Wahrscheinlichkeitstheorie und diskrete Finanzmathematik

Vorlesungsskript

Benedikt Jahnel

Technische Universität Braunschweig

27. Oktober 2026

Inhaltsverzeichnis

I. Wahrscheinlichkeitstheorie **4**

0. Wiederholung **5**
0.1. Wahrscheinlichkeitsräume . . . 5
0.2. Zufallsvariablen und Verteilungen . . . 5
0.3. Unabhängigkeit und Unkorreliertheit . . . 7
0.4. Elementare bedingte Wahrscheinlichkeiten . . . 8

1. Charakteristische Funktionen **10**

2. Konvergenz von Zufallsvariablen **22**
2.1. Konvergenzbegriffe . . . 22
2.2. Beweise des Stetigkeitssatzes und des Satzes von Cramér–Wold . . . 41

3. L^p-Räume **43**
3.1. Konvexe Funktionen und die Jensen-Ungleichung . . . 43
3.2. Die Räume L^p . . . 45
3.3. Konvergenz im p-ten Mittel . . . 47

4. Ergodentheorie **49**
4.1. Maßerhaltende Abbildungen und stationäre Folgen . . . 49
4.2. Invariante Mengen, Ergodizität und Mischung . . . 50
4.3. Die Ergodensätze . . . 53

5. Bedingte Erwartungswerte und Martingale **56**
5.1. Bedingte Erwartungswerte . . . 56
5.2. Filtrationen und stochastische Folgen . . . 62
5.3. Martingale . . . 63
5.4. Diskrete stochastische Integrale . . . 65
5.5. Doob-Zerlegung und quadratische Variation . . . 66

6. Optional Sampling und Optional Stopping **69**
6.1. Stoppzeiten . . . 69
6.2. Die τ-Vergangenheit . . . 70
6.3. Optional Sampling . . . 71
6.4. Gestoppte Prozesse und Optional Stopping . . . 72
6.5. Gleichgradige Integrierbarkeit . . . 74
6.6. Optional Sampling für unbeschränkte Stoppzeiten . . . 77

7. Martingalkonvergenzsätze **79**
7.1. Doobsche Ungleichungen . . . 79
7.2. Der Martingalkonvergenzsatz von Doob . . . 80

II. Diskrete Finanzmathematik **85**

8. Das Einperiodenmodell **86**
8.1. Modell, Portfolios und Arbitrage 86
8.2. Der erste Hauptsatz der Preistheorie 88
8.3. Preisbestimmung von Kontrakten 91
8.3.1. Vollständigkeit und arbitragefreie Preise 94
8.4. Der zweite Hauptsatz der Preistheorie 98

9. Das Mehrperiodenmodell **101**
9.1. Modell, Handelsstrategien und Arbitrage 101
9.2. Der erste Hauptsatz der Preistheorie 103
9.3. Preisbestimmung europäischer Kontrakte 104
9.4. Das Cox-Ross-Rubinstein-Modell 105
9.5. Vom CRR-Modell zur Black–Scholes-Formel 107
9.5.1. Die Folge der CRR-Modelle 108
9.5.2. Konvergenz der Optionspreise 109

Literatur

Die Vorlesung orientiert sich für den wahrscheinlichkeitstheoretischen Teil an [Kle13], [Bau02] und [Dur19], für den finanzmathematischen Teil an [FS02]. Eine elementare Einführung in die Stochastik, an die Kapitel 0 anknüpft, bietet [Geo15]. Zur Konvergenz von Wahrscheinlichkeitsmaßen ist [Bil99] das Standardwerk, zu Martingalen empfehlen wir [Wil91], zur Ergodentheorie [Wal82]. Für den finanzmathematischen Teil eignen sich ergänzend [Shr04] als elementare Darstellung des Binomialmodells und [DS06] für die allgemeine Arbitragetheorie. Hinweise auf Originalarbeiten finden sich an den entsprechenden Stellen im Text.

Teil I.

Wahrscheinlichkeitstheorie

0. Wiederholung

Eine ausführliche Darstellung des hier wiederholten Stoffs findet man etwa in [Geo15] sowie in [Kle13, Kapitel 1–5].

In diesem Kapitel stellen wir die Grundbegriffe aus der einführenden Stochastik zusammen, die wir im Folgenden ständig verwenden.

0.1. Wahrscheinlichkeitsräume

Definition 0.1 (Wahrscheinlichkeitsraum)**.** Ein *Wahrscheinlichkeitsraum* ist ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$, bestehend aus einer nichtleeren Menge Ω, dem *Ergebnisraum*, einer σ-Algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$, wobei $\mathcal{P}(\Omega)$ die Menge aller Teilmengen von Ω bezeichnet, und einem Wahrscheinlichkeitsmaß $\mathbb{P}\colon \mathcal{F} \to [0,1]$. Die Elemente von $\mathcal{F}$ heißen *Ereignisse*.

0.2. Zufallsvariablen und Verteilungen

Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\Phi, \mathcal{G})$ ein Messraum. Der Wahrscheinlichkeitsraum bleibt dabei oft abstrakt im Hintergrund; konkret arbeiten wir meist auf dem Messraum $(\Phi, \mathcal{G})$. Typische Beispiele sind $\Phi = \mathbb{N}_0$ mit der Potenzmenge $\mathcal{G} = \mathcal{P}(\mathbb{N}_0)$ sowie $\Phi = \mathbb{R}$ mit der Borel-σ-Algebra $\mathcal{G} = \mathcal{B}(\mathbb{R})$.

Definition 0.2 (Zufallsvariable)**.** Eine Abbildung $X\colon \Omega \to \Phi$ heißt *Zufallsvariable* (ZV), falls X messbar ist, das heißt, falls

$$X^{-1}(A) = \{\omega \in \Omega\colon X(\omega) \in A\} \in \mathcal{F} \qquad \text{für alle } A \in \mathcal{G}.$$

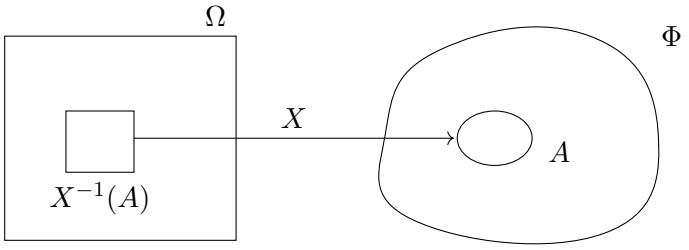

Abbildung 1.: Urbild einer Menge $A \in \mathcal{G}$ unter einer Zufallsvariablen X.

Definition 0.3 (Verteilung)**.** Die *Verteilung* einer Zufallsvariablen $X\colon \Omega \to \Phi$ ist das Wahrscheinlichkeitsmaß $\mathbb{P}^X$ auf $(\Phi, \mathcal{G})$, definiert durch

$$\mathbb{P}^X(B) \coloneqq \mathbb{P}\big(X^{-1}(B)\big) \qquad \text{für alle } B \in \mathcal{G}.$$

Beispiel 0.4 (Poisson-Verteilung)**.** Seien $\Phi = \mathbb{N}_0$, $\mathcal{G} = \mathcal{P}(\mathbb{N}_0)$ und $\lambda > 0$. Eine Zufallsvariable X heißt *Poisson-verteilt* mit Parameter λ, kurz $X \sim \mathrm{Poi}_\lambda$, falls

$$\mathbb{P}(X = k) = \mathbb{P}^X(\{k\}) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} \qquad \text{für alle } k \in \mathbb{N}_0.$$

Die Wahrscheinlichkeiten der Elementarereignisse $\{k\}$ legen $\mathbb{P}^X$ eindeutig fest; die Funktion $k \mapsto \mathbb{P}(X = k)$ heißt *Zähldichte*.

(a) Es handelt sich um eine Wahrscheinlichkeitsverteilung, denn

$$\sum_{k\in\mathbb{N}_0} \mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \sum_{k\in\mathbb{N}_0} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1.$$

(b) Für den *Erwartungswert* von X gilt

$$\mathbb{E}[X] = \sum_{k\in\mathbb{N}_0} k\,\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \sum_{k\geq 1} k\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \sum_{k\geq 1} \frac{\lambda^k}{(k-1)!} = \lambda\,\mathrm{e}^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!} = \lambda.$$

Für Zufallsvariablen mit Werten in $\mathbb{N}_0$ gilt allgemein auch $\mathbb{E}[X] = \sum_{k\in\mathbb{N}} \mathbb{P}(X \geq k)$.

(c) Für die *Varianz* von X gilt

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda.$$

Dies folgt wie in (b) aus $\mathbb{E}[X(X-1)] = \lambda^2$, also $\mathbb{E}[X^2] = \lambda^2 + \lambda$.

(d) *Anwendung: Modellierung seltener Ereignisse.* Wir betrachten schwere Thrombose-Fälle nach einer Impfung. Seien n die Anzahl der Impfungen (groß) und p das Thrombose-Risiko pro Impfung (klein). Dann ist die Anzahl der Thrombose-Fälle $\mathrm{Bin}(n, p)$-verteilt, also binomialverteilt. Für $n \to \infty$ und $p = p_n$ mit $np_n = \lambda$, wobei λ die erwartete Anzahl der Fälle ist, gilt

$$\mathrm{Bin}(n, p_n)(\{k\}) \longrightarrow \mathrm{Poi}_\lambda(\{k\}) \qquad \text{für alle } k \in \mathbb{N}_0.$$

Die Poisson-Verteilung approximiert also die Anzahl der Fälle.

Beispiel 0.5 (Exponentialverteilung)**.** Seien $\Phi = [0, \infty)$, $\mathcal{G} = \mathcal{B}([0, \infty))$ und $\lambda > 0$. Eine Zufallsvariable X heißt *exponentialverteilt* mit Parameter λ, kurz $X \sim \mathrm{Exp}_\lambda$, falls

$$\mathbb{P}(X \leq t) = \int_0^t \lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = 1 - \mathrm{e}^{-\lambda t} \qquad \text{für alle } t \geq 0.$$

Die Abbildung $t \mapsto \mathbb{P}(X \leq t)$ heißt *Verteilungsfunktion*; sie ist hier durch die *Dichte* $f(x) = \lambda\mathrm{e}^{-\lambda x}$ gegeben.

(a) Es gilt $\mathbb{P}(X \geq 0) = \int_0^\infty \lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \left[-\mathrm{e}^{-\lambda x}\right]_0^\infty = 0 + 1 = 1.$

(b) Mit partieller Integration folgt

$$\mathbb{E}[X] = \int_0^\infty x\lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}.$$

(c) Für die Varianz gilt

$$\mathrm{Var}[X] = \int_0^\infty (x - \mathbb{E}[X])^2\lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda^2}.$$

Hier ist die Varianz direkt über ihre Definition $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$ berechnet.

0.3. Unabhängigkeit und Unkorreliertheit

Definition 0.6 (Unabhängigkeit von Ereignissen)**.** Zwei Ereignisse $A, B \in \mathcal{F}$ heißen *stochastisch unabhängig* (stu), falls

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

Bemerkung 0.7. Disjunktheit und Unabhängigkeit sind verschiedene Begriffe. Für disjunkte Ereignisse A und B gilt $\mathbb{P}(A \cap B) = 0$; sie sind also nur dann unabhängig, wenn $\mathbb{P}(A) = 0$ oder $\mathbb{P}(B) = 0$ gilt.

Definition 0.8 (Unabhängigkeit von Zufallsvariablen)**.** Seien $X\colon \Omega \to \Phi$ und $Y\colon \Omega \to \Phi'$ Zufallsvariablen mit Werten in Messräumen $(\Phi, \mathcal{G})$ und $(\Phi', \mathcal{G}')$. Sie heißen *stochastisch unabhängig*, falls

$$\mathbb{P}(X \in A,\, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B) \qquad \text{für alle } A \in \mathcal{G},\ B \in \mathcal{G}'.$$

Bemerkung 0.9. In zwei wichtigen Spezialfällen genügt es, die Produktformel auf einer kleineren Klasse von Mengen nachzuprüfen.

(a) Nehmen X und Y nur Werte in $\mathbb{N}_0$ an, so sind sie genau dann unabhängig, wenn

$$\mathbb{P}(X = k,\, Y = \ell) = \mathbb{P}(X = k)\,\mathbb{P}(Y = \ell) \qquad \text{für alle } k, \ell \in \mathbb{N}_0.$$

(b) Nehmen X und Y Werte in $\mathbb{R}$ an, so sind sie genau dann unabhängig, wenn

$$\mathbb{P}(X \le t,\, Y \le s) = \mathbb{P}(X \le t)\,\mathbb{P}(Y \le s) \qquad \text{für alle } t, s \in \mathbb{R}.$$

Definition 0.10 (Kovarianz und Korrelation)**.** Seien X und Y reellwertige Zufallsvariablen mit $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$. Die *Kovarianz* von X und Y ist

$$\operatorname{Cov}(X, Y) \coloneqq \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Insbesondere gilt $\operatorname{Cov}(X, X) = \operatorname{Var}[X]$. Ist zusätzlich $\operatorname{Var}[X], \operatorname{Var}[Y] > 0$, so ist die *Korrelation* von X und Y definiert als

$$\operatorname{Corr}(X, Y) \coloneqq \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} \in [-1, 1].$$

Die Zufallsvariablen X und Y heißen *unkorreliert*, falls $\operatorname{Cov}(X, Y) = 0$.

Die Kovarianz ist symmetrisch, $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$, und nach der Cauchy–Schwarz-Ungleichung gilt $|\operatorname{Cov}(X, Y)| \le \sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}$, also liegt die Korrelation tatsächlich in $[-1, 1]$.

Satz 0.11. *Seien X und Y reellwertige Zufallsvariablen mit $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$. Sind X und Y stochastisch unabhängig, so sind sie unkorreliert.*

Die Umkehrung von Satz 0.11 ist falsch.

Beispiel 0.12. Seien $X \sim \mathcal{U}(-1, 1)$ gleichverteilt und $Y \coloneqq X^2$. Dann gilt $\operatorname{Cov}(X, Y) = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = 0$, da X symmetrisch verteilt ist. Andererseits gilt $\mathbb{P}(|X| \le 1/2,\, Y > 1/4) = 0 \neq 1/4 = \mathbb{P}(|X| \le 1/2)\,\mathbb{P}(Y > 1/4)$, also sind X und Y nicht unabhängig.

Satz 0.13 (Gleichung von Bienaymé)**.** *Seien $X_1, \dots, X_n$ paarweise unkorrelierte reellwertige Zufallsvariablen mit $\mathbb{E}[X_i^2] < \infty$, das heißt, es gilt $\operatorname{Cov}(X_i, X_j) = 0$ für alle $i \neq j$. Dann gilt*

$$\operatorname{Var}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \operatorname{Var}[X_i].$$

Für den Erwartungswert gilt die Linearität $\mathbb{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbb{E}[X_i]$ dagegen immer, das heißt ohne Annahme an die Abhängigkeitsstruktur.

Satz 0.14 (Schwaches Gesetz der großen Zahlen)**.** *Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge paarweise unkorrelierter, identisch verteilter reellwertiger Zufallsvariablen mit $\mathbb{E}[X_1^2] < \infty$. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{} \mathbb{E}[X_1] \qquad \textit{in Wahrscheinlichkeit,}$$

das heißt, $\mathbb{P}(|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_1]| > \varepsilon) \to 0$ für $n \to \infty$ und jedes $\varepsilon > 0$.

Beweis. Nach der Chebyshev-Ungleichung und Satz 0.13 ist die Wahrscheinlichkeit höchstens $\mathrm{Var}[X_1]/(n\varepsilon^2)$. □

Interpretation: Der Mittelwert über viele unabhängige (oder auch nur paarweise unkorrelierte) Experimente approximiert den Erwartungswert.

0.4. Elementare bedingte Wahrscheinlichkeiten

Definition 0.15 (Bedingte Wahrscheinlichkeit)**.** Seien $A, B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$. Dann heißt

$$\mathbb{P}(A|B) \coloneqq \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

die *elementare bedingte Wahrscheinlichkeit* von A gegeben B.

Die Abbildung $\mathbb{P}(\,\cdot|B)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$. Sie beschreibt die Wahrscheinlichkeiten unter der Information, dass das Ereignis B bereits eingetreten ist. Sind A und B stochastisch unabhängig, so gilt $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Für Ereignisse A, B mit $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(B^c) > 0$ gelten die *Formel von der totalen Wahrscheinlichkeit* und die *Formel von Bayes*,

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c), \tag{0.1}$$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} = \mathbb{P}(A|B)\frac{\mathbb{P}(B)}{\mathbb{P}(A)}. \tag{0.2}$$

Beispiel 0.16 (Corona-Schnelltest)**.** Ein Schnelltest habe die folgenden Kenngrößen. Die *Spezifität*, also die Wahrscheinlichkeit, dass eine gesunde Person als gesund erkannt wird, betrage 99,03 %. Die *Sensitivität*, also die Wahrscheinlichkeit, dass eine kranke Person als krank erkannt wird, betrage 96,45 %. Wir betrachten die Ereignisse

$$A = \{\text{Test negativ}\}, \qquad A^c = \{\text{Test positiv}\},$$
$$B = \{\text{Covid-gesund}\}, \qquad B^c = \{\text{Covid-krank}\}.$$

Dann ist die Spezifität gleich $\mathbb{P}(A|B)$ und die Sensitivität gleich $\mathbb{P}(A^c|B^c)$. Uns interessiert die Wahrscheinlichkeit, bei positivem Test tatsächlich krank zu sein. Nach der Formel von Bayes (0.2) gilt

$$\mathbb{P}(B^c|A^c) = \frac{\mathbb{P}(B^c\cap A^c)}{\mathbb{P}(A^c)} = \mathbb{P}(A^c|B^c)\,\frac{\mathbb{P}(B^c)}{\mathbb{P}(A^c)}.$$

Hier ist $\mathbb{P}(A^c|B^c)$ bekannt. Als Zusatzinformation benötigen wir den Anteil der Infizierten, $\mathbb{P}(B^c) \approx 0{,}003$. Die Wahrscheinlichkeit $\mathbb{P}(A^c)$ erhalten wir aus der Formel von der totalen Wahrscheinlichkeit (0.1), wobei $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$:

$$\mathbb{P}(A^c) = \mathbb{P}(A^c|B^c)\mathbb{P}(B^c) + \mathbb{P}(A^c|B)\mathbb{P}(B) = 0{,}9645\cdot 0{,}003 + 0{,}0097\cdot 0{,}997 \approx 0{,}0126.$$

Damit folgt

$$\mathbb{P}(B^{\mathrm{c}}|A^{\mathrm{c}}) \approx 0{,}9645 \cdot \frac{0{,}003}{0{,}0126} \approx 0{,}23 = 23\,\%.$$

Nur etwa ein Viertel der positiv Getesteten ist also tatsächlich krank; etwa drei von vier positiven Tests sind falsch positiv. Ebenso erhält man $\mathbb{P}(B|A) \approx 99{,}99\,\%$. Ein negativer Schnelltest bedeutet also eine große Sicherheit darüber, dass man gesund ist.

1. Charakteristische Funktionen

Zu diesem Kapitel vergleiche man [Kle13, Kapitel 15] und [Dur19, Abschnitte 3.3 und 3.4].

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $\boldsymbol{X} = (X_1, \dots, X_k)^\top$ eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $\mathbb{R}^k$, versehen mit der Borel-σ-Algebra $\mathcal{B}(\mathbb{R}^k)$. Wir bezeichnen mit $\mathbb{P}^{\boldsymbol{X}}$ die Verteilung von $\boldsymbol{X}$; sie ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.

Definition 1.1 (Charakteristische Funktion)**.** Die Funktion

$$\varphi^{\boldsymbol{X}}(\boldsymbol{t}) \coloneqq \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{X}}\big] = \mathbb{E}\big[\cos(\boldsymbol{t}^\top \boldsymbol{X}) + \mathrm{i}\sin(\boldsymbol{t}^\top \boldsymbol{X})\big],$$

mit $\boldsymbol{t} = (t_1, \dots, t_k)^\top \in \mathbb{R}^k$, heißt *charakteristische Funktion* (cf) von $\boldsymbol{X}$ beziehungsweise von $\mathbb{P}^{\boldsymbol{X}}$, oder auch *Fouriertransformierte* von $\mathbb{P}^{\boldsymbol{X}}$. Dabei bezeichnet i die imaginäre Einheit und $\boldsymbol{t}^\top \boldsymbol{X}$ das Skalarprodukt von $\boldsymbol{t}$ und $\boldsymbol{X}$.

Bemerkung 1.2. Zur Erinnerung stellen wir einige Fakten über Integrale komplexwertiger Funktionen zusammen.

(i) Eine Funktion $f\colon \Omega \to \mathbb{C}$ ist genau dann messbar, wenn $\operatorname{Re} f$ und $\operatorname{Im} f\colon \Omega \to \mathbb{R}$ messbar sind.

(ii) Sind $\operatorname{Re} f$ und $\operatorname{Im} f$ integrierbar, so definiert man

$$\int_\Omega f \,\mathrm{d}\mathbb{P} \coloneqq \int_\Omega \operatorname{Re} f \,\mathrm{d}\mathbb{P} + \mathrm{i}\int_\Omega \operatorname{Im} f \,\mathrm{d}\mathbb{P}.$$

(iii) Für integrierbare $f, f_1, f_2\colon \Omega \to \mathbb{C}$ und $c_1, c_2 \in \mathbb{C}$ gelten die folgenden Rechenregeln:

$$\operatorname{Re}\int f \,\mathrm{d}\mathbb{P} = \int \operatorname{Re} f \,\mathrm{d}\mathbb{P}, \qquad \operatorname{Im}\int f \,\mathrm{d}\mathbb{P} = \int \operatorname{Im} f \,\mathrm{d}\mathbb{P}, \qquad \overline{\int f \,\mathrm{d}\mathbb{P}} = \int \overline{f} \,\mathrm{d}\mathbb{P}, \tag{1.1}$$

$$\int (c_1 f_1 + c_2 f_2) \,\mathrm{d}\mathbb{P} = c_1 \int f_1 \,\mathrm{d}\mathbb{P} + c_2 \int f_2 \,\mathrm{d}\mathbb{P}, \qquad \Big|\int f \,\mathrm{d}\mathbb{P}\Big| \le \int |f| \,\mathrm{d}\mathbb{P}. \tag{1.2}$$

Beispiel 1.3. Seien $k = 1$ und $t \in \mathbb{R}$. Dann gilt:

$$X \sim \delta_a \text{ (Einpunktmaß in } a \in \mathbb{R}) \quad \Longrightarrow \quad \varphi^X(t) = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] = \mathrm{e}^{\mathrm{i}ta}. \tag{1.3}$$

$$X \sim \mathrm{Poi}_\lambda \quad \Longrightarrow \quad \varphi^X(t) = \sum_{n=0}^{\infty} \mathrm{e}^{\mathrm{i}tn}\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda\mathrm{e}^{\mathrm{i}t}} = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}.$$

$$X \sim \mathrm{Bin}(n,p) \quad \Longrightarrow \quad \varphi^X(t) = \sum_{k=0}^{n} \mathrm{e}^{\mathrm{i}tk}\binom{n}{k}p^k(1-p)^{n-k} = \big(\mathrm{e}^{\mathrm{i}t}p + (1-p)\big)^n,$$

wobei wir im letzten Schritt den binomischen Lehrsatz verwendet haben.

$$X \sim \mathcal{N}(0,1) \quad \Longrightarrow \quad \varphi^X(t) = \mathrm{e}^{-t^2/2}. \tag{1.4}$$

Zur Begründung entwickeln wir die Exponentialfunktion in ihre Taylorreihe. Wegen der Identität $\sum_{n\ge 0} |tX|^n/n! = \mathrm{e}^{|t||X|}$ und $\mathbb{E}[\mathrm{e}^{|t||X|}] < \infty$ dürfen wir Summe und Erwartungswert vertauschen. Wir erhalten

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] = \mathbb{E}\Big[\sum_{n=0}^{\infty}\frac{(\mathrm{i}tX)^n}{n!}\Big] = \sum_{n=0}^{\infty}\frac{(\mathrm{i}t)^n}{n!}\mathbb{E}[X^n].$$

Für die Momente der Standardnormalverteilung gilt

$$\mathbb{E}[X^n] = \begin{cases} 0, & n \text{ ungerade}, \\ \dfrac{n!}{2^{n/2}(n/2)!}, & n \text{ gerade}. \end{cases}$$

Damit folgt

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] = \sum_{n=0}^{\infty} \frac{(\mathrm{i}t)^{2n}}{(2n)!}\,\frac{(2n)!}{2^n n!} = \sum_{n=0}^{\infty} \frac{1}{n!}\Big(\frac{(\mathrm{i}t)^2}{2}\Big)^n = \sum_{n=0}^{\infty} \frac{1}{n!}\Big(-\frac{t^2}{2}\Big)^n = \mathrm{e}^{-t^2/2}.$$

Für $\mu \in \mathbb{R}$ und $\sigma > 0$ gilt weiter

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Longrightarrow \quad \varphi^X(t) = \mathrm{e}^{\mathrm{i}t\mu}\mathrm{e}^{-t^2\sigma^2/2}, \tag{1.5}$$

denn X hat dieselbe Verteilung wie $\mu + \sigma Z$ mit $Z \sim \mathcal{N}(0,1)$, und somit gilt $\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}] = \mathrm{e}^{\mathrm{i}t\mu}\mathbb{E}[\mathrm{e}^{\mathrm{i}t\sigma Z}]$. Für die Cauchy-Verteilung mit Dichte $x \mapsto (1/\pi)(1/(1+x^2))$ gilt

$$X \sim \mathrm{Cauchy}(0,1) \quad \Longrightarrow \quad \varphi^X(t) = \mathrm{e}^{-|t|}.$$

Der Erwartungswert $\mathbb{E}[X]$ existiert hier übrigens nicht; die Verteilung ist „heavy-tailed“. Schließlich gilt für die Gleichverteilung auf $[-1,1]$

$$X \sim \mathcal{U}(-1,1) \quad \Longrightarrow \quad \varphi^X(t) = \frac{\sin t}{t} \quad \text{für } t \neq 0,$$

und $\varphi^X(0) = 1$; der Nachweis ist eine Übungsaufgabe.

Lemma 1.4 (Grundlegende Eigenschaften charakteristischer Funktionen)**.** *Sei φ die charakteristische Funktion einer $\mathbb{R}^k$-wertigen Zufallsvariablen $\boldsymbol{X}$. Dann gilt:*

$$\varphi \text{ ist gleichmäßig stetig auf } \mathbb{R}^k. \tag{1.6}$$

$$\varphi(\mathbf{0}) = 1 \quad \text{und} \quad |\varphi(\boldsymbol{t})| \le \varphi(\mathbf{0}) \quad \text{für alle } \boldsymbol{t} \in \mathbb{R}^k. \tag{1.7}$$

Außerdem ist φ positiv semidefinit*, das heißt, für alle $n \in \mathbb{N}$, alle $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n \in \mathbb{R}^k$ und alle $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ gilt*

$$(\lambda_1, \dots, \lambda_n)\big(\varphi(\boldsymbol{t}_r - \boldsymbol{t}_s)\big)_{r,s=1,\dots,n} \begin{pmatrix} \overline{\lambda_1} \\ \vdots \\ \overline{\lambda_n} \end{pmatrix} \ge 0, \tag{1.8}$$

oder äquivalent

$$\sum_{r=1}^{n}\sum_{s=1}^{n} \lambda_r \overline{\lambda_s}\, \varphi(\boldsymbol{t}_r - \boldsymbol{t}_s) \ge 0. \tag{1.9}$$

Beweis. Zu (1.7): Es gilt $\varphi(\mathbf{0}) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\mathbf{0}^\top \boldsymbol{X}}] = 1$, und mit (1.2) folgt

$$|\varphi(\boldsymbol{t})| = \big|\mathbb{E}\big[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{X}}\big]\big| \le \mathbb{E}\big[\big|\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{X}}\big|\big] = 1.$$

Zu (1.8): Mit (1.2) und (1.1) gilt

$$\sum_{r=1}^{n}\sum_{s=1}^{n} \lambda_r \overline{\lambda_s}\, \mathbb{E}\big[\mathrm{e}^{\mathrm{i}(\boldsymbol{t}_r - \boldsymbol{t}_s)^\top \boldsymbol{X}}\big] = \mathbb{E}\Big[\sum_{r=1}^{n} \lambda_r \mathrm{e}^{\mathrm{i}\boldsymbol{t}_r^\top \boldsymbol{X}} \sum_{s=1}^{n} \overline{\lambda_s}\mathrm{e}^{-\mathrm{i}\boldsymbol{t}_s^\top \boldsymbol{X}}\Big] = \mathbb{E}\Big[\Big|\sum_{r=1}^{n} \lambda_r \mathrm{e}^{\mathrm{i}\boldsymbol{t}_r^\top \boldsymbol{X}}\Big|^2\Big] \ge 0.$$

Zu (1.6): Für alle $\boldsymbol{t}, \boldsymbol{h} \in \mathbb{R}^k$ gilt mit (1.2)

$$|\varphi(\boldsymbol{t}+\boldsymbol{h}) - \varphi(\boldsymbol{t})| \le \mathbb{E}\big[\big|\mathrm{e}^{\mathrm{i}(\boldsymbol{t}+\boldsymbol{h})^\top \boldsymbol{X}} - \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{X}}\big|\big] = \mathbb{E}\big[\big|\mathrm{e}^{\mathrm{i}\boldsymbol{h}^\top \boldsymbol{X}} - 1\big|\big].$$

Die rechte Seite hängt nicht von $\boldsymbol{t}$ ab. Da die komplexe Exponentialfunktion stetig in 0 ist, gilt $\lim_{\boldsymbol{h}\to\mathbf{0}} |\mathrm{e}^{\mathrm{i}\boldsymbol{h}^\top \boldsymbol{X}} - 1| = 0$ punktweise, und der Integrand ist durch 2 beschränkt. Mit dem Satz von der dominierten Konvergenz konvergiert die rechte Seite für $\boldsymbol{h} \to \mathbf{0}$ gegen 0, und zwar gleichmäßig in $\boldsymbol{t}$. □

Bemerkung. Weiterhin gilt $\varphi(-\boldsymbol{t}) = \overline{\varphi(\boldsymbol{t})}$ sowie

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top(a\boldsymbol{X}+\boldsymbol{b})}\big] = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{b}}\varphi(a\boldsymbol{t}) \qquad \text{für alle } a \in \mathbb{R},\ \boldsymbol{b} \in \mathbb{R}^k.$$

Die Eigenschaften (1.6), (1.7) und (1.8) charakterisieren charakteristische Funktionen vollständig.

Satz 1.5 (Satz von Bochner [Boc33])**.** *Zu jeder Abbildung $\varphi\colon \mathbb{R}^k \to \mathbb{C}$ mit den Eigenschaften (1.6)–(1.8) existiert ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, dessen charakteristische Funktion φ ist.*

Beweis. Wir führen den Beweis nur für $k = 1$. Die Idee ist, φ durch die Funktionen $\varphi(s)\mathrm{e}^{-s^2\sigma^2/2}$, $\sigma > 0$, zu approximieren, für diese explizit eine Dichte anzugeben, und dann $\sigma \downarrow 0$ mit Satz 1.6 zu behandeln.

Schritt 1: Positivität. Für alle $\sigma > 0$ und $x \in \mathbb{R}$ gilt

$$\iint_{\mathbb{R}^2} \mathrm{e}^{-u^2\sigma^2 - \mathrm{i}ux}\,\varphi(u-v)\,\mathrm{e}^{-v^2\sigma^2 + \mathrm{i}vx}\,\mathrm{d}u\,\mathrm{d}v \geq 0. \tag{1.10}$$

Mit $g(u) \coloneqq \mathrm{e}^{-u^2\sigma^2 - \mathrm{i}ux}$ lässt sich das Integral nämlich durch Riemann-Summen der Form

$$\sum_{r,s} g(t_r)\,\varphi(t_r - t_s)\,\overline{g(t_s)}\,\Delta^2$$

über endliche Gitter $\{t_r\}$ mit Maschenweite Δ approximieren; der Integrand ist stetig und durch $\mathrm{e}^{-\sigma^2(u^2+v^2)}$ beschränkt. Jede dieser Summen ist nach (1.9) mit $\lambda_r = g(t_r)\Delta$ nichtnegativ.

Schritt 2: Definition der Dichte. Wegen $u^2 + v^2 = (1/2)(u-v)^2 + (1/2)(u+v)^2$ lautet (1.10)

$$0 \leq \int_{\mathbb{R}} \Big\{ \int_{\mathbb{R}} \varphi(u-v)\,\mathrm{e}^{-\mathrm{i}(u-v)x - \frac{\sigma^2}{2}(u-v)^2 - \frac{\sigma^2}{2}(u+v)^2}\,\mathrm{d}u \Big\}\,\mathrm{d}v.$$

Mit der Substitution $s = u + v$, $t = u - v$ (Jacobi-Determinante vom Betrag $1/2$) folgt

$$0 \leq \frac{1}{2}\int_{\mathbb{R}} \Big\{ \int_{\mathbb{R}} \varphi(t)\,\mathrm{e}^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2 - \frac{\sigma^2}{2}s^2}\,\mathrm{d}s \Big\}\,\mathrm{d}t = \frac{1}{2}\int_{\mathbb{R}} \mathrm{e}^{-\frac{\sigma^2}{2}s^2}\,\mathrm{d}s \int_{\mathbb{R}} \varphi(t)\,\mathrm{e}^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2}\,\mathrm{d}t.$$

Da der erste Faktor positiv ist, folgt

$$f(x,\sigma) \coloneqq \frac{1}{2\pi}\int_{\mathbb{R}} \varphi(t)\,\mathrm{e}^{-\mathrm{i}tx - \frac{t^2\sigma^2}{2}}\,\mathrm{d}t \geq 0 \qquad \text{für alle } x \in \mathbb{R},\ \sigma > 0.$$

Schritt 3: $f(\cdot,\sigma)$ ist eine Wahrscheinlichkeitsdichte. Mit monotoner Konvergenz (wegen $f \geq 0$) und dem Satz von Fubini gilt

$$\int_{\mathbb{R}} f(x,\sigma)\,\mathrm{d}x = \lim_{\delta\downarrow 0}\int_{\mathbb{R}} f(x,\sigma)\,\mathrm{e}^{-\frac{x^2\delta^2}{2}}\,\mathrm{d}x = \frac{1}{2\pi}\lim_{\delta\downarrow 0}\int_{\mathbb{R}} \varphi(t)\,\mathrm{e}^{-\frac{t^2\sigma^2}{2}}\Big(\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx - \frac{x^2\delta^2}{2}}\,\mathrm{d}x\Big)\,\mathrm{d}t.$$

Nach (1.5), angewandt auf $\mathcal{N}(0,\delta^{-2})$, ist das innere Integral gleich $(\sqrt{2\pi}/\delta)\mathrm{e}^{-t^2/(2\delta^2)}$. Mit der Substitution $s = t/\delta$ folgt

$$\begin{aligned}\int_{\mathbb{R}} f(x,\sigma)\,\mathrm{d}x &= \frac{1}{\sqrt{2\pi}}\lim_{\delta\downarrow 0}\frac{1}{\delta}\int_{\mathbb{R}} \varphi(t)\,\mathrm{e}^{-\frac{t^2\sigma^2}{2} - \frac{t^2}{2\delta^2}}\,\mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}}\lim_{\delta\downarrow 0}\int_{\mathbb{R}} \varphi(\delta s)\,\mathrm{e}^{-\frac{s^2\sigma^2\delta^2}{2} - \frac{s^2}{2}}\,\mathrm{d}s = \varphi(0)\cdot 1 = 1.\end{aligned}$$

Im letzten Schritt haben wir den Satz von der dominierten Konvergenz verwendet, mit Majorante $\mathrm{e}^{-s^2/2}$ wegen (1.7), sowie die Stetigkeit von φ in 0. Also existiert zu $f(\cdot,\sigma)$ für jedes $\sigma > 0$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit dieser Dichte.

Schritt 4: Die charakteristische Funktion von $f(\cdot,\sigma)$. Für $s \in \mathbb{R}$ gilt mit dominierter Konvergenz (Majorante $f(\cdot,\sigma)$) und dem Satz von Fubini

$$\begin{aligned}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}sx} f(x,\sigma)\,\mathrm{d}x &= \lim_{\delta\downarrow 0} \int_{\mathbb{R}} f(x,\sigma)\,\mathrm{e}^{\mathrm{i}sx-\frac{x^2\delta^2}{2}}\,\mathrm{d}x \\ &= \lim_{\delta\downarrow 0} \int_{\mathbb{R}} \varphi(t)\,\mathrm{e}^{-\frac{t^2\sigma^2}{2}} \Big(\frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}x(s-t)-\frac{x^2\delta^2}{2}}\,\mathrm{d}x\Big)\,\mathrm{d}t.\end{aligned}$$

Nach (1.5) ist das innere Integral gleich $(1/(\delta\sqrt{2\pi}))\mathrm{e}^{-(t-s)^2/(2\delta^2)}$. Mit der Substitution $u = (t-s)/\delta$ und dominierter Konvergenz folgt

$$\begin{aligned}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}sx} f(x,\sigma)\,\mathrm{d}x &= \lim_{\delta\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(s+\delta u)\,\mathrm{e}^{-\frac{(s+\delta u)^2\sigma^2}{2}-\frac{u^2}{2}}\,\mathrm{d}u \\ &= \varphi(s)\,\mathrm{e}^{-\frac{s^2\sigma^2}{2}}\,\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u = \varphi(s)\,\mathrm{e}^{-\frac{s^2\sigma^2}{2}}.\end{aligned}$$

Schritt 5: Grenzübergang $\sigma \downarrow 0$. Zum Wahrscheinlichkeitsmaß mit Dichte $f(\cdot,\sigma)$ gehört also die charakteristische Funktion $s \mapsto \varphi(s)\mathrm{e}^{-s^2\sigma^2/2}$, welche für $\sigma \downarrow 0$ punktweise gegen φ konvergiert. Allerdings konvergiert $f(\cdot,\sigma)$ nicht unbedingt, und die Existenz eines Grenzmaßes ist damit noch nicht sichergestellt. Eine direkte Anwendung des folgenden Stetigkeitssatzes, Satz 1.6, mit $\sigma = 1/n$ komplettiert jedoch den Beweis, da φ nach (1.6) stetig in 0 ist. □

Satz 1.6 (Stetigkeitssatz für charakteristische Funktionen [Lév25])**.** *Sei $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit Werten in $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ und zugehörigen charakteristischen Funktionen $(\varphi_n)_{n\in\mathbb{N}}$. Existiert eine in $\mathbf{0}$ stetige Abbildung $\varphi_0\colon \mathbb{R}^k \to \mathbb{C}$ mit*

$$\varphi_n(\boldsymbol{t}) \xrightarrow[n\to\infty]{} \varphi_0(\boldsymbol{t}) \qquad \text{für alle } \boldsymbol{t} \in \mathbb{R}^k,$$

so ist φ_0 die charakteristische Funktion eines Wahrscheinlichkeitsmaßes Q_0 auf $\mathbb{R}^k$, und es gilt

$$\mathbb{P}^{\boldsymbol{X}_n}\big((-\boldsymbol{\infty}, \boldsymbol{x}]\big) \xrightarrow[n\to\infty]{} Q_0\big((-\boldsymbol{\infty}, \boldsymbol{x}]\big) \tag{1.11}$$

für alle $\boldsymbol{x} \in \mathbb{R}^k$, in denen Q_0 stetig ist, das heißt, für die $Q_0\big(\partial(-\boldsymbol{\infty}, \boldsymbol{x}]\big) = 0$ gilt.

Beweis. Wir führen den Beweis in Abschnitt 2.2, sobald uns mit dem Satz von Prohorov die nötigen Hilfsmittel zur Verfügung stehen. □

Bemerkung. Hier bezeichnet $(-\boldsymbol{\infty}, \boldsymbol{x}] \coloneqq (-\infty, x_1] \times \cdots \times (-\infty, x_k]$ den von $\boldsymbol{x} = (x_1, \ldots, x_k)^\top$ aus nach links unten verlaufenden Quadranten, siehe Abbildung 1.1 für $k = 2$. Die Bedingung $Q_0\big(\partial(-\boldsymbol{\infty}, \boldsymbol{x}]\big) = 0$ verlangt, dass der Rand dieses Quadranten keine Masse trägt. Für $k = 1$ ist dies gerade $Q_0(\{x\}) = 0$, also die Stetigkeit der Verteilungsfunktion von Q_0 in x.

Lemma 1.7 (Unabhängigkeit und Summen)**.** *Seien $\boldsymbol{X}_1$ und $\boldsymbol{X}_2$ stochastisch unabhängige Zufallsvariablen mit Werten in $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Dann gilt für alle $\boldsymbol{t}, \boldsymbol{t}_1, \boldsymbol{t}_2 \in \mathbb{R}^k$*

$$\varphi^{(\boldsymbol{X}_1,\boldsymbol{X}_2)}(\boldsymbol{t}_1, \boldsymbol{t}_2) = \mathbb{E}\Big[\mathrm{e}^{\mathrm{i}(\boldsymbol{t}_1^\top, \boldsymbol{t}_2^\top)\binom{\boldsymbol{X}_1}{\boldsymbol{X}_2}}\Big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\boldsymbol{t}_1^\top \boldsymbol{X}_1}\big]\,\mathbb{E}\big[\mathrm{e}^{\mathrm{i}\boldsymbol{t}_2^\top \boldsymbol{X}_2}\big] = \varphi^{\boldsymbol{X}_1}(\boldsymbol{t}_1)\,\varphi^{\boldsymbol{X}_2}(\boldsymbol{t}_2) \tag{1.12}$$

und

$$\varphi^{\boldsymbol{X}_1+\boldsymbol{X}_2}(\boldsymbol{t}) = \varphi^{\boldsymbol{X}_1}(\boldsymbol{t})\,\varphi^{\boldsymbol{X}_2}(\boldsymbol{t}). \tag{1.13}$$

Beweis. Da $\boldsymbol{X}_1$ und $\boldsymbol{X}_2$ unabhängig sind, sind auch die komplexwertigen Zufallsvariablen $\mathrm{e}^{\mathrm{i}\boldsymbol{t}_1^\top \boldsymbol{X}_1}$ und $\mathrm{e}^{\mathrm{i}\boldsymbol{t}_2^\top \boldsymbol{X}_2}$ unabhängig; ihr Produkt hat also als Erwartungswert das Produkt der Erwartungswerte, angewandt auf Real- und Imaginärteile. Dies ist die mittlere Gleichung in (1.12). Die Gleichung (1.13) folgt aus (1.12) mit $\boldsymbol{t}_1 = \boldsymbol{t}_2 = \boldsymbol{t}$, denn $\boldsymbol{t}^\top(\boldsymbol{X}_1 + \boldsymbol{X}_2) = (\boldsymbol{t}^\top, \boldsymbol{t}^\top)\binom{\boldsymbol{X}_1}{\boldsymbol{X}_2}$. □

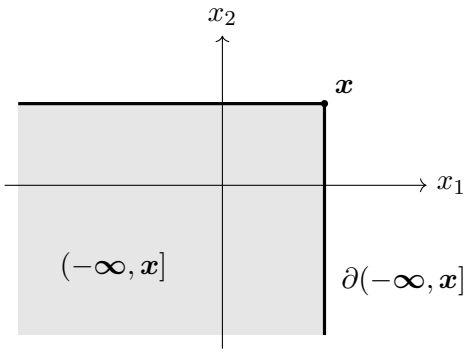

Abbildung 1.1.: Der Quadrant $(-\infty, \boldsymbol{x}]$ für $k = 2$; der Rand ist fett eingezeichnet.

Satz 1.8 (Eindeutigkeitssatz für charakteristische Funktionen)**.** *Seien $\boldsymbol{X}_1$ und $\boldsymbol{X}_2$ zwei Zufallsvariablen mit Werten in $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Dann gilt*

$$\varphi^{\boldsymbol{X}_1}(\boldsymbol{t}) = \varphi^{\boldsymbol{X}_2}(\boldsymbol{t}) \quad \textit{für alle } \boldsymbol{t} \in \mathbb{R}^k \qquad \Longleftrightarrow \qquad \mathbb{P}^{\boldsymbol{X}_1} = \mathbb{P}^{\boldsymbol{X}_2}.$$

Beweis. „⇐“: Gilt $\mathbb{P}^{\boldsymbol{X}_1} = \mathbb{P}^{\boldsymbol{X}_2}$, so folgt für alle $\boldsymbol{t} \in \mathbb{R}^k$

$$\varphi^{\boldsymbol{X}_1}(\boldsymbol{t}) = \int_{\mathbb{R}^k} \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{x}} \, \mathbb{P}^{\boldsymbol{X}_1}(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^k} \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{x}} \, \mathbb{P}^{\boldsymbol{X}_2}(\mathrm{d}\boldsymbol{x}) = \varphi^{\boldsymbol{X}_2}(\boldsymbol{t}).$$

„⇒“: Die Idee ist, Dichten durch Faltung mit einer Normalverteilung zu erzeugen. Seien dazu $Y_1, \ldots, Y_k \sim \mathcal{N}(0,1)$ unabhängig und identisch verteilt sowie unabhängig von $\boldsymbol{X}_j$, $j \in \{1,2\}$. Für jedes $\sigma > 0$ hat $\boldsymbol{Y}_\sigma \coloneqq (\sigma Y_1, \ldots, \sigma Y_k)^\top$ die Dichte

$$\boldsymbol{u} = (u_1, \ldots, u_k)^\top \longmapsto \frac{1}{\sqrt{(2\pi\sigma^2)^k}} \, \mathrm{e}^{-\boldsymbol{u}^\top \boldsymbol{u}/(2\sigma^2)} =: f_\sigma(\boldsymbol{u}).$$

Dann hat $\boldsymbol{Z}_{j,\sigma} \coloneqq \boldsymbol{X}_j + \boldsymbol{Y}_\sigma$, $j \in \{1,2\}$, die Dichte

$$\boldsymbol{x} \longmapsto g_{j,\sigma}(\boldsymbol{x}) \coloneqq \int_{\mathbb{R}^k} f_\sigma(\boldsymbol{u} - \boldsymbol{x}) \, \mathbb{P}^{\boldsymbol{X}_j}(\mathrm{d}\boldsymbol{u}).$$

Gilt nun $\mathbb{P}^{\boldsymbol{Z}_{1,\sigma}} = \mathbb{P}^{\boldsymbol{Z}_{2,\sigma}}$ für alle $\sigma > 0$, so folgt daraus $\mathbb{P}^{\boldsymbol{X}_1} = \mathbb{P}^{\boldsymbol{X}_2}$. Dies folgt aus der Eindeutigkeit des schwachen Grenzwerts, siehe Kapitel 2, genauer Satz 2.20 und die Bemerkung danach. Denn für $\sigma \downarrow 0$ gilt $\boldsymbol{Z}_{j,\sigma} \to \boldsymbol{X}_j$ punktweise, und mit dominierter Konvergenz folgt $\int f \, \mathrm{d}\mathbb{P}^{\boldsymbol{Z}_{j,\sigma}} \to \int f \, \mathrm{d}\mathbb{P}^{\boldsymbol{X}_j}$ für jede stetige und beschränkte Funktion $f \colon \mathbb{R}^k \to \mathbb{R}$.

Es bleibt also $g_{1,\sigma} = g_{2,\sigma}$ für alle $\sigma > 0$ zu zeigen. Es gilt

$$g_{j,\sigma}(\boldsymbol{x}) = \int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi\sigma^2)^k}} \, \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{u})^\top (\boldsymbol{x}-\boldsymbol{u})} \, \mathbb{P}^{\boldsymbol{X}_j}(\mathrm{d}\boldsymbol{u}).$$

Die Exponentialfunktion im Integranden ist die charakteristische Funktion von $\boldsymbol{Y}_{1/\sigma}$ an der Stelle $\boldsymbol{x} - \boldsymbol{u}$, vergleiche (1.5) und Lemma 1.7. Mit $\varphi^{\boldsymbol{Y}_{1/\sigma}}(\boldsymbol{x} - \boldsymbol{u}) = \int_{\mathbb{R}^k} \mathrm{e}^{\mathrm{i}(\boldsymbol{x}-\boldsymbol{u})^\top \boldsymbol{y}} f_{1/\sigma}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}$ gilt also

$$\begin{aligned}
g_{j,\sigma}(\boldsymbol{x}) &= \int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi\sigma^2)^k}} \, \varphi^{\boldsymbol{Y}_{1/\sigma}}(\boldsymbol{x} - \boldsymbol{u}) \, \mathbb{P}^{\boldsymbol{X}_j}(\mathrm{d}\boldsymbol{u}) \\
&= \int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi\sigma^2)^k}} \Big(\int_{\mathbb{R}^k} \mathrm{e}^{\mathrm{i}(\boldsymbol{x}-\boldsymbol{u})^\top \boldsymbol{y}} \, \frac{1}{\sqrt{(2\pi/\sigma^2)^k}} \, \mathrm{e}^{-\frac{\sigma^2}{2}\boldsymbol{y}^\top \boldsymbol{y}} \, \mathrm{d}\boldsymbol{y} \Big) \mathbb{P}^{\boldsymbol{X}_j}(\mathrm{d}\boldsymbol{u}) \\
&= \int_{\mathbb{R}^k} \Big(\int_{\mathbb{R}^k} \mathrm{e}^{-\mathrm{i}\boldsymbol{u}^\top \boldsymbol{y}} \, \mathbb{P}^{\boldsymbol{X}_j}(\mathrm{d}\boldsymbol{u}) \Big) \frac{1}{(2\pi)^k} \, \mathrm{e}^{\mathrm{i}\boldsymbol{x}^\top \boldsymbol{y} - \frac{\sigma^2}{2}\boldsymbol{y}^\top \boldsymbol{y}} \, \mathrm{d}\boldsymbol{y},
\end{aligned}$$

wobei wir im letzten Schritt den Satz von Fubini verwendet haben; der Integrand ist durch die integrierbare Funktion $\mathrm{e}^{-\sigma^2 \boldsymbol{y}^\top \boldsymbol{y}/2}$ beschränkt. Das innere Integral ist $\varphi^{\boldsymbol{X}_j}(-\boldsymbol{y})$ und nach Voraussetzung unabhängig von j. Damit folgt $g_{1,\sigma}(\boldsymbol{x}) = g_{2,\sigma}(\boldsymbol{x})$ für alle $\boldsymbol{x} \in \mathbb{R}^k$, also $\mathbb{P}^{\boldsymbol{Z}_{1,\sigma}} = \mathbb{P}^{\boldsymbol{Z}_{2,\sigma}}$ für alle $\sigma > 0$, und nach der Vorüberlegung $\mathbb{P}^{\boldsymbol{X}_1} = \mathbb{P}^{\boldsymbol{X}_2}$. □

Korollar 1.9 (Unabhängigkeit)**.** *Für $j = 1, \dots, n$ sei $\boldsymbol{X}_j$ eine Zufallsvariable mit Werten in $(\mathbb{R}^{k_j}, \mathcal{B}(\mathbb{R}^{k_j}))$. Dann gilt: $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ sind genau dann stochastisch unabhängig, wenn*

$$\varphi^{(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)}(\boldsymbol{t}_1, \dots, \boldsymbol{t}_n) = \prod_{j=1}^{n} \varphi^{\boldsymbol{X}_j}(\boldsymbol{t}_j) \qquad \textit{für alle } \boldsymbol{t}_j \in \mathbb{R}^{k_j},\ j = 1, \dots, n.$$

Beweis. „⇒“: Dies folgt durch wiederholte Anwendung von Lemma 1.7.

„⇐“: Nach Lemma 1.7 ist $(\boldsymbol{t}_1, \dots, \boldsymbol{t}_n) \mapsto \prod_{j=1}^n \varphi^{\boldsymbol{X}_j}(\boldsymbol{t}_j)$ die charakteristische Funktion der unabhängigen Kopplung $\bigotimes_{j=1}^n \mathbb{P}^{\boldsymbol{X}_j}$ der Wahrscheinlichkeitsmaße $\mathbb{P}^{\boldsymbol{X}_1}, \dots, \mathbb{P}^{\boldsymbol{X}_n}$. Insbesondere gilt dann nach Satz 1.8

$$\mathbb{P}^{(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)} = \bigotimes_{j=1}^{n} \mathbb{P}^{\boldsymbol{X}_j},$$

und somit sind $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ stochastisch unabhängig. □

Definition 1.10 (k-dimensionale Normalverteilung)**.** Eine Zufallsvariable $\boldsymbol{X}$ mit Werten in $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ und charakteristischer Funktion

$$\varphi^{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{\mu}}\, \mathrm{e}^{-\frac{1}{2}\boldsymbol{t}^\top S \boldsymbol{t}} \qquad \text{für alle } \boldsymbol{t} \in \mathbb{R}^k, \tag{1.14}$$

wobei $\boldsymbol{\mu} \in \mathbb{R}^k$ und S eine positiv semidefinite symmetrische $k \times k$-Matrix ist (das heißt, $\boldsymbol{x}^\top S \boldsymbol{x} \geq 0$ für alle $\boldsymbol{x} \in \mathbb{R}^k$), heißt *$k$-dimensional normalverteilt* mit Erwartungswert $\boldsymbol{\mu}$ und Kovarianzmatrix S, kurz $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, S)$.

Nach Satz 1.8 legt (1.14) die Verteilung von $\boldsymbol{X}$ eindeutig fest. Dass zu jedem Paar $(\boldsymbol{\mu}, S)$ eine solche Verteilung existiert, zeigt die folgende Konstruktion (2) mit $B = S^{1/2}$.

Bemerkung. (1) Nur falls S positiv definit ist, das heißt, falls $\boldsymbol{x}^\top S \boldsymbol{x} > 0$ für alle $\boldsymbol{x} \in \mathbb{R}^k \setminus \{\boldsymbol{0}\}$, besitzt $\boldsymbol{X}$ eine Dichte. Diese lautet

$$\boldsymbol{x} \longmapsto \frac{1}{\sqrt{(2\pi)^k \det(S)}}\, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top S^{-1}(\boldsymbol{x}-\boldsymbol{\mu})};$$

hier ist S^{-1} definiert, da S positiv definit und somit invertierbar ist.

(2) *Konstruktion.* Seien $X_1, \dots, X_k \sim \mathcal{N}(0,1)$ unabhängig und identisch verteilt und $\boldsymbol{X} = (X_1, \dots, X_k)^\top$. Seien weiter $\boldsymbol{\mu} \in \mathbb{R}^\ell$ und B eine $\ell \times k$-Matrix. Dann gilt für die Zufallsvariable

$$\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}, BB^\top),$$

wobei $BB^\top$ positiv semidefinit ist, denn $\boldsymbol{x}^\top BB^\top \boldsymbol{x} = |B^\top \boldsymbol{x}|^2 \geq 0$. Mit $\boldsymbol{u} = B^\top \boldsymbol{t}$ gilt nämlich

$$\begin{aligned}\varphi^{\boldsymbol{Y}}(\boldsymbol{t}) &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top(B\boldsymbol{X}+\boldsymbol{\mu})}\big] = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}}\, \mathbb{E}\big[\mathrm{e}^{\mathrm{i}(B^\top\boldsymbol{t})^\top \boldsymbol{X}}\big] = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}}\, \varphi^{\boldsymbol{X}}(B^\top \boldsymbol{t}) \\ &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}} \prod_{j=1}^{k} \varphi^{X_j}(u_j) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}}\, \mathrm{e}^{-\sum_{j=1}^k u_j^2/2} = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{u}^\top\boldsymbol{u}} = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}}\, \mathrm{e}^{-\frac{1}{2}\boldsymbol{t}^\top BB^\top \boldsymbol{t}},\end{aligned}$$

wobei wir Korollar 1.9 und (1.4) verwendet haben. Insbesondere gilt

$$\mathbb{E}[\boldsymbol{Y}] = B\,\mathbb{E}[\boldsymbol{X}] + \boldsymbol{\mu} = \boldsymbol{\mu}$$

und

$$\mathrm{Cov}(\boldsymbol{Y},\boldsymbol{Y}) = \mathbb{E}\big[(\boldsymbol{Y}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\mu})^\top\big] = B\,\mathbb{E}\big[\boldsymbol{X}\boldsymbol{X}^\top\big]B^\top = BB^\top,$$

denn $\mathbb{E}[X_iX_j] = 1$ für $i = j$ und $\mathbb{E}[X_iX_j] = 0$ für $i \neq j$.

(3) *Summen von Normalverteilungen.* Seien $X_1 \sim \mathcal{N}(\mu_1,\sigma_1^2)$ und $X_2 \sim \mathcal{N}(\mu_2,\sigma_2^2)$ stochastisch unabhängig. Mit Lemma 1.7 und (1.5) folgt

$$\varphi^{X_1+X_2}(t) = \varphi^{X_1}(t)\,\varphi^{X_2}(t) = \mathrm{e}^{\mathrm{i}t(\mu_1+\mu_2)}\,\mathrm{e}^{-\frac{1}{2}t^2(\sigma_1^2+\sigma_2^2)},$$

und mit Satz 1.8 folgt $X_1 + X_2 \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. Ebenso folgt aus (1.14): Für $\boldsymbol{c} \in \mathbb{R}^k$ und $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, S)$ gilt

$$\boldsymbol{c}^\top\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{c}^\top\boldsymbol{\mu}, \boldsymbol{c}^\top S\boldsymbol{c}), \tag{1.15}$$

denn $\varphi^{\boldsymbol{c}^\top\boldsymbol{X}}(t) = \varphi^{\boldsymbol{X}}(t\boldsymbol{c}) = \mathrm{e}^{\mathrm{i}t\boldsymbol{c}^\top\boldsymbol{\mu}}\mathrm{e}^{-(1/2)t^2\boldsymbol{c}^\top S\boldsymbol{c}}$. Dabei verstehen wir unter $\mathcal{N}(m,0)$ das Einpunktmaß δ_m, im Einklang mit (1.3).

Satz 1.11 (Unabhängigkeit von normalverteilten Zufallsvariablen)**.** *Seien $\boldsymbol{\mu} \in \mathbb{R}^k$ und $\boldsymbol{X} = (X_1,\dots,X_k)^\top \sim \mathcal{N}(\boldsymbol{\mu}, S)$. Dann gilt*

$$X_1,\dots,X_k \text{ sind stochastisch unabhängig} \iff S \text{ ist eine Diagonalmatrix.} \tag{1.16}$$

Beweis. Nach Korollar 1.9 gilt

$$X_1,\dots,X_k \text{ stu} \iff \varphi^{(X_1,\dots,X_k)}(t_1,\dots,t_k) = \prod_{j=1}^k \varphi^{X_j}(t_j) \quad \text{für alle } t_1,\dots,t_k \in \mathbb{R},$$

wobei „stu“ für „stochastisch unabhängig“ steht. Hier gilt einerseits

$$\varphi^{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{t}^\top S\boldsymbol{t}} \tag{1.17}$$

und andererseits, da $X_j \sim \mathcal{N}(\mu_j, S_{jj})$ nach (1.15) mit $\boldsymbol{c}$ gleich dem j-ten Einheitsvektor,

$$\prod_{j=1}^k \varphi^{X_j}(t_j) = \prod_{j=1}^k \mathrm{e}^{\mathrm{i}t_j\mu_j-\frac{1}{2}t_j^2S_{jj}} = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{t}^\top\,\mathrm{diag}(S)\boldsymbol{t}} \qquad \text{für alle } \boldsymbol{t} \in \mathbb{R}^k. \tag{1.18}$$

Dabei bezeichnet $\mathrm{diag}(S)$ die Diagonalmatrix mit denselben Diagonaleinträgen wie S. Folglich gilt (1.17) = (1.18) für alle $\boldsymbol{t}$ genau dann, wenn $\boldsymbol{t}^\top S\boldsymbol{t} = \boldsymbol{t}^\top\,\mathrm{diag}(S)\boldsymbol{t}$ für alle $\boldsymbol{t} \in \mathbb{R}^k$ gilt. Da beide Matrizen symmetrisch sind, ist dies durch Polarisierung äquivalent zu $S = \mathrm{diag}(S)$. □

Bemerkung. Aus (1.16) folgt wegen

$$S = \mathbb{E}\big[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top\big] = \big(\mathrm{Cov}(X_r,X_s)\big)_{r,s=1,\dots,k},$$

dass

$$X_1,\dots,X_k \text{ stochastisch unabhängig} \iff \mathrm{Cov}(X_r,X_s) = 0 \quad \text{für alle } r \neq s,$$

das heißt, genau dann, wenn $X_1,\dots,X_k$ paarweise unkorreliert sind. Dies gilt nur, weil der Vektor $\boldsymbol{X}$ gemeinsam normalverteilt ist; im Allgemeinen folgt aus Unkorreliertheit keine Unabhängigkeit, siehe Beispiel 0.12.

Satz 1.12 (Zusammenhang zwischen charakteristischer Funktion und Momenten)**.** *Sei μ ein endliches Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $\int |x|^m \, \mu(\mathrm{d}x) < \infty$ für ein $m \in \mathbb{N}$. Dann ist die Abbildung*

$$\varphi \colon \mathbb{R} \to \mathbb{C}, \qquad t \mapsto \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \, \mu(\mathrm{d}x),$$

m-mal differenzierbar mit gleichmäßig stetiger und beschränkter m-ter Ableitung. Es gilt

$$\varphi^{(m)}(0) = \mathrm{i}^m \int_{\mathbb{R}} x^m \, \mu(\mathrm{d}x). \tag{1.19}$$

Ist speziell X eine reellwertige Zufallsvariable mit $\mathbb{E}[|X|^m] < \infty$, so folgt für ihre charakteristische Funktion φ

$$\varphi^{(m)}(0) = \mathrm{i}^m \mathbb{E}[X^m]. \tag{1.20}$$

Außerdem gilt die Taylorentwicklung

$$\varphi(t) = \sum_{r=0}^{m} \frac{(\mathrm{i}t)^r}{r!} \mathbb{E}[X^r] + \int_0^t \frac{(t-y)^{m-1}}{(m-1)!} \bigl(\varphi^{(m)}(y) - \varphi^{(m)}(0)\bigr) \, \mathrm{d}y, \tag{1.21}$$

wobei das Integral für $t \to 0$ von der Ordnung $o(|t|^m)$ ist.

Beweis. Zunächst ist $t \mapsto \mathrm{e}^{\mathrm{i}tx}$ beliebig oft differenzierbar mit m-ter Ableitung $t \mapsto (\mathrm{i}x)^m \mathrm{e}^{\mathrm{i}tx}$, wobei $|(\mathrm{i}x)^m \mathrm{e}^{\mathrm{i}tx}| \le |x|^m$. Wegen $\int |x|^m \, \mu(\mathrm{d}x) < \infty$ hat die m-te Ableitung eine integrierbare Majorante. Dasselbe gilt für die r-te Ableitung, $r \le m$, wegen $|x|^r \le 1 + |x|^m$ und da μ endlich ist. Mit dem Satz von Lebesgue, genauer dessen Anwendung zum Beweis der Vertauschbarkeit von Integration und Differentiation, gilt

$$\varphi^{(m)}(t) = \int_{\mathbb{R}} (\mathrm{i}x)^m \mathrm{e}^{\mathrm{i}tx} \, \mu(\mathrm{d}x),$$

und insbesondere $\varphi^{(m)}(0) = \mathrm{i}^m \int x^m \, \mu(\mathrm{d}x)$ und $|\varphi^{(m)}(t)| \le \int |x|^m \, \mu(\mathrm{d}x) < \infty$. Außerdem gilt

$$|\varphi^{(m)}(t+h) - \varphi^{(m)}(t)| = \Bigl| \int_{\mathbb{R}} (\mathrm{i}x)^m \bigl(\mathrm{e}^{\mathrm{i}(t+h)x} - \mathrm{e}^{\mathrm{i}tx}\bigr) \, \mu(\mathrm{d}x) \Bigr| \le \int_{\mathbb{R}} |x|^m \bigl|\mathrm{e}^{\mathrm{i}hx} - 1\bigr| \, \mu(\mathrm{d}x) \xrightarrow[h \to 0]{} 0$$

mit dominierter Konvergenz, da $|\mathrm{e}^{\mathrm{i}hx} - 1| \le 2$; die Schranke hängt nicht von t ab. Die Gleichung (1.20) folgt aus (1.19), und die Taylorformel ist die übliche mit Integralrestglied. Für das Restglied gilt

$$\Bigl| \int_0^t \frac{(t-y)^{m-1}}{(m-1)!} \bigl(\varphi^{(m)}(y) - \varphi^{(m)}(0)\bigr) \, \mathrm{d}y \Bigr| \le \frac{|t|^m}{m!} \sup_{|y| \le |t|} \bigl|\varphi^{(m)}(y) - \varphi^{(m)}(0)\bigr| = o(|t|^m) \qquad \text{für } t \to 0$$

wegen der Stetigkeit von $\varphi^{(m)}$. □

Beispiel 1.13 (Höhere Momente der Normalverteilung)**.** Für $X \sim \mathcal{N}(0,1)$ gilt $\varphi^X(t) = \mathrm{e}^{-t^2/2}$, also $(\varphi^X)^{(m)}(0) = \mathrm{i}^m \mathbb{E}[X^m]$ nach (1.20). Ein Vergleich der Taylorentwicklung (1.21) mit der Potenzreihe $\mathrm{e}^{-t^2/2} = \sum_{n \ge 0} (-1)^n t^{2n} / (2^n n!)$ liefert

$$\mathbb{E}[X^m] = \begin{cases} 0, & m \text{ ungerade}, \\ \dfrac{m!}{2^{m/2}(m/2)!}, & m \text{ gerade}. \end{cases}$$

Satz 1.14 (Cramér–Wold-Technik)**.** *Sei $(\boldsymbol{X}_n)_{n \in \mathbb{N}_0}$ eine Folge von $\mathbb{R}^k$-wertigen Zufallsvariablen. Dann gilt*

$$\boldsymbol{X}_n \xrightarrow[n \to \infty]{\mathcal{D}} \boldsymbol{X}_0 \quad \Longleftrightarrow \quad \boldsymbol{c}^\top \boldsymbol{X}_n \xrightarrow[n \to \infty]{\mathcal{D}} \boldsymbol{c}^\top \boldsymbol{X}_0 \quad \textit{für alle } \boldsymbol{c} \in \mathbb{R}^k.$$

Hier bezeichnet $\xrightarrow{\mathcal{D}}$ die schwache Konvergenz (Konvergenz in Verteilung), die wir in Definition 2.15 präzise einführen; für reellwertige Zufallsvariablen entspricht sie der Konvergenz der Verteilungsfunktionen in allen Stetigkeitspunkten wie in (1.11).

Beweis. Siehe Abschnitt 2.2. □

Satz 1.15 (Zentraler Grenzwertsatz von Lindeberg–Lévy [Lin22, Lév25])**.** *Seien $(X_n)_{n\in\mathbb{N}}$ unabhängige und identisch verteilte Zufallsvariablen mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $\mathbb{E}[X_1] = \mu \in \mathbb{R}$, $\operatorname{Var}[X_1] = \sigma^2 \in (0,\infty)$. Dann gilt*

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\Big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2),$$

das heißt,

$$\mathbb{P}\Big(\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\Big) \le x\Big) \xrightarrow[n\to\infty]{} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{u^2}{2\sigma^2}}\, \mathrm{d}u \qquad \text{für alle } x \in \mathbb{R}. \tag{1.22}$$

Bemerkung. Die Konvergenz in (1.22) ist sogar gleichmäßig in x.

Beweis. Sei φ die charakteristische Funktion von $X_1 - \mu$. Nach Satz 1.12, Gleichung (1.21) mit $m = 2$, gilt die Taylorentwicklung

$$\varphi(t) = \underbrace{\varphi(0)}_{=1} + \mathrm{i}t\,\underbrace{\mathbb{E}[X_1-\mu]}_{=0} - \frac{t^2}{2}\,\underbrace{\mathbb{E}\big[(X_1-\mu)^2\big]}_{=\sigma^2} + o(t^2) = 1 - \frac{t^2}{2}\sigma^2 + o(t^2).$$

Aber nach Lemma 1.7, angewandt auf die unabhängigen Zufallsvariablen $(X_i - \mu)/\sqrt{n}$, gilt

$$\begin{aligned}\varphi^{\sum_{i=1}^n \frac{X_i-\mu}{\sqrt{n}}}(t) = \prod_{i=1}^{n} \varphi^{\frac{X_i-\mu}{\sqrt{n}}}(t) = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}(X_1-\mu)\frac{t}{\sqrt{n}}}\big]^n = \varphi\Big(\frac{t}{\sqrt{n}}\Big)^n \\ = \Big(1 - \frac{t^2}{2n}\sigma^2 + o\Big(\frac{t^2}{n}\Big)\Big)^n \xrightarrow[n\to\infty]{} \mathrm{e}^{-\frac{t^2\sigma^2}{2}},\end{aligned}$$

denn es gilt $(1 - a_n/n)^n \to \mathrm{e}^{-a}$, falls $a_n \to a$ (einfache Übung). Hier ist $a_n = t^2\sigma^2/2 + n\,o(t^2/n) \to t^2\sigma^2/2$ für festes t; die a_n sind im Allgemeinen komplex, und die Aussage der Übung gilt auch für komplexe Folgen. Da $\mathrm{e}^{-t^2\sigma^2/2}$ nach (1.5) die charakteristische Funktion von $\mathcal{N}(0,\sigma^2)$ ist, folgt die Behauptung aus dem Stetigkeitssatz, Satz 1.6. Da die Verteilungsfunktion von $\mathcal{N}(0,\sigma^2)$ stetig ist, gilt (1.11) dabei für alle $x \in \mathbb{R}$. □

Satz 1.16 (Diskrete Fourier-Inversionsformel)**.** *Sei X eine Zufallsvariable mit Werten in $\mathbb{Z}$ und sei φ^X die zugehörige charakteristische Funktion. Dann gilt für alle $k \in \mathbb{Z}$*

$$\mathbb{P}(X = k) = \frac{1}{2\pi}\int_{(-\pi,\pi]} \mathrm{e}^{-\mathrm{i}tk}\varphi^X(t)\,\mathrm{d}t. \tag{1.23}$$

Beweis. Es gilt $\varphi^X(t) = \sum_{j\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}tj}\mathbb{P}(X = j)$. Mit dem Satz von Lebesgue folgt

$$\begin{aligned}\int_{(-\pi,\pi]} \mathrm{e}^{-\mathrm{i}tk}\varphi^X(t)\,\mathrm{d}t &= \int_{(-\pi,\pi]} \mathrm{e}^{-\mathrm{i}tk} \lim_{n\to\infty} \sum_{j\colon |j|\le n} \mathrm{e}^{\mathrm{i}tj}\mathbb{P}(X=j)\,\mathrm{d}t \\ &= \lim_{n\to\infty} \sum_{j\colon |j|\le n} \mathbb{P}(X=j) \int_{(-\pi,\pi]} \mathrm{e}^{\mathrm{i}t(j-k)}\,\mathrm{d}t \\ &= \sum_{j\in\mathbb{Z}} \mathbb{P}(X=j)\,2\pi\,\mathbb{1}\{j=k\} = 2\pi\,\mathbb{P}(X=k).\end{aligned}$$

Die Vertauschung von Limes und Integral ist durch dominierte Konvergenz gerechtfertigt, denn für alle n gilt

$$\Big|\mathrm{e}^{-\mathrm{i}tk} \sum_{j\colon |j|\le n} \mathrm{e}^{\mathrm{i}tj}\mathbb{P}(X=j)\Big| \le \sum_{j\colon |j|\le n} \mathbb{P}(X=j) \le 1 \qquad \text{und} \qquad \int_{(-\pi,\pi]} 1\,\mathrm{d}t = 2\pi < \infty.$$

Für das verbleibende Integral gilt im Fall $j=k$ einfach $\int_{(-\pi,\pi]} 1\,\mathrm{d}t = 2\pi$, und im Fall $j \neq k$

$$\int_{(-\pi,\pi]} \mathrm{e}^{\mathrm{i}t(j-k)}\,\mathrm{d}t = \frac{1}{\mathrm{i}(j-k)}\,\mathrm{e}^{\mathrm{i}t(j-k)}\Big|_{-\pi}^{\pi} = \frac{1}{\mathrm{i}(j-k)}\big(\mathrm{e}^{\mathrm{i}\pi(j-k)} - \mathrm{e}^{-\mathrm{i}\pi(j-k)}\big) = 0. \qquad \square$$

Korollar 1.17 (Charakterisierung von endlichen Maßen)**.** *Jedes endliche Maß μ auf $\mathbb{Z}$ ist durch die Werte*

$$\int \mathrm{e}^{\mathrm{i}tx}\,\mu(\mathrm{d}x), \qquad t \in (-\pi,\pi],$$

eindeutig festgelegt.

Dies folgt aus Satz 1.16, angewandt auf das normierte Maß $\mu/\mu(\mathbb{Z})$, sofern $\mu \neq 0$; der Beweis von (1.23) verwendet nur die Endlichkeit des Maßes.

Satz 1.18 (Allgemeine Inversionsformel)**.** *Sei X eine reellwertige Zufallsvariable mit charakteristischer Funktion φ. Dann gilt für alle $a<b$*

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\,\varphi(t)\,\mathrm{d}t = \mathbb{P}(X\in(a,b)) + \frac{1}{2}\mathbb{P}(X\in\{a,b\}). \tag{1.24}$$

Bemerkung. Die Existenz des Grenzwerts ist Teil der Aussage.

Beweis. Sei

$$I_T := \int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx}\,\mathbb{P}^X(\mathrm{d}x)\,\mathrm{d}t.$$

Wegen $(\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb})/(\mathrm{i}t) = \int_a^b \mathrm{e}^{-\mathrm{i}ty}\,\mathrm{d}y$ ist dieser Bruch betragsmäßig durch $b-a$ beschränkt. Nach dem Satz von Fubini gilt daher

$$I_T = \int_{\mathbb{R}}\int_{-T}^{T} \frac{\mathrm{e}^{\mathrm{i}t(x-a)} - \mathrm{e}^{\mathrm{i}t(x-b)}}{\mathrm{i}t}\,\mathrm{d}t\,\mathbb{P}^X(\mathrm{d}x).$$

Wir schreiben $\mathrm{e}^{\mathrm{i}t(x-a)} = \cos(t(x-a)) + \mathrm{i}\sin(t(x-a))$ und ebenso für b. Wegen $\cos(s) = \cos(-s)$ ist $t \mapsto (\cos(t(x-a)) - \cos(t(x-b)))/t$ ungerade, sein Integral über $[-T,T]$ verschwindet also. Wegen $\sin(-s) = -\sin(s)$ ist $t \mapsto \sin(t\theta)/t$ gerade. Es folgt

$$\begin{aligned} I_T &= \int_{\mathbb{R}} \Big\{ \underbrace{\int_{-T}^{T} \frac{\sin(t(x-a))}{t}\,\mathrm{d}t}_{=:R(x-a,T)} - \int_{-T}^{T} \frac{\sin(t(x-b))}{t}\,\mathrm{d}t \Big\}\,\mathbb{P}^X(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \big(R(x-a,T) - R(x-b,T)\big)\,\mathbb{P}^X(\mathrm{d}x). \end{aligned}$$

Nun gilt mit der Substitution $x = \theta t$

$$R(\theta,T) = 2\int_0^{T\theta} \frac{\sin x}{x}\,\mathrm{d}x = 2\,\mathrm{sgn}(\theta)\int_0^{T|\theta|} \frac{\sin x}{x}\,\mathrm{d}x,$$

wobei $\operatorname{sgn}(\theta) = 1$ für $\theta > 0$, $\operatorname{sgn}(\theta) = -1$ für $\theta < 0$ und $\operatorname{sgn}(0) = 0$. Das letzte Integral konvergiert für $T \to \infty$ gegen $\pi/2$ (Integralsinus). Daraus folgt

$$R(x-a,T) - R(x-b,T) \xrightarrow[T\to\infty]{} \begin{cases} 2\pi, & a < x < b, \\ \pi, & x = a \text{ oder } x = b, \\ 0, & x < a \text{ oder } x > b, \end{cases}$$

siehe Abbildung 1.2. Da $y \mapsto \int_0^y (\sin x/x)\,\mathrm{d}x$ stetig ist und für $y \to \infty$ konvergiert, ist sie auf $[0,\infty)$ beschränkt; also ist $|R(x-a,T) - R(x-b,T)|$ gleichmäßig in x und T beschränkt. Mit dominierter Konvergenz folgt

$$\frac{1}{2\pi} I_T \xrightarrow[T\to\infty]{} \mathbb{P}^X\big((a,b)\big) + \frac{1}{2}\mathbb{P}^X\big(\{a,b\}\big). \qquad \square$$

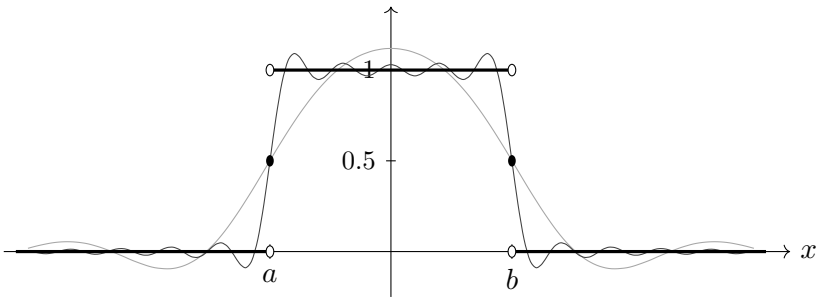

Abbildung 1.2.: Die Funktion $x \mapsto (1/(2\pi))\big(R(x-a,T) - R(x-b,T)\big)$ für $a = -1$, $b = 1$ und $T = 4$ (hellgrau), $T = 15$ (dunkelgrau) sowie ihr Grenzwert für $T \to \infty$ (schwarz). Der Wert $1/2$ an den Randpunkten erklärt den Term $(1/2)\mathbb{P}(X \in \{a,b\})$ in (1.24).

Satz 1.19 (Fourier-Inversion für Dichten und Satz von Riemann–Lebesgue)**.** *Sei X eine reellwertige Zufallsvariable mit charakteristischer Funktion φ^X. Dann gilt:*

(i) Ist φ^X integrierbar über $\mathbb{R}$, so besitzt $\mathbb{P}^X$ die Lebesgue-Dichte

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}tx}\varphi^X(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$$

(ii) Besitzt $\mathbb{P}^X$ eine Dichte f_X, so gilt

$$\varphi^X(t) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} f_X(x)\,\mathrm{d}x \xrightarrow[|t|\to\infty]{} 0.$$

Beweisideen. Zu (i): Wir leiten (1.24) ab und wählen dazu $a = x$ und $b = x + h$. Mit dominierter Konvergenz gilt

$$\begin{aligned}\lim_{h\downarrow 0}\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{1}{\mathrm{i}t}\big(\mathrm{e}^{-\mathrm{i}tx} - \mathrm{e}^{-\mathrm{i}t(x+h)}\big)\varphi^X(t)\,\mathrm{d}t &= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\lim_{h\downarrow 0}\frac{1}{h}\int_x^{x+h}\mathrm{e}^{-\mathrm{i}ty}\,\mathrm{d}y\,\varphi^X(t)\,\mathrm{d}t \\ &= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\mathrm{e}^{-\mathrm{i}tx}\varphi^X(t)\,\mathrm{d}t = f_X(x),\end{aligned}$$

wobei der letzte Grenzwert existiert, da φ^X integrierbar ist. Als Majorante dient dabei $|\varphi^X|$, denn $|(1/h)\int_x^{x+h}\mathrm{e}^{-\mathrm{i}ty}\,\mathrm{d}y| \le 1$ und $\int_{-\infty}^{\infty}|\varphi^X(t)|\,\mathrm{d}t < \infty$. Genauer zeigt die Rechnung zusammen mit (1.24): Die linke Seite von (1.24) ist für $a = x$, $b = x + h$ betragsmäßig

höchstens $(h/(2\pi))\int|\varphi^X|\,\mathrm{d}t$, also hat $\mathbb{P}^X$ keine Atome, und die Verteilungsfunktion F von X erfüllt $(1/h)(F(x+h)-F(x))\to f_X(x)$ für $h\downarrow 0$. Da f_X nach dem Satz von Lebesgue stetig ist, ist F stetig differenzierbar mit Ableitung f_X, und f_X ist eine Dichte von $\mathbb{P}^X$.

Zu (ii): Sei zunächst f_X stetig und $t\neq 0$. Mit der Substitution $x\mapsto x+\pi/t$ und $\mathrm{e}^{\mathrm{i}\pi}=-1$ gilt

$$\varphi^X(t)=\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}tx}f_X(x)\,\mathrm{d}x=\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}tx}\mathrm{e}^{\mathrm{i}\pi}f_X\Big(x+\frac{\pi}{t}\Big)\,\mathrm{d}x=-\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}tx}f_X\Big(x+\frac{\pi}{t}\Big)\,\mathrm{d}x.$$

Durch Mittelung der beiden Darstellungen folgt

$$\varphi^X(t)=\frac{1}{2}\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}tx}\Big(f_X(x)-f_X\Big(x+\frac{\pi}{t}\Big)\Big)\,\mathrm{d}x,$$

also

$$|\varphi^X(t)|\leq\frac{1}{2}\int_{\mathbb{R}}\Big|f_X(x)-f_X\Big(x+\frac{\pi}{t}\Big)\Big|\,\mathrm{d}x.$$

Der Integrand konvergiert für $|t|\to\infty$ wegen der Stetigkeit von f_X für jedes x gegen 0. Dass auch das Integral gegen 0 konvergiert, folgt allgemein aus der Stetigkeit der Translation in L^1: Für jede integrierbare Funktion f gilt $\int|f(x)-f(x+s)|\,\mathrm{d}x\to 0$ für $s\to 0$; dies sieht man durch Approximation von f durch stetige Funktionen mit kompaktem Träger. Insbesondere ist die Stetigkeit von f_X hier gar nicht nötig.

Alternativ kann man zu (ii) wie folgt argumentieren: Ist f_X differenzierbar mit integrierbarer Ableitung f_X', so gilt mit partieller Integration für $|t|>0$

$$\Big|\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}tx}f_X(x)\,\mathrm{d}x\Big|\leq\Big|f_X(x)\frac{1}{\mathrm{i}t}\mathrm{e}^{\mathrm{i}tx}\Big|_{-\infty}^{\infty}\Big|+\Big|\int_{\mathbb{R}}\frac{1}{\mathrm{i}t}\mathrm{e}^{\mathrm{i}tx}f_X'(x)\,\mathrm{d}x\Big|\leq 0+\frac{1}{|t|}\int_{\mathbb{R}}|f_X'(x)|\,\mathrm{d}x\xrightarrow[|t|\to\infty]{}0,$$

denn aus der Integrierbarkeit von f_X und f_X' folgt $f_X(x)\to 0$ für $|x|\to\infty$. □

2. Konvergenz von Zufallsvariablen

Zu diesem Kapitel vergleiche man [Kle13, Kapitel 6 und 13] sowie [Bil99].

2.1. Konvergenzbegriffe

Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X_0, X_1, \dots$ Zufallsvariablen mit Werten in einem vollständigen, separablen metrischen Raum (S, d), einem sogenannten *polnischen Raum*, zum Beispiel $S = \mathbb{R}$, $S = \mathbb{R}^k$ oder $S = C[0,1]$. Wir statten S mit der Borel-σ-Algebra $\mathcal{B}(S)$ aus, welche sich aus der Metrik ergibt: Die Metrik erzeugt die offenen Mengen, und diese erzeugen die Borel-σ-Algebra. Die offenen Mengen erzeugen zudem, als topologischer Raum, einen Konvergenzbegriff: Eine Folge $(x_n)_{n\in\mathbb{N}} \subset S$ konvergiert genau dann gegen $x \in S$, wenn $d(x_n, x) \to 0$ für $n \to \infty$. Ein wichtiger Spezialfall sind normierte Räume mit $d(x, y) = \|x - y\|$.

Definition 2.1 (Fast sichere Konvergenz)**.** Eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen mit Werten in S *konvergiert* $\mathbb{P}$*-fast sicher* (f.s.) gegen eine Zufallsvariable X_0, falls

$$\mathbb{P}\big(\{\omega \in \Omega\colon X_n(\omega) \xrightarrow[n\to\infty]{} X_0(\omega)\}\big) = 1.$$

Man schreibt $X_n \xrightarrow[n\to\infty]{} X_0$ f.s.

Bemerkung 2.2. Stets gilt $A \coloneqq \{\omega\colon X_n(\omega) \to X_0(\omega)\} \in \mathcal{F}$, denn

$$A = \bigcap_{k\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \underbrace{\Big\{\omega\colon d(X_n(\omega), X_0(\omega)) < \frac{1}{k}\Big\}}_{=:A_{n,k}},$$

und es gilt $A_{n,k} \in \mathcal{F}$, da S separabel ist, siehe [Kle13, Lemma 6.1]. Für separables S gilt nämlich $\mathcal{B}(S \times S) = \mathcal{B}(S) \otimes \mathcal{B}(S)$, sodass (X_n, X_0) messbar in $S \times S$ und damit $d(X_n, X_0)$ eine reelle Zufallsvariable ist. Also gilt

$$\begin{aligned} X_n \to X_0 \text{ f.s.} &\iff \mathbb{P}(A) = 1 \iff \mathbb{P}(A^{\mathrm{c}}) = 0 \iff \mathbb{P}\big(\{\omega\colon X_n(\omega) \not\to X_0(\omega)\}\big) = 0 \\ &\iff \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} \bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \Big\{\omega\colon d(X_n(\omega), X_0(\omega)) \geq \frac{1}{k}\Big\}\Big) = 0 \\ &\iff \mathbb{P}\Big(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \Big\{\omega\colon d(X_n(\omega), X_0(\omega)) \geq \frac{1}{k}\Big\}\Big) = 0 \quad \text{für alle } k \in \mathbb{N}. \end{aligned}$$

Lemma 2.3 (Eindeutigkeit des f.s. Grenzwerts)**.** *Gilt $X_n \to X_0$ f.s. und $X_n \to Y_0$ f.s., so folgt $X_0 = Y_0$ f.s., das heißt,* $\mathbb{P}(\{\omega\colon X_0(\omega) = Y_0(\omega)\}) = 1$.

Beweis. Es gilt

$$\{\omega\colon X_0(\omega) \neq Y_0(\omega)\} \subset \{\omega\colon X_n(\omega) \not\to X_0(\omega)\} \cup \{\omega\colon X_n(\omega) \not\to Y_0(\omega)\},$$

und beide Mengen auf der rechten Seite haben Wahrscheinlichkeit 0. □

Lemma 2.4 (F.s. Konvergenz unter stetigen Abbildungen)**.** *Gilt $X_n \to X_0$ f.s. und ist $f\colon S \to S'$ stetig, wobei S' ein weiterer polnischer Raum ist, so gilt*

$$f \circ X_n \xrightarrow[n\to\infty]{} f \circ X_0 \quad \text{f.s.}$$

Beweis. Wegen der Stetigkeit von f gilt $\{\omega\colon f(X_n(\omega)) \not\to f(X_0(\omega))\} \subset \{\omega\colon X_n(\omega) \not\to X_0(\omega)\}$. □

Beispiel. (i) Da $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto x+y$ (oder $x-y$ oder $x\cdot y$) stetig ist, gilt: Aus $X_n \to X_0$ f.s. und $Y_n \to Y_0$ f.s. folgt $X_n+Y_n \to X_0+Y_0$ f.s. Dabei wenden wir Lemma 2.4 auf $(X_n,Y_n) \to (X_0,Y_0)$ f.s. in $\mathbb{R}^2$ an; dies gilt, da der Schnitt zweier Ereignisse mit Wahrscheinlichkeit 1 wieder Wahrscheinlichkeit 1 hat.

(ii) Da $f\colon \mathbb{R}^k \to \mathbb{R}$, $(x_1,\dots,x_k) \mapsto \max(x_1,\dots,x_k)$ (oder min) stetig ist, gilt: Aus

$$\boldsymbol{X}_n = \begin{pmatrix} X_{n,1} \\ \vdots \\ X_{n,k} \end{pmatrix} \xrightarrow[n\to\infty]{} \boldsymbol{X}_0 \quad \text{f.s.}$$

folgt

$$\max(X_{n,j}\colon j=1,\dots,k) \xrightarrow[n\to\infty]{} \max(X_{0,j}\colon j=1,\dots,k) \quad \text{f.s.}$$

Satz 2.5 (Charakterisierung von f.s. Konvergenz)**.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen. Dann gilt*

$$X_n \xrightarrow[n\to\infty]{} X_0 \text{ f.s.} \quad \Longleftrightarrow \quad \lim_{m\to\infty} \mathbb{P}\Big(\sup_{n\geq m} d(X_n,X_0) > \varepsilon\Big) = 0 \quad \text{für alle } \varepsilon > 0.$$

Beweis. Es gilt

$$\mathbb{P}\Big(\sup_{n\geq m} d(X_n,X_0) > \varepsilon\Big) = \mathbb{P}\Big(\bigcup_{n\geq m} \{d(X_n,X_0) > \varepsilon\}\Big)$$

konvergiert für $m \to \infty$ wegen der Stetigkeit von Maßen von oben gegen

$$\mathbb{P}\Big(\bigcap_{m\geq 1}\bigcup_{n\geq m} \{d(X_n,X_0) > \varepsilon\}\Big) =: q_\varepsilon.$$

Nun gilt $q_\varepsilon = 0$ für alle $\varepsilon > 0$ genau dann, wenn $X_n \to X_0$ f.s., nach Bemerkung 2.2. Dabei ist es unerheblich, ob dort $> \varepsilon$ oder $\geq 1/k$ steht, da beide Bedingungen für alle $\varepsilon > 0$ beziehungsweise alle $k \in \mathbb{N}$ gefordert werden. □

Satz 2.6 (Cauchy-Kriterium für die f.s. Konvergenz)**.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen. Dann gilt*

$$\lim_{m\to\infty} \mathbb{P}\Big(\sup_{n\geq m} d(X_n,X_m) > \varepsilon\Big) = 0 \quad \text{für alle } \varepsilon > 0$$

genau dann, wenn eine Zufallsvariable X_0 auf $(\Omega,\mathcal{F},\mathbb{P})$ mit Werten in S existiert, sodass $X_n \to X_0$ f.s.

Beweis. „⇐“: Mit der Dreiecksungleichung und $\mathbb{P}(X+Y>c) \leq \mathbb{P}(X>c/2)+\mathbb{P}(Y>c/2)$ gilt

$$\begin{aligned} 0 \leq \mathbb{P}\Big(\sup_{n\geq m} d(X_n,X_m) > \varepsilon\Big) &\leq \mathbb{P}\Big(\sup_{n\geq m} d(X_n,X_0)+d(X_m,X_0) > \varepsilon\Big) \\ &\leq 2\,\mathbb{P}\Big(\sup_{n\geq m} d(X_n,X_0) > \frac{\varepsilon}{2}\Big) \xrightarrow[m\to\infty]{} 0 \end{aligned}$$

wegen Satz 2.5.

„⇒“: Für $n \in \mathbb{N}$ setzen wir

$$Y_n \coloneqq \sup_{\ell,k \geq n} d(X_\ell, X_k).$$

Dann ist $(Y_n)_{n\in\mathbb{N}}$ eine fallende Folge von Zufallsvariablen mit Werten in $[0,\infty]$. Für alle $\varepsilon > 0$ gilt

$$\begin{aligned}\mathbb{P}\Big(\sup_{n\geq m} |Y_n| > \varepsilon\Big) = \mathbb{P}(Y_m > \varepsilon) &\leq \mathbb{P}\Big(\sup_{\ell,k\geq m} d(X_\ell, X_m) + \sup_{\ell,k\geq m} d(X_k, X_m) > \varepsilon\Big) \\ &\leq 2\,\mathbb{P}\Big(\sup_{n\geq m} d(X_n, X_m) > \frac{\varepsilon}{2}\Big) \xrightarrow[m\to\infty]{} 0\end{aligned}$$

nach Voraussetzung. Also gilt $Y_n \to 0$ f.s. wie in Satz 2.5. (Direkt: $\mathbb{P}(\bigcap_m \{Y_m > 1/k\}) \leq \inf_m \mathbb{P}(Y_m > 1/k) = 0$ für alle $k \in \mathbb{N}$.) Das heißt, es existiert ein $N \in \mathcal{F}$ mit $\mathbb{P}(N) = 0$ und

$$\sup_{\ell,k\geq n} d\big(X_\ell(\omega), X_k(\omega)\big) \xrightarrow[n\to\infty]{} 0 \qquad \text{für alle } \omega \in N^{\mathrm{c}}.$$

Damit ist $(X_n(\omega))_{n\in\mathbb{N}}$ eine Cauchy-Folge für alle $\omega \in N^{\mathrm{c}}$. Da S vollständig ist, existiert ein $s(\omega) \in S$ mit $X_n(\omega) \to s(\omega)$ für $n \to \infty$. Für ein beliebiges $s_0 \in S$ definieren wir

$$X_0(\omega) \coloneqq s(\omega)\mathbb{1}\{\omega \in N^{\mathrm{c}}\} + s_0\mathbb{1}\{\omega \in N\}.$$

Als punktweiser Grenzwert messbarer Abbildungen auf N^{c} ist X_0 messbar. Dann gilt $X_n(\omega) \to X_0(\omega)$ für alle $\omega \in N^{\mathrm{c}}$, also $X_n \to X_0$ f.s. □

Definition 2.7 (Konvergenz nach Wahrscheinlichkeit)**.** Wir sagen, X_n konvergiert für $n \to \infty$ *nach Wahrscheinlichkeit* (auch *stochastisch*) gegen X_0, falls

$$\lim_{n\to\infty} \underbrace{\mathbb{P}\big(\{\omega\colon d(X_n(\omega), X_0(\omega)) > \varepsilon\}\big)}_{=\mathbb{P}(d(X_n,X_0)>\varepsilon)} = 0 \qquad \text{für alle } \varepsilon > 0,$$

und schreiben $X_n \xrightarrow[n\to\infty]{} X_0$ n.W.

Bemerkung. Für unabhängige und identisch verteilte reelle Zufallsvariablen $X_1, X_2, \ldots$ mit $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ und $\mathrm{Var}[X_1] = \sigma^2 < \infty$ gilt

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{} \mathbb{E}[X_1] = \mu \quad \text{f.s.} \qquad \text{(starkes Gesetz der großen Zahlen)}$$

und

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{} \mu \quad \text{n.W.} \qquad \text{(schwaches Gesetz der großen Zahlen).}$$

Für das schwache Gesetz der großen Zahlen genügt die paarweise Unkorreliertheit, siehe Satz 0.14.

Bemerkung 2.8. (i) Mit Satz 2.5 folgt

$$X_n \to X_0 \text{ f.s.} \quad \Longrightarrow \quad X_n \to X_0 \text{ n.W.},$$

denn $\mathbb{P}(d(X_m, X_0) > \varepsilon) \leq \mathbb{P}(\sup_{n\geq m} d(X_n, X_0) > \varepsilon)$. Die Umkehrung ist im Allgemeinen falsch, siehe das folgende Beispiel.

(ii) Ist $X_0 = s \in S$ konstant, so kann die Konvergenz nach Wahrscheinlichkeit auch für Zufallsvariablen auf verschiedenen Wahrscheinlichkeitsräumen erklärt werden, denn $\mathbb{P}(d(X_n, s) > \varepsilon)$ hängt nur von der Verteilung von X_n ab.

Beispiel (Wandernde Indikatoren)**.** Seien $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \mathrm{Leb}_{[0,1]})$ und für $n = 2^j + i$ mit $j \in \mathbb{N}_0$ und $i \in \{0, \dots, 2^j - 1\}$

$$X_n := \mathbb{1}_{[i2^{-j},(i+1)2^{-j}]}.$$

Dann gilt $\mathbb{P}(|X_n| > \varepsilon) \le 2^{-j} \to 0$ für $n \to \infty$, also $X_n \to 0$ n.W. Andererseits durchläuft das Intervall für jedes feste j ganz $[0,1]$, siehe Abbildung 2.1; für jedes ω gilt also $X_n(\omega) = 1$ für unendlich viele n und $X_n(\omega) = 0$ für unendlich viele n. Insbesondere konvergiert $(X_n(\omega))_{n\in\mathbb{N}}$ für kein ω.

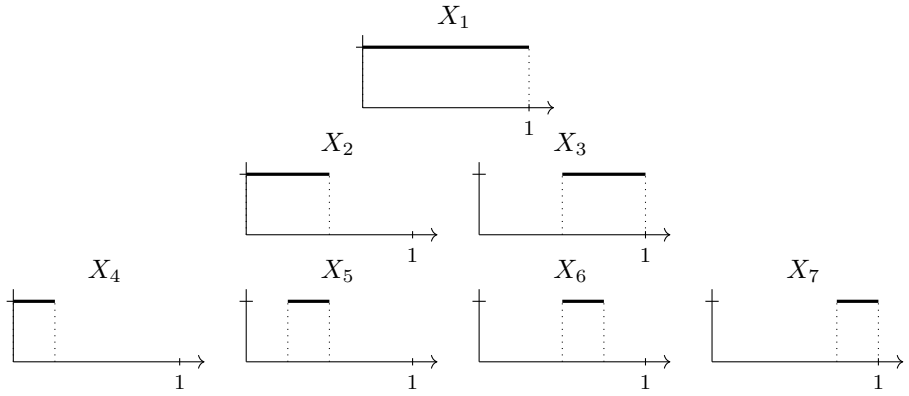

Abbildung 2.1.: Die wandernden Indikatoren $X_1, \dots, X_7$ auf $[0,1]$. Die Träger werden immer kleiner (Konvergenz nach Wahrscheinlichkeit gegen 0), überstreichen aber in jeder Zeile ganz $[0,1]$ (keine f.s. Konvergenz).

Lemma 2.9 (Teilfolgenkriterium)**.** *Die beiden folgenden Aussagen sind äquivalent.*

(i) Es gilt $X_n \to X_0$ n.W.

(ii) Zu jeder Teilfolge $(X_{n_k})_{k\in\mathbb{N}}$ von $(X_n)_{n\in\mathbb{N}}$ existiert eine Teil-Teilfolge $(X_{n_{k_m}})_{m\in\mathbb{N}}$ von $(X_{n_k})_{k\in\mathbb{N}}$ mit $X_{n_{k_m}} \xrightarrow[m\to\infty]{} X_0$ f.s.

Beweis. (i) $\Rightarrow$ (ii): Sei $(X_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge von $(X_n)_{n\in\mathbb{N}}$. Für alle $\varepsilon > 0$ gilt nach Voraussetzung $\lim_{k\to\infty} \mathbb{P}(d(X_{n_k}, X_0) > \varepsilon) = 0$. Wir wählen eine streng monoton wachsende Folge $(k_m)_{m\in\mathbb{N}}$ wie folgt:

$$\mathbb{P}\Big(d(X_{n_{k_m}}, X_0) > \frac{1}{m}\Big) \le \frac{1}{m^2} \qquad \text{für alle } m \in \mathbb{N}.$$

Sei nun $\varepsilon > 0$ und $m_0 \in \mathbb{N}$ mit $1/m_0 \le \varepsilon$. Dann gilt für alle $m \ge m_0$

$$\begin{aligned}\mathbb{P}\Big(\sup_{\ell\ge m} d(X_{n_{k_\ell}}, X_0) > \varepsilon\Big) &\le \mathbb{P}\Big(\bigcup_{\ell\ge m} \{d(X_{n_{k_\ell}}, X_0) > \varepsilon\}\Big) \le \sum_{\ell=m}^{\infty} \mathbb{P}\Big(d(X_{n_{k_\ell}}, X_0) > \frac{1}{\ell}\Big) \\ &\le \sum_{\ell=m}^{\infty} \frac{1}{\ell^2} \xrightarrow[m\to\infty]{} 0,\end{aligned}$$

da $\sum_{\ell\ge1} \ell^{-2} < \infty$. Nach Satz 2.5 folgt $X_{n_{k_m}} \to X_0$ f.s. für $m \to \infty$.

(ii) $\Rightarrow$ (i): Seien $\varepsilon > 0$ und $a_n := \mathbb{P}(d(X_n, X_0) > \varepsilon)$. Wir verwenden das Teilfolgenkriterium für reelle Folgen $(a_n)_{n\in\mathbb{N}}$: Es gilt $a_n \to a$ genau dann, wenn zu jeder Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$

eine Teil-Teilfolge $(a_{n_{k_\ell}})_{\ell\in\mathbb{N}}$ mit $a_{n_{k_\ell}} \to a$ existiert. Da f.s. Konvergenz die Konvergenz nach Wahrscheinlichkeit impliziert, Bemerkung 2.8 (i), gilt

$$X_{n_{k_\ell}} \xrightarrow[\ell\to\infty]{} X_0 \text{ f.s.} \quad \Longrightarrow \quad X_{n_{k_\ell}} \xrightarrow[\ell\to\infty]{} X_0 \text{ n.W.},$$

also $a_{n_{k_\ell}} \to 0$. Somit gilt $a_n \to 0$, also $X_n \to X_0$ n.W. □

Bemerkung 2.10. Mit Lemma 2.9 gelten

(1) Lemma 2.3 (Eindeutigkeit des f.s. Grenzwerts) auch für den Grenzwert nach Wahrscheinlichkeit und

(2) Lemma 2.4 (f.s. Konvergenz unter stetigen Abbildungen) auch für die Konvergenz nach Wahrscheinlichkeit.

Satz 2.11 (Satz von Pratt, Sandwich-Folgen)**.** *Seien $(X_n)_{n\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}}$ und $(G_n)_{n\in\mathbb{N}}$ Folgen von Zufallsvariablen auf $(\Omega,\mathcal{F},\mathbb{P})$ mit Werten in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Es gelte:*

(i) $X_n \to X_0$ n.W., $g_n \to g_0$ n.W. und $G_n \to G_0$ n.W.;

(ii) $g_n \le X_n \le G_n$ f.s. für alle $n \in \mathbb{N}$;

(iii) $(X_n)_{n\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}}$, $(G_n)_{n\in\mathbb{N}}$, g_0 und G_0 sind integrierbar mit

$$\int g_n \,\mathrm{d}\mathbb{P} \to \int g_0 \,\mathrm{d}\mathbb{P} \qquad \text{und} \qquad \int G_n \,\mathrm{d}\mathbb{P} \to \int G_0 \,\mathrm{d}\mathbb{P}.$$

Dann ist X_0 integrierbar, und es gilt $\int X_n \,\mathrm{d}\mathbb{P} \to \int X_0 \,\mathrm{d}\mathbb{P}$.

Bemerkung 2.12 (Satz von Lebesgue unter Konvergenz nach Wahrscheinlichkeit). Wir erhalten den folgenden Spezialfall von Satz 2.11. Seien $(X_n)_{n\in\mathbb{N}}$ eine Folge integrierbarer Zufallsvariablen und G_0 eine integrierbare Zufallsvariable auf $(\Omega,\mathcal{F},\mathbb{P})$ mit Werten in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Es gelte

(i) $X_n \to X_0$ n.W. (nicht notwendig f.s.) und

(ii) $|X_n| \le G_0$ f.s. für alle $n \in \mathbb{N}$.

Dann ist X_0 integrierbar, und es gilt $\int X_n \,\mathrm{d}\mathbb{P} \to \int X_0 \,\mathrm{d}\mathbb{P}$. Zur Anwendung von Satz 2.11 wählen wir $G_n \equiv G_0$ und $g_n \equiv -G_0$.

Hinweis: Es gilt genau dann $\mathbb{P}(|X_n| \le G_0) = 1$ für alle $n \in \mathbb{N}$, wenn

$$\mathbb{P}\big(|X_n| \le G_0 \text{ für alle } n \in \mathbb{N}\big) = 1.$$

Denn die Mengen $N_n := \{|X_n| > G_0\}$ sind genau dann für alle n Nullmengen, wenn $N = \bigcup_{n\in\mathbb{N}} N_n$ eine Nullmenge ist (abzählbare Vereinigung).

Beweis von Satz 2.11. Nach Lemma 2.9 (Teilfolgenkriterium) existiert zu jeder Teilfolge $(n_k)_{k\in\mathbb{N}}$ eine Teil-Teilfolge $(n_{k_\ell})_{\ell\in\mathbb{N}}$, sodass die zugehörigen Zufallsvariablen f.s. konvergieren. Dazu wenden wir Lemma 2.9 nacheinander auf (X_n), (g_n) und (G_n) an und gehen jeweils zu einer weiteren Teilfolge über. Für diese Teil-Teilfolge gelten die Voraussetzungen von Satz 2.11 sogar f.s., das heißt, für alle $\omega \in N^{\mathrm{c}}$ für ein $N \in \mathcal{F}$ mit $\mathbb{P}(N) = 0$, wobei wir alle Zufallsvariablen auf N gleich null setzen. Insbesondere bleiben dabei die Integrale unverändert. Es gilt

$$X_{n_{k_\ell}} \xrightarrow[\ell\to\infty]{} X_0, \qquad g_{n_{k_\ell}} \xrightarrow[\ell\to\infty]{} g_0 \qquad \text{und} \qquad G_{n_{k_\ell}} \xrightarrow[\ell\to\infty]{} G_0 \qquad \text{f.s.}$$

Insbesondere folgt aus (ii), dass $g_0 \le X_0 \le G_0$, und daher $\mathbb{E}[|X_0|] \le \mathbb{E}[|g_0|] + \mathbb{E}[|G_0|] < \infty$.

Wir erinnern an das *Lemma von Fatou*: Für eine Folge $(f_n)_{n\in\mathbb{N}}$ messbarer Funktionen mit $f_n \ge 0$ für alle n auf einem Maßraum $(\Omega, \mathcal{F}, \mu)$ gilt

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

Angewandt auf $f_\ell = X_{n_{k_\ell}} - g_{n_{k_\ell}} \ge 0$ folgt

$$\begin{aligned}\mathbb{E}[X_0] - \mathbb{E}[g_0] &= \mathbb{E}\Big[\liminf_{\ell\to\infty}(X_{n_{k_\ell}} - g_{n_{k_\ell}})\Big] \le \liminf_{\ell\to\infty} \mathbb{E}\big[X_{n_{k_\ell}} - g_{n_{k_\ell}}\big] \\ &= \liminf_{\ell\to\infty} \mathbb{E}\big[X_{n_{k_\ell}}\big] - \mathbb{E}[g_0],\end{aligned}$$

wobei wir im letzten Schritt $\mathbb{E}[g_{n_{k_\ell}}] \to \mathbb{E}[g_0]$ verwendet haben. Also gilt

$$\mathbb{E}[X_0] \le \liminf_{\ell\to\infty} \mathbb{E}\big[X_{n_{k_\ell}}\big].$$

Ebenso erhalten wir mit $G_n - X_n \ge 0$ anstelle von $X_n - g_n$

$$\limsup_{\ell\to\infty} \mathbb{E}\big[X_{n_{k_\ell}}\big] \le \mathbb{E}[X_0],$$

also $\lim_{\ell\to\infty} \mathbb{E}[X_{n_{k_\ell}}] = \mathbb{E}[X_0]$. Mit dem Teilfolgenkriterium für reelle Folgen folgt daraus $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X_0]$. □

Definition 2.13 (Stochastische Äquivalenz)**.** Zwei Folgen $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(S, \mathcal{B}(S))$ heißen *stochastisch äquivalent*, falls

$$\lim_{n\to\infty} \mathbb{P}\big(d(X_n, Y_n) > \varepsilon\big) = 0 \qquad \text{für alle } \varepsilon > 0.$$

Beispiel 2.14. Seien $(Z_n)_{n\in\mathbb{N}}$ Zufallsvariablen mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $\mathbb{E}[Z_n] = \mu \in \mathbb{R}$ sowie $\mathrm{Var}[Z_n] < \infty$ für alle $n \in \mathbb{N}$. Dann gilt für alle $\varepsilon > 0$ mit der Markov-Ungleichung

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n Z_i - \mu\Big| > \varepsilon\Big) \le \frac{1}{\varepsilon^2}\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=1}^n (Z_i - \mu)\Big)^2\Big] = \frac{1}{n^2\varepsilon^2}\sum_{i,j=1}^n \underbrace{\mathbb{E}\big[(Z_i - \mu)(Z_j - \mu)\big]}_{=\mathrm{Cov}(Z_i, Z_j)}.$$

Sind die Z_n paarweise unkorreliert, so gilt

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n Z_i - \mu\Big| > \varepsilon\Big) \le \frac{1}{n^2\varepsilon^2}\sum_{i=1}^n \mathrm{Var}[Z_i].$$

Gilt zusätzlich $\frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[Z_i] \to 0$ für $n \to \infty$, so folgt $\frac{1}{n}\sum_{i=1}^n Z_i \to \mu$ n.W. Gilt sogar $\mathrm{Var}[Z_n] < K$ für alle n und ein $K > 0$, so folgt

$$\mathbb{P}\Big(n^\alpha\Big|\frac{1}{n}\sum_{i=1}^n Z_i - \mu\Big| > \varepsilon\Big) \le \frac{n^{2\alpha}K}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0 \qquad \text{für } \alpha \in [0, \tfrac{1}{2}).$$

Für $\alpha = 1/2$ und unabhängige, identisch verteilte Z_i mit $\mathrm{Var}[Z_1] = \sigma^2 < \infty$ gilt dagegen nach dem zentralen Grenzwertsatz von Lindeberg–Lévy, Satz 1.15,

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n (Z_i - \mu) \le x\Big) \xrightarrow[n\to\infty]{} \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}y \qquad \text{für alle } x \in \mathbb{R}.$$

Wir betrachten diesen Konvergenzbegriff nun genauer.

Definition 2.15 (Schwache Konvergenz)**.** Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(S, \mathcal{B}(S))$. Wir sagen, X_n *konvergiert schwach* gegen X_0, in Zeichen $X_n \xrightarrow{\mathcal{D}} X_0$, wenn

$$\int_S f \,\mathrm{d}\mathbb{P}^{X_n} \xrightarrow[n\to\infty]{} \int_S f \,\mathrm{d}\mathbb{P}^{X_0} \qquad \text{für alle } f \in C_\mathrm{b}(S).$$

Dabei bezeichnet $C_\mathrm{b}(S)$ die Menge der stetigen und beschränkten Funktionen $f\colon S \to \mathbb{R}$; diese sind insbesondere messbar.

Bemerkung 2.16. Die schwache Konvergenz ist ein Konvergenzbegriff für die Verteilungen $(\mathbb{P}^{X_n})_{n\in\mathbb{N}}$, deshalb schreibt man auch $\mathbb{P}^{X_n} \xrightarrow{\mathcal{D}} \mathbb{P}^{X_0}$.

Um Eigenschaften der schwachen Konvergenz zu untersuchen, benötigen wir zwei Hilfsaussagen aus der Maßtheorie.

Lemma 2.17 (Sandwicheigenschaft)**.** *Seien S ein metrischer Raum mit Metrik d und $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(S, \mathcal{B}(S))$. Dann gilt: Für alle $B \in \mathcal{B}(S)$ und alle $\varepsilon > 0$ existieren eine offene Menge $G \subset S$ und eine abgeschlossene Menge $F \subset S$ mit*

$$F \subset B \subset G \qquad \textit{und} \qquad \mathbb{P}(G \setminus F) < \varepsilon. \tag{2.1}$$

Hieraus folgt

$$\mathbb{P}(B) = \sup\{\mathbb{P}(F)\colon F \subset B \textit{ abgeschlossen}\} = \inf\{\mathbb{P}(G)\colon B \subset G \textit{ offen}\}.$$

Beweisidee. Für $\delta > 0$ ist $G_\delta := \{s \in S\colon d(s, B) < \delta\}$ offen. Ist B abgeschlossen, so gilt $G_\delta \downarrow B$ für $\delta \downarrow 0$ und damit $\mathbb{P}(G_\delta) \to \mathbb{P}(B)$. Das heißt, (2.1) gilt für $F = B \subset G_{\delta_0}$, wenn $\delta_0 > 0$ klein genug ist. Setzt man $\mathcal{G} := \{B \in \mathcal{B}(S)\colon (2.1) \text{ gilt}\}$, so kann man zeigen, dass $\mathcal{G}$ eine σ-Algebra ist, die alle abgeschlossenen Mengen in $\mathcal{B}(S)$ enthält. Die abgeschlossenen Mengen erzeugen jedoch $\mathcal{B}(S)$, und damit gilt $\mathcal{G} = \mathcal{B}(S)$. □

Lemma 2.18 (Straffheit von Wahrscheinlichkeitsmaßen)**.** *In der Situation von Lemma 2.17, wobei S zusätzlich polnisch sei, gilt: Für alle $\varepsilon > 0$ existiert eine kompakte Menge $K_\varepsilon \subset S$ mit $\mathbb{P}(K_\varepsilon) \geq 1 - \varepsilon$. Man sagt, $\mathbb{P}$ ist* straff *(englisch* tight*). Für jedes $B \in \mathcal{B}(S)$ gilt sogar*

$$\mathbb{P}(B) = \sup\{\mathbb{P}(K)\colon K \subset B \textit{ kompakt}\}.$$

Beweisidee. Für $S = \mathbb{R}^k$ und $K_M = [-M, M]^k$ gilt $K_M \uparrow \mathbb{R}^k$ für $M \to \infty$, und die Aussage folgt aus der Stetigkeit von Maßen. □

Satz 2.19 (Approximation durch stetige Funktionen)**.** *Seien $F \subset S$ abgeschlossen und $\varepsilon > 0$. Dann existiert eine Abbildung $f \in C_\mathrm{b}(S)$ mit*

$$\begin{aligned} f(x) &= 1 && \textit{für alle } x \in F, \\ f(x) &= 0 && \textit{für alle } x \in S \textit{ mit } d(x, F) \geq \varepsilon, \\ 0 \leq f(x) &\leq 1 && \textit{für alle } x \in S. \end{aligned}$$

Die Abbildung f kann sogar gleichmäßig stetig gewählt werden.

Beweis. Wir definieren

$$\varphi(t) = \begin{cases} 1, & t \leq 0, \\ 1 - t, & 0 < t < 1, \\ 0, & t \geq 1. \end{cases} \tag{2.2}$$

Dann hat die Abbildung $f\colon S \to [0, 1]$, $x \mapsto f(x) := \varphi((1/\varepsilon)d(x, F))$, die gewünschten Eigenschaften, siehe Abbildung 2.2. Da $x \mapsto d(x, F)$ Lipschitz-stetig mit Konstante 1 und φ Lipschitz-stetig mit Konstante 1 ist, ist f Lipschitz-stetig mit Konstante $1/\varepsilon$, also insbesondere gleichmäßig stetig. □

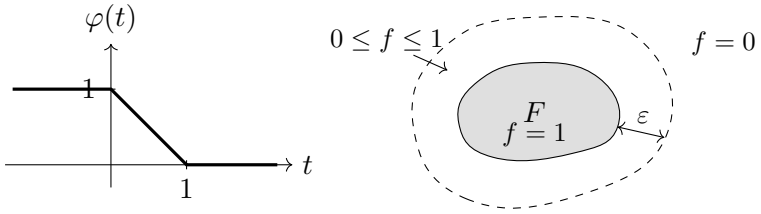

Abbildung 2.2.: Links: die Funktion φ aus (2.2). Rechts: Die Funktion $f = \varphi(d(\cdot, F)/\varepsilon)$ ist gleich 1 auf F, gleich 0 außerhalb der ε-Umgebung von F (gestrichelt) und fällt dazwischen linear im Abstand zu F ab.

Satz 2.20 (Eindeutigkeitssatz der schwachen Konvergenz)**.** *Seien $\mathbb{P}$ und Q zwei Wahrscheinlichkeitsmaße auf $(S, \mathcal{B}(S))$ mit*

$$\int_S f \,\mathrm{d}\mathbb{P} = \int_S f \,\mathrm{d}Q \qquad \textit{für alle } f \in C_\mathrm{b}(S).$$

Dann gilt $\mathbb{P} = Q$.

Beweis. Wegen Lemma 2.17 genügt es, $\mathbb{P}(F) = Q(F)$ für alle abgeschlossenen Mengen $F \subset S$ zu zeigen. Mit φ wie im Beweis von Satz 2.19 und

$$f_n(x) = \varphi(n\, d(x, F)) \qquad \text{für alle } n \in \mathbb{N}$$

gilt $f_n \in C_\mathrm{b}(S)$ und $f_n \downarrow \mathbb{1}_F$ für $n \to \infty$. Für $x \notin F$ gilt nämlich $d(x, F) > 0$, da F abgeschlossen ist, und somit $f_n(x) = 0$ für $n \ge 1/d(x, F)$. Mit dem Satz von Lebesgue folgt

$$\mathbb{P}(F) = \int_S \mathbb{1}_F \,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_S f_n \,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_S f_n \,\mathrm{d}Q = Q(F),$$

wobei wir im dritten Schritt die Voraussetzung verwendet haben. □

Insbesondere ist der schwache Grenzwert eindeutig: Gilt $X_n \xrightarrow{\mathcal{D}} X_0$ und $X_n \xrightarrow{\mathcal{D}} Y_0$, so folgt $\mathbb{P}^{X_0} = \mathbb{P}^{Y_0}$.

Bemerkung. Mit dieser Aussage ist der im Beweis von Satz 1.8 (Eindeutigkeitssatz für charakteristische Funktionen) skizzierte Schritt $\mathbb{P}^{\boldsymbol{Z}_{j,\sigma}} \to \mathbb{P}^{\boldsymbol{X}_j}$ vollständig begründet; die Details sind eine Übung.

Definition 2.21 (Straffheit)**.** Eine Folge von Wahrscheinlichkeitsmaßen $(\mathbb{P}_n)_{n\in\mathbb{N}}$ auf $(S, \mathcal{B}(S))$ heißt *straff*, falls für alle $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subset S$ existiert mit

$$\mathbb{P}_n(K_\varepsilon) \ge 1 - \varepsilon \qquad \text{für alle } n \in \mathbb{N}.$$

Beispiel 2.22. Sei $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \mathrm{Leb}_{[0,1]})$. Wir definieren für $n \in \mathbb{N}$

$$X_{2n}\colon [0,1] \to \mathbb{R}, \quad \omega \mapsto \begin{cases} 1 + \frac{1}{2n}, & \frac{1}{2} < \omega \le 1, \\ -(1 + \frac{1}{2n}), & 0 \le \omega \le \frac{1}{2}, \end{cases}$$

und

$$X_{2n+1}\colon [0,1] \to \mathbb{R}, \quad \omega \mapsto \begin{cases} 1 + \frac{1}{2n+1}, & 0 \le \omega < \frac{1}{2}, \\ -(1 + \frac{1}{2n+1}), & \frac{1}{2} \le \omega \le 1. \end{cases}$$

Hier ist also $(S, \mathcal{B}(S)) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, siehe Abbildung 2.3; die Folge ist für $n \ge 2$ definiert. Für $\omega \ne 1/2$ wechselt $X_n(\omega)$ ab einem Index ständig das Vorzeichen, die Folge $(X_n(\omega))_{n\ge 2}$

konvergiert also nicht; nur für $\omega = 1/2$ gilt $X_n(1/2) = -(1 + 1/n) \to -1$. Da $\{1/2\}$ eine Nullmenge ist, konvergiert (X_n) nicht f.s. Aber für alle $f \in C_\mathrm{b}(\mathbb{R})$ gilt

$$\int f \,\mathrm{d}\mathbb{P}^{X_n} = \frac{1}{2}\Big(f\Big(1 + \frac{1}{n}\Big) + f\Big(-\Big(1 + \frac{1}{n}\Big)\Big)\Big), \tag{2.3}$$

da $\mathbb{P}^{X_n} = (1/2)\big(\delta_{1+1/n} + \delta_{-(1+1/n)}\big)$, wobei das Gewicht für gerade und ungerade n auf verschiedenen Teilen des Wahrscheinlichkeitsraums liegt. Wegen der Stetigkeit von f gilt

$$(2.3) \xrightarrow[n\to\infty]{} \frac{1}{2}(f(1) + f(-1)) = \int f \,\mathrm{d}\mathbb{P}^{X_0}$$

mit

$$X_0 \colon [0,1] \to \mathbb{R}, \quad \omega \mapsto \begin{cases} 1, & 0 \le \omega < \frac{1}{2}, \\ -1, & \frac{1}{2} \le \omega \le 1. \end{cases}$$

Also gilt $X_n \xrightarrow{\mathcal{D}} X_0$. Hier gilt aber $\mathrm{Leb}(|X_{2n} - X_0| > 1/2) = 1$ für alle n, also konvergiert X_n nicht nach Wahrscheinlichkeit gegen X_0.

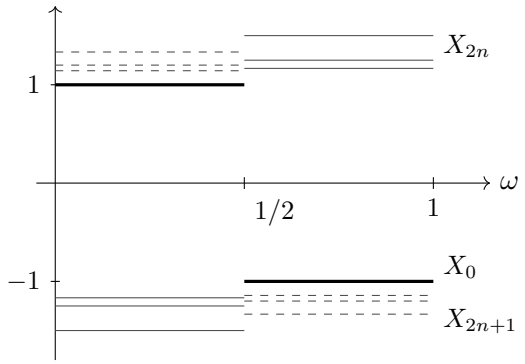

Abbildung 2.3.: Die Zufallsvariablen aus Beispiel 2.22: X_{2n} (durchgezogen, grau), X_{2n+1} (gestrichelt) für $n = 1, 2, 3$ und X_0 (schwarz). Die Werte springen zwischen den beiden Hälften von $[0, 1]$ hin und her, die Verteilungen konvergieren aber gegen $(1/2)(\delta_1 + \delta_{-1})$.

Wie bisher sei (S, d) ein metrischer Raum, versehen mit seiner Borel-σ-Algebra $\mathcal{B}(S)$. Für $x \in S$ und $F \subset S$ bezeichnet $d(x, F) \coloneqq \inf\{d(x, y) \colon y \in F\}$ den Abstand von x zu F. Für $B \subset S$ schreiben wir $\mathring{B}$ für das Innere, $\overline{B}$ für den Abschluss und $\partial B = \overline{B} \setminus \mathring{B}$ für den Rand von B.

Das folgende Resultat stellt mehrere äquivalente Charakterisierungen der schwachen Konvergenz aus Definition 2.15 zusammen.

Satz 2.23 (Portmanteau-Theorem, vgl. [Bil99, Theorem 2.1])**.** *Seien $(\mathbb{P}_n)_{n\in\mathbb{N}_0}$ Wahrscheinlichkeitsmaße auf $(S, \mathcal{B}(S))$. Dann sind die folgenden Aussagen äquivalent:*

(i) Es gilt $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$.

(ii) Für alle gleichmäßig stetigen $f \in C_\mathrm{b}(S)$ gilt $\int_S f \,\mathrm{d}\mathbb{P}_n \xrightarrow[n\to\infty]{} \int_S f \,\mathrm{d}\mathbb{P}_0$.

(iii) Für alle abgeschlossenen $F \subset S$ gilt $\limsup_{n\to\infty} \mathbb{P}_n(F) \le \mathbb{P}_0(F)$.

(iv) Für alle offenen $G \subset S$ gilt $\liminf_{n\to\infty} \mathbb{P}_n(G) \ge \mathbb{P}_0(G)$.

(v) Für alle $B \in \mathcal{B}(S)$ mit $\mathbb{P}_0(\partial B) = 0$ gilt $\lim_{n\to\infty} \mathbb{P}_n(B) = \mathbb{P}_0(B)$.

Eine Menge $B \in \mathcal{B}(S)$ mit $\mathbb{P}_0(\partial B) = 0$ heißt $\mathbb{P}_0$-*stetig* oder $\mathbb{P}_0$-*randlos*.

Beweis. Wir zeigen (i)⇒(ii)⇒(iii)⇒(i) sowie (iv)⇔(iii)⇔(v).

(i)⇒(ii): Dies ist ein Spezialfall der Definition, da gleichmäßig stetige Funktionen insbesondere stetig sind.

(ii)⇒(iii): Sei $F \subset S$ abgeschlossen. Für $\varepsilon > 0$ ist $G_\varepsilon := \{x \in S\colon d(x,F) < \varepsilon\}$ offen, und es gilt $G_\varepsilon \downarrow F$ für $\varepsilon \downarrow 0$, da F abgeschlossen ist. Da $\mathbb{P}_0$ stetig von oben ist, gibt es zu jedem $\delta > 0$ ein $\varepsilon > 0$ mit $\mathbb{P}_0(G_\varepsilon) < \mathbb{P}_0(F) + \delta$. Mit der Funktion φ aus (2.2) setzen wir $f(x) := \varphi((1/\varepsilon)d(x,F))$. Dann ist f gleichmäßig stetig, es gilt $f(x) = 1$ für alle $x \in F$, $f(x) = 0$ für alle $x \in G_\varepsilon^{\mathrm{c}}$ und $0 \le f(x) \le 1$ für alle $x \in S$. Also gilt

$$\mathbb{P}_n(F) = \int_F f \,\mathrm{d}\mathbb{P}_n \le \int_S f \,\mathrm{d}\mathbb{P}_n \xrightarrow[n\to\infty]{\text{(ii)}} \int_S f \,\mathrm{d}\mathbb{P}_0 = \int_{G_\varepsilon} f \,\mathrm{d}\mathbb{P}_0 \le \mathbb{P}_0(G_\varepsilon) < \mathbb{P}_0(F) + \delta.$$

Es folgt $\limsup_{n\to\infty} \mathbb{P}_n(F) \le \mathbb{P}_0(F) + \delta$, und da $\delta > 0$ beliebig war, gilt (iii).

(iii)⇒(i): Sei $f \in C_{\mathrm{b}}(S)$ gegeben. Wir zeigen zunächst

$$\limsup_{n\to\infty} \int_S f \,\mathrm{d}\mathbb{P}_n \le \int_S f \,\mathrm{d}\mathbb{P}_0. \tag{2.4}$$

Da f beschränkt ist und (2.4) unter Transformationen $f \mapsto af + b$ mit $a > 0$ und $b \in \mathbb{R}$ erhalten bleibt, dürfen wir ohne Einschränkung $0 < f(x) < 1$ für alle $x \in S$ annehmen. Für $k \in \mathbb{N}$ definieren wir die Niveaumengen

$$F_i := \Big\{x \in S\colon f(x) \ge \frac{i}{k}\Big\}, \qquad i = 0, 1, \dots, k.$$

Alle F_i sind abgeschlossen, da f stetig ist, und es gilt $F_0 = S$ sowie $F_k = \emptyset$. Mit $B_i := \{x \in S\colon (i-1)/k \le f(x) < i/k\} = F_{i-1} \setminus F_i$ gilt wegen $f(x) \in (0,1)$ für alle $x \in S$ und alle $n \in \mathbb{N}_0$

$$\sum_{i=1}^k \frac{i-1}{k}\, \mathbb{P}_n(B_i) \le \int_S f \,\mathrm{d}\mathbb{P}_n \le \sum_{i=1}^k \frac{i}{k}\, \mathbb{P}_n(B_i), \qquad \text{wobei } \mathbb{P}_n(B_i) = \mathbb{P}_n(F_{i-1}) - \mathbb{P}_n(F_i),$$

siehe Abbildung 2.4. Für die rechte Seite erhalten wir mit $F_0 = S$ und $F_k = \emptyset$

$$\frac{1}{k}\Big(\sum_{i=1}^k i\, \mathbb{P}_n(F_{i-1}) - \sum_{i=1}^k i\, \mathbb{P}_n(F_i)\Big) = \frac{1}{k}\Big(1 + \sum_{i=1}^k (i+1)\mathbb{P}_n(F_i) - \sum_{i=1}^k i\, \mathbb{P}_n(F_i)\Big) = \frac{1}{k} + \frac{1}{k}\sum_{i=1}^k \mathbb{P}_n(F_i).$$

Eine ähnliche Rechnung für die untere Schranke liefert insgesamt

$$\frac{1}{k}\sum_{i=1}^k \mathbb{P}_n(F_i) \le \int_S f \,\mathrm{d}\mathbb{P}_n \le \frac{1}{k}\sum_{i=1}^k \mathbb{P}_n(F_i) + \frac{1}{k}.$$

Mit (iii) gilt also für alle $k \in \mathbb{N}$

$$\begin{aligned}\limsup_{n\to\infty} \int_S f \,\mathrm{d}\mathbb{P}_n &\le \frac{1}{k} + \frac{1}{k}\sum_{i=1}^k \limsup_{n\to\infty} \mathbb{P}_n(F_i) \le \frac{1}{k} + \frac{1}{k}\sum_{i=1}^k \mathbb{P}_0(F_i) \\ &\le \frac{1}{k} + \int_S f \,\mathrm{d}\mathbb{P}_0 \xrightarrow[k\to\infty]{} \int_S f \,\mathrm{d}\mathbb{P}_0,\end{aligned}$$

also (2.4). Eine Anwendung von (2.4) auf $-f$ liefert

$$\liminf_{n\to\infty} \int_S f \,\mathrm{d}\mathbb{P}_n \ge \int_S f \,\mathrm{d}\mathbb{P}_0,$$

und daraus folgt $\int_S f \,\mathrm{d}\mathbb{P}_n \to \int_S f \,\mathrm{d}\mathbb{P}_0$, also (i).

(iii)$\Leftrightarrow$(iv): Dies ergibt sich durch Bildung von Komplementen, da $\mathbb{P}_n(G) = 1 - \mathbb{P}_n(G^\mathrm{c})$ gilt und G genau dann offen ist, wenn G^c abgeschlossen ist.

(iii)$\Rightarrow$(v): Sei $B \in \mathcal{B}(S)$ eine $\mathbb{P}_0$-stetige Menge. Dann gilt mit (iii) und der bereits gezeigten Aussage (iv)

$$\mathbb{P}_0(\overline{B}) \geq \limsup_{n\to\infty} \mathbb{P}_n(\overline{B}) \geq \limsup_{n\to\infty} \mathbb{P}_n(B) \geq \liminf_{n\to\infty} \mathbb{P}_n(B) \geq \liminf_{n\to\infty} \mathbb{P}_n(\mathring{B}) \geq \mathbb{P}_0(\mathring{B}).$$

Wegen $\mathbb{P}_0(\partial B) = 0$ gilt $\mathbb{P}_0(\overline{B}) = \mathbb{P}_0(B) = \mathbb{P}_0(\mathring{B})$, und damit folgt (v).

(v)$\Rightarrow$(iii): Sei $F \subset S$ abgeschlossen. Für alle $\delta > 0$ gilt

$$\partial\{x \in S \colon d(x, F) \leq \delta\} \subset \{x \in S \colon d(x, F) = \delta\}.$$

Die Mengen $\{x \in S \colon d(x, F) = \delta\}$ sind für verschiedene δ disjunkt, also können höchstens abzählbar viele dieser Mengen positives $\mathbb{P}_0$-Maß haben (siehe Bemerkung 2.25). Deshalb gibt es eine monoton fallende Nullfolge $(\delta_k)_{k\in\mathbb{N}}$, sodass die Mengen $F_k \coloneqq \{x \in S \colon d(x, F) \leq \delta_k\}$ $\mathbb{P}_0$-stetig sind. Also gilt für alle $k \in \mathbb{N}$

$$\limsup_{n\to\infty} \mathbb{P}_n(F) \leq \limsup_{n\to\infty} \mathbb{P}_n(F_k) \overset{\text{(v)}}{=} \mathbb{P}_0(F_k).$$

Da F abgeschlossen ist, gilt $F_k \downarrow F$ für $k \to \infty$, also $\mathbb{P}_0(F_k) \to \mathbb{P}_0(F)$, und es folgt (iii). □

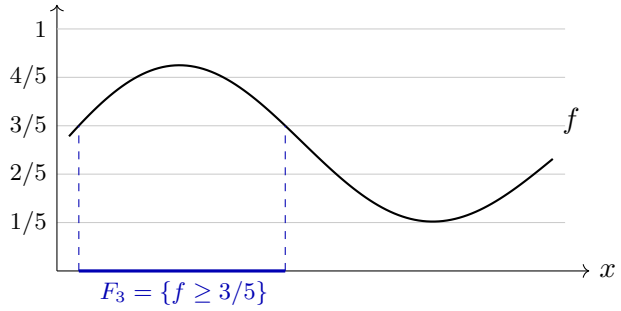

Abbildung 2.4.: Approximation von $\int f \,\mathrm{d}\mathbb{P}_n$ durch die Niveaumengen $F_i = \{f \geq i/k\}$ im Beweis von Satz 2.23, hier für $k = 5$. Auf $\{(i-1)/k \leq f < i/k\} = F_{i-1} \setminus F_i$ liegt f zwischen $(i-1)/k$ und i/k.

Bemerkung. Man bezeichnet die Klasse $\{A \in \mathcal{B}(S) \colon \mathbb{P}_0(\partial A) = 0\}$ als *konvergenzbestimmend*, da aus $\mathbb{P}_n(A) \to \mathbb{P}_0(A)$ für alle $\mathbb{P}_0$-randlosen A bereits $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$ folgt.

Satz 2.24 (Konvergenzbestimmende Mengensysteme)**.** *Sei $(\mathbb{P}_n)_{n\in\mathbb{N}_0}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(S, \mathcal{B}(S))$. Sei ferner $\mathcal{G} \subset \mathcal{B}(S)$ gegeben mit den folgenden Eigenschaften:*

(i) Für alle $A, B \in \mathcal{G}$ gilt $A \cap B \in \mathcal{G}$.

(ii) Zu jeder offenen Menge $G \subset S$ gibt es $A_1, A_2, \ldots \in \mathcal{G}$ mit $G = \bigcup_{i=1}^\infty A_i$.

(iii) Für alle $A \in \mathcal{G}$ gilt $\mathbb{P}_n(A) \xrightarrow[n\to\infty]{} \mathbb{P}_0(A)$.

Dann gilt $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$.

Bemerkung. Ist S separabel, so kann in Satz 2.24 die Bedingung (ii) ersetzt werden durch

(ii′) Für alle $x \in S$ und alle $\varepsilon > 0$ gibt es ein $A \in \mathcal{G}$ mit $x \in \mathring{A} \subset A \subset \{y \in S \colon d(x,y) < \varepsilon\}$.

Beweis von Satz 2.24. Sei $G \subset S$ offen. Nach (ii) gibt es $A_1, A_2, \ldots \in \mathcal{G}$ mit $G = \bigcup_{i=1}^{\infty} A_i$. Wegen der Stetigkeit von $\mathbb{P}_0$ von unten gibt es zu jedem $\varepsilon > 0$ ein $m \in \mathbb{N}$ mit $\mathbb{P}_0\big(\bigcup_{i=1}^{m} A_i\big) \geq \mathbb{P}_0(G) - \varepsilon$. Dann gilt

$$\mathbb{P}_0(G) - \varepsilon \leq \mathbb{P}_0\Big(\bigcup_{i=1}^{m} A_i\Big) = \lim_{n\to\infty} \mathbb{P}_n\Big(\bigcup_{i=1}^{m} A_i\Big) \leq \liminf_{n\to\infty} \mathbb{P}_n(G).$$

Für die Gleichheit in der Mitte verwenden wir das Ein-Ausschluss-Prinzip, etwa für $m = 2$

$$\mathbb{P}_n(A_1 \cup A_2) = \mathbb{P}_n(A_1) + \mathbb{P}_n(A_2) - \mathbb{P}_n(A_1 \cap A_2).$$

Allgemein ist $\mathbb{P}_n\big(\bigcup_{i=1}^{m} A_i\big)$ eine endliche Summe von Termen $\pm\mathbb{P}_n(A_{i_1} \cap \cdots \cap A_{i_j})$; nach (i) liegen alle diese Schnitte in $\mathcal{G}$, und nach (iii) konvergiert jeder Term für $n \to \infty$ gegen den entsprechenden Term mit $\mathbb{P}_0$. Da $\varepsilon > 0$ beliebig war, folgt $\liminf_{n\to\infty} \mathbb{P}_n(G) \geq \mathbb{P}_0(G)$, also gilt Satz 2.23 (iv). □

Bemerkung 2.25. (i) Ist $\mathcal{G} \subset \mathcal{B}(S)$ abgeschlossen unter endlicher Durchschnittsbildung, so gilt dies auch für $\{A \in \mathcal{G} \colon A \text{ ist } \mathbb{P}_0\text{-randlos}\}$, denn

$$\mathbb{P}_0\big(\partial(A \cap B)\big) \leq \mathbb{P}_0(\partial A \cup \partial B) \leq \mathbb{P}_0(\partial A) + \mathbb{P}_0(\partial B).$$

(ii) Ist $\mathcal{G} = \{A_i \colon i \in I\} \subset \mathcal{B}(S)$ eine Familie von Mengen mit paarweise disjunkten Rändern ∂A_i, so sind höchstens abzählbar viele Mengen aus $\mathcal{G}$ nicht $\mathbb{P}_0$-randlos. In der Tat, setzen wir $I_k \coloneqq \{A \in \mathcal{G} \colon \mathbb{P}_0(\partial A) \geq 1/k\}$, so gilt

$$\{A \in \mathcal{G} \colon \mathbb{P}_0(\partial A) > 0\} = \bigcup_{k\in\mathbb{N}} I_k \qquad \text{und} \qquad \#I_k \leq k,$$

wobei die Abschätzung $\#I_k \leq k$ aus der Disjunktheit der Ränder und $\mathbb{P}_0(S) = 1$ folgt.

Satz 2.26 (Teilfolgenkriterium für schwache Konvergenz)**.** *Es gilt $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$ genau dann, wenn es zu jeder Teilfolge $(\mathbb{P}_{n_k})_{k\in\mathbb{N}}$ von $(\mathbb{P}_n)_{n\in\mathbb{N}}$ eine weitere Teilfolge $(\mathbb{P}_{n_{k_m}})_{m\in\mathbb{N}}$ gibt mit $\mathbb{P}_{n_{k_m}} \xrightarrow[m\to\infty]{\mathcal{D}} \mathbb{P}_0$, wobei $\mathbb{P}_0$ nicht von der Wahl der Teilfolge abhängt.*

Beweis. Übung. □

Satz 2.27 (Schwache Konvergenz und Verteilungsfunktionen)**.** *Wir betrachten $S = \mathbb{R}^k$ mit der Metrik $d(\boldsymbol{x}, \boldsymbol{y}) = \big(\sum_{i=1}^{k}(x_i - y_i)^2\big)^{1/2}$; dieser Raum ist vollständig und separabel. Sei $(\mathbb{P}_n)_{n\in\mathbb{N}_0}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ mit Verteilungsfunktionen*

$$F_n(\boldsymbol{x}) \coloneqq \mathbb{P}_n\big(\{\boldsymbol{y} \in \mathbb{R}^k \colon \boldsymbol{y} \leq \boldsymbol{x}\}\big), \qquad \boldsymbol{x} \in \mathbb{R}^k,$$

wobei $\boldsymbol{y} \leq \boldsymbol{x}$ komponentenweise zu verstehen ist. Dann gilt

$$\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0 \quad \Longleftrightarrow \quad F_n(\boldsymbol{x}) \xrightarrow[n\to\infty]{} F_0(\boldsymbol{x}) \quad \text{für alle } \boldsymbol{x} \in C(F_0),$$

wobei $C(F_0) \coloneqq \{\boldsymbol{x} \in \mathbb{R}^k \colon F_0 \text{ ist stetig in } \boldsymbol{x}\}$.

Beweis. „⇒“: Nach Satz 2.23 (v) gilt

$$F_n(\boldsymbol{x}) = \mathbb{P}_n\big((-\boldsymbol{\infty}, \boldsymbol{x}]\big) \xrightarrow[n\to\infty]{} \mathbb{P}_0\big((-\boldsymbol{\infty}, \boldsymbol{x}]\big) = F_0(\boldsymbol{x}),$$

falls $\mathbb{P}_0(\partial(-\boldsymbol{\infty}, \boldsymbol{x}]) = 0$ gilt. Dies bedeutet aber gerade, dass F_0 stetig in $\boldsymbol{x}$ ist.

„$\Leftarrow$“: Wir wollen Satz 2.24 benutzen. Für $i = 1, \dots, k$ sei $D_i := \{c \in \mathbb{R} \colon \mathbb{P}_0(\{\boldsymbol{x} \in \mathbb{R}^k \colon x_i = c\}) > 0\}$. Die Hyperebenen $\{x_i = c\}$, $c \in \mathbb{R}$, sind paarweise disjunkt und abgeschlossen mit leerem Inneren, also ist D_i nach Bemerkung 2.25 höchstens abzählbar. Wir setzen

$$\mathcal{G} := \{(\boldsymbol{a}, \boldsymbol{b}] \subset \mathbb{R}^k \colon \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^k,\ a_i, b_i \notin D_i \text{ für } i = 1, \dots, k\} \cup \{\emptyset\}.$$

Dann gelten die Voraussetzungen von Satz 2.24:

(i) $\mathcal{G}$ ist durchschnittsstabil, denn der Schnitt zweier Quader aus $\mathcal{G}$ ist leer oder ein Quader, dessen Eckkoordinaten unter denen der beiden Quader vorkommen.

(ii) Da die Mengen $\mathbb{R} \setminus D_i$ dicht in $\mathbb{R}$ sind, gibt es zu jedem $\boldsymbol{x}$ in einer offenen Menge G einen Quader $(\boldsymbol{a}, \boldsymbol{b}] \in \mathcal{G}$ mit $\boldsymbol{x} \in (\boldsymbol{a}, \boldsymbol{b}] \subset G$ und Eckkoordinaten aus einer festen abzählbaren dichten Teilmenge von $\mathbb{R} \setminus D_i$; also ist G abzählbare Vereinigung von Mengen aus $\mathcal{G}$.

(iii) Für $(\boldsymbol{a}, \boldsymbol{b}] \in \mathcal{G}$ liegen alle Ecken $\boldsymbol{v}$ von $(\boldsymbol{a}, \boldsymbol{b}]$ in $C(F_0)$, da $\partial(-\boldsymbol{\infty}, \boldsymbol{v}] \subset \bigcup_{i=1}^k \{x_i = v_i\}$ eine $\mathbb{P}_0$-Nullmenge ist. Mit der Bezeichnung $\Delta_{\boldsymbol{a}}^{\boldsymbol{b}} F$ für die alternierende Summe der Werte von F an den Ecken von $(\boldsymbol{a}, \boldsymbol{b}]$ folgt

$$\mathbb{P}_n((\boldsymbol{a}, \boldsymbol{b}]) = \Delta_{\boldsymbol{a}}^{\boldsymbol{b}} F_n \xrightarrow[n\to\infty]{\text{Vor.}} \Delta_{\boldsymbol{a}}^{\boldsymbol{b}} F_0 = \mathbb{P}_0((\boldsymbol{a}, \boldsymbol{b}]).$$

Für $k = 2$ ist dies, siehe Abbildung 2.5,

$$\Delta_{\boldsymbol{a}}^{\boldsymbol{b}} F_n = F_n(b_1, b_2) - F_n(b_1, a_2) - F_n(a_1, b_2) + F_n(a_1, a_2),$$

und für allgemeines k ist $\Delta_{\boldsymbol{a}}^{\boldsymbol{b}} F_n$ das analoge Objekt.

Nach Satz 2.24 folgt $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$. □

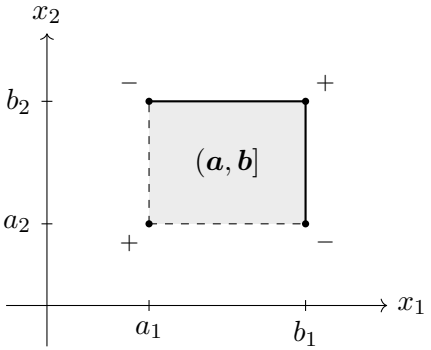

Abbildung 2.5.: Der Quader $(\boldsymbol{a}, \boldsymbol{b}]$ für $k = 2$. Seine Wahrscheinlichkeit ergibt sich aus den Werten der Verteilungsfunktion an den vier Ecken mit den eingezeichneten Vorzeichen.

Bemerkung. (i) Wegen Satz 2.27 spricht man auch von *Verteilungskonvergenz*.

(ii) Die halboffenen Intervalle bilden eine konvergenzbestimmende Klasse in dem Sinne, dass aus $\mathbb{P}_n((a, b]) \to \mathbb{P}_0((a, b])$ für alle $\mathbb{P}_0$-randlosen Intervalle $(a, b]$ bereits $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$ folgt. Die Einschränkung auf $\mathbb{P}_0$-randlose Intervalle ist nötig: Für $\mathbb{P}_n = \delta_{1/n}$ und $\mathbb{P}_0 = \delta_0$ gilt $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$, aber $\mathbb{P}_n((0, 1]) = 1 \not\to 0 = \mathbb{P}_0((0, 1])$.

(iii) Die Verteilungskonvergenz kommt beim zentralen Grenzwertsatz vor.

Satz 2.28 (Schwache Konvergenz gegen deterministische Größen)**.** *Seien $(X_n)_{n\in\mathbb{N}}$ Zufallsvariablen auf $(\Omega,\mathcal{F},\mathbb{P})$ mit Werten in $(S,\mathcal{B}(S))$ und sei $a\in S$. Dann gilt*

$$X_n \xrightarrow[n\to\infty]{} a \text{ in Wahrscheinlichkeit} \quad\Longleftrightarrow\quad \mathbb{P}^{X_n} \xrightarrow[n\to\infty]{\mathcal{D}} \delta_a,$$

wobei δ_a das Punktmaß in a bezeichnet.

Beweis. Übung. □

Satz 2.29 (Schwache Konvergenz und stochastische Äquivalenz)**.** *Gilt $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$ und ist $(X_n)_{n\in\mathbb{N}}$ stochastisch äquivalent zu $(Y_n)_{n\in\mathbb{N}}$ im Sinne von Definition 2.13, so gilt $Y_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$.*

Beweis. Übung. □

Satz 2.30 (Konvergenz in Wahrscheinlichkeit impliziert schwache Konvergenz)**.** *Seien $(X_n)_{n\in\mathbb{N}_0}$ Zufallsvariablen mit Werten in $(S,\mathcal{B}(S))$. Dann gilt*

$$X_n \xrightarrow[n\to\infty]{} X_0 \text{ in Wahrscheinlichkeit} \quad\Longrightarrow\quad X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0.$$

Beweis. Die Konvergenz $X_n \to X_0$ in Wahrscheinlichkeit bedeutet, dass $(X_n)_{n\in\mathbb{N}}$ stochastisch äquivalent zur konstanten Folge $(X_0, X_0, \dots)$ ist. Diese konstante Folge konvergiert schwach gegen X_0, da alle ihre Glieder die Verteilung $\mathbb{P}^{X_0}$ haben; also folgt die Behauptung mit Satz 2.29. □

Satz 2.31 (Satz von Slutsky)**.** *Seien $(\boldsymbol{X}_n)_{n\in\mathbb{N}_0}$ und $(\boldsymbol{Y}_n)_{n\in\mathbb{N}}$ Zufallsvariablen auf $(\Omega,\mathcal{F},\mathbb{P})$ mit Werten in $\mathbb{R}^k$, und es gelte*

$$\boldsymbol{X}_n \xrightarrow[n\to\infty]{\mathcal{D}} \boldsymbol{X}_0 \qquad \text{und} \qquad \boldsymbol{Y}_n \xrightarrow[n\to\infty]{} \boldsymbol{a}\in\mathbb{R}^k \text{ in Wahrscheinlichkeit.}$$

Dann gilt

(i) $\boldsymbol{X}_n + \boldsymbol{Y}_n \xrightarrow[n\to\infty]{\mathcal{D}} \boldsymbol{X}_0 + \boldsymbol{a}$,

(ii) $\boldsymbol{Y}_n^\top \boldsymbol{X}_n \xrightarrow[n\to\infty]{\mathcal{D}} \boldsymbol{a}^\top \boldsymbol{X}_0$.

Beispiel. Seien $X_1, X_2, \dots$ unabhängig und identisch verteilt mit $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ und $\operatorname{Var}[X_1] = \sigma^2 \in (0,\infty)$, und sei $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Dann gilt nach dem zentralen Grenzwertsatz, Satz 1.15,

$$\mathbb{P}(\sqrt{n}(\overline{X}_n - \mu) \le x) \xrightarrow[n\to\infty]{} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{y^2}{2\sigma^2}}\, \mathrm{d}y \qquad \text{für alle } x\in\mathbb{R},$$

also nach Satz 2.27

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow[n\to\infty]{\mathcal{D}} Z \qquad \text{mit } Z\sim\mathcal{N}(0,\sigma^2).$$

Wir betrachten nun den Varianzschätzer

$$S_{n-1}^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

Ist zusätzlich $\mathbb{E}[X_1^4] < \infty$, so gilt $S_{n-1}^2 \to \sigma^2$ in Wahrscheinlichkeit. Wegen

$$S_{n-1}^2 = \frac{n}{n-1}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\Big)$$

folgt dies nämlich aus dem schwachen Gesetz der großen Zahlen, Satz 0.14, angewandt auf $(X_i^2)_{i\in\mathbb{N}}$ und $(X_i)_{i\in\mathbb{N}}$, zusammen mit Korollar 2.34. Mit dem starken Gesetz der großen Zahlen in der Fassung für integrierbare Zufallsvariablen gilt dies sogar ohne die Annahme $\mathbb{E}[X_1^4] < \infty$. Mit Korollar 2.34 für $h(x) = 1/\sqrt{x}$, stetig in $\sigma^2 > 0$, folgt $1/S_{n-1} \to 1/\sigma$ in Wahrscheinlichkeit, wobei wir $1/S_{n-1} \coloneqq 0$ auf $\{S_{n-1} = 0\}$ setzen; wegen $\mathbb{P}(S_{n-1} = 0) \to 0$ ändert dies nichts an der Konvergenz. Mit dem Satz von Slutsky, Satz 2.31 (ii) für $k = 1$, folgt daher

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_{n-1}} \xrightarrow[n\to\infty]{\mathcal{D}} \frac{1}{\sigma} Z \sim \mathcal{N}(0,1).$$

Beweis von Satz 2.31. Wir beweisen nur (i), und zwar nur für $k = 1$. Zunächst sind $(X_n + a)_{n\in\mathbb{N}}$ und $(X_n + Y_n)_{n\in\mathbb{N}}$ stochastisch äquivalent, denn $|(X_n + Y_n) - (X_n + a)| = |Y_n - a| \to 0$ in Wahrscheinlichkeit. Nach Satz 2.29 genügt es also zu zeigen, dass

$$X_n + a \xrightarrow[n\to\infty]{\mathcal{D}} X_0 + a.$$

Dafür wollen wir die Konvergenz der Verteilungsfunktionen prüfen, siehe Satz 2.27. Zu zeigen ist also

$$F^{X_n}(x - a) = F^{X_n + a}(x) \xrightarrow[n\to\infty]{} F^{X_0 + a}(x) = F^{X_0}(x - a)$$

für alle $x \in C(F^{X_0+a})$. Nun ist F^{X_0+a} genau dann stetig in x, wenn $F^{X_0}(\cdot - a)$ stetig in x ist, also wenn F^{X_0} stetig in $x - a$ ist. Also gilt $C(F^{X_0+a}) = C\big(F^{X_0}(\cdot - a)\big)$, und die gewünschte Konvergenz folgt aus der Voraussetzung $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$ und Satz 2.27. □

Beispiel 2.32 (Schwache Konvergenz und endlich-dimensionale Verteilungen)**.** Der Raum $S = C[0,1] = \{f\colon [0,1] \to \mathbb{R}\colon f \text{ stetig}\}$ mit der Metrik

$$d(f,g) \coloneqq \sup_{t\in[0,1]} |f(t) - g(t)|$$

ist ein polnischer Raum. Zufallsvariablen $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{B}(S))$ sind zufällige Funktionen auf $[0,1]$, zum Beispiel Aktienkurse.

Frage: Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Folge von Zufallsvariablen mit Werten in $C[0,1]$. Folgt $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$ bereits aus

$$\big(X_n(t_1), \dots, X_n(t_k)\big)^\top \xrightarrow[n\to\infty]{\mathcal{D}} \big(X_0(t_1), \dots, X_0(t_k)\big)^\top \tag{2.5}$$

für alle $(t_1, \dots, t_k) \in [0,1]^k$ und alle $k \in \mathbb{N}$? Hier sind die Vektoren auf der linken Seite $\mathbb{R}^k$-wertige Zufallsvariablen. Die Aussage (2.5) heißt *schwache Konvergenz der endlich-dimensionalen Randverteilungen* (englisch „fidi-convergence“).

Antwort: Die Implikation „(2.5) $\Rightarrow X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$“ ist im Allgemeinen falsch!

Als Beispiel setzen wir $X_0 \equiv 0$, also $X_0(\omega, t) = 0$ für alle $\omega \in \Omega$ und $t \in [0,1]$, sowie

$$X_n(\omega, t) \coloneqq x_n(t) \coloneqq \begin{cases} nt, & 0 \le t \le \frac{1}{n}, \\ 2 - nt, & \frac{1}{n} < t \le \frac{2}{n}, \\ 0, & t > \frac{2}{n}, \end{cases}$$

für $n \geq 2$, siehe Abbildung 2.6. Dann ist $x_n \in C[0,1]$ unabhängig von ω, also gilt $\mathbb{P}^{X_n} = \delta_{x_n}$. Wir betrachten die Menge $A \coloneqq \{f \in C[0,1] \colon d(f,0) \leq 1/2\}$. Wegen $\partial A \subset \{f \colon d(f,0) = 1/2\}$ und $d(x_n, 0) = 1$ gilt

$$\mathbb{P}^{X_0}(\partial A) = 0, \qquad \mathbb{P}^{X_0}(A) = 1, \qquad \mathbb{P}^{X_n}(A) = 0 \quad \text{für alle } n \geq 2.$$

Nach Satz 2.23 (v) konvergiert X_n also nicht schwach gegen X_0. Sind aber $t_1, \dots, t_k \in [0,1]$ gegeben und ist n so groß, dass $2/n < t_i$ für alle i mit $t_i > 0$ gilt, so ist

$$\big(X_n(t_1), \dots, X_n(t_k)\big)^\top = \mathbf{0} = \big(X_0(t_1), \dots, X_0(t_k)\big)^\top,$$

also gilt (2.5).

Ziel: Wir suchen eine Zusatzbedingung, sodass aus (2.5) und dieser Bedingung $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$ folgt.

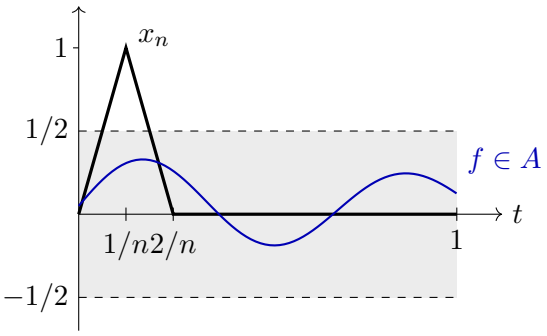

Abbildung 2.6.: Die Zeltfunktion x_n aus Beispiel 2.32 und die Menge A der Funktionen, deren Graph im grauen Streifen liegt. Für jedes feste $t > 0$ gilt $x_n(t) = 0$ für große n, aber $x_n \notin A$.

Satz 2.33 (Continuous Mapping Theorem)**.** *Seien $(X_n)_{n\in\mathbb{N}_0}$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(S, \mathcal{B}(S))$, und sei $(S', \mathcal{B}(S'))$ ein weiterer polnischer Raum mit Metrik d'. Sei $h \colon S \to S'$ messbar, und sei D_h die Menge der Unstetigkeitsstellen von h; es gilt $D_h \in \mathcal{B}(S)$. Dann gilt*

$$X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0 \quad \textit{und} \quad \mathbb{P}(X_0 \in D_h) = 0 \qquad \Longrightarrow \qquad h(X_n) \xrightarrow[n\to\infty]{\mathcal{D}} h(X_0).$$

Beweis. Zunächst gilt

$$D_h = \bigcup_{n\in\mathbb{N}} \bigcap_{k\in\mathbb{N}} \underbrace{\Big\{x \in S \colon \exists\, y, z \in S \text{ mit } d(x,y) < \tfrac{1}{k},\ d(x,z) < \tfrac{1}{k} \text{ und } d'(h(y),h(z)) \geq \tfrac{1}{n}\Big\}}_{=:A_{n,k}},$$

und die Mengen $A_{n,k}$ sind offen, also gilt $D_h \in \mathcal{B}(S)$.

Nach Satz 2.23 genügt es zu zeigen, dass

$$\limsup_{n\to\infty} \mathbb{P}^{h(X_n)}(F) \leq \mathbb{P}^{h(X_0)}(F) \qquad \text{für alle abgeschlossenen } F \subset S'.$$

Nun gilt nach Satz 2.23 (iii) und der Voraussetzung

$$\limsup_{n\to\infty} \mathbb{P}^{h(X_n)}(F) = \limsup_{n\to\infty} \mathbb{P}^{X_n}\big(h^{-1}(F)\big) \leq \limsup_{n\to\infty} \mathbb{P}^{X_n}\big(\overline{h^{-1}(F)}\big) \leq \mathbb{P}^{X_0}\big(\overline{h^{-1}(F)}\big).$$

Es bleibt zu zeigen, dass

$$\mathbb{P}^{h(X_0)}(F) = \mathbb{P}^{X_0}\big(h^{-1}(F)\big) = \mathbb{P}^{X_0}\big(\overline{h^{-1}(F)}\big).$$

Dazu genügt es, $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$ zu zeigen, denn es gilt $\mathbb{P}^{X_0}(D_h) = 0$. Sei also $(x_n)_{n\in\mathbb{N}}$ eine Folge in $h^{-1}(F)$ mit Grenzwert $x_0 \in \overline{h^{-1}(F)}$. Dann ist h entweder stetig in x_0, also gilt $h(x_n) \to h(x_0) \in F$, da F abgeschlossen ist, oder es gilt $x_0 \in D_h$. □

Korollar 2.34 (Konvergenz in Wahrscheinlichkeit unter stetigen Abbildungen)**.** *Seien (S', d') und h wie in Satz 2.33, und sei h stetig in $a \in S$. Aus $X_n \to a$ in Wahrscheinlichkeit folgt dann $h(X_n) \to h(a)$ in Wahrscheinlichkeit.*

Beweis. Nach Satz 2.28 gilt $X_n \to a$ in Wahrscheinlichkeit genau dann, wenn $X_n \xrightarrow[n\to\infty]{\mathcal{D}} \delta_a$. Da $a \notin D_h$, also $\delta_a(D_h) = 0$ gilt, folgt aus Satz 2.33 $h(X_n) \xrightarrow[n\to\infty]{\mathcal{D}} \delta_{h(a)}$, und das ist nach Satz 2.28 äquivalent zu $h(X_n) \to h(a)$ in Wahrscheinlichkeit. □

Die Straffheit aus Definition 2.21 ist eng mit der Existenz schwach konvergenter Teilfolgen verknüpft.

Satz 2.35 (Satz von Prohorov [Pro56])**.** *Sei S polnisch, und seien $(\mathbb{P}_n)_{n\in\mathbb{N}}$ Wahrscheinlichkeitsmaße auf $(S, \mathcal{B}(S))$. Dann sind die folgenden Aussagen äquivalent:*

(i) Die Folge $(\mathbb{P}_n)_{n\in\mathbb{N}}$ ist straff.

(ii) Die Folge $(\mathbb{P}_n)_{n\in\mathbb{N}}$ ist relativ kompakt*, das heißt, zu jeder Teilfolge $(n_k)_{k\in\mathbb{N}}$ gibt es eine weitere Teilfolge $(n_{k_m})_{m\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß Q auf $(S, \mathcal{B}(S))$ mit*

$$\mathbb{P}_{n_{k_m}} \xrightarrow[m\to\infty]{\mathcal{D}} Q.$$

Das Grenzmaß Q hängt im Allgemeinen von der Teilfolge $(n_k)_{k\in\mathbb{N}}$ ab. Für den Beweis benötigen wir das folgende Resultat.

Satz 2.36 (Auswahlsatz von Helly)**.** *Sei $(F_n)_{n\in\mathbb{N}}$ eine Folge von Verteilungsfunktionen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann existieren eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ und eine maßdefinierende Funktion F mit*

$$F_{n_k}(x) \xrightarrow[k\to\infty]{} F(x) \qquad \text{für alle } x \in C(F).$$

Dabei heißt eine Funktion $F\colon \mathbb{R} \to \mathbb{R}$ *maßdefinierend*, falls sie (i) rechtsseitig stetig und (ii) monoton steigend ist, siehe Abbildung 2.7. Zu jeder maßdefinierenden Funktion F gehört ein Maß μ_F auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $\mu_F((a, b]) = F(b) - F(a)$ für alle $a < b$.

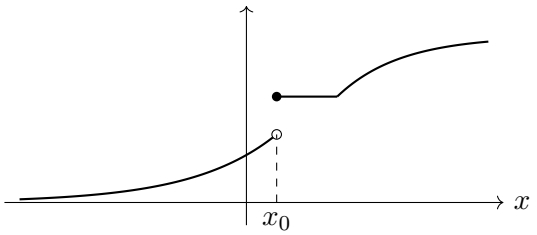

Abbildung 2.7.: Eine maßdefinierende Funktion mit einem Sprung in x_0; dort ist sie rechtsseitig stetig, und das zugehörige Maß μ_F hat ein Atom in x_0.

Beweis von Satz 2.36. Für jedes $q \in \mathbb{Q}$ besitzt die Folge $(F_n(q))_{n\in\mathbb{N}} \subset [0,1]$ nach dem Satz von Bolzano–Weierstraß eine konvergente Teilfolge. Da $\mathbb{Q}$ abzählbar ist, erhalten wir mit dem Diagonalfolgenargument sogar eine Teilfolge $(n_k)_{k\in\mathbb{N}}$, für die $(F_{n_k}(q))_{k\in\mathbb{N}}$ für alle $q \in \mathbb{Q}$ konvergiert. Wir setzen

$$F_0 \colon \mathbb{Q} \to [0,1], \qquad q \mapsto F_0(q) \coloneqq \lim_{k\to\infty} F_{n_k}(q).$$

Als Grenzwert nichtfallender Funktionen ist F_0 nichtfallend. Ferner setzen wir

$$F(x) \coloneqq \inf\{F_0(q) \colon x < q,\ q \in \mathbb{Q}\}, \qquad x \in \mathbb{R}.$$

Dann ist F monoton steigend. Außerdem ist F rechtsseitig stetig, denn für $x_k \downarrow x$ gilt

$$\lim_{k\to\infty} F(x_k) = \inf_{k\in\mathbb{N}} F(x_k) = \inf_{k\in\mathbb{N}} \inf_{q>x_k} F_0(q) = \inf_{q>x} F_0(q) = F(x).$$

Also ist F eine maßdefinierende Funktion. Es bleibt zu zeigen, dass $F_{n_k}(x) \to F(x)$ für alle $x \in C(F)$ gilt.

Sei also $x \in C(F)$ und $\varepsilon > 0$. Dann gibt es $r_1, r_2 \in \mathbb{Q}$ mit $r_1 < x < r_2$, sodass

$$F(x) - \varepsilon < F_0(r_1) \le F_0(r_2) < F(x) + \varepsilon.$$

In der Tat gibt es wegen der Stetigkeit von F in x ein $y < x$ mit $F(y) > F(x) - \varepsilon$; für jedes $r_1 \in \mathbb{Q} \cap (y, x)$ gilt dann $F_0(r_1) \ge F(y)$ nach Definition von F. Nach Definition des Infimums gibt es ein $r_2 \in \mathbb{Q}$ mit $r_2 > x$ und $F_0(r_2) < F(x) + \varepsilon$. Außerdem gilt für alle $n \in \mathbb{N}$ die Ungleichung $F_n(r_1) \le F_n(x) \le F_n(r_2)$. Also gilt

$$\begin{aligned} F(x) - \varepsilon \le \lim_{k\to\infty} F_{n_k}(r_1) \le \liminf_{k\to\infty} F_{n_k}(x) \qquad &\text{und} \\ F(x) + \varepsilon \ge \lim_{k\to\infty} F_{n_k}(r_2) \ge \limsup_{k\to\infty} F_{n_k}(x). & \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt $\lim_{k\to\infty} F_{n_k}(x) = F(x)$. □

Beweis von Satz 2.35 (nur für $S = \mathbb{R}$). (ii)⇒(i): Angenommen, $(\mathbb{P}_n)_{n\in\mathbb{N}}$ ist nicht straff. Dann gibt es ein $\varepsilon > 0$, sodass es für alle $k \in \mathbb{N}$ ein n_k gibt mit $\mathbb{P}_{n_k}([-k,k]) < 1 - \varepsilon$. Da jedes einzelne Wahrscheinlichkeitsmaß auf $\mathbb{R}$ straff ist, können wir $(n_k)_{k\in\mathbb{N}}$ streng monoton wachsend wählen. Wegen (ii) gibt es eine Teilfolge $(\mathbb{P}_{n_{k_m}})_{m\in\mathbb{N}}$ mit $\mathbb{P}_{n_{k_m}} \xrightarrow[m\to\infty]{\mathcal{D}} Q$, wobei Q ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist. Mit Satz 2.23 (iv) gilt also für alle $x > 0$

$$Q((-x,x)) \le \liminf_{m\to\infty} \mathbb{P}_{n_{k_m}}((-x,x)) \le \liminf_{m\to\infty} \mathbb{P}_{n_{k_m}}([-k_m,k_m]) \le 1 - \varepsilon,$$

wobei wir $k_m \ge x$ für große m benutzt haben. Andererseits gilt $\lim_{x\to\infty} Q((-x,x)) = 1$, ein Widerspruch.

(i)⇒(ii): Sei $(\mathbb{P}_{n_k})_{k\in\mathbb{N}}$ eine beliebige Teilfolge von $(\mathbb{P}_n)_{n\in\mathbb{N}}$ mit zugehörigen Verteilungsfunktionen F_{n_k}. Aufgrund des Auswahlsatzes von Helly, Satz 2.36, gibt es eine Teilfolge $(F_{n_{k_m}})_{m\in\mathbb{N}}$ und eine maßdefinierende Funktion F, sodass

$$F_{n_{k_m}}(x) \xrightarrow[m\to\infty]{} F(x) \qquad \text{für alle } x \in C(F). \tag{2.6}$$

Für F gilt

$$0 \le \lim_{x\to-\infty} F(x) \le \lim_{x\to\infty} F(x) \le 1,$$

und daher

$$\mu_F(\mathbb{R}) = \lim_{x\to\infty} \mu_F((-x,x]) = \lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x) \le 1.$$

Angenommen, es gilt $\mu_F(\mathbb{R}) < 1$. Wegen (i) gibt es zu jedem $\varepsilon > 0$ Punkte $a, b \in C(F)$ mit $a < b$ und

$$\mathbb{P}_n((a,b]) \geq 1 - \varepsilon \qquad \text{für alle } n \in \mathbb{N};$$

dabei verwenden wir, dass das Komplement von $C(F)$ höchstens abzählbar ist. Damit gilt

$$\mu_F((a,b]) = F(b) - F(a) = \lim_{m\to\infty} F_{n_{k_m}}(b) - \lim_{m\to\infty} F_{n_{k_m}}(a) = \lim_{m\to\infty} \mathbb{P}_{n_{k_m}}((a,b]) \geq 1 - \varepsilon.$$

Für $\varepsilon < 1 - \mu_F(\mathbb{R})$ ist dies ein Widerspruch zur Annahme $\mu_F(\mathbb{R}) < 1$. Also ist μ_F ein Wahrscheinlichkeitsmaß; insbesondere gilt $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to\infty} F(x) = 1$, das heißt, F ist die Verteilungsfunktion von μ_F. Mit (2.6) und Satz 2.27 folgt $\mathbb{P}_{n_{k_m}} \xrightarrow[m\to\infty]{\mathcal{D}} \mu_F$. □

Bemerkung. (i) Der Beweis für $S = \mathbb{R}^k$ ist sehr ähnlich.

(ii) Lediglich die Richtung (ii)⇒(i) benötigt, dass S polnisch ist; die Richtung (i)⇒(ii) gilt in jedem metrischen Raum, siehe [Bil99, Theoreme 5.1 und 5.2].

(iii) Als Gegenbeispiel ohne Straffheit betrachten wir die Folge $(\mathbb{P}_n)_{n\in\mathbb{N}}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $\mathbb{P}_n = \mathrm{Leb}_{[n,n+1]}$, der Gleichverteilung auf $[n, n+1]$. Dann gilt für alle kompakten $K \subset \mathbb{R}$ die Konvergenz $\mathbb{P}_n(K) \to 0$, und $(\mathbb{P}_n)_{n\in\mathbb{N}}$ ist nicht relativ kompakt (Masseverlust). Die Folge $(\mathbb{P}_n)_{n\in\mathbb{N}}$ konvergiert lediglich vage gegen das Sub-Wahrscheinlichkeitsmaß 0, das heißt, $\int f \,\mathrm{d}\mathbb{P}_n \to 0$ für alle stetigen f mit kompaktem Träger.

Beispiel 2.37 (Anwendung auf endlich-dimensionale Verteilungen)**.** Sei $(\mathbb{P}_n)_{n\in\mathbb{N}_0}$ eine Familie von Wahrscheinlichkeitsmaßen auf $C[0,1]$ mit

$$\mathbb{P}_n^{\pi_{t_1,\dots,t_k}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0^{\pi_{t_1,\dots,t_k}} \qquad \text{für alle } t_1,\dots,t_k \in [0,1] \text{ und alle } k \in \mathbb{N}. \tag{2.7}$$

Hier bezeichnet $\pi_{t_1,\dots,t_k} \colon C[0,1] \to \mathbb{R}^k$, $f \mapsto (f(t_1),\dots,f(t_k))^\top$ die Projektion auf die Zeitpunkte $t_1,\dots,t_k$. In Beispiel 2.32 haben wir gesehen, dass aus (2.7) nicht zwingend $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0$ folgt.

Sei nun $(\mathbb{P}_n)_{n\in\mathbb{N}}$ zusätzlich straff. Dann gilt nach Satz 2.35, dass es zu jeder Teilfolge $(n_k)_{k\in\mathbb{N}}$ eine weitere Teilfolge $(n_{k_m})_{m\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß Q gibt mit

$$\mathbb{P}_{n_{k_m}} \xrightarrow[m\to\infty]{\mathcal{D}} Q. \tag{2.8}$$

Hieraus folgt $\mathbb{P}_{n_{k_m}}^{\pi_{t_1,\dots,t_k}} \xrightarrow[m\to\infty]{\mathcal{D}} Q^{\pi_{t_1,\dots,t_k}}$ nach dem Continuous Mapping Theorem, Satz 2.33, da $\pi_{t_1,\dots,t_k}$ stetig ist. Also gilt mit dem Eindeutigkeitssatz des schwachen Grenzwerts, Satz 2.20, und (2.7)

$$Q^{\pi_{t_1,\dots,t_k}} = \mathbb{P}_0^{\pi_{t_1,\dots,t_k}} \qquad \text{für alle } t_1,\dots,t_k \in [0,1] \text{ und alle } k \in \mathbb{N}.$$

Die Zylindermengen $\pi_{t_1,\dots,t_k}^{-1}(B)$, $B \in \mathcal{B}(\mathbb{R}^k)$, in $C[0,1]$ bilden eine maßbestimmende Klasse (ohne Beweis), also gilt $Q = \mathbb{P}_0$. Also hängt der Grenzwert in (2.8) nicht von der konvergenten Teilfolge ab. Dann gilt aber wegen des Teilfolgenkriteriums, Satz 2.26,

$$\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}_0.$$

Wir fassen zusammen.

Satz 2.38 (Schwache Konvergenz in $C[0,1]$)**.** *Seien $(X_n)_{n\in\mathbb{N}_0}$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $C[0,1]$. Es gelte*

(i) $(X_n(t_1),\dots,X_n(t_k)) \xrightarrow[n\to\infty]{\mathcal{D}} (X_0(t_1),\dots,X_0(t_k))$ *für alle* $t_1,\dots,t_k \in [0,1]$ *und alle* $k \in \mathbb{N}$ *(Konvergenz der endlich-dimensionalen Randverteilungen),*

(ii) $(\mathbb{P}^{X_n})_{n\in\mathbb{N}}$ *ist straff.*

Dann folgt $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$.

Bemerkung. Umgekehrt folgen aus $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X_0$ auch (i) und (ii).

Satz 2.39 (Kriterium für Straffheit)**.** *Eine Folge* $(X_n)_{n\in\mathbb{N}}$ *von Zufallsvariablen mit Werten in* $C[0,1]$ *ist straff, falls gilt:*

(i) Für alle $\eta > 0$ *gibt es ein* $a > 0$ *mit* $\mathbb{P}(|X_n(0)| > a) < \eta$ *für alle* $n \in \mathbb{N}$*, das heißt,* $(X_n(0))_{n\in\mathbb{N}}$ *ist straff.*

(ii) Für alle $\varepsilon, \eta > 0$ *gibt es ein* $\delta \in (0,1)$ *und ein* $n_0 \in \mathbb{N}$*, sodass für alle* $n \geq n_0$ *und alle* $t \in [0,1]$ *gilt*

$$\frac{1}{\delta}\,\mathbb{P}\Big(\sup_{s\in[t,t+\delta]\cap[0,1]} |X_n(s) - X_n(t)| \geq \varepsilon\Big) \leq \eta.$$

Beweis. Ohne Beweis. □

2.2. Beweise des Stetigkeitssatzes und des Satzes von Cramér–Wold

Mit dem Satz von Prohorov können wir nun die beiden in Kapitel 1 offen gebliebenen Beweise nachholen. Die Idee für den Stetigkeitssatz ist, aus der Stetigkeit von φ_0 in $\mathbf{0}$ die Straffheit der Verteilungen zu gewinnen. Nach dem Satz von Prohorov besitzt dann jede Teilfolge eine schwach konvergente Teilfolge, und der Eindeutigkeitssatz identifiziert den Grenzwert. Die Straffheit folgt aus der folgenden Abschätzung.

Lemma 2.40 (Abschneide-Ungleichung)**.** *Sei* X *eine reellwertige Zufallsvariable mit charakteristischer Funktion* φ*. Dann gilt für alle* $u > 0$

$$\mathbb{P}(|X| \geq 2/u) \leq \frac{1}{u}\int_{-u}^{u} (1 - \varphi(t))\,\mathrm{d}t.$$

Beweis. Für $x \neq 0$ gilt

$$\frac{1}{u}\int_{-u}^{u} (1 - \mathrm{e}^{\mathrm{i}tx})\,\mathrm{d}t = 2 - \frac{1}{u}\int_{-u}^{u} \cos(tx)\,\mathrm{d}t = 2\Big(1 - \frac{\sin(ux)}{ux}\Big),$$

da der Imaginärteil eine ungerade Funktion von t ist; für $x = 0$ sind beide Seiten gleich 0, wenn wir $\sin(0)/0 \coloneqq 1$ setzen. Da der Integrand beschränkt ist, dürfen wir mit dem Satz von Fubini Integral und Erwartungswert vertauschen. Wegen $1 - \sin(y)/y \geq 0$ und $\sin(y)/y \leq 1/|y|$ für alle $y \neq 0$ folgt

$$\frac{1}{u}\int_{-u}^{u} (1 - \varphi(t))\,\mathrm{d}t = 2\,\mathbb{E}\Big[1 - \frac{\sin(uX)}{uX}\Big] \geq 2\,\mathbb{E}\Big[\Big(1 - \frac{1}{|uX|}\Big)\mathbb{1}\{|uX| \geq 2\}\Big] \geq \mathbb{P}(|uX| \geq 2).$$ □

Beweis des Stetigkeitssatzes, Satz 1.6. Wir schreiben $\mathbb{P}_n \coloneqq \mathbb{P}^{\boldsymbol{X}_n}$ und $X_n^{(j)} \coloneqq \boldsymbol{e}_j^\top \boldsymbol{X}_n$ für die j-te Koordinate, wobei $\boldsymbol{e}_j$ der j-te Einheitsvektor ist.

Schritt 1: Straffheit. Die charakteristische Funktion von $X_n^{(j)}$ ist $t \mapsto \varphi_n(t\boldsymbol{e}_j)$. Wegen $\varphi_n(\mathbf{0}) = 1$ gilt $\varphi_0(\mathbf{0}) = 1$. Sei $\varepsilon > 0$. Da φ_0 in $\mathbf{0}$ stetig ist, gibt es ein $u > 0$ mit $|1 - \varphi_0(t\boldsymbol{e}_j)| <$

$\varepsilon/2$ für alle $|t| \le u$ und alle $j = 1, \dots, k$. Wegen $|1 - \varphi_n| \le 2$ liefert der Satz von der dominierten Konvergenz

$$\frac{1}{u} \int_{-u}^{u} (1 - \varphi_n(t\boldsymbol{e}_j)) \, \mathrm{d}t \xrightarrow[n\to\infty]{} \frac{1}{u} \int_{-u}^{u} (1 - \varphi_0(t\boldsymbol{e}_j)) \, \mathrm{d}t,$$

und der Betrag der rechten Seite ist kleiner als ε. Mit Lemma 2.40 gibt es daher ein $n_0 \in \mathbb{N}$ mit $\mathbb{P}(|X_n^{(j)}| \ge 2/u) < \varepsilon$ für alle $n \ge n_0$ und alle j, also

$$\mathbb{P}_n(\mathbb{R}^k \setminus [-2/u, 2/u]^k) \le \sum_{j=1}^{k} \mathbb{P}(|X_n^{(j)}| \ge 2/u) < k\varepsilon \qquad \text{für alle } n \ge n_0.$$

Jedes einzelne der endlich vielen Maße $\mathbb{P}_1, \dots, \mathbb{P}_{n_0-1}$ ist straff, da die Würfel $[-a, a]^k$ für $a \to \infty$ gegen $\mathbb{R}^k$ aufsteigen. Durch Vergrößern des Würfels erhalten wir eine kompakte Menge K mit $\mathbb{P}_n(K) > 1 - k\varepsilon$ für alle $n \in \mathbb{N}$. Da $\varepsilon > 0$ beliebig war, ist $(\mathbb{P}_n)_{n\in\mathbb{N}}$ straff im Sinne von Definition 2.21.

Schritt 2: Identifikation des Grenzwerts. Da $\mathbb{R}^k$ polnisch ist, gibt es nach dem Satz von Prohorov, Satz 2.35, zu jeder Teilfolge $(n_l)_{l\in\mathbb{N}}$ eine weitere Teilfolge $(n_{l_m})_{m\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß Q mit $\mathbb{P}_{n_{l_m}} \xrightarrow[n\to\infty]{\mathcal{D}} Q$ für $m \to \infty$. Da $\boldsymbol{x} \mapsto \cos(\boldsymbol{t}^\top \boldsymbol{x})$ und $\boldsymbol{x} \mapsto \sin(\boldsymbol{t}^\top \boldsymbol{x})$ stetig und beschränkt sind, folgt

$$\varphi_{n_{l_m}}(\boldsymbol{t}) \xrightarrow[m\to\infty]{} \int_{\mathbb{R}^k} \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{x}} \, Q(\mathrm{d}\boldsymbol{x}) \qquad \text{für alle } \boldsymbol{t} \in \mathbb{R}^k.$$

Also ist φ_0 die charakteristische Funktion von Q. Nach dem Eindeutigkeitssatz, Satz 1.8, ist Q durch φ_0 eindeutig bestimmt und hängt daher nicht von der Wahl der Teilfolgen ab; wir schreiben $Q_0 \coloneqq Q$. Mit dem Teilfolgenkriterium, Satz 2.26, folgt $\mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} Q_0$.

Schritt 3: Konvergenz der Verteilungsfunktionen. Sei F_0 die Verteilungsfunktion von Q_0, und sei $\boldsymbol{x} \in \mathbb{R}^k$ mit $Q_0(\partial(-\boldsymbol{\infty}, \boldsymbol{x}]) = 0$. Für $\boldsymbol{y} \in \mathbb{R}^k$ mit $\max_i |y_i - x_i| \le h$ gilt

$$|F_0(\boldsymbol{y}) - F_0(\boldsymbol{x})| \le Q_0\big((-\boldsymbol{\infty}, \boldsymbol{x} + h\mathbf{1}] \setminus (-\boldsymbol{\infty}, \boldsymbol{x} - h\mathbf{1}]\big),$$

wobei $\mathbf{1} = (1, \dots, 1)^\top$. Für $h \downarrow 0$ fallen die Mengen auf der rechten Seite gegen $(-\boldsymbol{\infty}, \boldsymbol{x}] \setminus \{\boldsymbol{y} \colon y_i < x_i \text{ für alle } i\} = \partial(-\boldsymbol{\infty}, \boldsymbol{x}]$, sodass die rechte Seite gegen 0 konvergiert. Also ist F_0 stetig in $\boldsymbol{x}$, und Satz 2.27 liefert (1.11). □

Bemerkung. Der Beweis zeigt sogar die schwache Konvergenz $\mathbb{P}^{\boldsymbol{X}_n} \xrightarrow[n\to\infty]{\mathcal{D}} Q_0$.

Beweis des Satzes von Cramér–Wold, Satz 1.14. „⇒“: Die Abbildung $\boldsymbol{x} \mapsto \boldsymbol{c}^\top \boldsymbol{x}$ ist stetig, also folgt die Behauptung aus dem Continuous Mapping Theorem, Satz 2.33.

„⇐“: Für $\boldsymbol{c} \in \mathbb{R}^k$ gilt

$$\varphi^{\boldsymbol{X}_n}(\boldsymbol{c}) = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\boldsymbol{c}^\top \boldsymbol{X}_n}\big] = \varphi^{\boldsymbol{c}^\top \boldsymbol{X}_n}(1) \xrightarrow[n\to\infty]{} \varphi^{\boldsymbol{c}^\top \boldsymbol{X}_0}(1) = \varphi^{\boldsymbol{X}_0}(\boldsymbol{c}),$$

da cos und sin stetig und beschränkt sind. Die Grenzfunktion $\varphi^{\boldsymbol{X}_0}$ ist nach Lemma 1.4 stetig. Nach dem Stetigkeitssatz und der vorangehenden Bemerkung gilt $\mathbb{P}^{\boldsymbol{X}_n} \xrightarrow[n\to\infty]{\mathcal{D}} Q_0$, wobei Q_0 die charakteristische Funktion $\varphi^{\boldsymbol{X}_0}$ hat. Mit dem Eindeutigkeitssatz folgt $Q_0 = \mathbb{P}^{\boldsymbol{X}_0}$, also $\boldsymbol{X}_n \xrightarrow[n\to\infty]{\mathcal{D}} \boldsymbol{X}_0$. □

3. L^p-Räume

Zu diesem Kapitel vergleiche man [Kle13, Kapitel 7].

3.1. Konvexe Funktionen und die Jensen-Ungleichung

Definition 3.1 (Konvexe Funktion)**.** Seien $-\infty \le a < b \le \infty$. Eine Funktion $f\colon (a,b) \to \mathbb{R}$ heißt *konvex*, wenn

$$f\big((1-\lambda)x + \lambda y\big) \le (1-\lambda)f(x) + \lambda f(y) \qquad \text{für alle } x, y \in (a,b),\ \lambda \in [0,1] \tag{3.1}$$

gilt. Der Punkt $(1-\lambda)x + \lambda y$ heißt *Konvexkombination* von x und y. Die Funktion f heißt *strikt konvex*, falls in (3.1) „$<$“ statt „$\le$“ für alle $\lambda \in (0,1)$ und alle $x \ne y$ gilt.

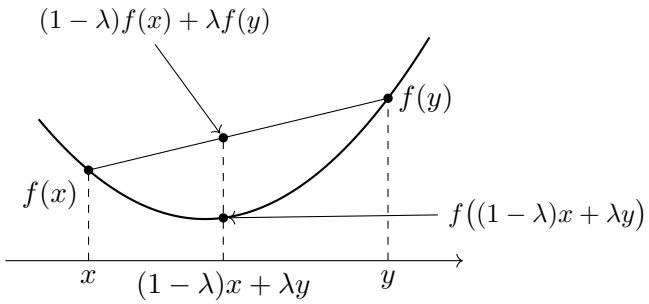

Abbildung 3.1.: Eine konvexe Funktion: Die Sehne zwischen $(x, f(x))$ und $(y, f(y))$ liegt oberhalb des Graphen.

Lemma 3.2. *Die Bedingung (3.1) ist äquivalent zu*

$$\frac{f(t)-f(x)}{t-x} \le \frac{f(y)-f(t)}{y-t} \qquad \textit{für alle } a < x < t < y < b. \tag{3.2}$$

Beweisidee. Es gilt

$$(3.2) \quad \Longleftrightarrow \quad f(t) \le \frac{t-x}{y-x} f(y) + \frac{y-t}{y-x} f(x).$$

Mit $\lambda := (t-x)/(y-x) \in (0,1)$ gilt $t = (1-\lambda)x + \lambda y$ und $(y-t)/(y-x) = 1-\lambda$, und die rechte Seite ist (3.1). □

Geometrisch besagt (3.2), dass die Steigungen der Sekanten von links nach rechts zunehmen, siehe Abbildung 3.2. Durch zweimalige Anwendung erhält man daraus auch

$$\frac{f(t)-f(x)}{t-x} \le \frac{f(y)-f(x)}{y-x} \le \frac{f(y)-f(t)}{y-t} \qquad \text{für alle } a < x < t < y < b.$$

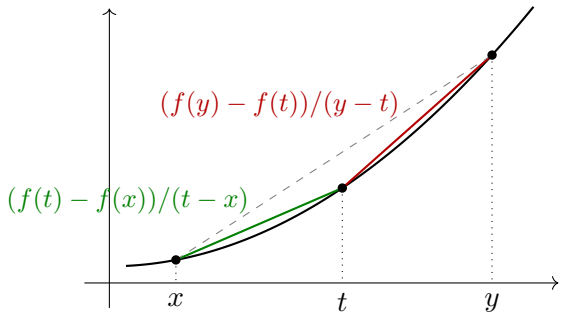

Abbildung 3.2.: Die Steigungsbedingung (3.2): Die linke (grüne) Sekante ist flacher als die rechte (rote). Die gestrichelte Sekante von x nach y liegt mit ihrer Steigung dazwischen.

Bemerkung 3.3. Für differenzierbare f gilt: Ist f' monoton wachsend, so ist f konvex; ist f' streng monoton wachsend, so ist f strikt konvex. Für zweimal differenzierbare f gilt: Aus $f'' \geq 0$ folgt, dass f konvex ist, und aus $f'' > 0$ folgt, dass f strikt konvex ist.

Satz 3.4. *Eine auf (a, b) konvexe Funktion ist stetig.*

Beweis. Seien $x_0 \in (a, b)$ und $a < u < x_0 < v < b$. Nach Lemma 3.2 und der anschließenden Drei-Sehnen-Ungleichung gilt für alle $z \in (u, v)$ mit $z \neq x_0$

$$\frac{f(x_0) - f(u)}{x_0 - u} \leq \frac{f(z) - f(x_0)}{z - x_0} \leq \frac{f(v) - f(x_0)}{v - x_0}.$$

Daraus folgt

$$\frac{|f(z) - f(x_0)|}{|z - x_0|} \leq L := \max\Big\{\frac{|f(x_0) - f(u)|}{x_0 - u}, \frac{|f(v) - f(x_0)|}{v - x_0}\Big\},$$

also $|f(z) - f(x_0)| \leq L|z - x_0|$ für alle $z \in (u, v)$, und somit ist f stetig in x_0. □

Satz 3.5 (Jensen-Ungleichung)**.** *Sei X eine integrierbare Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in (a, b). Sei f konvex auf (a, b). Dann gilt $\mathbb{E}[f(X)^-] < \infty$, wobei $f(X)^-$ den Negativteil von $f(X)$ bezeichnet, und*

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

Insbesondere ist $\mathbb{E}[f(X)] \in (-\infty, \infty]$ wohldefiniert.

Bemerkung. Wegen der Konvexität von $x \mapsto x^2$ liefert Satz 3.5 die bekannte Ungleichung $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$.

Beweis von Satz 3.5. Es gibt $\alpha, \beta \in \mathbb{R}$ mit $\alpha x + \beta \leq f(x)$ für alle $x \in (a, b)$, etwa eine Stützgerade wie unten. Daraus folgt

$$\mathbb{E}[f(X)^-] \leq \mathbb{E}[(\alpha X + \beta)^-] \leq |\beta| + |\alpha|\,\mathbb{E}[|X|] < \infty.$$

1. Fall: Es gilt $\mathbb{E}[X] \in (a, b)$. Sei $t^+ := D^+ f(\mathbb{E}[X])$ die rechtsseitige Ableitung von f in $\mathbb{E}[X]$, also die maximale Steigung einer Tangente an f in $\mathbb{E}[X]$; sie existiert wegen der Monotonie der Sekantensteigungen aus Lemma 3.2. Dann gilt

$$f(x) \geq t^+(x - \mathbb{E}[X]) + f(\mathbb{E}[X]) \qquad \text{für alle } x \in (a, b),$$

und daraus folgt

$$\mathbb{E}[f(X)] \geq t^+\,\mathbb{E}[X - \mathbb{E}[X]] + \mathbb{E}[f(\mathbb{E}[X])] = f(\mathbb{E}[X]).$$

2. Fall: Es gilt $\mathbb{E}[X] \in \{a, b\}$. Dann folgt $X = \mathbb{E}[X]$ fast sicher, etwa für $\mathbb{E}[X] = a$ aus $X - a \geq 0$ und $\mathbb{E}[X - a] = 0$. Dies widerspricht $X \in (a, b)$, also tritt dieser Fall nicht ein. □

3.2. Die Räume L^p

Bezeichnung 3.6. Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $p \in [1, \infty)$. Wir setzen

$$L^p(\Omega, \mathcal{F}, \mu) \coloneqq \Big\{ f \colon \Omega \to \mathbb{R} \text{ messbar} \colon \int_\Omega |f|^p \, \mathrm{d}\mu < \infty \Big\}.$$

Eine Funktion $f \in L^p(\Omega, \mathcal{F}, \mu)$ heißt *p-fach μ-integrierbar*. Oft schreiben wir $L^p(\mu)$ statt $L^p(\Omega, \mathcal{F}, \mu)$ oder nur L^p. Für $f \in L^p(\mu)$ setzen wir

$$\|f\|_p \coloneqq \Big(\int_\Omega |f|^p \, \mathrm{d}\mu \Big)^{1/p}.$$

Es gilt $\|f\|_p = 0$ genau dann, wenn $f = 0$ μ-fast-sicher gilt. Die Relation „$f = g$ μ-fast-sicher“ definiert eine Äquivalenzrelation $f \sim g$ und damit eine Äquivalenzklasse

$$\tilde{f} \coloneqq \{ g \in L^p \colon f = g \ \mu\text{-fast-sicher} \}$$

zu $f \in L^p(\mu)$. Im Raum der Äquivalenzklassen definiert $\|\tilde{f}\|_p \coloneqq \|f\|_p$ eine *Norm*, auf $L^p(\mu)$ hingegen nur eine *Pseudonorm*. In der Praxis wird dieser Unterschied oft ignoriert.

Bemerkung 3.7. (i) Für ein endliches Maß μ, das heißt $\mu(\Omega) < \infty$, gilt für $1 \le q < p$

$$f \in L^p \quad \Longrightarrow \quad f \in L^q,$$

denn $|f|^q \le \max\{1, |f|^p\} \le 1 + |f|^p$, also

$$\int |f|^q \, \mathrm{d}\mu \le \mu(\Omega) + \int |f|^p \, \mathrm{d}\mu < \infty.$$

(ii) Für $(\Omega, \mathcal{F}, \mu) = (\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0), \text{Zählmaß})$, also $\mu(\{k\}) = 1$ für alle $k \in \mathbb{N}_0$, ist

$$L^p(\mu) = \Big\{ (f_k)_{k \in \mathbb{N}_0} \text{ reelle Folge} \colon \sum_{k=0}^{\infty} |f_k|^p < \infty \Big\}.$$

Satz 3.8 (Hölder-Ungleichung)**.** *Seien $p, q \in (1, \infty)$ mit $1 = 1/p + 1/q$. Dann gilt für $f \in L^p$ und $g \in L^q$ die* Hölder-Ungleichung

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

beziehungsweise $\|fg\|_1 \le \|f\|_p \|g\|_q$. Insbesondere gilt die Cauchy–Schwarz-Ungleichung

$$\int |fg| \, \mathrm{d}\mu \le \Big(\int |f|^2 \, \mathrm{d}\mu \Big)^{1/2} \Big(\int |g|^2 \, \mathrm{d}\mu \Big)^{1/2}.$$

Bemerkung 3.9. (i) Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ und Zufallsvariablen X, Y schreibt sich Satz 3.8 als

$$\mathbb{E}[|XY|] \le \mathbb{E}[|X|^p]^{1/p} \, \mathbb{E}[|Y|^q]^{1/q}$$

beziehungsweise

$$\mathbb{E}[|XY|] \le \sqrt{\mathbb{E}[|X|^2]} \sqrt{\mathbb{E}[|Y|^2]}.$$

(ii) Es gilt

$$|\operatorname{Cov}(X,Y)| = \big|\mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big]\big| \le \mathbb{E}\big[|X-\mathbb{E}[X]|\,|Y-\mathbb{E}[Y]|\big]$$
$$\le \sqrt{\mathbb{E}\big[(X-\mathbb{E}[X])^2\big]}\sqrt{\mathbb{E}\big[(Y-\mathbb{E}[Y])^2\big]} = \sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]},$$

wobei wir zuerst die Jensen-Ungleichung (für $x \mapsto |x|$) und dann die Cauchy–Schwarz-Ungleichung verwendet haben. Also gilt

$$|\operatorname{Corr}(X,Y)| = \frac{|\operatorname{Cov}(X,Y)|}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} \le 1.$$

Beweis von Satz 3.8. 1. Fall: Es gilt $\|f\|_p = 0$ oder $\|g\|_q = 0$; ohne Einschränkung sei $\|f\|_p = 0$. Dann gilt $f = 0$ μ-fast-sicher, also $fg = 0$ μ-fast-sicher, und damit $\int |fg|\,\mathrm{d}\mu = 0$.
2. Fall: Es gilt $\|f\|_p, \|g\|_q \in (0,\infty)$. Mit $F := |f|/\|f\|_p$ und $G := |g|/\|g\|_q$ gilt

$$\int F^p\,\mathrm{d}\mu = \int G^q\,\mathrm{d}\mu = 1.$$

Wir zeigen zunächst

$$F(\omega)G(\omega) \le \frac{1}{p}F(\omega)^p + \frac{1}{q}G(\omega)^q \qquad \text{für alle } \omega \in \Omega. \tag{3.3}$$

Ist $F(\omega) = 0$ oder $G(\omega) = 0$, so gilt (3.3), da die rechte Seite nichtnegativ ist; ist $F(\omega) = \infty$ oder $G(\omega) = \infty$, so ist die rechte Seite unendlich. Für alle ω mit $0 < F(\omega) < \infty$ und $0 < G(\omega) < \infty$ definieren wir $F'(\omega)$ und $G'(\omega)$ durch

$$F(\omega) = \mathrm{e}^{\frac{1}{p}F'(\omega)} \qquad \text{und} \qquad G(\omega) = \mathrm{e}^{\frac{1}{q}G'(\omega)}.$$

Wir betrachten das Wahrscheinlichkeitsmaß ν_ω auf $\{F'(\omega), G'(\omega)\}$ mit $\nu_\omega(\{F'(\omega)\}) = 1/p$ und $\nu_\omega(\{G'(\omega)\}) = 1/q$ (im Fall $F'(\omega) = G'(\omega)$ ist ν_ω das Punktmaß in diesem Punkt). Ist X eine Zufallsvariable mit Verteilung ν_ω, so gilt mit der Jensen-Ungleichung, Satz 3.5, für die konvexe Funktion $x \mapsto \mathrm{e}^x$

$$\frac{1}{p}\mathrm{e}^{F'(\omega)} + \frac{1}{q}\mathrm{e}^{G'(\omega)} = \mathbb{E}_{\nu_\omega}\big[\mathrm{e}^X\big] \ge \mathrm{e}^{\mathbb{E}_{\nu_\omega}[X]} = \mathrm{e}^{\frac{1}{p}F'(\omega)+\frac{1}{q}G'(\omega)} = F(\omega)G(\omega),$$

und das ist (3.3). Integration von (3.3) liefert

$$\int FG\,\mathrm{d}\mu \le \frac{1}{p}\underbrace{\int F^p\,\mathrm{d}\mu}_{=1} + \frac{1}{q}\underbrace{\int G^q\,\mathrm{d}\mu}_{=1} = 1,$$

also $\int |fg|\,\mathrm{d}\mu \le \|f\|_p\|g\|_q$. □

Satz 3.10 (Minkowski-Ungleichung)**.** *Seien $p \in [1,\infty)$ und $f, g \in L^p$. Dann gilt $f + g \in L^p$ und die Dreiecksungleichung*

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

Bemerkung. Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ schreibt sich dies als

$$\mathbb{E}\big[|X+Y|^p\big]^{1/p} \le \mathbb{E}\big[|X|^p\big]^{1/p} + \mathbb{E}\big[|Y|^p\big]^{1/p}.$$

Beweis von Satz 3.10. Für $p = 1$ folgt die Aussage durch Integration der Dreiecksungleichung $|f + g| \leq |f| + |g|$; sei also $p > 1$. Es gilt

$$|f + g|^p \leq (2 \max\{|f|, |g|\})^p \leq 2^p (|f|^p + |g|^p),$$

also $f + g \in L^p$. Weiterhin gilt wegen $|f + g| \leq |f| + |g|$

$$|f + g|^p \leq |f|\,|f + g|^{p-1} + |g|\,|f + g|^{p-1}.$$

Sei q so gewählt, dass $1/p + 1/q = 1$; dann gilt $q(p - 1) = p$. Mit der Hölder-Ungleichung, Satz 3.8, folgt

$$\begin{aligned}
\int |f + g|^p \,\mathrm{d}\mu &\leq \int |f|\,|f + g|^{p-1} \,\mathrm{d}\mu + \int |g|\,|f + g|^{p-1} \,\mathrm{d}\mu \\
&\leq \Big(\int |f|^p \,\mathrm{d}\mu\Big)^{1/p} \Big(\int |f + g|^{q(p-1)} \,\mathrm{d}\mu\Big)^{1/q} \\
&\quad + \Big(\int |g|^p \,\mathrm{d}\mu\Big)^{1/p} \Big(\int |f + g|^{q(p-1)} \,\mathrm{d}\mu\Big)^{1/q} \\
&= \Big[\Big(\int |f|^p \,\mathrm{d}\mu\Big)^{1/p} + \Big(\int |g|^p \,\mathrm{d}\mu\Big)^{1/p}\Big] \Big(\int |f + g|^p \,\mathrm{d}\mu\Big)^{1/q}.
\end{aligned}$$

Ist $\int |f+g|^p \,\mathrm{d}\mu = 0$, so ist nichts zu zeigen. Andernfalls dividieren wir durch $\big(\int |f+g|^p \,\mathrm{d}\mu\big)^{1/q} \in (0, \infty)$ und erhalten wegen $1 - 1/q = 1/p$

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

□

Bemerkung 3.11. Die Räume L^p, $1 \leq p < \infty$, mit der Norm $\|\cdot\|_p$ sind (im Sinne von Äquivalenzklassen) reelle *Banachräume*, das heißt vollständige normierte Vektorräume.

3.3. Konvergenz im p-ten Mittel

Definition 3.12 (Konvergenz im p-ten Mittel)**.** Seien $(X_n)_{n\in\mathbb{N}_0}$ reelle Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $X_n \in L^p(\mathbb{P})$ für alle $n \in \mathbb{N}_0$, und sei $p \in [1, \infty)$. Wir sagen, X_n *konvergiert im p-ten Mittel* gegen X_0, falls

$$\|X_n - X_0\|_p \xrightarrow[n\to\infty]{} 0.$$

Für $p = 2$ sprechen wir auch von Konvergenz im *quadratischen Mittel*.

Beispiel 3.13. Sei $X_n \sim \mathrm{Bin}(n, p)$ mit $n \in \mathbb{N}$ und $p \in [0, 1]$. Dann konvergiert X_n/n im quadratischen Mittel gegen p, denn

$$\mathbb{E}\Big[\Big|\frac{X_n}{n} - p\Big|^2\Big] = \frac{1}{n^2}\mathbb{E}\big[(X_n - np)^2\big] = \frac{1}{n^2}\operatorname{Var}[X_n] = \frac{1}{n^2}\, np(1 - p) \xrightarrow[n\to\infty]{} 0.$$

Hier bezeichnet p ausnahmsweise den Parameter der Binomialverteilung und nicht den Exponenten.

Satz 3.14 (Markov-Ungleichung)**.** *Für jede reelle Zufallsvariable X und jedes $r \geq 0$ gilt für alle $\varepsilon > 0$*

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^r}\mathbb{E}\big[|X|^r\big].$$

Beweis. Siehe die Vorlesung „Einführung in die Stochastik“. □

Bemerkung. Die Ungleichung gilt sogar für monoton wachsende Funktionen $g\colon [0,\infty) \to [0,\infty)$ mit $g(\varepsilon) > 0$ statt nur für $x \mapsto x^r$, das heißt, es gilt $\mathbb{P}(|X| \geq \varepsilon) \leq \mathbb{E}[g(|X|)]/g(\varepsilon)$.

Satz 3.15 (Konvergenz im p-ten Mittel impliziert Konvergenz in Wahrscheinlichkeit)**.** *Für eine Folge $(X_n)_{n\in\mathbb{N}_0}$ reeller Zufallsvariablen gilt*

$$X_n \xrightarrow[n\to\infty]{} X_0 \text{ im } p\text{-ten Mittel} \quad \Longrightarrow \quad X_n \xrightarrow[n\to\infty]{} X_0 \text{ in Wahrscheinlichkeit.}$$

Beweis. Mit der Markov-Ungleichung, Satz 3.14, gilt für alle $\varepsilon > 0$

$$\mathbb{P}\big(|X_n - X_0| \geq \varepsilon\big) \leq \frac{1}{\varepsilon^p}\mathbb{E}\big[|X_n - X_0|^p\big] \xrightarrow[n\to\infty]{} 0. \qquad \square$$

Satz 3.16 (Satz von Riesz)**.** *Es gelte $X_n \to X_0$ fast sicher und $X_n, X_0 \in L^p$ für alle $n \in \mathbb{N}$. Dann gilt*

$$\mathbb{E}\big[|X_n|^p\big] \xrightarrow[n\to\infty]{} \mathbb{E}\big[|X_0|^p\big] \quad \Longleftrightarrow \quad X_n \xrightarrow[n\to\infty]{} X_0 \text{ im } p\text{-ten Mittel.}$$

Beweis. Ohne Beweis. $\square$

Die Beziehungen zwischen den bisher betrachteten Konvergenzarten sind in Abbildung 3.3 zusammengefasst.

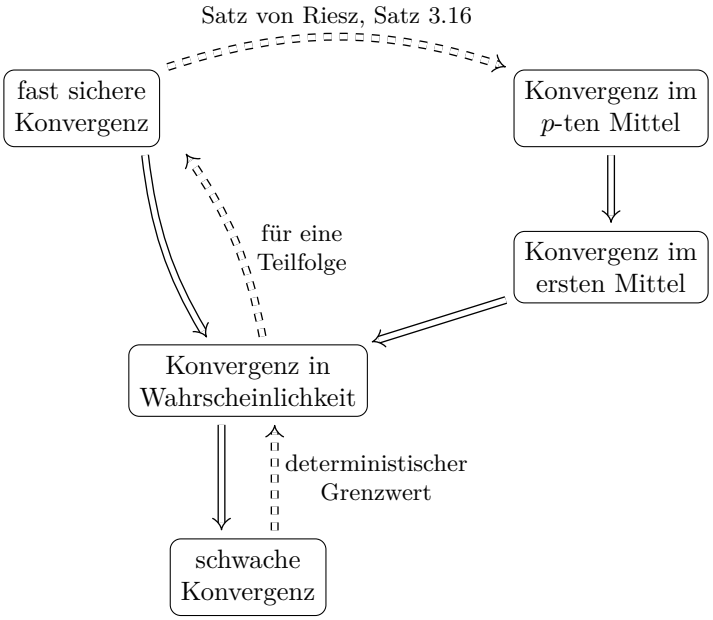

Abbildung 3.3.: Zusammenhang der Konvergenzarten für Zufallsvariablen auf einem Wahrscheinlichkeitsraum. Gestrichelte Pfeile gelten nur unter Zusatzbedingungen.

4. Ergodentheorie

Zu diesem Kapitel vergleiche man [Kle13, Kapitel 20], [Dur19, Kapitel 6] und [Wal82].

In diesem Kapitel behandeln wir eine Erweiterung des Gesetzes der großen Zahlen. Statt der Unabhängigkeit der Summanden setzen wir nur noch voraus, dass die Folge in einem geeigneten Sinne stationär ist und keine nichttrivialen invarianten Ereignisse besitzt.

4.1. Maßerhaltende Abbildungen und stationäre Folgen

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ stets ein Wahrscheinlichkeitsraum.

Definition 4.1 (Maßerhaltende Transformation)**.** Eine Abbildung $T\colon \Omega \to \Omega$ heißt *maßerhaltend*, falls T messbar bezüglich $\mathcal{F}$ und $\mathcal{F}$ ist und $\mathbb{P}$ invariant unter T ist, das heißt, falls

$$\mathbb{P} \circ T^{-1} = \mathbb{P}.$$

Definition 4.2 (Stationarität)**.** Sei $X = (X_n)_{n\in\mathbb{N}_0}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in einem Messraum $(E, \mathcal{E})$. Die Folge X heißt *stationär*, falls

$$X \overset{\mathrm{d}}{=} \widetilde{X} \coloneqq (X_{n+1})_{n\in\mathbb{N}_0},$$

das heißt, falls X und die um einen Schritt verschobene Folge $\widetilde{X}$ dieselbe Verteilung auf $(E^{\mathbb{N}_0}, \mathcal{E}^{\otimes\mathbb{N}_0})$ besitzen.

Bemerkung. Eine Folge X ist genau dann stationär, wenn für alle $n \in \mathbb{N}_0$ und alle $A \in \mathcal{E}^{\otimes(n+1)}$ gilt

$$\mathbb{P}\big((X_0, \dots, X_n) \in A\big) = \mathbb{P}\big((X_1, \dots, X_{n+1}) \in A\big).$$

Dies folgt, da die Zylindermengen $\{x \in E^{\mathbb{N}_0} \colon (x_0, \dots, x_n) \in A\}$ ein $\cap$-stabiles Erzeugendensystem von $\mathcal{E}^{\otimes\mathbb{N}_0}$ bilden und Wahrscheinlichkeitsmaße durch ihre Werte auf einem solchen eindeutig bestimmt sind. Durch Iteration folgt außerdem $X \overset{\mathrm{d}}{=} (X_{n+k})_{n\in\mathbb{N}_0}$ für alle $k \in \mathbb{N}_0$.

Für eine Abbildung $T\colon \Omega \to \Omega$ setzen wir $T^0 \coloneqq \mathrm{id}$ und $T^n \coloneqq T \circ \cdots \circ T$ (n-mal) für $n \in \mathbb{N}$.

Satz 4.3 (Maßerhaltung versus Stationarität)**.**

(i) Seien T eine maßerhaltende Abbildung und X_0 eine Zufallsvariable mit Werten in $(E, \mathcal{E})$. Dann ist

$$X_n \coloneqq X_0 \circ T^n, \qquad n \in \mathbb{N}_0,$$

eine stationäre Folge.

(ii) Sei $X = (X_n)_{n\in\mathbb{N}_0}$ eine stationäre Folge mit Werten in $(E, \mathcal{E})$. Dann ist

$$T\big((x_n)_{n\in\mathbb{N}_0}\big) \coloneqq (x_{n+1})_{n\in\mathbb{N}_0}, \qquad (x_n)_{n\in\mathbb{N}_0} \in E^{\mathbb{N}_0},$$

eine maßerhaltende Abbildung auf $(E^{\mathbb{N}_0}, \mathcal{E}^{\otimes\mathbb{N}_0}, \mathbb{P}^X)$. Für den Prozess der Projektionen $(\pi_n)_{n\in\mathbb{N}_0}$, $\pi_n\colon E^{\mathbb{N}_0} \to E$, $\pi_n(x) \coloneqq x_n$, gilt $\pi_n = \pi_0 \circ T^n$; er ist also von der Form in (i) und besitzt unter $\mathbb{P}^X$ dieselbe Verteilung wie X.

Beweis. (i) Seien $n \in \mathbb{N}_0$ und $B \in \mathcal{E}^{\otimes(n+1)}$. Wegen $X_{k+1} = X_k \circ T$ und da T maßerhaltend ist, gilt

$$\begin{aligned}\mathbb{P}\big((X_1, \dots, X_{n+1}) \in B\big) &= \mathbb{P}\big((X_0 \circ T, \dots, X_n \circ T) \in B\big) \\ &= \mathbb{P}\big(T^{-1}\{(X_0, \dots, X_n) \in B\}\big) = \mathbb{P}\big((X_0, \dots, X_n) \in B\big).\end{aligned}$$

Nach der Bemerkung nach Definition 4.2 ist $(X_n)_{n\in\mathbb{N}_0}$ also stationär.

(ii) Die Abbildung T ist messbar, denn für jedes $n \in \mathbb{N}_0$ ist $\pi_n \circ T = \pi_{n+1}$ messbar, und die Produkt-σ-Algebra wird von den Projektionen erzeugt. Außerdem gilt $T \circ X = (X_{n+1})_{n\in\mathbb{N}_0}$, und mit der Stationarität folgt für alle $A \in \mathcal{E}^{\otimes\mathbb{N}_0}$

$$\mathbb{P}^X\big(T^{-1}(A)\big) = \mathbb{P}\big(\{\omega\colon T(X(\omega)) \in A\}\big) = \mathbb{P}\big((T \circ X)^{-1}(A)\big) = \mathbb{P}\big(X^{-1}(A)\big) = \mathbb{P}^X(A).$$

Schließlich gilt $\pi_0 \circ T^n = \pi_n$ nach Definition von T, und $(\pi_n)_{n\in\mathbb{N}_0}$ ist unter $\mathbb{P}^X$ gerade wie X verteilt, da die Abbildung $x \mapsto (\pi_n(x))_{n\in\mathbb{N}_0}$ die Identität auf $E^{\mathbb{N}_0}$ ist. □

4.2. Invariante Mengen, Ergodizität und Mischung

Definition 4.4 (Invariante Menge)**.** Seien $T\colon \Omega \to \Omega$ eine Abbildung und $A \subset \Omega$. Die Menge A heißt *T-invariant*, falls $T^{-1}(A) = A$.

Satz 4.5 (Teil-σ-Algebra der invarianten Mengen)**.** *Sei T eine maßerhaltende Abbildung. Dann ist*

$$\mathcal{I} \coloneqq \{A \in \mathcal{F}\colon T^{-1}(A) = A\}$$

eine Teil-σ-Algebra von $\mathcal{F}$, die σ-Algebra der T-invarianten Mengen.

Beweis. Wir prüfen die drei Eigenschaften einer σ-Algebra. Es gilt $\Omega \in \mathcal{I}$, da $T^{-1}(\Omega) = \Omega$. Ist $A \in \mathcal{I}$, so gilt

$$T^{-1}(A^{\mathrm{c}}) = \{\omega\colon T(\omega) \in A^{\mathrm{c}}\} = \Omega \setminus T^{-1}(A) = A^{\mathrm{c}},$$

also $A^{\mathrm{c}} \in \mathcal{I}$. Sind schließlich $A_1, A_2, \dots \in \mathcal{I}$, so gilt

$$T^{-1}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \Big\{\omega\colon T(\omega) \in \bigcup_{i=1}^{\infty} A_i\Big\} = \bigcup_{i=1}^{\infty}\{\omega\colon T(\omega) \in A_i\} = \bigcup_{i=1}^{\infty} T^{-1}(A_i) = \bigcup_{i=1}^{\infty} A_i. \qquad \square$$

Definition 4.6 (Ergodizität und Mischung)**.** Sei T eine maßerhaltende Abbildung.

(i) Die Abbildung T heißt *ergodisch*, falls $\mathcal{I}$ eine $\mathbb{P}$-triviale σ-Algebra ist, also falls

$$\mathbb{P}(A) \in \{0, 1\} \qquad \text{für alle } A \in \mathcal{I}.$$

(ii) Die Abbildung T heißt *mischend*, falls für alle $A, B \in \mathcal{F}$ gilt

$$\lim_{n\to\infty} \mathbb{P}\big(A \cap T^{-n}(B)\big) = \mathbb{P}(A)\mathbb{P}(B).$$

Bemerkung. Mischung charakterisiert, dass die Ereignisse A und B asymptotisch unabhängig werden, wenn man sie „auseinanderschiebt“.

Satz 4.7 (Mischend impliziert ergodisch)**.** *Mischende Transformationen sind ergodisch.*

Beweis. Sei $A \in \mathcal{I}$. Dann gilt $T^{-n}(A) = A$ für alle $n \in \mathbb{N}$, und mit der Mischungseigenschaft folgt

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}\big(A \cap T^{-n}(A)\big) \xrightarrow[n\to\infty]{} \mathbb{P}(A)\mathbb{P}(A).$$

Also gilt $\mathbb{P}(A) = \mathbb{P}(A)^2$ und somit $\mathbb{P}(A) \in \{0, 1\}$. □

Bemerkung.

(1) Es genügt tatsächlich, die Mischungseigenschaft für die Mengen eines $\cap$-stabilen Erzeugers von $\mathcal{F}$ zu prüfen; siehe das Argument in Beispiel 4.8 (i).

(2) Die Abbildung $T\colon (x_n)_{n\in\mathbb{N}_0} \mapsto (x_{n+1})_{n\in\mathbb{N}_0}$ aus Satz 4.3 (ii) wird oft *Verschiebungsoperator* (englisch *shift*) genannt.

(3) Ein stationärer Prozess $X = (X_n)_{n\in\mathbb{N}_0}$ heißt *ergodisch* beziehungsweise *mischend*, falls der assoziierte Verschiebungsoperator auf $(E^{\mathbb{N}_0}, \mathcal{E}^{\otimes\mathbb{N}_0}, \mathbb{P}^X)$ ergodisch beziehungsweise mischend ist.

Beispiel 4.8.

(i) *Unabhängige, identisch verteilte Folgen.* Ist $X = (X_n)_{n\in\mathbb{N}_0}$ eine Folge unabhängiger, identisch verteilter (u.i.v.) Zufallsvariablen mit Werten in $(E, \mathcal{E})$, so ist X mischend. Wir arbeiten auf dem kanonischen Raum $(E^{\mathbb{N}_0}, \mathcal{E}^{\otimes\mathbb{N}_0}, \mathbb{P}^X)$ mit dem Verschiebungsoperator T. Wir betrachten zunächst Zylindermengen $A, B \in \mathcal{E}^{\otimes\mathbb{N}_0}$ mit

$$X \in A \iff (X_{m_1}, \dots, X_{m_1+t_1}) \in \widetilde{A} \quad \text{und} \quad X \in B \iff (X_{m_2}, \dots, X_{m_2+t_2}) \in \widetilde{B}$$

für gewisse $m_1, m_2, t_1, t_2 \in \mathbb{N}_0$ und messbare $\widetilde{A} \subset E^{t_1+1}$, $\widetilde{B} \subset E^{t_2+1}$. Wegen $X \in T^{-n}(B) \iff (X_{n+m_2}, \dots, X_{n+m_2+t_2}) \in \widetilde{B}$ und der Unabhängigkeit gilt

$$\mathbb{P}^X\big(A \cap T^{-n}(B)\big) = \mathbb{P}\big(\{X \in A\} \cap \{X \in T^{-n}(B)\}\big) = \mathbb{P}(X \in A)\,\mathbb{P}\big(X \in T^{-n}(B)\big)$$

für alle hinreichend großen n, nämlich sobald $n + m_2 > m_1 + t_1$. Wegen der identischen Verteilung ist die rechte Seite gleich $\mathbb{P}(X \in A)\mathbb{P}(X \in B) = \mathbb{P}^X(A)\mathbb{P}^X(B)$.

Für eine feste Zylindermenge A ist nun

$$\mathcal{D} \coloneqq \big\{B \in \mathcal{E}^{\otimes\mathbb{N}_0} \colon \mathbb{P}^X\big(A \cap T^{-n}(B)\big) \to \mathbb{P}^X(A)\mathbb{P}^X(B)\big\}$$

ein Dynkin-System, welches alle Zylindermengen enthält. Die Abgeschlossenheit unter Komplementen folgt aus $\mathbb{P}^X(A \cap T^{-n}(B^{\mathrm{c}})) = \mathbb{P}^X(A) - \mathbb{P}^X(A \cap T^{-n}(B))$; für paarweise disjunkte $B_1, B_2, \dots \in \mathcal{D}$ folgt sie aus $\mathbb{P}^X(A \cap T^{-n}(\bigcup_i B_i)) = \sum_i \mathbb{P}^X(A \cap T^{-n}(B_i))$ mit dem Satz von der dominierten Konvergenz für Reihen, da $\mathbb{P}^X(A \cap T^{-n}(B_i)) \le \mathbb{P}^X(T^{-n}(B_i)) = \mathbb{P}^X(B_i)$ summierbar ist. Da die Zylindermengen ein $\cap$-stabiles Erzeugendensystem bilden, folgt mit dem π-λ-Satz $\mathcal{D} = \mathcal{E}^{\otimes\mathbb{N}_0}$. Ebenso ist für jedes $B \in \mathcal{E}^{\otimes\mathbb{N}_0}$

$$\mathcal{C}_B \coloneqq \big\{A \in \mathcal{E}^{\otimes\mathbb{N}_0} \colon \mathbb{P}^X\big(A \cap T^{-n}(B)\big) \to \mathbb{P}^X(A)\mathbb{P}^X(B)\big\}$$

ein Dynkin-System, welches nach dem ersten Schritt alle Zylindermengen enthält. Also gilt $\mathcal{C}_B = \mathcal{E}^{\otimes\mathbb{N}_0}$ für alle $B \in \mathcal{E}^{\otimes\mathbb{N}_0}$, das heißt, T ist mischend.

(ii) *Rotationen des Kreises.* Wir betrachten den Einheitskreis $\mathbb{T} \coloneqq \{z \in \mathbb{C} \colon |z| = 1\}$ mit der Borel-σ-Algebra $\mathcal{B}(\mathbb{T})$ und dem normierten Bogenmaß $\mu \coloneqq (1/(2\pi))\mathrm{Leb}_{\mathbb{T}}$ als Wahrscheinlichkeitsmaß. Für $c \in \mathbb{T}$ ist die *Rotation*

$$T_c(w) \coloneqq c\,w, \qquad w \in \mathbb{T},$$

maßerhaltend. Schreiben wir nämlich $c = \mathrm{e}^{\mathrm{i}s}$ und $w = \mathrm{e}^{\mathrm{i}t}$, so gilt $T_c(w) = \mathrm{e}^{\mathrm{i}(t+s)}$; die Rotation verschiebt also nur das Argument um s, und das Bogenmaß ist invariant unter solchen Verschiebungen.

Satz 4.9 (Ergodizität und Rotation)**.** *Der Rotationsoperator T_c aus Beispiel 4.8 (ii) ist genau dann ergodisch, wenn c keine Einheitswurzel ist, das heißt, wenn $c^n \neq 1$ für alle $n \in \mathbb{N}$.*

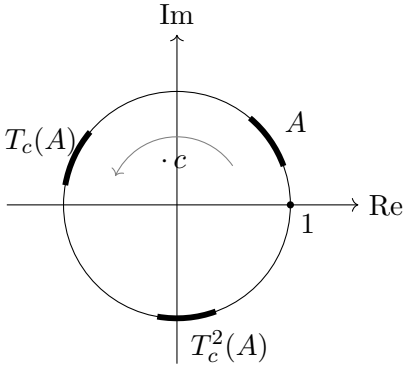

Abbildung 4.1.: Zum Beweis von Satz 4.9 für $c = \mathrm{e}^{2\pi\mathrm{i}/3}$, also $c^3 = 1$: Die Vereinigung $B = A \cup T_c(A) \cup T_c^2(A)$ der gedrehten Kopien eines kleinen Bogens A ist T_c-invariant, hat aber Maß in $(0, 1)$.

Beweis. „⇒“: Wir zeigen die Kontraposition. Sei also $c^n = 1$ für ein $n \in \mathbb{N}$. Dann gilt $T_c^n = \mathrm{id}$. Wir betrachten einen kleinen Kreisbogen $A = \{\mathrm{e}^{\mathrm{i}t} \colon 0 \le t \le \varepsilon\}$ und setzen

$$B \coloneqq \bigcup_{k=0}^{n-1} T_c^k(A),$$

siehe Abbildung 4.1. Mit $c = \mathrm{e}^{\mathrm{i}s}$ gilt

$$\begin{aligned} T_c^{-1}(B) &= \{\mathrm{e}^{\mathrm{i}t} \colon \mathrm{e}^{\mathrm{i}t}\mathrm{e}^{\mathrm{i}s} \in B\} = \{\mathrm{e}^{\mathrm{i}t} \colon \exists k \colon \mathrm{e}^{\mathrm{i}t}\mathrm{e}^{\mathrm{i}s} = \mathrm{e}^{\mathrm{i}sk}\mathrm{e}^{\mathrm{i}r} \text{ für ein } \mathrm{e}^{\mathrm{i}r} \in A\} \\ &= \{\mathrm{e}^{\mathrm{i}t} \colon \exists k \colon \mathrm{e}^{\mathrm{i}t} = \mathrm{e}^{\mathrm{i}s(k-1)}\mathrm{e}^{\mathrm{i}r} \text{ für ein } \mathrm{e}^{\mathrm{i}r} \in A\} = B, \end{aligned}$$

wobei wir $c^{-1} = c^{n-1}$ verwendet haben. Also ist $B \in \mathcal{I}$, aber es gilt $0 < \varepsilon/(2\pi) \le \mu(B) \le n\varepsilon/(2\pi) < 1$ für ε klein genug. Somit ist T_c nicht ergodisch.

„⇐“: Sei c keine Einheitswurzel. Dann gilt $c^n \ne c^m$ für alle $m \ne n$, und man kann zeigen, dass die Bahn $\{c^n w \colon n \in \mathbb{N}_0\}$ für jedes $w \in \mathbb{T}$ dicht in $\mathbb{T}$ liegt. Die Dichtheit der Bahnen allein liefert aber noch keine Aussage über messbare invariante Mengen; wir argumentieren daher mit Fourierkoeffizienten. Sei $A \in \mathcal{I}$, also $T_c^{-1}(A) = A$ und damit $\mathbb{1}_A(cw) = \mathbb{1}_A(w)$ für alle $w \in \mathbb{T}$. Für die Fourierkoeffizienten

$$a_k \coloneqq \int_{\mathbb{T}} \mathbb{1}_A(w)\, w^{-k}\, \mu(\mathrm{d}w), \qquad k \in \mathbb{Z},$$

gilt mit der Substitution $u = cw$ und der Rotationsinvarianz von μ

$$a_k = \int_{\mathbb{T}} \mathbb{1}_A(cw)\, w^{-k}\, \mu(\mathrm{d}w) = \int_{\mathbb{T}} \mathbb{1}_A(u)\, (c^{-1}u)^{-k}\, \mu(\mathrm{d}u) = c^k a_k.$$

Da $c^k \ne 1$ für $k \ne 0$, folgt $a_k = 0$ für alle $k \ne 0$. Da das trigonometrische System $\{w \mapsto w^k \colon k \in \mathbb{Z}\}$ eine Orthonormalbasis von $L^2(\mathbb{T}, \mu)$ ist, folgt $\mathbb{1}_A = a_0$ μ-fast sicher. Eine Indikatorfunktion ist aber nur dann fast sicher konstant, wenn $\mu(A) \in \{0, 1\}$. Also ist T_c ergodisch. □

Satz 4.10 (Rotation und Mischung)**.** *Für alle $c \in \mathbb{T}$ ist T_c nicht mischend.*

Beweis. Ist c eine Einheitswurzel, so ist T_c nach Satz 4.9 nicht ergodisch und nach Satz 4.7 somit auch nicht mischend. Es genügt also anzunehmen, dass c keine Einheitswurzel ist. Dann existiert eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $n_k \to \infty$ und $c^{n_k} \to 1$ für $k \to \infty$. Da nämlich $\mathbb{T}$ kompakt ist, besitzt die Folge $(c^m)_{m\in\mathbb{N}}$ eine konvergente Teilfolge $(c^{m_k})_{k\in\mathbb{N}}$, und $n_k \coloneqq m_{k+1} - m_k$

erfüllt $c^{n_k} = c^{m_{k+1}}/c^{m_k} \to 1$; dabei gilt $n_k \to \infty$, da sonst ein Wert n unendlich oft auftreten und $c^n = 1$ gelten würde. Sei nun J ein Kreisbogen mit $0 < \mu(J) < 1$. Da $T_c^{-n_k}(J) = c^{-n_k} J$ der um den Winkel $\arg c^{-n_k} \to 0$ gedrehte Bogen ist, gilt

$$\mu\big(J \cap T_c^{-n_k}(J)\big) = \mu\big(J \cap T_{c^{n_k}}^{-1}(J)\big) \xrightarrow[k\to\infty]{} \mu(J) \neq \mu(J)^2.$$

Also ist T_c nicht mischend. □

Beispiel 4.11 (Markovketten)**.** Jede irreduzible, aperiodische und stationäre Markovkette $X = (X_n)_{n\in\mathbb{N}_0}$ mit endlichem Zustandsraum E ist mischend, das heißt, ihr zugehöriger Verschiebungsoperator ist mischend. In der Tat, seien $(p_i^0)_{i\in E}$ die Startverteilung und $(p_{ij})_{i,j\in E}$ die Übergangsmatrix. Wegen der Stationarität gilt

$$\sum_{i\in E} p_i^0 p_{ij} = p_j^0 \qquad \text{für alle } j \in E,$$

und nach dem Konvergenzsatz für Markovketten gilt für die n-Schritt-Übergangswahrscheinlichkeiten

$$p_{ij}^{(n)} \xrightarrow[n\to\infty]{} p_j^0 \qquad \text{für alle } i, j \in E.$$

Mit dem Argument aus Beispiel 4.8 (i) genügt es, Zylindermengen

$$A = \{(i_n)_{n\in\mathbb{N}_0} \colon (i_0, \dots, i_r) \in \widetilde{A}\} \quad \text{und} \quad B = \{(i_n)_{n\in\mathbb{N}_0} \colon (i_0, \dots, i_s) \in \widetilde{B}\}$$

mit $r, s \in \mathbb{N}_0$, $\widetilde{A} \subset E^{r+1}$ und $\widetilde{B} \subset E^{s+1}$ zu betrachten; diese bilden ein $\cap$-stabiles Erzeugendensystem. Für alle $n > r$ gilt dann

$$\begin{aligned}\mathbb{P}^X\big(A \cap T^{-n}(B)\big) &= \mathbb{P}\big((X_0, \dots, X_r) \in \widetilde{A},\ (X_n, \dots, X_{n+s}) \in \widetilde{B}\big) \\ &= \sum_{(i_0,\dots,i_r)\in\widetilde{A}} \ \sum_{(j_0,\dots,j_s)\in\widetilde{B}} p_{i_0}^0 p_{i_0 i_1} \cdots p_{i_{r-1} i_r}\, p_{i_r j_0}^{(n-r)}\, p_{j_0 j_1} \cdots p_{j_{s-1} j_s}.\end{aligned}$$

Da die Summen endlich sind und $p_{i_r j_0}^{(n-r)} \to p_{j_0}^0$ gilt, folgt

$$\lim_{n\to\infty} \mathbb{P}^X\big(A \cap T^{-n}(B)\big) = \mathbb{P}\big((X_0, \dots, X_r) \in \widetilde{A}\big)\, \mathbb{P}\big((X_0, \dots, X_s) \in \widetilde{B}\big) = \mathbb{P}^X(A)\mathbb{P}^X(B).$$

4.3. Die Ergodensätze

Satz 4.12 (Ergodensatz von Birkhoff [Bir31])**.** *Sei T eine maßerhaltende und ergodische Abbildung auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt für alle $X_0 \in L^1(\mathbb{P})$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X_0 \circ T^k = \mathbb{E}[X_0] \qquad \mathbb{P}\text{-fast sicher.}$$

Bevor wir Satz 4.12 beweisen, formulieren wir die Folgerung für stationäre Folgen.

Korollar 4.13 (Ergodensatz für stationäre Prozesse)**.** *Sei $X = (X_n)_{n\in\mathbb{N}_0}$ eine ergodische stationäre Folge von Zufallsvariablen mit Zustandsraum $(E, \mathcal{E})$. Dann gilt für alle $m \in \mathbb{N}_0$ und alle messbaren Testfunktionen $f \colon E^{m+1} \to \mathbb{R}$ mit $f(X_0, \dots, X_m) \in L^1(\mathbb{P})$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, \dots, X_{k+m}) = \mathbb{E}\big[f(X_0, \dots, X_m)\big] \qquad \mathbb{P}\text{-fast sicher.}$$

Beweis. Nach Annahme ist der zu X assoziierte Verschiebungsoperator T ergodisch auf dem kanonischen Raum $(E^{\mathbb{N}_0}, \mathcal{E}^{\otimes\mathbb{N}_0}, \mathbb{P}^X)$. Sei $\pi_{[m]} \colon E^{\mathbb{N}_0} \to E^{m+1}$, $x \mapsto (x_0, \dots, x_m)$, die Projektion auf die ersten $m+1$ Koordinaten. Dann gilt

$$\mathbb{E}\big[|f(X_0, \dots, X_m)|\big] = \int |f \circ \pi_{[m]}| \, \mathrm{d}\mathbb{P}^X,$$

also $f \circ \pi_{[m]} \in L^1(\mathbb{P}^X)$. Der Ergodensatz von Birkhoff, Satz 4.12, impliziert, dass

$$\frac{1}{n} \sum_{k=0}^{n-1} (f \circ \pi_{[m]}) \circ T^k \xrightarrow[n\to\infty]{} \int f \circ \pi_{[m]} \, \mathrm{d}\mathbb{P}^X \qquad \mathbb{P}^X\text{-fast sicher.}$$

Wegen $(f \circ \pi_{[m]} \circ T^k)(X) = f(X_k, \dots, X_{k+m})$ folgt

$$\begin{aligned} 1 &= \mathbb{P}^X\Big(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (f \circ \pi_{[m]}) \circ T^k = \int f \circ \pi_{[m]} \, \mathrm{d}\mathbb{P}^X\Big) \\ &= \mathbb{P}\Big(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, \dots, X_{k+m}) = \mathbb{E}\big[f(X_0, \dots, X_m)\big]\Big). \end{aligned}$$ □

Bemerkung. Wir sehen, dass der Ergodensatz von Birkhoff eine Verallgemeinerung des starken Gesetzes der großen Zahlen darstellt, in der die Unabhängigkeit der Summanden durch Stationarität und Ergodizität ersetzt wird. Ist etwa $(X_n)_{n\in\mathbb{N}_0}$ u.i.v. mit $\mathbb{E}[|X_0|] < \infty$, so ist die Folge nach Beispiel 4.8 (i) mischend, nach Satz 4.7 also ergodisch, und Korollar 4.13 mit $m = 0$ und $f = \mathrm{id}$ liefert $\frac{1}{n} \sum_{k=0}^{n-1} X_k \to \mathbb{E}[X_0]$ fast sicher.

Für den Beweis von Satz 4.12 brauchen wir das folgende Resultat. Sei T maßerhaltend, sei $X_0 \in L^1(\mathbb{P})$ reellwertig, und seien

$$X_n \coloneqq X_0 \circ T^n \qquad \text{und} \qquad S_n \coloneqq \sum_{k=0}^{n-1} X_k, \qquad n \in \mathbb{N}_0,$$

insbesondere $S_0 = 0$. Nach Satz 4.3 (i) ist $(X_n)_{n\in\mathbb{N}_0}$ eine reelle stationäre Folge. Weiter sei

$$M_n \coloneqq \max\{0, S_1, \dots, S_n\}, \qquad n \in \mathbb{N}_0.$$

Lemma 4.14 (Lemma von Hopf)**.** *In der obigen Situation gilt für alle $n \in \mathbb{N}_0$*

$$\mathbb{E}\big[X_0 \mathbb{1}_{\{M_n > 0\}}\big] \geq 0.$$

Man kann das Lemma als eine Art „positive Korrelation“ zwischen X_0 und dem Ereignis $\{M_n > 0\}$ lesen.

Beweis. Für $0 \leq k \leq n$ gilt $M_n \circ T \geq S_k \circ T$, denn für $k \geq 1$ ist $S_k \circ T$ eines der Elemente im Maximum, und für $k = 0$ gilt $S_0 \circ T = 0 \leq M_n \circ T$. Also gilt

$$X_0 + M_n \circ T \geq X_0 + S_k \circ T = S_{k+1},$$

das heißt $X_0 \geq S_{k+1} - M_n \circ T$ für $k = 0, \dots, n-1$. Daraus folgt

$$X_0 \geq \max\{S_1, \dots, S_n\} - M_n \circ T. \tag{4.1}$$

Andererseits gilt

$$\{M_n > 0\}^{\mathrm{c}} = \{M_n = 0\} = \{M_n = 0\} \cap \{M_n \circ T \geq 0\} \subset \{M_n - M_n \circ T \leq 0\}, \tag{4.2}$$

wobei wir $M_n \circ T \geq 0$ verwendet haben. Auf $\{M_n > 0\}$ gilt $\max\{S_1, \dots, S_n\} = M_n$, da dort die 0 im Maximum hinzugefügt werden kann. Wegen $|M_n| \leq \sum_{k=0}^{n-1} |X_k|$ ist M_n integrierbar. Mit (4.1), (4.2) und da T maßerhaltend ist, folgt

$$\begin{aligned}\mathbb{E}\big[X_0 \mathbb{1}_{\{M_n>0\}}\big] &\overset{(4.1)}{\geq} \mathbb{E}\big[(\max\{S_1,\dots,S_n\} - M_n \circ T)\mathbb{1}_{\{M_n>0\}}\big] = \mathbb{E}\big[(M_n - M_n \circ T)\mathbb{1}_{\{M_n>0\}}\big] \\ &= \mathbb{E}[M_n - M_n \circ T] - \mathbb{E}\big[(M_n - M_n \circ T)\mathbb{1}_{\{M_n=0\}}\big] \\ &\overset{(4.2)}{\geq} \mathbb{E}[M_n - M_n \circ T] = \mathbb{E}[M_n] - \mathbb{E}[M_n] = 0. \qquad \square\end{aligned}$$

Beweis von Satz 4.12. Wir nehmen ohne Einschränkung an, dass $\mathbb{E}[X_0] = 0$; andernfalls ersetzen wir X_0 durch $X_0 - \mathbb{E}[X_0]$. Mit der Notation von oben genügt es zu zeigen, dass

$$Z \coloneqq \limsup_{n\to\infty} \frac{1}{n} S_n \leq 0 \qquad \mathbb{P}\text{-fast sicher.}$$

Die Anwendung dieser Aussage auf $-X_0$ statt X_0 liefert dann nämlich $\liminf_{n\to\infty}(1/n)S_n \geq 0$ fast sicher, also $\lim_{n\to\infty}(1/n)S_n = 0 = \mathbb{E}[X_0]$ fast sicher.

Seien $\varepsilon > 0$ und $F \coloneqq \{Z > \varepsilon\}$. Wir zeigen $\mathbb{P}(F) = 0$. Wegen $S_n \circ T = S_{n+1} - X_0$ gilt

$$Z \circ T = \limsup_{n\to\infty} \frac{1}{n}(S_{n+1} - X_0) = \limsup_{n\to\infty} \frac{n+1}{n} \frac{S_{n+1}}{n+1} = Z$$

punktweise, da $X_0/n \to 0$; also gilt $T^{-1}(F) = F$ und somit $F \in \mathcal{I}$. Seien

$$X_n^\varepsilon \coloneqq (X_n - \varepsilon)\mathbb{1}_F, \qquad S_n^\varepsilon \coloneqq \sum_{k=0}^{n-1} X_k^\varepsilon, \qquad M_n^\varepsilon \coloneqq \max\{0, S_1^\varepsilon, \dots, S_n^\varepsilon\}, \qquad F_n \coloneqq \{M_n^\varepsilon > 0\}.$$

Dann gilt $F_1 \subset F_2 \subset \cdots$ und

$$\bigcup_{n=1}^{\infty} F_n = \Big\{\sup_k \frac{1}{k} S_k^\varepsilon > 0\Big\} = \Big\{\sup_k \frac{1}{k} S_k > \varepsilon\Big\} \cap F = F,$$

wobei wir im letzten Schritt verwendet haben, dass auf F wegen $\limsup_k (1/k)S_k > \varepsilon$ auch $\sup_k (1/k)S_k > \varepsilon$ gilt. Also gilt $F_n \uparrow F$, und mit dem Satz von Lebesgue über die dominierte Konvergenz, mit der integrierbaren Majorante $|X_0| + \varepsilon$, folgt

$$\lim_{n\to\infty} \mathbb{E}\big[X_0^\varepsilon \mathbb{1}_{F_n}\big] = \mathbb{E}\big[X_0^\varepsilon \mathbb{1}_F\big] = \mathbb{E}\big[X_0^\varepsilon\big].$$

Da F invariant ist, gilt $X_n^\varepsilon = X_0^\varepsilon \circ T^n$ mit $X_0^\varepsilon \in L^1(\mathbb{P})$. Das Lemma von Hopf, Lemma 4.14, angewandt auf $(X_n^\varepsilon)_{n\in\mathbb{N}_0}$, ergibt daher $\mathbb{E}[X_0^\varepsilon \mathbb{1}_{F_n}] \geq 0$ für alle n, also

$$0 \leq \mathbb{E}[X_0^\varepsilon] = \mathbb{E}[(X_0 - \varepsilon)\mathbb{1}_F].$$

Da T ergodisch ist und $F \in \mathcal{I}$, gilt $\mathbb{P}(F) \in \{0, 1\}$. Falls $\mathbb{P}(F) = 1$, so folgt $\mathbb{E}[(X_0 - \varepsilon)\mathbb{1}_F] = \mathbb{E}[X_0 - \varepsilon] = -\varepsilon < 0$, ein Widerspruch. Also gilt $\mathbb{P}(F) = 0$. Da $\varepsilon > 0$ beliebig war, folgt $\mathbb{P}(Z > 0) = \lim_{j\to\infty} \mathbb{P}(Z > 1/j) = 0$, also $Z \leq 0$ fast sicher. $\square$

Wir erwähnen ohne Beweis die folgende Variante im L^p-Sinne. Für $p = 2$ ist dies der Ergodensatz von von Neumann [vN32], der historisch sogar früher gefunden wurde; die Fassung für $p \geq 1$ ist eine spätere Verallgemeinerung, siehe etwa [Kle13, Kapitel 20].

Satz 4.15 (L^p-Ergodensatz)**.** *Seien T eine maßerhaltende und ergodische Transformation und $X_0 \in L^p(\mathbb{P})$ für ein $p \geq 1$. Dann gilt*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X_0 \circ T^k = \mathbb{E}[X_0] \qquad \textit{im } p\textit{-ten Mittel.}$$

5. Bedingte Erwartungswerte und Martingale

Zu diesem Kapitel vergleiche man [Kle13, Kapitel 8 und 9] und [Wil91]; die Martingaltheorie geht im Wesentlichen auf [Doo53] zurück.

5.1. Bedingte Erwartungswerte

Zur Motivation betrachten wir einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Partition $(B_i)_{i\in\mathbb{N}} \subset \mathcal{F}$ von Ω, das heißt $B_i \neq \emptyset$ und $B_i \cap B_j = \emptyset$ für alle $i \neq j$ sowie $\bigcup_{i=1}^{\infty} B_i = \Omega$. Für alle $i \in \mathbb{N}$ mit $\mathbb{P}(B_i) > 0$ definieren wir

$$\mathbb{P}(A|B_i) \coloneqq \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} =: \mathbb{P}_i(A), \qquad A \in \mathcal{F},$$

ein neues, auf B_i konzentriertes Wahrscheinlichkeitsmaß. Sei X eine integrierbare Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Wir sehen

$$\mathbb{E}_i[X] \coloneqq \int X \,\mathrm{d}\mathbb{P}_i = \int_{B_i} X \,\mathrm{d}\mathbb{P}_i + \underbrace{\int_{B_i^c} X \,\mathrm{d}\mathbb{P}_i}_{=0} = \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X \,\mathrm{d}\mathbb{P}.$$

Betrachte nun eine neue Zufallsvariable X_0 mit

$$X_0 \coloneqq \sum_{i=1}^{\infty} \mathbb{E}_i[X]\, \mathbb{1}_{B_i}, \qquad \text{also} \qquad X_0(\omega) = \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X \,\mathrm{d}\mathbb{P} \quad \text{für alle } \omega \in B_i. \tag{5.1}$$

Dabei lassen wir die Summanden mit $\mathbb{P}(B_i) = 0$ weg. Die Zufallsvariable X_0 ist also konstant auf B_i für alle i, und es gilt

$$\int_{B_i} X_0 \,\mathrm{d}\mathbb{P} = \mathbb{E}_i[X]\, \mathbb{P}(B_i) = \int_{B_i} X \,\mathrm{d}\mathbb{P}.$$

Die Zufallsvariable X_0 ersetzt X also auf jeder Menge B_i durch ihren Mittelwert auf B_i, siehe Abbildung 5.1. Dieses Vorgehen kann verallgemeinert werden.

Wir schreiben $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{\pm\infty\}$ für die erweiterten reellen Zahlen mit der Borel-σ-Algebra $\mathcal{B}(\overline{\mathbb{R}})$.

Definition und Satz 5.1 (Existenz der bedingten Erwartung)**.** *Sei $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ eine nichtnegative oder integrierbare Zufallsvariable. Dann existiert zu jeder Unter-σ-Algebra $\mathcal{G} \subset \mathcal{F}$ eine nichtnegative beziehungsweise integrierbare Zufallsvariable X_0, welche $\mathcal{G}$-messbar ist (also $X_0^{-1}(C) \in \mathcal{G}$ für alle $C \in \mathcal{B}(\overline{\mathbb{R}})$) und die folgende Aussage erfüllt:*

$$\int_B X_0 \,\mathrm{d}\mathbb{P} = \int_B X \,\mathrm{d}\mathbb{P} \qquad \textit{für alle } B \in \mathcal{G}. \tag{5.2}$$

Die Zufallsvariable X_0 ist $\mathbb{P}$-fast sicher eindeutig bestimmt. Sie wird als $\mathbb{E}[X|\mathcal{G}]$ geschrieben und heißt bedingte Erwartung *oder* bedingter Erwartungswert von X unter $\mathcal{G}$.

Beispiel 5.2. Wir befinden uns in der Situation von Definition und Satz 5.1.

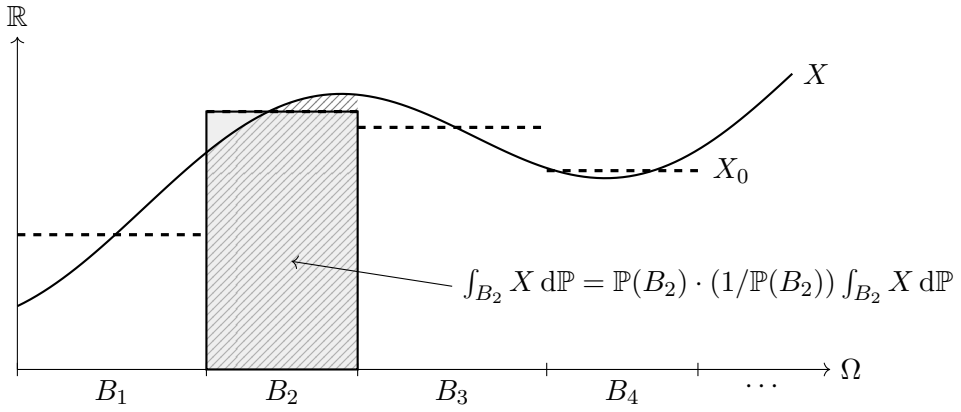

Abbildung 5.1.: Die Zufallsvariable X_0 aus (5.1) (gestrichelt) ist auf jeder Menge B_i der Partition konstant gleich dem Mittelwert von X auf B_i. Die schraffierte Fläche unter X über B_2 und das fett umrandete Rechteck über B_2 mit der Höhe des Mittelwerts haben denselben Inhalt.

(1) Für $\mathcal{G} = \mathcal{F}$ gilt $X_0 = X$ fast sicher.

(2) Für $\mathcal{G} = \{\emptyset, \Omega\}$ gilt $X_0 = \mathbb{E}[X]$, denn

$$\int_B \mathbb{E}[X]\,\mathrm{d}\mathbb{P} = \mathbb{E}[X]\,\mathbb{P}(B) = \begin{cases} 0, & \text{falls } B = \emptyset, \\ \mathbb{E}[X], & \text{falls } B = \Omega, \end{cases}$$

wie in (5.2) verlangt. Außerdem ist jede $\{\emptyset, \Omega\}$-messbare Zufallsvariable konstant.

Beweis von Definition und Satz 5.1. Wir führen den Beweis nur für $X \geq 0$. Für integrierbares X folgt die Existenz dann mit der Zerlegung $X = X^+ - X^-$.

Existenz. Wir nehmen zunächst zusätzlich $\mathbb{E}[X] < \infty$ an. Wir setzen

$$\mathbb{P}_0(B) \coloneqq \mathbb{P}(B) \qquad \text{und} \qquad Q(B) \coloneqq \int_B X\,\mathrm{d}\mathbb{P} \qquad \text{für alle } B \in \mathcal{G}.$$

Dann sind $\mathbb{P}_0$ und Q endliche Maße auf $\mathcal{G}$, und $\mathbb{P}_0$ ist sogar ein Wahrscheinlichkeitsmaß. Dabei gilt

$$\mathbb{P}_0(B) = 0 \implies \mathbb{P}(B) = 0 \implies Q(B) = 0,$$

also $Q \ll \mathbb{P}_0$. Nach dem Satz von Radon–Nikodym existiert also eine Dichte X_0 von Q bezüglich $\mathbb{P}_0$, also $Q(\mathrm{d}\omega) = X_0(\omega)\,\mathbb{P}_0(\mathrm{d}\omega)$, welche $\mathcal{G}$-messbar und nichtnegativ ist, und es gilt

$$\int_B X\,\mathrm{d}\mathbb{P} = Q(B) = \int_B X_0\,\mathrm{d}\mathbb{P}_0 \qquad \text{für alle } B \in \mathcal{G}.$$

Außerdem gilt

$$\int_B f\,\mathrm{d}\mathbb{P}_0 = \int_B f\,\mathrm{d}\mathbb{P} \qquad \text{für alle } \mathcal{G}\text{-messbaren Abbildungen } f \geq 0,$$

wie man mit der Definition des Integrals sieht: Für $f = \mathbb{1}_C$ mit $C \in \mathcal{G}$ gilt dies nach Definition von $\mathbb{P}_0$, für einfache Funktionen $f = \sum_j b_j \mathbb{1}_{C_j}$ folgt es mit der Linearität, und für allgemeine $f \geq 0$ mit dem Satz von der monotonen Konvergenz. Also erfüllt X_0 die Gleichung (5.2).

Ist $\mathbb{E}[X] = \infty$, so wenden wir das Vorige auf $X \wedge n$, $n \in \mathbb{N}$, an. Mit $Y_n \coloneqq \mathbb{E}[X \wedge n | \mathcal{G}]$ gilt $Y_n \leq Y_{n+1}$ fast sicher, wie das Argument im folgenden Eindeutigkeitsteil zeigt, angewandt

auf $\{Y_n > Y_{n+1}\}$. Ändern wir die Y_n auf einer Nullmenge aus $\mathcal{G}$ ab, so ist $(Y_n)_{n\in\mathbb{N}}$ monoton wachsend, und $X_0 \coloneqq \sup_n Y_n$ erfüllt (5.2) nach dem Satz von der monotonen Konvergenz.

Eindeutigkeit. Sei $f \geq 0$ eine weitere $\mathcal{G}$-messbare Abbildung mit (5.2). Für $n \in \mathbb{N}$ setzen wir $A_n \coloneqq \{X_0 > f\} \cap \{f \leq n\} \in \mathcal{G}$. Dann gilt $\int_{A_n} X_0 \,\mathrm{d}\mathbb{P} = \int_{A_n} f \,\mathrm{d}\mathbb{P} \leq n < \infty$ und damit

$$0 = \mathbb{E}[X_0 \mathbb{1}_{A_n}] - \mathbb{E}[f\mathbb{1}_{A_n}] = \mathbb{E}[\underbrace{(X_0 - f)\mathbb{1}_{A_n}}_{\geq 0}].$$

Also gilt $\mathbb{P}(A_n) = 0$ für alle n und wegen $\{X_0 > f\} = \bigcup_n A_n$ auch $\mathbb{P}(X_0 > f) = 0$, das heißt $X_0 \leq f$ $\mathbb{P}$-fast sicher. Analog gilt auch $X_0 \geq f$ $\mathbb{P}$-fast sicher. □

Definition 5.3 (Bedingte Erwartung unter Zufallsvariablen)**.** Seien $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ eine nichtnegative oder integrierbare Zufallsvariable und $Y_i\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\Omega_i, \mathcal{F}_i)$, $i \in I$, Zufallsvariablen. Sei

$$\mathcal{G} \coloneqq \sigma((Y_i)_{i\in I}) \coloneqq \sigma\Big(\bigcup_{i\in I} \{Y_i^{-1}(A_i)\colon A_i \in \mathcal{F}_i\}\Big)$$

die von $(Y_i)_{i\in I}$ erzeugte σ-Algebra. Dann bezeichnet

$$\mathbb{E}[X|(Y_i)_{i\in I}] \coloneqq \mathbb{E}[X|\mathcal{G}]$$

den *bedingten Erwartungswert von X bedingt auf $(Y_i)_{i\in I}$*.

Beispiel 5.4. Die Zufallsvariable $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to ([0,1], \mathcal{B}([0,1]))$ sei die Wahrscheinlichkeit für eine Person, schwer an Covid zu erkranken. Die Zufallsvariable $Y\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ sei das Alter der Person. Wir suchen $\mathbb{E}[X|Y]$. Da

$$\mathcal{G} = \sigma(Y) = \{Y^{-1}(M)\colon M \subset \mathbb{N}_0\} = \sigma(\{Y^{-1}(\{n\})\colon n \in \mathbb{N}_0\}),$$

genügt es, eine auf den Mengen $A_n \coloneqq Y^{-1}(\{n\}) \in \mathcal{F}$, $n \in \mathbb{N}_0$, konstante Zufallsvariable $\mathbb{E}[X|Y]$ zu finden mit

$$\int_{A_n} X \,\mathrm{d}\mathbb{P} = \int_{A_n} \mathbb{E}[X|Y] \,\mathrm{d}\mathbb{P} \qquad \text{für alle } n \in \mathbb{N}_0. \tag{5.3}$$

Jede Menge aus $\mathcal{G}$ ist nämlich eine abzählbare disjunkte Vereinigung von Mengen A_n, sodass (5.2) aus (5.3) mit der σ-Additivität des Integrals folgt. Die Zufallsvariable

$$\mathbb{E}[X|Y](\omega) \coloneqq \frac{\int_{A_n} X \,\mathrm{d}\mathbb{P}}{\mathbb{P}(A_n)} \qquad \text{falls } \omega \in A_n \text{ mit } \mathbb{P}(A_n) > 0,$$

und $\mathbb{E}[X|Y](\omega) \coloneqq 0$ auf den Nullmengen A_n mit $\mathbb{P}(A_n) = 0$, erfüllt dies. Der Wert $\int_{A_n} X \,\mathrm{d}\mathbb{P}/\mathbb{P}(A_n)$ ist das erwartete Risiko einer n Jahre alten Person.

Satz 5.5 (Faktorisierungslemma)**.** *Seien $T\colon (\Omega, \mathcal{F}) \to (\Phi, \mathcal{G})$ und $f\colon (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ messbare Abbildungen. Dann ist f genau dann messbar bezüglich $\sigma(T)$, wenn eine messbare Abbildung $g\colon (\Phi, \mathcal{G}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ existiert mit $f = g \circ T$, siehe Abbildung 5.2.*

Beweis. Ohne Beweis. □

Speziell ist in der Situation von Definition 5.3 die Zufallsvariable $\mathbb{E}[X|(Y_i)_{i\in I}]$ messbar bezüglich $\sigma((Y_i)_{i\in I})$. Wegen des Faktorisierungslemmas, Satz 5.5, angewandt auf $T = (Y_i)_{i\in I}$, existiert eine messbare Abbildung $g\colon (\bigtimes_{i\in I} \Omega_i, \bigotimes_{i\in I} \mathcal{F}_i) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ mit

$$\mathbb{E}[X|(Y_i)_{i\in I}] = g((Y_i)_{i\in I}) \qquad \mathbb{P}\text{-fast sicher.}$$

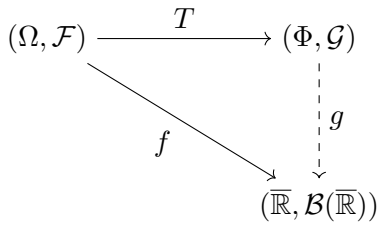

Abbildung 5.2.: Das Faktorisierungslemma: Eine $\sigma(T)$-messbare Abbildung f faktorisiert über T.

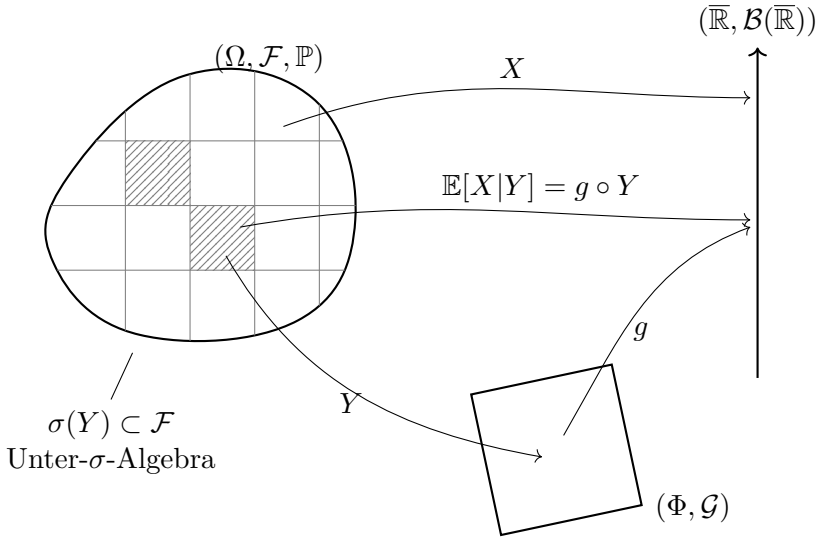

Abbildung 5.3.: Die bedingte Erwartung $\mathbb{E}[X|Y]$ ist $\sigma(Y)$-messbar und faktorisiert daher als $g \circ Y$; sie hängt von ω nur über die Beobachtung $Y(\omega)$ ab.

Insbesondere schreiben wir für eine Zufallsvariable $Y\colon (\Omega, \mathcal{F}) \to (\Phi, \mathcal{G})$ und $y \in \Phi$

$$\mathbb{E}[X|Y = y] \coloneqq g(y)$$

und nennen dies die *bedingte Erwartung von X unter dem Ereignis $\{Y = y\}$*. Man beachte, dass g nur $\mathbb{P}^Y$-fast sicher eindeutig bestimmt ist.

Die bedingte Erwartung $\mathbb{E}[X|Y]$ ist der gemittelte (erwartete) Ausgang von X bedingt auf die (zufällige) Beobachtung Y, siehe Abbildung 5.3. Für alle $A \in \mathcal{G}$ gilt

$$\int_{Y^{-1}(A)} X \,\mathrm{d}\mathbb{P} = \int_{Y^{-1}(A)} \mathbb{E}[X|Y] \,\mathrm{d}\mathbb{P} = \int_{Y^{-1}(A)} g \circ Y \,\mathrm{d}\mathbb{P} = \int_A g \,\mathrm{d}\mathbb{P}^Y.$$

Lemma 5.6 (Eigenschaften der bedingten Erwartung)**.** *Seien X, Y nichtnegative oder integrierbare Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, und seien $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ zwei Unter-σ-Algebren von $\mathcal{F}$. Dann gilt:*

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\big|\mathcal{H}\big] = \mathbb{E}[X|\mathcal{H}] \qquad \textit{(Turmeigenschaft)}, \tag{5.4}$$

$$X \ \mathcal{G}\textit{-messbar} \implies \mathbb{E}[X|\mathcal{G}] = X \quad \mathbb{P}\textit{-f.s.}, \tag{5.5}$$

$$X = Y \quad \mathbb{P}\textit{-f.s.} \implies \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[Y|\mathcal{G}] \quad \mathbb{P}\textit{-f.s.}, \tag{5.6}$$

$$X \equiv \alpha \quad \mathbb{P}\textit{-f.s. (konstant)} \implies \mathbb{E}[X|\mathcal{G}] = \alpha \quad \mathbb{P}\textit{-f.s.} \tag{5.7}$$

Für integrierbare X, Y gilt außerdem:

$$\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\,\mathbb{E}[X|\mathcal{G}] + \beta\,\mathbb{E}[Y|\mathcal{G}] \quad \mathbb{P}\text{-}f.s. \text{ für alle } \alpha, \beta \in \mathbb{R}, \tag{5.8}$$

$$\sigma(X) \text{ und } \mathcal{G} \text{ unabhängig} \implies \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X] \quad \mathbb{P}\text{-}f.s., \tag{5.9}$$

$$X \le Y \quad \mathbb{P}\text{-}f.s. \implies \mathbb{E}[X|\mathcal{G}] \le \mathbb{E}[Y|\mathcal{G}] \quad \mathbb{P}\text{-}f.s., \tag{5.10}$$

$$|\mathbb{E}[X|\mathcal{G}]| \le \mathbb{E}[|X||\mathcal{G}] \quad \mathbb{P}\text{-}f.s. \tag{5.11}$$

Ist $(X_n)_{n\in\mathbb{N}}$ eine monoton wachsende Folge nichtnegativer Zufallsvariablen, so gilt (Satz von der monotonen Konvergenz)

$$\sup_{n\in\mathbb{N}} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}\Big[\sup_{n\in\mathbb{N}} X_n \Big| \mathcal{G}\Big] \quad \mathbb{P}\text{-}f.s. \tag{5.12}$$

Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge mit $X_n \to X$ $\mathbb{P}$-fast sicher und $|X_n| \le Y$ $\mathbb{P}$-fast sicher für alle $n \in \mathbb{N}$, wobei Y integrierbar ist, so gilt (Satz von der majorisierten Konvergenz)

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] \quad \mathbb{P}\text{-}f.s. \tag{5.13}$$

Ist X eine integrierbare Zufallsvariable mit Werten in einem offenen Intervall $U \subset \mathbb{R}$ und $f\colon U \to \mathbb{R}$ konvex mit $f(X)$ integrierbar, so hat $\mathbb{E}[X|\mathcal{G}]$ $\mathbb{P}$-fast sicher Werte in U, und es gilt (Jensensche Ungleichung)

$$f(\mathbb{E}[X|\mathcal{G}]) \le \mathbb{E}[f(X)|\mathcal{G}] \quad \mathbb{P}\text{-}f.s. \tag{5.14}$$

Ist X $\mathcal{G}$-messbar und sind $X, Y \ge 0$, oder gilt $\mathbb{E}[|XY|] < \infty$ und $\mathbb{E}[|Y|] < \infty$ (etwa wenn $\mathbb{E}[|X|^p] < \infty$ und $\mathbb{E}[|Y|^q] < \infty$ für $p, q \in [1, \infty]$ mit $1/p + 1/q = 1$), so gilt

$$\mathbb{E}[XY|\mathcal{G}] = X\,\mathbb{E}[Y|\mathcal{G}] \quad \mathbb{P}\text{-}f.s. \tag{5.15}$$

Beweis. Zu (5.4): Für alle $A \in \mathcal{H}$ gilt wegen $A \in \mathcal{G}$

$$\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]\,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} = \int_A X\,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{H}]\,\mathrm{d}\mathbb{P}.$$

Da $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ $\mathcal{H}$-messbar ist, folgt die Behauptung aus der Eindeutigkeit in Definition und Satz 5.1.

Zu (5.5): Dies ist ein Spezialfall von (5.15) mit $Y = 1$ und (5.7); es folgt aber auch direkt, da X selbst $\mathcal{G}$-messbar ist und (5.2) erfüllt.

Zu (5.6): Es genügt zu zeigen, dass für alle $B \in \mathcal{G}$

$$\int_B \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} = \int_B X\,\mathrm{d}\mathbb{P} = \int_B Y\,\mathrm{d}\mathbb{P} = \int_B \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P}.$$

Zu (5.7): Für alle $B \in \mathcal{G}$ gilt

$$\int_B \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} = \int_B X\,\mathrm{d}\mathbb{P} = \int_B \alpha\,\mathrm{d}\mathbb{P},$$

und die Konstante α ist $\mathcal{G}$-messbar.

Zu (5.8): Für alle $B \in \mathcal{G}$ gilt

$$\begin{aligned}\int_B \mathbb{E}[\alpha X + \beta Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} &= \int_B (\alpha X + \beta Y)\,\mathrm{d}\mathbb{P} = \alpha \int_B X\,\mathrm{d}\mathbb{P} + \beta \int_B Y\,\mathrm{d}\mathbb{P} \\ &= \int_B (\alpha\,\mathbb{E}[X|\mathcal{G}] + \beta\,\mathbb{E}[Y|\mathcal{G}])\,\mathrm{d}\mathbb{P},\end{aligned}$$

und die rechte Seite ist $\mathcal{G}$-messbar.

Zu (5.9): Für alle $A \in \mathcal{G}$ gilt wegen der Unabhängigkeit von X und $\mathbb{1}_A$

$$\int_A \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X]\,\mathbb{P}(A) = \int_A \mathbb{E}[X]\,\mathrm{d}\mathbb{P}.$$

Zu (5.10): Sei $A \coloneqq \{\mathbb{E}[X|\mathcal{G}] > \mathbb{E}[Y|\mathcal{G}]\} \in \mathcal{G}$. Dann gilt

$$0 \le \int_A (\mathbb{E}[X|\mathcal{G}] - \mathbb{E}[Y|\mathcal{G}])\,\mathrm{d}\mathbb{P} = \int_A \underbrace{(X-Y)}_{\le 0}\,\mathrm{d}\mathbb{P} \le 0.$$

Da der Integrand links auf A strikt positiv ist, folgt $\mathbb{P}(A) = 0$.

Zu (5.11): Dies folgt aus (5.10) und (5.8) mit $X = X^+ - X^-$, wobei $X^+, X^- \ge 0$, denn

$$\big|\mathbb{E}[X|\mathcal{G}]\big| = \big|\mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}]\big| \le \mathbb{E}[X^+|\mathcal{G}] + \mathbb{E}[X^-|\mathcal{G}] = \mathbb{E}\big[|X||\mathcal{G}\big].$$

Die Aussagen (5.12), (5.13) und (5.14) sind eine Übungsaufgabe.

Zu (5.15): Seien zunächst $X, Y \ge 0$. Wir setzen $X_n \coloneqq 2^{-n}\lfloor 2^n X\rfloor$. Dann ist X_n $\mathcal{G}$-messbar mit Werten in $2^{-n}\mathbb{N}_0 \cup \{\infty\}$, und es gilt $X_n \uparrow X$ sowie $X_n\mathbb{E}[Y|\mathcal{G}] \uparrow X\mathbb{E}[Y|\mathcal{G}]$. Mit dem Satz von der monotonen Konvergenz folgt für alle $A \in \mathcal{G}$

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

Andererseits gilt, da $A \cap \{X_n = k2^{-n}\} \in \mathcal{G}$ für alle k,

$$\begin{aligned}\mathbb{E}\big[\mathbb{1}_A X_n \mathbb{E}[Y|\mathcal{G}]\big] &= \sum_{k=1}^{\infty} \mathbb{E}\big[\mathbb{1}_A \mathbb{1}_{\{X_n = k2^{-n}\}} k2^{-n}\mathbb{E}[Y|\mathcal{G}]\big] = \sum_{k=1}^{\infty} \mathbb{E}\big[\mathbb{1}_A \mathbb{1}_{\{X_n = k2^{-n}\}} k2^{-n} Y\big] \\ &= \mathbb{E}\big[\mathbb{1}_A X_n Y\big] \xrightarrow[n\to\infty]{} \mathbb{E}\big[\mathbb{1}_A XY\big].\end{aligned}$$

Dabei haben wir den Summanden zu $\{X_n = \infty\} = \{X = \infty\}$ noch nicht berücksichtigt. Mit der Konvention $0 \cdot \infty = 0$ gilt für $B \coloneqq A \cap \{X = \infty\} \in \mathcal{G}$ und jede Zufallsvariable $Z \ge 0$, dass $\mathbb{E}[\mathbb{1}_B \cdot \infty \cdot Z]$ gleich ∞ ist, falls $\mathbb{E}[\mathbb{1}_B Z] > 0$, und sonst gleich 0. Wegen $\mathbb{E}[\mathbb{1}_B \mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[\mathbb{1}_B Y]$ stimmen daher die Summanden für $Z = \mathbb{E}[Y|\mathcal{G}]$ und für $Z = Y$ überein, und die Rechnung gilt auch, wenn X den Wert ∞ annimmt. Also gilt $\mathbb{E}[\mathbb{1}_A X\mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[\mathbb{1}_A XY]$ für alle $A \in \mathcal{G}$, und da $X\mathbb{E}[Y|\mathcal{G}]$ $\mathcal{G}$-messbar ist, folgt die Behauptung. Im allgemeinen Fall nutzen wir die Zerlegungen $X = X^+ - X^-$, $Y = Y^+ - Y^-$ und die Linearität. □

Definition 5.7 (Bedingte Wahrscheinlichkeit)**.** Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{G} \subset \mathcal{F}$ eine Unter-σ-Algebra und $A \in \mathcal{F}$. Dann heißt

$$\mathbb{P}(A|\mathcal{G}) \coloneqq \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$$

die *bedingte Wahrscheinlichkeit von A unter $\mathcal{G}$*.

Speziell für $\mathcal{G} = \{\emptyset, B, B^{\mathrm{c}}, \Omega\}$ mit $\mathbb{P}(B) \in (0,1)$ ergibt sich

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}] = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}\mathbb{1}_B + \frac{\mathbb{P}(A\cap B^{\mathrm{c}})}{\mathbb{P}(B^{\mathrm{c}})}\mathbb{1}_{B^{\mathrm{c}}} = \mathbb{P}(A|B)\mathbb{1}_B + \mathbb{P}(A|B^{\mathrm{c}})\mathbb{1}_{B^{\mathrm{c}}},$$

denn die rechte Seite ist $\mathcal{G}$-messbar, und für $C \in \mathcal{G}$ gilt

$$\begin{aligned}\int_C \big(\mathbb{P}(A|B)\mathbb{1}_B + \mathbb{P}(A|B^{\mathrm{c}})\mathbb{1}_{B^{\mathrm{c}}}\big)\,\mathrm{d}\mathbb{P} &= \mathbb{P}(A|B)\,\mathbb{P}(C\cap B) + \mathbb{P}(A|B^{\mathrm{c}})\,\mathbb{P}(C\cap B^{\mathrm{c}}) \\ &= \begin{cases} 0, & C = \emptyset, \\ \mathbb{P}(A\cap B), & C = B, \\ \mathbb{P}(A\cap B^{\mathrm{c}}), & C = B^{\mathrm{c}}, \\ \mathbb{P}(A), & C = \Omega, \end{cases} \qquad = \mathbb{P}(A\cap C) = \int_C \mathbb{1}_A\,\mathrm{d}\mathbb{P}.\end{aligned}$$

Die bedingte Wahrscheinlichkeit unter $\mathcal{G}$ verallgemeinert also die elementare bedingte Wahrscheinlichkeit aus Definition 0.15.

Satz 5.8 (Variationelle Charakterisierung der bedingten Erwartung)**.** *Seien $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, das heißt $\mathbb{E}[X^2] < \infty$, und $\mathcal{G} \subset \mathcal{F}$ eine Unter-σ-Algebra. Dann ist $\mathbb{E}[X|\mathcal{G}]$ bis auf $\mathbb{P}$-fast sichere Gleichheit die einzige $\mathcal{G}$-messbare Zufallsvariable $X_0 \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, für die der Abstand*

$$\mathbb{E}\big[(X - X_0)^2\big]$$

minimal wird.

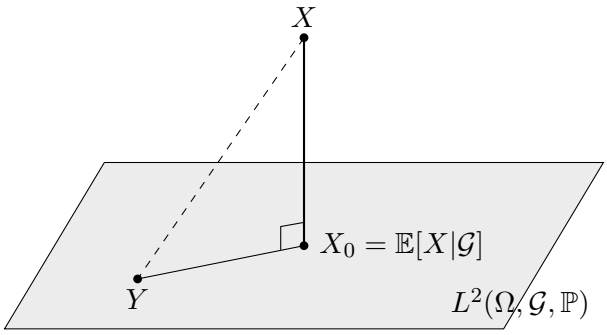

Abbildung 5.4.: Die bedingte Erwartung $\mathbb{E}[X|\mathcal{G}]$ als orthogonale Projektion von X auf den Unterraum der $\mathcal{G}$-messbaren quadratintegrierbaren Zufallsvariablen. Für jedes andere $\mathcal{G}$-messbare Y gilt nach Pythagoras $\mathbb{E}[(X - Y)^2] = \mathbb{E}[(X - X_0)^2] + \mathbb{E}[(Y - X_0)^2]$.

Beweis. Sei $X_0 := \mathbb{E}[X|\mathcal{G}]$. Nach der Jensenschen Ungleichung (5.14) mit $f(x) = x^2$ gilt $X_0^2 \le \mathbb{E}[X^2|\mathcal{G}]$, also $X_0 \in L^2$. Sei $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ $\mathcal{G}$-messbar. Nach der Cauchy–Schwarz-Ungleichung ist XY integrierbar, und mit (5.4) für $\mathcal{H} = \{\emptyset, \Omega\}$ sowie (5.15) gilt

$$\mathbb{E}[XY] = \mathbb{E}\big[\mathbb{E}[XY|\mathcal{G}]\big] = \mathbb{E}\big[Y\,\mathbb{E}[X|\mathcal{G}]\big] = \mathbb{E}[X_0 Y]. \tag{5.16}$$

Damit folgt, unter Verwendung von (5.16) für Y und für X_0 an Stelle von Y,

$$\begin{aligned}\mathbb{E}\big[(X - Y)^2\big] - \mathbb{E}\big[(X - X_0)^2\big] &= \mathbb{E}[X^2] - 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X^2] + 2\mathbb{E}[XX_0] - \mathbb{E}[X_0^2] \\ &\overset{(5.16)}{=} \mathbb{E}[Y^2] - 2\mathbb{E}[X_0 Y] + \mathbb{E}[X_0^2] = \mathbb{E}\big[(Y - X_0)^2\big] \ge 0.\end{aligned}$$

Also gilt $\mathbb{E}[(X - Y)^2] \ge \mathbb{E}[(X - X_0)^2]$ für alle $\mathcal{G}$-messbaren $Y \in L^2$, das heißt, X_0 ist minimal, siehe Abbildung 5.4. Falls $\mathbb{E}[(X - X_0)^2] = \mathbb{E}[(X - Y)^2]$, so folgt $\mathbb{E}[(Y - X_0)^2] = 0$, also $Y = X_0$ $\mathbb{P}$-fast sicher. Dies ist die gewünschte Eindeutigkeit. □

Bemerkung. Die bedingte Erwartung $\mathbb{E}[X|\mathcal{G}]$ ist die beste $\mathcal{G}$-messbare Vorhersage für X im quadratischen Mittel.

5.2. Filtrationen und stochastische Folgen

Definition 5.9 (Filtration, stochastische Folgen, Vorhersehbarkeit)**.** Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine Folge von σ-Algebren mit $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$. Dann nennt man $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine *Filtration*. Eine Folge $(X_n)_{n\in\mathbb{N}_0}$ von $\overline{\mathbb{R}}$-wertigen Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit der Eigenschaft, dass X_n messbar bezüglich $\mathcal{F}_n$ ist für alle $n \in \mathbb{N}_0$, heißt *stochastische Folge*, und auch $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ wird stochastische Folge genannt; man sagt auch, $(X_n)_{n\in\mathbb{N}_0}$ ist *adaptiert* an $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Gilt $X_0 = 0$ und $X_n \le X_{n+1}$ $\mathbb{P}$-f.s. für alle $n \in \mathbb{N}_0$, so heißt die stochastische Folge *wachsend*. Ist X_n sogar bezüglich $\mathcal{F}_{n-1}$ messbar für alle $n \in \mathbb{N}$, so heißt die stochastische Folge *vorhersehbar*.

5.3. Martingale

Definition 5.10 (Martingal)**.** Eine stochastische Folge $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ auf $(\Omega, \mathcal{F}, \mathbb{P})$ heißt *Martingal*, falls gilt:

$$\mathbb{E}\big[|X_n|\big] < \infty \qquad \text{für alle } n \in \mathbb{N}_0, \tag{5.17}$$

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n \quad \mathbb{P}\text{-f.s.} \qquad \text{für alle } n \in \mathbb{N}_0. \tag{5.18}$$

Die stochastische Folge heißt *Submartingal*, falls (5.17) gilt und

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n \quad \mathbb{P}\text{-f.s.} \qquad \text{für alle } n \in \mathbb{N}_0,$$

und *Supermartingal*, falls (5.17) gilt und

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n \quad \mathbb{P}\text{-f.s.} \qquad \text{für alle } n \in \mathbb{N}_0.$$

Bemerkung. Durch Bilden des Erwartungswerts in (5.18) sehen wir: Die Abbildung $n \mapsto \mathbb{E}[X_n]$ ist für Martingale konstant, für Submartingale wachsend und für Supermartingale fallend.

Beispiel 5.11.

(i) *Irrfahrt.* Seien $(X_n)_{n\in\mathbb{N}_0}$ stochastisch unabhängige, integrierbare Zufallsvariablen mit $\mathbb{E}[X_n] = 0$ für alle $n \in \mathbb{N}_0$. Dann ist

$$\Big(S_n \coloneqq \sum_{i=0}^{n} X_i,\ \mathcal{F}_n \coloneqq \sigma(X_0, \ldots, X_n)\Big)_{n\in\mathbb{N}_0}$$

ein Martingal. Es gilt nämlich $\mathbb{E}[|S_n|] \leq \sum_{i=0}^{n} \mathbb{E}[|X_i|] < \infty$, und mit (5.5) und (5.9) folgt

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \sum_{i=0}^{n+1} \mathbb{E}[X_i|\mathcal{F}_n] = \sum_{i=0}^{n} X_i + \mathbb{E}[X_{n+1}] = S_n.$$

(ii) *Produkte.* Sind die X_i wie in (i) stochastisch unabhängig, aber mit $\mathbb{E}[X_i] = 1$ und $\mathbb{E}[|X_i|] < \infty$, so ist

$$\Big(P_n \coloneqq \prod_{i=0}^{n} X_i,\ \mathcal{F}_n = \sigma(X_0, \ldots, X_n)\Big)_{n\in\mathbb{N}_0}$$

ein Martingal, denn wegen der Unabhängigkeit gilt $\mathbb{E}[|P_n|] = \prod_{i=0}^{n} \mathbb{E}[|X_i|] < \infty$, und mit (5.15) folgt

$$\mathbb{E}[P_{n+1}|\mathcal{F}_n] = \mathbb{E}\Big[\prod_{i=0}^{n+1} X_i \Big| \mathcal{F}_n\Big] = \prod_{i=0}^{n} X_i \cdot \mathbb{E}[X_{n+1}] = P_n.$$

(iii) *Summen nichtnegativer Zufallsvariablen.* Sind $X_n \geq 0$ für alle $n \in \mathbb{N}_0$ beliebige integrierbare Zufallsvariablen, so ist $S_n \coloneqq \sum_{i=0}^{n} X_i$ ein Submartingal bezüglich $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, denn

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \sum_{i=0}^{n} X_i + \underbrace{\mathbb{E}[X_{n+1}|\mathcal{F}_n]}_{\geq 0} \geq S_n.$$

(iv) *Doob-Martingal.* Ist X eine Zufallsvariable mit $\mathbb{E}[|X|] < \infty$ und $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$ eine Filtration, so ist $(\mathbb{E}[X|\mathcal{F}_n], \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Martingal. Dies folgt aus der Turmeigenschaft (5.4).

(v) *Konvexe Funktionen von Martingalen.* Ist $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Martingal und $g\colon \mathbb{R} \to \mathbb{R}$ konvex mit $\mathbb{E}[|g(X_n)|] < \infty$ für alle $n \in \mathbb{N}_0$, so ist $(g(X_n), \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Submartingal. Mit der Jensenschen Ungleichung (5.14) gilt nämlich $\mathbb{E}[g(X_{n+1})|\mathcal{F}_n] \geq g(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) = g(X_n)$.

Bemerkung 5.12.

(i) Für Zufallsvariablen $(X_n)_{n\in\mathbb{N}_0}$ heißt

$$\mathcal{F}_n \coloneqq \sigma(X_0, \ldots, X_n), \qquad n \in \mathbb{N}_0,$$

die *kanonische Filtration.* Insbesondere ist $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ dann stets eine stochastische Folge.

(ii) Mit Beispiel 5.11 (v) ist $(S_n^2)_{n\in\mathbb{N}_0}$ mit $(S_n)_{n\in\mathbb{N}_0}$ aus Beispiel 5.11 (i) ein Submartingal, sofern zusätzlich $\mathbb{E}[X_i^2] < \infty$ für alle $i \in \mathbb{N}_0$ gilt.

(iii) Bedingte Erwartungen haben wir bisher für nichtnegative oder integrierbare Zufallsvariablen erklärt. Sei nun X eine beliebige $\overline{\mathbb{R}}$-wertige Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$. Für eine Unter-$\sigma$-Algebra $\mathcal{G} \subset \mathcal{F}$ existieren stets $\mathbb{E}[X^+|\mathcal{G}], \mathbb{E}[X^-|\mathcal{G}] \in [0, \infty]$, wobei $X^+ = \max\{0, X\}$ und $X^- = -\min\{0, X\}$. Gilt für

$$A \coloneqq \{\omega \in \Omega\colon \mathbb{E}[X^+|\mathcal{G}](\omega) < \infty\} \quad \text{und} \quad B \coloneqq \{\omega \in \Omega\colon \mathbb{E}[X^-|\mathcal{G}](\omega) < \infty\},$$

dass $\mathbb{P}(A \cap B) = 1$, so ist

$$\mathbb{E}[X|\mathcal{G}] \coloneqq \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}] \quad \mathbb{P}\text{-f.s.}$$

wohldefiniert.

(iv) Seien $(X_n)_{n\in\mathbb{N}_0}$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine Filtration, an die (X_n) adaptiert ist. Wir nennen $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein *verallgemeinertes Martingal*, wenn $\mathbb{E}[X_{n+1}|\mathcal{F}_n]$ für alle $n \in \mathbb{N}_0$ im Sinne von (iii) $\mathbb{P}$-f.s. wohldefiniert ist und

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n \quad \mathbb{P}\text{-f.s.} \qquad \text{für alle } n \in \mathbb{N}_0.$$

Satz 5.13 (Weitere Eigenschaften von Martingalen)**.**

(i) Die Folge $(X_n)_{n\in\mathbb{N}_0}$ ist genau dann ein Sub- (Super-)Martingal, wenn $(-X_n)_{n\in\mathbb{N}_0}$ ein Super- (Sub-)Martingal ist.

(ii) Sind $(X_n)_{n\in\mathbb{N}_0}$ und $(Y_n)_{n\in\mathbb{N}_0}$ (Sub-, Super-)Martingale, so ist $(aX_n + bY_n)_{n\in\mathbb{N}_0}$ für alle $a, b \geq 0$ ein (Sub-, Super-)Martingal.

(iii) Sind $(X_n)_{n\in\mathbb{N}_0}$ und $(Y_n)_{n\in\mathbb{N}_0}$ Supermartingale, so ist auch $Z_n \coloneqq X_n \wedge Y_n$ ein Supermartingal.

(iv) Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal mit $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$ für ein $T \in \mathbb{N}_0$. Dann ist $(X_n)_{n=0,\ldots,T}$ ein Martingal. Gibt es sogar eine Folge $T_k \to \infty$ mit $\mathbb{E}[X_{T_k}] \geq \mathbb{E}[X_0]$ für alle k, so ist $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal.

Bemerkung. Mit (i) gelten die entsprechenden Aussagen (iii) und (iv) auch für Submartingale, mit $X_n \vee Y_n$ statt $X_n \wedge Y_n$ und $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ statt $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$.

Beweis. Die Aussagen (i) und (ii) folgen direkt aus der Linearität (5.8) der bedingten Erwartung.

Zu (iii): Wegen $|Z_n| \le |X_n| + |Y_n|$ gilt $\mathbb{E}[|Z_n|] < \infty$ für alle n, und Z_n ist $\mathcal{F}_n$-messbar. Mit der Monotonie (5.10) der bedingten Erwartung und der Supermartingaleigenschaft von (X_n) gilt

$$\mathbb{E}[Z_n|\mathcal{F}_{n-1}] \le \mathbb{E}[X_n|\mathcal{F}_{n-1}] \le X_{n-1},$$

und ebenso $\mathbb{E}[Z_n|\mathcal{F}_{n-1}] \le Y_{n-1}$. Also folgt $\mathbb{E}[Z_n|\mathcal{F}_{n-1}] \le \min\{X_{n-1}, Y_{n-1}\} = Z_{n-1}$.

Zu (iv): Für $n \le T$ setzen wir $Y_n := \mathbb{E}[X_T|\mathcal{F}_n]$. Dann ist $(Y_n)_{n=0,\dots,T}$ nach Beispiel 5.11 (iv) ein Martingal, und wegen der Supermartingaleigenschaft (iteriert mit der Turmeigenschaft) gilt $Y_n \le X_n$. Also folgt

$$\mathbb{E}[X_0] \le \mathbb{E}[X_T] = \mathbb{E}[Y_T] = \mathbb{E}[Y_n] \le \mathbb{E}[X_n] \le \mathbb{E}[X_0].$$

Daher gilt $\mathbb{E}[X_n - Y_n] = 0$ mit $X_n - Y_n \ge 0$, also $Y_n = X_n$ $\mathbb{P}$-f.s., und $(X_n)_{n=0,\dots,T}$ ist ein Martingal. Sei nun $T_k \uparrow \infty$. Dann existiert zu jedem $n \in \mathbb{N}$ ein k mit $T_k \ge n$, und nach dem ersten Teil gilt $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = X_{n-1}$. Also ist $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal. □

5.4. Diskrete stochastische Integrale

Definition 5.14 (Diskrete stochastische Integrale)**.** Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine stochastische Folge und $(H_n)_{n\in\mathbb{N}}$ vorhersehbar bezüglich $(\mathcal{F}_n)$. Wir definieren den stochastischen Prozess

$$(H \cdot X)_n := \sum_{m=1}^{n} H_m(X_m - X_{m-1}), \qquad n \in \mathbb{N}_0,$$

und nennen $((H \cdot X)_n)_{n\in\mathbb{N}_0}$ das *diskrete stochastische Integral* von H bezüglich (X_n). Ist (X_n) ein Martingal, so wird $((H \cdot X)_n)_{n\in\mathbb{N}_0}$ auch *Martingaltransformierte* von X genannt.

Bemerkung.

(i) Die Folge $((H \cdot X)_n)_{n\in\mathbb{N}_0}$ ist an $(\mathcal{F}_n)$ adaptiert.

(ii) Wir können das Integral als *Spiel* mit Spielstrategie $(H_n)_{n\in\mathbb{N}}$ interpretieren. Dabei repräsentiert $X_n - X_{n-1}$ den (zufälligen, eventuell fairen) Spielgewinn pro Einsatz in der n-ten Runde, und H_n ist der Einsatz in Runde n. Dieser muss mit der Information der Runden $0, \dots, n-1$ erfolgen und nicht mit der Information der Runde n; genau das drückt die Vorhersehbarkeit aus.

(iii) Ist (X_n) ein faires Spiel, also ein Martingal, und ist (H_n) *lokal beschränkt*, das heißt, es gilt $|H_n| \le K_n < \infty$ für Konstanten K_n und alle $n \in \mathbb{N}$, so ist $((H \cdot X)_n)_{n\in\mathbb{N}_0}$ ein Martingal. Wegen $|(H \cdot X)_n| \le \sum_{m=1}^{n} K_m(|X_m| + |X_{m-1}|)$ ist $(H \cdot X)_n$ integrierbar, und mit (5.15) gilt

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

Satz 5.15 (Stabilitätssatz für stochastische Integrale)**.** *Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine reellwertige stochastische Folge mit $\mathbb{E}[|X_0|] < \infty$.*

(i) Die Folge (X_n) ist genau dann ein Martingal, wenn für alle lokal beschränkten, vorhersehbaren Prozesse (H_n) das stochastische Integral $((H \cdot X)_n)$ ein Martingal ist.

(ii) Die Folge (X_n) ist genau dann ein Sub- (Super-)Martingal, wenn $((H \cdot X)_n)$ für alle lokal beschränkten, vorhersehbaren Prozesse (H_n) mit $H_n \geq 0$ für alle n ein Sub- (Super-)Martingal ist.

Beweis. Zu (i): Die Richtung „⇒“ haben wir in der vorangehenden Bemerkung gezeigt. Zu „⇐“: Wir wählen $n_0 \in \mathbb{N}$ und $H_n := \mathbb{1}\{n = n_0\}$. Dann ist $(H \cdot X)_{n_0-1} = 0$ und $(H \cdot X)_{n_0} = X_{n_0} - X_{n_0-1}$. Da $(H \cdot X)_{n_0}$ integrierbar ist, folgt induktiv aus $\mathbb{E}[|X_0|] < \infty$, dass $\mathbb{E}[|X_n|] < \infty$ für alle $n \in \mathbb{N}_0$. Weiter gilt

$$0 = \mathbb{E}\big[(H \cdot X)_{n_0} | \mathcal{F}_{n_0-1}\big] = \mathbb{E}[X_{n_0} | \mathcal{F}_{n_0-1}] - X_{n_0-1}.$$

Da n_0 beliebig war, ist (X_n) ein Martingal.

Aussage (ii) zeigen wir analog zu (i); für „⇒“ verwenden wir $H_{n+1} \geq 0$ in der vorangehenden Rechnung. □

Beispiel 5.16 (Petersburger Spiel)**.** Wir betrachten die symmetrische Irrfahrt $X_n = \sum_{i=1}^n D_i$, $X_0 = 0$, mit unabhängigen, auf $\{-1, 1\}$ gleichverteilten Zuwächsen D_i, das heißt $\mathbb{P}(D_i = 1) = \mathbb{P}(D_i = -1) = 1/2$, sowie die Strategie

$$H_1 := 1, \qquad H_n := 2^{n-1}\mathbb{1}\{D_1 = \cdots = D_{n-1} = -1\}, \quad n \geq 2.$$

Wir verdoppeln also den Einsatz, solange wir verlieren, und hören nach dem ersten Gewinn auf. Die Strategie (H_n) ist vorhersehbar und lokal beschränkt. Dann ist

$$S_n = \sum_{i=1}^n H_i D_i = (H \cdot X)_n$$

der Gewinn bis zur Runde n, und (S_n) ist nach Satz 5.15 ein Martingal. Insbesondere gilt $\mathbb{E}[S_n] = \mathbb{E}[S_0] = 0$ für alle $n \in \mathbb{N}_0$. Andererseits gilt $S_n = 1$, sobald das erste Mal $D_i = 1$ eintritt, denn $2^{n-1} - \sum_{i=1}^{n-1} 2^{i-1} = 1$. Da dies fast sicher irgendwann geschieht, folgt $S_n \to 1$ $\mathbb{P}$-f.s., obwohl $\mathbb{E}[S_n] = 0$ für alle n gilt. Auf $\{D_1 = \cdots = D_n = -1\}$, einem Ereignis der Wahrscheinlichkeit 2^{-n}, ist dagegen $S_n = -(2^n - 1)$.

5.5. Doob-Zerlegung und quadratische Variation

Satz 5.17 (Doob-Zerlegung)**.** *Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine adaptierte, integrierbare stochastische Folge. Dann existieren ein Martingal $(M_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ und eine vorhersehbare Folge $(A_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$, sodass*

$$X_n = M_n + A_n \quad \mathbb{P}\text{-f.s.} \qquad \text{für alle } n \in \mathbb{N}_0.$$

Mit der Festlegung $A_0 = 0$ ist diese sogenannte Doob-Zerlegung *eindeutig bestimmt; ihr Analogon in stetiger Zeit ist die* Doob–Meyer-Zerlegung*. Die Folge (A_n) heißt* Kompensator *von (X_n). Weiter ist (A_n) genau dann nicht-fallend, wenn (X_n) ein Submartingal ist.*

Beweis. Wir setzen $M_0 := X_0$, $A_0 := 0$ und für $n \in \mathbb{N}$

$$M_n := M_0 + \sum_{j=0}^{n-1} (X_{j+1} - \mathbb{E}[X_{j+1}|\mathcal{F}_j]), \qquad A_n := \sum_{j=0}^{n-1} (\mathbb{E}[X_{j+1}|\mathcal{F}_j] - X_j).$$

Dann ist $(M_n, \mathcal{F}_n)$ ein Martingal, denn M_n ist integrierbar und $\mathcal{F}_n$-messbar, und es gilt

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = M_{n-1} + \underbrace{\mathbb{E}[X_n - \mathbb{E}[X_n|\mathcal{F}_{n-1}]|\mathcal{F}_{n-1}]}_{=0} = M_{n-1}.$$

Die Folge $(A_n, \mathcal{F}_n)$ ist vorhersehbar, da A_n nur von $\mathcal{F}_{n-1}$-messbaren Größen abhängt, und

$$A_{n+1} - A_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n$$

ist genau dann für alle n nichtnegativ, wenn (X_n) ein Submartingal ist. Außerdem gilt $X_n = M_n + A_n$, da sich die Summen teleskopartig aufheben.

Zur Eindeutigkeit sei $X_n = M'_n + A'_n$ eine weitere solche Zerlegung mit $A'_0 = 0$. Dann gilt

$$A'_{n+1} - A'_n = (A_{n+1} - A_n) + (M_{n+1} - M_n) - (M'_{n+1} - M'_n),$$

und da $A'_{n+1} - A'_n$ und $A_{n+1} - A_n$ beide $\mathcal{F}_n$-messbar sind, folgt durch Bilden der bedingten Erwartung

$$A'_{n+1} - A'_n = \mathbb{E}[A'_{n+1} - A'_n|\mathcal{F}_n] = A_{n+1} - A_n \quad \mathbb{P}\text{-f.s.}$$

Wegen $A_0 = A'_0 = 0$ folgt $A'_n = A_n$ für alle $n \in \mathbb{N}_0$, und daraus $M_n = M'_n$ $\mathbb{P}$-f.s. □

Definition 5.18 (Quadratische Variation)**.** Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein quadratintegrierbares Martingal, das heißt, es gilt $\mathbb{E}[X_n^2] < \infty$ für alle $n \in \mathbb{N}_0$. Die eindeutige vorhersehbare stochastische Folge $(A_n)_{n\in\mathbb{N}_0}$ mit $A_0 = 0$, für die $(X_n^2 - A_n)_{n\in\mathbb{N}_0}$ ein Martingal ist, heißt *quadratischer Variationsprozess* von $(X_n)_{n\in\mathbb{N}_0}$ und wird mit $(\langle X\rangle_n)_{n\in\mathbb{N}_0}$ bezeichnet.

Bemerkung.

(1) Nach dem Beweis von Satz 5.17, angewandt auf (X_n^2), gilt

$$\begin{aligned}\langle X\rangle_n &= \sum_{i=1}^n (\mathbb{E}[X_i^2|\mathcal{F}_{i-1}] - X_{i-1}^2) = \sum_{i=1}^n \mathbb{E}[X_i^2 - X_{i-1}^2|\mathcal{F}_{i-1}] \\ &= \sum_{i=1}^n \Big(\mathbb{E}\big[(X_i - X_{i-1})^2|\mathcal{F}_{i-1}\big] - 2X_{i-1}^2 + \underbrace{2\mathbb{E}[X_{i-1}X_i|\mathcal{F}_{i-1}]}_{=2X_{i-1}^2}\Big) \\ &= \sum_{i=1}^n \mathbb{E}\big[(X_i - X_{i-1})^2|\mathcal{F}_{i-1}\big],\end{aligned}$$

und

$$\mathbb{E}[\langle X\rangle_n] = \mathbb{E}[X_n^2] - \mathbb{E}[X_0^2] = \mathrm{Var}[X_n - X_0].$$

(2) Da (X_n^2) nach Beispiel 5.11 (v) ein Submartingal ist, ist $(\langle X\rangle_n)_{n\in\mathbb{N}_0}$ nach Satz 5.17 nicht-fallend.

Beispiel 5.19 (Irrfahrt)**.** Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängige, quadratintegrierbare Zufallsvariablen mit $\mathbb{E}[X_n] = 0$ für alle n. Dann ist $S_n := \sum_{i=1}^n X_i$, $S_0 = 0$, ein quadratintegrierbares Martingal bezüglich der kanonischen Filtration, denn mit der Cauchy-Schwarz-Ungleichung gilt

$$\mathbb{E}[S_n^2] = \sum_{i,j=1}^n \mathbb{E}[X_iX_j] \le \sum_{i,j=1}^n \sqrt{\mathbb{E}[X_i^2]\,\mathbb{E}[X_j^2]} < \infty.$$

Weiter gilt mit der Unabhängigkeit (5.9)

$$\langle S\rangle_n = \sum_{i=1}^n \mathbb{E}\big[(S_i - S_{i-1})^2|\mathcal{F}_{i-1}\big] = \sum_{i=1}^n \mathbb{E}[X_i^2|\mathcal{F}_{i-1}] = \sum_{i=1}^n \mathbb{E}[X_i^2].$$

Beispiel 5.20 (Diskrete Itô-Formel)**.** Sei $X = (X_n)_{n\in\mathbb{N}_0}$ ein Martingal mit $|X_n - X_{n-1}| = 1$ $\mathbb{P}$-f.s. für alle $n \in \mathbb{N}$ und $X_0 \equiv x_0 \in \mathbb{Z}$. Sei $f\colon \mathbb{Z} \to \mathbb{R}$ beliebig. Dann ist $(f(X_n))_{n\in\mathbb{N}_0}$ ein integrierbarer, adaptierter Prozess, denn

$$|f(X_n)| \le \max\bigl\{|f(x)|\colon x \in \{x_0 - n, \dots, x_0 + n\}\bigr\}.$$

Wir bestimmen die Doob-Zerlegung von $(f(X_n))$. Dazu betrachten wir die diskreten Ableitungen

$$f'(x) \coloneqq \frac{f(x+1) - f(x-1)}{2}, \qquad f''(x) \coloneqq f(x-1) + f(x+1) - 2f(x),$$

und setzen $F'_n \coloneqq f'(X_{n-1})$ und $F''_n \coloneqq f''(X_{n-1})$. Dann gilt

$$f(X_n) - f(X_{n-1}) = \begin{cases} f(X_{n-1}+1) - f(X_{n-1}) & \text{für } X_n = X_{n-1} + 1, \quad (1) \\ f(X_{n-1}-1) - f(X_{n-1}) & \text{für } X_n = X_{n-1} - 1. \quad (2) \end{cases}$$

Im Fall (1) gilt mit der Abkürzung $x \coloneqq X_{n-1}$

$$\begin{aligned} f(x+1) - f(x) &= \tfrac{1}{2}\bigl(f(x+1) - f(x-1)\bigr) + \tfrac{1}{2}\bigl(f(x-1) + f(x+1)\bigr) - f(x) \\ &= f'(X_{n-1})(X_n - X_{n-1}) + \tfrac{1}{2}f''(X_{n-1}) = F'_n(X_n - X_{n-1}) + \tfrac{1}{2}F''_n, \end{aligned}$$

und im Fall (2) erhalten wir mit derselben Rechnung ebenfalls $f(X_{n-1} - 1) - f(X_{n-1}) = F'_n(X_n - X_{n-1}) + (1/2)F''_n$. Summation über n liefert die *diskrete Itô-Formel*

$$\begin{aligned} f(X_n) &= f(X_0) + \sum_{i=1}^{n} f'(X_{i-1})(X_i - X_{i-1}) + \sum_{i=1}^{n} \tfrac{1}{2}f''(X_{i-1}) \\ &= \underbrace{f(X_0) + (F' \cdot X)_n}_{\text{Martingal}} + \underbrace{\sum_{i=1}^{n} \tfrac{1}{2}F''_i}_{\text{Kompensator}}. \end{aligned}$$

Der Kompensator ist vorhersehbar, und $(F' \cdot X)_n$ ist ein diskretes stochastisches Integral mit vorhersehbarem und lokal beschränktem Integranden F', also nach Satz 5.15 ein Martingal. Gilt $f'' \ge 0$, ist also f (diskret) konvex, so ist $(f(X_n))$ ein Submartingal.

6. Optional Sampling und Optional Stopping

Zu diesem Kapitel vergleiche man [Kle13, Kapitel 10] und [Wil91].

Wir wollen im Folgenden anstelle von stochastischen Folgen etwas allgemeiner von *stochastischen Prozessen* sprechen, also von adaptierten Familien $(X_t, \mathcal{F}_t)_{t\in I}$, wobei $I \subset \mathbb{R}$ die Indexmenge ist (und nicht nur $\mathbb{N}_0$), und $(\mathcal{F}_t)_{t\in I}$ eine Filtration mit $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ für alle $s \leq t$. (Sub-, Super-)Martingale sind dann analog definiert, für Martingale etwa durch

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s \quad \mathbb{P}\text{-f.s.} \qquad \text{für alle } s \leq t.$$

6.1. Stoppzeiten

Definition 6.1 (Stoppzeit)**.** Eine Zufallsvariable τ mit Werten in $I \cup \{\infty\}$ heißt *Stoppzeit* bezüglich der Filtration $(\mathcal{F}_t)_{t\in I}$, falls für alle $t \in I$ gilt

$$\{\tau \leq t\} \in \mathcal{F}_t.$$

Bemerkung. Die Idee ist: $\mathcal{F}_t$ ist die Kenntnis zum Zeitpunkt t, und die Frage, ob das Ereignis $\{\tau \leq t\}$ eingetreten ist, soll mit dieser Kenntnis beantwortbar sein.

Beispiel 6.2 (Kontaktzeit)**.** Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine stochastische Folge mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und sei $K \in \mathcal{B}(\mathbb{R})$. Die Zufallsvariable

$$\tau_K \coloneqq \inf\{n \in \mathbb{N}_0 \colon X_n \in K\},$$

mit der Konvention $\inf \emptyset \coloneqq \infty$, ist eine Stoppzeit. Intuitiv kann der Erstkontakt mit K mit dem Wissen über den Pfad bis zum Erstkontakt entschieden werden. Formal gilt

$$\{\tau_K \leq n\} = \bigcup_{i=0}^{n} \{X_i \in K\} \in \mathcal{F}_n.$$

Dagegen ist der letzte Aufenthalt $\sup\{n \in \mathbb{N}_0 \colon X_n \in K\}$ im Allgemeinen *keine* Stoppzeit, denn um zu entscheiden, ob er bis zur Zeit n eingetreten ist, müssen wir die Zukunft nach n kennen; siehe Abbildung 6.1.

Lemma 6.3 (Kombination von Stoppzeiten)**.** *Seien σ, τ Stoppzeiten. Dann gilt:*

(i) Die Zufallsvariablen $\sigma \vee \tau$ und $\sigma \wedge \tau$ sind Stoppzeiten.

(ii) Ist $I = \mathbb{N}_0$ oder $I = [0, \infty)$, so ist $\sigma + \tau$ eine Stoppzeit.

(iii) Ist $I = \mathbb{N}_0$ und $s \in \mathbb{N}_0$ oder $I = [0, \infty)$ und $s \geq 0$, so ist $\tau + s$ eine Stoppzeit, jedoch im Allgemeinen nicht $\tau - s$.

Beweis. Zu (i): Für $t \in I$ gilt

$$\{\tau \vee \sigma \leq t\} = \{\tau \leq t\} \cap \{\sigma \leq t\} \in \mathcal{F}_t \quad \text{und} \quad \{\tau \wedge \sigma \leq t\} = \{\tau \leq t\} \cup \{\sigma \leq t\} \in \mathcal{F}_t.$$

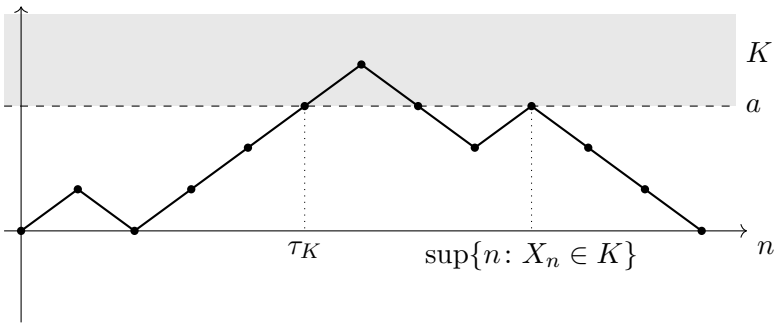

Abbildung 6.1.: Kontaktzeit τ_K eines Pfades mit $K = [a, \infty)$: Ob $\tau_K \leq n$ gilt, lässt sich am Pfad bis zur Zeit n ablesen. Für den letzten Aufenthalt in K ist das nicht der Fall: Erst die Zukunft entscheidet, ob der Pfad noch einmal nach K zurückkehrt.

Zu (ii): Sei $t \in I$. Nach (i) sind $\tau \wedge t$ und $\sigma \wedge t$ Stoppzeiten, da die Konstante t eine Stoppzeit ist. Also gilt für alle $s \in I$ mit $s < t$, dass $\{\tau \wedge t \leq s\} \in \mathcal{F}_s \subset \mathcal{F}_t$, und für $s \geq t$ gilt $\{\tau \wedge t \leq s\} = \Omega$. Für $I = [0, \infty)$ sind dies alle $s \geq 0$; für $I = \mathbb{N}_0$ genügt dies, da $\tau \wedge t$ dann Werte in $\mathbb{N}_0$ hat. Somit sind $\tau \wedge t$, und ebenso $\sigma \wedge t$, messbar bezüglich $\mathcal{F}_t$, und damit auch

$$\tau' \coloneqq (\tau \wedge t) + \mathbb{1}\{\tau > t\} \quad \text{und} \quad \sigma' \coloneqq (\sigma \wedge t) + \mathbb{1}\{\sigma > t\}$$

sowie $\tau' + \sigma'$. Wegen $\sigma, \tau \geq 0$ gilt

$$\{\tau' + \sigma' \leq t\} = \{(\tau \wedge t) + \mathbb{1}\{\tau > t\} + (\sigma \wedge t) + \mathbb{1}\{\sigma > t\} \leq t\} = \{\tau + \sigma \leq t\},$$

denn ist $\tau > t$ oder $\sigma > t$, so sind beide Seiten verletzt, und andernfalls gilt $\tau' = \tau$ und $\sigma' = \sigma$. Also folgt $\{\tau + \sigma \leq t\} \in \mathcal{F}_t$.

Zu (iii): Da die Konstante s eine Stoppzeit ist, folgt die erste Aussage aus (ii). Direkt sehen wir dies so: Für $t \in I$ mit $t \geq s$ gilt

$$\{\tau + s \leq t\} = \{\tau \leq t - s\} \in \mathcal{F}_{t-s} \subset \mathcal{F}_t,$$

und für $t < s$ ist $\{\tau + s \leq t\} = \emptyset \in \mathcal{F}_t$. Für $\tau - s$ erhalten wir dagegen nur

$$\{\tau - s \leq t\} = \{\tau \leq t + s\} \in \mathcal{F}_{t+s} \supset \mathcal{F}_t,$$

und $\mathcal{F}_{t+s}$ ist im Allgemeinen echt größer als $\mathcal{F}_t$. □

6.2. Die τ-Vergangenheit

Definition 6.4 (τ-Vergangenheit)**.** Sei τ eine Stoppzeit. Dann heißt

$$\mathcal{F}_\tau \coloneqq \{A \in \mathcal{F}\colon A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ für alle } t \in I\}$$

die *σ-Algebra der τ-Vergangenheit*.

Beispiel. Wir betrachten die Kontaktzeit $\tau \coloneqq \tau_{[a,\infty)}$ aus Beispiel 6.2 mit $K = [a, \infty)$ für ein $a \in \mathbb{R}$, sowie für ein $\delta > 0$ die Ereignisse

$$A \coloneqq \Big\{\sup_{n \in \mathbb{N}_0} X_n > a - \delta\Big\}, \qquad B \coloneqq \Big\{\sup_{n \in \mathbb{N}_0} X_n > a + \delta\Big\}.$$

Dann gilt $\{\tau \leq t\} \subset A$, also $A \cap \{\tau \leq t\} = \{\tau \leq t\} \in \mathcal{F}_t$ für alle t, und somit $A \in \mathcal{F}_\tau$. Zur Zeit τ wissen wir also, dass A eingetreten ist. Dagegen gilt im Allgemeinen $B \notin \mathcal{F}_\tau$, da zur Zeit τ noch unklar ist, ob B eintreten wird.

Lemma 6.5. *Sind σ, τ Stoppzeiten mit $\sigma \leq \tau$, so gilt $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.*

Beweis. Seien $A \in \mathcal{F}_\sigma$ und $t \in I$. Dann gilt $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$ und $\{\tau \leq t\} \in \mathcal{F}_t$. Wegen $\{\tau \leq t\} \subset \{\sigma \leq t\}$ folgt

$$A \cap \{\tau \leq t\} = \big(A \cap \{\sigma \leq t\}\big) \cap \{\tau \leq t\} \in \mathcal{F}_t. \qquad \square$$

Lemma 6.6. *Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine stochastische Folge und $\tau < \infty$ eine Stoppzeit. Dann ist*

$$X_\tau(\omega) \coloneqq X_{\tau(\omega)}(\omega), \qquad \omega \in \Omega,$$

messbar bezüglich $\mathcal{F}_\tau$.

Beweis. Seien $A \in \mathcal{B}(\mathbb{R})$ und $n \in \mathbb{N}_0$. Für alle $m \leq n$ ist $\{\tau = m\} \cap \{X_m \in A\}$ messbar bezüglich $\mathcal{F}_m$, und es gilt $\mathcal{F}_m \subset \mathcal{F}_n$. Also folgt

$$\{X_\tau \in A\} \cap \{\tau \leq n\} = \bigcup_{m\leq n} \underbrace{\big(\{\tau = m\} \cap \{X_m \in A\}\big)}_{\in \mathcal{F}_m} \in \mathcal{F}_n. \qquad \square$$

6.3. Optional Sampling

Lemma 6.7 (Gestoppte Martingale)**.** *Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Martingal, $T \in \mathbb{N}_0$ und τ eine Stoppzeit mit $\tau \leq T$. Dann gilt*

$$X_\tau = \mathbb{E}[X_T|\mathcal{F}_\tau] \quad \mathbb{P}\text{-f.s.} \qquad \text{und} \qquad \mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

Beweis. Nach Lemma 6.6 ist X_τ messbar bezüglich $\mathcal{F}_\tau$, und wegen $|X_\tau| \leq \sum_{n\leq T} |X_n|$ ist X_τ integrierbar. Sei $A \in \mathcal{F}_\tau$. Dann gilt $\{\tau = n\} \cap A \in \mathcal{F}_n$ für alle $n \in \mathbb{N}_0$, und mit der Martingaleigenschaft $X_n = \mathbb{E}[X_T|\mathcal{F}_n]$ für $n \leq T$ folgt

$$\begin{aligned}\mathbb{E}[X_\tau \mathbb{1}_A] &= \sum_{n\leq T} \mathbb{E}\big[X_n \mathbb{1}_{\{\tau=n\}\cap A}\big] = \sum_{n\leq T} \mathbb{E}\big[\mathbb{E}[X_T|\mathcal{F}_n]\mathbb{1}_{\{\tau=n\}\cap A}\big] \\ &= \sum_{n\leq T} \mathbb{E}\big[X_T \mathbb{1}_{\{\tau=n\}\cap A}\big] = \mathbb{E}[X_T \mathbb{1}_A].\end{aligned}$$

Also ist $X_\tau = \mathbb{E}[X_T|\mathcal{F}_\tau]$. Mit $A = \Omega$ und der Martingaleigenschaft folgt $\mathbb{E}[X_\tau] = \mathbb{E}[X_T] = \mathbb{E}[X_0]$. $\square$

Satz 6.8 (Optional Sampling Theorem)**.** *Sei $X = (X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und seien $\sigma \leq \tau$ Stoppzeiten. Dann gilt:*

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

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \leq X_\sigma$$

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

(ii) Ist $X \geq 0$ und $\tau < \infty$ $\mathbb{P}$-f.s., so gilt

$$\mathbb{E}[X_\tau] \leq \mathbb{E}[X_0] < \infty, \qquad \mathbb{E}[X_\sigma] \leq \mathbb{E}[X_0] < \infty \qquad \text{und} \qquad \mathbb{E}[X_\tau|\mathcal{F}_\sigma] \leq X_\sigma.$$

(iii) Ist X nur adaptiert und integrierbar, so ist X genau dann ein Martingal, wenn $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ für alle beschränkten Stoppzeiten τ gilt.

Beweis. Zu (i): Sei $X = M + A$ die Doob-Zerlegung aus Satz 5.17, wobei A vorhersehbar und nicht-steigend mit $A_0 = 0$ und M ein Martingal ist. (Nach Satz 5.17 und Satz 5.13 (i) ist der Kompensator eines Supermartingals nicht-steigend.) Wegen $\sigma \leq \tau$ gilt $A_\sigma \geq A_\tau$. Mit Lemma 6.7, angewandt auf M, der Monotonie der bedingten Erwartung, der Turmeigenschaft und Lemma 6.5 folgt

$$\begin{aligned} X_\sigma = A_\sigma + M_\sigma &= \mathbb{E}[A_\sigma + M_T|\mathcal{F}_\sigma] \geq \mathbb{E}[A_\tau + M_T|\mathcal{F}_\sigma] \\ &= \mathbb{E}\big[A_\tau + \mathbb{E}[M_T|\mathcal{F}_\tau]\big|\mathcal{F}_\sigma\big] = \mathbb{E}[A_\tau + M_\tau|\mathcal{F}_\sigma] = \mathbb{E}[X_\tau|\mathcal{F}_\sigma]. \end{aligned}$$

Dabei haben wir verwendet, dass A_σ nach Lemma 6.6 $\mathcal{F}_\sigma$-messbar ist. Ist X ein Martingal, so ist $A \equiv 0$, und es gilt überall Gleichheit.

Zu (ii): Wegen $\tau < \infty$ $\mathbb{P}$-f.s. gilt $X_{\tau\wedge n} \to X_\tau$ für $n \to \infty$ $\mathbb{P}$-f.s. Nach (i), angewandt auf die beschränkten Stoppzeiten $0 \leq \tau \wedge n$, gilt $\mathbb{E}[X_{\tau\wedge n}] \leq \mathbb{E}[X_0]$ für alle $n \in \mathbb{N}_0$. Mit dem Lemma von Fatou, anwendbar wegen $X \geq 0$, folgt

$$\mathbb{E}[X_\tau] \leq \liminf_{n\to\infty} \mathbb{E}[X_{\tau\wedge n}] \leq \mathbb{E}[X_0] < \infty.$$

Analog gilt $\mathbb{E}[X_\sigma] \leq \mathbb{E}[X_0]$. Seien nun $m, n \in \mathbb{N}_0$ mit $m \geq n$. Dann gilt $\tau \wedge m \geq \sigma \wedge n$, und mit (i) folgt

$$X_{\sigma\wedge n} \geq \mathbb{E}[X_{\tau\wedge m}|\mathcal{F}_{\sigma\wedge n}].$$

Für $A \in \mathcal{F}_\sigma$ gilt $\{\sigma < n\} \cap A \in \mathcal{F}_{\sigma\wedge n}$, also

$$\mathbb{E}\big[X_\sigma \mathbb{1}_{\{\sigma<n\}\cap A}\big] = \mathbb{E}\big[X_{\sigma\wedge n}\mathbb{1}_{\{\sigma<n\}\cap A}\big] \geq \mathbb{E}\big[\mathbb{E}[X_{\tau\wedge m}|\mathcal{F}_{\sigma\wedge n}]\mathbb{1}_{\{\sigma<n\}\cap A}\big] = \mathbb{E}\big[X_{\tau\wedge m}\mathbb{1}_{\{\sigma<n\}\cap A}\big].$$

Mit dem Lemma von Fatou für $m \to \infty$ folgt

$$\mathbb{E}\big[X_\tau \mathbb{1}_{\{\sigma<n\}\cap A}\big] \leq \liminf_{m\to\infty} \mathbb{E}\big[X_{\tau\wedge m}\mathbb{1}_{\{\sigma<n\}\cap A}\big] \leq \mathbb{E}\big[X_\sigma \mathbb{1}_{\{\sigma<n\}\cap A}\big],$$

und mit monotoner Konvergenz für $n \to \infty$ und $\sigma \leq \tau < \infty$ $\mathbb{P}$-f.s. erhalten wir $\mathbb{E}[X_\tau \mathbb{1}_A] \leq \mathbb{E}[X_\sigma \mathbb{1}_A]$. Da $A \in \mathcal{F}_\sigma$ beliebig war und X_σ $\mathcal{F}_\sigma$-messbar ist, folgt $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \leq X_\sigma$.

Zu (iii): Die Richtung „⇒“ folgt aus Lemma 6.7. Zu „⇐“: Es gelte $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ für alle beschränkten Stoppzeiten τ. Seien $t > s$ und $A \in \mathcal{F}_s$. Es genügt zu zeigen, dass $\mathbb{E}[X_t \mathbb{1}_A] = \mathbb{E}[X_s \mathbb{1}_A]$. Wir setzen $\tau := s\mathbb{1}_A + t\mathbb{1}_{A^c}$. Dann ist τ eine beschränkte Stoppzeit, denn $\{\tau \leq u\}$ ist gleich $\emptyset$ für $u < s$, gleich A für $s \leq u < t$ und gleich Ω für $u \geq t$. Wegen $\mathbb{E}[X_\tau] = \mathbb{E}[X_s \mathbb{1}_A] + \mathbb{E}[X_t \mathbb{1}_{A^c}]$ und $\mathbb{E}[X_t] = \mathbb{E}[X_0] = \mathbb{E}[X_\tau]$ folgt

$$\mathbb{E}[X_t \mathbb{1}_A] = \mathbb{E}[X_t] - \mathbb{E}[X_t \mathbb{1}_{A^c}] = \mathbb{E}[X_0] - \mathbb{E}[X_\tau] + \mathbb{E}[X_s \mathbb{1}_A] = \mathbb{E}[X_s \mathbb{1}_A]. \qquad \square$$

6.4. Gestoppte Prozesse und Optional Stopping

Definition 6.9 (Gestoppte Prozesse)**.** Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine stochastische Folge und τ eine Stoppzeit. Dann nennen wir

$$X_n^\tau := X_{\tau\wedge n}, \qquad n \in \mathbb{N}_0,$$

die *gestoppte stochastische Folge*. Wir schreiben $(\mathcal{F}_n^\tau)_{n\in\mathbb{N}_0} := (\mathcal{F}_{\tau\wedge n})_{n\in\mathbb{N}_0}$ für ihre Filtration.

Bemerkung. Die Folge $(X_n^\tau)_{n\in\mathbb{N}_0}$ ist sowohl an $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ als auch an $(\mathcal{F}_n^\tau)_{n\in\mathbb{N}_0}$ adaptiert, denn nach Lemma 6.6 ist $X_{\tau\wedge n}$ messbar bezüglich $\mathcal{F}_{\tau\wedge n}$, und nach Lemma 6.5 gilt $\mathcal{F}_{\tau\wedge n} \subset \mathcal{F}_n$.

Satz 6.10 (Optional Stopping)**.** *Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein (Sub-, Super-)Martingal und τ eine Stoppzeit. Dann ist $(X_n^\tau)_{n\in\mathbb{N}_0}$ ein (Sub-, Super-)Martingal bezüglich $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ und bezüglich $(\mathcal{F}_n^\tau)_{n\in\mathbb{N}_0}$.*

Beweis. Wir führen den Beweis nur für ein Submartingal $(X_n)_{n\in\mathbb{N}_0}$; die anderen Fälle folgen mit Satz 5.13 (i). Für alle $n \in \mathbb{N}_0$ gilt

$$\mathbb{E}\big[|X_n^\tau|\big] \le \mathbb{E}\big[\max\{|X_m|\colon m \le n\}\big] \le \sum_{m=0}^{n} \mathbb{E}\big[|X_m|\big] < \infty,$$

also ist X_n^τ integrierbar. Wegen $\{\tau > n-1\} \in \mathcal{F}_{n-1}$ gilt mit (5.15)

$$\begin{aligned}\mathbb{E}\big[X_n^\tau - X_{n-1}^\tau \big| \mathcal{F}_{n-1}\big] &= \mathbb{E}\big[X_{\tau\wedge n} - X_{\tau\wedge(n-1)} \big| \mathcal{F}_{n-1}\big] \\ &= \mathbb{E}\big[(X_n - X_{n-1})\mathbb{1}\{\tau > n-1\} \big| \mathcal{F}_{n-1}\big] \\ &= \mathbb{1}\{\tau > n-1\} \underbrace{\mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}]}_{\ge 0} \ge 0.\end{aligned}$$

Also ist $(X_n^\tau)_{n\in\mathbb{N}_0}$ ein $(\mathcal{F}_n)$-Submartingal. Da (X_n^τ) auch an $(\mathcal{F}_n^\tau)$ adaptiert ist und $\mathcal{F}_n^\tau \subset \mathcal{F}_n$ gilt, ist $(X_n^\tau)_{n\in\mathbb{N}_0}$ mit der Turmeigenschaft auch ein $(\mathcal{F}_n^\tau)$-Submartingal:

$$\mathbb{E}[X_n^\tau|\mathcal{F}_{n-1}^\tau] = \mathbb{E}\big[\mathbb{E}[X_n^\tau|\mathcal{F}_{n-1}]\big|\mathcal{F}_{n-1}^\tau\big] \ge \mathbb{E}[X_{n-1}^\tau|\mathcal{F}_{n-1}^\tau] = X_{n-1}^\tau. \qquad \square$$

Beispiel 6.11 (Treffzeiten der einfachen Irrfahrt)**.** Sei $(X_n)_{n\in\mathbb{N}_0}$ die symmetrische einfache Irrfahrt auf $\mathbb{Z}$ mit $X_0 = 0$ und seien $a < 0 < b$ ganze Zahlen. Dann sind

$$\tau_a \coloneqq \inf\{n \ge 0\colon X_n = a\}, \qquad \tau_b \coloneqq \inf\{n \ge 0\colon X_n = b\} \qquad \text{und} \qquad \tau_{a,b} \coloneqq \tau_a \wedge \tau_b$$

Stoppzeiten. Sei $A \coloneqq \{\tau_{a,b} = \tau_a\}$ das Ereignis, dass (X_n) zuerst a trifft. Wir suchen $\mathbb{P}(A)$; siehe Abbildung 6.2.

Es gilt $\tau_a, \tau_b < \infty$ $\mathbb{P}$-f.s., da $\limsup_{n\to\infty} X_n = \infty$ und $\liminf_{n\to\infty} X_n = -\infty$ $\mathbb{P}$-f.s. gilt (Übung). Nach Satz 6.10 ist $X^{\tau_{a,b}}$ ein Martingal. Da $\tau_{a,b}\wedge n \to \tau_{a,b}$ für $n \to \infty$ $\mathbb{P}$-f.s. gilt, folgt $X_n^{\tau_{a,b}} \to X_{\tau_{a,b}}$ $\mathbb{P}$-f.s. Da außerdem $|X_n^{\tau_{a,b}}| \le b - a$ $\mathbb{P}$-f.s. gilt, folgt mit dominierter Konvergenz $X_n^{\tau_{a,b}} \to X_{\tau_{a,b}}$ in L^1, und damit

$$0 = \lim_{n\to\infty} \mathbb{E}\big[X_n^{\tau_{a,b}}\big] = \mathbb{E}[X_{\tau_{a,b}}] = a\,\mathbb{P}(\tau_{a,b} = \tau_a) + b\,\mathbb{P}(\tau_{a,b} = \tau_b) = b + (a-b)\,\mathbb{P}(\tau_{a,b} = \tau_a).$$

Also gilt

$$\mathbb{P}(\tau_{a,b} = \tau_a) = \frac{b}{b-a}.$$

Wir berechnen nun $\mathbb{E}[\tau_{a,b}]$ und $\mathbb{E}[\tau_a]$. Wegen $|X_i - X_{i-1}| = 1$ gilt

$$\langle X\rangle_n = \sum_{i=1}^{n} \mathbb{E}\big[(X_i - X_{i-1})^2|\mathcal{F}_{i-1}\big] = n,$$

also ist $(X_n^2 - n)_{n\in\mathbb{N}_0}$ ein Martingal, und mit Satz 6.10 gilt

$$0 = \mathbb{E}\big[X_{\tau_{a,b}\wedge n}^2 - (\tau_{a,b}\wedge n)\big] \qquad \text{für alle } n \in \mathbb{N}_0.$$

Mit monotoner Konvergenz für $\mathbb{E}[\tau_{a,b}\wedge n]$ und dominierter Konvergenz für $\mathbb{E}[X_{\tau_{a,b}\wedge n}^2]$ folgt

$$\mathbb{E}[\tau_{a,b}] = \mathbb{E}\big[X_{\tau_{a,b}}^2\big] = a^2\,\mathbb{P}(\tau_{a,b} = \tau_a) + b^2\,\mathbb{P}(\tau_{a,b} = \tau_b) = \frac{a^2 b - b^2 a}{b-a} = |a|\,b.$$

Da $\tau_{a,b} \uparrow \tau_a$ für $b \uparrow \infty$ gilt, folgt mit monotoner Konvergenz $\mathbb{E}[\tau_a] = \infty$.

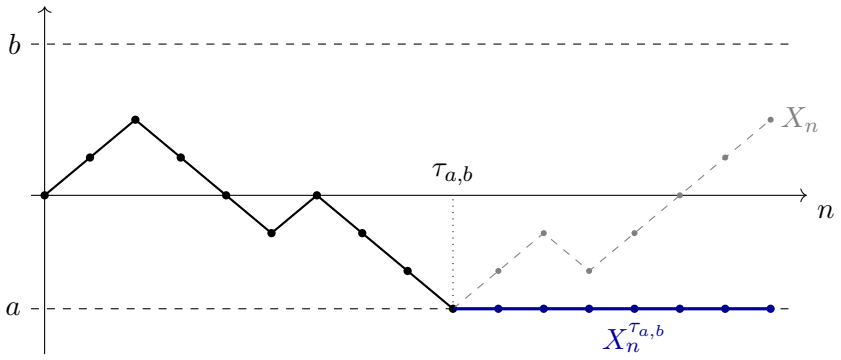

Abbildung 6.2.: Ein Pfad der einfachen Irrfahrt, der den Streifen zwischen a und b in a verlässt. Ab $\tau_{a,b}$ bleibt der gestoppte Pfad $X_n^{\tau_{a,b}} = X_{\tau_{a,b}\wedge n}$ konstant (blau), während X_n weiterläuft (grau gestrichelt). Insbesondere gilt $|X_n^{\tau_{a,b}}| \leq b - a$.

6.5. Gleichgradige Integrierbarkeit

Sei in diesem Abschnitt $(\Omega, \mathcal{F}, \mu)$ ein Maßraum.

Definition 6.12 (Gleichgradige Integrierbarkeit)**.** Eine Familie $\mathcal{U} \subset L^1(\mu)$ heißt *gleichgradig integrierbar*, falls

$$\inf_{0\leq g\in L^1(\mu)} \sup_{f\in\mathcal{U}} \int (|f| - g)^+ \,\mathrm{d}\mu = 0.$$

Satz 6.13 (Kriterien für gleichgradige Integrierbarkeit)**.** *Sei $\mu(\Omega) < \infty$. Dann ist die gleichgradige Integrierbarkeit von $\mathcal{U} \subset L^1(\mu)$ äquivalent zu jeder der folgenden Bedingungen:*

(1) $\inf\limits_{0\leq\tilde{g}\in L^1(\mu)} \sup\limits_{f\in\mathcal{U}} \int_{\{|f|>\tilde{g}\}} |f| \,\mathrm{d}\mu = 0,$

(2) $\inf\limits_{a\geq 0} \sup\limits_{f\in\mathcal{U}} \int (|f| - a)^+ \,\mathrm{d}\mu = 0,$

(3) $\inf\limits_{a\geq 0} \sup\limits_{f\in\mathcal{U}} \int_{\{|f|>a\}} |f| \,\mathrm{d}\mu = 0.$

Beweis. Zu (1): Die Richtung „⇐“ folgt aus $(|f| - g)^+ \leq |f|\mathbb{1}\{|f| > g\}$. Zu „⇒“: Für alle $\varepsilon > 0$ existiert ein $0 \leq g_\varepsilon \in L^1(\mu)$ mit

$$\sup_{f\in\mathcal{U}} \int (|f| - g_\varepsilon)^+ \,\mathrm{d}\mu \leq \varepsilon.$$

Wir setzen $\tilde{g}_\varepsilon := 2g_{\varepsilon/2}$. Wegen $|f| \leq (|f| - g_{\varepsilon/2})^+ + g_{\varepsilon/2}$ gilt dann für alle $f \in \mathcal{U}$

$$\int_{\{|f|>\tilde{g}_\varepsilon\}} |f| \,\mathrm{d}\mu \leq \underbrace{\int_{\{|f|>\tilde{g}_\varepsilon\}} (|f| - g_{\varepsilon/2})^+ \,\mathrm{d}\mu}_{=:\,(\mathrm{I})} + \underbrace{\int_{\{|f|>\tilde{g}_\varepsilon\}} g_{\varepsilon/2} \,\mathrm{d}\mu}_{=:\,(\mathrm{II})},$$

wobei (I) $\leq \varepsilon/2$ nach Konstruktion gilt. Auf $\{|f| > \tilde{g}_\varepsilon\}$ gilt $|f| - g_{\varepsilon/2} > g_{\varepsilon/2}$, also

$$g_{\varepsilon/2}\mathbb{1}\{|f| > \tilde{g}_\varepsilon\} \leq (|f| - g_{\varepsilon/2})^+\mathbb{1}\{|f| > \tilde{g}_\varepsilon\},$$

und somit auch (II) $\leq \varepsilon/2$.

Zu (2) und (3): Ebenso gilt (3) $\Rightarrow$ (2) wegen $(|f|-a)^+ \le |f|\mathbb{1}\{|f|>a\}$, und (2) impliziert die gleichgradige Integrierbarkeit, da wegen $\mu(\Omega)<\infty$ die Konstanten $a \ge 0$ in $L^1(\mu)$ liegen und das Infimum dort über eine kleinere Menge gebildet wird.

Es bleibt zu zeigen, dass die gleichgradige Integrierbarkeit (3) impliziert. Sei $\varepsilon > 0$ und $\tilde{g}_{\varepsilon/2}$ wie im Beweis von (1), sodass $\sup_{f\in\mathcal{U}} \int_{\{|f|>\tilde{g}_{\varepsilon/2}\}} |f| \,\mathrm{d}\mu \le \varepsilon/2$. Wir wählen a_ε so, dass

$$\int_{\{\tilde{g}_{\varepsilon/2}>a_\varepsilon\}} \tilde{g}_{\varepsilon/2} \,\mathrm{d}\mu < \frac{\varepsilon}{2};$$

dies ist wegen $\tilde{g}_{\varepsilon/2} \in L^1(\mu)$ mit dominierter Konvergenz möglich. Mit $\mathbb{1} = \mathbb{1}\{|f|>\tilde{g}_{\varepsilon/2}\} + \mathbb{1}\{|f|\le\tilde{g}_{\varepsilon/2}\}$ folgt für alle $f \in \mathcal{U}$

$$\int_{\{|f|>a_\varepsilon\}} |f| \,\mathrm{d}\mu \le \int_{\{|f|>\tilde{g}_{\varepsilon/2}\}} |f| \,\mathrm{d}\mu + \int_{\{\tilde{g}_{\varepsilon/2}>a_\varepsilon\}} \tilde{g}_{\varepsilon/2} \,\mathrm{d}\mu < \varepsilon. \qquad \square$$

Satz 6.14 (Weitere Eigenschaften)**.**

(1) Ist $\mathcal{U} \subset L^1(\mu)$ endlich, so ist $\mathcal{U}$ gleichgradig integrierbar.

(2) Sind $\mathcal{U}, \mathcal{V} \subset L^1(\mu)$ gleichgradig integrierbar, so sind auch

$$\{f+g\colon f\in\mathcal{U},\ g\in\mathcal{V}\}, \qquad \{f-g\colon f\in\mathcal{U},\ g\in\mathcal{V}\} \qquad \text{und} \qquad \{|f|\colon f\in\mathcal{U}\}$$

gleichgradig integrierbar.

(3) Ist $\mathcal{U}$ gleichgradig integrierbar und existiert zu jedem $g \in \mathcal{V}$ ein $f \in \mathcal{U}$ mit $|g| \le |f|$, so ist $\mathcal{V}$ gleichgradig integrierbar.

Beweis. Der Beweis ist eine Übungsaufgabe. $\square$

Satz 6.15 (Ein weiteres nützliches Kriterium)**.** *Sei $\mu(\Omega) < \infty$. Dann ist $\mathcal{U} \subset L^1(\mu)$ genau dann gleichgradig integrierbar, wenn es eine Funktion $H\colon [0,\infty) \to [0,\infty)$ gibt mit*

$$\lim_{x\to\infty} \frac{H(x)}{x} = \infty \qquad \text{und} \qquad \sup_{f\in\mathcal{U}} \int H(|f|) \,\mathrm{d}\mu < \infty.$$

Bemerkung. Die Funktion H kann monoton wachsend und konvex gewählt werden; siehe den Beweis.

Beweis. Zu „$\Leftarrow$“: Es gilt

$$K_a := \inf_{x\ge a} \frac{H(x)}{x} \uparrow \infty \qquad \text{für } a \uparrow \infty.$$

Wegen $|f| \le H(|f|)/K_a$ auf $\{|f| \ge a\}$ folgt

$$\sup_{f\in\mathcal{U}} \int_{\{|f|\ge a\}} |f| \,\mathrm{d}\mu \le \frac{1}{K_a} \sup_{f\in\mathcal{U}} \int_{\{|f|\ge a\}} H(|f|) \,\mathrm{d}\mu \le \frac{1}{K_a} \sup_{f\in\mathcal{U}} \int H(|f|) \,\mathrm{d}\mu \xrightarrow[a\to\infty]{} 0,$$

und mit Satz 6.13 (3) ist $\mathcal{U}$ gleichgradig integrierbar.

Zu „$\Rightarrow$“: Nach Satz 6.13 (2) existiert eine Folge $0 \le a_n \uparrow \infty$ mit

$$\sup_{f\in\mathcal{U}} \int (|f|-a_n)^+ \,\mathrm{d}\mu < 2^{-n} \qquad \text{für alle } n \in \mathbb{N}.$$

Wir setzen

$$H(x) := \sum_{n=1}^{\infty} (x-a_n)^+, \qquad x \ge 0;$$

für jedes x sind nur endlich viele Summanden ungleich null. Dann ist H monoton wachsend und konvex als Summe konvexer Funktionen, siehe Abbildung 6.3. Für alle $n \in \mathbb{N}$ und $x \geq 2a_n$ gilt $a_k \leq a_n \leq x/2$ für $k \leq n$, also

$$\frac{H(x)}{x} \geq \sum_{k=1}^{n} \Big(1 - \frac{a_k}{x}\Big)^+ \geq \frac{n}{2}.$$

Also gilt $H(x)/x \to \infty$ für $x \to \infty$. Mit dem Satz von der monotonen Konvergenz folgt für alle $f \in \mathcal{U}$

$$\int H(|f|)\,\mathrm{d}\mu = \sum_{n=1}^{\infty} \int (|f| - a_n)^+\,\mathrm{d}\mu \leq \sum_{n=1}^{\infty} 2^{-n} = 1. \qquad \square$$

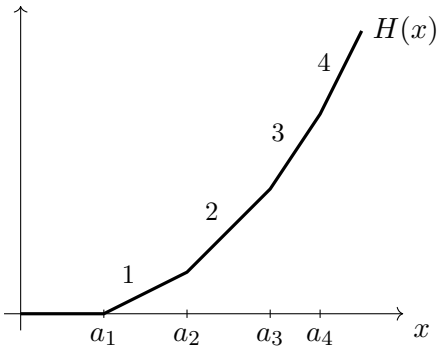

Abbildung 6.3.: Die Funktion $H(x) = \sum_n (x - a_n)^+$ aus dem Beweis von Satz 6.15 (vertikal gestaucht): Sie ist stückweise linear, und an jeder Stelle a_n wächst die Steigung um 1 (Zahlen an den Segmenten). Deshalb ist H konvex und $H(x)/x \to \infty$.

Bemerkung.

(1) Gilt $\sup\{\|f\|_p \colon f \in \mathcal{U}\} < \infty$ für ein $p > 1$, so ist $\mathcal{U}$ gleichgradig integrierbar; dies folgt aus Satz 6.15 mit $H(x) = x^p$.

(2) Gilt für eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen

$$\sup\{|\mathbb{E}[X_i]| \colon i \in I\} < \infty \qquad \text{und} \qquad \sup\{\mathrm{Var}[X_i] \colon i \in I\} < \infty,$$

so ist $(X_i)_{i \in I}$ gleichgradig integrierbar, denn dann ist $\sup_{i \in I} \mathbb{E}[X_i^2] < \infty$, und wir können (1) mit $p = 2$ anwenden.

Im Folgenden sei $(\Omega, \mathcal{F}, \mathbb{P})$ wieder ein Wahrscheinlichkeitsraum.

Satz 6.16 (Konvergenz in Wahrscheinlichkeit und gleichgradige Integrierbarkeit implizieren L^1-Konvergenz)**.** *Sei $(f_n)_{n \in \mathbb{N}}$ eine gleichgradig integrierbare Folge mit $f_n \to f$ in Wahrscheinlichkeit für $n \to \infty$. Dann gilt $f \in L^1(\mathbb{P})$ und $f_n \to f$ in L^1 für $n \to \infty$.*

Beweis. Angenommen, es gilt nicht $f_n \to f$ in L^1. Dann existieren ein $\varepsilon > 0$ und eine Teilfolge $(f_{n_k})_{k \in \mathbb{N}}$ mit

$$\|f - f_{n_k}\|_1 > 2\varepsilon \qquad \text{für alle } k \in \mathbb{N},$$

wobei wir $\|f - f_{n_k}\|_1 = \infty$ setzen, falls $f \notin L^1$. Nach Lemma 2.9 existiert eine Teilfolge $(f_{n_{k_m}})_{m \in \mathbb{N}}$ von (f_{n_k}) mit $f_{n_{k_m}} \to f$ $\mathbb{P}$-f.s. für $m \to \infty$. Mit dem Lemma von Fatou folgt

$$\int |f|\,\mathrm{d}\mathbb{P} \leq \liminf_{m \to \infty} \int |f_{n_{k_m}}|\,\mathrm{d}\mathbb{P} < \infty,$$

wobei die letzte Ungleichung aus der gleichgradigen Integrierbarkeit folgt, denn mit a wie in Satz 6.13 (2) gilt $|f_n| \le (|f_n| - a)^+ + a$. Also ist $f \in L^1$. Nach Satz 6.14 (1) und (2) ist $(|f - f_{n_{k_m}}|)_{m\in\mathbb{N}}$ gleichgradig integrierbar, das heißt, es existiert ein $0 \le g \in L^1$ mit

$$\int (|f - f_{n_{k_m}}| - g)^+ \,\mathrm{d}\mathbb{P} \le \varepsilon \qquad \text{für alle } m \in \mathbb{N}.$$

Wir setzen $g_m := |f_{n_{k_m}} - f| \wedge g$ für $m \in \mathbb{N}$. Dann gilt $g_m \to 0$ $\mathbb{P}$-f.s. für $m \to \infty$ und $g - g_m \ge 0$. Mit dem Lemma von Fatou folgt

$$\limsup_{m\to\infty} \int g_m \,\mathrm{d}\mathbb{P} = \int g \,\mathrm{d}\mathbb{P} - \liminf_{m\to\infty} \int (g - g_m) \,\mathrm{d}\mathbb{P} \le \int g \,\mathrm{d}\mathbb{P} - \int \liminf_{m\to\infty} (g - g_m) \,\mathrm{d}\mathbb{P} = 0.$$

Da $|f - f_{n_{k_m}}| = (|f - f_{n_{k_m}}| - g)^+ + g_m$ gilt (für $h := |f - f_{n_{k_m}}| \ge g$ ist $h = (h-g)^+ + g$, für $h < g$ ist $h = (h-g)^+ + h$), folgt

$$\limsup_{m\to\infty} \|f - f_{n_{k_m}}\|_1 \le \limsup_{m\to\infty} \int (|f - f_{n_{k_m}}| - g)^+ \,\mathrm{d}\mathbb{P} + \underbrace{\limsup_{m\to\infty} \int g_m \,\mathrm{d}\mathbb{P}}_{\le 0} \le \varepsilon,$$

im Widerspruch zu $\|f - f_{n_k}\|_1 > 2\varepsilon$ für alle k. □

6.6. Optional Sampling für unbeschränkte Stoppzeiten

Wir erweitern nun das Optional Sampling Theorem, Satz 6.8, auf unbeschränkte Stoppzeiten.

Lemma 6.17. *Ist $X = (X_n)_{n\in\mathbb{N}_0}$ ein gleichgradig integrierbares Martingal, so ist die Familie*

$$(X_\tau \colon \tau \text{ endliche Stoppzeit})$$

gleichgradig integrierbar.

Beweis. Nach Satz 6.15 und der anschließenden Bemerkung existiert eine monoton wachsende, konvexe Funktion $f \colon [0,\infty) \to [0,\infty)$ mit $\lim_{x\to\infty} f(x)/x = \infty$ und

$$L := \sup_{n\in\mathbb{N}_0} \mathbb{E}[f(|X_n|)] < \infty.$$

Sei $\tau < \infty$ eine Stoppzeit und $n \in \mathbb{N}_0$. Nach dem Optional Sampling Theorem, Satz 6.8 (i), angewandt auf die Stoppzeiten $\tau \wedge n \le n$, gilt

$$\mathbb{E}[X_n | \mathcal{F}_{\tau\wedge n}] = X_{\tau\wedge n}.$$

Da $\{\tau \le n\} \in \mathcal{F}_{\tau\wedge n}$ gilt, folgt mit der Jensenschen Ungleichung (5.14) und der Monotonie von f

$$\begin{aligned}\mathbb{E}[f(|X_\tau|)\mathbb{1}\{\tau \le n\}] &= \mathbb{E}[f(|X_{\tau\wedge n}|)\mathbb{1}\{\tau \le n\}] \le \mathbb{E}[\mathbb{E}[f(|X_n|)|\mathcal{F}_{\tau\wedge n}]\mathbb{1}\{\tau \le n\}] \\ &= \mathbb{E}[f(|X_n|)\mathbb{1}\{\tau \le n\}] \le L.\end{aligned}$$

Mit monotoner Konvergenz für $n \to \infty$ folgt $\mathbb{E}[f(|X_\tau|)] \le L$ für alle endlichen Stoppzeiten τ. Mit Satz 6.15 folgt die Behauptung. □

Satz 6.18 (Optional Sampling bei gleichgradiger Integrierbarkeit)**.** *Sei $(X_n)_{n\in\mathbb{N}_0}$ ein gleichgradig integrierbares Martingal (beziehungsweise Supermartingal) und seien $\sigma \le \tau$ endliche Stoppzeiten. Dann gilt $\mathbb{E}[|X_\tau|] < \infty$ und*

$$X_\sigma = \mathbb{E}[X_\tau | \mathcal{F}_\sigma] \qquad (\textit{beziehungsweise } X_\sigma \ge \mathbb{E}[X_\tau|\mathcal{F}_\sigma]).$$

Beweis. Sei zunächst X ein Martingal. Nach Lemma 6.17 ist X_τ integrierbar. Für $A \in \mathcal{F}_\sigma$ gilt $A \cap \{\sigma \le n\} \in \mathcal{F}_{\sigma\wedge n}$, und mit Satz 6.8 (i) folgt

$$\mathbb{E}\big[X_{\tau\wedge n}\mathbb{1}\{\sigma \le n\}\mathbb{1}_A\big] = \mathbb{E}\big[X_{\sigma\wedge n}\mathbb{1}\{\sigma \le n\}\mathbb{1}_A\big].$$

Nach Lemma 6.17 und Satz 6.14 (3) sind $(X_{\sigma\wedge n})_{n\in\mathbb{N}}$ und $(X_{\sigma\wedge n}\mathbb{1}\{\sigma \le n\}\mathbb{1}_A)_{n\in\mathbb{N}}$ gleichgradig integrierbar, und ebenso die entsprechenden Folgen mit τ statt σ. Außerdem gilt, da $\sigma \le \tau < \infty$, für alle $\varepsilon > 0$

$$\mathbb{P}\big(|X_\tau - X_{\tau\wedge n}\mathbb{1}\{\sigma \le n\}|\mathbb{1}_A > \varepsilon\big) \xrightarrow[n\to\infty]{} 0, \qquad \mathbb{P}\big(|X_\sigma - X_{\sigma\wedge n}\mathbb{1}\{\sigma \le n\}|\mathbb{1}_A > \varepsilon\big) \xrightarrow[n\to\infty]{} 0.$$

Mit Satz 6.16 folgt

$$\mathbb{E}[X_\tau\mathbb{1}_A] = \lim_{n\to\infty}\mathbb{E}\big[X_{\tau\wedge n}\mathbb{1}\{\sigma \le n\}\mathbb{1}_A\big] = \lim_{n\to\infty}\mathbb{E}\big[X_{\sigma\wedge n}\mathbb{1}\{\sigma \le n\}\mathbb{1}_A\big] = \mathbb{E}[X_\sigma\mathbb{1}_A].$$

Da X_σ nach Lemma 6.6 $\mathcal{F}_\sigma$-messbar ist, folgt $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] = X_\sigma$.

Sei nun X ein Supermartingal und $X = M + A$ seine Doob-Zerlegung, wobei M ein Martingal und A vorhersehbar und nicht-steigend mit $A_0 = 0$ ist. Dann gilt

$$\mathbb{E}[|A_n|] = \mathbb{E}[-A_n] = \mathbb{E}[M_n - X_n] \le \mathbb{E}[|X_0|] + \sup_{m\in\mathbb{N}_0}\mathbb{E}[|X_m|] < \infty,$$

denn $\mathbb{E}[M_n] = \mathbb{E}[M_0] = \mathbb{E}[X_0]$, und das Supremum ist wegen der gleichgradigen Integrierbarkeit endlich. Nach dem Satz von der monotonen Konvergenz existiert daher A_∞ mit $A_n \downarrow A_\infty$, $A_\infty \le 0$ und $\mathbb{E}[-A_\infty] < \infty$. Wegen $|A_n| \le |A_\infty|$ für alle n ist A nach Satz 6.14 (1) (für die Familie $\{A_\infty\}$) und (3) gleichgradig integrierbar, und damit nach Satz 6.14 (2) auch $M = X - A$. Also folgt mit dem ersten Teil

$$\mathbb{E}[|X_\tau|] \le \mathbb{E}[-A_\tau] + \mathbb{E}[|M_\tau|] \le \mathbb{E}[-A_\infty] + \mathbb{E}[|M_\tau|] < \infty$$

und

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] = \mathbb{E}[M_\tau|\mathcal{F}_\sigma] + \mathbb{E}[A_\tau|\mathcal{F}_\sigma] = M_\sigma + A_\sigma + \mathbb{E}\big[\underbrace{A_\tau - A_\sigma}_{\le 0}\big|\mathcal{F}_\sigma\big] \le M_\sigma + A_\sigma = X_\sigma. \quad \square$$

7. Martingalkonvergenzsätze

Zu diesem Kapitel vergleiche man [Kle13, Kapitel 11], [Dur19, Kapitel 4] und [Doo53].

Sei $I \subset \mathbb{N}_0$ und $X = (X_n)_{n\in I}$ eine stochastische Folge. Wir setzen

$$X_n^* \coloneqq \sup\{X_k \colon k \le n\} \qquad \text{und} \qquad |X|_n^* \coloneqq \sup\{|X_k| \colon k \le n\},$$

wobei das Supremum jeweils über $k \in I$ gebildet wird.

7.1. Doobsche Ungleichungen

Lemma 7.1. *Ist X ein Submartingal, so gilt für alle $\lambda > 0$*

$$\lambda\,\mathbb{P}(X_n^* \ge \lambda) \le \mathbb{E}\big[X_n \mathbb{1}\{X_n^* \ge \lambda\}\big] \le \mathbb{E}\big[|X_n| \mathbb{1}\{X_n^* \ge \lambda\}\big].$$

Beweis. Die zweite Ungleichung folgt aus $X_n \le |X_n|$. Für die erste betrachten wir die beschränkte Stoppzeit

$$\tau \coloneqq \inf\{k \in I \colon X_k \ge \lambda\} \wedge n;$$

siehe Abbildung 7.1. Auf $\{X_n^* \ge \lambda\}$ gilt $X_\tau \ge \lambda$, und auf $\{X_n^* < \lambda\}$ gilt $\tau = n$. Nach dem Optional Sampling Theorem, Satz 6.8 (i) (angewandt auf das Supermartingal $-X$), gilt daher

$$\begin{aligned}\mathbb{E}[X_n] \ge \mathbb{E}[X_\tau] &= \mathbb{E}\big[X_\tau \mathbb{1}\{X_n^* \ge \lambda\}\big] + \mathbb{E}\big[X_\tau \mathbb{1}\{X_n^* < \lambda\}\big] \\ &\ge \lambda\,\mathbb{P}(X_n^* \ge \lambda) + \mathbb{E}\big[X_n \mathbb{1}\{X_n^* < \lambda\}\big].\end{aligned}$$

Subtrahieren wir $\mathbb{E}[X_n \mathbb{1}\{X_n^* < \lambda\}]$ auf beiden Seiten, so folgt die Behauptung. □

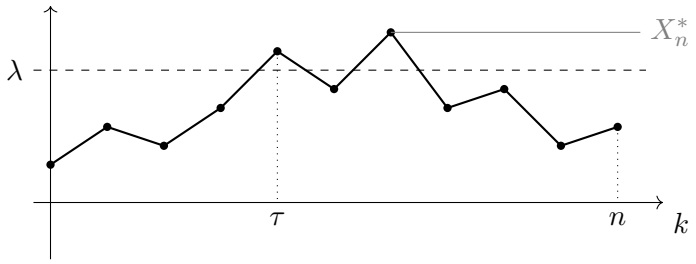

Abbildung 7.1.: Zum Beweis von Lemma 7.1: Überschreitet der Pfad bis zur Zeit n das Niveau λ, so stoppt τ beim ersten Überschreiten mit $X_\tau \ge \lambda$; andernfalls ist $\tau = n$.

Satz 7.2 (Doobsche Ungleichungen)**.** *Sei X ein Martingal oder ein positives Submartingal.*

(1) Für alle $p \ge 1$ und $\lambda > 0$ gilt

$$\lambda^p\,\mathbb{P}\big(|X|_n^* \ge \lambda\big) \le \mathbb{E}\big[|X_n|^p\big].$$

(2) Für alle $p > 1$ gilt

$$\mathbb{E}\big[|X_n|^p\big] \le \mathbb{E}\big[(|X|_n^*)^p\big] \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_n|^p\big].$$

Beweis. Zu (1): Ist $\mathbb{E}[|X_n|^p] = \infty$, so ist nichts zu zeigen. Andernfalls ist $(|X_k|^p)_{k\le n}$ nach der Jensenschen Ungleichung, vergleiche Beispiel 5.11 (v), ein Submartingal. Die Aussage folgt daher aus Lemma 7.1 mit λ^p und $|X|^p$ statt λ und X, denn $\{|X|_n^* \ge \lambda\} = \{\sup_{k\le n} |X_k|^p \ge \lambda^p\}$.

Zu (2): Die erste Ungleichung folgt aus $|X_n| \le |X|_n^*$. Die zweite Ungleichung ist trivial, falls $\mathbb{E}[|X_n|^p] = \infty$. Sei also $\mathbb{E}[|X_n|^p] < \infty$. Nach Lemma 7.1, angewandt auf das Submartingal $(|X_k|)$, gilt

$$\lambda\,\mathbb{P}(|X|_n^* \ge \lambda) \le \mathbb{E}\big[|X_n| \mathbb{1}\{|X|_n^* \ge \lambda\}\big].$$

Damit folgt für alle $K > 0$ mit dem Satz von Fubini

$$\begin{aligned}
\mathbb{E}\big[(|X|_n^* \wedge K)^p\big] &= \mathbb{E}\Big[\int_0^{|X|_n^* \wedge K} p\lambda^{p-1}\,\mathrm{d}\lambda\Big] = \mathbb{E}\Big[\int_0^K p\lambda^{p-1}\mathbb{1}\{|X|_n^* \ge \lambda\}\,\mathrm{d}\lambda\Big] \\
&= \int_0^K p\lambda^{p-1}\,\mathbb{P}(|X|_n^* \ge \lambda)\,\mathrm{d}\lambda \\
&\le \int_0^K p\lambda^{p-2}\,\mathbb{E}\big[|X_n|\mathbb{1}\{|X|_n^* \ge \lambda\}\big]\,\mathrm{d}\lambda = p\,\mathbb{E}\Big[|X_n| \int_0^{|X|_n^* \wedge K} \lambda^{p-2}\,\mathrm{d}\lambda\Big] \\
&= \frac{p}{p-1}\mathbb{E}\big[|X_n|\,(|X|_n^* \wedge K)^{p-1}\big] \\
&\le \frac{p}{p-1}\mathbb{E}\big[(|X|_n^* \wedge K)^p\big]^{\frac{p-1}{p}}\,\mathbb{E}\big[|X_n|^p\big]^{\frac{1}{p}},
\end{aligned}$$

wobei wir im letzten Schritt die Höldersche Ungleichung verwendet haben. Da $\mathbb{E}[(|X|_n^* \wedge K)^p] \le K^p < \infty$ gilt, folgt nach Division (der Fall $\mathbb{E}[(|X|_n^* \wedge K)^p] = 0$ ist trivial)

$$\mathbb{E}\big[(|X|_n^* \wedge K)^p\big] \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_n|^p\big].$$

Für $K \uparrow \infty$ folgt die Behauptung mit monotoner Konvergenz. □

7.2. Der Martingalkonvergenzsatz von Doob

Wir beginnen mit der Idee hinter den Martingalkonvergenzsätzen. Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine reellwertige stochastische Folge, und sei

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

Wir fassen X_n als Modell für einen Aktienkurs auf. Entspricht dieser einem fairen Spiel, so ist $(X_n)_{n\in\mathbb{N}_0}$ ein $(\mathcal{F}_n)$-Martingal. Wir betrachten zwei Niveaus $a < b$ und die folgende Spielstrategie: Wir kaufen, sobald $X_n \le a$ gilt, und verkaufen, sobald $X_n \ge b$ gilt. Jede solche *Aufkreuzung* des Intervalls $[a, b]$ führt zu einem Gewinn von mindestens $b - a$. Ist der maximale Gewinn endlich, so gibt es für alle $a < b$ nur endlich viele solcher Aufkreuzungen, und wir erwarten, dass X_n fast sicher konvergiert.

Um dies zu präzisieren, definieren wir die Stoppzeiten $\sigma_0 \equiv 0$ und rekursiv für $k \in \mathbb{N}$

$$\begin{aligned}
\tau_k &:= \inf\{n \ge \sigma_{k-1} \colon X_n \le a\}, \\
\sigma_k &:= \inf\{n \ge \tau_k \colon X_n \ge b\},
\end{aligned}$$

wobei $\inf\emptyset = \infty$; insbesondere folgt aus $\sigma_{k-1} = \infty$, dass $\tau_k = \infty$ und damit $\sigma_k = \infty$ gilt. Ist $\sigma_k < \infty$, so sagen wir, dass X zwischen τ_k und σ_k seine *k-te Aufkreuzung* über $[a, b]$ hat. Die Zufallsvariable

$$U_n^{a,b} := \sup\{k \in \mathbb{N}_0 \colon \sigma_k \le n\}$$

ist die *Anzahl der Aufkreuzungen* bis zur Zeit n; sie ist $\mathcal{F}_n$-messbar. Im Folgenden schreiben wir $x^+ := x \vee 0$ und $x^- := -(x \wedge 0)$, sodass $x = x^+ - x^-$ gilt.

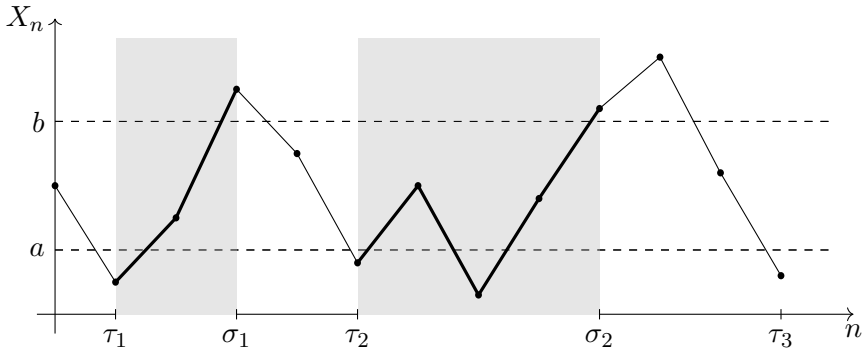

Abbildung 7.2.: Ein Pfad von X mit den Stoppzeiten τ_k und σ_k. Die beiden Aufkreuzungen über $[a, b]$ sind fett gezeichnet; grau unterlegt sind die Zeiten $m \in \{\tau_k + 1, \dots, \sigma_k\}$, zu denen die Strategie H aus dem Beweis von Lemma 7.3 investiert ist. Hier gilt $U_{12}^{a,b} = 2$.

Lemma 7.3 (Aufkreuzungsungleichung)**.** *Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal und seien $a < b$. Dann gilt für alle $n \in \mathbb{N}_0$*

$$\mathbb{E}[U_n^{a,b}] \le \frac{1}{b-a}\Big(\mathbb{E}[(X_n - a)^+] - \mathbb{E}[(X_0 - a)^+]\Big).$$

Beweis. Wir betrachten das stochastische Integral $H \cdot X$, also $(H \cdot X)_n = \sum_{m=1}^n H_m(X_m - X_{m-1})$, mit

$$H_m := \begin{cases} 1, & \text{falls } m \in \{\tau_k + 1, \dots, \sigma_k\} \text{ für ein } k \in \mathbb{N}, \\ 0, & \text{sonst.} \end{cases}$$

Es gilt $H_m \ge 0$, und H ist vorhersehbar, denn

$$\{H_m = 1\} = \bigcup_{k=1}^{\infty} \big(\underbrace{\{\tau_k \le m-1\} \cap \{\sigma_k > m-1\}}_{\in \mathcal{F}_{m-1}}\big).$$

Sei $Y_n := \max\{X_n, a\}$. Ist $\sigma_k < \infty$, so gilt $Y_{\sigma_i} - Y_{\tau_i} = Y_{\sigma_i} - a \ge b - a$ für alle $i \le k$ und daher

$$(H \cdot Y)_{\sigma_k} = \sum_{i=1}^{k} \sum_{j=\tau_i+1}^{\sigma_i} (Y_j - Y_{j-1}) = \sum_{i=1}^{k} (Y_{\sigma_i} - Y_{\tau_i}) \ge k(b-a).$$

Für $j \in \{\sigma_k, \dots, \tau_{k+1}\}$ gilt $(H \cdot Y)_j = (H \cdot Y)_{\sigma_k}$, und für $j \in \{\tau_k + 1, \dots, \sigma_k\}$ gilt wegen $Y_j \ge a = Y_{\tau_k}$

$$(H \cdot Y)_j \ge (H \cdot Y)_{\tau_k} = (H \cdot Y)_{\sigma_{k-1}}.$$

Wenden wir dies mit $k = U_n^{a,b}$ an, so folgt

$$(H \cdot Y)_n \ge (b-a)\, U_n^{a,b}.$$

Der Prozess $Y = (X-a)^+ + a$ ist ein Submartingal, da $X - a$ ein Submartingal ist und damit auch $(X-a)^+ = (X-a) \vee 0$. Da H und $1 - H$ vorhersehbar und durch 0 und 1 beschränkt sind, sind $H \cdot Y$ und $(1-H) \cdot Y$ Submartingale, die in 0 starten. Nun gilt

$$Y_n - Y_0 = (1 \cdot Y)_n = (H \cdot Y)_n + \big((1-H) \cdot Y\big)_n,$$

also wegen $\mathbb{E}[((1-H) \cdot Y)_n] \ge 0$

$$\mathbb{E}[(X_n - a)^+] - \mathbb{E}[(X_0 - a)^+] = \mathbb{E}[Y_n - Y_0] \ge \mathbb{E}[(H \cdot Y)_n] \ge (b-a)\, \mathbb{E}[U_n^{a,b}]. \qquad \square$$

Satz 7.4 (Martingalkonvergenzsatz [Doo53]). *Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Submartingal, und es gelte $\sup_{n\in\mathbb{N}_0}\mathbb{E}[X_n^+]<\infty$. Dann existiert eine $\mathcal{F}_\infty$-messbare Zufallsvariable X_∞ mit $\mathbb{E}[|X_\infty|]<\infty$ und*

$$X_n \xrightarrow[n\to\infty]{} X_\infty \qquad \mathbb{P}\text{-f.s.}$$

Beweis. Seien $a<b$. Nach Lemma 7.3 gilt

$$\mathbb{E}[U_n^{a,b}] \le \frac{1}{b-a}\mathbb{E}[(X_n-a)^+] \le \frac{1}{b-a}(|a|+\mathbb{E}[X_n^+]).$$

Da $U_n^{a,b}$ monoton wachsend in n ist, existiert $U^{a,b} \coloneqq \lim_{n\to\infty} U_n^{a,b}$, und nach dem Lemma von Fatou und der Voraussetzung gilt

$$\mathbb{E}[U^{a,b}] \le \liminf_{n\to\infty}\mathbb{E}[U_n^{a,b}] < \infty,$$

also $\mathbb{P}(U^{a,b}<\infty)=1$. Wir definieren die $\mathcal{F}_\infty$-messbaren Ereignisse

$$C^{a,b} \coloneqq \Big\{\liminf_{n\to\infty} X_n < a\Big\} \cap \Big\{\limsup_{n\to\infty} X_n > b\Big\} \subset \{U^{a,b}=\infty\} \qquad \text{und} \qquad C \coloneqq \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} C^{a,b}.$$

Es folgt $\mathbb{P}(C^{a,b})=0$ für alle rationalen $a<b$ und damit $\mathbb{P}(C)=0$. Auf $C^{\mathrm{c}}=\Omega\setminus C$ gilt $\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n$, das heißt, X_n konvergiert in $[-\infty,\infty]$. Also existiert $X_\infty \coloneqq \lim_{n\to\infty} X_n$ $\mathbb{P}$-fast sicher, und da X_n für alle n $\mathcal{F}_\infty$-messbar ist, ist auch X_∞ $\mathcal{F}_\infty$-messbar. Mit dem Lemma von Fatou gilt

$$\mathbb{E}[X_\infty^+] \le \liminf_{n\to\infty}\mathbb{E}[X_n^+] \le \sup\{\mathbb{E}[X_n^+]\colon n\in\mathbb{N}_0\} < \infty$$

und, da X ein Submartingal ist und daher $\mathbb{E}[X_n]\ge\mathbb{E}[X_0]$ gilt,

$$\mathbb{E}[X_\infty^-] \le \liminf_{n\to\infty}\mathbb{E}[X_n^-] = \liminf_{n\to\infty}\big(\mathbb{E}[X_n^+]-\mathbb{E}[X_n]\big) \le \sup\{\mathbb{E}[X_n^+]\colon n\in\mathbb{N}_0\} - \mathbb{E}[X_0] < \infty.$$

Also gilt $\mathbb{E}[|X_\infty|]<\infty$; insbesondere ist X_∞ fast sicher endlich. □

Korollar 7.5 (Version für nichtnegative Supermartingale). *Sei $X\ge 0$ ein Supermartingal. Dann existiert eine $\mathcal{F}_\infty$-messbare Zufallsvariable $X_\infty\ge 0$ mit $\mathbb{E}[X_\infty]\le\mathbb{E}[X_0]$ und $X_n \xrightarrow[n\to\infty]{} X_\infty$ fast sicher.*

Beweis. Der Prozess $-X$ ist ein Submartingal mit $\mathbb{E}[(-X_n)^+]=0$ für alle $n\in\mathbb{N}_0$. Nach Satz 7.4 existiert also $X_\infty\ge 0$ mit $X_n\to X_\infty$ fast sicher, und mit dem Lemma von Fatou gilt

$$\mathbb{E}[X_\infty] \le \liminf_{n\to\infty}\mathbb{E}[X_n] \le \mathbb{E}[X_0].$$ □

Beispiel (Petersburger Spiel). Wir betrachten das Petersburger Spiel aus Beispiel 5.16. Der Gewinnprozess $S_n=\sum_{i=1}^n H_iD_i$ ist ein Martingal mit $S_n\le 1$ $\mathbb{P}$-fast sicher für alle $n\in\mathbb{N}$; dabei wird der Einsatz so lange verdoppelt, bis zum ersten Mal gewonnen wird (siehe Abbildung 7.3). Nach Satz 7.4 gilt

$$S_n \xrightarrow[n\to\infty]{} S_\infty = 1 \qquad \mathbb{P}\text{-f.s.},$$

aber $\mathbb{E}[S_n]=0$ für alle n. Also konvergiert S_n nicht in L^1, und $(S_n)_{n\in\mathbb{N}}$ ist nicht gleichgradig integrierbar (zur Übung).

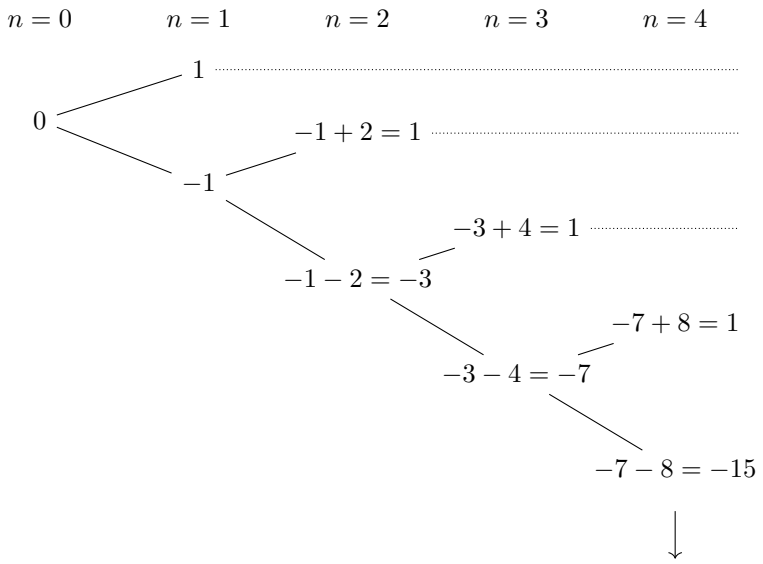

Abbildung 7.3.: Mögliche Werte von S_n im Petersburger Spiel. Nach dem ersten Gewinn bleibt $S_n = 1$; der Pfad, auf dem nur verloren wird, hat Wahrscheinlichkeit 2^{-n} und den Wert $1 - 2^n$.

Satz 7.6 (Version für gleichgradig integrierbare Martingale)**.** *Sei $(X_n)_{n\in\mathbb{N}_0}$ ein gleichgradig integrierbares Sub-, Super- beziehungsweise Martingal. Dann existiert eine $\mathcal{F}_\infty$-messbare Zufallsvariable X_∞ mit $X_n \xrightarrow[n\to\infty]{} X_\infty$ $\mathbb{P}$-fast sicher* und in L^1*. Außerdem gilt für alle $n \in \mathbb{N}_0$:*

(1) $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$, falls X ein Martingal ist;

(2) $X_n \le \mathbb{E}[X_\infty|\mathcal{F}_n]$, falls X ein Submartingal ist;

(3) $X_n \ge \mathbb{E}[X_\infty|\mathcal{F}_n]$, falls X ein Supermartingal ist.

Bemerkung. Kurz gesagt ist $(X_n)_{n\in\mathbb{N}_0\cup\{\infty\}}$ ein Sub-, Super- beziehungsweise Martingal bezüglich der Filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0\cup\{\infty\}}$.

Beweis von Satz 7.6. Wir betrachten nur den Fall, dass X ein Submartingal ist; der Fall eines Supermartingals folgt durch Übergang zu $-X$, und ein Martingal ist beides. Da X gleichgradig integrierbar ist, gilt $\sup\{\mathbb{E}[X_n^+]\colon n \in \mathbb{N}_0\} < \infty$. Nach Satz 7.4 existiert also X_∞ mit $X_n \to X_\infty$ $\mathbb{P}$-fast sicher, also auch in Wahrscheinlichkeit, und nach Satz 6.16 folgt

$$\mathbb{E}\big[|X_n - X_\infty|\big] \xrightarrow[n\to\infty]{} 0.$$

Sei $m \in \mathbb{N}_0$. Mit der Jensen-Ungleichung für bedingte Erwartungen erhalten wir

$$\mathbb{E}\Big[\big|\mathbb{E}[X_n|\mathcal{F}_m] - \mathbb{E}[X_\infty|\mathcal{F}_m]\big|\Big] \le \mathbb{E}\Big[\mathbb{E}\big[|X_n - X_\infty||\mathcal{F}_m\big]\Big] = \mathbb{E}\big[|X_n - X_\infty|\big] \xrightarrow[n\to\infty]{} 0.$$

Wegen $|u^- - v^-| \le |u - v|$ für $u, v \in \mathbb{R}$ folgt daraus

$$\Big|\mathbb{E}\big[(\mathbb{E}[X_\infty|\mathcal{F}_m] - X_m)^-\big] - \mathbb{E}\big[(\mathbb{E}[X_n|\mathcal{F}_m] - X_m)^-\big]\Big| \le \mathbb{E}\Big[\big|\mathbb{E}[X_\infty|\mathcal{F}_m] - \mathbb{E}[X_n|\mathcal{F}_m]\big|\Big] \xrightarrow[n\to\infty]{} 0.$$

Da X ein Submartingal ist, gilt aber $(\mathbb{E}[X_n|\mathcal{F}_m] - X_m)^- = 0$ für alle $n \ge m$. Also gilt $\mathbb{E}[(\mathbb{E}[X_\infty|\mathcal{F}_m] - X_m)^-] = 0$ und damit $\mathbb{E}[X_\infty|\mathcal{F}_m] - X_m \ge 0$ $\mathbb{P}$-fast sicher. □

Korollar 7.7. *Sei $X \geq 0$ ein Martingal und $X_\infty = \lim_{n\to\infty} X_n$ (dieser Grenzwert existiert nach Korollar 7.5 fast sicher). Dann gilt*

$$\mathbb{E}[X_\infty] = \mathbb{E}[X_0] \quad \Longleftrightarrow \quad X \text{ ist gleichgradig integrierbar.}$$

Beweis. Die Richtung „$\Leftarrow$“ folgt aus Satz 7.6, denn aus der L^1-Konvergenz folgt $\mathbb{E}[X_\infty] = \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X_0]$. Für „$\Rightarrow$“ beachten wir, dass $X_n \geq 0$, $X_n \to X_\infty$ fast sicher und $\mathbb{E}[X_n] = \mathbb{E}[X_0] = \mathbb{E}[X_\infty]$ gelten; nach dem Lemma von Scheffé konvergiert X_n daher sogar in L^1, woraus die gleichgradige Integrierbarkeit folgt (zur Übung). □

Satz 7.8 (Version für L^p-Martingale)**.** *Seien $p > 1$ und $(X_n)_{n\in\mathbb{N}_0}$ ein L^p-beschränktes Martingal, das heißt $\sup\{\mathbb{E}[|X_n|^p] \colon n \in \mathbb{N}_0\} < \infty$. Dann existiert eine $\mathcal{F}_\infty$-messbare Zufallsvariable X_∞ mit $\mathbb{E}[|X_\infty|^p] < \infty$ und $X_n \xrightarrow[n\to\infty]{} X_\infty$ $\mathbb{P}$-fast sicher und in L^p. Insbesondere ist $(|X_n|^p)_{n\in\mathbb{N}_0}$ gleichgradig integrierbar.*

Beweis. Zunächst ist X nach Satz 6.15 mit $H(x) = x^p$ gleichgradig integrierbar, siehe auch die Bemerkung im Anschluss an diesen Satz; Nach der Doob-Ungleichung (Satz 7.2) gilt

$$\mathbb{E}[\sup\{|X_k|^p \colon k \leq n\}] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|X_n|^p],$$

also mit monotoner Konvergenz

$$\mathbb{E}[\sup\{|X_k|^p \colon k \geq 0\}] \leq \Big(\frac{p}{p-1}\Big)^p \sup\{\mathbb{E}[|X_n|^p] \colon n \geq 0\} < \infty.$$

Daher ist $(|X_n|^p)_{n\in\mathbb{N}_0}$ gleichgradig integrierbar. Da $|X_n - X_\infty|^p \leq 2^p \sup\{|X_k|^p \colon k \geq 0\}$ integrierbar ist, folgt mit dem Satz von der dominierten Konvergenz

$$\mathbb{E}\big[|X_\infty|^p\big] < \infty \qquad \text{und} \qquad \mathbb{E}\big[|X_n - X_\infty|^p\big] \xrightarrow[n\to\infty]{} 0. \qquad \square$$

Korollar 7.9 (Version für L^2-Martingale)**.** *Sei X ein quadratintegrierbares Martingal mit quadratischer Variation $\langle X\rangle$. Dann sind die folgenden Aussagen äquivalent:*

(i) $\sup\{\mathbb{E}[X_n^2] \colon n \geq 0\} < \infty$*;*

(ii) $\lim_{n\to\infty} \mathbb{E}[\langle X\rangle_n] < \infty$*;*

(iii) X konvergiert in L^2;

(iv) X konvergiert fast sicher und in L^2.

Beweis. Die Äquivalenz von (i) und (ii) folgt aus

$$\mathrm{Var}[X_n - X_0] = \mathbb{E}\big[(X_n - X_0)^2\big] = \mathbb{E}[X_n^2] - \mathbb{E}[X_0^2] = \mathbb{E}[\langle X\rangle_n],$$

wobei die rechte Seite wachsend in n ist; im zweiten Schritt haben wir die Orthogonalität der Martingalzuwächse benutzt. Die Implikationen (iv)⇒(iii)⇒(i) sind klar, da L^2-konvergente Folgen in L^2 beschränkt sind. Die Implikation (i)⇒(iv) ist Satz 7.8 mit $p = 2$. □

Teil II.

Diskrete Finanzmathematik

8. Das Einperiodenmodell

Zu diesem Kapitel vergleiche man [FS02, Kapitel 1] sowie [Shr04].

Mit diesem Kapitel beginnt der finanzmathematische Teil der Vorlesung. Wir betrachten ein Finanzgut, etwa ein Konto, eine Aktie, einen Rohstoff, Strom oder eine Währung, mit Preisen $(S_t)_{t\in\mathbb{N}_0}$. Ein Beispiel ist eine *festverzinsliche Anleihe* mit Startpreis S_0 und Zinsrate $r \geq 0$, für die

$$S_t = (1+r)^t S_0$$

gilt. Ein anderes Beispiel ist eine *Aktie* in einem einfachen Modell: Seien $(Z_s)_{s\in\mathbb{N}}$ unabhängig und identisch Bernoulli(p)-verteilt auf $\{0,1\}$, und seien $r_g > r_b$ eine „gute“ und eine „schlechte“ Rate. Wir setzen

$$S_t = S_0\,(1+r_g)^{\sum_{s=1}^t Z_s}\,(1+r_b)^{t-\sum_{s=1}^t Z_s} = (1+r_g)^{Z_t}(1+r_b)^{1-Z_t}\,S_{t-1}.$$

Werden Anleihe und Aktie gemeinsam gehandelt und gilt $p \in (0,1)$, so muss zusätzlich $r_b < r < r_g$ gelten. Denn im Fall $r \leq r_b$ oder $r \geq r_g$ entsteht ein risikoloser Gewinn, das heißt eine *Arbitrage*: Im ersten Fall leiht man Geld zum Zins r und kauft davon die Aktie, im zweiten Fall verkauft man die Aktie und legt den Erlös zum Zins r an. Bei der ersten Strategie (mit Startkapital 0) beträgt der Gewinn nach einer Periode $(1+Z_1 r_g + (1-Z_1)r_b - (1+r))S_0 \geq 0$ und mit Wahrscheinlichkeit $p > 0$ wegen $r_g > r_b \geq r$ sogar strikt positiv; der zweite Fall ist symmetrisch, mit Wahrscheinlichkeit $1-p > 0$ für einen strikt positiven Gewinn.

8.1. Modell, Portfolios und Arbitrage

Definition 8.1 (Das Einperiodenmodell)**.** Das *Einperiodenmodell* mit $d+1$ Basisgütern besteht aus

(1) einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$;

(2) einem Vektor
$$\boldsymbol{S}_0 = (S_0^0, S_0^1, \ldots, S_0^d) \in \mathbb{R}^{d+1} \qquad \text{mit } S_0^0 > 0,$$
der Preise S_0^i des i-ten Basisguts zum Zeitpunkt 0;

(3) einem Zufallsvektor
$$\boldsymbol{S}_1 = (S_1^0, S_1^1, \ldots, S_1^d)\colon \Omega \to \mathbb{R}^{d+1},$$
wobei S_1^0 deterministisch und strikt positiv ist, also $S_1^0(\omega) = S_1^0 > 0$ für alle $\omega \in \Omega$. Das 0-te Basisgut ist also eine risikolose Anlage (*Numéraire*). Die S_1^i, $i = 1, \ldots, d$, modellieren den (zufälligen) Preis des i-ten Basisguts zum Zeitpunkt 1.

Häufig gilt $S_1^0 = (1+r)S_0^0$ mit einer Zinsrate $r > -1$. Außerdem betrachten wir oft die *diskontierten Basisgüter*

$$\tilde{S}_t^i \coloneqq \frac{S_t^i}{S_t^0}, \qquad i = 0, \ldots, d,\ t \in \{0,1\},$$

also das *diskontierte Modell*; insbesondere gilt $\tilde{S}_t^0 = 1$. Den Vektor der diskontierten Preise bezeichnen wir mit $\tilde{\boldsymbol{S}}_t = (\tilde{S}_t^0, \ldots, \tilde{S}_t^d)$.

Definition 8.2 (Portfolio)**.** Ein Vektor $\boldsymbol{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$ heißt *Portfolio* im Einperiodenmodell; dabei ist x_i die *Stückzahl* des i-ten Basisguts. Die Größe

$$V_t(\boldsymbol{x}) \coloneqq \sum_{i=0}^{d} x_i S_t^i = \boldsymbol{x}^\top \boldsymbol{S}_t, \qquad t = 0, 1,$$

ist das zum Portfolio $\boldsymbol{x}$ gehörige *Vermögen* zum Zeitpunkt t.

Bemerkung 8.3. (1) Die Stückzahlen x_i müssen nicht ganzzahlig sein.

(2) Die Stückzahlen x_i können negativ sein. Dabei bedeutet $x_0 < 0$ eine Kreditaufnahme in Höhe von $|x_0|S_0^0$ zum Zeitpunkt 0, die zum Zeitpunkt 1 eine Tilgung in Höhe von $(1+r)S_0^0|x_0| = |x_0|S_1^0$ erfordert. Für $i = 1, \ldots, d$ bedeutet $x_i < 0$, dass man $|x_i|$ Einheiten des i-ten Basisguts verkauft, ohne sie zu besitzen; man tätigt einen *Leerverkauf* (*short sale*) beziehungsweise geht eine *Short-Position* ein. Zum Zeitpunkt 1 muss man den Preis $|x_i|S_1^i$ bezahlen, um das verkaufte Gut liefern zu können. Mit einem Kauf ($x_i > 0$) geht man dagegen zum Zeitpunkt 0 eine sogenannte *Long-Position* ein.

Definition 8.4 (Arbitrage)**.** Ein Portfolio $\boldsymbol{x} \in \mathbb{R}^{d+1}$ im Einperiodenmodell heißt *Arbitragemöglichkeit*, falls gilt:

(i) $\boldsymbol{x}^\top \boldsymbol{S}_0 = V_0(\boldsymbol{x}) \leq 0$;

(ii) $\boldsymbol{x}^\top \boldsymbol{S}_1 = V_1(\boldsymbol{x}) \geq 0$ $\mathbb{P}$-fast sicher und $\mathbb{P}(\boldsymbol{x}^\top \boldsymbol{S}_1 > 0) > 0$.

Gibt es keine Arbitragemöglichkeit, so heißt das Modell *arbitragefrei*.

Lemma 8.5. *Wir betrachten das Einperiodenmodell mit $d+1$ Basisgütern $(S_t^0, \ldots, S_t^d)$ und die zugehörigen diskontierten Basisgüter $(\tilde{S}_t^0, \ldots, \tilde{S}_t^d)$. Sei $\boldsymbol{x} \in \mathbb{R}^{d+1}$ ein Portfolio. Dann ist $\boldsymbol{x}$ genau dann eine Arbitragemöglichkeit, wenn $\boldsymbol{x}$ eine Arbitragemöglichkeit im diskontierten Modell ist. Genauer sind die folgenden Aussagen äquivalent:*

(1) Es existiert eine Arbitragemöglichkeit.

(2) Es existiert eine Arbitragemöglichkeit im diskontierten Modell.

(3) Es existieren $c \in \mathbb{R}$ und ein Portfolio $\boldsymbol{y} \in \mathbb{R}^{d+1}$ mit

$$\boldsymbol{y}^\top \tilde{\boldsymbol{S}}_0 = c, \qquad \boldsymbol{y}^\top \tilde{\boldsymbol{S}}_1 \geq c \quad \mathbb{P}\text{-f.s.} \qquad \text{und} \qquad \mathbb{P}(\boldsymbol{y}^\top \tilde{\boldsymbol{S}}_1 > c) > 0. \tag{8.1}$$

(4) Für alle $c \in \mathbb{R}$ existiert ein $\boldsymbol{y} \in \mathbb{R}^{d+1}$, sodass (8.1) gilt.

Beweis. Zur Übung. □

Beispiel 8.6. Eine *europäische Put-Option* ist das Recht, aber nicht die Pflicht, zu einem vereinbarten Fälligkeitszeitpunkt ein spezifiziertes Basisgut zu einem festgelegten Ausübungspreis (*strike price*) K zu *verkaufen*. Ein Put legt also das Basisgut auf den Markt; die Idee dahinter sind Gewinnmöglichkeiten und Absicherung. Eine *europäische Call-Option* ist das Recht, aber nicht die Pflicht, zu einem vereinbarten Fälligkeitszeitpunkt ein spezifiziertes Basisgut zum Ausübungspreis K zu *kaufen*. Ein Call nimmt also das Basisgut vom Markt; auch hier sind Gewinn und Absicherung die Idee.

Wir betrachten als Basisgüter ein Konto mit Zinsrate r, eine Aktie mit Preisen S_0^1 und S_1^1 sowie einen Call und einen Put auf die Aktie mit Fälligkeit 1 und demselben Strike K. Wir vergleichen die folgenden beiden Portfolios:

- Portfolio I: Wir erwerben die Call-Option zum Strike K und den abgezinsten Betrag $K/(1+r)$ auf dem Konto.
- Portfolio II: Wir erwerben die Put-Option zum Strike K und die Aktie zum Preis S_0^1.

Zum Zeitpunkt 1 ergibt sich folgendes Vermögen der beiden Portfolios:

Szenario	Vermögen von Portfolio I	Vermögen von Portfolio II
$S_1^1 > K$	übe Call aus: S_1^1	übe Put nicht aus: S_1^1
$S_1^1 \leq K$	übe Call nicht aus: K	übe Put aus: K

Beide Portfolios haben also zum Zeitpunkt 1 in jedem Szenario den gleichen Wert $\max\{S_1^1, K\}$ (siehe Abbildung 8.1). Unter welcher Bedingung folgt aus dem gleichen Wert zum Zeitpunkt 1 auch der gleiche Wert zum Zeitpunkt 0, das heißt

$$\text{Preis Call}(1,K) + \frac{K}{1+r} = \text{Preis Put}(1,K) + S_0^1 \,?$$

Angenommen, es gilt

$$\text{Preis Call}(1,K) + \frac{K}{1+r} > \text{Preis Put}(1,K) + S_0^1.$$

Dann kaufen wir zum Zeitpunkt 0 den Put $\text{Put}(1,K)$ und die Aktie zum Preis S_0^1, verkaufen den Call $\text{Call}(1,K)$ und leihen uns den Betrag $K/(1+r)$ zur Zinsrate r. Dabei entsteht ein Profit $c > 0$, nämlich die Differenz der beiden Seiten. Zum Zeitpunkt $t = 1$ unterscheiden wir zwei Fälle. Falls $S_1^1 > K$, übt der Käufer den Call aus; wir geben die Aktie ab und erhalten K, das wir zur Tilgung des Darlehens nutzen. Den Put lassen wir verfallen. Falls $S_1^1 \leq K$, übt der Käufer den Call nicht aus, aber wir üben den Put aus und erhalten K, das wir zur Tilgung verwenden. In beiden Fällen stehen wir zum Zeitpunkt 1 mit leeren Händen da und haben den risikolosen Gewinn $c > 0$ erzielt; dies ist eine Arbitragemöglichkeit. Legen wir c zum Zeitpunkt 0 auf dem Konto an, so ist das Anfangsvermögen 0 und das Endvermögen $(1+r)c > 0$, also liegt eine Arbitragemöglichkeit im Sinne von Definition 8.4 vor. Ebenso erhalten wir eine Arbitragemöglichkeit, falls

$$\text{Preis Call}(1,K) + \frac{K}{1+r} < \text{Preis Put}(1,K) + S_0^1.$$

In einem arbitragefreien Modell gilt also die *Put-Call-Parität*

$$\boxed{\text{Preis Call}(1,K) + \frac{K}{1+r} = \text{Preis Put}(1,K) + S_0^1.}$$

8.2. Der erste Hauptsatz der Preistheorie

Definition 8.7 (Risikoneutrale Maße)**.** Wir betrachten das Einperiodenmodell mit $d+1$ Basisgütern $S^0, S^1, \ldots, S^d$ und Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) Ein Wahrscheinlichkeitsmaß Q auf $(\Omega, \mathcal{F})$ heißt *risikoneutrales Maß* oder *Martingalmaß*, falls

$$\mathbb{E}_Q\Big[\Big|\frac{S_1^i}{S_1^0}\Big|\Big] < \infty \qquad \text{und} \qquad \mathbb{E}_Q\Big[\frac{S_1^i}{S_1^0}\Big] = \frac{S_0^i}{S_0^0} \qquad \text{für alle } i = 0, \ldots, d,$$

also $\mathbb{E}_Q[\tilde{S}_1^i] = \tilde{S}_0^i$ für alle $i = 0, \ldots, d$.

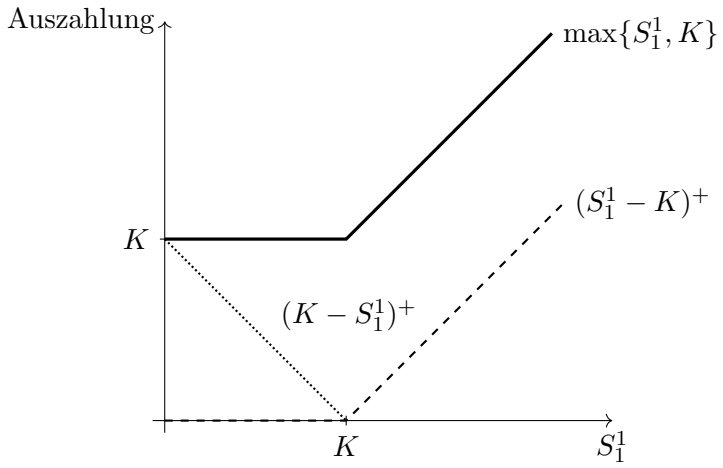

Abbildung 8.1.: Auszahlungen von Call und Put zum Strike K. Portfolio I liefert $(S_1^1 - K)^+ + K$, Portfolio II liefert $(K - S_1^1)^+ + S_1^1$; beide stimmen mit $\max\{S_1^1, K\}$ überein.

(b) Gilt zusätzlich $Q \ll \mathbb{P}$ sowie $\mathbb{P} \ll Q$, so heißt Q ein (zu $\mathbb{P}$) *äquivalentes risikoneutrales Maß*.

Satz 8.8 (Erster Hauptsatz der Preistheorie [HK79, HP81])**.** *Für das Einperiodenmodell sind die beiden folgenden Aussagen äquivalent:*

(i) Das Modell ist arbitragefrei.

(ii) Es existiert ein äquivalentes risikoneutrales Maß.

Gilt (i) oder (ii), so existiert sogar ein äquivalentes risikoneutrales Maß Q, dessen Dichte $\mathrm{d}Q/\mathrm{d}\mathbb{P}$ *(nach dem Satz von Radon–Nikodým) beschränkt ist.*

In der Literatur findet man diesen Satz zum Beispiel in [FS02, Abschnitt 1.2].

Beweis von (ii)⇒(i). Wir führen einen Widerspruchsbeweis. Seien Q ein äquivalentes Martingalmaß und $\boldsymbol{x} \in \mathbb{R}^{d+1}$ eine Arbitragemöglichkeit. Nach Lemma 8.5 ist $\boldsymbol{x}$ auch eine Arbitragemöglichkeit im diskontierten Modell, also gilt $\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_0 \le 0$, $\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1 \ge 0$ $\mathbb{P}$-fast sicher und $\mathbb{P}(\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1 > 0) > 0$; insbesondere ist $\mathbb{E}_{\mathbb{P}}[\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1] > 0$. Wegen $\mathbb{P} \ll Q$ gilt $Q(\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1 > 0) > 0$, und wegen $Q \ll \mathbb{P}$ gilt $Q(\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1 \ge 0) = 1$. Daraus folgt $\mathbb{E}_Q[\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1] > 0$. Da Q ein Martingalmaß ist, erhalten wir

$$0 \ge \boldsymbol{x}^\top \tilde{\boldsymbol{S}}_0 = \boldsymbol{x}^\top \mathbb{E}_Q[\tilde{\boldsymbol{S}}_1] = \mathbb{E}_Q[\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1] > 0,$$

ein Widerspruch zur Existenz der Arbitragemöglichkeit. □

Für die Richtung (i)⇒(ii) müssen wir weiter ausholen; wir führen den Beweis in der nächsten Sitzung.

Für den Beweis der Richtung (i)⇒(ii) in Satz 8.8 benötigen wir einen Trennungssatz aus der konvexen Analysis.

Satz (Trennungssatz)**.** *Sei $\emptyset \ne C \subset \mathbb{R}^n$ konvex mit $0 \notin C$. Dann existiert ein $\boldsymbol{\eta} \in \mathbb{R}^n$ mit $\boldsymbol{\eta}^\top \boldsymbol{x} \ge 0$ für alle $\boldsymbol{x} \in C$ und $\boldsymbol{\eta}^\top \boldsymbol{x}_0 > 0$ für mindestens ein $\boldsymbol{x}_0 \in C$. Gilt zudem $\inf_{\boldsymbol{x} \in C} |\boldsymbol{x}| > 0$, so kann $\boldsymbol{\eta}$ so gewählt werden, dass $\inf_{\boldsymbol{x} \in C} \boldsymbol{\eta}^\top \boldsymbol{x} > 0$ gilt.*

Für den Beweis verweisen wir auf die konvexe Analysis, siehe etwa [FS02, Anhang A.1].

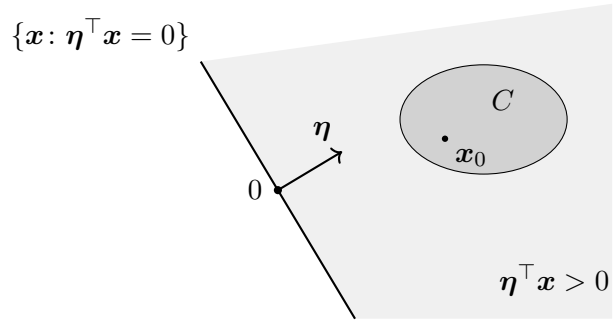

Abbildung 8.2.: Trennungssatz: Die konvexe Menge C mit $0 \notin C$ liegt im Halbraum $\{\boldsymbol{\eta}^\top \boldsymbol{x} \geq 0\}$, und mindestens ein Punkt $\boldsymbol{x}_0 \in C$ liegt nicht auf der trennenden Hyperebene.

Beweis von Satz 8.8, (i)⇒(ii). Für $i = 0, \dots, d$ setzen wir

$$Y^i \coloneqq \frac{S_1^i}{S_1^0} - \frac{S_0^i}{S_0^0} = \tilde{S}_1^i - \tilde{S}_0^i$$

und bezeichnen mit $\boldsymbol{Y} = (Y^0, Y^1, \dots, Y^d)^\top$, wobei $Y^0 = S_1^0/S_1^0 - S_0^0/S_0^0 = 0$ gilt, den Vektor der diskontierten Nettogewinne. Mit $|\cdot|$ bezeichnen wir die euklidische Norm.

Nach Annahme ist das Modell arbitragefrei. Mit Lemma 8.5 (3) gilt also für alle $\boldsymbol{x} \in \mathbb{R}^{d+1}$

$$\boldsymbol{x}^\top \boldsymbol{Y} \geq 0 \quad \mathbb{P}\text{-f.s.} \qquad \Longrightarrow \qquad \boldsymbol{x}^\top \boldsymbol{Y} = 0 \quad \mathbb{P}\text{-f.s.} \tag{8.2}$$

Denn wäre $\boldsymbol{x}^\top \boldsymbol{Y} \geq 0$ fast sicher und $\boldsymbol{x}^\top \boldsymbol{Y} > 0$ mit positiver Wahrscheinlichkeit, so erfüllten $\boldsymbol{y} = \boldsymbol{x}$ und $c = \boldsymbol{x}^\top \tilde{\boldsymbol{S}}_0$ die Bedingung aus Lemma 8.5 (3), denn $\mathbb{P}(\boldsymbol{x}^\top \tilde{\boldsymbol{S}}_1 > c) = \mathbb{P}(\boldsymbol{x}^\top \boldsymbol{Y} > 0) > 0$; dies ist ein Widerspruch zur Arbitragefreiheit.

Unser Ziel ist es, ein zu $\mathbb{P}$ äquivalentes Maß Q mit beschränkter Dichte $\mathrm{d}Q/\mathrm{d}\mathbb{P}$ zu finden, für das

$$\mathbb{E}_Q\big[|\boldsymbol{Y}|\big] < \infty \qquad \text{und} \qquad \mathbb{E}_Q[\boldsymbol{Y}] = 0 \tag{8.3}$$

gilt; ein solches Q ist ein äquivalentes risikoneutrales Maß. Wir behandeln hier nur den Fall $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{Y}|] < \infty$ (für den allgemeinen Fall siehe die Bemerkung im Anschluss an den Beweis). Wir setzen

$$\mathcal{A} \coloneqq \Big\{ Q \text{ Wahrscheinlichkeitsmaß auf } (\Omega, \mathcal{F})\colon Q \ll \mathbb{P},\ \mathbb{P} \ll Q,\ \tfrac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}} \text{ beschränkt} \Big\}$$

und

$$\mathcal{C} \coloneqq \Big\{ \mathbb{E}_Q[\boldsymbol{Y}] = \int \boldsymbol{Y} \frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}\, \mathrm{d}\mathbb{P}\colon Q \in \mathcal{A} \Big\} \subset \mathbb{R}^{d+1},$$

wobei für $Q \in \mathcal{A}$ mit $\mathrm{d}Q/\mathrm{d}\mathbb{P} \leq \alpha$

$$\mathbb{E}_Q\big[|\boldsymbol{Y}|\big] = \int |\boldsymbol{Y}|\, \mathrm{d}Q = \int |\boldsymbol{Y}| \frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}\, \mathrm{d}\mathbb{P} \leq \alpha\, \mathbb{E}_{\mathbb{P}}\big[|\boldsymbol{Y}|\big] < \infty$$

gilt. Wir beachten, dass (8.3) für ein $Q \in \mathcal{A}$ genau dann gilt, wenn $0 \in \mathcal{C}$ ist. Weiter ist $\mathcal{C}$ konvex, denn für alle $Q_1, Q_2 \in \mathcal{A}$ und $\lambda \in [0,1]$ gilt $\lambda Q_1 + (1-\lambda) Q_2 \in \mathcal{A}$ und

$$\mathbb{E}_{\lambda Q_1 + (1-\lambda) Q_2}[\boldsymbol{Y}] = \lambda \mathbb{E}_{Q_1}[\boldsymbol{Y}] + (1-\lambda) \mathbb{E}_{Q_2}[\boldsymbol{Y}] \in \mathcal{C}.$$

Angenommen, es gilt $0 \notin \mathcal{C}$. Aufgrund des Trennungssatzes existieren dann $\boldsymbol{x} \in \mathbb{R}^{d+1}$ und $\boldsymbol{c}_0 \in \mathcal{C}$ mit $\boldsymbol{x}^\top \boldsymbol{c} \geq 0$ für alle $\boldsymbol{c} \in \mathcal{C}$ und $\boldsymbol{x}^\top \boldsymbol{c}_0 > 0$. Also gilt

$$\mathbb{E}_Q[\boldsymbol{x}^\top \boldsymbol{Y}] = \boldsymbol{x}^\top \underbrace{\mathbb{E}_Q[\boldsymbol{Y}]}_{\in \mathcal{C}} \geq 0 \qquad \text{für alle } Q \in \mathcal{A}, \tag{8.4}$$

und es existiert ein $Q_0 \in \mathcal{A}$ mit $\mathbb{E}_{Q_0}[\boldsymbol{Y}] = \boldsymbol{c}_0$ (als Vektor), also mit $\mathbb{E}_{Q_0}[\boldsymbol{x}^\top \boldsymbol{Y}] > 0$. Daher gilt $\boldsymbol{x}^\top \boldsymbol{Y} > 0$ mit positiver Q_0-Wahrscheinlichkeit und wegen $Q_0 \ll \mathbb{P}$ auch $\mathbb{P}(\boldsymbol{x}^\top \boldsymbol{Y} > 0) > 0$.

Es genügt daher zu zeigen, dass $\boldsymbol{x}^\top \boldsymbol{Y} \geq 0$ $\mathbb{P}$-fast sicher gilt, denn dies wäre ein Widerspruch zu (8.2). Dafür sei

$$A \coloneqq \{\omega \in \Omega \colon \boldsymbol{x}^\top \boldsymbol{Y}(\omega) < 0\},$$

und wir approximieren $\mathbb{1}_A$ durch

$$f_n \coloneqq \Big(1 - \frac{1}{n}\Big)\mathbb{1}_A + \frac{1}{n}\mathbb{1}_{A^\mathrm{c}}, \qquad n \geq 2.$$

Wir setzen

$$Q_n \colon \mathcal{F} \ni B \mapsto Q_n(B) \coloneqq \int \frac{f_n(\omega)}{\mathbb{E}_{\mathbb{P}}[f_n]} \mathbb{1}_B(\omega)\, \mathbb{P}(\mathrm{d}\omega),$$

wobei $0 < \mathbb{E}_{\mathbb{P}}[f_n] < \infty$ gilt. Dann ist Q_n ein Wahrscheinlichkeitsmaß. Wegen $1/n \leq f_n \leq 1 - 1/n$ ist die Dichte $\mathrm{d}Q_n/\mathrm{d}\mathbb{P}$ beschränkt und strikt positiv, also gilt $Q_n \in \mathcal{A}$. Mit (8.4) folgt

$$0 \leq \mathbb{E}_{Q_n}\big[\boldsymbol{x}^\top \boldsymbol{Y}\big] = \frac{1}{\mathbb{E}_{\mathbb{P}}[f_n]} \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{x}^\top \boldsymbol{Y}\, f_n\big] \qquad \text{für alle } n \geq 2.$$

Da $|\boldsymbol{x}^\top \boldsymbol{Y}\, f_n| \leq |\boldsymbol{x}^\top \boldsymbol{Y}|$ und $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{x}^\top \boldsymbol{Y}|] \leq |\boldsymbol{x}|\, \mathbb{E}_{\mathbb{P}}[|\boldsymbol{Y}|] < \infty$ gelten, liefert der Satz von Lebesgue

$$0 \leq \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{x}^\top \boldsymbol{Y}\, f_n\big] = \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{x}^\top \boldsymbol{Y} \lim_{n\to\infty} f_n\Big] = \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{x}^\top \boldsymbol{Y}\, \mathbb{1}_A\big] = \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{x}^\top \boldsymbol{Y}\, \mathbb{1}\{\boldsymbol{x}^\top \boldsymbol{Y} < 0\}\big] \leq 0,$$

wobei im letzten Schritt Gleichheit genau dann gilt, wenn $\mathbb{P}(\boldsymbol{x}^\top \boldsymbol{Y} < 0) = 0$ ist. Also gilt $\mathbb{P}(\boldsymbol{x}^\top \boldsymbol{Y} < 0) = 0$, das heißt $\mathbb{P}(\boldsymbol{x}^\top \boldsymbol{Y} \geq 0) = 1$. Damit ist $0 \in \mathcal{C}$, und es existiert ein $Q \in \mathcal{A}$ mit (8.3). □

Bemerkung. Der Fall $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{Y}|] = \infty$ lässt sich auf den behandelten Fall zurückführen. Dazu definieren wir ein Wahrscheinlichkeitsmaß $\mathbb{P}'$ durch $\mathrm{d}\mathbb{P}'/\mathrm{d}\mathbb{P} \coloneqq \kappa/(1 + |\boldsymbol{Y}|)$, wobei $\kappa > 0$ eine Normierungskonstante ist. Dann sind $\mathbb{P}$ und $\mathbb{P}'$ äquivalent, es gilt $\mathbb{E}_{\mathbb{P}'}[|\boldsymbol{Y}|] < \infty$, und das Modell ist auch unter $\mathbb{P}'$ arbitragefrei, da Definition 8.4 nur von den Nullmengen des Maßes abhängt. Der obige Beweis liefert ein zu $\mathbb{P}'$ äquivalentes Maß Q mit beschränkter Dichte $\mathrm{d}Q/\mathrm{d}\mathbb{P}'$ und (8.3). Dann ist Q auch zu $\mathbb{P}$ äquivalent, und $\mathrm{d}Q/\mathrm{d}\mathbb{P} = \mathrm{d}Q/\mathrm{d}\mathbb{P}' \cdot \kappa/(1 + |\boldsymbol{Y}|)$ ist beschränkt.

8.3. Preisbestimmung von Kontrakten

Definition 8.9 (Kontrakt)**.** Sei S^0 ein festverzinsliches Konto mit Rate r im Einperiodenmodell auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, also $S_1^0 = (1 + r)S_0^0$. Ein *Kontrakt* (auch *Claim*) ist eine $\mathcal{F}$-messbare Zufallsvariable $\xi \colon \Omega \to \mathbb{R}$, welche die Auszahlung repräsentiert, die der Käufer des Kontrakts zum Zeitpunkt $t = 1$ erhält. Gilt $\mathbb{P}(\xi \geq 0) = 1$, so heißt ξ ein *bedingtes Finanzgut* (oder *contingent claim*). Die Menge aller Kontrakte bezeichnen wir mit $L^0(\Omega, \mathcal{F}, \mathbb{P})$.

Beispiel 8.10. (1) Die europäische Put-Option zum Zeitpunkt 1 mit Strike K, bezogen auf das Basisgut S^i, ist der Kontrakt

$$\xi \coloneqq \mathrm{Put}(1, K, i) \coloneqq (K - S_1^i)^+.$$

(2) Die europäische Call-Option $\mathrm{Call}(1, K, i) \coloneqq (S_1^i - K)^+$ ist ein Kontrakt.

(3) Ein *Forward* (Kontrakt) mit Fälligkeit 1 und Ausübungspreis (*Forwardpreis*) K: Der Käufer des Forwards *muss* zum Zeitpunkt 1 das Basisgut S^i zum Preis K kaufen. Der zugehörige Kontrakt ist

$$\xi = \mathrm{Fwd}(1, K, i) \coloneqq S_1^i - K.$$

Es stellt sich die Frage: Was ist ein *fairer Preis* für einen Kontrakt zum Zeitpunkt $t = 0$?

Definition 8.11 (Derivat)**.** (i) Ein Kontrakt ξ heißt *Derivat* (der Basisgüter $(S^0, \dots, S^d)$), falls ξ messbar bezüglich $\sigma(S_1^0, \dots, S_1^d)$ ist, also

$$\xi = f(S_1^0, \dots, S_1^d) \qquad \text{für eine messbare Funktion } f\colon \mathbb{R}^{d+1} \to \mathbb{R}.$$

(ii) Ein Kontrakt ξ heißt *replizierbar* oder *absicherbar* im Einperiodenmodell, falls ein $\boldsymbol{x} \in \mathbb{R}^{d+1}$ existiert mit

$$V_1(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{S}_1 = \xi \qquad \mathbb{P}\text{-f.s.}$$

Ein solches Portfolio $\boldsymbol{x}$ heißt *perfekter Hedge* für ξ. Den Raum aller replizierbaren Kontrakte bezeichnen wir mit

$$\mathcal{H} \coloneqq \{\xi \text{ Zufallsvariable}\colon \text{es existiert } \boldsymbol{x} \in \mathbb{R}^{d+1} \text{ mit } V_1(\boldsymbol{x}) = \xi \ \mathbb{P}\text{-f.s.}\}.$$

Bemerkung 8.12. (1) Derivate hängen oft nur von wenigen Basisgütern ab, ihren sogenannten *Underlyings*. Europäische Put- und Call-Optionen sind Derivate.

(2) Die Verkäuferin eines replizierbaren Kontrakts ξ muss die Auszahlung ξ zum Zeitpunkt $t = 1$ garantieren. Dafür kauft sie das Portfolio $\boldsymbol{x}$ eines perfekten Hedges. Ein fairer Preis des Kontrakts scheint also $V_0(\boldsymbol{x})$ zu sein.

Definition und Satz 8.13 (Law of one price)**.** *Wir betrachten ein arbitragefreies Einperiodenmodell mit $d+1$ Basisgütern. Seien $\xi \in \mathcal{H}$ und $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{d+1}$ zwei perfekte Hedges für ξ. Dann gilt das* Law of one price

$$V_0(\boldsymbol{x}) = V_0(\boldsymbol{y}).$$

Man nennt $\hat{\pi}(\xi) \coloneqq V_0(\boldsymbol{x})$, wobei $\boldsymbol{x}$ ein perfekter Hedge für ξ ist, den Hedgepreis *von ξ. Für jedes äquivalente Martingalmaß Q gilt $\mathbb{E}_Q[|\xi|] < \infty$ und*

$$\hat{\pi}(\xi) = \frac{S_0^0}{S_1^0} \mathbb{E}_Q[\xi]. \tag{8.5}$$

Beweis. Die Arbitragefreiheit impliziert nach Satz 8.8 die Existenz eines äquivalenten Martingalmaßes Q. Sei Q ein solches. Da $\mathbb{E}_Q[|S_1^i|] < \infty$ für alle $i = 0, \dots, d$ gilt, folgt $\mathbb{E}_Q[|\boldsymbol{x}^\top \boldsymbol{S}_1|] < \infty$ für alle $\boldsymbol{x} \in \mathbb{R}^{d+1}$ und damit $\mathbb{E}_Q[|\xi|] < \infty$ für alle replizierbaren Kontrakte ξ. Seien $\boldsymbol{x}, \boldsymbol{y}$ zwei perfekte Hedges für ξ. Dann gilt

$$\begin{aligned}
V_0(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{S}_0 &= S_0^0 \, \boldsymbol{x}^\top \tilde{\boldsymbol{S}}_0 && \text{(Diskontierung)} \\
&= S_0^0 \, \boldsymbol{x}^\top \mathbb{E}_Q[\tilde{\boldsymbol{S}}_1] && \text{(Martingalmaßeigenschaft)} \\
&= \frac{S_0^0}{S_1^0} \mathbb{E}_Q[\boldsymbol{x}^\top \boldsymbol{S}_1] \\
&= \frac{S_0^0}{S_1^0} \mathbb{E}_Q[\xi] && \text{(Hedging).}
\end{aligned}$$

Dieselbe Rechnung für $\boldsymbol{y}$ ergibt $V_0(\boldsymbol{y}) = (S_0^0/S_1^0)\mathbb{E}_Q[\xi] = V_0(\boldsymbol{x})$. □

Beispiel 8.14. Sei $S_0^0 = 1$, also $S_1^0 = 1 + r$. Wir betrachten den Forward $\xi = \mathrm{Fwd}(1, K, i) = S_1^i - K$. Er ist replizierbar, denn für das Portfolio

$$x_j \coloneqq \begin{cases} -\frac{K}{1+r}, & j = 0 \quad (\text{leihe } \frac{K}{1+r}), \\ 1, & j = i \quad (\text{kaufe eine Einheit des Basisguts } S^i), \\ 0, & \text{sonst}, \end{cases}$$

gilt

$$V_1(\boldsymbol{x}) = -\frac{K}{1+r} S_1^0 + S_1^i = -K + S_1^i = \xi,$$

wobei

$$\hat{\pi}(\xi) = V_0(\boldsymbol{x}) = S_0^i - \frac{K}{1+r}.$$

Alternativ erhalten wir mit (8.5)

$$\begin{aligned} \hat{\pi}(\xi) &= \frac{S_0^0}{S_1^0} \mathbb{E}_Q[\xi] = \frac{1}{1+r} \mathbb{E}_Q\big[S_1^i - K\big] = \mathbb{E}_Q\Big[\frac{S_1^i}{1+r} - \frac{K}{1+r}\Big] \\ &= \mathbb{E}_Q\big[\tilde{S}_1^i\big] - \frac{K}{1+r} = \tilde{S}_0^i - \frac{K}{1+r} = S_0^i - \frac{K}{1+r}, \end{aligned}$$

wobei wir im vorletzten Schritt die Martingalmaßeigenschaft und im letzten Schritt $S_0^0 = 1$ benutzt haben.

Beispiel 8.15. Wir betrachten eine Aktie mit zwei Zuständen, also $d = 1$, mit $S_0^0 = 1$ (und damit $S_1^0 = 1 + r$) und $S_0^1 = c > 0$ für eine Konstante c sowie

$$S_1^1 = \begin{cases} a & \text{mit Wahrscheinlichkeit } p, \\ b & \text{mit Wahrscheinlichkeit } 1 - p, \end{cases} \qquad p \in (0,1),\ a < b.$$

Der einfachste Wahrscheinlichkeitsraum ist hier $\Omega = \{0, 1\}$, $\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ mit $\mathbb{P}(\{0\}) = p$, $\mathbb{P}(\{1\}) = 1 - p$ sowie $S_1^1(0) = a$ und $S_1^1(1) = b$ (siehe Abbildung 8.3).

Wir suchen ein äquivalentes Martingalmaß Q. Es gilt $Q \ll \mathbb{P}$ und $\mathbb{P} \ll Q$ genau dann, wenn $Q(\{0\}) = q \in (0, 1)$ ist. Zudem gilt $\mathbb{E}_Q[S_1^0/S_1^0] = 1 = S_0^0/S_0^0$, und damit ist Q genau dann ein Martingalmaß, wenn

$$\mathbb{E}_Q\Big[\frac{S_1^1}{S_1^0}\Big] = \mathbb{E}_Q\Big[\frac{S_1^1}{1+r}\Big] \overset{!}{=} \frac{S_0^1}{S_0^0} = S_0^1 = c$$

gilt, also

$$c(1+r) = \mathbb{E}_Q\big[S_1^1\big] = aq + b(1-q).$$

Diese Gleichung ist genau dann mit $q \in (0, 1)$ lösbar, wenn $c \in \big(a/(1+r), b/(1+r)\big)$ gilt, und dann ist

$$q = \frac{b - c(1+r)}{b - a}.$$

Sei nun ξ ein beliebiger Kontrakt mit $\xi(0) = \alpha$ und $\xi(1) = \beta$. Da $S_1^1(\omega) = a + (b-a)\mathbb{1}\{\omega = 1\}$ für $\omega \in \Omega$ gilt, folgt

$$\xi(\omega) = \alpha + (\beta - \alpha)\mathbb{1}\{\omega = 1\} = \alpha + (\beta - \alpha)\frac{S_1^1(\omega) - a}{b - a} = \frac{1}{1+r}\,\frac{\alpha b - a\beta}{b - a}\,S_1^0 + \frac{\beta - \alpha}{b - a}\,S_1^1(\omega).$$

Also ist

$$\boldsymbol{x} = \Big(\frac{1}{1+r}\,\frac{\alpha b - a\beta}{b - a},\ \frac{\beta - \alpha}{b - a}\Big) \in \mathbb{R}^2$$

ein perfekter Hedge. Insbesondere sind hier alle Kontrakte replizierbar, und es gilt

$$\hat{\pi}(\xi) = V_0(\boldsymbol{x}) = \frac{1}{1+r}\,\frac{\alpha b - a\beta}{b-a}\cdot 1 + \frac{\beta-\alpha}{b-a}\,c.$$

Hätten wir dies bereits gewusst, so hätten wir den Hedgepreis auch direkt mit (8.5) berechnen können:

$$\begin{aligned}\hat{\pi}(\xi) &= \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi] = \frac{1}{1+r}\mathbb{E}_Q[\xi] = \frac{1}{1+r}\big(q\alpha + (1-q)\beta\big)\\ &= \frac{1}{1+r}\Big(\frac{b-c(1+r)}{b-a}\,\alpha + \frac{c(1+r)-a}{b-a}\,\beta\Big) = \frac{1}{1+r}\,\frac{\alpha b - a\beta}{b-a} + \frac{\beta-\alpha}{b-a}\,c.\end{aligned}$$

Übung: Berechnen Sie den Hedgepreis von europäischen Put- und Call-Optionen.

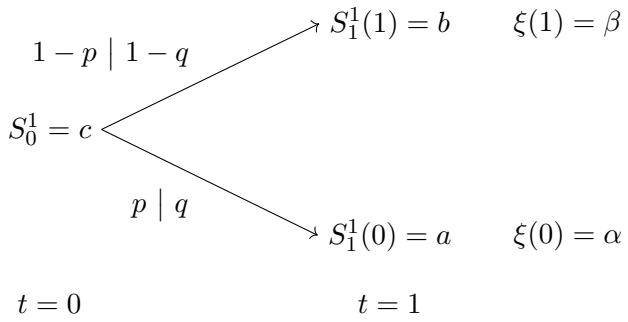

Abbildung 8.3.: Das Einperiodenmodell mit zwei Zuständen aus Beispiel 8.15. An den Kanten stehen die Wahrscheinlichkeiten unter $\mathbb{P}$ und unter dem äquivalenten Martingalmaß Q.

8.3.1. Vollständigkeit und arbitragefreie Preise

Wir betrachten weiterhin ein Einperiodenmodell mit $d+1$ Basisgütern $S^0, \dots, S^d$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, wobei S^0 ein festverzinsliches Konto mit Zinsrate r ist, also $S_1^0 = (1+r)S_0^0$. Wie in Definition 8.11 bezeichnet $\mathcal{H}$ den Raum der replizierbaren Kontrakte, und für $\xi \in \mathcal{H}$ ist $\hat{\pi}(\xi)$ der Hedgepreis aus Definition und Satz 8.13.

Definition 8.16 (Vollständigkeit des Einperiodenmodells)**.** Ein Einperiodenmodell heißt *vollständig*, falls jeder Kontrakt replizierbar ist.

Im Allgemeinen ist ein Einperiodenmodell nicht vollständig. Damit stellt sich die Frage, wie der Preis eines Kontrakts $\xi \notin \mathcal{H}$ bestimmt werden soll. Die Idee ist, ξ als neues Basisgut aufzufassen und dessen Preis so zu bestimmen, dass die Arbitragefreiheit nicht verletzt wird. Für replizierbare Kontrakte führt diese Idee auf den bekannten Hedgepreis.

Satz 8.17 (Hedgepreis und Arbitragefreiheit)**.** *Gegeben sei ein arbitragefreies Einperiodenmodell mit $d+1$ Basisgütern $S^0, S^1, \dots, S^d$. Sei $\xi \in \mathcal{H}$, und für eine Konstante $a \in \mathbb{R}$ setzen wir*

$$S_1^{d+1}(\omega) \coloneqq \xi(\omega) \quad \textit{und} \quad S_0^{d+1}(\omega) \coloneqq a \qquad \textit{für alle } \omega \in \Omega.$$

Dann ist das um S^{d+1} erweiterte Einperiodenmodell genau dann arbitragefrei, wenn $a = \hat{\pi}(\xi)$ gilt, also wenn a der Hedgepreis von ξ ist.

Beweis. „⇐“: Sei Q ein äquivalentes Martingalmaß (ÄMM) im Modell $(S^0, S^1, \dots, S^d)$; ein solches existiert nach dem ersten Hauptsatz der Preistheorie, Satz 8.8. Mit (8.5) gilt

$$\mathbb{E}_Q[\tilde{S}_1^{d+1}] = \frac{1}{S_1^0}\mathbb{E}_Q[\xi] = \frac{1}{S_0^0}\hat{\pi}(\xi) = \frac{a}{S_0^0} = \tilde{S}_0^{d+1}.$$

Also ist Q auch ein ÄMM im erweiterten Modell, und dieses ist nach Satz 8.8 arbitragefrei.

„⇒“: Wegen $\xi \in \mathcal{H}$ existiert ein $\boldsymbol{x} \in \mathbb{R}^{d+1}$ mit $V_1(\boldsymbol{x}) = \xi$ $\mathbb{P}$-fast sicher. Dann sind $\boldsymbol{x}' \coloneqq (\boldsymbol{x}, 0) \in \mathbb{R}^{d+2}$ und $\boldsymbol{y}' \coloneqq (\boldsymbol{0}, 1) \in \mathbb{R}^{d+2}$ zwei perfekte Hedges für ξ im erweiterten Modell. Da dieses nach Voraussetzung arbitragefrei ist, liefert das Law of One Price, Definition und Satz 8.13, angewandt auf den Preisvektor $\boldsymbol{S}_0 = (S_0^0, \dots, S_0^{d+1})^\top$ des erweiterten Modells,

$$\hat{\pi}(\xi) = V_0(\boldsymbol{x}) = (\boldsymbol{x}')^\top \boldsymbol{S}_0 = (\boldsymbol{y}')^\top \boldsymbol{S}_0 = S_0^{d+1} = a. \qquad \square$$

Vereinbarung 8.18 (No-Arbitrage-Prinzip)**.** Der Preis eines Kontrakts ξ in einem arbitragefreien Einperiodenmodell mit $d+1$ Basisgütern wird so festgelegt, dass das erweiterte Einperiodenmodell mit $d+2$ Basisgütern, wobei $S_1^{d+1} = \xi$ gilt und der Preis S_0^{d+1} konstant ist, arbitragefrei ist. Jeder solche Preis für ξ (im Allgemeinen ist er nicht eindeutig) heißt *arbitragefreier Preis* für ξ und wird mit $\pi(\xi)$ bezeichnet. Nach Satz 8.17 gilt für $\xi \in \mathcal{H}$ also $\pi(\xi) = \hat{\pi}(\xi)$.

Satz 8.19 (Intervalleigenschaft arbitragefreier Preise)**.** *Gegeben sei ein arbitragefreies Einperiodenmodell mit $d+1$ Basisgütern auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Dann ist die Menge $\Pi(\xi)$ aller arbitragefreien Preise $\pi(\xi)$ eines Kontrakts ξ ein nichtleeres Intervall, und es gilt*

$$\Pi(\xi) = \Big\{ \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi] \colon Q \text{ ÄMM in } (S^0, \dots, S^d) \text{ mit } \mathbb{E}_Q[|\xi|] < \infty \Big\}. \tag{8.6}$$

Definition 8.20 (Arbitrageschranken)**.** Die Intervallgrenzen

$$\pi_{\inf}(\xi) \coloneqq \inf \Pi(\xi) \quad \text{und} \quad \pi_{\sup}(\xi) \coloneqq \sup \Pi(\xi)$$

von (8.6) heißen *Arbitrageschranken* für ξ.

Beispiel 8.21 (Abschätzung der Arbitrageschranken für Call und Put)**.** Wir betrachten die europäische Call- beziehungsweise Put-Option auf das i-te Basisgut zum Strike $K > 0$ in einem arbitragefreien Einperiodenmodell. Es gelte $S_0^i \geq 0$, $S_1^i \geq 0$ $\mathbb{P}$-fast sicher und $S_1^0 = (1+r)S_0^0$. Wegen

$$\mathrm{Call}(1, K, i) = (S_1^i - K)^+ \leq S_1^i$$

folgt für jedes ÄMM Q

$$\frac{S_0^0}{S_1^0}\mathbb{E}_Q[\mathrm{Call}(1, K, i)] \leq \frac{S_0^0}{S_1^0}\mathbb{E}_Q[S_1^i] = S_0^i.$$

Andererseits ist die Abbildung $u \mapsto c(u-K)^+$ für $c \geq 0$ konvex, und mit der Jensen-Ungleichung folgt

$$\frac{S_0^0}{S_1^0}\mathbb{E}_Q[\mathrm{Call}(1, K, i)] = \mathbb{E}_Q\Big[\frac{S_0^0}{S_1^0}(S_1^i - K)^+\Big] \geq \frac{S_0^0}{S_1^0}(\mathbb{E}_Q[S_1^i] - K)^+ = \Big(S_0^i - \frac{K}{1+r}\Big)^+,$$

wobei wir $\mathbb{E}_Q[S_1^i] = (S_1^0/S_0^0)S_0^i$ verwendet haben. Mit (8.6) erhalten wir also

$$\Big(S_0^i - \frac{K}{1+r}\Big)^+ \leq \pi_{\inf}(\mathrm{Call}(1, K, i)) \leq \pi_{\sup}(\mathrm{Call}(1, K, i)) \leq S_0^i. \tag{8.7}$$

Analog gilt für den Put

$$\Big(\frac{K}{1+r} - S_0^i\Big)^+ \le \pi_{\inf}\big(\mathrm{Put}(1,K,i)\big) \le \pi_{\sup}\big(\mathrm{Put}(1,K,i)\big) \le \frac{K}{1+r}.$$

Ein Preis außerhalb dieser Schranken würde eine Arbitragemöglichkeit eröffnen; siehe Abbildung 8.4.

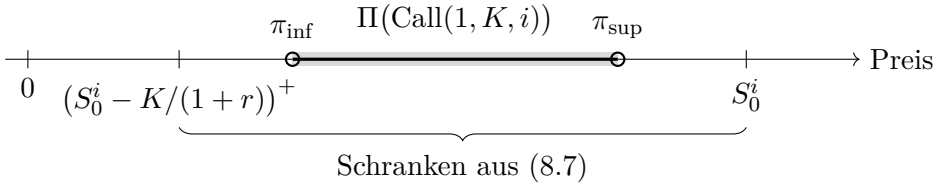

Abbildung 8.4.: Die Menge der arbitragefreien Preise eines Calls ist ein Intervall innerhalb der Schranken aus (8.7). Ist der Call nicht replizierbar, so ist das Intervall nach Satz 8.26 offen.

Beweis von Satz 8.19. „$\subseteq$“: Sei $\pi(\xi)$ ein arbitragefreier Preis. Dann ist das erweiterte Modell arbitragefrei, und im erweiterten Modell gilt $\xi \in \mathcal{H}'$, wobei $\mathcal{H}'$ den Raum der im erweiterten Modell replizierbaren Kontrakte bezeichnet. Folglich existiert ein ÄMM Q im erweiterten Modell, und mit Satz 8.17 und (8.5) gilt

$$\pi(\xi) = \hat{\pi}(\xi) = \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi],$$

wobei $\hat{\pi}$ hier den Hedgepreis im erweiterten Modell bezeichnet; insbesondere ist $\mathbb{E}_Q[|\xi|] < \infty$. Da Q auch ein ÄMM im ursprünglichen Modell ist, folgt „$\subseteq$“.

„$\supseteq$“: Sei Q ein ÄMM im ursprünglichen Modell mit $\mathbb{E}_Q[|\xi|] < \infty$. Wir setzen

$$S_1^{d+1} \coloneqq \xi \quad \text{und} \quad S_0^{d+1} \coloneqq \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi].$$

Dann gilt

$$\mathbb{E}_Q\Big[\frac{S_1^{d+1}}{S_1^0}\Big] = \frac{S_0^{d+1}}{S_0^0},$$

also ist Q auch ein ÄMM im erweiterten Modell. Nach Satz 8.8 ist dieses somit arbitragefrei, und $\pi(\xi) = (S_0^0/S_1^0)\mathbb{E}_Q[\xi]$ ist ein arbitragefreier Preis.

$\Pi(\xi) \neq \emptyset$: Das Maß $\tilde{\mathbb{P}}$ mit

$$\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}} = \frac{1}{1+|\xi|}\,\mathbb{E}_{\mathbb{P}}\Big[\frac{1}{1+|\xi|}\Big]^{-1}$$

ist ein zu $\mathbb{P}$ äquivalentes Wahrscheinlichkeitsmaß. Da $(S^0, S^1, \dots, S^d)$ arbitragefrei ist und Arbitragefreiheit nur von den Nullmengen abhängt, existiert nach Satz 8.8 ein ÄMM Q, das äquivalent zu $\tilde{\mathbb{P}}$ und damit zu $\mathbb{P}$ ist und eine beschränkte Dichte $f = \mathrm{d}Q/\mathrm{d}\tilde{\mathbb{P}}$ besitzt. Daraus folgt

$$\mathbb{E}_Q\big[|\xi|\big] = \mathbb{E}_{\mathbb{P}}\Big[\frac{1}{1+|\xi|}\Big]^{-1} \int_\Omega |\xi|\, f\, \frac{1}{1+|\xi|}\,\mathrm{d}\mathbb{P} < \infty,$$

und mit „$\supseteq$“ folgt $(S_0^0/S_1^0)\mathbb{E}_Q[\xi] \in \Pi(\xi)$.

Intervalleigenschaft: Seien Q_1, Q_2 zwei ÄMM in $(S^0, S^1, \dots, S^d)$ mit $\mathbb{E}_{Q_i}[|\xi|] < \infty$ für $i = 1, 2$. Für alle $\alpha \in [0, 1]$ ist dann auch $\alpha Q_1 + (1-\alpha)Q_2$ ein ÄMM, und es gilt

$$\mathbb{E}_{\alpha Q_1 + (1-\alpha)Q_2}\big[|\xi|\big] = \alpha \mathbb{E}_{Q_1}\big[|\xi|\big] + (1-\alpha)\mathbb{E}_{Q_2}\big[|\xi|\big] < \infty$$

sowie

$$\alpha \underbrace{\frac{S_0^0}{S_1^0}\mathbb{E}_{Q_1}[\xi]}_{\in \Pi(\xi)} + (1-\alpha) \underbrace{\frac{S_0^0}{S_1^0}\mathbb{E}_{Q_2}[\xi]}_{\in \Pi(\xi)} = \underbrace{\frac{S_0^0}{S_1^0}\mathbb{E}_{\alpha Q_1 + (1-\alpha)Q_2}[\xi]}_{\in \Pi(\xi)}.$$

Also ist $\Pi(\xi) \subseteq \mathbb{R}$ konvex und damit ein Intervall. □

Satz 8.22 (Monotonie von Kontrakten)**.** *Gegeben sei ein arbitragefreies Einperiodenmodell mit $d+1$ Basisgütern $S^0, S^1, \dots, S^d$ sowie Kontrakte $\xi_1, \dots, \xi_n$. Wir setzen*

$$S_1^{d+k} \coloneqq \xi_k \qquad (k = 1, \dots, n)$$

und wählen arbitragefreie Preise $\pi(\xi_k)$ im erweiterten Einperiodenmodell mit $d+n+1$ Basisgütern $S^0, \dots, S^{d+n}$. (Dies gelingt induktiv, und zwar zu einem gemeinsamen ÄMM Q.) Bezeichnet $\mathcal{H}'$ den Raum der im erweiterten Modell replizierbaren Kontrakte, so gilt für alle $j \neq l \in \{1, \dots, n\}$:

(i) Aus $\xi_j \leq \xi_l$ $\mathbb{P}$-fast sicher folgt $\pi(\xi_j) \leq \pi(\xi_l)$.

(ii) Es gilt $\xi_j + \xi_l \in \mathcal{H}'$ mit $\hat{\pi}(\xi_j + \xi_l) = \pi(\xi_j + \xi_l) = \pi(\xi_j) + \pi(\xi_l)$.

Beweis. Es existiert ein ÄMM Q im erweiterten Modell, für das

$$\pi(\xi_j) = \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi_j], \qquad \pi(\xi_l) = \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi_l] \qquad \text{und} \qquad \pi(\xi_j + \xi_l) = \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi_j + \xi_l]$$

gilt; dabei ist $\xi_j + \xi_l = S_1^{d+j} + S_1^{d+l}$ im erweiterten Modell replizierbar. Die Linearität und die Monotonie des Erwartungswerts implizieren die Behauptung. □

Vereinbarung 8.23 (No-Arbitrage-Prinzip für mehrere Kontrakte)**.** Die Preise von mehreren Kontrakten werden so festgelegt, dass das erweiterte Modell arbitragefrei ist.

Korollar 8.24 (Put-Call-Parität)**.** *Im arbitragefreien Modell gilt für die Preise von europäischen Put- und Call-Optionen auf das i-te Basisgut*

$$\pi\big(\mathrm{Call}(1, K, i)\big) = S_0^i - \frac{1}{1+r}K + \pi\big(\mathrm{Put}(1, K, i)\big).$$

Beweis. Es gilt

$$\mathrm{Call}(1, K, i) - \mathrm{Put}(1, K, i) = (S_1^i - K)^+ - (K - S_1^i)^+ = S_1^i - K.$$

Sei Q ein ÄMM des um Call und Put erweiterten Modells. Mit Satz 8.22 und (8.6) folgt

$$\pi(\mathrm{Call}) - \pi(\mathrm{Put}) = \pi(S_1^i) - \pi(K) = \frac{S_0^0}{S_1^0}\mathbb{E}_Q[S_1^i] - \frac{S_0^0}{S_1^0}\mathbb{E}_Q[K] = S_0^i - \frac{K}{1+r}.$$ □

8.4. Der zweite Hauptsatz der Preistheorie

In diesem Abschnitt zeigen wir, dass ein arbitragefreies Einperiodenmodell genau dann vollständig ist, wenn es genau ein äquivalentes Martingalmaß gibt. Wir betrachten wie in Abschnitt 8.3 ein Einperiodenmodell mit $d+1$ Basisgütern $S^0, \dots, S^d$ auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $S_1^0 = (1+r)S_0^0$ und schreiben $\tilde{S}_t^i = S_t^i / S_t^0$ für die diskontierten Preise.

Lemma 8.25 (Charakterisierung von $\pi_{\inf}$ und $\pi_{\sup}$)**.** *Gegeben sei ein arbitragefreies Einperiodenmodell mit $d+1$ Basisgütern. Sei $\xi \in L^0(\Omega, \mathcal{F}, \mathbb{P})$ ein Kontrakt. Mit*

$$Y^i := \tilde{S}_1^i - \tilde{S}_0^i \quad (i = 0, \dots, d) \qquad \text{und} \qquad \boldsymbol{Y} := (Y^0, \dots, Y^d)^\top$$

gilt dann

$$\pi_{\inf}(\xi) = \max\Big\{m \in [-\infty, \infty) \colon \exists\, \boldsymbol{x} \in \mathbb{R}^{d+1} \text{ mit } m + \boldsymbol{x}^\top \boldsymbol{Y} \le \frac{S_0^0}{S_1^0}\xi \ \mathbb{P}\text{-f.s.}\Big\},$$

$$\pi_{\sup}(\xi) = \min \underbrace{\Big\{m \in (-\infty, \infty] \colon \exists\, \boldsymbol{x} \in \mathbb{R}^{d+1} \text{ mit } m + \boldsymbol{x}^\top \boldsymbol{Y} \ge \frac{S_0^0}{S_1^0}\xi \ \mathbb{P}\text{-f.s.}\Big\}}_{=:M}.$$

Beweis. Wir führen den Beweis nur für $\pi_{\sup}(\xi)$; die Aussage für $\pi_{\inf}(\xi)$ folgt analog. Wir verwenden $S_0^0/S_1^0 = 1/(1+r)$.

Schritt 1: $\pi_{\sup}(\xi) \le \inf M$. Sei Q ein ÄMM mit $\mathbb{E}_Q[|\xi|] < \infty$; insbesondere gilt $\mathbb{E}_Q[\boldsymbol{Y}] = \boldsymbol{0}$. Für alle $m \in M$ folgt durch Bilden des Erwartungswerts unter Q

$$m \ge \frac{S_0^0}{S_1^0}\mathbb{E}_Q[\xi].$$

Nach Satz 8.19 durchläuft die rechte Seite ganz $\Pi(\xi)$, also gilt $m \ge \pi_{\sup}(\xi)$.

Schritt 2: $\pi_{\sup}(\xi) \ge \inf M$. Gilt $\pi_{\sup}(\xi) = \infty$, so folgt $\inf M \le \infty = \pi_{\sup}(\xi)$. Sei also $\pi_{\sup}(\xi) < \infty$, und sei $m \in \mathbb{R}$ mit $\pi_{\sup}(\xi) < m < \infty$. Da $m \notin \Pi(\xi)$ gilt, besitzt das erweiterte Modell mit $S_1^{d+1} = \xi$ und $S_0^{d+1} = m$ eine Arbitragemöglichkeit. Der diskontierte Nettogewinn des neuen Basisguts ist $\xi/S_1^0 - m/S_0^0 = (1/S_0^0)(\xi/(1+r) - m)$. Nach Lemma 8.5 existiert also ein $(\boldsymbol{x}, x_{d+1}) \in \mathbb{R}^{d+2}$ mit

$$\boldsymbol{x}^\top \boldsymbol{Y} + x_{d+1}\Big(\frac{\xi}{1+r} - m\Big) \begin{cases} \ge 0 & \mathbb{P}\text{-f.s.}, \\ > 0 & \text{mit positiver Wahrscheinlichkeit}; \end{cases} \tag{8.8}$$

dabei haben wir den positiven Faktor $1/S_0^0$ in x_{d+1} aufgenommen. Da das Modell mit $d+1$ Basisgütern arbitragefrei ist, folgt $x_{d+1} \neq 0$. Es gilt sogar $x_{d+1} < 0$, denn für ein ÄMM Q mit $\mathbb{E}_Q[|\xi|] < \infty$ folgt aus (8.8)

$$x_{d+1}\Big(\underbrace{\frac{\mathbb{E}_Q[\xi]}{1+r}}_{\in \Pi(\xi)} - m\Big) \ge 0,$$

und die Klammer ist wegen $m > \pi_{\sup}(\xi)$ negativ. Mit $\boldsymbol{z} := -(1/x_{d+1})\boldsymbol{x}$ folgt aus (8.8)

$$\boldsymbol{z}^\top \boldsymbol{Y} + m \ge \frac{S_0^0}{S_1^0}\xi \qquad \mathbb{P}\text{-f.s.},$$

also $m \in M$. Da $m > \pi_{\sup}(\xi)$ beliebig war, folgt $\inf M \le \pi_{\sup}(\xi)$.

Schritt 3: $\inf M = \min M$. Dies ist nur für $\pi_{\sup}(\xi) < \infty$ zu zeigen. Wir wählen eine Folge $(m_n)_{n\in\mathbb{N}} \subseteq M$ mit $m_n \downarrow \inf M = \pi_{\sup}(\xi)$. Wegen $m_n \in M$ und $Y^0 = 0$ existieren $\boldsymbol{x}_n \in \mathbb{R}^{d+1}$ mit $x_{n,0} = 0$, also $\boldsymbol{x}_n = (0, x_{n,1}, \dots, x_{n,d})$, sodass

$$m_n + \boldsymbol{x}_n^\top \boldsymbol{Y} \geq \frac{\xi}{1+r} \qquad \mathbb{P}\text{-f.s. für alle } n \in \mathbb{N}. \tag{8.9}$$

Wir nehmen zunächst $\sup_n \|\boldsymbol{x}_n\| < \infty$ an. Dann existieren eine Teilfolge $(\boldsymbol{x}_{n_k})_{k\in\mathbb{N}}$ und ein $\boldsymbol{x} \in \mathbb{R}^{d+1}$ mit $\boldsymbol{x}_{n_k} \to \boldsymbol{x}$ für $k \to \infty$, und aus (8.9) folgt

$$\pi_{\sup}(\xi) + \boldsymbol{x}^\top \boldsymbol{Y} \geq \frac{\xi}{1+r} \qquad \mathbb{P}\text{-f.s.},$$

also $\pi_{\sup}(\xi) \in M$. Diese Annahme ist keine Einschränkung. Sei dazu $\mathcal{N} := \{\boldsymbol{x} \in \mathbb{R}^{d+1} : \boldsymbol{x}^\top \boldsymbol{Y} = 0\ \mathbb{P}\text{-f.s.}\}$. Ersetzen wir $\boldsymbol{x}_n$ durch seine Orthogonalprojektion auf $\mathcal{N}^\perp$, so ändert sich $\boldsymbol{x}_n^\top \boldsymbol{Y}$ nur auf einer Nullmenge; wir dürfen also $\boldsymbol{x}_n \in \mathcal{N}^\perp$ annehmen. Wäre $(\boldsymbol{x}_n)$ unbeschränkt, so gäbe es eine Teilfolge mit $\|\boldsymbol{x}_{n_k}\| \to \infty$ und $\boldsymbol{x}_{n_k}/\|\boldsymbol{x}_{n_k}\| \to \boldsymbol{\eta} \in \mathcal{N}^\perp$ mit $\|\boldsymbol{\eta}\| = 1$. Division von (8.9) durch $\|\boldsymbol{x}_{n_k}\|$ und $k \to \infty$ liefert $\boldsymbol{\eta}^\top \boldsymbol{Y} \geq 0$ $\mathbb{P}$-fast sicher. Wegen der Arbitragefreiheit folgt mit Lemma 8.5 $\boldsymbol{\eta}^\top \boldsymbol{Y} = 0$ $\mathbb{P}$-fast sicher, also $\boldsymbol{\eta} \in \mathcal{N} \cap \mathcal{N}^\perp = \{\boldsymbol{0}\}$, ein Widerspruch zu $\|\boldsymbol{\eta}\| = 1$. □

Satz 8.26 (Replizierbarkeit und Eindeutigkeit des Preises)**.** *Gegeben sei ein arbitragefreies Einperiodenmodell mit $d+1$ Basisgütern, und sei $\xi \in L^0(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt:*

(i) Es gilt genau dann $\xi \in \mathcal{H}$, wenn ξ genau einen arbitragefreien Preis besitzt.

(ii) Ist $\xi \notin \mathcal{H}$, so gilt $\pi_{\inf}(\xi) < \pi_{\sup}(\xi)$; genauer ist $\Pi(\xi) = (\pi_{\inf}(\xi), \pi_{\sup}(\xi))$ ein offenes Intervall.

Beweis. „(ii)“: Sei $\xi \notin \mathcal{H}$, und angenommen, es gelte $\pi_{\inf}(\xi) \in \Pi(\xi)$. Nach Lemma 8.25 existiert ein $\boldsymbol{x} \in \mathbb{R}^{d+1}$ mit

$$\pi_{\inf}(\xi) + \boldsymbol{x}^\top \boldsymbol{Y} \leq \frac{S_0^0}{S_1^0}\xi \qquad \mathbb{P}\text{-f.s.} \tag{8.10}$$

Hier kann nicht $\mathbb{P}$-fast sicher Gleichheit gelten, da sich ξ sonst als Linearkombination von $S_1^0, \dots, S_1^d$ darstellen ließe, also $\xi \in \mathcal{H}$ wäre. Wir setzen

$$\boldsymbol{z} = (z_0, \dots, z_{d+1})^\top := (-x_0, \dots, -x_d, S_0^0)^\top \qquad \text{und} \qquad S_1^{d+1} := \xi, \quad S_0^{d+1} := \pi_{\inf}(\xi).$$

Mit dem Vektor der diskontierten Nettogewinne im erweiterten Modell

$$\tilde{\boldsymbol{Y}} := \Big(\boldsymbol{Y}, \frac{\xi}{S_1^0} - \frac{\pi_{\inf}(\xi)}{S_0^0}\Big)$$

folgt aus (8.10)

$$\boldsymbol{z}^\top \tilde{\boldsymbol{Y}} = -\boldsymbol{x}^\top \boldsymbol{Y} + S_0^0\Big(\frac{\xi}{S_1^0} - \frac{\pi_{\inf}(\xi)}{S_0^0}\Big) \begin{cases} \geq 0 & \mathbb{P}\text{-f.s.}, \\ > 0 & \text{mit positiver Wahrscheinlichkeit.} \end{cases}$$

Also ist $\boldsymbol{z}$ im erweiterten Modell eine Arbitragemöglichkeit, im Widerspruch zur Annahme $\pi_{\inf}(\xi) \in \Pi(\xi)$. Folglich gilt $\pi_{\inf}(\xi) \notin \Pi(\xi)$, und ebenso folgt $\pi_{\sup}(\xi) \notin \Pi(\xi)$. Da $\Pi(\xi)$ nach Satz 8.19 ein nichtleeres Intervall ist, folgt $\Pi(\xi) = (\pi_{\inf}(\xi), \pi_{\sup}(\xi))$ und insbesondere $\pi_{\inf}(\xi) < \pi_{\sup}(\xi)$.

„(i)“: Die Richtung „⇒“ ist bekannt aus Satz 8.17, denn für $\xi \in \mathcal{H}$ gilt $\Pi(\xi) = \{\hat{\pi}(\xi)\}$. Die Richtung „⇐“ folgt mit (ii). □

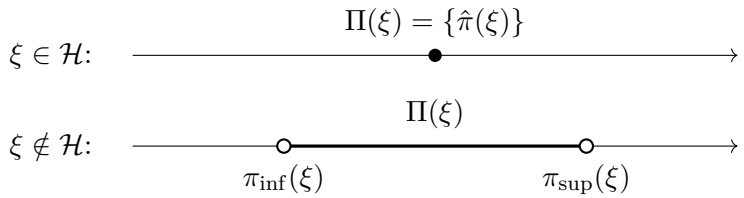

Abbildung 8.5.: Die Menge der arbitragefreien Preise nach Satz 8.26: ein Punkt für replizierbare Kontrakte, ein offenes Intervall sonst.

Satz 8.27 (Zweiter Hauptsatz der Preistheorie im Einperiodenmodell)**.** *Gegeben sei ein arbitragefreies Einperiodenmodell mit $d+1$ Basisgütern. Dann sind die folgenden Aussagen äquivalent:*

(i) Das Modell ist vollständig, das heißt, es gilt $\mathcal{H} = L^0(\Omega, \mathcal{F}, \mathbb{P})$ bis auf $\mathbb{P}$-Nullmengen.

(ii) Es gibt genau ein ÄMM.

Beweis. „(i)⇒(ii)“: Seien Q_1, Q_2 zwei ÄMM. Für alle $A \in \mathcal{F}$ gilt nach (i) $\mathbb{1}_A \in L^0(\Omega, \mathcal{F}, \mathbb{P}) = \mathcal{H}$. Nach Satz 8.19 gilt

$$\frac{S_0^0}{S_1^0}\mathbb{E}_{Q_i}[\mathbb{1}_A] \in \Pi(\mathbb{1}_A) \qquad (i = 1, 2),$$

und nach Satz 8.26 gilt $\#\Pi(\mathbb{1}_A) = 1$. Also folgt $Q_1(A) = Q_2(A)$ für alle $A \in \mathcal{F}$, das heißt $Q_1 = Q_2$.

„(ii)⇒(i)“: Sei Q das eindeutige ÄMM und ξ ein Kontrakt. Nach Satz 8.19 ist $\Pi(\xi)$ nichtleer; nach (8.6) gilt also $\mathbb{E}_Q[|\xi|] < \infty$ und $\#\Pi(\xi) = 1$. Mit Satz 8.26 folgt $\xi \in \mathcal{H}$. □

Bemerkung 8.28. Vollständigkeit heißt

$$L^0(\Omega, \mathcal{F}, \mathbb{P}) = \{\boldsymbol{x}^\top \boldsymbol{S}_1 \colon \boldsymbol{x} \in \mathbb{R}^{d+1}\} = L^0(\Omega, \sigma(S_1^1, \dots, S_1^d), \mathbb{P}),$$

und jeder Kontrakt ist dann ein Derivat. Insbesondere gilt $\dim \mathcal{H} = \dim L^0(\Omega, \mathcal{F}, \mathbb{P}) \leq d+1$. Daraus folgt, dass eine Partition von Ω in höchstens $d+1$ Atome von $(\Omega, \mathcal{F}, \mathbb{P})$ existiert. Dabei heißt $A \in \mathcal{F}$ ein *Atom*, falls $\mathbb{P}(A) > 0$ gilt und $\mathbb{P}(B) \in \{0, \mathbb{P}(A)\}$ für alle $B \in \mathcal{F}$ mit $B \subseteq A$. Für einen Beweis siehe [FS02, Abschnitt 1.4].

9. Das Mehrperiodenmodell

Zu diesem Kapitel vergleiche man [FS02, Kapitel 5] sowie [Shr04] und [DS06].

Wir übertragen nun die Begriffe aus Kapitel 8 auf Modelle mit endlich vielen Handelszeitpunkten.

9.1. Modell, Handelsstrategien und Arbitrage

Definition 9.1 (Mehrperiodenmodell)**.** Das *Mehrperiodenmodell* mit $d+1$ Basisgütern und Zeitpunkten $t = 0, 1, \dots, T$ besteht aus

(1) einem filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,\dots,T}, \mathbb{P})$ mit $\mathcal{F}_0 = \{\emptyset, \Omega\}$ und $\mathcal{F}_T = \mathcal{F}$,

(2) einer $\mathbb{R}^{d+1}$-wertigen, an $(\mathcal{F}_t)$ adaptierten stochastischen Folge $(\boldsymbol{S}_t)_{t=0,\dots,T}$ mit

$$\boldsymbol{S}_t = (S_t^0, \dots, S_t^d) \colon \Omega \to \mathbb{R}^{d+1},$$

(3) dem nullten Basisgut S_t^0, das für alle t deterministisch ist und eine risikolose Anlage (Numéraire) mit Zinsrate r beschreibt, also

$$S_t^0 = (1+r)^t S_0^0.$$

Für $i > 0$ beschreibt S_t^i den (zufälligen) Preis des i-ten Basisguts zum Zeitpunkt t, und

$$(\tilde{S}_t^0, \dots, \tilde{S}_t^d) \coloneqq \Big(1, \frac{S_t^1}{S_t^0}, \dots, \frac{S_t^d}{S_t^0}\Big)$$

heißt der Vektor der *diskontierten Basisgüter*.

Definition 9.2 (Portfolio)**.** Ein *Portfolio* oder eine *Handelsstrategie* im Mehrperiodenmodell ist eine $\mathbb{R}^{d+1}$-wertige, vorhersehbare stochastische Folge $(\boldsymbol{\varphi}_t)_{t=1,\dots,T}$ mit $\boldsymbol{\varphi}_t = (\varphi_t^0, \dots, \varphi_t^d)$. Dabei gibt φ_t^i die Stückzahl des i-ten Basisguts an, welche der Investor in der Zeit $[t-1, t)$ hält. Zum Zeitpunkt t kann (und muss) er sich für eine Umstrukturierung $\boldsymbol{\varphi}_{t+1}$ entscheiden, welche dann in der Zeit $[t, t+1)$ gilt. Insbesondere ist $\boldsymbol{\varphi}_t$ also $\mathcal{F}_{t-1}$-messbar.

Der *Vermögensprozess* $(V_t)_{t=0,\dots,T}$ zu einem Portfolio $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_t)_{t=1,\dots,T}$ ist

$$V_t(\boldsymbol{\varphi}) \coloneqq \begin{cases} \boldsymbol{\varphi}_1^\top \boldsymbol{S}_0 = \sum_{i=0}^d \varphi_1^i S_0^i, & t = 0, \\ \boldsymbol{\varphi}_t^\top \boldsymbol{S}_t = \sum_{i=0}^d \varphi_t^i S_t^i, & t = 1, \dots, T; \end{cases}$$

er beschreibt das Vermögen zur Zeit t *vor* der Umstrukturierung. Dagegen beschreibt

$$V_{t+}(\boldsymbol{\varphi}) \coloneqq \begin{cases} \boldsymbol{\varphi}_{t+1}^\top \boldsymbol{S}_t, & t = 0, \dots, T-1, \\ \boldsymbol{\varphi}_T^\top \boldsymbol{S}_T, & t = T, \end{cases}$$

den Vermögenswert zur Zeit t *nach* der Umstrukturierung; siehe Abbildung 9.1.

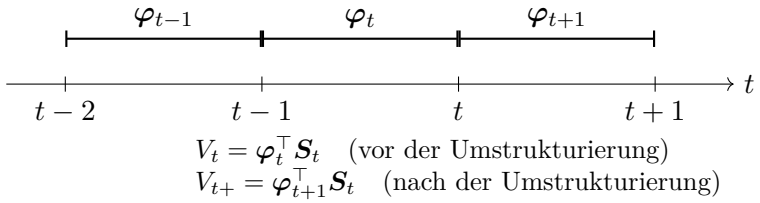

Abbildung 9.1.: Das Portfolio φ_t wird zum Zeitpunkt $t-1$ auf Basis der Information $\mathcal{F}_{t-1}$ gewählt und in $[t-1,t)$ gehalten. Zum Zeitpunkt t wird zu φ_{t+1} umgeschichtet; V_t und V_{t+} sind die Vermögenswerte vor und nach der Umstrukturierung.

Bemerkung. Es gilt $V_0(\varphi) = V_{0+}(\varphi)$ und $V_T(\varphi) = V_{T+}(\varphi)$.

Definition 9.3 (Selbstfinanzierung)**.** Ein Portfolio $\varphi = (\varphi_1, \dots, \varphi_T)$ im Mehrperiodenmodell heißt *selbstfinanzierend*, falls

$$V_t(\varphi) = V_{t+}(\varphi) \qquad \text{für alle } t = 1, \dots, T-1.$$

Das Vermögen vor und nach der Umstrukturierung ist also identisch; es gibt keine Zu- und Abflüsse.

Bemerkung 9.4. Für eine Handelsstrategie $\varphi = (\varphi_1, \dots, \varphi_T)$ sind äquivalent:

(i) φ ist selbstfinanzierend.

(ii) Für alle $t = 1, \dots, T$ gilt

$$V_t(\varphi) = V_0(\varphi) + \sum_{k=1}^{t} \varphi_k^\top (S_k - S_{k-1}).$$

Für $t = 1$ gilt $V_1(\varphi) - V_0(\varphi) = \varphi_1^\top (S_1 - S_0)$ stets. Für $t = 2, \dots, T$ gilt

$$\begin{aligned} V_t(\varphi) - V_{t-1}(\varphi) = \varphi_t^\top (S_t - S_{t-1}) &\iff \varphi_t^\top S_t - \varphi_{t-1}^\top S_{t-1} = \varphi_t^\top S_t - \varphi_t^\top S_{t-1} \\ &\iff V_{t-1}(\varphi) = V_{(t-1)+}(\varphi). \end{aligned}$$

Die Aussage (ii) ist äquivalent dazu, dass die linke Seite für alle $t = 1, \dots, T$ gilt, und die rechte Seite für alle $t = 2, \dots, T$ ist gerade (i).

Definition 9.5 (Arbitragemöglichkeit)**.** Ein Portfolio $(\varphi_1, \dots, \varphi_T)$ im Mehrperiodenmodell heißt *Arbitragemöglichkeit*, falls gilt:

(1) $\varphi = (\varphi_1, \dots, \varphi_T)$ ist selbstfinanzierend,

(2) $V_0(\varphi) = \varphi_1^\top S_0 \le 0$,

(3) $V_T(\varphi) = \varphi_T^\top S_T \ge 0$ $\mathbb{P}$-fast sicher,

(4) $V_T(\varphi) > 0$ mit positiver Wahrscheinlichkeit.

Gibt es keine Arbitragemöglichkeit, so heißt das Modell *arbitragefrei*.

Lemma 9.6 (Arbitragefreiheit und Zwischenzeitpunkte)**.** *Im Mehrperiodenmodell sind äquivalent:*

(i) Das Modell ist arbitragefrei.

(ii) Es gibt keine Handelsstrategie $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_T)$ und keinen Zeitpunkt $t \in \{0, \dots, T\}$, sodass $\boldsymbol{\varphi}$ selbstfinanzierend ist und

$$V_0(\boldsymbol{\varphi}) \le 0, \qquad V_t(\boldsymbol{\varphi}) \ge 0 \ \mathbb{P}\text{-}f.s. \qquad und \qquad \mathbb{P}\big(V_t(\boldsymbol{\varphi}) > 0\big) > 0$$

gilt.

Beweisidee. Die Richtung (ii)⇒(i) ist klar, denn wir können $t = T$ wählen. Für (i)⇒(ii) zeigen wir ¬(ii)⇒ ¬(i). Dazu schichten wir das Vermögen zum Zeitpunkt t vollständig auf das festverzinsliche Konto um. Genauer setzen wir $\boldsymbol{\psi}_k := \boldsymbol{\varphi}_k$ für $k \le t$ und $\boldsymbol{\psi}_k := (V_t(\boldsymbol{\varphi})/S_t^0, 0, \dots, 0)$ für $k > t$. Dann ist $\boldsymbol{\psi}$ selbstfinanzierend, und es gilt $V_0(\boldsymbol{\psi}) = V_0(\boldsymbol{\varphi}) \le 0$ sowie $V_T(\boldsymbol{\psi}) = (S_T^0/S_t^0)V_t(\boldsymbol{\varphi})$; also ist $\boldsymbol{\psi}$ eine Arbitragemöglichkeit. □

9.2. Der erste Hauptsatz der Preistheorie

Definition 9.7 (Äquivalentes Martingalmaß im Mehrperiodenmodell)**.** Wir betrachten ein Mehrperiodenmodell mit $d+1$ Basisgütern $\boldsymbol{S}_t = (S_t^0, \dots, S_t^d)_{t=0,\dots,T}$ auf $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,\dots,T}, \mathbb{P})$ und den diskontierten Größen $\tilde{S}_t^i = S_t^i/S_t^0$. Ein Wahrscheinlichkeitsmaß Q auf $(\Omega, \mathcal{F})$ heißt *risikoneutrales Maß* oder *Martingalmaß*, falls $(\tilde{\boldsymbol{S}}_t)_{t=0,\dots,T}$ bezüglich $(\mathcal{F}_t)_{t=0,\dots,T}$ ein Q-Martingal ist. Gilt zusätzlich $\mathbb{P} \ll Q$ und $Q \ll \mathbb{P}$, so heißt Q *äquivalentes Martingalmaß* (ÄMM).

Bemerkung 9.8. Wegen $\mathcal{F}_0 = \{\emptyset, \Omega\}$ gilt für alle $t = 0, \dots, T$

$$\tilde{\boldsymbol{S}}_0 = \mathbb{E}_Q[\tilde{\boldsymbol{S}}_t | \mathcal{F}_0] = \mathbb{E}_Q[\tilde{\boldsymbol{S}}_t] \qquad \mathbb{P}\text{-f.s.},$$

wie im Einperiodenmodell.

Satz 9.9 (Diskontierter Vermögensprozess als Martingal)**.** *Im Mehrperiodenmodell auf $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,\dots,T}, \mathbb{P}, (\boldsymbol{S}_t)_{t=0,\dots,T})$ mit $d+1$ Basisgütern sei Q ein ÄMM. Sei $(\boldsymbol{\varphi}_t)_{t=1,\dots,T}$ eine selbstfinanzierende, lokal beschränkte Handelsstrategie, das heißt, es gelte $\sup_{\omega\in\Omega} |\varphi_t^i(\omega)| < \infty$ für alle t und i. Dann ist der diskontierte Vermögensprozess $\big(V_t(\boldsymbol{\varphi})/S_t^0\big)_{t=0,\dots,T}$ bezüglich $(\mathcal{F}_t)_{t=0,\dots,T}$ ein Q-Martingal.*

Beweis. Wir setzen $\tilde{V}_t := (1/S_t^0)V_t(\boldsymbol{\varphi})$. Da $\boldsymbol{\varphi}$ selbstfinanzierend ist, gilt auch $\tilde{V}_{t+} = \tilde{V}_t$, das heißt, $\boldsymbol{\varphi}$ ist auch im diskontierten Modell selbstfinanzierend. Mit Bemerkung 9.4, angewandt auf das diskontierte Modell, folgt

$$\tilde{V}_t(\boldsymbol{\varphi}) = \tilde{V}_0(\boldsymbol{\varphi}) + \sum_{k=1}^{t} \boldsymbol{\varphi}_k^\top (\tilde{\boldsymbol{S}}_k - \tilde{\boldsymbol{S}}_{k-1}) = \tilde{V}_0(\boldsymbol{\varphi}) + (\boldsymbol{\varphi} \cdot \tilde{\boldsymbol{S}})_t,$$

wobei $(\boldsymbol{\varphi} \cdot \tilde{\boldsymbol{S}})_t$ das stochastische Integral (die Martingaltransformation) bezeichnet. Da $(\tilde{\boldsymbol{S}}_t)_{t=0,\dots,T}$ ein Q-Martingal und $\boldsymbol{\varphi}$ vorhersehbar und beschränkt ist, folgt mit Satz 5.15, dass $(\tilde{V}_t(\boldsymbol{\varphi}))_{t=0,\dots,T}$ ein Q-Martingal ist. □

Satz 9.10 (Erster Hauptsatz der Preistheorie im Mehrperiodenmodell [DMW90])**.** *Im Mehrperiodenmodell $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,\dots,T}, \mathbb{P}, (\boldsymbol{S}_t)_{t=0,\dots,T})$ mit $d+1$ Basisgütern gilt: Das Modell ist genau dann arbitragefrei, wenn ein ÄMM existiert. In diesem Fall existiert sogar ein ÄMM mit beschränkter Dichte bezüglich $\mathbb{P}$.*

Beweis. „⇐“: Sei Q ein ÄMM und $\boldsymbol{\varphi}$ selbstfinanzierend mit $V_T(\boldsymbol{\varphi}) \ge 0$ $\mathbb{P}$-fast sicher und $\mathbb{P}(V_T(\boldsymbol{\varphi}) > 0) > 0$. Zu zeigen ist $V_0(\boldsymbol{\varphi}) > 0$, das heißt, $\boldsymbol{\varphi}$ ist keine Arbitragemöglichkeit. Für $\tilde{V}_t := (1/S_t^0)V_t$ gilt $\tilde{V}_T(\boldsymbol{\varphi}) \ge 0$ $\mathbb{P}$-fast sicher und $\mathbb{P}(\tilde{V}_T(\boldsymbol{\varphi}) > 0) > 0$, da $S_T^0 > 0$. Wegen

$\mathbb{P} \ll Q \ll \mathbb{P}$ folgt $\tilde{V}_T(\boldsymbol{\varphi}) \geq 0$ Q-fast sicher und $Q(\tilde{V}_T(\boldsymbol{\varphi}) > 0) > 0$, also $\mathbb{E}_Q[\tilde{V}_T(\boldsymbol{\varphi})] > 0$. Da $(\tilde{V}_t)_{t=0,\dots,T}$ nach Satz 9.9 ein Q-Martingal bezüglich $(\mathcal{F}_t)$ ist, folgt

$$\tilde{V}_0(\boldsymbol{\varphi}) = \mathbb{E}_Q\big[\tilde{V}_T(\boldsymbol{\varphi})|\mathcal{F}_0\big] = \mathbb{E}_Q\big[\tilde{V}_T(\boldsymbol{\varphi})\big] > 0,$$

also $V_0(\boldsymbol{\varphi}) > 0$, da $S_0^0 > 0$. Hier haben wir Satz 9.9 verwendet, also implizit die Beschränktheit von $\boldsymbol{\varphi}$ vorausgesetzt; für allgemeine Handelsstrategien siehe [FS02, Abschnitt 5.2].

„⇒“: Siehe [FS02, Abschnitt 5.2]. □

9.3. Preisbestimmung europäischer Kontrakte

Definition 9.11 (Europäischer Kontrakt)**.** Wir betrachten ein Mehrperiodenmodell mit $d+1$ Basisgütern auf $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,\dots,T}, \mathbb{P})$ mit Preisprozess $(\boldsymbol{S}_t)_{t=0,\dots,T}$. Ein *(europäischer) Kontrakt* mit Fälligkeit T ist eine $\mathcal{F}_T$-messbare (also $\mathcal{F}$-messbare) Zufallsvariable ξ. Gilt $\mathbb{P}(\xi \geq 0) = 1$, so heißt ξ *bedingter europäischer Finanztitel*. Der Kontrakt ξ heißt *Derivat* der Underlyings $S^0, S^1, \dots, S^d$, falls $\xi = f(S^1, \dots, S^d)$ für eine messbare Funktion f gilt.

Definition 9.12 (Hedging)**.** Ein europäischer Kontrakt ξ heißt *absicherbar* oder *replizierbar*, falls ein selbstfinanzierendes Portfolio $\boldsymbol{\varphi}$ existiert mit

$$V_T(\boldsymbol{\varphi}) = \xi \qquad \mathbb{P}\text{-f.s.}$$

Das Portfolio $\boldsymbol{\varphi}$ wird dann *Hedge* für ξ genannt. Wir schreiben

$$\begin{aligned}\mathcal{H} &:= \{\xi \ \mathcal{F}\text{-messbar}\colon \xi \text{ ist replizierbar}\},\\ \mathcal{H}_+ &:= \{\xi \in \mathcal{H}\colon \xi \geq 0 \ \mathbb{P}\text{-f.s.}\}.\end{aligned}$$

Ein Mehrperiodenmodell heißt *vollständig*, falls alle Kontrakte replizierbar sind.

Satz 9.13 (Law of One Price)**.** *Im arbitragefreien Mehrperiodenmodell sei $\xi \in \mathcal{H}_+$, und seien $\boldsymbol{\varphi}, \boldsymbol{\psi}$ zwei perfekte Hedges für ξ. Dann gilt*

$$V_0(\boldsymbol{\varphi}) = V_0(\boldsymbol{\psi}),$$

und man nennt $\hat{\pi}(\xi) := V_0(\boldsymbol{\varphi})$ den Hedgepreis *von ξ. Ist Q ein ÄMM, so gilt:*

(i) $\mathbb{E}_Q\big[|\xi|\big] < \infty$,

(ii) $V_0(\boldsymbol{\varphi}) = \hat{\pi}(\xi) = (S_0^0/S_T^0)\mathbb{E}_Q[\xi]$,

(iii) $V_t(\boldsymbol{\varphi}) = V_t(\boldsymbol{\psi}) = (S_t^0/S_T^0)\mathbb{E}_Q[\xi|\mathcal{F}_t]$ *für alle* $t = 0, \dots, T$.

Solange $\boldsymbol{\varphi}$ und $\boldsymbol{\psi}$ beschränkt sind, gelten alle Aussagen auch für $\xi \in \mathcal{H}$.

Beweis. Ohne Beweis. □

Bemerkung. Ist Ω endlich, so ist jede Handelsstrategie $\boldsymbol{\varphi}$ beschränkt.

Vereinbarung 9.14 (No-Arbitrage-Prinzip im Mehrperiodenmodell)**.** Wir betrachten ein arbitragefreies Mehrperiodenmodell und einen Kontrakt $\xi \geq 0$. Eine Zahl $a \in \mathbb{R}$ heißt *arbitragefreier Preis* für ξ, bezeichnet mit $\pi(\xi)$, falls ein adaptierter stochastischer Prozess $(S_t^{d+1})_{t=0,\dots,T}$ mit

$$a = S_0^{d+1} \qquad \text{und} \qquad \xi = S_T^{d+1}$$

existiert, sodass der um S^{d+1} erweiterte Markt mit $d+2$ Basisgütern arbitragefrei ist. Die Menge der arbitragefreien Preise wird mit $\Pi(\xi)$ bezeichnet. Als Preis für ξ ist grundsätzlich ein arbitragefreier Preis festzulegen.

Satz 9.15. *Sei $\xi \in \mathcal{H}_+$ im arbitragefreien Mehrperiodenmodell. Dann gilt*

$$\Pi(\xi) = \{\hat{\pi}(\xi)\}.$$

Beweis. Ohne Beweis. □

Satz 9.16. *Sei $\xi \geq 0$ ein Kontrakt im arbitragefreien Mehrperiodenmodell. Dann gilt*

$$\emptyset \neq \Pi(\xi) = \Big\{ \frac{S_0^0}{S_T^0} \mathbb{E}_Q[\xi] \colon Q \text{ ÄMM mit } \mathbb{E}_Q[|\xi|] < \infty \Big\}.$$

Beweis. Ohne Beweis. □

Bemerkung 9.17. Die Größen

$$\pi_{\inf}(\xi) \coloneqq \inf \Pi(\xi) \qquad \text{und} \qquad \pi_{\sup}(\xi) \coloneqq \sup \Pi(\xi)$$

heißen *untere* beziehungsweise *obere Arbitrageschranke* für ξ. Wie im Einperiodenmodell gilt für $\xi \geq 0$:

(i) Es gilt genau dann $\xi \in \mathcal{H}_+$, wenn $\Pi(\xi)$ aus genau einem Element besteht; in diesem Fall ist $\Pi(\xi) = \{\hat{\pi}(\xi)\}$.

(ii) Ist $\xi \notin \mathcal{H}_+$, so gilt $\pi_{\inf}(\xi) < \pi_{\sup}(\xi)$ und $\Pi(\xi) = (\pi_{\inf}(\xi), \pi_{\sup}(\xi))$.

Wir verzichten auf einen Beweis; siehe [FS02, Abschnitt 5.3].

Satz 9.18 (Zweiter Hauptsatz der Preistheorie im Mehrperiodenmodell)**.** *Im arbitragefreien Mehrperiodenmodell sind die folgenden Aussagen äquivalent:*

(i) Das Mehrperiodenmodell ist vollständig.

(ii) Zu jedem bedingten Kontrakt existiert ein perfekter Hedge.

(iii) Es gibt genau ein ÄMM.

Beweis. Ohne Beweis. □

Bemerkung 9.19. Wie im Einperiodenmodell, vergleiche Satz 8.22, gilt

$$0 \leq \xi_1 \leq \xi_2 \ \mathbb{P}\text{-f.s.} \quad \Longrightarrow \quad \pi(\xi_1) \leq \pi(\xi_2)$$

und $\pi(\xi_1 + \xi_2) = \pi(\xi_1) + \pi(\xi_2)$, sofern die Preise wie dort im gemeinsam erweiterten Modell festgelegt werden. Außerdem gilt die Put-Call-Parität, vergleiche Korollar 8.24.

9.4. Das Cox-Ross-Rubinstein-Modell

Das *Cox-Ross-Rubinstein-Modell* (CRR-Modell) [CRR79] ist ein Mehrperiodenmodell mit einem festverzinslichen Konto und einer Aktie, die sich zu jedem Zeitpunkt in zwei Zustände verzweigen kann. Genauer besteht es aus den folgenden Bestandteilen.

(1) Die Zeitpunkte sind $t = 0, \ldots, T$.

(2) Das Konto ist $S_t^0 = (1 + r)^t S_0^0$ mit $S_0^0 = 1$ (ohne Einschränkung).

(3) Der Modellraum ist $\Omega \coloneqq \{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_T)^\top \colon \omega_i \in \{0, 1\}\}$ mit $\mathcal{F} \coloneqq \mathcal{P}(\Omega)$ und einem beliebigen Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{F})$ mit $\mathbb{P}(\{\boldsymbol{\omega}\}) > 0$ für alle $\boldsymbol{\omega} \in \Omega$. Dabei bedeutet $\omega_t = 1$, dass sich der Markt zum Zeitpunkt t gut entwickelt.

(4) Wir setzen $Z_t(\boldsymbol{\omega}) \coloneqq \omega_t$ und

$$D_t(\boldsymbol{\omega}) \coloneqq \sum_{s=1}^{t} Z_s(\boldsymbol{\omega}),$$

also die Anzahl der guten Marktentwicklungen bis zum Zeitpunkt t.

(5) Der Anfangspreis $S_0^1 > 0$ der Aktie ist konstant, und

$$S_t^1(\boldsymbol{\omega}) = S_0^1(1+r_g)^{D_t(\boldsymbol{\omega})}(1+r_b)^{t-D_t(\boldsymbol{\omega})}.$$

Dabei ist r_g die Wachstumsrate bei guter und r_b die Wachstumsrate bei schlechter Marktentwicklung.

Wir setzen dabei $-1 < r_b < r_g$ und $r > -1$ voraus und versehen Ω mit der Filtration $\mathcal{F}_t \coloneqq \sigma(Z_1, \ldots, Z_t)$, $t = 0, \ldots, T$; insbesondere gilt $\mathcal{F}_0 = \{\emptyset, \Omega\}$ und $\mathcal{F}_T = \mathcal{F}$. Abbildung 9.2 zeigt die möglichen Kursverläufe der Aktie als Binomialbaum.

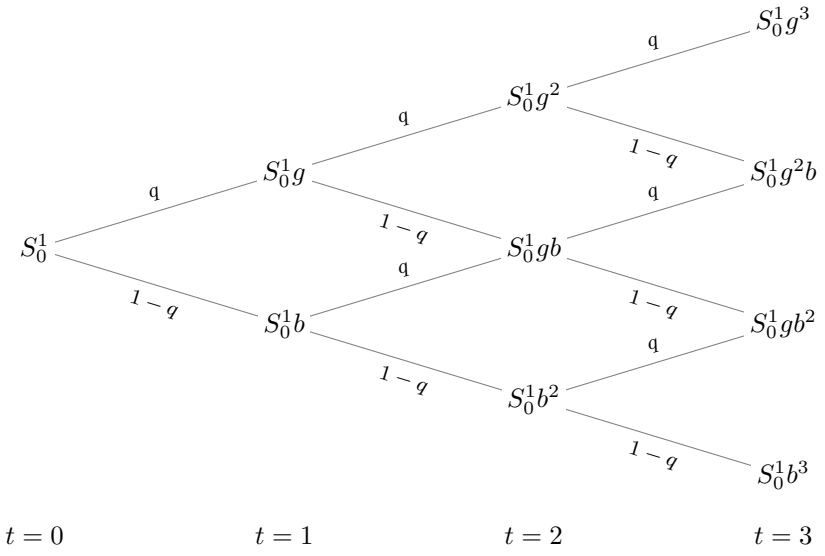

Abbildung 9.2.: Kursverlauf S_t^1 der Aktie im CRR-Modell für $T = 3$, mit den Abkürzungen $g \coloneqq 1 + r_g$ und $b \coloneqq 1 + r_b$. Der Baum ist rekombinierend: S_t^1 hängt nur von der Anzahl D_t der Aufwärtsschritte ab. An den Kanten stehen die Übergangswahrscheinlichkeiten unter dem ÄMM Q aus Satz 9.20.

Satz 9.20 (Arbitragefreiheit von CRR-Modellen)**.** *Das CRR-Modell ist genau dann arbitragefrei, wenn $r_b < r < r_g$ gilt. In diesem Fall ist das Modell auch vollständig. Das eindeutig bestimmte ÄMM lautet*

$$Q(\{\boldsymbol{\omega}\}) = q^{D_T(\boldsymbol{\omega})}(1-q)^{T-D_T(\boldsymbol{\omega})} \tag{9.1}$$

mit

$$q \coloneqq \frac{r - r_b}{r_g - r_b} \in (0,1). \tag{9.2}$$

Beweis. Es gilt

$$\tilde{S}_t^1 = \frac{S_t^1}{S_t^0} = \Big(\frac{1+r_g}{1+r}\Big)^{D_t}\Big(\frac{1+r_b}{1+r}\Big)^{t-D_t} S_0^1.$$

Ein Wahrscheinlichkeitsmaß Q ist genau dann ein ÄMM, wenn

(1) $Q(\{\boldsymbol{\omega}\}) > 0$ für alle $\boldsymbol{\omega} \in \Omega$ gilt und

(2) für alle $t = 1, \dots, T$ gilt

$$\mathbb{E}_Q\big[\tilde{S}^1_t | \mathcal{F}_{t-1}\big] = \tilde{S}^1_{t-1}. \tag{9.3}$$

Dabei ist (9.3) wegen $\tilde{S}^1_t = \tilde{S}^1_{t-1}\big((1+r_g)/(1+r)\big)^{Z_t}\big((1+r_b)/(1+r)\big)^{1-Z_t}$ äquivalent zu

$$\mathbb{E}_Q\Big[\tilde{S}^1_{t-1}\Big(\frac{1+r_g}{1+r}\Big)^{Z_t}\Big(\frac{1+r_b}{1+r}\Big)^{1-Z_t}\Big|\mathcal{F}_{t-1}\Big] = \tilde{S}^1_{t-1} \qquad \text{für alle } t.$$

Da $\tilde{S}^1_{t-1} > 0$ und $\mathcal{F}_{t-1}$-messbar ist, ist dies äquivalent zu

$$\begin{aligned}
&\mathbb{E}_Q\Big[\Big(\frac{1+r_g}{1+r}\Big)^{Z_t}\Big(\frac{1+r_b}{1+r}\Big)^{1-Z_t}\Big|\mathcal{F}_{t-1}\Big] = 1 && \text{für alle } t\\
\iff\ &\mathbb{E}_Q[\mathbb{1}\{Z_t = 1\}|\mathcal{F}_{t-1}]\frac{1+r_g}{1+r} + \mathbb{E}_Q[\mathbb{1}\{Z_t = 0\}|\mathcal{F}_{t-1}]\frac{1+r_b}{1+r} = 1 && \text{für alle } t\\
\iff\ &\mathbb{E}_Q[\mathbb{1}\{Z_t = 1\}|\mathcal{F}_{t-1}]\frac{(1+r_g)-(1+r_b)}{1+r} + \frac{1+r_b}{1+r} = 1 && \text{für alle } t\\
\iff\ &Q(\{Z_t = 1\}|\mathcal{F}_{t-1})\frac{r_g - r_b}{1+r} = 1 - \frac{1+r_b}{1+r} = \frac{r-r_b}{1+r} && \text{für alle } t.
\end{aligned}$$

Die bedingte Wahrscheinlichkeit $Q(\{Z_t = 1\}|\mathcal{F}_{t-1})$ ist also für alle t gleich der Konstanten q aus (9.2). Durch sukzessives Bedingen folgt, dass dies genau dann gilt, wenn $Z_1, \dots, Z_T$ unter Q unabhängig und identisch Bernoulli-verteilt sind mit

$$Q(Z_t = 1) = \frac{r - r_b}{r_g - r_b} \qquad \text{für alle } t,$$

also genau dann, wenn Q durch (9.1) gegeben ist. Die Bedingung (1) ist genau dann erfüllt, wenn $q \in (0,1)$ gilt, das heißt, wenn $r_b < r < r_g$ gilt. Mit Satz 9.10 folgt die Charakterisierung der Arbitragefreiheit. Da Q eindeutig ist, ist das CRR-Modell nach Satz 9.18 vollständig. □

Satz 9.21 (Preise von Derivaten im CRR-Modell)**.** *Wir betrachten im CRR-Modell ein Derivat ξ der Form $\xi = f(S^1_T)$ mit $f\colon [0,\infty) \to \mathbb{R}$, und es gelte $r_b < r < r_g$. Dann lautet der Hedgepreis von ξ*

$$\hat{\pi}(\xi) = \mathbb{E}_Q\Big[\frac{f(S^1_T)}{(1+r)^T}\Big] = \frac{1}{(1+r)^T}\sum_{j=0}^{T}\binom{T}{j}q^j(1-q)^{T-j}f\big(S^1_0(1+r_g)^j(1+r_b)^{T-j}\big), \tag{9.4}$$

mit Q und q wie in Satz 9.20. Ist $f \geq 0$, so ist $\hat{\pi}(\xi)$ zugleich der eindeutige arbitragefreie Preis, das heißt $\Pi(\xi) = \{\hat{\pi}(\xi)\}$.

Beweis. Nach Satz 9.20 ist das Modell vollständig, also $\xi \in \mathcal{H}$; da Ω endlich ist, sind alle Hedges beschränkt. Mit Satz 9.13 und $S^0_0/S^0_T = (1+r)^{-T}$ folgt die erste Darstellung in (9.4), und für $f \geq 0$ folgt $\Pi(\xi) = \{\hat{\pi}(\xi)\}$ aus Satz 9.15. Unter Q ist D_T nach (9.1) binomialverteilt mit Parametern T und q, und es gilt $S^1_T = S^1_0(1+r_g)^{D_T}(1+r_b)^{T-D_T}$; daraus folgt die zweite Darstellung. □

9.5. Vom CRR-Modell zur Black–Scholes-Formel

Wir lassen nun die Anzahl der Handelszeitpunkte gegen unendlich gehen und halten dabei die Laufzeit fest. Die Preise europäischer Optionen im CRR-Modell konvergieren dann gegen die Preise im Modell von Black, Scholes und Merton [BS73, Mer73]. Für eine ausführlichere Darstellung verweisen wir auf [FS02, Abschnitt 5.6].

9.5.1. Die Folge der CRR-Modelle

Abweichend von Abschnitt 9.4 bezeichne $T > 0$ ab jetzt die (reelle) Laufzeit, und $N \in \mathbb{N}$ sei die Anzahl der Perioden. Wir fixieren einen Zinssatz $\rho \geq 0$ in stetiger Verzinsung, eine *Volatilität* $\sigma > 0$ und einen Anfangspreis $S_0^1 > 0$. Für jedes $N \in \mathbb{N}$ betrachten wir das CRR-Modell mit N Perioden der Länge $\Delta_N := T/N$ und den Parametern

$$1 + r_N := \mathrm{e}^{\rho\Delta_N}, \qquad 1 + r_{g,N} := \mathrm{e}^{\sigma\sqrt{\Delta_N}}, \qquad 1 + r_{b,N} := \mathrm{e}^{-\sigma\sqrt{\Delta_N}}. \tag{9.5}$$

Für alle hinreichend großen N gilt $\rho\Delta_N < \sigma\sqrt{\Delta_N}$, also $r_{b,N} < r_N < r_{g,N}$. Nach Satz 9.20 ist das N-te Modell dann arbitragefrei und vollständig mit eindeutigem ÄMM Q_N, und unter Q_N sind $Z_1, \ldots, Z_N$ unabhängig und Bernoulli-verteilt mit Parameter

$$q_N = \frac{\mathrm{e}^{\rho\Delta_N} - \mathrm{e}^{-\sigma\sqrt{\Delta_N}}}{\mathrm{e}^{\sigma\sqrt{\Delta_N}} - \mathrm{e}^{-\sigma\sqrt{\Delta_N}}}.$$

Wir schreiben $S_N^{1,N}$ für den Aktienpreis zum Endzeitpunkt im N-ten Modell. Wegen (9.5) gilt

$$S_N^{1,N} = S_0^1 \exp(X_N) \qquad \text{mit} \qquad X_N := \sum_{k=1}^{N} Y_{N,k}, \quad Y_{N,k} := \sigma\sqrt{\Delta_N}\,(2Z_k - 1).$$

Der Logarithmus des Endpreises ist also eine Summe von N unter Q_N unabhängigen und identisch verteilten Zufallsvariablen mit Werten in $\{\pm\sigma\sqrt{\Delta_N}\}$. Im Unterschied zu Satz 1.15 hängt ihre Verteilung von N ab; wir argumentieren daher direkt mit charakteristischen Funktionen.

Lemma 9.22 (Grenzverteilung des Log-Preises)**.** *Unter Q_N gilt*

$$X_N \xrightarrow[N\to\infty]{\mathcal{D}} X \qquad \textit{mit} \qquad X \sim \mathcal{N}\Big((\rho - \sigma^2/2)T,\, \sigma^2 T\Big).$$

Beweis. Wir schreiben kurz $\Delta = \Delta_N$. Taylorentwicklung von Zähler und Nenner von q_N liefert für $N \to \infty$

$$q_N = \frac{\sigma\sqrt{\Delta} + (\rho - \sigma^2/2)\Delta + O(\Delta^{3/2})}{2\sigma\sqrt{\Delta} + O(\Delta^{3/2})} = \frac{1}{2} + \frac{\rho - \sigma^2/2}{2\sigma}\sqrt{\Delta} + O(\Delta).$$

Für $Y := Y_{N,1}$ folgt

$$\mathbb{E}_{Q_N}[Y] = \sigma\sqrt{\Delta}\,(2q_N - 1) = (\rho - \sigma^2/2)\Delta + O(\Delta^{3/2})$$

sowie $\mathbb{E}_{Q_N}[Y^2] = \sigma^2\Delta$ und $\mathbb{E}_{Q_N}[|Y|^3] = \sigma^3\Delta^{3/2}$. Mit der Abschätzung $|\mathrm{e}^{\mathrm{i}x} - 1 - \mathrm{i}x + x^2/2| \leq |x|^3/6$ für $x \in \mathbb{R}$ erhalten wir für festes $t \in \mathbb{R}$

$$\mathbb{E}_{Q_N}\big[\mathrm{e}^{\mathrm{i}tY}\big] = 1 + \mathrm{i}t\,\mathbb{E}_{Q_N}[Y] - \frac{t^2}{2}\mathbb{E}_{Q_N}[Y^2] + O(|t|^3\Delta^{3/2}) = 1 + \frac{a_N}{N}$$

mit

$$a_N := T\Big(\mathrm{i}t(\rho - \sigma^2/2) - \frac{t^2\sigma^2}{2}\Big) + O(N^{-1/2}) \xrightarrow[N\to\infty]{} a := \mathrm{i}t(\rho - \sigma^2/2)T - \frac{t^2\sigma^2 T}{2}.$$

Wegen der Unabhängigkeit und Lemma 1.7 gilt $\varphi^{X_N}(t) = (1 + a_N/N)^N$, und dies konvergiert für $N \to \infty$ gegen e^a, denn $(1 + a_N/N)^N \to \mathrm{e}^a$ gilt für jede komplexe Folge $a_N \to a$ wie im Beweis von Satz 1.15. Nach (1.5) ist $t \mapsto \mathrm{e}^a$ die charakteristische Funktion von $\mathcal{N}((\rho - \sigma^2/2)T, \sigma^2 T)$, und der Stetigkeitssatz, Satz 1.6, liefert die Behauptung. □

9.5.2. Konvergenz der Optionspreise

Für ein Derivat $\xi_N = f(S_N^{1,N})$ bezeichne $\pi_N(f)$ seinen Preis im N-ten CRR-Modell. Nach Satz 9.21 und wegen $(1+r_N)^N = \mathrm{e}^{\rho T}$ gilt

$$\pi_N(f) = \mathrm{e}^{-\rho T}\,\mathbb{E}_{Q_N}\big[f\big(S_0^1 \mathrm{e}^{X_N}\big)\big].$$

Satz 9.23 (Konvergenz gegen den Black–Scholes-Preis)**.** *Sei $f\colon [0,\infty) \to \mathbb{R}$ stetig und beschränkt. Dann gilt*

$$\lim_{N\to\infty} \pi_N(f) = \mathrm{e}^{-\rho T}\,\mathbb{E}\Big[f\Big(S_0^1 \exp\big((\rho - \sigma^2/2)T + \sigma\sqrt{T}\,Z\big)\Big)\Big], \tag{9.6}$$

wobei $Z \sim \mathcal{N}(0,1)$. Insbesondere gilt dies für den europäischen Put mit $f(x) = (K-x)^+$ und Ausübungspreis $K > 0$.

Beweis. Die Funktion $x \mapsto f(S_0^1 \mathrm{e}^x)$ ist stetig und beschränkt auf $\mathbb{R}$. Nach Lemma 9.22 und der Definition der schwachen Konvergenz, Definition 2.15, konvergiert $\mathbb{E}_{Q_N}[f(S_0^1\mathrm{e}^{X_N})]$ gegen $\mathbb{E}[f(S_0^1\mathrm{e}^X)]$. Da X dieselbe Verteilung wie $(\rho-\sigma^2/2)T+\sigma\sqrt{T}\,Z$ hat, folgt die Behauptung. □

Bemerkung. Die Voraussetzung der Beschränktheit lässt sich zu $|f(x)| \le C(1+x)$ abschwächen, da die Familie $(\mathrm{e}^{X_N})_N$ unter $(Q_N)_N$ gleichgradig integrierbar ist. Für den Call umgehen wir dies im Folgenden mit der Put-Call-Parität.

Der Grenzwert in (9.6) lässt sich für Call und Put explizit berechnen. Dazu bezeichne $F_{0,1}(x) \coloneqq \mathbb{P}(Z \le x)$ die Verteilungsfunktion der Standardnormalverteilung.

Lemma 9.24. *Sei $X \sim \mathcal{N}(\mu, s^2)$ mit $\mu \in \mathbb{R}$, $s > 0$, und sei $c \in \mathbb{R}$. Dann gilt*

$$\mathbb{P}(X \le c) = F_{0,1}\Big(\frac{c-\mu}{s}\Big) \qquad \text{und} \qquad \mathbb{E}\big[\mathrm{e}^X \mathbb{1}\{X \le c\}\big] = \mathrm{e}^{\mu+s^2/2} F_{0,1}\Big(\frac{c-\mu-s^2}{s}\Big).$$

Beweis. Die erste Identität folgt aus $(X-\mu)/s \sim \mathcal{N}(0,1)$. Für die zweite ergänzen wir quadratisch, $x - (x-\mu)^2/(2s^2) = \mu + s^2/2 - (x-\mu-s^2)^2/(2s^2)$, und erhalten

$$\mathbb{E}\big[\mathrm{e}^X \mathbb{1}\{X \le c\}\big] = \int_{-\infty}^{c} \frac{\mathrm{e}^x}{\sqrt{2\pi s^2}}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2s^2}}\,\mathrm{d}x = \mathrm{e}^{\mu+s^2/2}\int_{-\infty}^{c} \frac{1}{\sqrt{2\pi s^2}}\,\mathrm{e}^{-\frac{(x-\mu-s^2)^2}{2s^2}}\,\mathrm{d}x,$$

und das letzte Integral ist $\mathbb{P}(X' \le c)$ für $X' \sim \mathcal{N}(\mu+s^2, s^2)$. □

Korollar 9.25 (Black–Scholes-Formel)**.** *Sei $K > 0$, und seien*

$$d_\pm \coloneqq \frac{\log(S_0^1/K) + (\rho \pm \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Dann konvergieren die Preise des europäischen Puts und des europäischen Calls mit Ausübungspreis K im N-ten CRR-Modell für $N \to \infty$ gegen

$$\pi_{\text{Put}}^{\text{BS}} = K\mathrm{e}^{-\rho T} F_{0,1}(-d_-) - S_0^1 F_{0,1}(-d_+), \tag{9.7}$$

$$\pi_{\text{Call}}^{\text{BS}} = S_0^1 F_{0,1}(d_+) - K\mathrm{e}^{-\rho T} F_{0,1}(d_-). \tag{9.8}$$

Beweis. Sei $X \sim \mathcal{N}(\mu, s^2)$ mit $\mu = (\rho - \sigma^2/2)T$ und $s = \sigma\sqrt{T}$, und sei $c \coloneqq \log(K/S_0^1)$. Dann gilt $(c-\mu)/s = -d_-$ und $(c-\mu-s^2)/s = -d_+$ sowie $\mu + s^2/2 = \rho T$. Mit Satz 9.23 und Lemma 9.24 folgt für den Put

$$\begin{aligned}\lim_{N\to\infty} \pi_N\big((K-x)^+\big) &= \mathrm{e}^{-\rho T}\Big(K\,\mathbb{P}(X \le c) - S_0^1\,\mathbb{E}\big[\mathrm{e}^X \mathbb{1}\{X \le c\}\big]\Big) \\ &= K\mathrm{e}^{-\rho T} F_{0,1}(-d_-) - S_0^1 F_{0,1}(-d_+).\end{aligned}$$

Für den Call verwenden wir $(x-K)^+ = (K-x)^+ + x - K$. Da der diskontierte Preisprozess ein Q_N-Martingal ist, gilt $\mathrm{e}^{-\rho T}\mathbb{E}_{Q_N}[S_N^{1,N}] = S_0^1$, also nach Satz 9.21

$$\pi_N\big((x-K)^+\big) = \pi_N\big((K-x)^+\big) + S_0^1 - K\mathrm{e}^{-\rho T};$$

dies ist die Put-Call-Parität, vergleiche Korollar 8.24. Mit $1 - F_{0,1}(-d) = F_{0,1}(d)$ folgt (9.8) aus (9.7). □

Abbildung 9.3 zeigt die Konvergenz der CRR-Preise eines Calls gegen den Black–Scholes-Preis. Die Konvergenz ist nicht monoton; die Preise oszillieren je nach Parität von N, weil der Ausübungspreis abwechselnd näher an und weiter weg von einem Knoten des Baumes liegt.

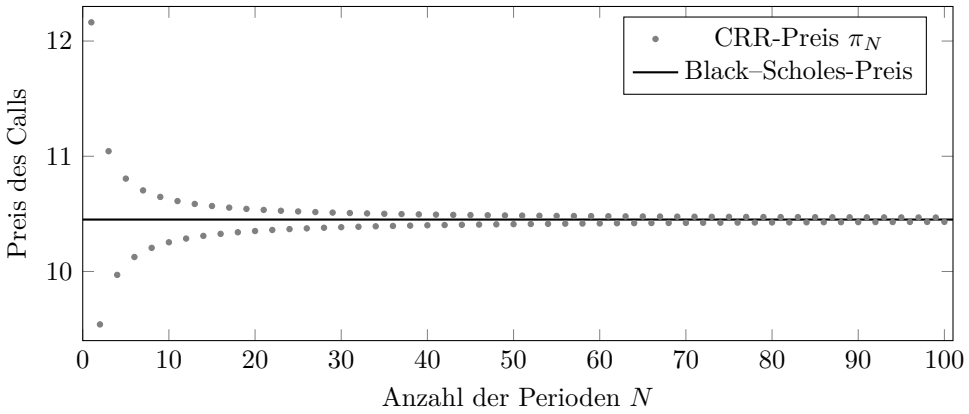

Abbildung 9.3.: CRR-Preise eines europäischen Calls in Abhängigkeit von der Anzahl N der Perioden und der Grenzwert (9.8), für $S_0^1 = K = 100$, $T = 1$, $\rho = 0{,}05$ und $\sigma = 0{,}2$; der Black–Scholes-Preis beträgt hier etwa 10,45.

Literaturverzeichnis

[Bau02] H. Bauer, *Wahrscheinlichkeitstheorie*, 5. ed., de Gruyter, Berlin, 2002.

[Bil99] P. Billingsley, *Convergence of probability measures*, 2. ed., Wiley, New York, 1999.

[Bir31] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. USA **17** (1931), 656–660.

[Boc33] S. Bochner, *Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse*, Math. Ann. **108** (1933), 378–410.

[BS73] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ. **81** (1973), 637–654.

[CRR79] J. C. Cox, S. A. Ross, and M. Rubinstein, *Option pricing: a simplified approach*, J. Financ. Econ. **7** (1979), 229–263.

[DMW90] R. C. Dalang, A. Morton, and W. Willinger, *Equivalent martingale measures and no-arbitrage in stochastic securities market models*, Stochastics Stochastics Rep. **29** (1990), 185–201.

[Doo53] J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.

[DS06] F. Delbaen and W. Schachermayer, *The mathematics of arbitrage*, Springer, Berlin, 2006.

[Dur19] R. Durrett, *Probability: Theory and examples*, 5. ed., Cambridge University Press, Cambridge, 2019.

[FS02] H. Föllmer and A. Schied, *Stochastic finance: An introduction in discrete time*, de Gruyter, Berlin, 2002.

[Geo15] H.-O. Georgii, *Stochastik: Einführung in die wahrscheinlichkeitstheorie und statistik*, 5. ed., de Gruyter, Berlin, 2015.

[HK79] J. M. Harrison and D. M. Kreps, *Martingales and arbitrage in multiperiod securities markets*, J. Econom. Theory **20** (1979), 381–408.

[HP81] J. M. Harrison and S. R. Pliska, *Martingales and stochastic integrals in the theory of continuous trading*, Stochastic Process. Appl. **11** (1981), 215–260.

[Kle13] A. Klenke, *Wahrscheinlichkeitstheorie*, 3. ed., Springer Spektrum, Berlin, Heidelberg, 2013.

[Lév25] P. Lévy, *Calcul des probabilités*, Gauthier-Villars, Paris, 1925.

[Lin22] J. W. Lindeberg, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, Math. Z. **15** (1922), 211–225.

[Mer73] R. C. Merton, *Theory of rational option pricing*, Bell J. Econ. Manag. Sci. **4** (1973), 141–183.

[Pro56] Yu. V. Prohorov, *Convergence of random processes and limit theorems in probability theory*, Theory Probab. Appl. **1** (1956), 157–214.

[Shr04] S. E. Shreve, *Stochastic calculus for finance I: The binomial asset pricing model*, Springer, New York, 2004.

[vN32] J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Natl. Acad. Sci. USA **18** (1932), 70–82.

[Wal82] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer, New York, 1982.

[Wil91] D. Williams, *Probability with martingales*, Cambridge University Press, Cambridge, 1991.