infty} X_n(\omega)$, also $X_n(\omega)$ konvergent. Definiere nun die Zufallsvariable

$$X_\infty := \liminf_{n\to\infty} X_n .$$

Dann gilt nach den obigen Vorüberlegungen $X_\infty = \lim_{n\to\infty} X_n$ P-f.s. Mit Hilfe des Lemmas von Fatou folgt nun

$$E(X_\infty^+) \leq \liminf_{n\to\infty} E(X_n^+) \leq \sup_{n\geq 0} E(X_n^+) < \infty$$

nach Annahme. Auf der anderen Seite folgt aus der Submartingaleigenschaft von (X_n) dass

$$E(X_n^-) = E(X_n^+) - E(X_n) \leq \sup_{n\geq 0} E(X_n^+) - E(X_0)$$

und damit im Grenzwert für $n \to \infty$ mit Hilfe des Lemmas von Fatou

$$E(X_\infty^-) \leq \liminf_{n\to\infty} E(X_n^-) \leq \sup_{n\geq 0} E(X_n^+) - E(X_0) < \infty .$$

Zusammengenommen also $E(|X_\infty|) = E(X^+) + E(X^-) < \infty$. □

Korollar 4.15. *Es sei $(X_n)_{n\in\mathbb{N}_0}$ ein nichtnegatives Supermartingal. Dann existiert eine Zufallsvariable $X_\infty \geq 0$ mit $E(X_\infty) \leq E(X_0)$ und*

$$X_\infty = \lim_{n\to\infty} X_n \quad \mathsf{P} - f.s.$$

Beweis. Aus dem P-f.s. Martingalkonvergenzsatz Korollar 4.14 folgt durch Anwendung auf das Submartingal $(-X_n)_{n\in\mathbb{N}_0}$ die Existenz des f.s.-Grenzwertes $X_\infty := \lim_{n\to\infty} X_n$. Mit Hilfe des Lemmas von Fatou folgt dann $E(X_\infty) \leq \liminf_{n\to\infty} E(X_n) \leq E(X_0)$. $\square$

Korollar 4.16. *($\mathcal{L}^p$-Martingalkonvergenzsatz, $p \geq 1$)*
Es sei $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal.

(i) ("$p = 1$ Fall") Äquivalent sind:

(a) (X_n) ist gleichgradig integrierbar.

(b) $X_\infty := \lim_{n\to\infty} X_n$ existiert in $\mathcal{L}^1(P)$.

(c) $\exists\, Y \in \mathcal{L}^1(P)$ mit $X_n = E(Y \mid \mathcal{A}_n)$ für alle $n \geq 0$.

In diesem Falle gilt: $X_\infty = E(Y \mid \bigvee_{n\geq 0} \mathcal{A}_n)$ P-f.s. und für alle $(\mathcal{A}_n)_{n\in[0,\infty]}$-Stoppzeiten S, T mit $S \leq T$ gilt

$$X_S = E(X_T \mid \mathcal{A}_S)\,.$$

Insbesondere ist $(X_n)_{n\in[0,\infty]}$ ein $(\mathcal{A}_n)_{n\in[0,\infty]}$-Martingal. Hierbei ist $\mathcal{A}_\infty := \bigvee_{n\geq 0} \mathcal{A}_n = \sigma\left(\bigcup_{n\geq 0} \mathcal{A}_n\right)$.

(ii) ("$p > 1$ Fall") Äquivalent sind:

(a) $\sup_{n\in\mathbb{N}_0} E\left(|X_n|^p\right) < \infty$.

(b) $X_\infty := \lim_{n\to\infty} X_n$ existiert in $\mathcal{L}^p(P)$.

Beweis. (i) $(a) \Rightarrow (b)$ Nach Annahme ist $(X_n)_{n\in\mathbb{N}_0}$ gleichgradig integrierbar und damit beschränkt in $\mathcal{L}^1(P)$. In der Tat: wegen

$$\limsup_{M\uparrow\infty} \sup_{n\geq 0} E\left(|X_n|, |X_n| \geq M\right) = 0$$

gibt es ein $M > 0$ mit $\sup_{n\geq 0} E\left(|X_n|, |X_n| \geq M\right) \leq 1$, und damit

$$E(|X_n|) \leq E(|X_n|, |X_n| \geq M) + M \leq 1 + M$$

für alle $n \geq 0$.
Also folgt aus dem P-f.s. Martingalkonvergenzsatz Korollar 4.14

$$\exists\, X_\infty := \lim_{n\to\infty} X_n \quad \in \mathbb{R} \quad \mathsf{P} - f.s.$$

Da (X_n) gleichgradig integrierbar, folgt mit Hilfe des Lebesgueschen Konvergenzsatzes nun auch $X_\infty = \lim_{n\to\infty} X_n$ in $\mathcal{L}^1(P)$ und damit die Behauptung.

$(b) \Rightarrow (c)$ Es sei $n \geq 0$ fest gewählt, also $X_n = E(X_m \mid \mathcal{A}_n)$ für alle $m \geq n$. Die Annahme (b) impliziert nun für alle $A \in \mathcal{A}_n$, dass

$$E(X_n, A) = \lim_{m\to\infty} E(E(X_m \mid \mathcal{A}_n), A) = \lim_{m\to\infty} E(X_m, A) = E(X_\infty, A)\,.$$

Hieraus folgt $X_n = E(X_\infty \mid \mathcal{A}_n)$.

$(c) \Rightarrow (a)$ Es sei $\varepsilon > 0$. Da $|Y|$ gleichgradig integrierbar, existiert nach dem $\varepsilon - \delta$-Kriterium (Lemma 1.62) ein $\delta > 0$ mit $P(A) < \delta$ impliziert $E(|Y|, A) < \varepsilon$. Mithilfe der Markovschen Ungleichung folgt für alle $M > 0$ und alle $n \geq 0$

$$P(|X_n| \geq M) \leq \frac{1}{M} E(|X_n|) \leq \frac{1}{M} E(E(|Y| \mid \mathcal{A}_n)) = \frac{1}{M} E(|Y|)\,.$$

Hieraus folgt für alle $M \geq \frac{1}{\delta} E(|Y|)$

$$E(|X_n|, \underbrace{|X_n| \geq M}_{\in \mathcal{A}_n}) = E(|E(Y \mid \mathcal{A}_n)|, |X_n| \geq M) \leq E(|Y|, |X_n| \geq M) < \varepsilon\,,$$

und (a) ist gezeigt.
Falls (a) (und damit (b) und (c)) gilt, so folgt

$$\lim_{n\to\infty} E(Y \mid \mathcal{A}_n) = \lim_{n\to\infty} X_n = X_\infty \quad \text{P-f.s. und in } \mathcal{L}^1(P)\,.$$

Sei nun $A \in \bigcup_{n\geq 0} \mathcal{A}_n$ also $A \in \mathcal{A}_{n_0}$ für ein $n_0 \geq 0$. Dann gilt

$$E(X_\infty, A) = \lim_{\substack{n\to\infty \\ n\geq n_0}} E(E(Y \mid \mathcal{A}_{n_0}), A) = E(Y, A)$$

und damit $X_\infty = E(Y \mid \bigvee_{n\geq 0} \mathcal{A}_n)$, da X_∞ $\bigvee_{n\geq 0} \mathcal{A}_n$-messbar.
Weiterhin gilt dann für alle $(\mathcal{A}_n)_{n\in[0,\infty]}$-Stoppzeiten S, T mit $S \leq T$ zunächst, falls $T \equiv \infty$ und S beschränkt durch $M > 0$,

$$E(X_\infty \mid \mathcal{A}_S) = \lim_{\substack{n\to\infty \\ n\geq M}} E(X_n \mid \mathcal{A}_S) = X_S$$

nach dem Stoppsatz 4.8 (für beschränkte Stoppzeiten). Falls S nicht beschränkt und $T \equiv \infty$ schließt man so: für alle $A \in \mathcal{A}_S$

$$\begin{aligned}
E(X_S, A) &= E(X_S, A \cap \{S < \infty\}) + E(X_\infty, A \cap \{S = \infty\}) \\
&= \sum_{k=0}^{\infty} E(X_{S\wedge k}, \underbrace{A \cap \{S = k\}}_{\in \mathcal{A}_{S\wedge k}}) + E(X_\infty, A \cap \{S = \infty\}) \\
&\underset{s.oben}{=} \sum_{k=0}^{\infty} E(X_\infty, A \cap \{S = k\}) + E(X_\infty, A \cap \{S = \infty\}) \\
&= E(X_\infty, A)\,.
\end{aligned}$$

Folglich ist $E(X_\infty \mid \mathcal{A}_S) = X_S$ (insbesondere ist X_S $\mathcal{A}_S$-messbar). Für beliebiges T mit $S \leq T$ folgt

$$E(X_T \mid \mathcal{A}_S) = E((E(X_\infty \mid \mathcal{A}_T) \mid \mathcal{A}_S) \underset{\mathcal{A}_S \subseteq \mathcal{A}_T}{=} E(X_\infty \mid \mathcal{A}_S) = X_S\,.$$

(ii) $(b) \Rightarrow (a)$ Klar.
$(a) \Rightarrow (b)$ Nach Annahme folgt aus der Maximalungleichung Theorem 4.5 (für $p > 1$) dass $\sup_{n\geq 0} |X_n|^p \in \mathcal{L}^1(P)$ und damit ist $\{|X_n|^p \mid n \geq 0\}$ gleichgradig integrierbar. Damit folgt (b) aus (i). □

Rückwärtsmartingalkonvergenzsatz
Statt $I = \mathbb{N}_0$ betrachten wir im folgenden als Indexmenge $I = -\mathbb{N} = \{-n \mid n \in \mathbb{N}\}$ mit der natürlichen Ordnung "$\leq$" und hierzu eine Filtration $(\mathcal{A}_n)_{n\in-\mathbb{N}}$.

Theorem 4.17. *(Rückwärtsmartingalkonvergenzsatz)*
Es sei $(X_n)_{n\in-\mathbb{N}}$ ein $(\mathcal{A}_n)_{n\in-\mathbb{N}}$-Submartingal. Dann gilt:

(i) $\exists\, X_{-\infty} := \lim_{n\to-\infty} X_n \in \mathbb{R} \cup \{-\infty\}\; P-$ *f.s.*

(ii) Falls

$$\sup_{n\in-\mathbb{N}} E(|X_n|) < \infty\, (\Leftrightarrow \inf_{n\in-\mathbb{N}} E(X_n) > -\infty\,)\,,$$

so ist $(X_n)_{n\in-\mathbb{N}}$ gleichgradig integrierbar und

$$X_{-\infty} = \lim_{n\to-\infty} X_n \quad \textit{in } \mathcal{L}^1(P)\,.$$

Insbesondere gilt:

$$\lim_{n\to-\infty} E(Y \mid \mathcal{A}_n) = E(Y \mid \bigcap_{n\in-\mathbb{N}} \mathcal{A}_n)\; P - \textit{f.s. und in } \mathcal{L}^1(P)\,.$$

Beweis. (i) Analog zum Fall $I = \mathbb{N}$ im P-f.s. Martingalkonvergenzsatz 4.14 zeigt man für $a > b$

$$U(a,b;-N) \uparrow U(a,b;-\infty) \quad \text{für } N \uparrow \infty\,.$$

Hierbei bezeichnet $U(a,b;-N)(\omega)$ die Anzahl der Aufwärtsüberquerungen von $[a,b]$ auf $[-N,0]$ und $U(a,b;-\infty)(\omega)$ die Anzahl aller Aufwärtsüberquerungen von $[a,b]$ auf $]-\infty,0]$. Aus Doobs Upcrossing Lemma 4.13 folgt mit hilfe einer montonen Integration

$$\begin{aligned} E(U(a,b;-\infty)) = \lim_{N\uparrow\infty} E(U(a,b;-N)) &\leq \lim_{N\uparrow\infty} \frac{E\left((X_0-a)^+\right)}{(b-a)^+} \\ &\leq \frac{E\left((X_0)^+\right) + |a|}{(b-a)^+} < \infty\,. \end{aligned}$$

Wie im P-f.s. Martingalkonvergenzsatz 4.14 folgert man hieraus, dass $\liminf_{n\to-\infty} X_n(\omega) = \limsup_{n\to-\infty} X_n(\omega)$ P-f.s., also $X_n(\omega)$ konvergent für $n \to -\infty$ P-f.s. Definiere nun die Zufallsvariable

$$X_{-\infty} := \liminf_{n\to-\infty} X_n\,.$$

Dann gilt $X_{-\infty} = \lim_{n\to-\infty} X_n$ P-f.s. Mit Hilfe des Lemmas von Fatou folgt nun

$$E(X_{-\infty}^+) \le \liminf_{n\to-\infty} E(X_n^+) \le E(X_0^+) < \infty\,,$$

also $X_{-\infty} < \infty$ P-f.s. (Man vergleiche dazu den entsprechenden Fall des Vorwärtsmartingals: Im Falle des Rückwärtsmartingals ist die im Vorwärtsfall benötigte Annahme der $\mathcal{L}^1(P)$-Beschränkheit immer erfüllt!).
(ii) Da $E(X_n)$ aufsteigend in n, folgt

$$\lim_{n\to-\infty} E(X_n) = \inf_{n\in-\mathbb{N}} E(X_n) > -\sup_{n\in-\mathbb{N}} E(|X_n|)\,.$$

Es folgt hieraus für $M > 0$

$$\begin{aligned}
E(|X_n|, |X_n| > M) &= E(X_n, X_n > M) - E(X_n, X_n < -M)\\
&= E(X_n, X_n > M) - E(X_n) + E(X_n, X_n \ge -M)\\
&\underbrace{\le}_{\text{für } n\le n_0}\ E(X_{n_0}, X_n > M) - E(X_n) + E(X_{n_0}, X_n \ge -M)\,.
\end{aligned}$$

Zu $\varepsilon > 0$ findet man nun ein n_0 mit der Eigenschaft, dass

$$E(X_n) \ge E(X_{n_0}) - \varepsilon \ \text{ für alle } n \le n_0\,.$$

Damit folgt dann aber

$$\begin{aligned}
E(|X_n|, |X_n| > M) &\le E(X_{n_0}, X_n > M) - E(X_{n_0}) + \varepsilon + E(X_{n_0}, X_n \ge -M)\\
&= E(X_{n_0}, |X_n| > M) + \varepsilon\\
&\le E(|X_{n_0}|, |X_n| > M) + \varepsilon \to \varepsilon \quad \text{ für } M \to \infty\,.
\end{aligned}$$

Damit folgt (ii) aus dem verallgemeinerten Konvergenzsatz von Lebesgue (Satz 1.59). Zum Beweis der letzten Aussage beachte man, dass für $Y \in \mathcal{L}^1(P)$ $X_n := E(Y \mid \mathcal{A}_n)$ ein Rückwärtsmartingal ist mit $\sup_{n\in-\mathbb{N}} E(|X_n|) \le E(|Y|) < \infty$. Daher existiert $X_{-\infty} := \lim_{n\to-\infty} X_n$ P-f.s. und in $\mathcal{L}^1(P)$. Beachte, dass $X_{-\infty}$ trivialerweise $\mathcal{A}_{-\infty}$-messbar ist, wobei zur Vereinfachung der Notation $\mathcal{A}_{-\infty} := \bigcap_{n\in-\mathbb{N}} \mathcal{A}_n$ gesetzt wurde. Für $A \in \mathcal{A}_{-\infty}$ folgt nun aber

$$E(X_{-\infty}, A) = \lim_{n\to-\infty} E(X_n, A) = \lim_{n\to-\infty} E(E(Y \mid \mathcal{A}_n), A) = \lim_{n\to-\infty} E(Y, A) = E(Y, A)\,.$$

Hieraus folgt die Behauptung. □

4.4 Anwendungen des Martingalkonvergenzsatzes

(1) Klassische symmetrische Irrfahrt. $X_n = Y_1 + \ldots + Y_n$ mit $(Y_k)_{k\in\mathbb{N}}$ unabhängig und identisch verteilt mit $P(Y_k = \pm 1) = \frac{1}{2}$.
Für $c > 0$ sei $T_c := \min\{n \ge 0 \mid X_n \ge c\}$, Stoppzeit. Dann ist $(X_{T_c\wedge n})_{n\in\mathbb{N}}$ Martingal nach dem Stoppsatz 4.8 und $X_{T_c\wedge n} \le c + 1$ für alle n. Also P-f.s. konvergent nach dem f.s.-Martingalkonvergenzsatz 4.14. Da X_n nur Inkremente ± 1 besitzt, ist aber $X_{T_c\wedge n}$ auf

$\{T_c = \infty\}$ nicht konvergent, also muss $P(T_c < \infty) = 1$ gelten. Da diese für alle $c > 0$ gilt, ist $P(\limsup_{n\to\infty} X_n = \infty) = 1$. Analog zeigt man $P(\liminf_{n\to\infty} X_n = -\infty) = 1$. Damit folgt

$$P(\liminf_{n\to\infty} X_n = -\infty, \limsup_{n\to\infty} X_n = +\infty) = 1\,,$$

d.h., (X_n) ist von unbeschränkter Variation.

(2) Martingale mit **beschränkten Inkrementen**

Satz 4.18. *Es sei $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal mit beschränkten Inkrementen, d.h.*

$$\sup_{n\in\mathbb{N}_0} |X_{n+1} - X_n| =: Y \;\in\; \mathcal{L}^1(P)\,.$$

Dann gilt $P(C \cup E) = 1$, wobei

$$C := \{\omega \in \Omega \mid \exists \lim_{n\to\infty} X_n(\omega) \in \mathbb{R}\}$$
$$E := \{\liminf_{n\to\infty} X_n = -\infty, \limsup_{n\to\infty} X_n = +\infty\}\,.$$

Beweis. Zu $c \in \mathbb{R}$ betrachte wieder die Stoppzeit

$$T_c := \min\{n \geq 0 \mid X_n \geq c\}\,.$$

Dann gilt

$$X_{T_c\wedge n} = X_{T_c} 1_{\{T_c<n\}} + X_n 1_{\{T_c\geq n\}} \leq Y + c\,.$$

Aus dem f.s.-Martingalkonvergenzsatz 4.14 folgt nun wieder

$$\exists \lim_{n\to\infty} X_{T_c\wedge n} \in \mathbb{R}\, P-\text{f.s.}$$

und damit

$$\exists \lim_{n\to\infty} X_n \in \mathbb{R}\, P-\text{f.s. auf der Menge } \{T_c = \infty\}\,.$$

Folglich ist

$$P\left(\left(\bigcup_{c\in\mathbb{N}}\{T_c = \infty\}\right) \cap C^c\right) = 0\,.$$

Andererseits gilt $\left(\bigcup_{c\in\mathbb{N}})\{T_c = \infty\}\right)^c = \bigcap_{c\in\mathbb{N}}\{T_c < \infty\} = \{\limsup_{n\to\infty} X_n = +\infty\}$, so dass

$$P\left(\{\liminf_{n\to\infty} X_n = +\infty\} \cup C\right) = 1\,.$$

Übergang zu $(-X_n)_{n\in\mathbb{N}_0}$ ergibt analog

$$P\left(\{\limsup_{n\to\infty} X_n = +\infty\} \cup C\right) = 1$$

und damit die Behauptung. □

Korollar 4.19. *(verallgemeinertes Lemma von Borel-Cantelli)*
Es sei $(\mathcal{A}_n)_{n\geq 0}$ eine Filtration und $A_n \in \mathcal{A}_n$, $n \geq 0$, beliebige Ereignisse. Dann gilt

$$\limsup_{n\to\infty} A_n = \bigcap_n \bigcup_{m\geq n} A_m = \left\{ \sum_{n\geq 0} P\left(A_{n+1} \mid \mathcal{A}_n\right) = \infty \right\} \quad P - \textit{f.s.}$$

Beweis. Übung! □

(3) Lebesgue-Zerlegung
Es seien P, Q Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$, $\mathcal{A}_0 \subseteq \mathcal{A}$ eine Unter-σ-Algebra. Existiert dann eine Zufallsvariable $X_0 \in \mathcal{L}^1(\Omega, \mathcal{A}_0, P)$, $X_0 \geq 0$, mit

$$\begin{aligned} Q(A) &= \int_A X_0 \, dP + Q(A \cap \{X_0 = \infty\}) \\ &=: Q_a(A) + Q_s(A) \qquad \forall\, A \in \mathcal{A}_0 \,. \end{aligned} \tag{4.3}$$

so heisst (4.3) **Lebesgue-Zerlegung** von Q bzgl. P (auf der σ-Algebra $\mathcal{A}_0$). Notation:

$$X_0 := L - \frac{dQ}{dP}_{|\mathcal{A}_0} \,.$$

Lemma 4.20. *Ist $X_0 = L - \frac{dQ}{dP}_{|\mathcal{A}_0}$ dann folgt: $X_0 > 0$ Q-f.s. und*

$$\frac{1}{X_0} = L - \frac{dP}{dQ}_{|\mathcal{A}_0} \,.$$

Beweis. Folgt aus

$$Q(X_0 = 0) = \int_{\{X_0=0\}} X_0 \, dP + Q(\{X_0 = 0\} \cap \{X_0 = \infty\}) = 0$$

sowie aus

$$P(A) = P(A \cap \underbrace{\{X_0 = 0\}}_{\frac{1}{X_0}=\infty}) + \underbrace{\int_{A\cap\{X_0>0\}} \frac{1}{X_0} X_0 dP}_{\int_A \frac{1}{X_0}\, dQ} \quad \forall\, A \in \mathcal{A}_0 \,.$$

□

Eindeutigkeit Es sei $\tilde{X}_0 \in \mathcal{L}^1(\Omega, \mathcal{A}_0, P)$, $\tilde{X}_0 \geq 0$ mit (4.3) gegeben. Anwendung von (4.3) auf $A = \{X_0 = \infty\}$ ergibt

$$\begin{aligned} Q(X_0 = \infty) &= \underbrace{\int_{\{X_0=\infty\}} \tilde{X}_0 \, dP}_{=0 \text{ da } X_0<\infty P-\text{f.s.}} + Q(X_0 = \infty, \tilde{X}_0 = \infty) \\ &= Q(X_0 = \infty, \tilde{X}_0 = \infty) \,. \end{aligned}$$

Ebenso gilt
$$Q(\tilde{X}_0 = \infty) = Q(X_0 = \infty, \tilde{X}_0 = \infty)\,.$$
Somit folgt
$$\int_A X_0\, dP = \int_A \tilde{X}_0\, dP \qquad \forall A \in \mathcal{A}_0$$
also $X_0 = \tilde{X}_0$ P-f.s., da beide Zufallsvariablen, X_0 und $\tilde{X}_0$, $\mathcal{A}_0$-messbar sind.

Existenz

Elementarer Fall: $\mathcal{A}_0 = \sigma(\{B_n \mid n \in \mathbb{N}\})$ für eine $\mathcal{A}$-messbare Partition von Ω. Dann gilt für alle $A \in \mathcal{A}_0$, also $A = \bigcup_{n\in\mathbb{N}:B_n\subseteq A} B_n$,
$$\begin{aligned} Q(A) &= \sum_{n:B_n\subseteq A} Q(B_n) \\ &= \sum_{n:B_n\subseteq A, P(B_n)>0} \frac{Q(B_n)}{P(B_n)} P(B_n) + \sum_{n:B_n\subseteq A, P(B_n)=0} Q(B_n) \\ &= \int_A \underbrace{\sum_{n:P(B_n)>0} \frac{Q(B_n)}{P(B_n)} P(B_n) 1_{B_n}}_{=:X_0 \text{ auf } \bigcup_{n:P(B_n)>0} B_n} dP + Q(A \cap \{X_0 = \infty\}) \end{aligned}$$
mit $X_0 = \infty$ auf $\bigcup_{n:P(B_n)>0} B_n$.
Wir beweisen nun im folgenden weiteren Spezialfall die Existenz der Lebesgue-Zerlegung: Gegeben sei eine Filtration $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ und $\mathcal{A}_\infty := \sigma\left(\bigcup_{n\in\mathbb{N}_0} \mathcal{A}_n\right)$.
Wir nehmen an, dass auf $\mathcal{A}_n$ die Lebesgue-Zerlegung mit "lokaler Dichte" $X_n := L - \frac{dQ}{dP}_{|\mathcal{A}_n}$ existiert. Dann gilt:

Satz 4.21. *(i) (a) $(X_n)_{n\in\mathbb{N}_0}$ ist ein (nichtnegatives) Supermartingal bzgl. P,*

(b) $(\frac{1}{X_n})_{n\in\mathbb{N}_0}$ ist ein (nichtnegatives) Supermartingal bzgl. Q.

Insbesondere existiert $X_\infty := \lim_{n\to\infty} X_n$ mit $P(X_\infty \in [0,\infty[) = 1$ und $Q(X_\infty \in]0,\infty]) = 1$.

(ii) $X_\infty = L - \frac{dQ}{dP}_{|\mathcal{A}_\infty}$ und $\frac{1}{X_\infty} = L - \frac{dP}{dQ}_{|\mathcal{A}_\infty}$.

Beweis. (i) Wie im Beweis der Eindeutigkeit folgt
$$Q(X_n = \infty) = Q(X_n = \infty, X_{n+1} = \infty)$$
also $\{X_n = \infty\} \subseteq \{X_{n+1} = \infty\}$ P-f.s. Somit folgt für alle $A \in \mathcal{A}_n$
$$\begin{aligned} \int_A X_{n+1}\, dP &= Q(A) - Q(A \cap \{X_{n+1} = \infty\}) \\ &\leq Q(A) - Q(A \cap \{X_n = \infty\}) = \int_A X_n dP\,. \end{aligned}$$
Damit ist $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal bzgl. P. Da $\frac{1}{X_n} = L - \frac{dP}{dQ}_{|\mathcal{A}_n}$ (Lemma 4.20), folgt analog dass $(\frac{1}{X_n})_{n\in\mathbb{N}_0}$ ein Supermartingal bzgl. Q ist. Nach dem

f.s.-Martingalkonvergenzsatz 4.14 existieren $X_\infty := \lim_{n\to\infty} X_n \in [0,\infty[$ P-f.s. und $\frac{1}{X_\infty} := \lim_{n\to\infty} \frac{1}{X_n} \in]0,\infty]$ Q-f.s.
(ii) Definiere

$$Q_a(A) := \int_A X_\infty \, dP \, , Q_s(A) := Q(A) - Q_a(A) \, , A \in \mathcal{A}_\infty \, .$$

Zu zeigen ist dann, dass $Q_s(A) := Q(A \cap \{X_\infty = \infty\})$. Dafür ist es hinreichend, zu zeigen, dass $Q_s(X_\infty < \infty) = 0$, also $Q(X_\infty < \infty) = Q_a(X_\infty < \infty)$.
Aus dem Lebesgueschen Konvergenzsatz folgt nun aber

$$\begin{aligned} Q(X_\infty < \infty) &= \lim_{c\uparrow\infty} Q(X_\infty \leq c) = \lim_{c\uparrow\infty} \lim_{n\to\infty} Q(X_n \leq c) \\ &= \lim_{c\uparrow\infty} \lim_{n\to\infty} \int_{\{X_n \leq c\}} X_n \, dP + Q(\underbrace{\{X_n \leq c\} \cap \{X_n = \infty\}}_{=\emptyset}) \\ &= \lim_{c\uparrow\infty} \int_{\{X_\infty \leq c\}} X_\infty \, dP = Q_a(X_\infty < c) \, . \end{aligned}$$

Der zweite Teil von (ii) folgt nun wieder aus dem Lemma 4.20. □

(4) Asymptotik quadratisch integrierbarer Martingale
Es sei $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal in $\mathcal{L}^2(P)$, $X_0 := 0$. Dann ist X_n^2, $n \geq 0$, ein nichtnegatives Submartingal, denn

$$X_n^2 = (E(X_{n+1} \mid \mathcal{A}_n))^2 \leq E\left(X_{n+1}^2 \mid \mathcal{A}_n\right) \, .$$

Die Doob-Meyer Zerlegung Satz 4.4

$$X_n^2 = \underbrace{M_n}_{\text{Martingal}} + \underbrace{V_n}_{\nearrow\text{,previsibel}} \quad , n \geq 0$$

ist gegeben durch

$$\begin{aligned} V_n &= \sum_{k=1}^{n} E\left(X_k^2 - X_{k-1}^2 \mid \mathcal{A}_{k-1}\right) \, , n \geq 1, \, V_0 = 0 \\ &= \sum_{k=1}^{n} E\left((X_k - X_{k-1})^2 \mid \mathcal{A}_{k-1}\right) \qquad (\, = n\sigma^2, \text{ falls } Y_k := X_k - X_{k-1} \text{ u.i.d.}) \end{aligned}$$

$(V_n)_{n\in\mathbb{N}_0}$ heißt **Varianzprozess (von $(X_n)_{n\in\mathbb{N}_0}$)**. Setze nun

$$V_\infty := \sum_{k=1}^{\infty} E\left((X_k - X_{k-1})^2 \mid \mathcal{A}_{k-1}\right) \, .$$

Dann gilt offenbar:

$$E(V_\infty) < \infty \Leftrightarrow \sum_{k=1}^{\infty} E\left((X_k - X_{k-1})^2\right) < \infty \Leftrightarrow \lim_{n\to\infty} E(X_n^2) < \infty \, . \tag{4.4}$$

Begründung:

$$\sum_{k=1}^{n} E\left((X_k - X_{k-1})^2\right) = E(V_n) = E(X_n^2 - M_n) \underbrace{=}_{E(M_n)=0} E(X_n^2)\,.$$

Eine Verfeinerung liefert folgender

Satz 4.22.

$$\exists \lim_{n\to\infty} X_n \in \mathbb{R} \quad P-\textit{f.s. auf}\ \{V_\infty < \infty\}\,.$$

Beweis. Für $c > 0$ definiere die Stoppzeit

$$T_c := \min\{n \geq 0 \mid V_{n+1} > c\}$$

Beachte, dass T_c in der Tat eine $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$-Stoppzeit ist, da V_n $\mathcal{A}_{n-1}$-messbar ist. Dann ist der gestoppte Varianzprozess $V_n^{T_c} := V_{T_c\wedge n}$, $n \in \mathbb{N}_0$, gerade der Varianzprozess des gestoppten Martingals $(X_{T_c\wedge n})_{n\in\mathbb{N}_0}$, denn V^{T_c} ist aufsteigend, previsibel und $M_{T_c\wedge n} = X^2_{T_c\wedge n} - V_{T_c\wedge n}$, $n \geq 0$, ist nach dem Stoppsatz 4.8 ein $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$-Martingal. Wegen $V_\infty^{T_c} \leq c$ folgt aus den Vorüberlegungen

$$\exists \lim_{n\to\infty} X_n^{T_c} \in \mathbb{R} \quad P-\text{f.s.}$$

und für $c \uparrow \infty$ ergibt sich

$$\exists \lim_{n\to\infty} X_n \in \mathbb{R} \quad P-\text{f.s. auf}\ \{V_\infty < \infty\}\,.$$

□

Was passiert nun auf der Menge $\{V_\infty = \infty\}$?

Satz 4.23. *Es sei* $0 = C_0 \leq C_1 \leq C_2 \leq \ldots$ *previsibel. Dann gilt:*

$$\sum_{k=1}^{\infty} E\left(\left(\frac{X_k - X_{k-1}}{C_k}\right)^2\right) < \infty \quad \Rightarrow \quad \frac{X_n}{C_n} \to 0 \quad P-\textit{f.s. auf}\ \{C_\infty := \sup_{n\geq 0} C_n = \infty\}\,.$$

Beweis. Zur Vereinfachung der Notation setzen wir wieder $Y_k := X_k - X_{k-1}$. Dann folgt aus Beispiel 4.2 (4), dass auch

$$\tilde{X}_n := \sum_{k=1}^{n} \frac{Y_k}{C_k}\,, n \geq 0\,,$$

wieder ein $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$-Martingal ist und aus der Annahme ergibt sich aufgrund von (4.4), dass

$$\exists \lim_{n\to\infty} \sum_{k=1}^{n} \frac{Y_k}{C_k} \in \mathbb{R} \quad P-\text{f.s.}$$

Damit folgt die Behauptung aus Kroneckers Lemma 4.24. □

Lemma 4.24. *Es seien c_n und y_n, $n \geq 1$, Folgen reeller Zahlen mit $0 < c_1 \leq c_2 \leq \ldots \uparrow \infty$ und $\sum_{k=1}^{\infty} \frac{y_k}{c_k} < \infty$. Dann folgt*

$$\lim_{n\to\infty} \frac{1}{c_n} \sum_{k=1}^{n} y_k = 0 \,.$$

Beweis. Es sei $s_n := \sum_{k=1}^{n} \frac{y_k}{c_k}$. Dann folgt mithilfe einer "partiellen Integration"

$$\sum_{k=1}^{n} y_k = \sum_{k=1}^{n} c_k(s_k - s_{k-1}) = \sum_{k=1}^{n} (c_k - c_{k-1})(s_n - s_{k-1}) \,.$$

Da s_n konvergent nach Annahme, gibt es zu $\varepsilon > 0$ ein $N_\varepsilon \in \mathbb{N}$ mit $|s_n - s_m| < \varepsilon$ für $n, m \geq N_\varepsilon$. Dann aber folgt für $n \geq N_\varepsilon$

$$\begin{aligned}
\Big| \sum_{k=1}^{n} y_k \Big| &\leq \Big| \sum_{k=1}^{n} (c_k - c_{k-1})(s_n - s_{N_\varepsilon}) \Big| + \Big| \sum_{k=1}^{n} (c_k - c_{k-1})(s_{N_\varepsilon} - s_{k-1}) \Big| \\
&\leq \underbrace{\Big| \sum_{k=1}^{N_\varepsilon} (c_k - c_{k-1})(s_{N_\varepsilon} - s_{k-1}) \Big|}_{\text{beschränkt in } n} \\
&\quad + 2 \underbrace{\sup_{m \geq N_\varepsilon} |s_n - s_m|}_{<\varepsilon} \cdot \underbrace{\sum_{k=N_\varepsilon+1}^{n} (c_k - c_{k-1})}_{\leq c_n}
\end{aligned}$$

woraus die Behauptung folgt. □

Korollar 4.25. *Es sei $Y_k \in \mathcal{L}^1(P)$, $k \geq 1$ mit $E(Y_k \mid \mathcal{A}_{k-1}) = 0$ gegeben, und $X_n := \sum_{k=1}^{n} Y_k$, $n \geq 1$. Dann gilt:*

(i) $p > 0$ mit $\sum_{k=1}^{\infty} E\left(\left(\frac{Y_k}{k^p}\right)^2\right) < \infty \quad \Rightarrow \quad \frac{X_n}{n^p} \to 0$ P-f.s. (Im Spezialfall $p = 1$ ergibt sich also hiermit eine Verallgemeinerung des starken Gesetzes der großen Zahlen von Kolmogorov 2.15!)

(ii) $\sup_{n\geq 1} E(Y_n^2) < \infty \quad \Rightarrow \quad \frac{X_n}{n^p} \to 0$ P-f.s. für alle $p > \frac{1}{2}$.

Beweis. (i) Zunächst impliziert die Annahme, dass X_n, $n \geq 1$, ein quadratintegrierbares Martingal ist. Somit folgt die Behauptung aus Satz 4.23 mit $C_k := k^p$.
(ii) folgt aus (i), da

$$\sum_{k=1}^{\infty} E\left(\left(\frac{Y_k}{k^p}\right)^2\right) \leq \sum_{k=1}^{\infty} \frac{1}{k^{2p}} \sup_{k\geq 1} E\left(Y_k^2\right) < \infty$$

wegen der Annahme und wegen $p > \frac{1}{2}$. □

Korollar 4.26. *Für alle $p > \frac{1}{2}$ gilt $\lim_{n\to\infty} \frac{X_n}{V_n^p} = 0$ P-f.s. auf $\{V_\infty = \infty\}$ und in Verbindung mit Satz 4.22*

$$\exists \lim_{n\to\infty} \frac{X_n}{V_n^p} \in \mathbb{R} \quad P-\text{f.s.}$$

Beweis. Setze $f(x) := x^p$, $C_n := 1 + f(V_n)$, $n \geq 1$, previsibel. Dann gilt

$$E\left(\left(\frac{Y_n}{C_n}\right)^2 \mid \mathcal{A}_{n-1}\right) = \frac{V_n - V_{n-1}}{(1+f(V_n))^2} \leq \int_{V_{n-1}}^{V_n} \frac{1}{(1+f(x))^2}\, dx\,.$$

Daraus folgt offenbar

$$\sum_{n\geq 1} E\left(\left(\frac{Y_n}{C_n}\right)^2\right) \leq \int_0^\infty \frac{1}{(1+x^p)^2}\, dx < \infty\,.$$

Satz 4.23 impliziert nun $\lim_{n\to\infty} \frac{X_n}{1+V_n^p} = 0$ P-f.s. auf $\{V_\infty = \infty\}$, also $\lim_{n\to\infty} \frac{X_n}{V_n^p} = 0$ P-f.s. auf $\{V_\infty = \infty\}$. □

Im "kritischen Fall" $p = \frac{1}{2}$ ist die Aussage des Korollars 4.26 falsch, wie der klassische zentrale Grenzwertsatz zeigt. Der folgende Satz, den wir ohne Beweis angeben, gibt schließlich das exakte Wachstum quadratintegrierbarer Martingale an.

Satz 4.27. *(Satz vom iterierten Logarithmus)*
Es sei $(X_n)_{n\in\mathbb{N}_0}$, $X_0 := 0$, ein Martingal mit gleichmäßig beschränkten Inkrementen, d.h., $\exists M > 0$ mit $|X_k - X_{k-1}| \leq M$, insbesondere sind die X_n also quadratintegrierbar. Dann gilt:

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2V_n \log\log V_n}} = +1 \quad \textit{und} \quad \liminf_{n\to\infty} \frac{X_n}{\sqrt{2V_n \log\log V_n}} = -1$$

P-f.s. auf $\{V_\infty = \infty\}$.

Klassische Version *$Y_k := X_k - X_{k-1}$ unabhängig und identisch verteilt, $E(Y_k) = 0$, Var$(Y_k) = \sigma^2 \in]0,\infty[$, $X_n := \sum_{k=1}^n Y_k$. Dann gilt $V_n = n\sigma^2$ und*

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2\sigma^2 n \log\log n}} = +1 \quad \textit{und} \quad \liminf_{n\to\infty} \frac{X_n}{\sqrt{2\sigma^2 n \log\log n}} = -1 \quad P-\textit{f.s.}$$

5 Irrfahrten und Brownsche Bewegung

5.1 Die symmetrische Irrfahrt

Im ganzen Abschnitt seien $(X_k)_{k\in\mathbb{N}}$ unabhängig und identisch verteilt mit $P(X_k = \pm 1) = \frac{1}{2}$, $S_0 := 0$ sowie $S_n := X_1 + \ldots + X_n$, $n \geq 1$, die zugehörige Irrfahrt. Man beachte, dass in diesem Spezialfall für alle n die Verteilung des Pfades $(S_0, S_1, \ldots, S_n)$ gerade die Gleichverteilung auf der Menge

$$\Omega_n := \{(s_0, s_1, \ldots, s_n) \mid s_0 = 0, |s_k - s_{k-1}| = 1 \text{ für } 1 \leq k \leq n\}, |\Omega_n| = 2^n ,$$

ist. Wir werden diesen Sachverhalt im folgenden nutzen, um wesentliche Rechnungen im Zusammenhang mit der symmetrischen Irrfahhrt auf elementare kombinatorische Überlegungen zurückzuführen.
Wir untersuchen im folgenden die Verteilungen

(1) von S_n,

(2) des Maximums $M_n := \max\{S_0, S_1, \ldots, S_n\}$ und des Minimums $m_n := \min\{S_0, S_1, \ldots, S_n\}$.

Als wesentliches Hilfsmittel für Asymptotiken der Verteilungen für große n verwenden wir dabei im folgenden des öfteren die Stirlingsche Formel

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n , n \to \infty .$$

Hierbei bedeutet $a_n \sim b_n$ für zwei reelle Zahlenfolgen (a_n) und (b_n) deren asymptotische Äquivalenz, d.h. $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Lemma 5.1.

$$P(S_n = i) = \begin{cases} 0 & \text{für } n+i \text{ ungerade oder } |i| > n \\ 2^{-n}\binom{n}{\frac{n+i}{2}} & \text{sonst.} \end{cases}$$

Beweis. Damit $S_n = i$ gilt, müssen genau $(n \pm i)/2$ Inkremente ± 1 sein. Insbesondere muss also $n + i$ gerade und $n - i \geq 0$ sein. Es gibt nun aber aber offenbar $\binom{n}{\frac{n+i}{2}}$ viele solcher Pfade, woraus die Behauptung folgt. □

Für das Verständnis der Asymptotik beachte man, dass die Binomialkoeffizienten $\binom{n}{k}$ für $k = \frac{n}{2}$ (bzw. $k = \frac{n+1}{2}$) ihr Maximum annehmen. Insbesondere gilt

$$P(S_{2n} = i) \leq P(S_{2n} = 0) =: u_{2n} = 2^{-2n}\binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}} , n \to \infty .$$

Wir kommen nun zur Verteilung des Minimums m_n der Pfade.

Lemma 5.2. *(Reflektionsprinzip) Für alle $n \in \mathbb{N}$ und alle $i, j \in \mathbb{Z}$ mit $j \leq 0$ und $i \geq j$ gilt*

$$P(m_n \leq j, S_n = i) = P(S_n = i - 2j)\,.$$

Beweis. Wir können o.B.d.A annehmen, dass $n + i$ gerade ist, sonst sind beide Ereignisse leer. Für einen Pfad in $s \in \{m_n \leq j, S_n = i\}$ sei k der kleinste Zeitpunkt mit $s_k = j$. Spiegele nun das Pfadstück $(s_k, s_{k+1}, \ldots)$ an j und erhalte einen gespiegelten Pfad $\hat{s}$ mit $\hat{s}_m = 2j - s_m$ für $m \geq k$, insbesondere also $s_n = 2j - i$. Die so erhaltene Abbildung ist eine Bijektion zwischen den beiden Ereignissen $\{m_n \leq j, S_n = i\}$ und $\{S_n = 2j - i\}$. Die Umkehrabbildung ist dabei gegeben durch Spiegelung eines Pfades in $\{S_n = 2j - i\}$ an j ab dem Zeitpunkt des ersten Erreichens des Niveaus j. Dass das Niveau j bis zur Zeit n mindestens einmal erreicht ist, folgt aus der Tatsache, dass $j \leq i$, denn $2j - i \leq j \leq 0$. Es folgt

$$P(m_n \leq j, S_n = i) = P(S_n = 2j - i) = P(S_n = i - 2j)\,.$$

□

Der folgende Satz zur gemeinsamen Verteilung von Minimum m_n und S_n ist eine Übungsaufgabe:

Satz 5.3. *Für alle $n \in \mathbb{N}$ und alle $i, j \in \mathbb{Z}$ mit $j \leq 0$ und $i \geq j$ gelten*

$$\begin{aligned} P(m_n = j, S_n = i) &= P(S_n = i - 2j) - P(S_n = i - 2j + 2) \\ P(m_n = j) &= P(S_n \in \{j, j-1\})\,. \end{aligned}$$

Satz 5.4. *Für alle $n \in \mathbb{N}$ und alle $j \in \mathbb{Z}$ mit $j \leq 0$ gelten*

(i) $P(m_n \leq j) = P(S_n \leq j) + P(S_n \leq j - 1) = 2P(S_n \leq j) + P(S_n = j - 1)$.

(ii) $P(M_n \geq -j) = P(S_n \geq -j) + P(S_n \geq -j + 1) = 2P(S_n \geq -j) + P(S_n = -j + 1)$.

Beweis. Wir zeigen nur (i). Der Beweis von (ii) folgt aus Symmetrieeigenschaften der Verteilung von (S_n). Aus dem Reflektionsprinzip Lemma 5.2 ergibt sich

$$\begin{aligned} P(m_n \leq j) &= \sum_k P(m_n \leq j, S_n = k) \\ &= \sum_{k=-\infty}^{j} P(S_n = k) + \sum_{k=j+1}^{\infty} \underbrace{P(m_n \leq j, S_n = k)}_{=P(S_n = k - 2j)} \\ &= P(S_n \leq j) + P(S_n \geq 1 - j) = P(S_n \leq j) + P(S_n \leq j - 1)\,. \end{aligned}$$

□

5.2 Brownsche Bewegung

Motivation Im folgenden betrachten wir für unabhängig und identisch verteilte Zufallsvariablen $(X_k)_{k\in\mathbb{N}}$ mit $E(X_k) = 0$ und Var $(X_k) = 1$ die zugehörige Irrfahrt

$$S_n := X_1 + \ldots + X_n \,, n \geq 1 \,, \quad S_0 := 0 \,.$$

Wenn wir den Pfad $(S_0, S_1, S_2, \ldots)$ dieser Irrfahrt durch einen Polygonzug linear interpolieren, erhalten wir eine zufällige stetige Funktion $[0, \infty[\to \mathbb{R}$. Zum Verständnis ihrer asymptotischen Verteilung beachte man, dass aufgrund des zentralen Grenzwertsatzes

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0,1) \,.$$

Wir werden daher im folgenden den Polygonzug mit $\sqrt{n}$ standardisieren, d.h. wir betrachten

$$B_t^{(n)} := \frac{1}{\sqrt{n}} \left(S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1} \right) \,, t \geq 0 \,. \tag{5.1}$$

Allgemeiner gilt nun:

Satz 5.5. *Es sei $m \in \mathbb{N}$ und $0 = t_0 < t_1 < t_2 < \ldots < t_m$. Dann konvergiert der Zufallsvektor $(B_{t_1}^{(n)}, B_{t_2}^{(n)}, \ldots, B_{t_m}^{(n)})$ in Verteilung gegen die m-dimensionale Normalverteilung mit Erwartungswertvektor 0 und Kovarianzmatrix $(\min(t_i, t_j))_{1\leq i,j\leq m}$. Insbesondere sind die Inkremente*

$$B_{t_1}^{(n)} - B_{t_0}^{(n)}, B_{t_2}^{(n)} - B_{t_1}^{(n)}, \ldots, B_{t_m}^{(n)} - B_{t_{m-1}}^{(n)}$$

im Grenzwert $n \to \infty$ unabhängig normalverteilt mit Erwartungswert 0 und Varianz $t_i - t_{i-1}$ für $1 \leq i \leq m$.

Zur Vorbereitung auf den Beweis des Satzes zeigen wir zunächst:

Lemma 5.6. *Es sei $m \in \mathbb{N}$ und $0 = t_0 < t_1 < t_2 < \ldots < t_m$ und $B = (B_{t_1}, \ldots, B_{t_m})$ ein Zufallsvektor. Weiter sei $B_{t_0} := 0$. Dann sind äquivalent:*

(i) B ist normalverteilt mit Erwartungswertvektor 0 und Kovarianzmatrix $(\min(t_i, t_j))_{1\leq i,j\leq m}$.

(ii) Die Inkremente

$$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}}$$

sind unabhängig normalverteilt mit Erwartungswert 0 und Varianz $t_i - t_{i-1}$ für $1 \leq i \leq m$.

Beweis. Zu Vereinfachung der Notation sei $X_i := B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq m$, $X = (X_1, \ldots, X_m)$.

$(i) \Rightarrow (ii)$ Zunächst ist klar, dass

$$X = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix} B\,.$$

Insbesondere besitzt also auch X eine m-dimensionale Normalverteilung mit Erwartungswertvektor 0 und Kovarianzmatrix

$$\mathsf{Cov}(X_i, X_i) = E(X_i X_i) = E((B_{t_i} - B_{t_{i-1}})^2) = (t_i - t_{i-1})$$

und für $i \neq j$

$$\begin{aligned} \mathsf{Cov}(X_i, X_j) &= E(X_i X_j) = E((B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})) \\ &= E(B_{t_i} B_{t_j}) - E(B_{t_{i-1}} B_{t_j}) - E(B_{t_i} B_{t_{j-1}}) + E(B_{t_{i-1}} B_{t_{j-1}}) \\ &= t_i \wedge t_j - t_{i-1} \wedge t_j - t_i \wedge t_{j-1} + t_{i-1} \wedge t_{j-1} = 0\,. \end{aligned}$$

Hieraus folgt (ii).
Zum Beweis der umgekehrten Implikation $(ii) \Rightarrow (i)$ beachte, dass

$$B = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \dots & \dots & 1 \end{bmatrix} X\,.$$

Insbesondere ist also B normalverteilt mit Erwartungswertvektor 0 und Kovarianzmatrix

$$\begin{aligned} \mathsf{Cov}(B_{t_i}, B_{t_j}) &= E((X_1 + \ldots + X_i)(X_1 + \ldots + X_j)) \\ &= E(X_1^2) + E(X_2^2) + \ldots + E(X_{i\wedge j}^2) \\ &= (t_1 - t_0) + (t_2 - t_1) + \ldots + (t_i \wedge t_j - t_{i-1} \wedge t_{j-1}) = t_j \wedge t_j\,. \end{aligned}$$

□

Beweis. (von Satz 5.5) Nach dem vorangegangenen Lemma reicht es aus zu zeigen, dass die Inkremente

$$B_{t_i}^{(n)} - B_{t_{i-1}}^{(n)}\,, 1 \leq i \leq m$$

in Verteilung gegen unabhängig und normalverteilte Zufallsvariablen mit Erwartungswert 0 und Varianz $t_i - t_{i-1}$ konvergieren.

Zu gegebenem n sei $k_i^{(n)} := \lfloor t_i n \rfloor$, $1 \leq i \leq m$, $k_0^{(n)} = 0$, und $t_i^{(n)} = \frac{k_i^{(n)}}{n}$. Beachte, dass $|t_i - t_i^{(n)}| \leq \frac{1}{n}$. Dann gilt für

$$\Delta_i^{(n)} := |B_{t_i}^{(n)} - B_{t_{i-1}}^{(n)} - (B_{t_i^{(n)}}^{(n)} - B_{t_{i-1}^{(n)}}^{(n)})|$$

dass

$$P\left(\Delta_i^{(n)} \geq \varepsilon\right) \leq P\left(\frac{X_{k_i^{(n)}}}{\sqrt{n}} \geq \frac{\varepsilon}{2}\right) + P\left(\frac{X_{k_{i-1}^{(n)}}}{\sqrt{n}} \geq \frac{\varepsilon}{2}\right)$$
$$\leq \frac{4}{\varepsilon^2 n} E(X_{k_i^{(n)}}^2) + \frac{4}{\varepsilon^2 n} E(X_{k_{i-1}^{(n)}}^2) \to 0\,, n \to \infty\,.$$

Damit sind die gemeinsamen Verteilungen von

$$B_{t_i}^{(n)} - B_{t_{i-1}}^{(n)}\,, 1 \leq i \leq m \quad \text{und von} \quad B_{t_i^{(n)}}^{(n)} - B_{t_{i-1}^{(n)}}^{(n)}\,, 1 \leq i \leq m$$

asymptotisch gleich. Nun gilt aber

$$B_{t_i^{(n)}}^{(n)} - B_{t_{i-1}^{(n)}}^{(n)} = \frac{1}{\sqrt{n}} \sum_{k=k_{i-1}^{(n)}+1}^{k_i^{(n)}} X_k = \sqrt{\frac{k_i^{(n)} - k_{i-1}^{(n)}}{n}} \frac{1}{\sqrt{k_i^{(n)} - k_{i-1}^{(n)}}} \sum_{k=k_{i-1}^{(n)}+1}^{k_i^{(n)}} X_k$$
$$= \underbrace{\sqrt{t_i^{(n)} - t_{i-1}^{(n)}}}_{\to \sqrt{t_i - t_{i-1}}} \underbrace{\frac{1}{\sqrt{k_i^{(n)} - k_{i-1}^{(n)}}} \sum_{k=k_{i-1}^{(n)}+1}^{k_i^{(n)}} X_k}_{\stackrel{(d)}{\to} N(0,1)}$$

nach dem klassischen zentralen Grenzwertsatz. Weiterhin sind die standardisierten Summen

$$\frac{1}{\sqrt{k_i^{(n)} - k_{i-1}^{(n)}}} \sum_{k=k_{i-1}^{(n)}+1}^{k_i^{(n)}} X_k\,, 1 \leq i \leq m\,,$$

unabhängig für alle n. Damit konvergiert ihre gemeinsame Verteilung schwach gegen das Produktmaß

$$\bigotimes_{i=1}^{m} N(0, t_i - t_{i-1})\,.$$

Hieraus folgt die Behauptung. □

Wir haben also gezeigt, dass die endlichdimensionalen Randverteilungen des standardisierten Polygonzuges $(B_t^{(n)})_{t\geq 0}$ für $n \to \infty$ gegen die Verteilungen normalverteilter Zufallsvariablen $(B_t)_{t\geq 0}$ konvergieren. Wir werden noch zeigen, dass nicht nur die endlichdimensionalen Randverteilungen konvergieren, sondern sogar der ganze Prozess $B^{(n)}$ als (zufällige) Funktion gegen B. Aufgefasst als stochastischer Prozess nennt man $(B_t)_{t\geq 0}$ eine Brownsche Bewegung im Sinne der folgenden Definition:

Definition 5.7. Eine **Brownsche Bewegung** (BB) (mit Start in 0) ist ein $\mathbb{R}$-wertiger stochastischer Prozess $(B_t)_{t\geqslant 0}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit folgenden Eigenschaften:

(i) $B_0 = 0$ P-f.s.

(ii) Für $0 = t_0 < t_1 < \ldots < t_m$ sind die Inkremente

$$B_{t_i} - B_{t_{i-1}} \qquad (i = 1, \ldots, m)$$

unabhängig, $N(0, t_i - t_{i-1})$-verteilt.

Bemerkung 5.8. *Die Brownsche Bewegung ist benannt nach Brown (Schottischer Botaniker, 19. Jahrhundert). Er beobachtete und beschrieb erstmals die irreguläre Bewegung von in einer Lösungsflüssigkeit aufgelösten Blütenpollen - später wurde diese Bewegung durch zufällige Kollisionen mit den Molekülen der Lösungsflüssigkeit erklärt. Später wurde diese "regellose" Bewegung der Blütenpollen aufgegriffen von*

Bachelier (1900): "Sur la théorie de la spéculation"

Einstein (1905): "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von (in ruhenden Flüssigkeiten) suspendierten Teilchen"

Wiener (1923): lieferte eine erste rigorose Konstruktion eines mathematischen Modells.

Definition 5.9. Eine BB $(B_t)_{t\geqslant 0}$ heißt **stetig**, falls die Trajektorie $\underbrace{t \longmapsto B_t(\omega)}_{\text{("Pfad" oder "Trajektorie")}}$ stetig ist für alle $\omega \in \Omega$.

Der Zustandsraum einer stetigen Brownschen Bewegung ist also der Raum $C([0, \infty[)$ aller stetigen Funktionen $f : [0, \infty[\to \mathbb{R}$. Wir werden zur Vereinfachung im folgenden $B^{(n)}$ und B nur auf dem Zeitintervall $[0, 1]$ betrachten. Die Verallgemeinerung auf endliche Zeitintervalle $[0, T]$ und somit schließlich auf ganz $[0, \infty[$ ist eine leichte Übung. Als Zustandsraum einer stetigen Brownschen Bewegung bis zur Zeit 1 erhalten wir somit den Raum $C([0, 1])$ aller stetigen Funktionen $f : [0, 1] \to \mathbb{R}$. Der Raum $C([0, 1])$ ist bezüglich der Supremumsnorm

$$\|f\|_\infty := \sup_{t\in[0,1]} |f(t)| = \max_{t\in[0,1]} |f(t)|$$

ein Banachraum, d.h. ein vollständiger normierter Vektorraum.

Für $t \in [0, 1]$ bezeichne

$$\Pi_t : C([0, 1]) \to \mathbb{R}\,, f \mapsto f(t)\,,$$

die "Auswertung" der Funktion f in t. Wenn wir f als Vektor in $\mathbb{R}^{[0,1]}$ auffassen, so ist $\Pi_t(f)$ also die Projektion des Vektors f auf die t-te Koordinate. Wir betrachten nun auf $C([0, 1])$ als σ-Algebra $\mathcal{F}$ die kleinste σ-Algebra, bzgl. der alle Projektionen Π_t messbar sind, d.h.

$$\mathcal{F} := \sigma(\{\Pi_t \mid t \in [0, 1]\})\,.$$

Das folgende Lemma gibt eine Charakterisierung der σ-Algebra $\mathcal{F}$.

Lemma 5.10. *(i) $\mathcal{F} = \mathcal{B}(C([0, 1]))$, wobei $\mathcal{B}(C([0, 1]))$ die Borelsche σ-Algebra auf $C([0, 1])$ bezeichnet.*

(ii) Ist $(B_t)_{t\in[0,1]}$ eine stetige Brownsche Bewegung bis zur Zeit 1 auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, so ist die Abbildung

$$\Phi : \Omega \to C([0,1])\,, \omega \mapsto (t \mapsto B_t(\omega))\,, \quad \mathcal{A}/\mathcal{F}-\textit{messbar}.$$

Beweis. (i) π_t ist stetig, also $\mathcal{F} \subseteq \mathcal{B}(C([0,1]))$.
Für den Beweis der umgekehrten Inklusion $\mathcal{B}(C([0,1])) \subseteq \mathcal{F}$ sei $U \subseteq C([0,1])$ offen. Da $C([0,1])$ separabler metrischer Raum (benutze zum Beispiel den Approximationssatz von Weierstraß), gibt es eine abzählbare Teilmenge $(f_n)_n \subseteq C([0,1])$ mit

$$U = \bigcup_n \overline{B}_{\epsilon_n}(f_n) \quad \text{wobei } \overline{B}_{\epsilon_n}(f_n) = \{g \in C([0,1]) : \|g - f_n\| \le \epsilon_n\}$$

den abgeschlossenen ϵ_n-Ball um f_n bezeichnet. Damit ist der Beweis der umgekehrten Inklusion reduziert auf den Fall

$$\overline{B}_\epsilon(f) \in \mathcal{F}\,.$$

Dieser folgt aber aus

$$\overline{B}_\epsilon(f) = \bigcap_{t\in\mathbb{Q}\cap[0,1]} \{g : |g(t) - f(t)| \le \epsilon\}\,.$$

(ii) Das System aller Mengen

$$\bigcap_{i=1}^m \{f : f(t_i) \in A_i\}$$

für $0 \leqslant t_1 \leqslant \ldots \leqslant t_m \leqslant 1$ und $A_i \in \mathcal{B}(\mathbb{R})$ ist ein Erzeugendensystem von $\mathcal{F}$. Da

$$\Phi^{-1}\left(\bigcap_{i=1}^m \{f : f(t_i) \in A_i\}\right) = \bigcap_{i=1}^m \underbrace{\{\omega : B_{t_i}(\omega) \in A_i\}}_{\in\mathcal{A}}\,,$$

folgt die Behauptung. □

Wienermaß

Definition 5.11. Für eine stetige Brownsche Bewegung $(B_t)_{t\in[0,1]}$ bis zur Zeit 1, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, heißt das nach Lemma 5.10 wohldefinierte Bildmaß $W := P \circ \Phi^{-1}$ von P unter der Transformation Φ auf $(C([0,1]), \mathcal{F})$ **Wienermaß**, $(C([0,1]), \mathcal{F}, W)$ heißt **(klasssischer) Wienerraum**, und $(C([0,1]), \mathcal{F}, (\Pi_t)_{t\in[0,1]}, W)$ **kanonisches Modell der BB.**

Wir kommen nun auf die Konvergenz der reskalierten Irrfahrten $B^{(n)}$ zurück und formulieren das Hauptergebnis dieses Kapitels:

Theorem 5.12. *(Donskers Invarianzprinzip) Es sei $S_n := X_1 + \ldots + X_n$, $n \geq 1$, $S_0 := 0$, eine Irrfahrt auf $\mathbb{R}$ mit unabhängig und identisch verteilten Inkrementen $(X_k)_{kge1}$ mit $E(X_k) = 0$ und Var$(X_k) = 1$. Weiter sei $(B_t^{(n)})_{t\in[0,1]}$ die in* (5.1) *definierte (lineare Interpolation der) reskalierte(n) Irrfahrt. Dann folgt*

$$\lim_{n\to\infty} P \circ (B^{(n)})^{-1} = W \ \textit{schwach}\,,$$

d.h., die Verteilung der reskalierten Irrfahrt konvergiert schwach gegen die Verteilung einer stetigen Brownschen Bewegung.

Bemerkung 5.13. *Donskers Invarianzprinzip 5.12 heißt so, weil der Grenzwert der Verteilungen der reskalierten Irrfahrten, also das Wienermaß W, unabhängig ist von den Verteilungen der Inkremente X_k, solange diese standardisiert sind. In diesem Sinne ist die Aussage des Invarianzprinzips analog zur Aussage des klassischen zentralen Grenzwertsatzes. Man nennt deshalb auch Satz 5.12 einen* **funktionalen zentralen Grenzwertsatz**. *Wir können insbesondere analog zur Normalapproximation der Verteilung der Summe unabhängig und identisch verteilter Zufallsvariablen, das Wienermaß zur Approximation der Verteilung gewisser Funktionale einfacher Irrfahrten verwenden (siehe Satz 5.21 unten). Weitere wichtige Anwendungen sind die numerische Approximation stochastischer Differentialgleichungen mit Hilfe von Irrfahrten (siehe Vorlesungen zur stochastischen Analysis) oder die Approximation der Black-Scholes Preisformel für Finanzderivate durch Verteilungen gewisser Funktionale im Cox-Ross-Rubinstein Modell (siehe Vorlesungen zur Finanzmathematik).*

Zum Beweis des Invarianzprinzips von Donsker müssen wir zunächst die schwache Konvergenz von Wahrscheinlichkeitsmaßen auf $C([0,1])$ besser verstehen. Dazu formulieren wir im folgenden Lemma zunächst eine notwendige Bedingung:

Lemma 5.14. *Es seien μ_n, $n \geq 1$, und μ Wahrscheinlichkeitsmaße auf $(C([0,1]), \mathcal{B}(C([0,1])))$ mit $\mu_n \to \mu$ schwach. Dann gilt für alle $J \subseteq [0,1]$ endlich*

$$\lim_{n\to\infty} \mu_n^J = \mu^J \text{ auf } \mathbb{R}^{|J|} \text{ schwach}.$$

Hierbei bezeichnet μ^J die gemeinsame Verteilung von Π_t, $t \in J$, auf $\mathbb{R}^{|J|}$ (versehen mit der Borelschen σ-Algebra.

Beweis. Es sei $|J| = m$ und $J = \{t_1, \ldots t_m\}$ eine Aufzählung der Elemente in J. Weiter sei $f : \mathbb{R}^m \to \mathbb{R}$ stetig und beschränkt. Dann ist die auch die Komposition

$$f \circ \Pi^J : C([0,1]) \to \mathbb{R}\,, \omega \mapsto f(\omega(t_1), \ldots, \omega(t_n))\,,$$

stetig auf $C([0,1])$ (und beschränkt). Damit ergibt sich

$$\lim_{n\to\infty} \int f\, d\mu_n^J = \lim_{n\to\infty} \int f \circ \Pi^J\, d\mu_n = \int f \circ \Pi^J\, d\mu = \int f\, d\mu^J\,.$$

Hieraus folgt die Behauptung. □

Beispiel 5.15. Das folgende Beispiel zeigt, dass umgekehrt die Konvergenz der endlichdimensionalen Randverteilungen von μ_n gegen die endlichdimensionalen Randverteilungen von μ nicht hinreichend für schwache Konvergenz ist. Dazu betrachte zu $n \in \mathbb{N}$ die stetige Funktion

$$f_n(t) := \begin{cases} n(t - \frac{1}{n}) \wedge (\frac{2}{n} - t) & \text{für } t \in [0, \frac{2}{n}] \\ 0 & \text{sonst.} \end{cases}$$

Weiterhin sei $\mu_n := \delta_{f_n}$ die Einpunktverteilung in f_n. Offenbar folgt dann für alle $J \subset [0,1]$ endlich, dass $\mu_n^J \to \delta_0$ konvergiert. M.a.W. die endlichdimensionalen

Randverteilungen von μ_n konvergieren gegen die endlichdimensionalen Randverteilungen der Einpunktverteilung δ_0 in der 0-Funktion. Dennoch gilt nicht $\lim_{n\to\infty} \delta_{f_n} = \delta_0$, denn z.B. für die gestutzte Normfunktion $\|\cdot\|_\infty \wedge 1$ (stetig und beschränkt auf $C([0,1])$) gilt

$$\int \|\cdot\|_\infty \wedge 1 \, d\delta_{f_n} = \|f_n\|_\infty \wedge 1 = 1 \quad \text{aber} \quad \int \|\cdot\|_\infty \wedge 1 \, d\delta_0 = 0\,.$$

Der folgende Satz gibt eine Charakterisierung der schwachen Konvergenz, die in der Theorie stochastischer Prozesse von fundamentaler Bedeutung ist:

Satz 5.16. *Es seien μ_n, $n \geq 1$, und μ Wahrscheinlichkeitksmaße auf $(C([0,1]), \mathcal{B}(C([0,1])))$. Dann sind äquivalent:*

(i) $\lim_{n\to\infty} \mu_n = \mu$ *schwach.*

(ii) *(a) Die endlichdimensionalen Randverteilungen von μ_n konvergieren schwach gegen die endlichdimensionalen Randverteilungen von μ und*

(b) Die Menge $\{\mu_n \mid n \geq 1\}$ ist **straff**, *d.h. für allle $\varepsilon > 0$ gibt es eine kompakte Teilmenge $K_\varepsilon \subset C([0,1])$ mit $\mu_n(K_\varepsilon) \geq 1 - \varepsilon$ für alle n.*

Bemerkung 5.17. *Die Straffheit der Folge $\{\mu_n \mid n \geq 1\}$ bedeutet also nichts anderes, als dass bis auf beliebig kleines $\varepsilon > 0$, alle Wahrscheinlichkeitsmaße μ_n dieselbe kompakte Menge K_ε als Träger besitzt. Aus diesem Grunde ist es hilfreich, die kompakten Teilmengen in $C([0,1])$ zu verstehen. Dazu dient der folgende Satz von Arzelà-Ascoli.*

Satz 5.18. *Eine Teilmenge $K \subset C([0,1])$ ist relativ kompakt, d.h. der (topologische) Abschluss $\overline{K}$ von K in $C([0,1])$ ist kompakt, genau dann wenn gilt:*

(a) K *ist* **punktweise beschränkt in** 0, *d.h.* $\sup_{f\in K} |f(0)| < \infty$.

(b) K *ist* **gleichgradig gleichmäßig stetig**, *d.h.* $\limsup_{\delta\downarrow 0} \sup_{f\in K} w_\delta(f) = 0$, *wobei*

$$w_\delta(f) := \sup_{\substack{s,t\in[0,1] \\ |s-t|<\delta}} |f(s) - f(t)|\,.$$

Einen Beweis für diesen Satz findet man in Lehrbüchern zur Funktionalanalysis. Mithilfe des Satzes von Arzelà-Ascoli können wir nun folgendes Straffheitskriterium angeben:

Satz 5.19. *Es sei $\mathcal{M}$ eine beliebige Teilmenge von Wahrscheinlichkeitsmaßen auf $(C([0,1]), \mathcal{B}(C([0,1])))$ mit*

(a) $\mu(\{f \mid f(0) = 0\}) = 1$ *für alle* $\mu \in \mathcal{M}$, *und*

(b) $\lim_{\delta\downarrow 0} \sup_{\mu\in\mathcal{M}} \mu(w_\delta \geq \varepsilon) = 0$ *für alle* $\varepsilon > 0$.

Dann ist die Menge $\mathcal{M}$ straff.

Beweis. Es sei $\varepsilon > 0$ fest. Nach Annahme gibt es eine monoton fallende Folge $(\delta_k)_{k\in\mathbb{N}}$ mit

$$\sup_{\mu\in\mathcal{M}} \mu\left(w_{\delta_k} > \frac{1}{k}\right) \leq \varepsilon 2^{-k}\,.$$

Setze nun

$$K_\varepsilon := \{f \mid f(0) = 0\} \cap \left(\bigcap_{k\geq 1} \{f : w_{\delta_k}(f) \leq \frac{1}{k}\}\right)\,.$$

Dann ist die Menge $K_\varepsilon \subset C([0,1])$ relativ kompakt nach dem Kriterium von Arzelà-Ascoli und für $\mu \in \mathcal{M}$ gilt nach Konstruktion

$$\mu(K_\varepsilon) \geq 1 - \sum_{k=1}^{\infty} \mu\left(\{f \mid w_{\delta_k}(f) > \frac{1}{k}\}\right) \geq 1 - \varepsilon\,.$$

□

Anwendung auf den Beweis von 5.12
Zur Vereinfachung der Notation sei $\mu_n := P \circ (B^{(n)})^{-1}$.

Satz 5.20. *Die Folge* $\{\mu_n \mid n \geq 1\}$ *ist straff.*

Beweis. Wir wollen Satz 5.19 anwenden. Da $B_0^{(n)} = 0$ nach Konstruktion, folgt

$$\mu_n(\{f \mid f(0) = 0\}) = 1 \quad \forall n\,.$$

Es bleibt also noch zu zeigen, dass für alle $\varepsilon > 0$

$$\limsup_{n\to\infty} \mu_n\left(w_\delta > \varepsilon\right) = 0\,.$$

Dazu sei $\delta > 0$ fest gewählt und $n \in \mathbb{N}$ so dass $\frac{1}{n} < \delta$. Weiter sei $k_n \in \mathbb{N}$ minimal mit $\delta_n := \frac{k_n}{n} \geq \delta$, also $\delta_n \leq \delta + \frac{1}{n} \leq 2\delta$.

Für $s, t \in [0,1]$ mit $|s-t| < \delta$ gibt es dann $l \in \mathbb{N}$ mit $(l-1)\delta_n \leq s \wedge t \leq s \vee t < l\delta_n$ und somit

$$\begin{aligned}
|B_s^{(n)} - B_t^{(n)}| &\leq |B_{s\wedge t}^{(n)} - B_{l\delta_n}^{(n)}| + |B_{l\delta_n}^{(n)} - B_{s\vee t}^{(n)}| \\
&\leq 2|B_{l\delta_n}^{(n)} - B_{(l-1)\delta_n}^{(n)}| + |B_{l\delta_n}^{(n)} - B_{(l+1)\delta_n}^{(n)}| \\
&\leq \frac{3}{\sqrt{n}} \max_{l:l\delta_n\leq 1} \underbrace{\max_{(l-1)k_n < m \leq lk_n} |S_m - S_{(l-1)k_n}|}_{=:M_l^{(n)}}\,.
\end{aligned}$$

Es folgt

$$\mu_n\left(w_\delta > \varepsilon\right) \leq P\left(\max_{l:l\delta_n\leq 1} M_l^{(n)} > \frac{\varepsilon\sqrt{n}}{3}\right) \leq \frac{1}{\delta_n} P\left(M_1^{(n)} > \frac{\varepsilon\sqrt{n}}{3}\right)$$

da $M_l^{(n)}$, $l = 1, 2, 3, \ldots$, identisch verteilt. Aus der Maximalungleichung für Martingale 4.5 folgt

$$\begin{aligned} P\left(M_1^{(n)} > \frac{\varepsilon\sqrt{n}}{3}\right) &\leq \frac{3}{\varepsilon\sqrt{n}} E\left(|S_{k_n}|\right) \leq \frac{3}{\varepsilon\sqrt{n}} \left(E\left(|S_{k_n}|^2\right)\right)^{\frac{1}{2}} \\ &= \frac{3}{\varepsilon}\sqrt{\frac{k_n}{n}} = \frac{3}{\varepsilon}\sqrt{\delta_n} \leq \frac{3}{\varepsilon}\sqrt{2\delta} \to 0 \end{aligned}$$

für $\delta \to 0$. □

Der Beweis von Donskers Invarianzprinzip 5.12 folgt nun aus Satz 5.16 in Verbindung mit der Konvergenz der endlichdimensionalen Randverteilungen Satz 5.5 und der Straffheit aus dem letzten Satz 5.20.

Anwendung

Es sei $F(\omega) := \max_{t \in [0,1]} \omega(t)$, $\omega \in C([0,1])$. F ist offensichtlich stetig und damit konvergiert die Verteilung ν_n von F unter μ_n schwach gegen die Verteilung von F unter dem Wienermaß W.

Wir betrachten nun speziell im Satz von Donsker die einfache symmetrische Irrfahrt mit Inkrementen X_k mit $P(X_k = \pm 1) = \frac{1}{2}$. Dann gilt aufgrund des Reflektionsprinzips für die einfache Irrfahrt Satz 5.2

$$\begin{aligned} \mu_n(F \geq \alpha) &= P\left(\max_{1 \leq k \leq n} \frac{S_k}{\sqrt{n}} \geq \alpha\right) = P\left(\max_{1 \leq k \leq n} S_k \geq \alpha\sqrt{n}\right) \\ &= 2P\left(S_n \geq \alpha\sqrt{n}\right) + P\left(S_n = \lceil \alpha\sqrt{n} \rceil - 1\right) \\ &= P\left(\frac{|S_n|}{\sqrt{n}} \geq \alpha\right) + P\left(S_n = \lceil \alpha\sqrt{n} \rceil - 1\right) . \end{aligned}$$

Der Zentrale Grenzwertsatz impliziert nun, dass

$$\lim_{n \to \infty} P\left(\frac{|S_n|}{\sqrt{n}} \geq \alpha\right) = P\left(|Z| \geq \alpha\right)$$

wobei Z eine $N(0,1)$-verteilte Zufallsvariable ist, zum Beispiel B_1. Weiterhin gilt

$$P\left(S_n = \lceil \alpha\sqrt{n} \rceil - 1\right) \leq \frac{const}{\sqrt{n}} \to 0\,, n \to \infty\,.$$

Wir haben damit folgenden Satz beweisen:

Satz 5.21. *Es sei $(B_t)_{t \in [0,1]}$ eine stetige Brownsche Bewegung. Dann gilt:*

$$P\left(\max_{t \in [0,1]} B_t \geq \alpha\right) = P\left(|B_1| \geq \alpha\right)\,, \alpha \in \mathbb{R}\,.$$

Arkussinus Gesetz der Brownschen Bewegung

Satz 5.22. *Es sei* $T := \sup\{t \leqslant 1 : B_t = 0\}$ *der Zeitpunkt des letzten Besuches in* 0 *(keine Stoppzeit!). Dann gilt für* $t \in [0,1]$*:*

$$P(T \leqslant t) = \frac{2}{\pi} \arcsin(\sqrt{t}) .$$

Beweis. Es sei $\tilde{B}$ eine weitere Brownsche Bewegung unabhängig von B. Aus dem Reflektionsprinzip ergibt sich:

$$\begin{aligned}
P(T \leqslant t) &= P(B_s \neq 0 \ \forall s \in [t,1]) \\
&= \int_{-\infty}^{+\infty} P(B_s \neq 0 \ \forall s \in [t,1] \mid B_t = x) \underbrace{P(B_t \in dx)}_{=N(0,t)(dx)} \\
&= \int_{-\infty}^{+\infty} \underbrace{P(\tilde{B}_s \geq -|x| \ \forall s \in [t,1])}_{=P(\tilde{B}_s \leq |x| \ \ \forall s \in [t,1])} P(B_t \in dx) \\
&= \int_{-\infty}^{+\infty} P(|\tilde{B}_{1-t}| \leq |x|) \, P(B_t \in dx) \\
&= P(|\tilde{B}_{1-t}| \leqslant |B_t|) .
\end{aligned}$$

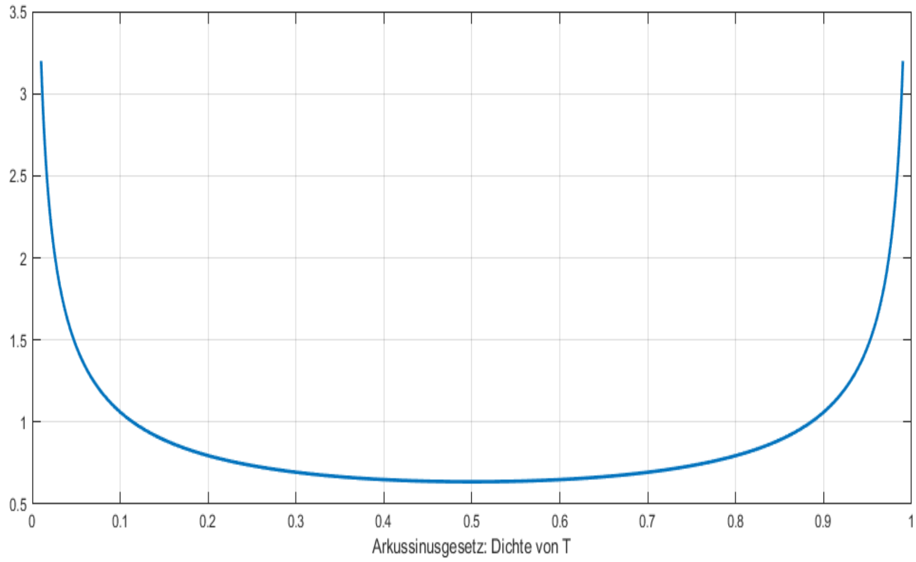
Nun sind $|\tilde{B}_{1-t}|$ und $|B_t|$ gemeinsam verteilt wie $\sqrt{1-t}|Y|$ und $\sqrt{t}|X|$ für zwei unabhängig $N(0,1)$-verteilte Zufallsvariablen X und Y. Daher folgt

$$\begin{aligned}
P(|\tilde{B}_{1-t}| \leqslant |B_t|) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} 1_{\{(1-t)y^2 \leqslant tx^2\}} \, dx \, dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} 1_{\{y^2 \leqslant t(x^2+y^2)\}} \, dx \, dy \\
&= \frac{1}{2\pi} \int_0^{\infty} r \, dr e^{-\frac{r^2}{2}} \int_0^{2\pi} d\varphi 1_{\{\sin(\varphi)^2 \leq t\}} = \frac{2}{\pi} \arcsin(\sqrt{t}) .
\end{aligned}$$

Also ist

$$\begin{aligned}
P(a \leq T \leq b) &= \frac{2}{\pi} \arcsin(\sqrt{b}) - \frac{2}{\pi} \arcsin(\sqrt{a}) \\
&= \frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} \, dx .
\end{aligned}$$

□



Interpretation S_n bezeichne den Spielstand in einem fairen Spiel für zwei Spieler A und B, wobei das Ereignis $X_i = \pm 1$ dafür steht, dass Spieler A das i-te Spiel gewinnt/verliert. Dann bedeutet $S_k > 0$, $k = m, m+1, \ldots, n$, dass Spieler A in den Spielen $m, m+1, \ldots, n$ in Führung liegt. Nun gilt aber

$$\{S_k > 0 : k = m, m+1, \ldots, n\} = \{B_s^{(n)} > 0 : s \geq \frac{km}{n}\}$$

und

$$P(B_s^{(n)} > 0 : s \geq \frac{km}{n}) \to P(T \leq t) \text{ für } \frac{km}{n} \to t\,.$$

Also bedeutet $T \ll 1$, dass Spieler A die meiste Zeit in Führung liegt und das Arkussinus Gesetz besagt nun, dass solche Spielverläufe viel wahrscheinlicher als ausgeglichene Spielverläufe sind.

Literaturverzeichnis

[Ba91] H. Bauer, Maß- und Integrationstheorie, 2. Auflage, Walter de Gruyter, Berlin, 1991.

[Ba02] H. Bauer, Wahrscheinlichkeitstheorie, 5. Auflage, Walter de Gruyter, Berlin, 2002.

[Bi12] P. Billingsley, Probability and Measure, Anniversary Edition, Wiley, New York, 2012.

[Bi99] P. Billingsley, Convergence of probability measures, Wiley, New York, 1999.

[Ch78] K.L. Chung, Elementare Wahrscheinlichkeitstheorie und stochastische Prozesse, Springer, Berlin, 1978.

[Du02] R.M. Dudley, Real analysis and probability, Cambridge University Press, 2002.

[Dur10] R. Durrett, Probability: Theory and Examples, 4. Auflage, Cambridge University Press, 2010.

[El05] J. Elstrodt, Maß- und Integrationstheorie, Springer, Berlin, 2005.

[Fe68] W. Feller, An Introduction to Probability Theory and its Applications, Vol. I, 3. Auflage, Wiley, New York, 1968.

[Fe71] W. Feller, An Introduction to Probability Theory and its Applications, Vol. II, Wiley, New York, 1971.

[Kl08] A. Klenke, Wahrscheinlichkeitstheorie, 2. Auflage, Springer, Berlin, 2008.

[No97] J. Norris, Markov Chains, Cambridge University Press, 1997.

[Sh96] A.N. Shiryaev, Probability, Springer, Berlin, 1996.

[St13] W. Stannat, Wahrscheinlichkeitstheorie I, *VL-Skript, TU Berlin*, 2015.