# Stochastische Modelle

Peter Eichelsbacher

Sommersemester 2009

# Inhaltsverzeichnis

KAPITEL 1

# Die Theorie der Martingale

Die von J. L. Doob begründete Theorie der Martingale ist zentral für die Untersuchung der fast sicheren Konvergenz von Folgen von Zufallsvariablen. Martingale werden mittels bedingter Erwartungen definiert.

Es sei $(X_n)_n$ eine unabhängige Folge integrierbarer reeller Zufallsvariablen, $S_n := X_1 + \cdots + X_n$. Offenbar gilt $\sigma(S_1, \ldots, S_n) = \sigma(X_1, \ldots, X_n)$. Da $X_{n+1}$ unabhängig von $\sigma(X_1, \ldots, X_n)$ ist, also auch von $\sigma(S_1, \ldots, S_n)$, folgt

$$\mathbb{E}\big(X_{n+1}|S_1, \ldots, S_n\big) = \mathbb{E}(X_{n+1}) \quad \text{P-f.s.}$$

Weiter gilt $\mathbb{E}(X_i|S_1, \ldots, S_n) = X_i$ P-f.s. für alle $i = 1, \ldots, n$, also bei Summation

$$\mathbb{E}(S_{n+1}|S_1, \ldots, S_n) = S_n + \mathbb{E}(X_{n+1}) \quad \text{P-f.s.}$$

Sind alle $X_n$ zentriert, so folgt

$$\mathbb{E}(S_{n+1}|S_1, \ldots, S_n) = S_n \quad \text{P-f.s.}$$

Gilt $\mathbb{E}(X_n) \leq 0$ (bzw. $\geq 0$) für alle $n$, so folgt entsprechend

$$\mathbb{E}(S_{n+1}|S_1, \ldots, S_n) \leq S_n \quad (\geq S_n) \quad \text{P-f.s.}$$

jeweils für alle $n \in \mathbb{N}$.

Sei speziell $X_n$ $\pm 1$-wertig mit $P(X_n = 1) = p$ und somit $\mathbb{E}(X_n) = 2p - 1$ (Spiel zwischen Spieler und Bank eines Spielcasinos). Der Spieler verfolge die folgende Strategie:

Vor Eintritt in das Spiel wählt er eine Folge $(e_n)_n$, $e_n : \{-1, +1\}^n \to \mathbb{R}^+$. Für die $(n+1)$-te Spielrunde wird der vom Ausgang der vorangegangenen $n$ Runden abhängige Einsatz $e_n(X_1, \ldots, X_n)$ geleistet. Die erste Runde wird ohne vorherigen Einsatz gespielt. Der Gegenspieler leiste keine Einsätze. $S_n$ bezeichne den Gesamtgewinn des Spielers nach der $n$-ten Runde:

$$\begin{aligned} S_1 &:= X_1 \\ S_{n+1} &:= S_n + e_n(X_1, \ldots, X_n) \cdot X_{n+1}\,. \end{aligned}$$

Es gilt:

$$\begin{aligned} \mathbb{E}(S_{n+1}|X_1, \ldots, X_n) &= S_n + e_n(X_1, \ldots, X_n)\mathbb{E}(X_{n+1}|X_1, \ldots, X_n) \\ &= S_n + e_n(X_1, \ldots, X_n)\mathbb{E}(X_{n+1}) \\ &= S_n + (2p-1)e_n(X_1, \ldots, X_n). \end{aligned}$$

Für $p = 1/2$ folgt

$$\mathbb{E}(S_{n+1}|X_1, \dots, X_n) = S_n \quad \text{P-f.s.}$$

Analog für $p < 1/2$ ($p > 1/2$):

$$\mathbb{E}(S_{n+1}|X_1, \dots, X_n) \leq S_n \quad (\geq S_n) \quad \text{P-f.s.}$$

Wir nehmen diese Beobachtungen zum Anlass für die folgenden Definitionen. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum.

**1.1 Definition**

(a) Eine *Filtrierung* von $\mathcal{A}$ ist eine Folge $(\mathcal{A}_n)_n$ von Teil-$\sigma$-Algebren von $\mathcal{A}$ mit $\mathcal{A}_m \subset \mathcal{A}_n$ für $m \leq n$. Eine Folge von Zufallsvariablen $(X_n)_n$ auf $(\Omega, \mathcal{A}, P)$ heißt $(\mathcal{A}_n)_n$-*adaptiert* (oder *angepasst*), wenn $X_n$ für jedes $n \in \mathbb{N}$ $\mathcal{A}_n$-messbar ist.

(b) Sei $(X_n)_n$ eine $(\mathcal{A}_n)_n$-adaptierte Folge von integrierbaren Zufallsgrößen. $(X_n)_n$ heißt

- $(\mathcal{A}_n)_n$-*Martingal*, falls $\mathbb{E}(X_n|\mathcal{A}_m) = X_m$ P-f.s. für alle $m, n \in \mathbb{N}$ mit $m \leq n$ gilt.
- $(\mathcal{A}_n)_n$-*Submartingal*, falls $\mathbb{E}(X_n|\mathcal{A}_m) \geq X_m$ P-f.s. für alle $m, n \in \mathbb{N}$ mit $m \leq n$ gilt.
- $(\mathcal{A}_n)_n$-*Supermartingal*, falls $\mathbb{E}(X_n|\mathcal{A}_m) \leq X_m$ P-f.s. für alle $m, n \in \mathbb{N}$ mit $m \leq n$ gilt.

**1.2 Bemerkungen**

(a) $(X_n)_n$ ist adaptiert an $(\mathcal{F}_n^X)_n$ mit $\mathcal{F}_n^X := \sigma(X_j, j \leq n)$. Man spricht bei dieser Filtrierung auch von der *kanonischen Filtrierung*.

(b) Um nachzuprüfen, dass $(X_n)_n$ ein Martingal ist, reicht es natürlich, dass

$$\mathbb{E}(X_{n+1}|\mathcal{A}_n) = X_n \quad \text{P-f.s.}$$

für alle $n \in \mathbb{N}$ gilt. (Entsprechendes für Sub- bzw. Supermartingale.) Dies folgt aus der P-fast sicheren Gleichheit

$$\mathbb{E}(X_{n+2}|\mathcal{A}_n) = E\big(E(X_{n+2}|\mathcal{A}_{n+1})|\mathcal{A}_n\big) = E(X_{n+1}|\mathcal{A}_n) = X_n.$$

(c) Das Wort „Martingal“ hat mehrere Bedeutungen. Es steht für „Hilfszügel beim Zaumzeug“ eines Reitpferdes, welches zu starke Kopfbewegungen des Pferdes verhindert, aber auch für ein die Takelage bei Segelschiffen absicherndes Seil. Vor allem bedeutet es aber wohl eine Spielstrategie beim Roulettespiel, im Provenzalischen genannt „jouga a la martegalo“. Diese Strategie besteht in der jeweiligen Verdoppelung des beim vorausgegangenen Spiel verlorenen Einsatzes, ein Spezialfall des in den Vorbemerkungen betrachteten Beispiels, welches wir in Beispiel 1.3 (g) genauer betrachten.

Bei der Wahl der kanonischen Filtrierung $(\mathcal{F}_n^X)_n$ nennen wir die Filtrierung nicht extra.

Wir sammeln Beispiele:

### 1.3 Beispiele

(a) Die eindimensionale symmetrische Irrfahrt auf $\mathbb{Z}$ mit Zeitachse $\mathbb{N}_0$ ist ein Martingal.

(b) Sei $(X_n)_n$ die symmetrische Irrfahrt auf $\mathbb{N}$ mit Start in 1 und Absorption in 0. Also $p_{i,i+1} = p_{i,i-1} = 1/2$ für $i \in \mathbb{N}$ und $p_{0,0} = 1$. Also gilt für $n \in \mathbb{N}_0$ fast sicher

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n^X) = \mathbb{E}(X_{n+1}|X_1, \ldots, X_n) = \sum_{j \in \mathbb{N}_0} j p_{X_n,j} = X_n\,,$$

also ist diese Irrfahrt ein Martingal.

(c) In den Vorbemerkungen haben wir schon die Verallgemeinerung von (a) gesehen: Sind $(\xi_n)_n$, $n \in \mathbb{N}$, unabhängige und identisch verteilte, integrierbare Zufallsgrößen mit $\mathbb{E}(\xi_j) = 0$, so sei $X_0 := 0$, $X_n := \sum_{j=1}^n \xi_j$, $n \in \mathbb{N}$, und $\mathcal{A}_0 := \{\emptyset, \Omega\}$ und $\mathcal{A}_n := \sigma(\xi_j,\, j \leq n)$. Dann ist $(X_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal.

(d) (Levy's *Martingal*) Es sei $(\mathcal{A}_n)_n$ eine beliebige Filtrierung und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Dann ist die Folge der bedingten Erwartungswerte

$$X_n := \mathbb{E}(X|\mathcal{A}_n)\,, \quad n \in \mathbb{N}\,,$$

offensichtlich ein Martingal, denn es gilt P-f.s.

$$\mathbb{E}(X_{n+1}|\mathcal{A}_n) = \mathbb{E}\big(\mathbb{E}(X|\mathcal{A}_{n+1})|\mathcal{A}_n\big) = \mathbb{E}(X|\mathcal{A}_n)\,.$$

Nicht jedes Martingal kann in dieser Weise dargestellt werden, wie wir noch sehen werden.

(e) (*exponentielles Martingal*) Die $\xi_n$, $n \in \mathbb{N}$, und die Filtrierung seien wie in (c) gewählt. Es sei $\mathbb{E}\xi_j \neq 0$ zugelassen. Es existiere ein $\lambda_0 > 0$ mit $M(\lambda) := \mathbb{E}(e^{\lambda \xi_j}) < \infty$ für $|\lambda| \leq \lambda_0$.

Sei $X_0 \equiv 1$ und

$$X_n := \exp\Big(\lambda \sum_{j=1}^n \xi_j\Big) M(\lambda)^{-n}\,, \quad n \in \mathbb{N}\,.$$

Dann ist für die entsprechenden Werte von $\lambda$ die Folge $(X_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal, denn es gilt P-f.s.

$$\begin{aligned}
\mathbb{E}(X_{n+1}|\mathcal{A}_n) &= M(\lambda)^{-n-1}\exp\big(\lambda\sum_{j=1}^{n}\xi_j\big)\mathbb{E}(e^{\lambda\xi_{n+1}}|\mathcal{A}_n)\\
&= M(\lambda)^{-n-1}\exp\big(\lambda\sum_{j=1}^{n}\xi_j\big)\mathbb{E}(e^{\lambda\xi_{n+1}})\\
&= M(\lambda)^{-n}\exp\big(\lambda\sum_{j=1}^{n}\xi_j\big) = X_n\,.
\end{aligned}$$

(f) Durch Differentiation nach $\lambda$ in $\lambda = 0$ lassen sich aus Beispiel (e) neue Martingale gewinnen. Einmalige Differentiation liefert (c), zweimaliges Differenzieren:

$$\frac{d^2}{d\lambda^2}X_n\Big|_{\lambda=0} = S_n^2 - 2nS_nM'(0) + n(n+1)M'(0)^2 - nM''(0)$$

mit $S_n := \sum_{j=1}^n \xi_j,\ M'(0) = \mathbb{E}\xi_j,\ M''(0) = \mathbb{E}(\xi_j^2)$. Die rechte Seite ist ein $(\mathcal{A}_n)_n$-Martingal (Übung), sofern $\xi_j \in \mathcal{L}^2(\Omega,\mathcal{A},P)$.

(g) (*Martingal-Wettstrategie*) Sei $(X_n)_n$ die Folge $\pm1$-wertiger unabhängiger Zufallsgrößen mit $P(X_i = 1) = 1/2$. Der Spieler[1] setzt 1 Einheit Einsatz zu Beginn und verdoppelt seinen Einsatz so lange, bis zum ersten Mal 1 erscheint, um dann aufzuhören! $M_n$ bezeichne den Gewinn nach $n$ Würfen mit $M_0 = 0$. Im Fall eines Gewinns hört der Spieler auf: $P(M_{n+1} = 1|M_n = 1) = 1$. Erscheint $n$ Mal stets $-1$, beträgt der Verlust $1+2+4+\cdots+2^{n-1} = 2^n - 1$ Einheiten. Dann wird der Einsatz auf $2^n$ verdoppelt. Es gilt:

$$P\big(M_{n+1} = 1|M_n = -(2^n-1)\big) = P(X_{n+1} = 1) = \frac{1}{2}$$

und

$$P\big(M_{n+1} = -(2^{n+1}-1)|M_n = -(2^n-1)\big) = P(X_{n+1} = -1) = \frac{1}{2}\,,$$

also

$$\mathbb{E}\big(M_{n+1}|M_n = -(2^n-1)\big) = 1\cdot\frac{1}{2} + (-2^{n+1}+1)\cdot\frac{1}{2} = -2^n+1\,,$$

also ist $(M_n)_n$ ein Martingal bezüglich $(\mathcal{A}_n)_n$ mit $\mathcal{A}_n := \sigma(M_0, X_1, \ldots, X_n)$.

[1] siehe einleitendes Beispiel

**1.4 Lemma** *Sei $(X_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal (bzw. Submartingal).*

(a) *Ist $(\mathcal{B}_n)_n$ eine weitere Filtrierung von $\mathcal{A}$ mit $\mathcal{B}_n \subset \mathcal{A}_n$ für alle $n \in \mathbb{N}$ und ist $(X_n)_n$ an $(\mathcal{B}_n)_n$ adaptiert, so ist $(X_n)_n$ auch ein $(\mathcal{B}_n)_n$-Martingal (bzw. Submartingal)*

(b) *Es gilt $\mathbb{E}(X_n) = \mathbb{E}(X_m)$ (bzw. $\mathbb{E}(X_n) \geq \mathbb{E}(X_m)$) für alle $m, n \in \mathbb{N}$ mit $m \leq n$.*

(c) *Ist $\varphi$ eine konvexe Funktion mit $\mathbb{E}\,|\varphi(X_n)| < \infty$, $n \in \mathbb{N}$, und $(X_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal, so ist $\big(\varphi(X_n)\big)_n$ ein Submartingal bzgl. $(\mathcal{A}_n)_n$.*

*Beweis:* Definition 1.1 und in (c) verwende man die Jensensche Ungleichung für bedingte Erwartungen, siehe Satz 13.7., Wahrscheinlichkeitstheorie. □

Martingale beschreiben im Sinne der einführenden Beispiele faire Spiele. Beispiel 2.3 (g) beschreibt eine Strategie, mit der in einem fairen Spiel Gewinn gemacht werden kann! Es gilt für $\tau := \inf\{n \geq 0; M_n = 1\}$ stets $M_\tau = 1$, während $M_0 = 0$ ist. $\tau$ ist dabei eine zufällige Zeit. Bei gewissen Einschränkungen an diese zufällige Zeit $\tau$ werden wir aber sehen, dass man „in einem fairen Spiel" keinen Gewinn machen kann. Dies wird der Inhalt des sogenannten „Optional-Sampling-Theorems" werden.

**1.5 Definition** Sei $(\mathcal{A}_n)_n$ eine Filtrierung von $\mathcal{A}$. Eine Abbildung $\tau : \Omega \to \mathbb{N} \cup \{+\infty\} =: \bar{\mathbb{N}}$ heißt *Stoppzeit*, falls $\{\tau \leq n\} \in \mathcal{A}_n$ für alle $n \in \mathbb{N}$ gilt.

**1.6 Bemerkungen**

(a) $\tau : \Omega \to \bar{\mathbb{N}}$ ist genau dann eine Stoppzeit, wenn $\{\tau = n\} \in \mathcal{A}_n$ für alle $n \in \mathbb{N}$ gilt. Denn

$$\{\tau = n\} = \{\tau \leq n\} \backslash \{\tau \leq n-1\} \in \mathcal{A}_n\,,$$

falls $\tau$ eine Stoppzeit ist. Außerdem ist $\{\tau \leq n\} = \bigcup_{m: m \leq n} \{\tau = m\} \in \mathcal{A}_n$, falls alle $\{\tau = n\} \in \mathcal{A}_n$, $n \in \mathbb{N}$.

(b) Ist $\tau$ eine Stoppzeit, so gilt

$$\{\tau = \infty\} = \Big(\bigcup_{n \in \mathbb{N}} \{\tau \leq n\}\Big)^c \in \mathcal{A}_\infty := \sigma\Big(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n\Big)\,.$$

**1.7 Beispiele**

(a) Konstante Abbildungen $\tau : \Omega \to \bar{\mathbb{N}}$ sind Stoppzeiten.

(b) $(X_n)_n$ sei eine Folge von Zufallsvariablen mit Werten in einem Messraum $(E, \mathcal{E})$, $\mathcal{A}_n := \sigma(X_j, j \leq n)$, $n \in \mathbb{N}$, und $A \in \mathcal{E}$. $\tau_A : \Omega \to \bar{\mathbb{N}}$ sei definiert durch

$$\tau_A(\omega) := \inf\big\{n \in \mathbb{N} : X_n(\omega) \in A\big\}\,.$$

Setze $\inf \emptyset = \infty$. Dann ist $\tau_A$ eine Stoppzeit, denn

$$\{\tau_A \le n\} = \bigcup_{m=1}^{n} \{X_m \in A\} \in \mathcal{A}_n\,, \quad n \in \mathbb{N}\,.$$

Sei weiter $\tau$ eine Stoppzeit, so setzen wir

$$\mathcal{A}_\tau := \Big\{A \in \mathcal{A}_\infty = \sigma\big(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n\big) | A \cap \{\tau \le n\} \in \mathcal{A}_n \text{ für alle } n \in \mathbb{N}\Big\}\,.$$

Ein Ereignis liegt in $\mathcal{A}_\tau$, wenn es durch den Ablauf des „Prozesses“ bis zum Zeitpunkt $\tau$ festgelegt ist.

**1.8 Lemma** *$\mathcal{A}_\tau$ ist eine $\sigma$-Algebra.*

*Beweis:* $\Omega \in \mathcal{A}_\tau$, da $\tau$ eine Stoppzeit ist. Ist $A \in \mathcal{A}_\tau$, so gilt für jedes $n \in \mathbb{N}$

$$A^c \cap \{\tau \le n\} = \{\tau \le n\} \backslash \big(A \cap \{\tau \le n\}\big) \in \mathcal{A}_n\,.$$

Sind $A_k \in \mathcal{A}_\tau$ für jedes $k \in \mathbb{N}$, so folgt für jedes $n \in \mathbb{N}$

$$\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) \cap \{\tau \le n\} = \bigcup_{k\in\mathbb{N}} \big(A_k \cap \{\tau \le n\}\big) \in \mathcal{A}_n\,. \qquad \square$$

**1.9 Lemma** *Ist $\tau \equiv n$ für ein $n \in \mathbb{N}$, so ist $\mathcal{A}_\tau = \mathcal{A}_n$.*

*Beweis:* Übung. $\square$

**1.10 Lemma** *Es seien $\tau$ und $\sigma$ zwei Stoppzeiten.*

(a) *Das Minimum $\tau \wedge \sigma := \min\{\tau, \sigma\}$ und das Maximum $\tau \vee \sigma := \max\{\tau, \sigma\}$ sind Stoppzeiten.*

(b) *Gilt $\sigma(\omega) \le \tau(\omega)$ für alle $\omega \in \Omega$, so gilt $\mathcal{A}_\sigma \subset \mathcal{A}_\tau$.*

(c) *Die Stoppzeit $\tau$ ist $\mathcal{A}_\tau$-messbar.*

(d) *$\{\sigma \le \tau\} \in \mathcal{A}_\sigma \cap \mathcal{A}_\tau$ und allgemeiner ist für alle $A \in \mathcal{A}_\sigma$ das Ereignis $A \cap \{\sigma \le \tau\}$ in $\mathcal{A}_\sigma \cap \mathcal{A}_\tau$.*

(e) *Ist $\tau' : \Omega \to \bar{\mathbb{N}}$ eine $\mathcal{A}_\tau$-messbare Abbildung und gilt $\tau' \ge \tau$, so ist $\tau'$ eine Stoppzeit.*

(f) *$\tau + \sigma$ ist eine Stoppzeit.*

*Beweis:*

(a) $n \in \mathbb{N}$: $\{\tau \vee \sigma \leq n\} = \{\tau \leq n\} \cap \{\sigma \leq n\}$ und $\{\tau \wedge \sigma \leq n\} = \{\tau \leq n\} \cup \{\sigma \leq n\}$.

(b) Sei $A \in \mathcal{A}_\sigma$. Mit $\{\tau \leq n\} \subset \{\sigma \leq n\}$ und $A \cap \{\sigma \leq n\} \in \mathcal{A}_n$ folgt $A \cap \{\tau \leq n\} = \big(A \cap \{\sigma \leq n\}\big) \cap \{\tau \leq n\} \in \mathcal{A}_n$ für jedes $n \in \mathbb{N}$.

(c) Für $n, m \in \mathbb{N}$ gilt

$$\{\tau \leq n\} \cap \{\tau \leq m\} = \{\tau \leq n \wedge m\} \in \mathcal{A}_{n \wedge m} \subset \mathcal{A}_m \,.$$

Dies gilt für alle $m$, also $\{\tau \leq n\} \in \mathcal{A}_\tau$ für alle $n$. Damit ist $\tau$ $\mathcal{A}_\tau$-messbar.

(d) Wir zeigen: $A \cap \{\sigma \leq \tau\} \in \mathcal{A}_{\sigma \wedge \tau}$ für alle $A \in \mathcal{A}_\sigma$. Mit (a) und (b) ist $\mathcal{A}_{\sigma \wedge \tau} \subset \mathcal{A}_\sigma \cap \mathcal{A}_\tau$. Für $n \in \mathbb{N}$ ist

$$\begin{aligned} A \cap \{\sigma \leq \tau\} \cap \{\sigma \wedge \tau \leq n\} &= A \cap \{\sigma \leq \tau\} \cap \{\sigma \leq n\} \\ &= A \cap \{\sigma \leq n\} \cap \bigcup_{m=1}^{n} \{\sigma \leq m\} \cap \{\tau \geq m\} \in \mathcal{A}_n \,, \end{aligned}$$

denn $A \cap \{\sigma \leq n\} \in \mathcal{A}_n$ und $\{\sigma \leq m\} \cap \{\tau \geq m\} \in \mathcal{A}_m \subset \mathcal{A}_n$. Also $A \cap \{\sigma \leq \tau\} \in \mathcal{A}_{\sigma \wedge \tau}$.

(e) Für $n \in \mathbb{N}$ gilt $\{\tau' \leq n\} = \{\tau' \leq n\} \cap \{\tau \leq n\} \in \mathcal{A}_n$, da $\{\tau' \leq n\}$ nach Voraussetzung in $\mathcal{A}_\tau$ ist.

(f)

$$\{\tau + \sigma \leq n\} = \bigcup_{\substack{k,l \in \mathbb{N}: \\ k+l \leq n}} \{\tau \leq k\} \cap \{\sigma \leq l\} \in \mathcal{A}_n.$$

□

Sei $(\mathcal{A}_n)_{n \in \mathbb{N}_0}$ eine Filtrierung von $\mathcal{A}$.

**1.11 Definition**

(a) Eine Folge $\mathbb{V} := (V_n)_{n \in \mathbb{N}}$ von Zufallsgrößen heißt *vorhersehbar* bzgl. $(\mathcal{A}_n)_{n \in \mathbb{N}_0}$, wenn $V_n$ für jedes $n \in \mathbb{N}$ bezüglich $\mathcal{A}_{n-1}$ messbar ist.

(b) Sind $\mathbb{V}$ und $\mathbb{X} := (X_n)_{n \in \mathbb{N}_0}$ zwei Folgen von Zufallsgrößen, so definieren wir eine neue Folge $\mathbb{V} \bullet \mathbb{X} = (Y_n)_{n \in \mathbb{N}_0}$ durch $Y_0 = 0$ und $Y_n := \sum_{k=1}^n V_k(X_k - X_{k-1})$ für alle $n \in \mathbb{N}$. $\mathbb{V} \bullet \mathbb{X}$ heißt *Martingaltransformation* (von $\mathbb{X}$ unter $\mathbb{V}$), wenn $\mathbb{X}$ ein Martingal ist.

**1.12 Bemerkung** Ist $\mathbb{X}$ ein Martingal, so werden die Zuwächse (Martingaldifferenzen) $X_k - X_{k-1}$ durch $V_k$ gewichtet. In einer Spielsituation gibt $X_k - X_{k-1}$ den Ausgang der $k$-ten Spielrunde an und $V_k$ den Einsatz des Spielers, den er vorher festzulegen hat („vorhersehbar“).

**1.13 Lemma**

(a) *Sei* $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ *ein Supermartingal und* $\mathbb{V} = (V_n)_{n\in\mathbb{N}}$ *ein vorhersehbarer „Prozess“ mit* $V_n \geq 0$ *und* $\|V_n\|_\infty < \infty$ *für alle* $n \in \mathbb{N}$*. Dann ist* $\mathbb{V} \bullet \mathbb{X}$ *ein Supermartingal. Ist* $\mathbb{X}$ *ein Martingal, so auch* $\mathbb{V} \bullet \mathbb{X}$*.*

(b) *In (a) kann die Voraussetzung* $\|V_n\|_\infty < \infty$ *ersetzt werden durch* $\|V_n\|_2 < \infty$*, falls* $\|X_n\|_2 < \infty$ *für alle* $n \in \mathbb{N}_0$ *gilt.*

*Beweis:* Sei $\mathbb{V} \bullet \mathbb{X} = (Y_n)_{n\in\mathbb{N}_0}$. Für $n \in \mathbb{N}$ folgt unter den Voraussetzungen in (a) oder (b) die Integrierbarkeit von $Y_n$. Weiter gilt P-f.s.

$$\mathbb{E}(Y_n|\mathcal{A}_{n-1}) = Y_{n-1} + V_n\mathbb{E}(X_n - X_{n-1}|\mathcal{A}_{n-1}) \leq Y_{n-1}\,.$$

Die letzte Ungleichung ist eine Gleichung, falls $\mathbb{X}$ ein Martingal ist. □

Wir untersuchen den wichtigen Spezialfall von Martingalstransformationen mittels einer Stoppzeit $\tau$:

Sei $V_n^\tau := 1_{\{n\leq\tau\}}$, $n \in \mathbb{N}$. Da $\{n \leq \tau\} = \{\tau \leq n-1\}^c \in \mathcal{A}_{n-1}$, ist $\mathbb{V}^\tau = (V_n^\tau)_{n\in\mathbb{N}}$ vorhersehbar. Für jede Folge $(X_n)_{n\in\mathbb{N}_0}$ folgt dann

$$\begin{aligned}(\mathbb{V}^\tau \bullet \mathbb{X})_n &= \sum_{k=1}^{n} 1_{\{\tau>k-1\}}(X_k - X_{k-1})\\ &= \sum_{k=1}^{\tau\wedge n}(X_k - X_{k-1}) = X_{\tau\wedge n} - X_0\,.\end{aligned}$$

Wir definieren den zur Zeit $\tau$ *gestoppten Prozess* $\mathbb{X}^\tau$ durch $\mathbb{X}^\tau = (X_{\tau\wedge n})_{n\in\mathbb{N}}$.

**1.14 Satz (Optional-Sampling-Theorem)** *Es seien* $\mathbb{X}$ *ein Supermartingal und* $\tau$ *eine Stoppzeit. Dann ist* $\mathbb{X}^\tau$ *ein Supermartingal. Ist* $\mathbb{X}$ *ein Martingal, so auch* $\mathbb{X}^\tau$*.*

Es folgt insbesondere $\mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_0)$ für alle $n \in \mathbb{N}_0$, wenn $\mathbb{X}$ ein Martingal und $\tau$ eine Stoppzeit ist. Aber daraus folgt im Allgemeinen nicht $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$, siehe etwa Beispiel 1.3 (g) oder:

**1.15 Beispiel** $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ sei die symmetrische Irrfahrt auf $\mathbb{N}_0$ mit Start in 1 und Absorbtion in 0. Nach Beispiel 1.3 (b) ist $\mathbb{X}$ ein $(\mathcal{A}_n^{\mathbb{X}})_{n\in\mathbb{N}_0}$-Martingal. Sei $\tau := \inf\{n \in \mathbb{N}_0 | X_n = 0\}$. Wegen der Rekurrenz der eindimensionalen Irrfahrt gilt $P(\tau < \infty) = 1$. Für jedes $n \in \mathbb{N}$ gilt

$$\{\tau \leq n\} = \bigcup_{k=0}^{n}\{X_k = 0\} \in \mathcal{A}_n^{\mathbb{X}}\,,$$

also ist $\tau$ eine $(\mathcal{A}_n^{\mathbb{X}})_{n\in\mathbb{N}_0}$-Stoppzeit. Es gilt $\mathbb{E}(X_0) = 1$, also $\mathbb{E}(X_{\tau\wedge n}) = 1$ für alle $n \in \mathbb{N}_0$ mittels Satz 1.14, aber offensichtlich ist $\mathbb{E}(X_\tau) = 0$.

Es gibt hinreichende Kriterien für $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$ („du kannst in einem fairen Spiel keinen Gewinn machen"):

**1.16 Satz** *Es seien $\tau$ eine Stoppzeit und $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal (bzw. ein Martingal), die eine der folgenden Bedingungen erfüllen:*

(a) *$\tau$ ist beschränkt,*

(b) *$\tau$ ist P-f.s. endlich und $(X_n)_{n\in\mathbb{N}_0}$ ist beschränkt:*

$$P(\tau < \infty) = 1\,, \quad \sup_{n\in\mathbb{N}_0} \|X_n\|_\infty < \infty\,,$$

(c) *$\mathbb{E}(\tau) < \infty$ und $\sup_{n\in\mathbb{N}} \|X_n - X_{n-1}\|_\infty < \infty$.*

*Dann ist $X_\tau$ integrierbar und es gilt*

$$\mathbb{E}(X_\tau) \le \mathbb{E}(X_0) \quad (\textit{bzw. } \mathbb{E}(X_\tau) = \mathbb{E}(X_0))\,.$$

*Beweis:* Nach Satz 1.14 gilt $\mathbb{E}(X_{\tau\wedge n}) \le \mathbb{E}(X_0)$ (bzw. $\mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_0)$) für jedes $n \in \mathbb{N}$. Ist (a) erfüllt, so ist $\tau \wedge n = \tau$ für ein genügend großes $n \in \mathbb{N}$. Unter (b) gilt $\lim_{n\to\infty} \mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_\tau)$ nach dem Satz von der majorisierten Konvergenz. Ist (c) erfüllt, so existiert ein $K > 0$ mit $P(|X_n - X_{n-1}| \le K) = 1$ für alle $n \in \mathbb{N}$. Also $|X_{\tau\wedge n} - X_0| \le K\tau$ fast sicher. Wegen $\mathbb{E}(\tau) < \infty$ folgt $\lim_{n\to\infty} \mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X_\tau)$ ebenfalls nach dem Satz von der majorisierten Konvergenz. □

**1.17 Korollar** *Seien $(X_n)_{n\in\mathbb{N}_0}$ ein positives Supermartingal und $\tau$ eine endliche Stoppzeit. Dann gilt $\mathbb{E}(X_\tau) \le \mathbb{E}(X_0)$.*

*Beweis:* Nach Satz 1.14 gilt $\mathbb{E}(X_{\tau\wedge n}) \le \mathbb{E}(X_0)$ und mittels des Lemmas von Fatou folgt

$$\mathbb{E}(X_\tau) = \mathbb{E}(\lim_{n\to\infty} X_{\tau\wedge n}) \le \lim_{n\to\infty} \mathbb{E}(X_{\tau\wedge n}) \le \mathbb{E}(X_0)\,. \qquad \square$$

Die betrachteten Stop- und Transformationstechniken werden wir auch zum Beweis des sogenannten „upcrossing"-Lemmas von Doob verwenden.

Seien $\alpha = (\alpha_n)_{n\in\mathbb{N}_0}$ eine Folge reeller Zahlen und $a, b \in \mathbb{R}$ mit $a < b$. Für $n \in \mathbb{N}_0$ ist die Anzahl der *aufsteigenden Überschreitungen* des Intervalls $[a, b]$ durch die Folge $\alpha$ bis zum Zeitpunkt $n$ definiert durch

$$U_n[a,b](\alpha) := \sup\big\{k \in \mathbb{N} \,|\, \text{es ex. } 0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \le n \\ \text{mit } \alpha_{s_i} < a,\ \alpha_{t_i} > b \text{ für alle } i \in \{1, \ldots, k\}\big\}\,,$$

wobei $\sup \emptyset := 0$ gesetzt wird.

Natürlich ist $U_n[a,b](\alpha) \leq n$. Wir setzen

$$U_\infty[a,b](\alpha) := \sup_{n\in\mathbb{N}_0} U_n[a,b](\alpha)\,.$$

**1.18 Lemma** *Eine Folge $\alpha = (\alpha_n)_{n\in\mathbb{N}_0}$ konvergiert genau dann in $\bar{\mathbb{R}}$, wenn $U_\infty[a,b](\alpha) < \infty$ für alle $a, b \in \mathbb{Q}$ mit $a < b$ gilt.*

*Beweis:* Wir argumentieren indirekt:

Es gilt:

$$\begin{aligned} & \liminf_{n\to\infty} \alpha_n < \limsup_{n\to\infty} \alpha_n \\ \Leftrightarrow \quad & \exists a, b \in \mathbb{Q} \text{ mit } \liminf_{n\to\infty} \alpha_n < a < b < \limsup_{n\to\infty} \alpha_n \\ \Leftrightarrow \quad & \exists a, b \in \mathbb{Q} \text{ mit } a < b \text{ und } U_\infty[a,b](\alpha) = \infty\,. \end{aligned}$$ □

**1.19 Satz (Ungleichung von Doob)** *Es sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und $a < b$. Dann gilt für jedes $n \in \mathbb{N}_0$:*

$$\mathbb{E}\big(U_n[a,b](\mathbb{X})\big) \leq \frac{1}{b-a}\mathbb{E}\big((X_n - a)^-\big)\,,$$

*wobei $x^- := \max\{0, -x\}$ für alle $x \in \mathbb{R}$.*

Bevor wir den Beweis geben, ziehen wir Konsequenzen:

**1.20 Satz (Doobscher Konvergenzsatz)** *Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal mit $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$. Dann existiert $X_\infty := \lim_{n\to\infty} X_n$ $P$-f.s. und ist integrierbar.*

*Beweis von Satz 1.20:* Seien $a < b$ zwei reelle Zahlen. Nach dem Satz von der monotonen Konvergenz gilt

$$(b-a)\mathbb{E}\big(U_\infty[a,b](\mathbb{X})\big) = (b-a)\lim_{n\to\infty} \mathbb{E}\big(U_n[a,b](\mathbb{X})\big)\,.$$

Die rechte Seite kann mit 1.19 nach oben durch

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}\big((X_n - a)^-\big) \leq |a| + \sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$$

abgeschätzt werden. Also ist $P\big(U_\infty[a,b](\mathbb{X}) < \infty\big) = 1$, also

$$P\Big(\bigcap_{\substack{a,b\in\mathbb{Q}\\ a<b}} \big\{U_\infty[a,b](\mathbb{X}) < \infty\big\}\Big) = 1\,.$$

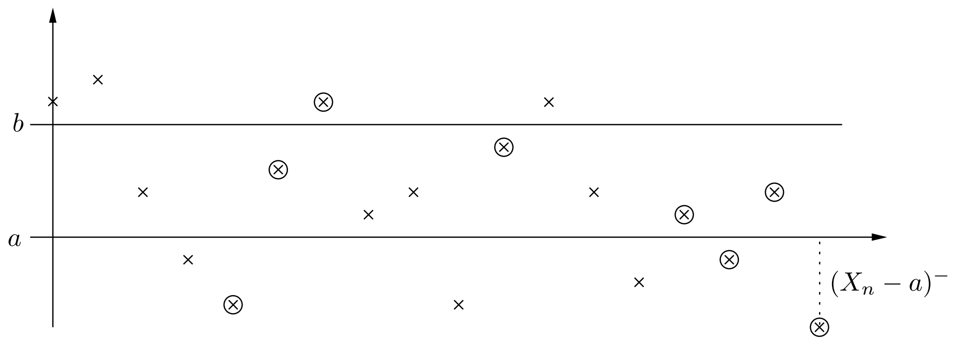



Abb. 1.1: Pfad $\mathbb{N}_0 \ni n \mapsto X_n(\omega)$ eines Supermartingals.

Somit existiert nach Lemma 1.18 $X_\infty = \lim_{n\to\infty} X_n$ in $\bar{\mathbb{R}}$ f.s. Nach dem Lemma von FATOU ist

$$\mathbb{E}|X_\infty| \le \liminf_{n\to\infty} \mathbb{E}|X_n| \le \sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty\,. \qquad \square$$

**1.21 Korollar** *Jedes nicht-negative Supermartingal $(X_n)_{n\in\mathbb{N}_0}$ konvergiert f.s. gegen eine integrierbare Zufallsgröße.*

*Beweis:* Aus $X_n \ge 0$ folgt $\mathbb{E}|X_n| = \mathbb{E}(X_n) \le \mathbb{E}(X_0)$, also $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$. $\square$

**1.22 Bemerkung** Die Bedingung $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$ reicht im Allgemeinen nicht für $\mathcal{L}^1$-Konvergenz aus. Siehe dazu Beispiel 1.15: Dort gilt $\mathbb{E}X_n = 1$ für alle $n \in \mathbb{N}$, $X_n \to 0$ f.s. (siehe Korollar), aber die Erwartungswerte konvergieren nicht, also gilt nicht $\mathcal{L}^1$-Konvergenz.

*Beweis von 1.19:* Wir konstruieren eine Martingaltransformation $\mathbb{Y} = (Y_n)_{n\in\mathbb{N}_0}$ von $\mathbb{X}$. Mit Hilfe der nachfolgenden Regeln nutzt dabei $\mathbb{Y}$ alle aufsteigenden Überschreitungen des Supermartingals $\mathbb{X}$, um möglichst weit nach oben zu gelangen:

(a) Starte mit $Y_0(\omega) = 0$. Ist $X_0 \ge a$, benutze Regel (b), andernfalls Regel (c).

(b) Warte so lange, das heißt setze $Y_n(\omega) = Y_{n-1}(\omega)$, bis $X_n(\omega) < a$ ist. Benutze dann für den nächsten Schritt Regel (c).

(c) Nutze die Zuwächse, das heißt setze $Y_n(\omega) = Y_{n-1}(\omega) + X_n(\omega) - X_{n-1}(\omega)$, bis $X_n > b$. Benutze dann für den nächsten Schritt Regel (b).

Da der Prozess $\mathbb{Y}$ jedes Mal mindestens die Höhe $(b-a)$ gewinnt, wenn $\mathbb{X}$ das Intervall $[a,b]$ aufsteigend überschreitet, und $\mathbb{Y}$ seit der letzten Überschreitung höchstens die Höhe $(X_n - a)^-$ verloren haben kann, gilt für jedes

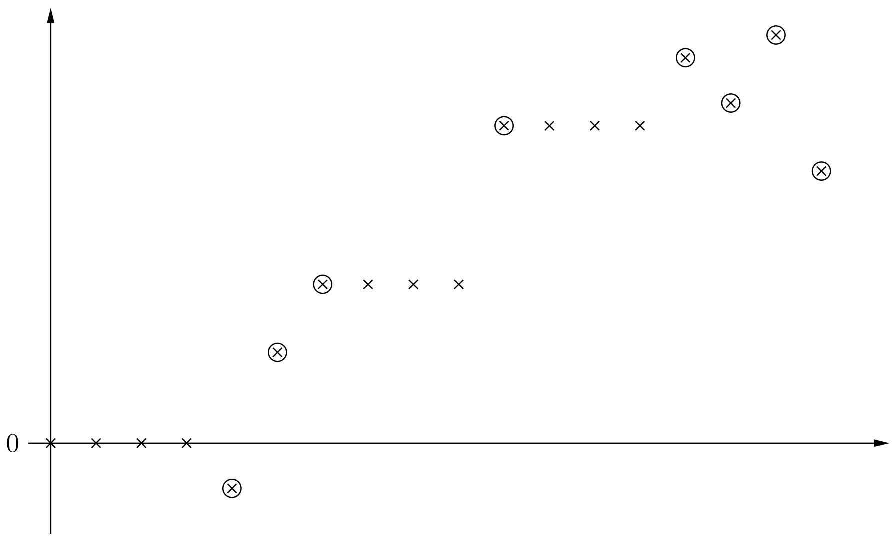



Abb. 1.2: Der zu Abb. 1.1 gehörige Pfad $\mathbb{N}_0 \ni n \mapsto Y_n(\omega)$ der Martingaltransformation.

$n \in \mathbb{N}_0$

$$Y_n \geq (b-a)U_n[a,b](\mathbb{X}) - (X_n - a)^- \,. \tag{1.1}$$

$\mathbb{Y}$ ist eine Martingaltransformation von $\mathbb{X}$:

$$V_n := \begin{cases} 1_{\{X_0<a\}} & \text{für } n = 1\,, \\ 1_{\{V_{n-1}=1, X_{n-1}\leq b\}} + 1_{\{V_{n-1}=0, X_{n-1}<a\}} & \text{für } n \geq 2\,. \end{cases}$$

Damit ist nach Definition $\mathbb{V} = (V_n)_{n\in\mathbb{N}}$ vorhersehbar und

$$Y_n = \sum_{k=1}^{n} V_k(X_k - X_{k-1})$$

für alle $n \in \mathbb{N}_0$, also ist $\mathbb{Y}$ ein Supermartingal nach Lemma 1.13 (a), also gilt:

$$\mathbb{E}(Y_n) \leq \mathbb{E}(Y_0) = 0\,,$$

und dies liefert mit (1.1) die behauptete Ungleichung. □

**1.23 Beispiel (Polyas Urnenschema)** In einer Urne liegen $R_n$ rote und $S_n$ schwarze Kugeln zum Zeitpunkt $n \in \mathbb{N}_0$. Im Zeitintervall $(n, n+1)$ wird die Urne gut gemischt, eine Kugel zufällig gezogen und zusammen mit einer zusätzlichen Kugel der gleichen Farbe zurückgelegt. Zum Zeitpunkt 0 sei $R_0 = S_0 = 1$. Dann ist $\big\{(R_n, S_n)\big\}_{n\in\mathbb{N}_0}$ eine Markovkette mit Zustandsraum $\mathbb{N}^2$ und Übergangswahrscheinlichkeiten $p\big((r,s),(r+1,s)\big) = r/(r+s)$ und

$p\big((r,s),(r,s+1)\big) = s/(r+s)$, $(r,s) \in \mathbb{N}^2$. $R_n + S_n = n+2$, $n \in \mathbb{N}_0$. Sei $\mathcal{A}_n := \sigma\big((R_i,S_i), i \le n\big)$ und

$$X_n := \frac{R_n}{n+2}\,, \quad n \in \mathbb{N}_0\,.$$

Dann ist $(X_n)_{n\in\mathbb{N}_0}$ ein $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$-Martingal, denn für jedes $n \in \mathbb{N}_0$ gilt:

$$\begin{aligned}\mathbb{E}(X_{n+1}\,|\,\mathcal{A}_n) &= \mathbb{E}\Big(\frac{R_{n+1}}{n+3}\,|(R_n,S_n)\Big)\\ &= \frac{R_n}{n+2}\frac{R_n+1}{n+3} + \frac{(n+2-R_n)}{n+2}\frac{R_n}{n+3}\\ &= \frac{R_n}{n+2} = X_n\,.\end{aligned}$$

Es gilt $X_n \ge 0$, also existiert $X_\infty := \lim_{n\to\infty} X_n$ f.s. nach Korollar 1.21.

Etwas Kombinatorik führt zu $P(R_n = j) = 1/(n+1)$ für jedes $j \in \{1,\dots,n+1\}$. Ist $f : [0,1] \to \mathbb{R}$ stetig, so gilt

$$\mathbb{E}f(X_n) = \frac{1}{n+1}\sum_{j=1}^{n+1} f\Big(\frac{j}{n+2}\Big)\,,$$

was für $n \to \infty$ gegen $\int_0^1 f(x)\,dx$ konvergiert.

Andererseits folgt aus $X_n \to X_\infty$ und dem Satz von Lebesgue von der majorisierten Konvergenz

$$\mathbb{E}f(X_\infty) = \lim_{n\to\infty} \mathbb{E}f(X_n)\,.$$

Also ist $\mathbb{E}f(X_\infty) = \int_0^1 f(x)\,dx$. Dies gilt für jede stetige Funktion, und somit ist die Verteilung von $X_\infty$ das Lebesgue-Maß auf $[0,1]$.

Wir diskutieren nun die $\mathcal{L}^2$-Konvergenz von Martingalen.

**1.24 Definition** Ein Martingal $(X_n)_{n\in\mathbb{N}_0}$ heißt $\mathcal{L}^2$-*Martingal*, wenn für jedes $n \in \mathbb{N}_0$ die Zufallsgröße $X_n$ quadratisch integrierbar ist.

**1.25 Satz** *Sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein $\mathcal{L}^2$-Martingal. Folgende Aussagen sind äquivalent:*

(a) $\sup_{n\in\mathbb{N}_0} \mathbb{E}(X_n^2) < \infty$,

(b) $\sum_{k=1}^\infty \mathbb{E}\big((X_k - X_{k-1})^2\big) < \infty$,

(c) $(X_n)_{n\in\mathbb{N}_0}$ *konvergiert $P$-f.s. und in $\mathcal{L}^2$.*

*Beweis:* Quadratisch integrierbare Martingale haben unkorrelierte Zuwächse: Für alle $m, n \in \mathbb{N}_0$ mit $m < n$ gilt

$$\mathbb{E}\big((X_n - X_m)^2\big) = \sum_{k=m+1}^{n} \mathbb{E}\big((X_k - X_{k-1})^2\big)\,. \tag{1.2}$$

Wir zeigen (1.2) via Induktion nach $n$: Für $n = m + 1$ ist (1.2) klar und für $n \geq m + 2$ gilt

$$\begin{aligned}\mathbb{E}\big((X_n - X_m)^2\big) = \mathbb{E}\big((X_n - X_{n-1})^2\big) + \mathbb{E}\big((X_{n-1} - X_m)^2\big) \\ + 2\mathbb{E}\big((X_n - X_{n-1})(X_{n-1} - X_m)\big)\,.\end{aligned}$$

Für den dritten Summanden liefert das Einschieben eines bedingten Erwartungswertes:

$$\begin{aligned}&\mathbb{E}\big((X_n - X_{n-1})(X_{n-1} - X_m)\big) \\ &\qquad = \mathbb{E}\Big(\mathbb{E}\big((X_n - X_{n-1})(X_{n-1} - X_m)|\mathcal{A}_{n-1}\big)\Big) \\ &\qquad = \mathbb{E}\big((X_{n-1} - X_m)\mathbb{E}(X_n - X_{n-1}|\mathcal{A}_{n-1})\big) = 0.\end{aligned}$$

Mit der Induktionsvoraussetzung für $n - 1$ folgt (1.2) für $n$.

Für jedes $n \in \mathbb{N}_0$ folgt aus $\mathbb{E}(X_n\, X_0) = \mathbb{E}\big(\mathbb{E}(X_n|\mathcal{A}_0)X_0\big) = \mathbb{E}(X_0^2)$ und (1.2):

$$\mathbb{E}(X_n^2) - \mathbb{E}(X_0^2) = \mathbb{E}\big((X_n - X_0)^2\big) = \sum_{k=1}^{n} \mathbb{E}\big((X_k - X_{k-1})^2\big)\,,$$

also die Äquivalenz von (a) und (b). (a) folgt aus der $\mathcal{L}^2$-Konvergenz in (c).

Wir zeigen noch, dass (c) aus (a) und (b) folgt: Aus (a) folgt $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$ und somit mit Satz 1.20, dass $X_\infty := \lim_{n\to\infty} X_n$ P-f.s. existiert. Aus dem Lemma von Fatou und (1.2) folgt

$$\begin{aligned}\mathbb{E}\big((X_\infty - X_m)^2\big) &\leq \liminf_{n\to\infty} \mathbb{E}\big((X_n - X_m)^2\big) \\ &= \sum_{k=m+1}^{\infty} \mathbb{E}\big((X_k - X_{k-1})^2\big)\,,\end{aligned}$$

was nach (b) für $m \to \infty$ gegen Null konvergiert. □

**1.26 Bemerkung** Aus der $\mathcal{L}^2$-Beschränktheit eines Martingals folgt Konvergenz in $\mathcal{L}^2$, aus der $\mathcal{L}^1$-Beschränktheit folgt P-fast sichere Konvergenz, aber nicht $\mathcal{L}^1$-Konvergenz, siehe Beispiel 1.15.

Wir diskutieren nun, unter welchen Zusatzbedingungen neben $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty$ ein Martingal in $\mathcal{L}^1$ konvergiert.

**1.27 Definition** Eine Teilmenge $\Gamma$ von $\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{A}, P)$ heißt *gleichgradig integrierbar*, falls

$$\lim_{n\to\infty} \sup_{X\in\Gamma} \int_{\{|X|\geq n\}} |X|\, dP = 0$$

gilt.

*Notation:* Wir schreiben $\mathbb{E}(X; A)$ für $\int_A X\, dP$ also etwa

$$\mathbb{E}\big(|X|; |X| \geq n\big) \quad \text{für} \quad \int_{\{|X|\geq n\}} |X|\, dP\,.$$

**1.28 Lemma** *Es sei $X \in \mathcal{L}^1$. Dann gilt*

$$\lim_{\varepsilon\searrow 0} \sup\big\{\mathbb{E}(|X|; A) | A \in \mathcal{A}\,,\ P(A) \leq \varepsilon\big\} = 0\,.$$

*Beweis:* Angenommen, es existiert eine Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ mit $P(A_n) \leq 2^{-n}$ für alle $n \in \mathbb{N}$ und $\limsup_{n\to\infty} \mathbb{E}(|X|; A_n) > 0$. Für jedes $n \in \mathbb{N}$ sei $B_n := \bigcup_{k=n}^{\infty} A_k$. Dann gilt $B_n \downarrow \emptyset$ für $n \to \infty$ und mit dem Satz von der majorisierten Konvergenz folgt

$$\limsup_{n\to\infty} \mathbb{E}(|X|; A_n) \leq \lim_{n\to\infty} \mathbb{E}(|X|; B_n) = 0\,.$$

Dies ist ein Widerspruch. □

**1.29 Lemma** *Jede der folgenden Bedingungen ist hinreichend für die gleichgradige Integrierbarkeit einer Familie $\Gamma \subset \mathcal{L}^1$:*

(a) *Es existiert $p \in (1, \infty)$ mit $\sup_{X\in\Gamma} \|X\|_p < \infty$ .*

(b) *Es existiert eine Zufallsgröße $Y \in \mathcal{L}^1$ mit $|X| \leq Y$ f.s. für alle $X \in \Gamma$.*

(c) *Es existiert $Y \in \mathcal{L}^1$ und eine Familie $\Phi$ von Teil-$\sigma$-Algebren von $\mathcal{A}$ mit*

$$\Gamma = \{\mathbb{E}(Y|\mathcal{G}) : \mathcal{G} \in \Phi\}\,.$$

*Beweis:*

(a) Für alle $n \in \mathbb{N}$ und $X \in \Gamma$ gilt

$$\begin{aligned}\mathbb{E}(|X|; |X| \geq n) &\leq \mathbb{E}\Big(|X|\frac{|X|^{p-1}}{n^{p-1}}; |X| \geq n\Big)\\ &\leq \frac{\|X\|_p^p}{n^{p-1}} \leq \frac{1}{n^{p-1}} \sup_{X\in\Gamma} \|X\|_p^p\end{aligned}$$

und dies konvergiert gegen Null für $n \to \infty$.

(b) Für alle $X \in \Gamma$ und $n \in \mathbb{N}$ gilt

$$\mathbb{E}(|X|;|X| \geq n) \leq \mathbb{E}(Y;Y \geq n)\,,$$

und dies konvergiert nach dem Satz von der majorisierten Konvergenz gegen Null für $n \to \infty$.

(c) Für $n \in \mathbb{N}$ und $\mathcal{G} \in \Phi$ sei

$$A_n(\mathcal{G}) := \big\{|\mathbb{E}(Y|\mathcal{G})| \geq n\big\}\,.$$

Es gilt $|\mathbb{E}(Y|\mathcal{G})| \leq \mathbb{E}(|Y|\,|\,\mathcal{G})$. Somit liefert die MARKOVsche Ungleichung

$$P\big(A_n(\mathcal{G})\big) \leq \frac{1}{n}\mathbb{E}\big(|\mathbb{E}(Y|\mathcal{G})|\big) \leq \frac{1}{n}\mathbb{E}\big(\mathbb{E}(|Y|\,|\,\mathcal{G})\big) = \frac{1}{n}\mathbb{E}(|Y|)\,.$$

Also gilt $\lim_{n\to\infty}\sup_{\mathcal{G}\in\Phi} P\big(A_n(\mathcal{G})\big) = 0$. Nun ist $A_n(\mathcal{G}) \in \mathcal{G}$, also

$$\begin{aligned}\mathbb{E}\big(|\mathbb{E}(Y|\mathcal{G})|;A_n(\mathcal{G})\big) &\leq \mathbb{E}\big(\mathbb{E}(|Y|\,|\,\mathcal{G});A_n(\mathcal{G})\big)\\ &= \mathbb{E}\big(|Y|;A_n(\mathcal{G})\big)\,.\end{aligned}$$

Somit liefert Lemma 1.28 die gleichgradige Integrierbarkeit von $\Gamma$. □

**1.30 Satz** *Es seien $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}^1$ und $X \in \mathcal{L}^1$. Die Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert genau dann in $\mathcal{L}^1$ gegen $X$, wenn $(X_n)_{n\in\mathbb{N}}$ in W.keit gegen $X$ konvergiert und $(X_n)_n$ gleichgradig integrierbar ist.*

*Beweis:*

(a) $(X_n)_n$ konvergiere in $\mathcal{L}^1$ gegen $X$. Dann konvergiert $(X_n)_n$ in W.keit gegen $X$ (Satz 9.9 in Warscheinlichkeitstheorie). Für $k, n \in \mathbb{N}$ ist

$$P(|X_n| \geq k) \leq \frac{\|X_n\|_1}{k}\,.$$

Konvergiert $(X_n)_n$ im ersten Mittel, so ist $\sup_{n\in\mathbb{N}}\|X_n\|_1 < \infty$. Damit folgt

$$\lim_{k\to\infty}\sup_{n\in\mathbb{N}} P(|X_n| \geq k) = 0\,. \tag{1.3}$$

Ist $N \in \mathbb{N}$, so gilt

$$\begin{aligned}\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|;|X_n| \geq k) \leq{}& \sup_{n\leq N}\mathbb{E}(|X_n|;|X_n| \geq k)\\ &+ \sup_{n\geq N}\|X_n - X\|_1 + \sup_{n\geq N}\mathbb{E}(|X|;|X_n| \geq k)\,.\end{aligned}$$

Mit (1.3) und Lemma 1.28 konvergiert der dritte Summand für $k \to \infty$ gegen Null. Jede endliche Familie von integrierbaren Zufallsgrößen ist natürlich gleichgradig integrierbar, also konvergiert der erste Summand für $k \to \infty$ gegen Null. Da $N$ beliebig war, folgt die gleichgradige Integrierbarkeit von $(X_n)_n$.

(b) Sei nun $(X_n)_n$ gleichgradig integrierbar und konvergiere in W.keit gegen $X$. Für $k \in \mathbb{N}$ und $x \in \mathbb{R}$ sei

$$\varphi_k(x) := (-k) \vee (x \wedge k)\,.$$

Für $\varepsilon > 0$ und $k, n \in \mathbb{N}$ gilt dann

$$\|\varphi_k(X_n) - \varphi_k(X)\|_1 \leq \varepsilon + 2kP(|X_n - X| > \varepsilon)\,.$$

Die Konvergenz in W.keit liefert

$$\lim_{n\to\infty} \|\varphi_k(X_n) - \varphi_k(X)\|_1 = 0\,,$$

da $\varepsilon > 0$ beliebig war, und dies für jedes $k \in \mathbb{N}$. Nun ist

$$\begin{aligned}\|X_n - X\|_1 &\leq \|X_n - \varphi_k(X_n)\|_1 + \|\varphi_k(X_n) - \varphi_k(X)\|_1 + \|\varphi_k(X) - X\|_1 \\ &\leq 2\,\mathbb{E}(|X_n|; |X_n| \geq k) + \|\varphi_k(X_n) - \varphi_k(X)\|_1 + 2\,\mathbb{E}(|X|; |X| \geq k)\,,\end{aligned}$$

also

$$\limsup_{n\to\infty} \|X_n - X\|_1 \leq 2 \sup_{n\in\mathbb{N}} E(|X_n|; |X_n| \geq k) + 2\,\mathbb{E}(|X|; |X| \geq k)\,.$$

Da $k$ beliebig ist, folgt aus der gleichgradigen Integrierbarkeit, dass $\|X_n - X\|_1$ für $n \to \infty$ gegen Null konvergiert. □

Nun zur $\mathcal{L}^1$-Konvergenz von Martingalen:

**1.31 Satz** *Es sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ ein $(\mathcal{A}_n)_n$-Martingal. Dann sind die folgenden Aussagen äquivalent:*

(a) *Das Martingal $(X_n)_n$ ist gleichgradig integrierbar.*

(b) *Das Martingal $(X_n)_n$ konvergiert P-f.s. und in $\mathcal{L}^1$.*

(c) *Es existiert ein $Y \in \mathcal{L}^1$ mit $X_n = \mathbb{E}(Y|\mathcal{A}_n)$ P-f.s. für alle $n \in \mathbb{N}$.*

*Erfüllt $(X_n)_n$ eine dieser Bedingungen, so kann für $Y$ in (c) insbesondere der P-fast sichere und $\mathcal{L}^1$-Limes des Martingals $(X_n)_n$ gewählt werden.*

*Beweis:* (c) ⇒ (a) folgt aus Lemma 1.29 (c).

(a) ⇒ (b) folgt aus Satz 1.30, wenn wir zeigen, dass $(X_n)_n$ P-f.s. und somit auch in W.keit konvergiert: Es existiert ein $k \in \mathbb{N}$ mit $\mathbb{E}(|X_n|; |X_n| > k) \leq 1$ für alle $n \in \mathbb{N}$. Also

$$\begin{aligned}\sup_{n\in\mathbb{N}} \|X_n\|_1 &\leq \sup_{n\in\mathbb{N}} \big(\mathbb{E}(|X_n|; |X_n| \leq k) + \mathbb{E}(|X_n|; |X_n| > k)\big) \\ &\leq k + 1 < \infty\,.\end{aligned}$$

Nun liefert Satz 1.20 die P-fast sichere Konvergenz von $(X_n)_n$.

(b) $\Rightarrow$ (c) Seien $X_\infty$ der $\mathcal{L}^1$-Limes der Folge $(X_n)_n$ und $n \in \mathbb{N}$. Dann gilt für jedes $m \geq n$

$$\begin{aligned}\mathbb{E}\big(|X_n - \mathbb{E}(X_\infty|\mathcal{A}_n)|\big) &= \mathbb{E}\big(|\mathbb{E}(X_m - X_\infty)|\mathcal{A}_n)|\big) \\ &\leq \mathbb{E}\big(\mathbb{E}(|X_m - X_\infty| \,|\, \mathcal{A}_n)\big) \\ &= \|X_m - X_\infty\|_1 \,.\end{aligned}$$

Für $m \to \infty$ folgt $X_n = \mathbb{E}(X_\infty|\mathcal{A}_n)$ P-f.s. Hiermit ist auch die weitere Aussage bewiesen. □

**1.32 Korollar** *Für $Y \in \mathcal{L}^1$ ist $\big(\mathbb{E}(Y|\mathcal{A}_n)\big)_n$ ein $(\mathcal{A}_n)_n$-Martingal, das fast sicher und in $\mathcal{L}^1$ gegen $\mathbb{E}(Y|\mathcal{A}_\infty)$ konvergiert.*

*Beweis:* Für $n \in \mathbb{N}$ sei $X_n := \mathbb{E}(Y|\mathcal{A}_n)$. Dann ist $(X_n)_n$ ein Martingal. Nach Satz 1.31 konvergiert dieses Martingal P-f.s. und in $\mathcal{L}^1$ gegen eine Zufallsgröße $X_\infty$, die $\mathcal{A}_\infty$-messbar ist. Zu zeigen: $X_\infty = \mathbb{E}(Y|\mathcal{A}_\infty)$. Zeige $\mathbb{E}(X_\infty; A) = \mathbb{E}(Y; A)$ für alle $A \in \mathcal{A}_\infty$. Mit Satz 1.31 folgt dies aus $X_n = \mathbb{E}(X_\infty|\mathcal{A}_n)$ für alle $A \in \mathcal{A}_n$. Da $\bigcup_{n\in\mathbb{N}} \mathcal{A}_n$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{A}_\infty$ ist, folgt $\mathbb{E}(X_\infty; A) = \mathbb{E}(Y; A)$ für alle $A \in \mathcal{A}_\infty$. □

Im folgenden Kapitel erleben wir Martingale in Aktion!

KAPITEL 2

# Martingale in Aktion

Wir wollen in diesem Kapitel Anwendungen der Martingaltheorie behandeln. Wir betrachten *Produktmartingale*, den *Satz von* KAKUTANI und Anwendungen in der *asymptotischen Statistik*; wir beweisen den *Satz von* RADON-NIKODYM mittels Martingaltheorie; wir diskutieren KOLMOGOROVS *Kriterium* zu starken Gesetzen der großen Zahlen, untersuchen *verrauschte Beobachtungen* (ein einfaches Modell der sogenannten *Filtertheorie*) und betrachten einfache *Verzweigungsprozesse*.

## 2.1 Der Satz von Kakutani

Es sei $(\xi_n)_n$ eine Folge von unabhängigen nicht-negativen Zufallsgrößen mit $\mathbb{E}(\xi_n) = 1$. Sei ferner $\mathcal{A}_n := \sigma(\xi_i, i \leq n)$, $n \in \mathbb{N}$, und $M_n := \prod_{i=1}^n \xi_i$. Die Folge $(M_n)_n$ ist ein Martingal, denn P-f.s. gilt

$$\begin{aligned}\mathbb{E}(M_{n+1}|\mathcal{A}_n) &= \mathbb{E}(M_n\xi_{n+1}|\mathcal{A}_n) \\ &= M_n\mathbb{E}(\xi_{n+1}|\mathcal{A}_n) = M_n\mathbb{E}(\xi_{n+1}) = M_n \ .\end{aligned}$$

Korollar 1.21 besagt, dass $(M_n)_n$ $P$-f.s. gegen eine nicht-negative Zufallsgröße $M_\infty$ konvergiert. Das Lemma von FATOU liefert $\mathbb{E}(M_\infty) \leq \liminf_{n\to\infty} \mathbb{E}(M_n) = 1$. Nach Satz 1.31 konvergiert $(M_n)_n$ genau dann in $\mathcal{L}^1$ gegen $M_\infty$, wenn $(M_n)_n$ gleichgradig integrierbar ist. In diesem Fall ist dann $\mathbb{E}(M_\infty) = 1$.

**2.1 Lemma**

(a) *Das Martingal $(M_n)_n$ ist genau dann gleichgradig integrierbar, wenn*

$$\prod_{n\in\mathbb{N}} \mathbb{E}\big(\sqrt{\xi_n}\big) > 0 \ .$$

(b) *Ist $(M_n)_n$ nicht gleichgradig integrierbar, so gilt $M_\infty = 0$ $P$-f.s.*

(c) *Ist $(M_n)_n$ gleichgradig integrierbar und $\xi_n > 0$ $P$-f.s. für alle $n \in \mathbb{N}$, so gilt auch $M_\infty > 0$ $P$-f.s.*

*Beweis:* Für $n \in \mathbb{N}$ seien $a_n := \mathbb{E}\big(\sqrt{\xi_n}\big)$ und $b_n := \prod_{i=1}^n a_i$. Mittels CAUCHY-SCHWARZ folgt $a_n \leq 1$, wegen $P(\xi_n = 0) \neq 1$ ist $a_n > 0$.

Für $n \in \mathbb{N}$ sei $N_n := \prod_{i=1}^n \frac{\sqrt{\xi_i}}{a_i}$. Dann ist $(N_n)_n$ ein $(\mathcal{A}_n)_n$-Martingal (s.o.), das nach Korollar 1.21 $P$-f.s. gegen eine nicht-negative Zufallsgröße $N_\infty$ konvergiert. Ist $b_\infty := \prod_{i=1}^\infty a_i > 0$, so folgt

$$K := \sup_{n\in\mathbb{N}} \mathbb{E}(N_n^2) = \sup_{n\in\mathbb{N}} \prod_{i=1}^n \frac{\mathbb{E}(\xi_i)}{a_i^2} = \prod_{i=1}^\infty \frac{1}{a_i^2} = \frac{1}{b_\infty^2} < \infty \ ,$$

nach Satz 1.25 konvergiert $(N_n)_n$ gegen $N_\infty$ in $\mathcal{L}^2$. Es gilt $M_n = b_n^2 N_n^2$, $n \in \mathbb{N}$, also $M_\infty = b_\infty^2 N_\infty^2$ $P$-f.s. Weiter liefert CAUCHY-SCHWARZ

$$\begin{aligned} ||M_n - M_\infty||_1 &= ||(b_n N_n + b_\infty N_\infty)(b_n N_n - b_\infty N_\infty)||_1 \\ &\leq ||b_n N_n + b_\infty N_\infty||_2 \ ||b_n N_n - b_\infty N_\infty||_2 \\ &\leq 2\sqrt{K}\big(b_n ||N_n - N_\infty||_2 + |b_n - b_\infty| \ ||N_\infty||_2\big) \ . \end{aligned}$$

In der letzten Ungleichung haben wir $\|b_n N_n + b_\infty N_\infty\|_2 \leq \|b_n N_n\|_2 + \|b_\infty N_\infty\|_2$ sowie

$$\|b_n N_n - b_\infty N_\infty\|_2 \leq \|b_n N_n - b_n N_\infty\|_2 + \|b_n N_\infty - b_\infty N_\infty\|_2$$

verwendet. Also konvergiert $(M_n)_n$ gegen $M_\infty$ im ersten Mittel. Nach Satz 1.31 ist $(M_n)_n$ gleichgradig integrierbar. Ist $\prod_{i=1}^\infty a_i = 0$, folgt aus der $P$-f.s. Konvergenz von $(N_n)_n$ gegen $N_\infty$ sofort, dass $(M_n)_n$ $P$-f.s. gegen Null konvergiert. Also gilt $M_\infty = 0$ $P$-f.s. und $(M_n)_n$ konvergiert nicht im ersten Mittel, kann also nach Satz 1.31 nicht gleichgradig integrierbar sein. Somit sind (a) und (b) bewiesen.

(c): $(M_n)_n$ konvergiert $P$-f.s. und in $\mathcal{L}^1$ gegen $M_\infty$. Also ist $\mathbb{E}(M_\infty) = 1$ und $P(M_\infty = 0) \neq 1$. Sei $B_n := \{\prod_{i=n}^\infty \xi_i = 0\}$. Wegen $\prod_{i=1}^{n-1} \xi_i > 0$ $P$-f.s. gilt $P\Big(\{M_\infty = 0\}\Delta B_n\Big) = 0$. Da $(B_n)_n$ eine absteigende Folge ist, gilt

$$\bigcap_{n\in\mathbb{N}} B_n = \limsup_{n\to\infty} B_n \in \bigcap_{k\in\mathbb{N}} \sigma(\xi_j, j \geq k) \ .$$

Nach KOLMOGOROVs 0-1-Gesetz ist $P(\bigcap_{n\in\mathbb{N}} B_n) \in \{0,1\}$ und

$$P\Big(\{M_\infty = 0\}\Delta\big(\bigcap_{n\in\mathbb{N}} B_n\big)\Big) \leq \sum_{n\in\mathbb{N}} P\Big(\{M_\infty = 0\}\Delta B_n\Big) = 0 \ .$$

Aus $P(M_\infty = 0) \neq 1$ folgt somit $P(M_\infty = 0) = 0$. $\square$

Wir betrachten nun die folgende Situation: Es seien $(E, \mathcal{E})$ ein Messraum und $(\mu_n)_n$ und $(\nu_n)_n$ zwei Folgen von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{E})$. Für jedes $n \in \mathbb{N}$ besitze $\nu_n$ eine Dichte $f_n$ bezüglich $\mu_n$ und umgekehrt $\mu_n$ eine Dichte bezüglich $\nu_n$. Wir können ohne Einschränkung annehmen, dass

$f_n(x) > 0$ für alle $x \in E$ gilt. Auf $(\Omega, \mathcal{A}) := (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ seien $P := \otimes_{n\in\mathbb{N}}\mu_n$ und $Q := \otimes_{n\in\mathbb{N}}\nu_n$ die zugehörigen Produktmaße.

Hat $P$ eine Dichte bezüglich $Q$ und umgekehrt $Q$ eine Dichte bezüglich $P$? Kakutanis Satz gibt ein notwendiges und hinreichendes Kriterium.

Für $n \in \mathbb{N}$ seien $\pi_n : \Omega \to E$ die Projektion auf den $n$-ten Faktor und $\mathcal{A}_n = \sigma(\pi_i, i \leq n)$. Dann ist $(\mathcal{A}_n)_n$ eine Filtrierung von $\mathcal{A}$.

Sei $\xi_n := f_n \circ \pi_n$. Unter $P$ ist $(\xi_n)_n$ eine Folge unabhängiger Zufallsgrößen mit $\mathbb{E}_P(\xi_n) = 1$ für alle $n \in \mathbb{N}$. Sei $a_n := \mathbb{E}_P\big(\sqrt{\xi_n}\big) = \int_E \sqrt{f_n}\, d\mu_n$.

**2.2 Satz (von Kakutani)**

(a) *Ist $\prod_{n=1}^\infty a_n > 0$, so sind $P$ und $Q$ äquivalent, d.h. es existiert eine Dichte von $P$ bzgl. $Q$ und eine Dichte von $Q$ bzgl. $P$. Es gilt*

$$\frac{dQ}{dP} = M_\infty := \lim_{n\to\infty}\prod_{i=1}^n \xi_i \quad P\text{-f.s.} \quad \textit{und}$$
$$\frac{dP}{dQ} = M_\infty^{-1} \quad P\text{-f.s.}$$

(b) *Ist $\prod_{n=1}^\infty a_n = 0$, so sind $P$ und $Q$ zueinander singulär, das heißt, es existiert ein $A \in \mathcal{A}$ mit $P(A) = 1$ und $Q(A) = 0$.*

*Beweis:*

(a) Für jedes $n \in \mathbb{N}$ seien $M_n := \prod_{i=1}^n \xi_i$; $P_n := P|_{\mathcal{A}_n}$, $Q_n := Q|_{\mathcal{A}_n}$. Dann sind $P_n$ und $Q_n$ zueinander äquivalent mit $\frac{dQ_n}{dP_n} = M_n$ $P$-f.s. (zum Beispiel nach Satz 3.7, Wahrscheinlichkeitstheorie). Nach Lemma 2.1 und Satz 1.31 konvergiert $(M_n)_n$ $P$-f.s. in $\mathcal{L}^1(P)$ gegen $M_\infty$. Weiter gilt $M_n = \mathbb{E}_P(M_\infty|\mathcal{A}_n)$ $P$-f.s. für jedes $n \in \mathbb{N}$. Also gilt für $A \in \mathcal{A}_n$

$$Q(A) = Q_n(A) = \int_A M_n\, dP_n = \int_A M_n\, dP = \int_A M_\infty\, dP\ .$$

Dies gilt für alle $A \in \bigcup_{n\in\mathbb{N}} \mathcal{A}_n$, und $\bigcup_{n\in\mathbb{N}} \mathcal{A}_n$ ist ein durchschnittstabiles Erzeugendensystem von $\mathcal{A}$, also gilt die Gleichheit für alle $A \in \mathcal{A}$. Damit ist $\frac{dQ}{dP} = M_\infty$ $P$-f.s. Wegen $M_\infty > 0$ $P$-f.s., siehe Lemma 2.1(c), ist $\frac{dP}{dQ} = M_\infty^{-1}$.

(b) Aus Lemma 2.1 wissen wir, dass $(M_n)_n$ $P$-f.s. gegen Null konvergiert. Wir zeigen nun, dass $(M_n)_n$ $Q$-f.s. gegen unendlich konvergiert. Da $(N_n)_n$ mit $N_n := \frac{\sqrt{M_n}}{\prod_{i=1}^n a_i}$ ein $P$-Martingal ist und $a_i \leq 1$ für alle $i \in \mathbb{N}$ gilt, ist $(\sqrt{M_n})_n$ ein $P$-Supermartingal: Es gilt P-f.s.

$$\mathbb{E}_P\big(\sqrt{M_n}|\mathcal{A}_{n-1}\big) = \Big(\prod_{i=1}^n a_i\Big)\, \mathbb{E}_P(N_n|\mathcal{A}_{n-1}) = a_n\sqrt{M_{n-1}} \leq \sqrt{M_{n-1}}.$$

Für $B \in \mathcal{A}_n$ gilt

$$\begin{aligned}\int_B \frac{1}{\sqrt{M_n}}\,dQ &= \int_B \frac{1}{\sqrt{M_n}} M_n\,dP = \int_B \sqrt{M_n}\,dP \\ &\geq \int_B \sqrt{M_{n+1}}\,dP = \int_B \frac{1}{\sqrt{M_{n+1}}}\,dQ\end{aligned}$$

($M_n$ und $M_{n+1}$ sind überall strikt positiv gewählt). Also ist $(\frac{1}{\sqrt{M_n}})_n$ ein positives $Q$-Supermartingal und konvergiert gemäß Korollar 1.21 $Q$-f.s. Wegen $\int_\Omega \frac{1}{\sqrt{M_n}}\,dQ = \int_\Omega \sqrt{M_n}\,dP = \prod_{i=1}^n a_i \searrow 0$ für $n \to \infty$ folgt

$$Q\Big(\lim_{n\to\infty} \frac{1}{\sqrt{M_n}} = 0\Big) = 1 \text{ , also } Q\Big(\lim_{n\to\infty} M_n = \infty\Big) = 1 \text{ .}$$

Somit ist $A := \{\omega \in \Omega : \lim_{n\to\infty} M_n(\omega) = 0\}$ eine $Q$-Nullmenge mit $P(A) = 1$. □

**2.3 Beispiel** Sei $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$, $\mu_n$ sei $N(0,1)$- und $\nu_n$ $N(\alpha_n, 1)$-verteilt, $\alpha_n \in \mathbb{R}$, $n \in \mathbb{N}$. Dann gilt für jedes $n \in \mathbb{N}$

$$\frac{d\nu_n}{d\mu_n}(x) = \frac{\exp\big(-(x-\alpha_n)^2/2\big)}{\exp\big(-x^2/2\big)} = \exp\big(\alpha_n x - \alpha_n^2/2\big) \text{ ,} \quad x \in \mathbb{R} \text{ .}$$

Weiter ist

$$\begin{aligned} a_n &:= \int_{\mathbb{R}} \sqrt{\frac{d\nu_n}{d\mu_n}}\,d\mu_n \\ &= \int_{\mathbb{R}} \exp\Big(\frac{\alpha_n x}{2} - \frac{\alpha_n^2}{4}\Big) \frac{1}{\sqrt{2\pi}} \exp\big(-x^2/2\big)\,dx \\ &= \exp\big(-\alpha_n^2/8\big) \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}\big(x - \frac{\alpha_n}{2}\big)^2\Big)\,dx \\ &= \exp\big(-\alpha_n^2/8\big) \text{ .}\end{aligned}$$

Somit ist $\prod_{n=1}^\infty a_n > 0 \Leftrightarrow \sum_{n=1}^\infty \alpha_n^2 < \infty$. Nach Satz 2.2 sind $P$ und $Q$ genau dann äquivalent, wenn $(\alpha_n)_n \in \ell_2$. Dann ist

$$\frac{dQ}{dP}(x) = \exp\Big(\sum_{n=1}^\infty \alpha_n x_n - \frac{1}{2}\sum_{n=1}^\infty \alpha_n^2\Big)$$

für $P$-fast alle $x = (x_n)_n \in \mathbb{R}^{\mathbb{N}}$. Die Reihe $\sum_{n=1}^\infty \alpha_n x_n$ konvergiert nur $P$-f.s. Außerhalb dieser Menge von $P$- und $Q$-Maß 1 können wir die Dichtefunktion nach Belieben setzen, zum Beispiel 1.

Als Anwendung machen wir einen kleinen Abstecher in die asymptotische Statistik. Wir untersuchen das Verhalten von Folgen von Dichtequotienten. Sind $P$ und $Q$ die obigen Produktmaße auf $(\Omega, \mathcal{A})$, so sind die $(\pi_n)_n$ unabhängige Zufallsgrößen. In der Statistik möchte man auf der Basis von Beobachtungen entscheiden, ob $P$ oder $Q$ vorliegt! Es sei nun $(\rho_n)_n$ eine weitere Folge von W-Maßen auf $(E, \mathcal{E})$ mit: $\mu_n$ besitzt eine strikt positive Dichte $f_n$ bzgl. $\rho_n$ und $\nu_n$ besitzt eine strikt positive Dichte $g_n$ bzgl. $\rho_n$ für jedes $n \in \mathbb{N}$. Dann ist (siehe Notation wie im Beweis von Satz 2.2)

$$\frac{dQ_n}{dP_n} = \prod_{i=1}^{n} \frac{g_i}{f_i} \circ \pi_i =: \prod_{i=1}^{n} \xi_i \quad \text{f.s.}$$

und $\mathbb{E}_P(\xi_i) = \int_E \frac{g_i(x)}{f_i(x)} f_i(x)\, d\rho_i(x) = \int_E g_i(x)\, d\rho_i(x) = \nu_i(E) = 1$, wie bereits gesehen. Weiter ist

$$\mathbb{E}_P(\sqrt{\xi_i}) = \int \frac{\sqrt{g_i(x)}}{\sqrt{f_i(x)}} f_i(x)\, d\rho_i(x) = \int \sqrt{g_i(x) f_i(x)}\, d\rho_i(x) \ ,$$

die Bedingung $\prod_{i=1}^{\infty} a_n > 0$ im Satz von Kakutani liest sich hier also zu

$$\prod_{i=1}^{\infty} \int_E \sqrt{g_i(x) f_i(x)}\, d\rho_i(x) > 0 \ , \tag{2.1}$$

was übrigens zu $\sum_{i \geq 1} \int_E \left( f_i(x)^{1/2} - g_i(x)^{1/2} \right)^2 d\rho_i(x) < \infty$ äquivalent ist.

**2.4 Korollar** *$P$ und $Q$ sind, wenn nicht äquivalent, bereits singulär und*

$$Q(M_n \to \infty) = 1 \quad \textit{und} \quad P(M_n \to 0) = 1 \ .$$

*Beweis:* Sind $P$ und $Q$ nicht äquivalent, so existiert keine Dichte von $Q$ bzgl. $P$, also ist das Martingal $(M_n)_n$ nicht gleichgradig integrierbar bzgl. $P$, also $M_n \to 0$ $P$-fast sicher. Analog zu Satz 2.2 ergibt sich $M_n^{-1} \to 0$ $Q$-fast sicher. □

Wir betrachten den Fall $\mu_1 = \mu_2 = \cdots$ und $\nu_1 = \nu_2 = \cdots$; somit sind alle Dichten $f_n$, sowie $g_n$, gleich, in Notation $f$ und $g$. Also ist

$$M_n = \prod_{k=1}^{n} \frac{g(\pi_k)}{f(\pi_k)} \quad , \ n \geq 0 \ .$$

Somit ist $P$ singulär zu $Q$, falls $\mu_1 \neq \nu_1$, siehe (2.1).

Aus der Statistik ist bekannt: Zu $\alpha \in (0,1)$ definiert

$$\varphi_n(x_1, \ldots, x_n) := 1_{(k_n, \infty)} \Big( \prod_{k=1}^{n} \frac{g(x_k)}{f(x_k)} \Big)$$

einen „besten“ Test zum Niveau $\alpha$ für $\mu_1$ gegen $\nu_1$ bei Beobachtung von $X_1, \dots, X_n$, wobei $k_n \in [0, \infty)$ so bestimmt wird, dass

$$\mu_1^{\otimes n}\Big(\Big\{(x_1, \dots, x_n) \in \mathbb{R}^{\mathbb{N}} : \prod_{k=1}^{n} \frac{g(x_k)}{f(x_k)} > k_n\Big\}\Big) = P(M_n > k_n) = \alpha$$

(Fehler 1. Art).

Da mit Korollar 2.4 $P(M_n \to 0) = 1$, folgt $k_n \to 0$. Damit konvergiert der Fehler 2. Art $Q(M_n \le k_n)$ für $n \to \infty$ gegen 0, denn

$$Q(M_n \le k_n) = \int_{\{M_n \le k_n\}} M_n \, dP \le k_n \to 0 \ .$$

Dies nennt man *Konsistenz* der Testfolge $(\varphi_n)_n$. Analog folgt durch Vertauschen der Rollen von $\mu_1$ und $\nu_1$ die Konsistenz des analogen Testes zum Niveau $\alpha$ für $\nu_1$ gegen $\mu_1$.

Die Folge der Dichtequotienten heißt auch *Likelihood-Prozess*.

## 2.2 Der Satz von Radon-Nikodym

Wir zeigen hier eine einfache Version des Satzes von Radon-Nikodym, der in Bemerkung 13.10 in der Wahrscheinlichkeitstheorie bereits Erwähnung fand:

**2.5 Satz** *Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\mathcal{A}$ sei separabel, das heißt $\mathcal{A} = \sigma(\{A_n, n \ge 1\})$ mit $A_n \subset \Omega$, $n \in \mathbb{N}$. Weiter sei $Q$ ein endliches Maß auf $(\Omega, \mathcal{A})$ und jede $P$-Nullmenge sei eine $Q$-Nullmenge. Dann existiert eine Zufallsvariable $X$ auf $(\Omega, \mathcal{A})$ mit*

$$Q(A) = \int_A X(\omega) \, dP(\omega) \quad \textit{für alle } A \in \mathcal{A} \ .$$

**2.6 Bemerkung** Viele $\sigma$-Algebren sind separabel, zum Beispiel die Borel-$\sigma$-Algebra eines jeden separablen metrischen Raums. Im Buch von Williams kann man lesen, dass diese Bedingung weggelassen werden kann.

Sei vorbereitend $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}_3 \subset \cdots$ eine Folge von Zerlegungen von $\Omega$ in disjunkte messbare Mengen derart, dass $\mathcal{A}_n = \{A_{n1}, \dots, A_{nk_n}\}$ eine Zerlegung von $\Omega$ ist, bei der jedes Element als Vereinigung von Elementen der Zerlegung $\mathcal{A}_{n+1}$ geschrieben werden kann:

$$A_{ni} = \bigcup_{j \in I(n,i)} A_{n+1,j} \ .$$

Zu $P$, $Q$ und $\mathcal{A}_n$ definieren wir nun

$$M_n := \sum_{i=1}^{k_n} \frac{Q(A_{ni})}{P(A_{ni})} 1_{A_{ni}} \; .$$

Dann ist $(M_n)_n$ unter $P$ ein Martingal bzgl. $(\sigma(\mathcal{A}_n))_n$. $M_n$ ist $\sigma(\mathcal{A}_n)$-messbar. Jedes $G \in \sigma(\mathcal{A}_n)$ ist disjunkte Vereinigung von $A_{ni}$'s. Ohne Einschränkung ist $G = A_{ni}$ für ein $i$. Dann gilt

$$\int_G M_{n+1}\, dP = \sum_{j \in I(n,i)} \frac{Q(A_{n+1,j})}{P(A_{n+1,j})} P(A_{n+1,j}) = Q(A_{ni}) = \int_G M_n\, dP \; .$$

**2.7 Lemma** *Es seien $P$ und $Q$ zwei endliche Maße auf $(\Omega, \mathcal{A})$. Dann ist jede $P$-Nullmenge eine $Q$-Nullmenge genau dann, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $P(A) \leq \delta$ impliziert $Q(A) \leq \varepsilon$ für jedes $A \in \mathcal{A}$.*

*Beweis:* „$\Leftarrow$“Aus der Bedingung folgt $Q(A) \leq \varepsilon$ für jede $P$-Nullmenge $A \in \mathcal{A}$ und jedes $\varepsilon > 0$. Daher ist $Q(A) = 0$. „$\Rightarrow$“Angenommen, die Bedingung gilt nicht. Dann existiert ein $\varepsilon > 0$ und eine Folge $(A_n)_n$ in $\mathcal{A}$ mit $P(A_n) \leq 2^{-n}$ und $Q(A_n) > \varepsilon$, $n \in \mathbb{N}$. Wähle

$$A := \limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \; ,$$

so ist $A \in \mathcal{A}$ und

$$P(A) \leq P(\bigcup_{m=n}^{\infty} A_m) \leq \sum_{m=n}^{\infty} P(A_m) \leq \sum_{m=n}^{\infty} 2^{-m} = 2^{-n+1} \; ,$$

$n \in \mathbb{N}$, also $P(A) = 0$ und andererseits wegen der Endlichkeit von $Q$

$$Q(A) \geq \limsup_{n \to \infty} Q(A_n) \geq \varepsilon > 0 \; ,$$

im Widerspruch zur Annahme, dass jede $P$-Nullmenge eine $Q$-Nullmenge ist. $\square$

*Beweis von 2.5:* Wir wählen $\mathcal{A}_n := \sigma(A_1, \ldots, A_n)$, $n \in \mathbb{N}$. Diese $\sigma$-Algebren werden erzeugt von den sogenannten Atomen $A_{n1}, \ldots, A_{nk_n}$ (erzeugt bedeutet: alle $2^{k_n}$ möglichen Vereinigungen von den Atomen), wobei jedes Atom der Form $H_1 \cap \cdots \cap H_n$ mit $H_i = A_i$ oder $H_i = A_i^c$, $i = 1, \ldots, n$, ist. Nun wählen wir

$$M_n := \sum_{i=1}^{k_n} \frac{Q(A_{ni})}{P(A_{ni})} 1_{A_{ni}} \; .$$

Für $A \in \mathcal{A}_n$ wähle die Darstellung

$$A = \bigcup_{i \in I} A_{ni} \ , \ I \text{ endlich} \ , \ A_{ni} \text{ Atom in } \mathcal{A}_n.$$

Dann ist

$$Q(A) = \sum_{i \in I} Q(A_{ni}) = \sum_{i \in I} \frac{Q(A_{ni})}{P(A_{ni})} P(A_{ni}) = \sum_{i \in I} \int_{A_{ni}} M_n \, dP = \int_A M_n \, dP \ ,$$

wobei wir beachten, dass im Fall $P(A_{ni}) = 0$ auch $Q(A_{ni}) = 0$ folgt, $\frac{Q(A_{ni})}{P(A_{ni})}$ also beliebig gesetzt werden kann.

$$\text{Also ist } M_n = \frac{dQ|_{\mathcal{A}_n}}{dP|_{\mathcal{A}_n}} \ .$$

$(M_n)_n$ ist ein nicht-negatives Martingal, konvergiert also f.s. gegen ein $M_\infty$. Zu $\varepsilon > 0$ wähle $\delta > 0$ wie in Lemma 2.7 und sei $K \in (0, \infty)$ so, dass $K > \frac{Q(\Omega)}{\delta}$. Dann ist $P(M_n > K) \leq \frac{1}{K} \mathbb{E}(M_n) = \frac{1}{K} Q(\Omega) < \delta$ und somit

$$\int_{\{M_n > K\}} M_n \, dP = Q(M_n > K) \leq \varepsilon \ .$$

Also ist $(M_n)_n$ gleichgradig integrierbar und somit konvergiert $(M_n)_n$ in $\mathcal{L}^1$ gegen $M_\infty$. Also gilt für alle $A \in \mathcal{A}_n$

$$\int_A M_\infty \, dP = \lim_{n \to \infty} \int_A M_n \, dP = Q(A) \ ,$$

und da $\bigcup_n \mathcal{A}_n$ ein durchschnittstabiler Erzeuger von $\mathcal{A}$ ist, folgt die Aussage für alle $A \in \mathcal{A}$, also ist $M_\infty$ die gesuchte Dichte. □

## 2.3 Kolmogorovs Kriterium

Wir erinnern an Satz 6.10 aus der Wahrscheinlichkeitstheorie: Es sei $(X_n)_n$ eine unabhängige Folge von Zufallsgrößen und $(a_n)_n$ eine Zahlenfolge mit $0 < a_n \nearrow \infty$. Dann folgt aus $\sum_{n \geq 1} a_n^{-2} \operatorname{Var}(X_n) < \infty$:

$$\lim_{n \to \infty} \frac{1}{a_n} \sum_{k=1}^{n} (X_k - \mathbb{E} X_k) = 0 \quad P\text{-f.s.}$$

Wir diskutieren dieses Kriterium nun mittels der Martingaltheorie: Ist $(X_n)_n$ fortan eine Folge integrierbarer reeller Zufallsvariablen mit

$$\mathbb{E}(X_{n+1} | X_1, \dots, X_n) \leq 0 \ , \ n \in \mathbb{N} \ , \ P\text{-f.s.}, \tag{2.2}$$

so ist $S_n := X_1 + \cdots + X_n$ ein Supermartingal, denn zu (2.2) addiere $\mathbb{E}(X_i|X_1,\dots,X_n) = X_i$, $i = 1,\dots,n$, und beachte $\sigma(S_1,\dots,S_n) = \sigma(X_1,\dots,X_n)$. Sind die $X_i$ auch zentriert, so ist $(S_n)_n$ ein Martingal. Wir hatten dies im Spezialfall einer Folge unabhängiger Zufallsgrößen $(X_n)_n$ zu Beginn von Kapitel 2 gesehen.

Die Bedingung $\sup_{n\in\mathbb{N}} \mathbb{E}|S_n| < \infty$ ist nach Satz 1.20 hinreichend für fast sichere Konvergenz. Im Fall der Unabhängigkeit ist $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$ hinreichend, denn

$$\big(\mathbb{E}|S_n|\big)^2 \le \mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \le \sum_{i=1}^{\infty} \mathrm{Var}(X_i)\ .$$

Ist $(X_n)_n$ eine Folge integrierbarer, zentrierter, reeller Zufallsvariablen mit

$$\mathbb{E}(X_{n+1}|X_1,\dots,X_n) = 0 \tag{2.3}$$

und ist $(a_n)_n$ wie in KOLMOGOROVs Kriterium (also $a_n > 0$ und $a_n \nearrow \infty$), so ist

$$\mathbb{E}\left(\frac{1}{a_{n+1}} X_{n+1} \Big| \frac{1}{a_1} X_1, \dots, \frac{1}{a_n} X_n\right) = 0 \quad \text{f.s. für alle } n \in \mathbb{N}\ .$$

Folglich ist $\left(\sum_{i=1}^{n} \frac{1}{a_i} X_i\right)_n$ ein Martingal. Konvergiert dieses f.s. (und eine Bedingung ist $\sup_{n\in\mathbb{N}} \mathbb{E}\left(\left|\sum_{i=1}^{n} \frac{1}{a_i} X_i\right|\right) < \infty$ ), so liefert das KRONECKERsche Lemma

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{i=1}^{n} X_i = 0 \quad \text{f.s.}$$

(2.3) ist schwächer als die geforderte Unabhängigkeit im Satz 6.10.

## 2.4 Verrauschte Beobachtungen (Filtertheorie)

Es seien $X$ und $(\eta_n)_n$ unabhängige normalverteilte Zufallsgrößen mit $\mathcal{L}(X) = N(0,\sigma^2)$ und $\mathcal{L}(\eta_n) = N(0,1)$, $n \in \mathbb{N}$. Es sei $X$ die Zufallsgröße, an der wir interessiert sind. Angenommen, sie kann nicht direkt beobachtet werden, aber in *verrauschter* Form. Wir nehmen dazu an, dass man die Folge $(Y_n)_n$ mit

$$Y_n := X + c_n \eta_n \quad ,\ n \in \mathbb{N}\ ,$$

beobachten kann, wobei $(c_n)_n$ reelle Zahlen in $(0,\infty)$ sind. Ein *natürlicher Schätzer*, der auf den Beobachtungen $Y_1,\dots,Y_n$ basieren soll, ist vermutlich $M_n := \mathbb{E}(X|\mathcal{A}_n)$ mit $\mathcal{A}_n := \sigma(Y_i, i \le n)$. $(M_n)_n$ ist nun ein quadratisch integrierbares Martingal mit

$$\mathbb{E}(M_n^2) \le \mathbb{E}\big(\mathbb{E}(X^2|\mathcal{A}_n)\big) = \mathbb{E}(X^2) = \sigma^2$$

für alle $n \in \mathbb{N}$. Mit Satz 1.25 und Korollar 1.32 folgt $\lim_{n\to\infty} M_n = M_\infty := \mathbb{E}(X|\mathcal{A}_\infty)$ in $\mathcal{L}^2$ und $P$-f.s., $\mathcal{A}_\infty := \sigma(Y_n, n \in \mathbb{N})$.

Wann ist $M_\infty = X$ $P$-fast sicher?

Dazu werden wir $M_n$ und die Varianzen von $X - M_n$ berechnen. Wir bestimmen dazu $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ so, dass

$$X - \sum_{i=1}^n \lambda_i Y_i =: Z$$

und $\mathcal{A}_n$ unabhängig sind. Da alle Zufallsgrößen gemeinsam normalverteilt sind, ist die gewünschte Unabhängigkeit gleichbedeutend mit

$$\mathrm{Cov}(Z, Y_i) = 0 \quad \text{für jedes } i \in \{1, \ldots, n\} \ .$$

Nun ist

$$\mathrm{Cov}(Z, Y_i) = \mathrm{Cov}\Big(X - \sum_{j=1}^n \lambda_j (X + c_j \eta_j), X + c_i \eta_i\Big) \ .$$

Eine kleine Nebenrechnung entlang der Definition der Kovarianz ergibt

$$\mathrm{Cov}(Z, Y_i) = \sigma^2 \Big(1 - \sum_{j=1}^n \lambda_j\Big) - \lambda_i c_i^2 \ .$$

Wir wollen also das Gleichungssystem

$$\sigma^2 \Big(1 - \sum_{j=1}^n \lambda_j\Big) - \lambda_i c_i^2 = 0 \ , \quad i = 1, \ldots, n$$

lösen: Es ergibt sich $\lambda_i = \frac{\sigma^2 c_i^{-2}}{1+\sigma^2 \sum_{j=1}^n c_j^{-2}}$, $i \in \{1, \ldots, n\}$. Nun liefert die Unabhängigkeit von $Z$ und $\mathcal{A}_n$: $\mathbb{E}(Z|\mathcal{A}_n) = \mathbb{E}(Z) = 0$, also

$$M_n = \mathbb{E}(X|\mathcal{A}_n) = \sum_{i=1}^n \lambda_i Y_i \quad \text{sowie}$$

$$\mathbb{E}\big((X - M_n)^2\big) = \mathbb{E}Z^2 = \sigma^2 \Big(1 - \sum_{j=1}^n \lambda_j\Big)^2 + \sum_{i=1}^n \lambda_i^2 c_i^2 = \sigma^2 \Big(1 + \sigma^2 \sum_{i=1}^n c_i^{-2}\Big)^{-1} \ .$$

Dies konvergiert genau dann gegen Null für $n \to \infty$, wenn $\sum_{i=1}^n c_i^{-2} = \infty$ gilt. Somit erhalten wir

**2.8 Satz** $(M_n)_n$ *konvergiert gegen* $X$ $P$*-f.s. und in* $\mathcal{L}^2$ *genau dann, wenn*

$$\sum_{i=1}^\infty \frac{1}{c_i^2} = \infty \ .$$

Allgemeine Filterprobleme lesen sich etwa so:

Es seien $\alpha, \beta, \gamma, \rho, \nu$ reelle Konstanten und $X_0$, $(\varepsilon_n)_n$, $(\eta_n)_n$ stochastisch unabhängige Zufallsgrößen mit $X_0 \sim N(\mu_0, \sigma_0^2)$; $\varepsilon_n \sim \eta_n \sim N(0,1)$, $n \in \mathbb{N}$. Der Zustand $X_n$ eines linearen stochastischen Systems zum Zeitpunkt $n \geq 1$ sei durch

$$X_n - X_{n-1} = \alpha X_{n-1} + \beta\varepsilon_n + \gamma$$

beschrieben. Statt $X_n$ selbst kann aber nur eine gestörte (verrauschte) Version $Y_n$, beschrieben durch $Y_n - Y_{n-1} = \rho X_n + \nu\eta_n$, $Y_0 = 0$, beobachtet werden. Der sogenannte Kalman-Bucy-Filter ist ein Verfahren, mit dessen Hilfe der tatsächliche Systemzustand erstaunlich gut geschätzt werden kann. Dies findet Anwendungen bei der Analyse von Zeitreihen oder der Signalverarbeitung (Ortung beweglicher Objekte).

## 2.5 Verzweigungsprozesse

Seien $(X_{n,i})_{n\in\mathbb{N}_0, i\in\mathbb{N}}$ unabhängige und identisch verteilte Zufallsvariablen mit Werten in $\mathbb{N}_0$. Setze $p_k := P(X_{1,1} = k)$, $k \in \mathbb{N}_0$. Seien $m := \mathbb{E}(X_{1,1}) < \infty$ und $\sigma^2 := \mathrm{Var}(X_{1,1}) \in (0,\infty)$. Wir definieren den stochastischen Prozess $Z = (Z_n)_{n\in\mathbb{N}_0}$ durch

$$Z_0 = 1\ , \quad Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}\ .$$

Eine mögliche Interpretation ist: $Z_n$ ist die Größe einer Population zur Zeit $n$. Das $i$'te Individuum hat $X_{n,i}$ Nachkommen in der $n+1$'ten Generation. $Z$ heißt Galton-Watson-*Prozess* oder *Verzweigungsprozess* mit Nachkommenverteilung $p = (p_k)_{k\in\mathbb{N}_0}$. Manchmal lernt man im dritten Semester: Ist $q := \lim_{n\to\infty} P(Z_n = 0)$, die *Aussterbewahrscheinlichkeit* , so gilt mit $p_1 \neq 1$:

$$q < 1 \quad \Leftrightarrow \quad \sum_{k=1}^{\infty} k\, p_k = \mathbb{E}(X_{1,1}) > 1$$

(Methode: erzeugende Funktion).

Wir betrachten dieses Modell jetzt im Rahmen der Martingaltheorie: Sei $\mathcal{A}_n := \sigma(X_{k,i}, k < n, i \in \mathbb{N})$. Dann ist $Z = (Z_n)_{n\in\mathbb{N}_0}$ an $(\mathcal{A}_n)_{n\in\mathbb{N}_0}$ adaptiert. Setze

$$W_n := \frac{Z_n}{m^n}\ .$$

**2.9 Lemma** *$W = (W_n)_{n\in\mathbb{N}_0}$ ist ein Martingal.*

*Beweis:*

$$\begin{aligned}
\mathbb{E}(W_{n+1}|\mathcal{A}_n) &= m^{-(n+1)}\mathbb{E}(Z_{n+1}|\mathcal{A}_n) = m^{-(n+1)}\mathbb{E}\Big(\sum_{i=1}^{Z_n} X_{n,i}|\mathcal{A}_n\Big) \\
&= m^{-(n+1)}\sum_{k=1}^{\infty}\mathbb{E}\big(1_{\{Z_n=k\}}\,k\,X_{n,1}|\mathcal{A}_n\big) \\
&= m^{-n}\sum_{k=1}^{\infty}\mathbb{E}\big(k\cdot 1_{\{Z_n=k\}}|\mathcal{A}_n\big) \\
&= m^{-n}Z_n = W_n \quad \text{f.s.}
\end{aligned}$$

□

**2.10 Satz** *Es existiert der fast sichere Limes $W_\infty = \lim_{n\to\infty} W_n$. Es sind äquivalent:*

(a) $m > 1$ ,

(b) $\mathbb{E}(W_\infty) = 1$ ,

(c) $\mathbb{E}(W_\infty) > 0$ .

*Beweis:* Die Limes-Zufallsgröße $W_\infty$ existiert, da $(W_n)_{n\in\mathbb{N}_0}$ ein nicht-negatives Martingal ist, siehe Korollar 1.21. Ist $m \le 1$, so folgt, dass $(Z_n)_n$ f.s. gegen ein $Z_\infty$ konvergiert. Wegen $\sigma^2 > 0$ kommt nur $Z_\infty = 0$ in Frage.

Es sei nun $m > 1$. Dann wollen wir die Varianz von $W_n$ bestimmen, also die von $Z_n$. Da gibt es eine Formel, die man auch den Satz von Blackwell-Girshick nennt, den wir im Anschluss formulieren und beweisen:

$$\begin{aligned}
\mathrm{Var}(W_n) &= m^{-2n}\big(\sigma^2\mathbb{E}(Z_{n-1}) + m^2\,\mathrm{Var}(Z_{n-1})\big) \\
&= \sigma^2 m^{-(n+1)} + \mathrm{Var}(W_{n-1})\ .
\end{aligned}$$

Induktiv folgt

$$\mathrm{Var}(W_n) = \sigma^2\sum_{k=2}^{n+1} m^{-k} \le \frac{\sigma^2 m}{m-1} < \infty\ .$$

Also ist $(W_n)_n$ in $\mathcal{L}^2$ beschränkt, also folgt mit Satz 1.25 $W_n \to W_\infty$ in $\mathcal{L}^2$, somit auch in $\mathcal{L}^1$ und speziell $\mathbb{E}(W_\infty) = \mathbb{E}(W_0) = 1$. □

**2.11 Satz (Waldsche Gleichheit, Satz von Blackwell-Girshick)** *Es sei $(X_n)_n$ eine an eine Filtrierung $(\mathcal{A}_n)_n$ in $\mathcal{A}$ adaptierte Folge identisch verteilter, integrierbarer, reeller Zufallsvariablen über $(\Omega, \mathcal{A}, P)$, wobei $X_{n+1}$ von $\mathcal{A}_n$ unabhängig sei, $n \in \mathbb{N}$. $(S_n)_n$ sei die Folge der Partialsummen $S_n := X_1 + \cdots + X_n$. $T$ sei eine integrierbare Stoppzeit bzgl. $(\mathcal{A}_n)_n$ mit Werten in $\mathbb{N}_0$. Dann gilt:*

(a) $S_T$ *ist integrierbar und*

$$\mathbb{E}(S_T) = \mathbb{E}(T)\,\mathbb{E}(X_1)\ . \quad (\text{WALD}\textit{sche Identität})$$

(b) *Sind die* $X_n$ *sogar quadratisch integrierbar und zentriert, so trifft dies auch auf* $S_T$ *zu und es gilt*

$$\mathbb{E}(S_T^2) = \mathbb{E}(T)\,\mathbb{E}(X_1^2)\ . \quad (\text{BLACKWELL-GIRSHICK})$$

Wäre die Stoppzeit $T$ beschränkt, so würde man das Martingal $S_n^* := S_n - n\mathbb{E}(X_1)$ betrachten und erhielt mit Satz 1.16 $\mathbb{E}(S_T^*) = \mathbb{E}(S_1^*) = 0$, also (a). Die Bedingung $\mathbb{E}(T) < \infty$ erzwingt ein anderes Vorgehen:

*Beweis:*

(a) Mit der Notation, die wir nach Definition 1.27 eingeführt hatten, gilt

$$\begin{aligned}\mathbb{E}(|S_T|) &= \sum_{n=1}^{\infty}\mathbb{E}(|S_n|;T=n) \le \sum_{n=1}^{\infty}\sum_{i=1}^{n}\mathbb{E}(|X_i|;T=n)\\ &= \sum_{i=1}^{\infty}\sum_{n=i}^{\infty}\mathbb{E}(|X_i|;T=n) = \sum_{i=1}^{\infty}\mathbb{E}(|X_i|;T\ge i)\ . \end{aligned}\tag{2.4}$$

$X_i$ ist von $\mathcal{A}_{i-1}$ unabhängig, also von $1_{\{T\ge i\}}$ $\big($beachte $\{T<i\} = \{T \le i-1\} \in \mathcal{A}_{i-1}\big)$, $i \ge 2$. Es gilt $\{T\ge 1\} = \Omega$. Somit

$$\mathbb{E}(|X_i|;T\ge i) = P(T\ge i)\,\mathbb{E}(|X_i|) = P(T\ge i)\,\mathbb{E}(|X_1|)\ ,$$

also

$$\mathbb{E}(|S_T|) \le \sum_{i=1}^{\infty} P(T\ge i)\,\mathbb{E}(|X_1|) = \mathbb{E}(T)\,\mathbb{E}(|X_1|) < \infty\ ,$$

da $P(T<\infty) = 1$.

Also ist $S_T$ integrierbar. Es folgt die WALDsche Identität, wenn man $S_T$ anstelle von $|S_T|$ betrachtet. Wegen der bewiesenen Konvergenz der Reihe $\sum_{n=1}^{\infty}\sum_{i=1}^{n}\big|\mathbb{E}(X_i;T=n)\big|$ ist die Vertauschung der Summen in (2.4) gerechtfertigt.

(b) Sei $Y_n := X_n 1_{\{n\le T\}}$, also $Y_n \in \mathcal{L}^2$. Nun gilt $\mathbb{E}(Y_mY_n) = 0$ für $m \ne n$: Für $m < n$ ist $Y_mY_n = X_nX_m1_{\{n\le T\}}$ wegen $\{n\le T\} \subset \{m \le T\}$. $X_n$ und $1_{\{n\le T\}}X_m$ sind wieder unabhängig, also

$$\mathbb{E}(Y_mY_n) = \mathbb{E}(X_n)\,\mathbb{E}(X_m1_{\{n\le T\}}) = 0\ .$$

Auch $X_n^2$ und $1_{\{n\le T\}}$ sind unabhängig, also

$$\mathbb{E}(Y_n^2) = \mathbb{E}(X_n^2)\,P(n\le T) = \mathbb{E}(X_1^2)\,P(n\le T)\ , \quad n\in\mathbb{N}\ .$$

Somit

$$\sum_{n\geq 1} \mathbb{E}(Y_n^2) = \mathbb{E}(X_1^2)\,\mathbb{E}(T) < \infty \ . \tag{2.5}$$

Weiter ist für $m < n$

$$\mathbb{E}\Big((Y_m + \cdots + Y_n)^2\Big) = \mathbb{E}\Big(\sum_{i=m}^{n} Y_i^2 + \sum_{i\neq j} Y_i Y_j\Big) = \sum_{i=m}^{n} \mathbb{E}(Y_i^2) \ ,$$

also

$$||Y_m + \cdots + Y_n||_2 = \Big(\sum_{i=m}^{n} \mathbb{E}(Y_i^2)\Big)^{1/2},$$

und somit folgt mit (2.5) die $\mathcal{L}^2$-Konvergenz der Reihe $\sum_{i=1}^{\infty} Y_i$. Es gilt $S_T = \sum_{i=1}^{\infty} X_i 1_{\{i\leq T\}} = \sum_{i=1}^{\infty} Y_i$ fast sicher, denn dies gilt für $\{T < \infty\}$. Also ist $S_T \in \mathcal{L}^2$ und

$$\lim_{n\to\infty} \sum_{i=1}^{n} Y_i = S_T \quad \text{in } \mathcal{L}^2 \ .$$

Schließlich gilt nach der majorisierten Konvergenz

$$\mathbb{E}(S_T^2) = \lim_{n\to\infty} \mathbb{E}\Big(\big(\sum_{i=1}^{n} Y_i\big)^2\Big) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}(Y_i^2) = \sum_{i=1}^{\infty} \mathbb{E}(Y_i^2) \ ,$$

also folgt die Identität in (b) aus (2.5) und die WALDsche Identität liefert die Zentriertheit von $S_T$. □

Ist $X_1$ in Teil (b) von Satz 2.11 nicht zentriert und bezeichnet $\mu$ den endlichen Erwartungswert und $\sigma^2$ die endliche Varianz, kann man mittels eines analogen Beweises zeigen:

$$\mathrm{Var}(S_T) = \sigma^2\,\mathbb{E}(T) + \mu^2\,\mathrm{Var}(T).$$

Tatsächlich haben wir diese Formel zur Bestimmung von $\mathrm{Var}(W_n)$ verwendet.

KAPITEL 3

# Stationäre Prozesse

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

**3.1 Definition** Sei $(E, \mathcal{E})$ ein messbarer Raum. Eine Folge $\mathbb{X} = (X_n)_n$ von $(E, \mathcal{E})$-wertigen Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ heißt $(E, \mathcal{E})$-*wertiger stochastischer Prozess*. Im Fall $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ sagen wir einfach *stochastischer Prozess*.

Die Verteilung eines stochastischen Prozesses $\mathbb{X}$ ist einfach seine Verteilung als $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$-wertige Zufallsvariable, d. h. das W-Maß $P\mathbb{X}^{-1}$ auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$. Ist $\mathbb{X}$ ein stochastischer Prozess, so ist die Folge $(\pi_n)_n$ der Projektionen $\pi_n : E^{\mathbb{N}} \to E$ ein auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}, P\mathbb{X}^{-1})$ definierter stochastischer Prozess, der dieselbe Verteilung wie $\mathbb{X}$ hat.

**3.2 Definition** Ist $Q$ ein W-Maß auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ und ist $\pi^{(n)} : E^{\mathbb{N}} \to E^n$, $n \in \mathbb{N}$, durch

$$\pi^{(n)} = (\pi_1, \dots, \pi_n)$$

definiert, so ist

$$Q^{(n)} := Q\big(\pi^{(n)}\big)^{-1}$$

ein W-Maß auf $(E^n, \mathcal{E}^n)$. Die Maße $Q^{(n)}$ heißen *endlichdimensionale Verteilungen* von $Q$.

**3.3 Satz** *Die Folge der endlichdimensionalen Verteilungen $(Q^{(n)})_n$ bestimmt das W-Maß $Q$ eindeutig.*

*Beweis:* Wir geben ein durchschnittstabiles Erzeugendensystem $\mathcal{D}$ von $\mathcal{E}^{\mathbb{N}}$ an:

$$\mathcal{D} := \{A_J,\ J = \{1, \dots, n\}, n \in \mathbb{N};\ A_j \in \mathcal{E},\ j \in J\}$$

mit

$$A_J := \left\{(x_i)_{i \in \mathbb{N}} \in E^{\mathbb{N}} : x_j \in A_j \text{ für } j \in J\right\}.$$

Dann gilt $\bigotimes_{i \in \mathbb{N}} \mathcal{E} = \mathcal{E}^{\mathbb{N}} = \sigma(\mathcal{D})$. $A_J$ nennt man auch „endlichdimensionale Zylinder". Nach Satz 2.9, Kapitel 2, Wahrscheinlichkeitstheorie, stimmen zwei W-Maße auf $\mathcal{E}^{\mathbb{N}}$ überein, wenn sie auf $\mathcal{D}$ übereinstimmen. $Q$ ist also auf $\mathcal{D}$ festgelegt. Daraus folgt die Behauptung. $\square$.

Im Fall $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ folgt

**3.4 Proposition** Sind $\mathbb{X} = (X_n)_n$ und $\mathbb{X}' = (X'_n)_n$ zwei $(\mathbb{R}, \mathcal{B})$-wertige stochastische Prozesse auf $(\Omega, \mathcal{A}, P)$ bzw. $(\Omega', \mathcal{A}', P')$, so ist $\mathcal{L}(\mathbb{X}) = \mathcal{L}(\mathbb{X}')$ genau dann, wenn

$$P(X_1 \leq t_1, \ldots, X_n \leq t_n) = P'(X'_1 \leq t_1, \ldots, X'_n \leq t_n)$$

für alle $n \in \mathbb{N}$ und alle $t_1, \ldots, t_n \in \mathbb{R}$ gilt.

*Beweis:* $\{x \in \mathbb{R}^{\mathbb{N}} : x_1 \leq t_1, \ldots, x_n \leq t_n\}$ bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}^{\mathbb{N}}$. Nach Voraussetzung stimmen $P\mathbb{X}^{-1}$ und $P'\mathbb{X}'^{-1}$ auf diesen Mengen überein. Also folgt $P\mathbb{X}^{-1} = P'\mathbb{X}'^{-1}$. □

**3.5 Definition** Eine $\mathcal{A}$-$\mathcal{A}$-messbare Abbildung $T : \Omega \to \Omega$ heißt *maßerhaltende Transformation*, wenn

$$PT^{-1} = P$$

gilt. Man sagt auch, dass $P$ *$T$-invariant* (oder *invariant unter* $T$) ist. Ein stochastischer Prozess $\mathbb{X}$ heißt *stationär*, wenn $\mathbb{X}$ dieselbe Verteilung wie der Prozess $(X_2, X_3, \ldots)$ hat.

**3.6 Lemma** *Es seien $Y$ eine Zufallsgröße und $T$ eine maßerhaltende Transformation auf $(\Omega, \mathcal{A}, P)$. $X_1 := Y$; $X_{n+1} := X_n \circ T$, $n \in \mathbb{N}$. Dann ist $(X_n)_n$ ein stationärer Prozess.*

*Beweis:* Es seien $n \in \mathbb{N}$ und $t_1, \ldots, t_n \in \mathbb{R}$. Dann gilt

$$\begin{aligned} P(X_1 \leq t_1, \ldots, X_n \leq t_n) &= PT^{-1}(X_1 \leq t_1, \ldots, X_n \leq t_n) \\ &= P(X_1 \circ T \leq t_1, \ldots, X_n \circ T \leq t_n) \\ &= P(X_2 \leq t_1, \ldots, X_{n+1} \leq t_n)\,. \end{aligned}$$

Also folgt die Behauptung mit Proposition 3.4. □

**3.7 Bemerkung** Zu jedem auf $(\Omega, \mathcal{A}, P)$ definierten, stationären Prozess $\mathbb{X}$ existiert eine maßerhaltende Transformation, die einen Prozess liefert, der in Verteilung mit $\mathbb{X}$ übereinstimmt. $\mathbb{X}$ ist eine $\mathcal{A}$-$\mathcal{B}^{\mathbb{N}}$-messbare Abbildung, sei $P' := P\mathbb{X}^{-1}$. Sei $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ definiert durch $T\big((x_1, x_2, \ldots)\big) = (x_2, x_3, \ldots)$. Dann ist $T$ auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$ maßerhaltend, wenn $\mathbb{X}$ stationär ist:

Für $A \in \mathcal{B}^{\mathbb{N}}$ gilt

$$P'(A) = P(\mathbb{X} \in A) = P(T \circ \mathbb{X} \in A) = P\mathbb{X}^{-1}T^{-1}(A) = P'T^{-1}(A)\,.$$

$P'$ ist auch die Verteilung der Folge der Projektionen $\pi_1, \pi_2, \dots : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, denn $\pi = (\pi_n)_n$ ist die identische Abbildung auf $\mathbb{R}^{\mathbb{N}}$. Weiter gilt $\pi_n = \pi_{n-1} \circ T$, $n \geq 2$. Der Prozess $(\pi_n)_n$ auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ ist somit von der in Lemma 3.6 definierten Form und besitzt dieselbe Verteilung wie $\mathbb{X}$.

**3.8 Beispiele**

(a) Eine Folge von unabhängigen, identisch verteilten Zufallsgrößen ist ein stationärer Prozess. Ist $\mu$ ein W-Maß auf $(\mathbb{R}, \mathcal{B})$, so ist $\mu^{\otimes \mathbb{N}}$ invariant unter der Verschiebungsabbildung $T$, wie in Bemerkung 3.7 definiert.

(b) $\Omega = D := \{z \in \mathbb{C} : |z| = 1\}$. Wir übertragen vom Intervall $[0, 2\pi)$ die Borel-$\sigma$-Algebra und die gleichförmige Verteilung (= Lebesgue-Maß/$2\pi$) auf $D$ und nennen dies $(D, \mathcal{B}_D, \lambda_D)$. Für $c \in D$ sei $T_c : D \to D$ durch $T_c(\omega) := c\omega$ definiert. Ist $A \subset D$ ein Intervall, so gilt

$$\lambda_D\big(T_c^{-1}(A)\big) = \lambda_D(A)\,.$$

Intervalle bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_D$, woraus folgt, dass $T_c$ maßerhaltend ist.

(c) Auf $([0,1), \mathcal{B}_{[0,1)}, \lambda_{[0,1)})$ betrachten wir $T : [0,1) \to [0,1)$, definiert durch

$$T(\omega) := \begin{cases} 2\omega & \text{für } \omega \in [0, \frac{1}{2})\,, \\ 2\omega - 1 & \text{für } \omega \in [\frac{1}{2}, 1)\,. \end{cases}$$

Für $0 \leq a < b \leq 1$ ist

$$T^{-1}\big([a,b)\big) = \Big[\frac{a}{2}, \frac{b}{2}\Big) \cup \Big[\frac{a+1}{2}, \frac{b+1}{2}\Big)\,,$$

also $PT^{-1}\big([a,b)\big) = b - a = P\big([a,b)\big)$. Die Mengen $[a,b)$ mit $0 \leq a < b \leq 1$ zusammen mit $\emptyset$ bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_{[0,1)}$, also folgt $PT^{-1} = P$.

**3.9 Definition** Für eine Abbildung $T : \Omega \to \Omega$ heißt eine Teilmenge $A$ von $\Omega$ $T$-*invariant*, wenn $T^{-1}(A) = A$ gilt. Eine maßerhaltende Abbildung $T$ auf $(\Omega, \mathcal{A}, P)$ heißt *ergodisch*, wenn für jede $T$-invariante Menge $A \in \mathcal{A}$ gilt:

$$P(A) \in \{0, 1\}\,.$$

**3.10 Bemerkung** Die Familie der $T$-invarianten Mengen aus $\mathcal{A}$ bilden eine Teil-$\sigma$-Algebra von $\mathcal{A}$, die hier mit $\mathcal{J}$ bezeichnet wird. Ist $T$ ergodisch, so sagt man, dass $\mathcal{J}$ $P$-*trivial* ist. Jede $\mathcal{J}$-messbare Zufallsgröße ist fast sicher konstant (im Fall $T$ ergodisch). Dies folgt aus Korollar 5.20, Wahrscheinlichkeitstheorie.

**3.11 Satz** *Es sei $\mathbb{X} = (X_n)_n$ eine Folge von unabhängigen, identisch verteilten Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ mit $P' := P\mathbb{X}^{-1}$. Dann ist die Verschiebungsabbildung $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, $\{x_n\}_{n \in \mathbb{N}} \mapsto \{x_{n+1}\}_{n \in \mathbb{N}}$ ergodisch auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$.*

*Beweis:* Sei $A \in \mathcal{B}^{\mathbb{N}}$ eine $T$-invariante Menge, so gilt $A = T^{-n}(A) := (T^n)^{-1}(A) = \{(x_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} : (x_{n+k})_{k\in\mathbb{N}} \in A\}$. Also gilt

$$\begin{aligned}\mathbb{X}^{-1}(A) &= \{\omega : \big(X_{n+1}(\omega), X_{n+2}(\omega), \dots\big) \in A\} \in \tilde{\mathcal{A}}_{n+1} \\ &:= \sigma\Big(\bigcup_{k=n+1}^{\infty} \big(X_k^{-1}(A)\big)\Big).\end{aligned}$$

Dies gilt für alle $n \in \mathbb{N}$, also

$$\mathbb{X}^{-1}(A) \in \mathcal{A}_\infty := \bigcap_{n=1}^{\infty} \tilde{\mathcal{A}}_n .$$

Nach Kolmogorovs 0-1-Gesetz folgt

$$P'(A) = P\big(\mathbb{X}^{-1}(A)\big) \in \{0,1\} . \qquad \square$$

**3.12 Bemerkung** Die $\sigma$-Algebra $\mathcal{J}$ der verschiebungsinvarianten messbaren Mengen in $\mathbb{R}^{\mathbb{N}}$ ist eine Teil-$\sigma$-Algebra der terminalen $\sigma$-Algebra

$$\mathcal{A}_\infty = \bigcap_{n=1}^{\infty} \sigma(\pi_k : k \geq n) .$$

Es gilt jedoch keinesfalls $\mathcal{J} = \mathcal{A}_\infty$ (ohne Beweis).

**3.13 Satz** *Es seien $D = \{z \in \mathbb{C} : |z| = 1\}$, $\mathcal{B}_D$ die Borel-$\sigma$-Algebra auf $D$ und $\lambda_D$ das normierte Lebesgue-Maß auf $(D, \mathcal{B}_D)$, $T_c : D \to D$ definiert durch $T_c(\omega) := c\,\omega$ für $c, \omega \in D$. Dann ist $T_c$ genau dann ergodisch auf $(D, \mathcal{B}_D, \lambda_D)$, wenn $c$ keine Einheitswurzel ist.*

*Beweis:* Setze $T := T_c$. Ist $c$ eine Einheitswurzel, existiert also ein $n \in \mathbb{N}$ mit $c^n = 1$, so ist $T^n$ die Identität. Dann ist für jedes $A \in \mathcal{B}_D$ die Menge $A \cup T^{-1}A \cup \dots \cup T^{-n+1}A$ invariant unter $T$.

Wähle ein $A \in \mathcal{B}_D$ mit $0 < \lambda_D(A) < 1/n$, so gilt

$$0 < \lambda_D(A \cup \dots \cup T^{-n+1}A) \leq \sum_{k=0}^{n-1} \lambda_D(T^{-k}A) = n\lambda_D(A) < 1 .$$

Somit ist $T$ nicht ergodisch. Für die Umkehrung treffen wir zunächst Vorbereitungen:

**3.14 Lemma** *Sei $c \in D$ keine Einheitswurzel. Dann liegt $\{c^n, n \in \mathbb{N}_0\}$ dicht in $D$.*

*Beweis:* Da die $c^n$, $n \in \mathbb{N}$, alle verschieden sind, besitzt diese Folge mindestens einen Häufungspunkt $\omega_0 \in D$. Seien $\varepsilon > 0$ und $m > n$ in $\mathbb{N}$ mit $|c^n - \omega_0| < \varepsilon$ und $|c^m - \omega_0| < \varepsilon$. Dann gilt $|c^{m-n} - 1| \in (0, 2\varepsilon)$. Daraus folgt, dass für jedes $\omega \in D$ ein $k \in \mathbb{N}$ existiert mit $|\omega - c^{k(m-n)}| < 2\varepsilon$. Da $\varepsilon > 0$ beliebig war, folgt die Behauptung. $\square$

**3.15 Lemma** *Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\mathcal{F}$ eine Algebra mit $\mathcal{A} = \sigma(\mathcal{F})$. Zu $\varepsilon > 0$ und $A \in \mathcal{A}$ existiert dann ein $B \in \mathcal{F}$ mit $P(A \triangle B) < \varepsilon$.*

*Beweis:* maßtheoretische Übung. $\square$

Wir führen den *Beweis von Satz 3.13* fort: Sei $c$ keine Einheitswurzel. $\mathcal{F}$ sei die Familie der endlichen Vereinigungen paarweise disjunkter Intervalle in $D$; dann ist $\mathcal{F}$ eine Algebra. Sei nun $A$ invariant unter $T_c$ mit $\lambda_D(A) > 0$ und $\varepsilon \in (0, 1/2)$, dann existiert nach 3.15 eine endliche Vereinigung $B = \bigcup_{i=1}^n I_i$ disjunkter Intervalle $I_1, \dots, I_n \subset D$ mit $\lambda_D(A \triangle B) \leq \varepsilon \lambda_D(A)$.

Ohne Einschränkung können wir $\lambda_D(I_i) \leq \varepsilon$ für $i = 1, \dots, n$ annehmen. Es folgt nun

$$\begin{aligned}\lambda_D(A \triangle B) &\leq \varepsilon \lambda_D(A) \leq 2\varepsilon(1-\varepsilon)\lambda_D(A)\\ &\leq 2\varepsilon\big(\lambda_D(A) - \lambda_D(A \triangle B)\big) \leq 2\varepsilon \lambda_D(B)\,,\end{aligned}$$

wobei wir $|\lambda_D(A) - \lambda_D(B)| \leq \lambda_D(A \triangle B)$ verwendet haben. Dies gilt für jedes Maß und folgt aus $A \subset B \cup (A \setminus B)$, womit $\lambda_D(A) - \lambda_D(B) \leq \lambda(A \setminus B) \leq \lambda_D(A \triangle B)$, und dies ist symmetrisch in $A$ und $B$. Es folgt nun

$$\begin{aligned}\sum_{i=1}^n \lambda_D(A \cap I_i) = \lambda_D(A \cap B) &\geq \lambda_D(B) - \lambda_D(A \triangle B)\\ &\geq (1-2\varepsilon)\lambda_D(B)\\ &= (1-2\varepsilon)\sum_{i=1}^n \lambda_D(I_i).\end{aligned}$$

Mindestens eines der $I_i$ erfüllt also die Ungleichung

$$\lambda_D(A \cap I_i) \geq (1-2\varepsilon)\lambda_D(I_i)\,.$$

Mit Lemma 3.14 kann gefolgert werden, dass ein $k \in \mathbb{N}$ und $n_1, \dots, n_k \in \mathbb{N}$ existieren, so dass die Intervalle

$$T^{-n_1} I_i\,,\, T^{-n_2} I_i\,,\, \dots\,,\, T^{-n_k} I_i$$

paarweise disjunkt sind und $D$ bis auf eine Menge von kleinerem Maß als $2\varepsilon$ ausfüllen. Wegen der $T$-Invarianz von $A$ und $\lambda_D$ gilt für $j = 1, \dots, k$:

$$\begin{aligned}\lambda_D(A \cap T^{-n_j} I_i) &= \lambda_D(T^{-n_j} A \cap T^{-n_j} I_i) = \lambda_D T^{-n_j}(A \cap I_i)\\ &= \lambda_D(A \cap I_i) \geq (1 - 2\varepsilon)\lambda_D(I_i)\\ &= (1 - 2\varepsilon)\lambda_D(T^{-n_j} I_i)\,,\end{aligned}$$

also

$$\begin{aligned}\lambda_D(A) &\geq \sum_{j=1}^{k} \lambda_D(A \cap T^{-n_j} I_i) \geq (1 - 2\varepsilon) \sum_{j=1}^{k} \lambda_D(T^{-n_j} I_i)\\ &\geq (1 - 2\varepsilon)^2\,.\end{aligned}$$

Daraus folgt $\lambda_D(A) = 1$. □

**3.16 Definition** Eine maßerhaltende Transformation $T$ auf $(\Omega, \mathcal{A}, P)$ heißt *mischend*, wenn für alle $A, B \in \mathcal{A}$ gilt

$$\lim_{n \to \infty} P(A \cap T^{-n} B) = P(A)P(B)\,. \tag{3.1}$$

**3.17 Lemma** *Jede mischende Transformation ist ergodisch.*

*Beweis:* Ist $A \in \mathcal{A}$ eine $T$-invariante Menge, so gilt

$$P(A) = P(A \cap A) = P(A \cap T^{-n} A) \to P(A)P(A)$$

für $n \to \infty$, also $P(A) = P(A)^2$, also $P(A) \in \{0, 1\}$. □

**3.18 Lemma** *Es sei $\mathcal{F}$ eine Algebra, die $\mathcal{A}$ erzeugt. Falls die Gleichung* (3.1) *für alle $A, B \in \mathcal{F}$ gilt, so ist $T$ mischend.*

*Beweis:* $A, B \in \mathcal{A}$, $\varepsilon > 0$. Nach Lemma 3.15 existieren $A_0, B_0 \in \mathcal{F}$ mit $P(A \triangle A_0) < \varepsilon$ und $P(B \triangle B_0) < \varepsilon$. Daraus folgt

$$\begin{aligned}\big|P(A \cap T^{-n} B) - P(A_0 \cap T^{-n} B_0)\big| &\leq P(A \triangle A_0) + P\big(T^{-n} B \triangle T^{-n} B_0\big)\\ &\leq P(A \triangle A_0) + P(B \triangle B_0) < 2\varepsilon\,.\end{aligned}$$

Wieder haben wir $|P(A) - P(B)| \leq P(A \triangle B)$ verwendet. Die Folge $\big(P(A_0 \cap T^{-n} B_0)\big)_n$ konvergiert gegen $P(A_0)P(B_0)$ und es gilt $|P(A_0) - P(A)| \leq P(A \triangle A_0) < \varepsilon$ und analog $|P(B_0) - P(B)| < \varepsilon$. Also folgt

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

**3.19 Satz** *Die in Beispiel 3.8 (c) definierte Transformation $T$ auf $([0,1), \mathcal{B}_{[0,1)}, \lambda_{[0,1)})$ ist mischend, also ergodisch.*

*Beweis:* Für jede Menge $A \subset [0,1]$ sind $T^{-1}A \cap [0,1/2)$ und $T^{-1}A \cap [1/2,1)$ nur um 1/2 gegeneinander verschobene Mengen. Also folgt mit $P := \lambda_{[0,1)}$

$$\begin{aligned} P\Big(T^{-1}A \cap \big[\frac{1}{2},1)\Big) &= P\Big(T^{-1}A \cap \big[0,\frac{1}{2})\Big) = \frac{1}{2}P(T^{-1}A) \\ &= P(A)P\Big(\big[0,\frac{1}{2})\Big). \end{aligned}$$

Analog folgt für $n \in \mathbb{N}$ und $k = 0, 1, \ldots, 2^n - 1$:

$$P\big(T^{-n}A \cap [k2^{-n}, (k+1)2^{-n})\big) = P(A)P\big([k2^{-n}, (k+1)2^{-n})\big).$$

$\mathcal{A}_0$ sei die Familie der Intervalle der Form $\big[k2^{-n}, (k+1)2^{-n}\big)$ mit $n \in \mathbb{N}$ und $0 \le k \le 2^n - 1$. Dann gilt also

$$\lim_{m\to\infty} P(T^{-m}A \cap I) = P(A)P(I)$$

für alle $I \in \mathcal{A}_0$. Diese Gleichung bleibt richtig, wenn $I$ eine endliche Vereinigung paarweise disjunkter Intervalle dieser Form ist. Diese Figuren bilden eine Algebra, die die BOREL-$\sigma$-Algebra auf $[0,1)$ erzeugt. Somit folgt der Satz mit Lemma 3.18. □

**3.20 Satz (Ergodensatz von Birkhoff, 1931)** *Es sei $T$ eine maßerhaltende Transformation auf $(\Omega, \mathcal{A}, P)$ und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Dann konvergiert*

$$S_n := \frac{1}{n}\sum_{j=0}^{n-1} X \circ T^j$$

*fast sicher für $n \to \infty$ gegen eine $\mathcal{J}$-messbare Zufallsgröße $Y$, für die*

$$\int X\, dP = \int Y\, dP$$

*gilt.*

**3.21 Bemerkungen**

(a) Ist $T$ ergodisch, so ist jede $\mathcal{J}$-messbare Zufallsgröße fast sicher konstant. Wegen $\int X\,dP = \int Y\,dP$ muss $Y = \int X\,dP = \mathbb{E}X$ gelten. Insbesondere folgt aus dem Ergodensatz zusammen mit Satz 3.11 das starke Gesetz der großen Zahlen, siehe Korollar 6.5, Wahrscheinlichkeitstheorie.

(b) Mit $A \in \mathcal{J}$ ist $1_A X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Mit Satz 3.20 existiert eine $\mathcal{J}$-messbare Zufallsgröße $Y_A$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (1_A X) \circ T^j = Y_A$$

und

$$\int Y_A \, dP = \int 1_A X \, dP = \int_A X \, dP \, .$$

Ist $A$ invariant, so folgt

$$\frac{1}{n} \sum_{j=0}^{n-1} (1_A X) \circ T^j = \frac{1}{n} 1_A \sum_{j=1}^{n-1} X \circ T^j \, ,$$

also $Y_A = 1_A Y$ P-f. s. Somit haben wir im nicht-ergodischen Fall gezeigt, dass

$$\int_A X \, dP = \int_A Y \, dP \quad \forall A \in \mathcal{J} \, ,$$

also $Y = \mathbb{E}(X \mid \mathcal{J})$.

(c) Die Konvergenz in Satz 3.20 gilt auch im 1. Mittel. Siehe etwa Billingsley, „Ergodic Theory and Information“.

*Beweis von Satz 3.20* (nach Katznelson, Weiss, 1982)*:* Mit der üblichen Zerlegung $X = X^+ - X^-$ reicht es, den Satz für $X \geq 0$ zu zeigen. Es sei

$$\overline{X}(\omega) := \limsup_{n\to\infty} S_n(\omega) \, , \quad \underline{X}(\omega) := \liminf_{n\to\infty} S_n(\omega) \, .$$

Es gilt $\underline{X} \circ T = \underline{X}$ und $\overline{X} \circ T = \overline{X}$, also sind $\underline{X}$ und $\overline{X}$ $\mathcal{J}$-messbar. Man zeigt nun

$$\int \overline{X} \, dP \leq \int X \, dP \leq \int \underline{X} \, dP \, , \tag{3.2}$$

denn mit $\underline{X} \leq \overline{X}$ folgt dann $\underline{X} = \overline{X}$ fast sicher, also konvergiert $(S_n)_n$ f. s. gegen eine $\mathcal{J}$-messbare Zufallsgröße mit den gewünschten Eigenschaften.

Für den Beweis der 1. Ungleichung in (3.2) verfolge man die Idee: Man schaue auf die „Zeitpunkte“ $0 = n_0 < n_1 < n_2 \ldots$, für die der durchschnittliche Zuwachs

$$\frac{1}{n_{j+1} - n_j} \left[ X(T^{n_j}\omega) + \cdots + X(T^{n_{j+1}-1}\omega) \right]$$

dem lim sup „nahe“ kommt. Es ergeben sich zwei technische Probleme. Erstens kann man nicht ausschließen, dass $\overline{X} = \infty$ ist. Man schneide ab: Für $M \in (0, \infty)$ sei $\overline{X}_M = \min(\overline{X}, M)$. $\varepsilon > 0$ sei beliebig. Für $\omega \in \Omega$ sei $n(\omega) \in \mathbb{N}$ die kleinste Zahl $k \in \mathbb{N}$ mit $S_k(\omega) \geq \overline{X}_M(\omega) - \varepsilon$. Die Abbildung $n : \Omega \to \mathbb{N}$ ist $\mathcal{A}$-$\mathcal{P}(\mathbb{N})$-messbar, denn für $k \in \mathbb{N}$ gilt

$$\{\omega : n(\omega) = k\} = \bigcap_{j=1}^{k-1} \{S_j < \overline{X}_M - \varepsilon\} \cap \{S_k \geq \overline{X}_M - \varepsilon\} .$$

$n$ als Funktion von $\omega$ braucht nicht beschränkt zu sein – das zweite technische Problem. Allerdings gilt

$$\lim_{N \to \infty} P\big(\{\omega : n(\omega) > N\}\big) = 0 .$$

Also existiert ein $N_\varepsilon \in \mathbb{N}$ mit $P\big(\{\omega : n(\omega) > N_\varepsilon\}\big) \leq \varepsilon$. Wir schneiden auch $n(\omega)$ und $X$ ab: Mit $A := \{\omega : n(\omega) \leq N_\varepsilon\}$ seien

$$\tilde{X}(\omega) := \begin{cases} X(\omega) , & \omega \in A , \\ M , & \omega \notin A \end{cases} \qquad \text{und} \qquad \tilde{n}(\omega) := \begin{cases} n(\omega) , & \omega \in A , \\ 1 , & \omega \notin A . \end{cases}$$

Gilt $X(\omega) > M$, so ist $n(\omega) = 1$, wegen $S_1(\omega) = X(\omega)$. Somit ist dann $\omega \in A$. Dann ist

$$X(\omega) \leq \tilde{X}(\omega) \quad \text{für alle } \omega , \tag{3.3}$$

denn im Fall $X(\omega) \leq M$ folgt die Ungleichung für $\omega \in A$ ohnehin und für $\omega \in A^c$ ist $M = \tilde{X}(\omega)$. Für $\omega \notin A$ gilt wegen $\tilde{n}(\omega) = 1$ und $\tilde{X}(\omega) = M$

$$\frac{1}{\tilde{n}(\omega)} \sum_{j=0}^{\tilde{n}(\omega)-1} \tilde{X}\big(T^j(\omega)\big) \geq \overline{X}_M(\omega) - \varepsilon . \tag{3.4}$$

Diese Ungleichung gilt auch für $\omega \in A$, denn dort ist $\tilde{n}(\omega) = n(\omega)$ und mit (3.3) ist die linke Seite von (3.4) nicht kleiner als $S_{n(\omega)}(\omega)$, also auch nicht kleiner als $\overline{X}_M(\omega) - \varepsilon$ nach der Definition von $n(\omega)$. Weiter gilt

$$\int \tilde{X}\, dP = \int_A \tilde{X}\, dP + \int_{A^c} \tilde{X}\, dP \leq \int X\, dP + M\varepsilon . \tag{3.5}$$

Nun definiert man rekursiv:

$$\begin{aligned} &n_0(\omega) := 0 , \quad n_1(\omega) := \tilde{n}(\omega) , \\ &n_k(\omega) := n_{k-1}(\omega) + \tilde{n}\big(T^{n_{k-1}(\omega)}(\omega)\big) \quad \text{für } k \in \mathbb{N} \backslash \{1\} . \end{aligned}$$

Für $m \in \mathbb{N}$ sei $K(\omega)$ die größte Zahl $k \in \mathbb{N}$ mit $n_k(\omega) \leq m$. $K(\omega)$ hängt somit auch von $m$ ab. Mit $\tilde{n}(\omega) \leq N_\varepsilon$ folgt $m - n_{K(\omega)}(\omega) \leq N_\varepsilon$. Dann folgt

$$\begin{aligned}\sum_{j=0}^{m-1} \tilde{X}(T^j\omega) &\geq \sum_{j=0}^{n_{K(\omega)}(\omega)-1} \tilde{X}(T^j\omega)\\ &= \sum_{j=0}^{n_1(\omega)-1} \tilde{X}(T^j\omega) + \sum_{j=n_1(\omega)}^{n_2(\omega)-1} \tilde{X}(T^j\omega)\\ &\quad + \cdots + \sum_{j=n_{K(\omega)-1}(\omega)}^{n_{K(\omega)}(\omega)-1} \tilde{X}(T^j\omega)\,.\end{aligned}$$

Nun verwenden wir (3.4), angewandt auf

$$\omega, T^{n_1(\omega)}(\omega), T^{n_2(\omega)}(\omega), \ldots, T^{n_{K(\omega)-1}(\omega)}(\omega)\,,$$

und erhalten

$$\begin{aligned}&\geq n_1(\omega)\big(\overline{X}_M(\omega) - \varepsilon\big)\\ &\quad + \big(n_2(\omega) - n_1(\omega)\big)\Big(\overline{X}_M\big(T^{n_1(\omega)}(\omega)\big) - \varepsilon\Big)\\ &\quad + \cdots + \big(n_{K(\omega)}(\omega) - n_{K(\omega)-1}(\omega)\big)\\ &\qquad \cdot \Big(\overline{X}_M\big(T^{n_{K(\omega)-1}(\omega)}(\omega)\big) - \varepsilon\Big)\,.\end{aligned}$$

Für alle $\omega \in \Omega$ und $j \in \mathbb{N}$ ist $\overline{X}_M(T^j\omega) = \overline{X}_M(\omega)$. Also folgt

$$\begin{aligned}&= n_{K(\omega)}(\omega)\overline{X}_M(\omega) - n_{K(\omega)}(\omega)\varepsilon\\ &\geq m\overline{X}_M(\omega) + \big(n_{K(\omega)}(\omega) - m\big)\overline{X}_M(\omega) - m\varepsilon\\ &\geq m\overline{X}_M(\omega) - N_\varepsilon M - m\varepsilon\,.\end{aligned}$$

Wir dividieren nun durch $m$ und beachten

$$\int \tilde{X}(T^j\omega)\,P(d\omega) = \int \tilde{X}\,dP \quad (T \text{ maßerhaltend})\,,$$

also

$$\int \tilde{X}\,dP \geq \int \overline{X}_M\,dP - \frac{N_\varepsilon M}{m} - \varepsilon\,.$$

Mit (3.5) folgt

$$\int X\,dP \geq \int \overline{X}_M\,dP - \frac{N_\varepsilon M}{m} - \varepsilon - M\varepsilon\,.$$

Da dies für alle $m \in \mathbb{N}$ und $\varepsilon > 0$ gilt, folgt

$$\int X\,dP \geq \int \overline{X}_M\,dP$$

für alle $M > 0$.

Der Satz von der monotonen Konvergenz liefert

$$\int X\,dP \geq \int \overline{X}\,dP\,.$$

Die zweite Ungleichung folgt analog und ist etwas einfacher, da $\underline{X}$ nicht nach unten abgeschnitten werden muss. $\square$

Der Begriff *Ergodensatz* hat seine Ursprünge in der Physik. Man denke sich $\Omega$ als den Phasenraum eines dynamischen Systems, zur Zeit $n = 0$ sei $\omega$ der Zustand des Systems, zur Zeit $n$ sei er gegeben durch $T^n\omega$, wobei $T : \Omega \to \Omega$ eine Bewegung beschreibe. Es sei nun $X(\omega)$ eine Beobachtung von $\omega$. Tatsächlich misst man in der Regel nicht $X(\omega)$ (die Beobachtungszeit ist häufig zu lang im Vergleich zur Zeitskala, in der molekulare Wechselwirkungen stattfinden), sondern durchschnittliche Werte von $X(\omega)$, die $\omega$ im Laufe seiner dynamischen Entwicklung annimmt, also

$$\frac{1}{n}\sum_{j=0}^{n-1} X \circ T^j(\omega).$$

J.W. Gibbs stellte in den letzten Jahren des 19. Jahrhunderts die Vermutung auf, dass $T^n(\omega)$ den ganzen Phasenraum durchlaufen sollte und die Dichte der Punkte $T^n(\omega)$ in einer kleinen Umgebung gegen eine Limesdichte konvergieren sollte. Diese Limesverteilung sollte dann invariant unter $T$ sein. Wenn die Limesverteilung $P$ ist, sollte man das obige diskrete *Zeitmittel* duch das *Phasenmittel*

$$\int X(\omega)\,P(d\omega)$$

ersetzen können. Birkhoff hat 1931 gezeigt, dass diese Vermutungen von Gibbs in einer geeigneten Formulierung stimmen. Dabei ist ein wichtiger Punkt, dass $T^n\omega$ niemals über den ganzen Phasenraum verteilt sein wird, wenn es Teilbereiche $A \subset \Omega$ gibt, die unter $T$ invariant sind und $P(A) > 0$ und $P(A^c) > 0$ gilt. Daher ist natürlich die einzige Hoffnung auf Bestätigung der Gibbsschen Vermutung die Annahme, dass invariante Mengen Wahrscheinlichkeit 0 oder 1 haben. Dies führte zu der gegebenen Definition ergodischer maßerhaltender Transformationen.

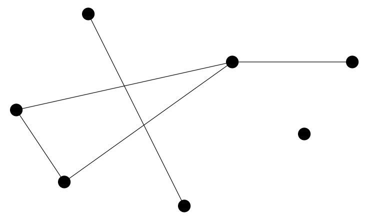
KAPITEL 4

# Zufallsgraphen

In diesem Kapitel wollen wir eine Einführung in die Theorie der zufälligen Graphen geben.

Wir betrachten ein paar Grundbegriffe der Graphentheorie, beschreiben das Modell eines Zufallsgraphen und führen sogenannte Schwellenfunktionen ein. Dieser Begriff wird an Hand einiger Grapheneigenschaften untersucht. Historisch sind die zu beschreibenden Objekte 1959 von PAUL ERDŐS eingeführt worden, und zwar zur Beantwortung einiger Fragen der Graphentheorie. Wir schließen das Kapitel mit dem Beweis eines fundamentalen Satzes von ERDŐS.

Ein *Graph* ist eine Ansammlung von Punkten (Knoten), wobei manche Punkte durch eine Kante verbunden sind. Etwa



**4.1 Definition** Ein *Graph* ist ein Paar $G = (V, E)$ disjunkter Mengen mit $E \subseteq [V]^2$ (bezeichne die Menge aller 2-elementigen Teilmengen von $V$). Elemente von $V$ nennt man *Ecken* oder *Knoten* (engl. *vertices*) des Graphen, die Elemente von $E$ heißen *Kanten* (engl. *edges*). Der Graph $G$ heißt *endlich* bzw. *unendlich* je nachdem, ob $V$ endlich oder unendlich ist. Für $|V|$ schreiben wir auch $|G|$ und nennen es die *Ordnung* von $G$. Zwei Knoten von $G$ sind *benachbart* in $G$, wenn $xy \in E$ ist. $xy$ bezeichnet kurz $\{x, y\}$; $x, y \in V$.

**Zur Historie:** Es ist manchmal schwer – warum auch immer –, einen Graphen mit einer speziellen Eigenschaft zu konstruieren: Gibt es Graphen, die gleichzeitig beliebig große *Taillenweite* und beliebig hohe *chromatische Zahl* haben? ERDŐS definierte für jedes $n \in \mathbb{N}$ auf der Menge aller Graphen der Ordnung $n$ einen Wahrscheinlichkeitsraum und zeigte, dass bei geeigneter Wahl der Wahrscheinlichkeitsmaße die Wahrscheinlichkeit, dass der gesuchte Graph existiert, für hinreichend große $n$ positiv wird! Man spricht von der *probabilistischen*

*Methode*. Wir werden diese historisch bedeutende Frage später genau untersuchen. Um die Frage aber zumindest formal verstehen zu können, hier die Definition der Begriffe:

**4.2 Definition**

(a) Ein *Weg* ist ein nicht leerer Graph $P = (V, E)$ mit

$$V = \{x_0, x_1, \ldots, x_k\} \quad \text{und} \quad E = \{x_0x_1, x_1x_2, \ldots, x_{k-1}x_k\},$$

wobei die $x_i$ paarweise verschieden sind. Die Anzahl der Kanten eines Weges ist seine *Länge*. Ist $P = x_0 \cdots x_{k-1}$ (Kurzschreibweise) ein Weg und $k \geq 3$, so ist $C := P + x_{k-1}x_0$ ein *Kreis* (wobei + bedeutet: wir vereinen die Kantenmenge bei gleichen $V = \{x_0, \ldots, x_{k-1}\}$). $C = x_0 \cdots x_{k-1}x_0$ (Kurzschreibweise). Ein Kreis der Länge $k$ wird mit $C^k$ bezeichnet.

(b) Die Länge eines kürzesten Kreises in einem Graphen $G$ (bezüglich $\subseteq$: ein Graph $G' = (V', E')$ ist ein Teilgraph von $G$, wenn $V' \subseteq V$ und $E' \subseteq E$ zu $G = (V, E)$) ist die *Taillenweite* $g(G)$ von $G$.

(c) Eine *Eckenfärbung* eines Graphen $G = (V, E)$ ist eine Abbildung $c : V \to S$ mit $c(v) \neq c(w)$ für je zwei benachbarte Ecken $v, w$. $S$ ist die Farbenmenge. Das kleinste $k \in \mathbb{N}$, so dass $G$ eine Eckenfärbung hat ($c : V \to \{1, \ldots, k\}$), nennt man die (Ecken-) *chromatische Zahl* von $G$, $\chi(G)$ in Bezeichnung.

Die oben genannte Frage, dessen Antwort Erdős fand, kann nun wie folgt gestellt werden: Existiert zu jedem $k \in \mathbb{N}$ ein Graph $G$ mit $g(G) > k$ *und* $\chi(G) > k$? Man benötigt Kenntnisse aus der Graphentheorie, um genauer zu verstehen, warum beide Forderungen gleichzeitig einer typischen Konstruktion entgegenstehen. Aus dem Bauch heraus verstehen wir jetzt aber, warum sich die beiden Forderungen an einen zu konstruierenden Graphen in die Quere kommen: Kreise (wie groß sie auch immer sind) benötigen zur Eckenfärbung nur 2 oder 3 Farben. Sie stören also, wenn man eine große Mindestzahl von Farben haben möchte.

Wir wenden uns den Zufallsgraphen zu: Zu $V = \{0, 1, \ldots, n-1\}$ wollen wir die Menge $\mathcal{G}$ der Graphen auf $V$ in einen Wahrscheinlichkeitsraum verwandeln. Dann machen Fragen der Art „mit welcher Wahrscheinlichkeit hat ein Graph $G \in \mathcal{G}$ eine spezielle Eigenschaft“ Sinn. Wir betrachten das sogenannte $\mathcal{G}(n, p)$-Modell:

**$\mathcal{G}(n, p)$-Modell:**

Für alle Eckenpaare $e \in [V]^2$ entscheidet man unabhängig voneinander, ob $e$ eine Kante von $G$ sein soll oder nicht. Die Wahrscheinlichkeit, dass $e$

eine Kante sei, soll $0 < p < 1$ sein (zeitliche Interpretation: starte mit einem leeren Graphen auf $n$ Knoten. Jede der $\binom{n}{2}$ Kanten wird mit Wahrscheinlichkeit $p$ eingefügt). Ist $G_0$ ein konkreter Graph auf $V$ mit $m$ Kanten, so hat das Elementarereignis $\{G_0\}$ die Wahrscheinlichkeit

$$p^m\, q^{\binom{n}{2}-m}, \;\; q := 1 - p.$$

Der Wahrscheinlichkeitsraum $\mathcal{G} = \mathcal{G}(n,p)$ ist nun ganz einfach zu konstruieren: zu $e \in [V]^2$ sei $\Omega_e := \{0,1\}$ und $P_e(0) = 1 - p$ und $P_e(1) = p$. Dann bezeichne $\mathcal{G}(n,p)$ den Produktraum $(\Omega, P)$ mit

$$\Omega := \prod_{e\in[V]^2} \Omega_e, \quad P \;\text{ Produktmaß.}$$

Ein $\omega \in \Omega$ identifiziert man mit einem Graphen $G$ auf $V$ und Kantenmenge $E = \{e : \omega(e) = 1\}$. Man sagt nun, $G$ ist ein *Zufallsgraph* auf $V$ mit Kantenwahrscheinlichkeit $p$. Es sei weiter

$$A_e := \{\omega : \omega(e) = 1\}$$

( $e$ ist eine Kante in $G$).

**4.3 Korollar** *Die Ereignisse $(A_e)_e$ sind unabhängige Ereignisse mit Wahrscheinlichkeit $p$.*

Dies ist nach Konstruktion des Produktraumes klar.

Es sei $X : \mathcal{G}(n,p) \to \mathbb{N}$ die Zufallsgröße, die jedem Zufallsgraphen $G$ die Anzahl seiner zu $C^k$ isomorphen Teilgraphen zuordnet. Wir sagen genauer, was dies bedeutet:

**4.4 Definition** $G = (V,E)$ und $G' = (V', E')$ seien zwei Graphen. Gilt $V' \subseteq V$ und $E' \subseteq E$, so ist $G'$ ein *Teilgraph* von $G$. $G$ heißt *isomorph* zu $G'$, $G \simeq G'$ in Zeichen, wenn es eine Bijektion $\varphi : V \to V'$ gibt mit

$$xy \in E \Leftrightarrow \varphi(x)\varphi(y) \in E' \quad \text{für alle } x, y \in V.$$

$\varphi$ heißt *Isomorphismus.*

Um mit den neuen Begriffen ein wenig Übung zu bekommen, zeigen wir

**4.5 Lemma** *Die mittlere Anzahl von Kreisen der Länge $k(\geq 3)$ in $G \in \mathcal{G}(n,p)$ beträgt*

$$\mathbb{E}(X) = \frac{(n)_k}{2k} p^k,$$

*wobei wir $(n)_k := n(n-1)(n-2)\cdots(n-k+1)$ setzen.*

*Beweis:* Für $C := C^k$ mit Ecken aus $V = \{0, \ldots, n-1\}$ sei

$$X_C : G \mapsto \begin{cases} 1, & \text{falls } C \subseteq G \text{ ist,} \\ 0, & \text{sonst.} \end{cases}$$

Dann ist

$$\mathbb{E}(X_C) = P(X_C = 1) = P(\text{ Menge aller Graphen aus } \mathcal{G}(n,p), \text{ die } C \text{ enthalten}).$$

Dies ist nichts anderes als die Wahrscheinlichkeit, dass $C$ ein Kreis in $G$ ist, also $P(C \subseteq G) = p^k$ (da $C$ ein Kreis der Länge $k$ ist). Wie viele dieser Kreise $C = v_0 \cdots v_{k-1} v_0$ gibt es? Es gibt $(n)_k$ Folgen $v_0 \cdots v_{k-1}$ mit (unterschiedlichen) Ecken aus $V$. Jeder Kreis wird durch $2\,k$ dieser Folgen beschrieben: also gibt es $\frac{(n)_k}{2k}$ solche Kreise. Nun ist $X$ die Summe aller $X_C$, wobei $C$ die $\frac{(n)_k}{2k}$ Kreise der Länge $k$ mit Ecken aus $V$ durchläuft, also

$$\mathbb{E}(X) = \sum_C \mathbb{E}(X_C) = \frac{(n)_k}{2k} p^k.$$

Damit ist das Lemma bewiesen. □

**Schwellenfunktionen:**

Man sagt, dass ein Graph $G$ eine *Eigenschaft* $\mathcal{A} \subseteq \mathcal{G}(n,p)$ hat, wenn $G \in \mathcal{A} \subseteq \mathcal{G}(n,p)$ gilt. Erdős und Rényi entdeckten 1960, dass viele Grapheneigenschaften ab einem gewissen $p(n)$ eintreten. Wir lassen nun also $p$ in $n$ variieren. Geht $P(G \in \mathcal{A}) \to 1$ für $n \to \infty$, so sagt man, dass *fast alle* $G$ in $\mathcal{G}(n,p)$ die Eigenschaft $\mathcal{A}$ haben. Strebt die Wahrscheinlichkeit gegen 0, so hat *fast kein* $G$ die Eigenschaft $\mathcal{A}$. (Vorsicht im Vergleich zur üblichen Definition von „fast alle“ (fast sicher).)

Ein kleiner Eindruck der *Evolution* eines Zufallsgraphen: Ist $p(n)$ wesentlich unterhalb von $1/n^2$, hat fast jeder Graph nur isolierte Ecken. Ab $p(n) = 1/(n\sqrt{n})$ hat fast jeder Graph die ersten Komponenten mit mehr als zwei Ecken (dabei ist eine Komponente ein maximal zusammenhängender nicht leerer Teilgraph, ohne Details). Ab $p(n) = 1/n$ tauchen die ersten Kreise auf. Eine spezielle Komponente beginnt stärker zu wachsen als die anderen und bei etwa $p(n) = \log n/n$ „verschlingt“ sie andere und die Graphen werden zusammenhängend.

Die Entwicklung vollzieht sich in Schüben: Die genannten $p(n)$-Werte sind Schwellenwerte, unterhalb derer fast kein Graph und oberhalb derer fast jeder Graph die betreffende Eigenschaft hat.

**4.6 Definition** Eine reelle Funktion $t = t(n)$ mit $t(n) \neq 0$ für alle $n \in \mathbb{N}$ heißt *Schwellenfunktion* für eine Grapheneigenschaft $\mathcal{A}$, wenn für jedes $p = p(n)$ und

$G \in \mathcal{G}(n,p)$ gilt:

$$\lim_{n\to\infty} P(G \in \mathcal{A}) = \begin{cases} 0, & \text{falls } p/t \underset{n\to\infty}{\longrightarrow} 0, \\ 1, & \text{falls } p/t \underset{n\to\infty}{\longrightarrow} \infty. \end{cases}$$

Ist $t$ Schwellenfunktion, so natürlich auch $c\,t$ mit $c \in \mathbb{R}_+$. Schwellenfunktionen sind also nur bis auf einen konstanten Faktor eindeutig bestimmt (wenn es sie gibt).

Wir wollen nun eine sehr zentrale Schwellenfunktion bestimmen: Zu einem gegebenen Graphen $H$ bezeichnen wir mit $\mathcal{A}_H$ die Grapheneigenschaft, eine Kopie von $H$ als Teilgraphen zu enthalten:

$$\mathcal{A}_H := \{G : H \subseteq G\}.$$

**4.7 Definition**

(a) Die *Dichte* eines Graphen $G = (V, E)$ ist definiert durch

$$\varrho(G) = \frac{|E|}{|V|}.$$

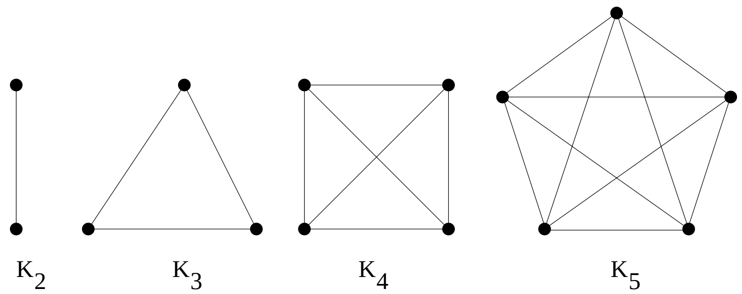
(b) Man nennt $G$ *ausgewogen* oder *balanciert*, wenn für alle Teilgraphen $G'$ von $G$ gilt:

$$\varrho(G') \leq \varrho(G)$$

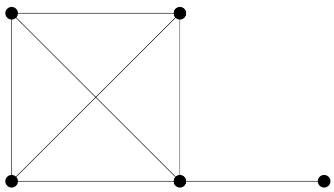
(*strikt ausgewogen, strikt balanciert*, wenn $<$ gilt).

**4.8 Beispiele**

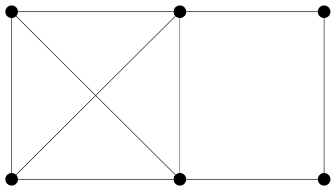
(a) Sind je zwei Ecken von $G$ benachbart, so heißt $G$ *vollständig*. Hat er $k$ Ecken, bezeichnen wir diesen Graphen mit $K_k$. $K_k$ ist strikt balanciert.



(b)



nicht balanciert



balanciert

Der erste Graph ist nicht balanciert, denn die Dichte ist $7/5$ und $K_4$ hat Dichte $\frac{3}{2}$. Der zweite ist balanciert, aber nicht strikt (liegt an $K_4$).

**4.9 Satz (von Erdős und Rényi)** *Es sei $H$ ein balancierter Graph mit $k$ Ecken und $l \geq 1$ Kanten. Dann ist $t(n) := n^{-k/l}$ eine Schwellenfunktion für $\mathcal{A}_H$.*

Der Beweis ist recht umfangreich und wird eine sehr wichtige, vielfach angewendete Methode präsentieren. Zuvor sammeln wir ein paar Folgerungen aus diesem Satz:

**4.10 Korollar** *Ist $k \geq 3$ und $\mathcal{A}$ die Eigenschaft, einen Kreis der Länge $k$ als Teilgraph zu enthalten, so ist $t(n) = 1/n$ eine Schwellenfunktion für $\mathcal{A}$.*

*Beweis:* Hier ist $l = k$ und $C^k$ ist balanciert. □

**4.11 Definition** Ein Graph heißt *zusammenhängend*, wenn er für je zwei seiner Ecken $x, y$ einen Weg von $x$ nach $y$ enthält. Ein Graph, der keinen Kreis enthält und zusammenhängend ist, heißt *Baum*. Eine alternative Definition (ohne Beweis): Ein zusammenhängender Graph mit $k$ Ecken ist genau dann ein Baum, wenn er $k-1$ Kanten hat.

**4.12 Korollar** *Ist $T$ ein Baum der Ordnung $k \geq 2$ und $\mathcal{A}$ die Eigenschaft, eine Kopie von $T$ als Teilgraph zu enthalten, so ist $t(n) = n^{-k/k-1}$ eine Schwellenfunktion für $\mathcal{A}$.*

*Beweis:* $k/k-1$ folgt aus der Definition eines Baums. Ein Baum ist balanciert. □

Bemerke, dass die Schwelle beim Kreis unabhängig ist von der Länge des Kreises. Beim Baum ist sie unabhängig von der Gestalt des Baumes.

**4.13 Korollar** *Ist $k \geq 2$ und $\mathcal{A}$ die Eigenschaft, einen $K_k$ als Teilgraph zu enthalten, so ist $t(n) = n^{-2/k-1}$ eine Schwellenfunktion.*

*Beweis:* Es gilt immer

$$\varrho(K_i) = \frac{1}{2}(i-1) < \frac{1}{2}(k-1) = \varrho(K_k)$$

für alle $i < k$. Nun ist $l = \binom{k}{2}$ für $K^k$. Es ist $l = \frac{1}{2}k(k-1)$ und daher ist die Schwelle $n^{-k/l} = n^{-2/k-1}$. □

Die *Cliquenzahl* eines Graphen $G$ ist definiert durch

$$w(G) := \max\{k : \text{ es existiert ein } K_k \text{ in } G\}.$$

Damit ist zum Beispiel die Schwellenfunktion der Eigenschaft

$$\mathcal{A} = \{\text{Existenz einer Clique der Größe mindestens } 4\} = \{w(G) \geq 4\}$$

gleich $n^{-2/3}$. Dies ist eine einfache Monotonie-Überlegung.

Es folgt nun der *Beweis von Satz 4.9*:

*Vorbereitung:* Es sei $X(G)$ die Anzahl der zu $H$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$. Weiter sei $\mathcal{H}$ die Menge aller zu $H$ isomorphen Graphen auf Teilmengen von $\{0,\dots,n-1\}$ (Eckenmenge der Graphen aus $\mathcal{G}(n,p)$). Formalisiert liefert dies:

$$\mathcal{H} = \{H' : H' \simeq H, V(H') \subseteq \{0,\dots,n-1\}\}$$

Es sei nun $H' \in \mathcal{H}$ und $G \in \mathcal{G}(n,p)$. Ist $H' \subseteq G$, so bezeichnet dies die Tatsache, dass $H'$ ein Teilgraph von $G$ ist.

Sei $h$ die Anzahl der zu $H$ isomorphen Graphen auf einer festen Menge von $k$ Ecken. Dann ist $h \leq k!$ und somit

$$|\mathcal{H}| = \binom{n}{k} h \leq \binom{n}{k} k! \leq n^k.$$

Zu $p = p(n)$ sei $\gamma = \frac{p}{t}$, also

$$p = \gamma\, n^{-k/l}.$$

Wir müssen nun zeigen:

(a) fast kein $G \in \mathcal{G}(n,p)$ liegt in $\mathcal{A}_H$, falls $\gamma \to 0$ für $n \to \infty$ und

(b) fast alle $G \in \mathcal{G}(n,p)$ liegen in $\mathcal{A}_H$, falls $\gamma \to \infty$ für $n \to \infty$.

*Beweis von Punkt (a):* Hier führen wir den Beweis mit Hilfe der *Methode der ersten Momente*: Dazu eine Erinnerung:

**4.14 Lemma** *Ist $X$ eine nicht-negative, ganzzahlige Zufallsgröße, so ist*

$$P(X > 0) \leq \mathbb{E}(X).$$

*Beweis:* Es gilt

$$\mathbb{E}(X) = \sum_{i \geq 0} i\, P(X = i) \geq \sum_{i \geq 1} P(X = i) = P(X > 0).$$

□

Gilt für eine Folge von nicht-negativen Zufallsgrößen $(X_n)_n$, dass $\lim_{n \to \infty} \mathbb{E}(X_n) = 0$, so folgt

$$\lim_{n \to \infty} P(X_n = 0) = 1,$$

und man sagt hierzu, $X_n$ ist fast sicher gleich Null. Wir wollen also $\mathbb{E}(X)$ bestimmen. Es gilt

$$\mathbb{E}(X) = \sum_{H' \in \mathcal{H}} P(H' \subseteq G) = |\mathcal{H}|\, p^l \leq n^k (\gamma n^{-k/l})^l = \gamma^l$$

und mit $\gamma$ geht also auch $\mathbb{E}(X)$ für wachsende $n$ gegen Null. Also liegt fast kein $G$ aus $\mathcal{G}(n,p)$ in $\mathcal{A}_H$.

*Beweis von Punkt (b):* Hier wird uns die *Methode der zweiten Momente* helfen:

**4.15 Lemma** *Ist $X$ eine Zufallsgröße mit Werten in $\mathbb{N}_0$, so gilt*

$$P(X = 0) \leq \frac{\operatorname{Var}(X)}{(\mathbb{E}X)^2}.$$

*Beweis:*

$$P(X = 0) \leq P\big(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)\big) \leq \frac{\operatorname{Var}(X)}{(\mathbb{E}X)^2}$$

nach der TSCHEBYSCHEV-Ungleichung. □

Ist nun $\operatorname{Var}(X) = o((\mathbb{E}X)^2)$, so ist $X > 0$ fast sicher. Wir untersuchen also $\operatorname{Var}(X)$: Zunächst ist

$$\mathbb{E}(X^2) = \sum_{(H',H'') \in \mathcal{H}^2} P(H' \cup H'' \subseteq G) = \sum_{(H',H'') \in \mathcal{H}^2} p^{2l - |E(H' \cap H'')|}.$$

Nun ist $\varrho(H' \cap H'') \le \varrho(H) = l/k$ (balanciert!). Ist $|V(H' \cap H'')| = i$, so folgt $|E(H' \cap H'')| \le \frac{il}{k}$. Also ist wegen $p < 1$

$$P(H' \cup H'' \subseteq G) \le p^{2l - \frac{il}{k}}.$$

Nun betrachten wir

$$\mathcal{H}_i^2 := \{(H', H'') \in \mathcal{H}^2 : |V(H' \cap H'')| = i\}$$

für $i = 0, \ldots, k$. Wir bestimmen nun

$$\sum_i P(H' \cup H'' \subseteq G) =: A_i,$$

wobei $\sum_i$ die Summe über alle $(H', H'') \in \mathcal{H}_i^2$ bezeichne. Im Fall $i = 0$ sind $H'$ und $H''$ disjunkt und $\{H' \subseteq G\}$ ist somit unabhängig von $\{H'' \subseteq G\}$. Also

$$A_0 \le \sum_{(H', H'') \in \mathcal{H}^2} P(H' \subseteq G)\, P(H'' \subseteq G) = (\mathbb{E}X)^2.$$

Für $i \ge 1$ ist

$$\sum_i = \sum_{H', H'' \in \mathcal{H} : |V(H' \cap H'')| = i}.$$

Ist $H'$ fest gewählt, so hat die Summe über die $H''$

$$\binom{k}{i}\binom{n-k}{k-i} h$$

Summanden. Damit ist

$$\begin{aligned}
A_i &\le \sum_{H' \in \mathcal{H}} \binom{k}{i}\binom{n-k}{k-i} h\, p^{2l} p^{-il/k} \\
&= |\mathcal{H}| \binom{k}{i}\binom{n-k}{k-i} h\, p^{2l} \left(\gamma\, n^{-k/l}\right)^{-il/k} \\
&\le |\mathcal{H}|\, p^l\, c_1\, n^{k-i}\, h\, p^l\, \gamma^{-il/k}\, n^i \\
&= \mathbb{E}(X)\, c_1\, n^k\, h\, p^l\, \gamma^{-il/k} \\
&\le \mathbb{E}(X)\, c_2 \binom{n}{k} h\, p^l\, \gamma^{-il/k} \\
&= \mathbb{E}(X)^2\, c_2\, \gamma^{-il/k} \\
&\le \mathbb{E}(X)^2\, c_2\, \gamma^{-l/k}.
\end{aligned}$$

Also folgt mit $c_3 = k\, c_2$:

$$\frac{\mathbb{E}(X^2)}{(\mathbb{E}X)^2} = \left(\frac{A_0}{(\mathbb{E}X)^2} + \sum_{i=1}^{k} A_i \frac{1}{(\mathbb{E}X)^2}\right) \le 1 + c_3\, \gamma^{-l/k},$$

also

$$\frac{\mathbb{E}(X^2) - (\mathbb{E}X)^2}{(\mathbb{E}X)^2} \leq c_3\, \gamma^{-l/k},$$

und die rechte Seite geht für $\gamma \to \infty$ gegen Null. Somit ist $X(G) > 0$ für fast alle $G \in \mathcal{G}(n,p)$, das heißt fast alle $G \in \mathcal{G}(n,p)$ enthalten eine Kopie von $H$ als Teilgraphen und liegen somit in $\mathcal{A}_H$. Damit ist der Satz bewiesen. $\square$

Für die Anzahl $X(G)$ der zu $H$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$ hatten wir $\mathbb{E}(X) = |\mathcal{H}|\, p^l \sim n^k\, p^l$ gesehen. Somit folgt bei der Wahl $\lim_{n\to\infty} p/n^{-k/l} = 0$ für $p = p(n)$, dass $\lim_{n\to\infty} \mathbb{E}(X) = 0$, und für die Wahl $\lim_{n\to\infty} p/n^{-k/l} = \infty$, dass $\lim_{n\to\infty} \mathbb{E}(X) = \infty$ gilt. Dies gilt für alle $H$ mit $k$ Ecken und $l$ Kanten. Trotzdem gilt der folgende Satz:

**4.16 Satz** *Es sei $H$ nicht balanciert mit $k$ Ecken und $l \geq 1$ Kanten. Dann ist $t(n) = n^{-k/l}$ keine Schwellenfunktion für $\mathcal{A}_H$.*

*Beweis:* Es sei $H' \subseteq H$ so gewählt, dass $\varrho(H') > \frac{l}{k}$ ($H$ nicht balanciert). Dann ist

$$\begin{aligned} \mathbb{E}(X') &\sim n^{|V(H')|}\, p^{|E(H')|} \\ &= (n^{|V(H')|/|E(H')|}\, p)^{|E(H')|}, \end{aligned}$$

wobei $X'(G)$ die Anzahl der zu $H'$ isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$ bezeichnet und $|V(H')|$ ($|E(H')|$) die Knotenanzahl (Kantenanzahl) von $H'$. Wir wählen nun $p = p(n)$ so, dass

$$p(n)/n^{-k/l} \to \infty \quad \text{und} \quad p(n)/n^{-|V(H')|/|E(H')|} \to 0$$

(dies ist möglich, da $\frac{|E(H')|}{|V(H')|} > \frac{l}{k}$). Damit folgt aber $\lim\limits_{n\to\infty} \mathbb{E}(X') = 0$, also $\lim\limits_{n\to\infty} P(X' = 0) = 1$. Da weiter $\{X' = 0\} \subseteq \{X = 0\}$, kann $t(n)$ keine Schwellenfunktion für $\mathcal{A}_H$ sein. $\square$

Wir stellen nun die zu Beginn des Kapitels erwähnte *probabilistische Methode* exemplarisch mittels des Beweises des Satzes von Erdős über die Existenz von Graphen mit hoher Taillenweite und chromatischer Zahl dar.

**4.17 Satz (von Erdős, 1959)** *Zu jedem $k \in \mathbb{N}$ gibt es einen Graphen $H$ mit Taillenweite $g(H) > k$ und chromatischer Zahl $\chi(H) > k$.*

Wir bereiten den Beweis vor. Zu einem Graph $G = (V,E)$ heißt eine Teilmenge von $V$ (Eckenmenge) *unabhängig*, wenn ihre Elemente paarweise nicht benachbart sind. Die größte Mächtigkeit einer unabhängigen Eckenmenge in $G$ ist die *Unabhängigkeitszahl* $\alpha(G)$ von $G$.

**4.18 Proposition** Für jedes $k \geq 2$ gilt für die Wahrscheinlichkeit, dass $G \in \mathcal{G}(n,p)$ eine unabhängige Eckenmenge der Mächtigkeit $k$ enthält,

$$P\big(\alpha(G) \geq k\big) \leq \binom{n}{k}(1-p)^{\binom{k}{2}}.$$

*Beweis:* Die Wahrscheinlichkeit, dass eine fest gewählte Eckenmenge $U \in [V]^k$ in $G$ unabhängig ist, beträgt $(1-p)^{\binom{k}{2}}$. Es gibt nur $\binom{n}{k}$ solche Mengen $U$. □

**4.19 Bemerkung** (analog zur Proposition zu beweisen)

$$P(\omega(G) \geq k) \leq \binom{n}{k} p^{\binom{k}{2}}.$$

(Wahrscheinlichkeit, dass $G \in \mathcal{G}(n,p)$ einen $K_k$ enthält; Erinnerung: $\omega(G)$ ist die Cliquenzahl).

Wir nennen Kreise der Länge $\leq k$ *kurz* und Mengen von $\frac{|G|}{k}$ oder mehr Knoten von $G$ *groß*. Wir suchen (Erdős suchte) einen Graphen $G$ *ohne kurze Kreise* und *ohne große unabhängige Eckenmengen*: die letzt genannte Eigenschaft impliziert, dass wir mehr als $k$ Farben brauchen, um $G$ zu färben (denn: $\alpha(G) < \frac{|G|}{k}$, also $k < \frac{|G|}{\alpha(G)} \leq \chi(G)$). Die letzte Ungleichung ist eine bekannte Schranke für die chromatische Zahl, die man wie folgt einsehen kann: Die Knoten einer Knotenmenge $\tilde{V}$ von $G$ können genau dann mit derselben Farbe belegt werden, wenn $\tilde{V}$ eine unabhängige Knotenmenge ist. Es sei eine Färbung von $G$ mit $k = \chi(G)$ Farben gegeben. Wir ordnen jeder Farbe $i$ die Menge $V_i \subset V$ der mit $i$ gefärbten Knoten zu. Dann sind alle Mengen $V_1, \ldots, V_k$ unabhängig. Es gilt jeweils $|V_i| \leq \alpha(G)$, also

$$|G| = \sum_{i=1}^{k} |V_i| \leq \alpha(G)\,\chi(G),$$

woraus die behauptete Ungleichung folgt. Wählt man $p$, die Kantenwahrscheinlichkeit, klein genug, so werden die Zufallsgraphen aus $\mathcal{G}(n,p)$ mit hoher Wahrscheinlichkeit keine (kurzen) Kreise haben (wir kennen bereits die Schwelle $p(n) = 1/n$). Intuitiv erwartet man bei $p$ groß genug, dass kaum große unabhängige Eckenmengen auftreten. Es ergibt sich somit die folgende Frage: Kann man $p$ so wählen, dass es für große $n$ gleichzeitig klein genug für $P(g \leq k) < \frac{1}{2}$ und groß genug für $P(\alpha \geq n/k) < \frac{1}{2}$ ist? Wenn ja, so enthält $\mathcal{G}(n,p)$ mindestens einen Graphen ohne kurze Kreise und ohne große unabhängige Eckenmengen. Tatsächlich kann man $p$ so nicht wählen, da $1/n$ auch eine Schwellenfunktion für große unabhängige Mengen ist. Der Beweis des Satzes verfolgt eine andere Idee.

*Beweis von Satz 4.17:* Es sei $\varepsilon > 0$ und $\varepsilon < 1/k$ fest gewählt. Sei $p := p(n) = \frac{n^\varepsilon}{n} = \frac{1}{n^{1-\varepsilon}}$. Wir liegen oberhalb der Schwelle $1/n$ und kümmern uns um die Anzahl $X(G)$ der kurzen Kreise in einem Zufallsgraphen $G \in \mathcal{G}(n,p)$ (Anzahl der Kreise der Länge $\leq k$). Es gilt mit Lemma 4.5:

$$\mathbb{E}(X) = \sum_{i=3}^{k} \frac{(n)_i}{2i} p^i \leq \frac{1}{2} \sum_{i=3}^{k} p^i\, n^i \leq \frac{1}{2}(k-2) n^k\, p^k,$$

denn $(n\,p)^i \leq (n\,p)^k$ wegen $n\,p = n^\varepsilon \geq 1$. Daraus folgt

$$P\left(X \geq \frac{n}{2}\right) \leq \frac{\mathbb{E}(X)}{n/2} \leq (k-2)\, n^{k-1}\, p^k = (k-2)\, n^{k-1}\, n^{(\varepsilon-1)k} = (k-2)\, n^{k\,\varepsilon-1}.$$

Nun ist $k\,\varepsilon - 1 < 0$ nach obiger Wahl, also

$$\lim_{n\to\infty} P\left(X \geq \frac{n}{2}\right) = 0.$$

Wir wählen $n$ so groß, dass $P\big(X \geq n/2\big) < 1/2$. Wir finden also einen Graphen $G \in \mathcal{G}(n,p)$ mit weniger als $n/2$ kurzen Kreisen.

*Behauptung:* $n$ kann so groß gewählt werden, dass $P\big(\alpha \geq \frac{n}{2k}\big) < 1/2$ bei obiger Wahl von $p(n) = \frac{1}{n^{1-\varepsilon}}$ gilt. Wir finden dann einen Graphen $G \in \mathcal{G}(n,p)$ mit $\alpha(G) < \frac{n}{2k}$ und mit weniger als $n/2$ kurzen Kreisen. Dann sind wir aber bereits am Ziel:

Aus *jedem* kurzen Kreis von $G$ entfernen wir eine Ecke. Der entstehende Graph $H$ hat dann noch mindestens $n/2$ Ecken. Er enthält keine kurzen Kreise mehr, also ist $g(H) > k$ und weiter gilt:

$$\chi(H) \geq \frac{|H|}{\alpha(H)} \geq \frac{n/2}{\alpha(G)} > k$$

(die Unabhängigkeitszahl ist nach Konstruktion nicht größer als die von $G$). Damit ist der Satz bewiesen, wenn obige Behauptung nachgeliefert wird. □

**4.20 Lemma** *Es seien $k > 0$ und $p = p(n)$ gegeben. Ist $p \geq \frac{6k\,\log n}{n}$ für große $n$, so gilt*

$$\lim_{n\to\infty} P\left(\alpha \geq \frac{n}{2k}\right) = 0.$$

Mit Hilfe dieses Lemmas folgt dann für $p = n^{\varepsilon-1}$ für beliebig kleines $\varepsilon > 0$, dass ein $G \in \mathcal{G}(n,p)$ existiert mit $\alpha(G) < n/(2k)$.

*Beweis von Lemma 4.20:* Es seien $n, r \in \mathbb{N}$ mit $n \geq r \geq 2$. Weiter sei $G \in \mathcal{G}(n,p)$. Mit $q := 1 - p$ gilt nach Proposition 4.18

$$\begin{aligned} P(\alpha \geq r) &\leq \binom{n}{r} q^{\binom{r}{2}} \leq n^r q^{\binom{r}{2}} \\ &= \left(n\, q^{\frac{r-1}{2}}\right)^r \leq \left(n\, e^{-\frac{p(r-1)}{2}}\right)^r, \end{aligned}$$

wobei wir $(1-p) \leq e^{-p}$, $0 < p < 1$, verwendet haben. Es sei nun $p \geq \frac{6k \log n}{n}$ und $r \geq \frac{n}{2k}$, so folgt

$$\begin{aligned} n\, e^{-p(r-1)/2} &= n\, e^{-pr/2+p/2} \leq n\, e^{-\frac{3}{2}\log n + p/2} \\ &\leq n\, n^{-3/2} e^{1/2} = \frac{\sqrt{e}}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

Mit $r := [\frac{n}{2k}]$ ist

$$\lim_{n\to\infty} P\left(\alpha \geq \frac{n}{2k}\right) = \lim_{n\to\infty} P(\alpha \geq r) = 0,$$

womit das Lemma bewiesen ist. $\square$

KAPITEL 5

# Perkolation

Wir werfen einen porösen Stein in Wasser und fragen nach der Wahrscheinlichkeit, dass das Zentrum des Steins nass wird. Broadbent und Hammersley haben 1957 ein stochastisches Modell zu dieser Frage entwickelt. Es ist die Geburtsstunde der Untersuchung von Perkolation.

In $\mathbb{Z}^d$, $d \geq 1$, sei

$$\delta(x,y) := \sum_{i=1}^{d} |x_i - y_i| \,, \quad x, y \in \mathbb{Z}^d \,.$$

Wir verbinden zwei „Knoten" $x, y \in \mathbb{Z}^d$ durch eine Kante, wenn $\delta(x,y) = 1$. Es entsteht der Graph $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$, wenn $\mathbb{E}^d$ die Menge aller Kanten bezeichnet. Es sei $p$ und $q$ durch $0 \leq p \leq 1$ und $p + q = 1$ festgelegt. Wir nennen eine Kante in $\mathbb{L}^d$ *offen* mit Wahrscheinlichkeit $p$ und *geschlossen* sonst, unabhängig von allen anderen Kanten. Formal meinen wir

$$\Omega = \prod_{e \in \mathbb{E}^d} \{0, 1\} \,;$$

Punkte in $\Omega$ heißen Konfigurationen $\omega = \big(\omega(e) : e \in \mathbb{E}^d\big)$, $\omega(e) = 0$ korrespondiere zu einer geschlossenen Kante $e$, $\omega(e) = 1$ zu einer offenen Kante $e$. $\mathcal{A}$ sei die $\sigma$-Algebra erzeugt von den endlich-dimensionalen Projektionen, und das W-Maß sei gegeben durch

$$P_p = \prod_{e \in \mathbb{E}^d} \mu_e \,, \quad \mu_e\big(\omega(e) = 0\big) = q \,, \quad \mu_e\big(\omega(e) = 1\big) = p \,.$$

Es liegt eine eins-zu-eins Korrespondenz zwischen $\Omega$ und der Menge der Teilmengen von $\mathbb{E}^d$ vor.

Für $\omega \in \Omega$ sei

$$K(\omega) := \{e \in \mathbb{E}^d : \omega(e) = 1\}$$

die Menge der offenen Kanten zu $\omega$. Es gilt $\omega_1 \leq \omega_2$ genau dann, wenn $K(\omega_1) \subseteq K(\omega_2)$. Es liegt uns also ein spezieller Zufallsgraph vor (siehe Abb. 5.1).

Angenommen, $\big(X(e), e \in \mathbb{E}^d\big)$ sei eine Familie unabhängiger Zufallsgrößen und $X(e)$ ist gleichmäßig verteilt auf $[0,1]$. Wir „koppeln" nun *alle* Kanten-Perkolationsprozesse auf $\mathbb{L}^d$, wenn $p$ das Intervall $[0,1]$ durchläuft. Sei $0 \leq p \leq 1$ und

$$\eta_p(e) = \begin{cases} 1 \,, & \text{wenn } X(e) < p \,, \\ 0 \,, & \text{wenn } X(e) \geq p \,. \end{cases}$$

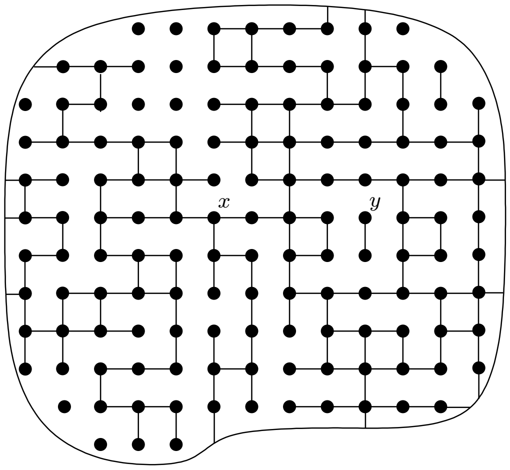



Abb. 5.1: Skizze der Struktur eines zweidimensionalen porösen Steins. Die Linien deuten die offenen Kanten an; geschlossene Kanten wurden weggelassen. Beim Eintauchen des Steins in Wasser wird der Knoten $x$ durch das Eindringen des Wassers nass, während der Knoten $y$ trocken bleibt.

$e$ nennen wir $p$-*offen*, wenn $\eta_p(e) = 1$. Es gilt $P\big(\eta_p(e) = 0\big) = 1 - p$, $P\big(\eta_p(e) = 1\big) = p$, und die Komponenten von $\eta_p$ sind unabhängig. Wenn $p_1 \le p_2$, so ist $\eta_{p_1} \le \eta_{p_2}$. Mit Kopplung ist hierbei gemeint, dass wir den Grundraum $\Omega^{[0,1]}$ mit der Produkt-$\sigma$-Algebra und dem Produktmaß $\otimes_{p\in[0,1]} P_p$ versehen und diesen gemeinsamen W-Raum verwenden.

Wir nennen einen Weg oder Kreis *offen*, wenn alle seine Kanten offen sind, und *geschlossen*, wenn alle Kanten geschlossen sind. Teilgraphen von $\mathbb{L}^d$ heißen *Kanten-disjunkt*, wenn sie keine gemeinsamen Kanten haben, disjunkt, wenn sie keine gemeinsamen Kanten und keine gemeinsamen Knoten haben.

Nun betrachten wir den Zufallsgraph in $\mathbb{L}^d$, der aus den Knoten $\mathbb{Z}^d$ und den offenen Kanten besteht. Zusammenhangskomponenten dieses Graphen heißen *offene Cluster*.

$C(x)$ nennen wir den offenen Cluster, der $x$ (Knoten) enthält. $C := C(0)$. $C(x)$ steht auch für die Knotenmenge des Clusters, $|C(x)|$ ist die Ordnung des Clusters. Für zwei Teilmengen $A$ und $B$ von Knoten in $\mathbb{L}^d$ schreiben wir $A \leftrightarrow B$, wenn ein offener Pfad/Weg existiert, der einen Knoten in $A$ mit einem in $B$ verbindet.

$$C(x) = \{y \in \mathbb{Z}^d : x \leftrightarrow y\}\,.$$

Wir interessieren uns für die *Perkolations-Wahrscheinlichkeit* $\Theta(p)$. Dies ist die Wahrscheinlichkeit, dass ein gegebener Knoten zu einem unendlichen Cluster gehört. Das Gitter und $P_p$ sind translationsinvariant. Daher setzen wir

$$\Theta(p) := P_p(|C| = \infty) = 1 - \sum_{n=1}^{\infty} P_p(|C| = n)\,.$$

**5.1 Lemma** $\Theta$ *wächst monoton und es gilt*

$$\Theta(0) = 0 \quad \textit{und} \quad \Theta(1) = 1\,.$$

*Beweis:* Die obige Kopplung durch Konstruktion der $\eta_p$ liefert: Für $p_1 \leq p_2$ folgt $\eta_{p_1} \leq \eta_{p_2}$. Wir nennen eine Zufallsgröße $N$ auf $(\Omega, \mathcal{A})$ wachsend, wenn $N(\omega) \leq N(\omega')$ für $\omega \leq \omega'$ folgt. Für $p_1 \leq p_2$ folgt dann $N(\eta_{p_1}) \leq N(\eta_{p_2})$. Nun wählen wir den Produktraum

$$(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, P_{p_1} \otimes P_{p_2}).$$

Dann folgt

$$\mathbb{E}_{P_{p_1}}(N) \leq \mathbb{E}_{P_{p_2}}(N).$$

Nun ist $\{|C| = \infty\}$ ein wachsendes Ereignis in dem Sinne, dass $\omega \in \{|C| = \infty\}$ impliziert dass $\omega' \in \{|C| = \infty\}$ für jedes $\omega \leq \omega'$. Somit ist $1_{\{|C|=\infty\}}$ wachsend, also folgt

$$\Theta(p_1) = P_{p_1}(|C_{p_1}| = \infty) \leq P_{p_2}(|C_{p_2}| = \infty) = \Theta(p_2)\,. \qquad \square$$

Sei nun

$$p_c(d) := \sup\{p : \Theta(p) = 0\}$$

die *kritische Wahrscheinlichkeit*. Es gilt $\Theta(p) = 0$ für $p < p_c$ und $\Theta(p) > 0$ für $p > p_c$. Betrachten wir den Fall $d = 1$. Für $p < 1$ existieren unendlich viele geschlossene Kanten in $\mathbb{L}^1$ links und rechts der Null fast sicher, also $\Theta(p) = 0$ für $p < 1$, also $p_c(1) = 1$. Sei nun fortan $d \geq 2$.

$\mathbb{L}^d$ können wir in $\mathbb{L}^{d+1}$ einbetten (Projektion auf die ersten $d$ Koordinaten). Dann gehört die Null in $\mathbb{L}^{d+1}$ einem unendlichen offenen Cluster für ein $p$ an, wenn die Null dies in $\mathbb{L}^d$ erfüllt. Also ist $\Theta(p) = \Theta_d(p)$ wachsend in $d$, also

$$p_c(d+1) \leq p_c(d)$$

für $d \geq 1$. Es gilt sogar strikte $<$-Relation (ohne Beweis).

Wir wollen die folgenden Sätze betrachten:

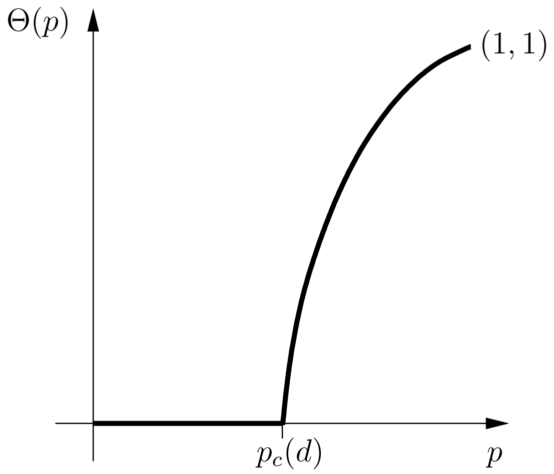
**5.2 Satz** *Für* $d \geq 2$ *ist* $0 < p_c(d) < 1$.

**5.3 Satz** *Die Wahrscheinlichkeit* $\psi(p)$ *für die Existenz eines unendlichen offenen Clusters genügt*

$$\psi(p) = \begin{cases} 0\,, & \textit{wenn } \Theta(p) = 0\,, \\ 1\,, & \textit{wenn } \Theta(p) > 0\,. \end{cases}$$

### 5.4 Bemerkungen

(a) Man glaubt den folgenden Verlauf der Perkolations-Wahrscheinlichkeit $\Theta(p)$:




$\Theta$ ist stetig in $p$, außer in $p_c(d)$. Für $3 \le d < 19$ ist es offen, ob $\Theta$ in $p_c(d)$ springt!

(b) In der *subkritischen Phase* $p < p_c(d)$ ist jeder Knoten (fast sicher) in einem endlichen Cluster. In der *superkritischen Phase* $p > p_c(d)$ hat jeder Knoten strikt positive Wahrscheinlichkeit, in einem unendlichen offenen Cluster zu sein. Es existiert also fast sicher mindestens ein unendlicher offener Cluster. In $p = p_c(d)$ existiert ein unendlicher offener Cluster, wenn $\Theta\big(p_c(d)\big) > 0$. Es ist bekannt, dass für $d = 2$ und $d \ge 19$ *kein* unendlicher offener Cluster existiert, die verbleibenden Dimensionen sind ungeklärt! Man erwartet, dass kein unendlicher offener Cluster existiert.

(c) Wir werden Satz 5.3 mittels eines 0-1-Gesetzes beweisen. Der Satz sagt nichts über die Anzahl von unendlichen offenen Clustern im Fall $\Theta(p) > 0$ aus. Tatsächlich ist aber bekannt, dass der unendliche offene Cluster fast sicher eindeutig ist, wenn er existiert.

(d) $p_c(1) = 1$. Man kennt auch $p_c(2) = 1/2$ (H. Kesten, 1980); $p_c(d)$, $d \ge 3$, sind nicht bekannt! Wir werden im Laufe des Beweises von Satz 5.2 untere und obere Schranken für $p_c(d)$, $d \ge 2$, kennenlernen.

Wir bereiten den Beweis von Satz 5.2 vor. Dazu führen wir den Begriff *selbst-meidender Pfade (self-avoiding)* ein. Ein $n$-Schritt selbst-meidender Pfad $\omega$ in $\mathbb{Z}^d$ mit Start in 0 ist durch $\big(\omega(0) = 0, \omega(1), \dots, \omega(n)\big)$ mit $\delta\big(\omega(j+1), \omega(j)\big) = 1$ und $\omega(i) \ne \omega(j)$ für alle $i \ne j$ gegeben. $\sigma(n)$ bezeichne die Anzahl der $n$-Schritt selbst-meidenden Pfade mit Start in 0, so ist

$$\sigma(1) = 2d\,, \quad \sigma(2) = 2d(2d-1)\,, \quad \sigma(3) = 2d(2d-1)^2$$

und

$$\sigma(4) = 2d(2d-1)^3 - 2d(2d-2) \quad (!)$$

$\sigma(n)$ ist allgemein schwierig zu bestimmen, aber es gilt

$$\sigma(n) \leq 2d(2d-1)^{n-1}, \tag{5.1}$$

denn in jedem neuen Schritt eines selbst-meidenden Pfades hat man maximal $2d-1$ Schrittmöglichkeiten, da man die letzte Position meiden muss. Weiter gilt

$$2d^n \leq \sigma(n). \tag{5.2}$$

Dazu betrachten wir einfach die Zahl der Pfade, bei denen jeder Schritt in eine der $d$ positiven (oder $d$ negativen) Koordinatenrichtungen erfolgt. Diese Pfade sind selbst-meidend. Sollte der Limes $\lim_{n\to\infty}\big(\sigma(n)^{1/n}\big)$ existieren, so besagen (5.1) und (5.2)

$$d \leq \lim_{n\to\infty} \sigma(n)^{1/n} =: \lambda(d) \leq 2d-1. \tag{5.3}$$

**5.5 Lemma**

(a) *Für $n, m \in \mathbb{N}$ gilt*

$$\sigma(n+m) \leq \sigma(n) \cdot \sigma(m),$$

*also*

$$\log \sigma(n+m) \leq \log \sigma(n) + \log \sigma(m).$$

(b) $\lambda(d) = \lim_{n\to\infty} \sigma(n)^{1/n}$ *existiert und*

$$\log \lambda(d) = \inf_{n\geq 1} \frac{1}{n} \log \sigma(n),$$

*also*

$$\lambda(d)^n \leq \sigma(n), \quad n \geq 1.$$

*Beweis:* Das Produkt $\sigma(n)\cdot\sigma(m)$ ist die Anzahl der Menge der Pfade der Länge $n+m$, die selbst-meidend bei den ersten $n$ Schritten und bei den letzten $m$ Schritten sind, die aber nicht notwendig insgesamt selbst-meidend sind, also

$$\sigma(n+m) \leq \sigma(n)\sigma(m).$$

Der Teil (b) folgt aus einem einfachen Argument der Analysis und geht auf Hammersley und Morton (1954) zurück:

**5.6 Lemma** *Sei $(a_n)_n$ eine Folge reeller Zahlen, die subadditiv ist, d. h. $a_{n+m} \leq a_n + a_m$, $n, m \in \mathbb{N}$. Dann existiert $\lim_{n\to\infty} a_n/n$ in $[-\infty, +\infty)$ und*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n}.$$

Somit ist für $a_n = \log\sigma(n)$ Lemma 5.5 bewiesen. $\square$

*Beweis von Lemma 5.6:* Wir zeigen

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \frac{a_k}{k} \quad \forall k\,, \tag{5.4}$$

denn dann folgt mit $\liminf_{k\to\infty}$ in (5.4) die Existenz des Grenzwertes und die Wahl von $\inf_{k\ge 1}$ in (5.4) liefert die zweite Behauptung.

Sei $k$ fest und $A_k := \max_{1\le r\le k} a_r$. Für $n \in \mathbb{N}\backslash\{0\}$ sei $j$ die größte Zahl strikt kleiner $n/k$. Dann ist $n = jk+r$ mit $1 \le r \le k$, $r \in \mathbb{N}$. Die Subadditivität liefert

$$\frac{a_n}{n} \le \frac{ja_k + a_r}{jk+r}.$$

Dann folgt im $\limsup_{n\to\infty}$

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \limsup_{j\to\infty} \Big(\frac{k}{a_k} + \frac{r}{ja_k}\Big)^{-1} = \frac{a_k}{k}.$$

Die Übereinstimmung mit dem Infimum zeigt $\lim_{n\to\infty} n^{-1}a_n < \infty$. Die Wahl $a_n = -n^2$ zeigt, dass $-\infty$ nicht ausgeschlossen werden kann. $\square$

*Beweis von Satz 5.2:* Wir sahen bereits $p_c(d+1) \le p_c(d)$, $d \ge 1$. Daher zeigen wir $p_c(d) > 0$, $d \ge 2$, und $p_c(2) < 1$.

*Teil (a)* (Beweis $p_c(d) > 0$): Wir zeigen $\Theta(p) = 0$ für $p$ hinreichend nahe bei 0:

$N(n)$ sei die Anzahl der offenen Pfade der Länge $n$ mit Start in 0, dann ist

$$\mathbb{E}_p\big(N(n)\big) = p^n\sigma(n)\,.$$

Wenn 0 einem unendlichen Cluster angehört, so existieren offene Pfade jeder Länge in 0:

$$\Theta(p) \le P_p\big(N(n) \ge 1\big) \le \mathbb{E}_p\big(N(n)\big) = p^n\sigma(n)$$

für alle $n \in \mathbb{N}$. Nun ist nach Lemma 5.5

$$\sigma(n) = \big(\lambda(d) + o(1)\big)^n \quad \text{für } n \to \infty\,,$$

also

$$\Theta(p) \le \big(p\lambda(d) + o(1)\big)^n \stackrel{n\to\infty}{\longrightarrow} 0\,,$$

falls $p\,\lambda(d) < 1$. Also ist $p_c(d) \ge 1/\lambda(d)$ mit $\lambda(d) \le 2d-1 < \infty$ (siehe Broadbent, Hammersley).

*Teil (b)* (Beweis $p_c(2) < 1$): Wir wollen $\Theta(p) > 0$ für $p$ hinreichend nahe bei 1 zeigen.

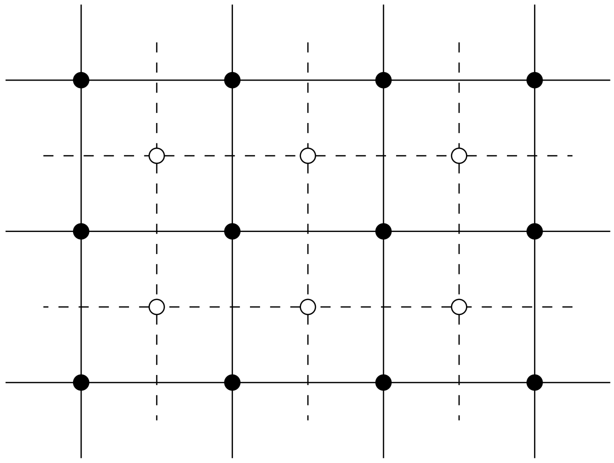


Abb. 5.2: Teil des quadratischen Gitters $\mathbb{L}^2$ zusammen mit seinem Dualen.

Nun betrachten wir das sogenannte *duale Gitter*. Die Knoten sind

$$\left\{x + \left(\frac{1}{2}, \frac{1}{2}\right) : x \in \mathbb{Z}^2\right\},$$

die Kanten sind geradlinige Verbindungen in $\mathbb{R}^2$, die zwei benachbarte Knoten verbinden (siehe Abb. 5.2).

Dies ergibt eine eins-zu-eins Korrespondenz zwischen den Kanten in $\mathbb{L}^2$ und denen des dualen Graphen. Eine Kante im dualen Gitter heißt *offen*, bzw. *geschlossen*, wenn sie eine offene bzw. geschlossene Kante in $\mathbb{L}^2$ schneidet. Dies liefert ein Kanten-Perkolationsmodell auf dem dualen Graphen mit Kanten-Wahrscheinlichkeit $p$.

Angenommen, der offene Cluster im Ursprung von $\mathbb{L}^2$ sei endlich. Dann ist der Ursprung von einer geschlossenen Kette geschlossener Kanten im dualen Graphen umgeben. Man versuche als Übung, dies rigoros zu notieren (siehe Abb. 5.3).

Es gilt sogar: $|C| < \infty$ genau dann, wenn der Ursprung in $\mathbb{L}^2$ im Inneren eines geschlossenen Kreises im Dualgraph liegt.

Sei nun $\varrho(n)$ die Anzahl der Kreise im Dualgraph mit Länge $n$, die im Inneren den Ursprung von $\mathbb{L}^2$ enthalten. Ein solcher Kreis passiert einen Knoten der Form $(k + 1/2, 1/2)$ für ein $k$ mit $0 \leq k < n$, denn er umrundet den Ursprung, also existiert so ein $(k + 1/2, 1/2)$ für ein $k \geq 0$, und er passiert nicht $(k + 1/2, 1/2)$ für ein $k \geq n$, denn sonst wäre seine Länge mindestens $2n$. Also enthält ein solcher Kreis einen selbst-meidenden Pfad der Länge $n - 1$, der in einem Knoten der Form $(k + 1/2, 1/2)$, $0 \leq k < n$, startet. Die Anzahl dieser Pfade ist maximal $n\,\sigma(n-1)$, also

$$\varrho(n) \leq n\,\sigma(n-1)\,.$$

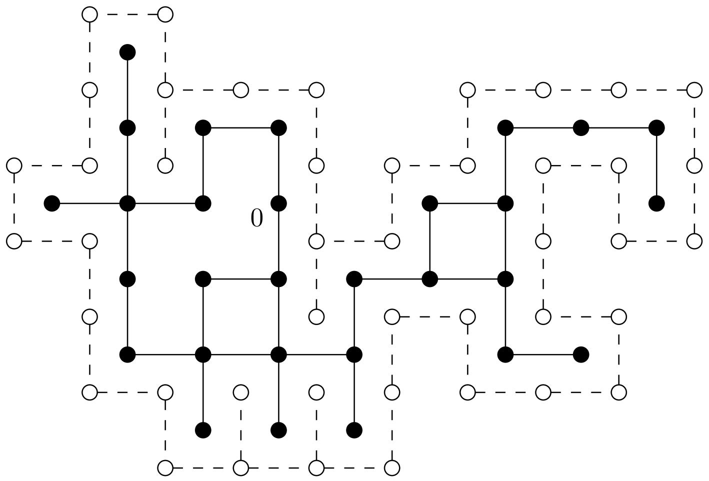



Abb. 5.3: Ein endlicher offener Cluster im Ursprung, umgeben von einer geschlossenen Kette im dualen Gitter.

Nun folgt

$$\begin{aligned}\sum_{\gamma} P_p(\gamma \text{ ist geschlossen}) &\leq \sum_{n=1}^{\infty} q^n \, n \, \sigma(n-1) \\ &= \sum_{n=1}^{\infty} q \, n \, \{q\lambda(2) + o(1)\}^{n-1} < \infty\end{aligned}$$

für $q\,\lambda(2) < 1$, wobei $q = 1-p$ ist und wir über alle Kreise $\gamma$ des Dualgraphen, die den Ursprung von $\mathbb{L}^2$ im Inneren haben, summieren.

Weiter ist

$$\sum_{\gamma} P_p(\gamma \text{ ist geschlossen}) \to 0$$

für $q = 1-p \to 0$, also existiert ein $0 < r < 1$ mit

$$\sum_{\gamma} P_p(\gamma \text{ ist geschlossen}) \leq \frac{1}{2} \quad \text{für } p > r\,,$$

Es bezeichne weiter $M(n)$ die Anzahl der geschlossenen Kreise im Dualgraph mit Länge $n$, die im Inneren den Ursprung von $\mathbb{L}^2$ enthalten. Es folgt

$$\begin{aligned}P_p(|C| = \infty) &= P_p\big(M(n) = 0 \,\forall n \in \mathbb{N}\big) \\ &= 1 - P_p\big(M(n) \geq 1 \text{ für ein } n \in \mathbb{N}\big) \\ &\geq 1 - \sum_{\gamma} P_p(\gamma \text{ ist geschlossen}) \geq \frac{1}{2}\end{aligned}$$

für $p > r$, also $p_c(2) \leq r$.

Tatsächlich kann man sogar

$$\frac{1}{\lambda(2)} \leq p_c(2) \leq 1 - \frac{1}{\lambda(2)}$$

zeigen. Der Beweis der rechten Ungleichung wird in den Übungen diskutiert. $\square$

*Beweis von Satz 5.3:* Mit dem 0-1-Gesetz von Kolmogorov folgt, dass $\psi$ nur die Werte 0 und 1 annimmt. Wenn $\Theta(p) = 0$ ist, so folgt

$$\psi(p) = \sum_{x \in \mathbb{Z}^d} P_p(|C(x)| = \infty) = 0\,.$$

Wenn $\Theta(p) > 0$, so folgt

$$\psi(p) \geq P_p(|C| = \infty) > 0\,,$$

also $\psi(p) = 1$. $\square$

**5.7 Bemerkungen**

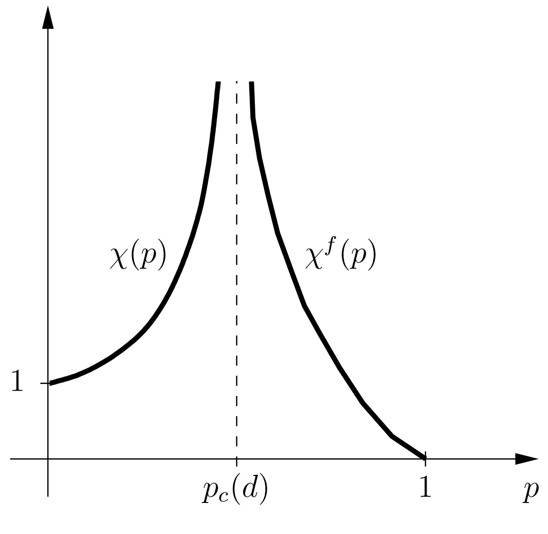
(a) Eine weitere interessante Größe ist

$$\chi(p) := \mathbb{E}_p(|C|)\,,$$

die mittlere Größe eines offenen Clusters. Es ist

$$\begin{aligned}\chi(p) &= \infty P_p(|C| = \infty) + \sum_{n \geq 1} n P_p(|C| = n) \\ &= \infty \Theta(p) + \sum_{n \geq 1} n P_p(|C| = n)\,,\end{aligned}$$

also $\chi(p) = \infty$ für $p > p_c$. Dass $\chi(p) < \infty$ ist für $p < p_c$, ist keinesfalls einfach zu zeigen. Qualitativ sieht $\chi(p)$ so aus:



Wobei $\chi^f(p) := \mathbb{E}_p(|C|; |C| < \infty)$.

(b) Für $p < p_c$ sind die offenen Cluster fast sicher endlich. Weiter gilt dort: Es existiert ein $\alpha(p)$ mit

$$P_p(|C| = n) \approx e^{-n\,\alpha(p)}$$

für $n \to \infty$, $\alpha(p) > 0$. Hierbei bezeichnet $\approx$ die Eigenschaft, dass $\log P_p(|C| = n)/(-n\,\alpha(p)) \to 1$ für $n \to \infty$. Somit hat $|C|$ endliche Momente jeder Ordnung.

(c) Für $p > p_c$ existiert ein unendlicher Cluster fast sicher: Man kann zeigen, dass dieser fast sicher eindeutig ist (keinesfalls einfach!). Wenn nun $|C| < \infty$, so gibt es $\beta_1(p)$ und $\beta_2(p)$ mit $0 < \beta_2(p) \le \beta_1(p) < \infty$, so dass

$$\exp\bigl(-\beta_1(p)n^{(d-1)/d}\bigr) \le P_p(|C| = n) \le \exp\bigl(-\beta_2(p)n^{(d-1)/d}\bigr)$$

für alle $n \in \mathbb{N}$ gilt. Man vermutet, dass

$$\delta(p) := \lim_{n\to\infty} \{-n^{-(d-1)/d} \log P_p(|C| = n)\}$$

existiert und strikt positiv ist für $p > p_c$. $n^{(d-1)/d}$ ist die Ordnung der Oberfläche der Sphäre in $\mathbb{R}^d$ mit Volumen $n$. $\delta(p)$ existiert für $d = 2$ (im Jahre 1990 bewiesen von Alexander, Chayes, Chayes) und für $d = 3$ (bewiesen von Cerf).

(d) Existiert ein unendlicher Cluster in $p = p_c$? Für $d = 2$, $d \ge 19$ ist die Antwort nein, für alle anderen Dimensionen vermutet man die gleiche Antwort. Weiter vermutet man

$$P_{p_c}(|C| = n) \approx n^{-1/\delta}$$

für $n \to \infty$ und ein $\delta = \delta(d) > 0$.

(e) Im Fall $d = 1$ sei für ein $k \in \mathbb{Z}$

$$\begin{aligned} r(k) &= \max\{n : k \leftrightarrow k + n\} \quad \text{und} \\ R_n &= \max\{r(k); 1 \le k \le n\}\,. \end{aligned}$$

Dann gilt für den maximalen Run

$$P\Bigl(R_n > \frac{(1+\varepsilon)\log n}{\log(1/p)}\Bigr) \to 0$$

für $n \to \infty$ und

$$P\Bigl(R_n < \frac{(1-\varepsilon)\log n}{\log(1/p)}\Bigr) \to 0$$

für $n \to \infty$, $\varepsilon > 0$. Dies folgt unmittelbar aus einem starken Gesetz für längste Erfolgs-Runs, welches wir aber nicht näher betrachtet haben.

KAPITEL 6

# Charakteristische Funktionen und Verteilungskonvergenz

Die Theorie charakteristischer Funktionen soll hier nur ganz kurz angerissen werden.

**6.1 Definition** Es sei $\mu$ ein W-Maß auf $(\mathbb{R}^n, \mathcal{B}^n)$. Die *charakteristische Funktion* $\hat{\mu}$ von $\mu$ ist die Abbildung von $\mathbb{R}^n$ nach $\mathbb{C}$, die durch

$$\hat{\mu}(t) = \int e^{i\langle t,\, x\rangle}\, \mu(dx) = \int \cos\bigl(\langle t,\, x\rangle\bigr)\, \mu(dx) + i \int \sin\bigl(\langle t,\, x\rangle\bigr)\, \mu(dx),\ \ t \in \mathbb{R}^n,$$

definiert ist. Hier ist $\langle t,\, x\rangle = \sum_{j=1}^n t_j x_j$. Die charakteristische Funktion eines Zufallsvektors $X$ ist die charakteristische Funktion der Verteilung von $X$, sie kann daher als $\mathbb{E}\bigl(\exp(i\,\langle t,\, X\rangle)\bigr)$ geschrieben werden.

$\hat{\mu}$ ist für alle $t \in \mathbb{R}^n$ definiert, denn Sinus und Cosinus sind beschränkt. Weiter ist $\hat{\mu}$ in $t$ stetig nach dem Satz von der dominierten Konvergenz.

**6.2 Satz (Eindeutigkeitssatz)** *Es seien $\mu, \nu$ zwei W-Maße auf $(\mathbb{R}^n, \mathcal{B}^n)$. Gilt $\hat{\mu}(t) = \hat{\nu}(t)$ für alle $t \in \mathbb{R}^n$, so gilt $\mu = \nu$.*

*Beweis:* Kompakte Mengen in $\mathbb{R}^n$ sind ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}^n$, also genügt es $\mu(K) = \nu(K)$ für alle kompakten Mengen $K$ zu zeigen (siehe Satz 2.9, Wahrscheinlichkeitstheorie). Für eine kompakte Menge $K$ und $m \in \mathbb{N}$ sei

$$f_m(x) = \begin{cases} 1 & , x \in K \\ 0 & , \text{für } d(x,\, K) := \inf\bigl\{|x-y|,\, y \in K\bigr\} \geq 1/m \\ 1 - m\, d(x,\, K) & , \text{sonst.} \end{cases}$$

Dann ist

(a) $0 \leq f_m(x) \leq 1$ für alle $x \in \mathbb{R}^n$,

(b) $f_m$ ist stetig und

(c) $f_m(x) \downarrow 1_K(x)$ für $m \to \infty$.

Falls $\int f_m\, d\mu = \int f_m\, d\nu$ für alle $m \in \mathbb{N}$, so folgt $\mu(K) = \nu(K)$ mit dem Satz von der dominierten Konvergenz aus (c). Es genügt also nachzuweisen, dass $\int f\, d\mu = \int f\, d\nu$ für alle $f$ gilt, die (a) und (b) erfüllen und einen kompakten Träger haben (der *Träger* von $f$ ist die Menge $\overline{\{x : f(x) \neq 0\}}$).

Sei $\varepsilon > 0$ und $N > 0$ so groß, dass

$$B_N := [-N,\, N]^n \supset \{x : f(x) \neq 0\}$$

und $\max\bigl\{\mu(B_N^c),\, \nu(B_N^c)\bigr\} \leq \varepsilon$ gelten. Nach dem WEIERSTRASSschen Approximationssatz gibt es eine Funktion $g : \mathbb{R}^n \to \mathbb{C}$ der Form

$$g(x) = \sum_{j=1}^{m} c_j \exp\Bigl(i\bigl\langle \frac{\pi}{N} t_j,\, x\bigr\rangle\Bigr)$$

mit $c_j \in \mathbb{C}$ und $t_j \in \mathbb{Z}^n$, die periodisch in jeder Komponente ist und $f$ in $B_N$ bis auf $\varepsilon$ approximiert:

$$\sup\bigl\{|f(x) - g(x)|;\, x \in B_N\bigr\} \leq \varepsilon\,.$$

Es folgen $\sup_{x\in\mathbb{R}^n} |g(x)| \leq 1 + \varepsilon$ und

$$\begin{aligned}\Bigl|\int f\,d\mu - \int f\,d\nu\Bigr| \leq \Bigl|\int f\,d\mu - \int g\,d\mu\Bigr| + \Bigl|\int g\,d\mu - \int g\,d\nu\Bigr|\\ + \Bigl|\int g\,d\nu - \int f\,d\nu\Bigr|\,.\end{aligned}$$

Nach Voraussetzung ist $\hat{\mu} = \hat{\nu}$, also ist der zweite Summand gleich Null. Es gilt

$$\begin{aligned}\Bigl|\int f\,d\mu - \int g\,d\mu\Bigr| &\leq \Bigl|\int_{B_N} f\,d\mu - \int_{B_N} g\,d\mu\Bigr| + \Bigl|\int_{B_N^c} f\,d\mu\Bigr| + \Bigl|\int_{B_N^c} g\,d\mu\Bigr|\\ &\leq \int_{B_N} |f - g|\,d\mu + 0 + (1+\varepsilon)\mu(B_N^c)\\ &\leq \varepsilon\mu(B_N) + (1+\varepsilon)\mu(B_N^c)\\ &\leq \varepsilon(2+\varepsilon)\,.\end{aligned}$$

Der dritte Summand wird analog abgeschätzt. Da $\varepsilon > 0$ beliebig war, folgt $\int f\,d\mu = \int g\,d\nu$. □

### 6.3 Beispiele

(a) Für $a \in \mathbb{R}^n$ betrachte das DIRAC-Maß $\delta_a$, dann ist $\hat{\delta}_a = e^{i\langle t,\, a\rangle}$, $t \in \mathbb{R}^n$, also gilt für $\mu = \sum_{i=1}^{\infty} a_i \delta_{x_i}$ mit $\sum_{i=1}^{\infty} a_i = 1$ und $x_i \in \mathbb{R}^n$:

$$\hat{\mu}(t) = \sum_{i=1}^{\infty} a_i e^{i\langle t, x_i\rangle}\,, \quad t \in \mathbb{R}^n\,.$$

Insbesondere

$$\begin{aligned}\widehat{b(n,\, p)}(t) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{ikt}\\ &= \bigl((1-p) + pe^{it}\bigr)^n\,, \quad t \in \mathbb{R}\end{aligned}$$

und

$$\hat{\pi}_\alpha(t) = e^{-\alpha} \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} e^{itn} = e^{\alpha(e^{it}-1)}\,, \quad t \in \mathbb{R}\,.$$

(b) $\mu$ sei standard-normalverteilt. Dann gilt

$$\hat{\mu}(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_{\mathbb{R}} e^{-(x-it)^2/2}\, dx\,, \quad t \in \mathbb{R}\,.$$

Es gilt $\int_{\mathbb{R}} e^{-(x-it)^2/2}\, dx = \sqrt{2\pi}$ (Übung), also $\hat{\mu}(t) = e^{-t^2/2}$.

(c) $\mu$ sei die Cauchy-Verteilung zum Parameter $\alpha > 0$. Dann gilt

$$\hat{\mu}(t) = \frac{\alpha}{\pi} \int_{\mathbb{R}} e^{itx} \frac{dx}{\alpha^2 + x^2}\,, \quad t \in \mathbb{R}\,.$$

Die Funktion $z \mapsto \frac{1}{\alpha^2+z^2}$ hat Pole in $\pm i\alpha$. Mit Hilfe des Residuensatzes ergibt sich $\hat{\mu}(t) = e^{-\alpha|t|}$ .

(d) $\mu$ sei die Standardnormalverteilung in $(\mathbb{R}^n,\, \mathcal{B}^n)$. Mit (b) folgt

$$\hat{\mu}(t) = \exp\Big(-\frac{1}{2} \sum_{j=1}^{n} t_j^2\Big) = e^{-\langle t,\, t\rangle/2}$$

für $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$.

Gemäß Definition 5.30, Wahrscheinlichkeitstheorie, ist die allgemeine Normalverteilung das Bildmaß $\nu = \mu\phi^{-1}$ der Standardnormalverteilung $\mu$ unter einer affinen Transformation $\mathbb{R}^n \ni x \mapsto \phi(x) = Ax + b$. Es gilt

$$\begin{aligned}
\hat{\nu}(t) = \int e^{i\langle t,\, x\rangle}\, \nu(dx) &= \int e^{i\langle t,\, \phi(x)\rangle}\, \mu(dx) \\
&= e^{i\langle t,\, b\rangle} \int e^{i\left\langle A^t t,\, x\right\rangle}\, \mu(dx) \\
&= e^{i\langle t,\, b\rangle} \hat{\mu}(A^t t) = e^{i\langle t,\, b\rangle} e^{-\left\langle A^t t,\, A^t t\right\rangle/2} \\
&= \exp\Big(i\, \langle t,\, b\rangle - \frac{1}{2}\, \langle t,\, \Sigma t\rangle\Big)
\end{aligned}$$

mit $\Sigma = AA^t$ als der Kovarianzmatrix von $\nu$ (siehe Satz 5.31 und Beispiel 7.28 (e), Wahrscheinlichkeitstheorie).

**6.4 Satz** *Für jedes $b \in \mathbb{R}^n$ und jede positiv semi-definite, symmetrische Matrix $\Sigma$ gibt es genau eine Normalverteilung mit $b$ als Erwartungswert und $\Sigma$ als Kovarianzmatrix.*

*Beweis:* Eindeutigkeit: Satz 6.2 und Beispiel 6.3 (d). Existenz: Es existiert mindestens eine $n \times n$-Matrix $A$ mit $AA^t = \Sigma$. $\square$

**6.5 Satz** *Es seien $X, Y$ zwei unabhängige Zufallsgrößen mit charakteristischen Funktionen $\chi_X$ bzw. $\chi_Y$. Dann ist $\chi_X \cdot \chi_Y$ die charakteristische Funktion von $X + Y$.*

*Beweis:* Für alle $t \in \mathbb{R}$ gilt

$$\mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX} e^{itY}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}) \,,$$

da $e^{itX}$ unabhängig von $e^{itY}$ ist. Aber Satz 10.21, Wahrscheinlichkeitstheorie, wurde nur für reellwertige Zufallsvariablen bewiesen. Eine Zerlegung in Real- und Imaginärteil liefert jedoch die entsprechende Aussage für komplexwertige Zufallsvariablen. □

**6.6 Beispiele**

(a) Sind $X$, $Y$ unabhängig und Cauchy-verteilt, so ist für $\lambda \in (0, 1)$

$$\lambda X + (1 - \lambda) Y$$

auch Cauchy-verteilt, denn für $t \in \mathbb{R}$ gilt

$$\begin{aligned}
\chi_{\lambda X + (1-\lambda) Y}(t) &= \mathbb{E}\Big(\exp\big(it(\lambda X + (1-\lambda)Y)\big)\Big) \\
&= \chi_X(\lambda t) \cdot \chi_Y\big((1-\lambda)t\big) \\
&= \exp(-|\lambda t|) \exp\big(-|(1-\lambda)t|\big) = e^{-|t|} \,.
\end{aligned}$$

(b) Beweis von Satz 12.9, Wahrscheinlichkeitstheorie (elegant!):

$$\begin{aligned}
\chi_{X_1+X_2}(t) &= \chi_{X_1}(t)\chi_{X_2}(t) \\
&= \exp\Big(i\mu t - \frac{\sigma^2}{2} t^2\Big) \cdot \exp\Big(i\nu t - \frac{t^2}{2} t^2\Big) \\
&= \exp\Big(i(\mu + \nu)t - \frac{1}{2}(\sigma^2 + \tau^2)t^2\Big) \,.
\end{aligned}$$ □

Sind $(X_n)_n$ unabhängig und Cauchy-verteilt, so folgt mit Beispiel 6.6 (a), dass $S_n/n$ auch Cauchy-verteilt ist. Für $c \in \mathbb{R}$ gilt

$$0 < \int_c^\infty \frac{1}{\pi} \frac{1}{1+x^2}\, dx = P\Big(\frac{S_n}{n} \geq c\Big) \leq P\Big(\sup_{k \geq n} \frac{S_k}{k} \geq c\Big) \stackrel{n\to\infty}{\longrightarrow} P\Big(\limsup_{n\to\infty} \frac{S_n}{n} \geq c\Big) \,.$$

Dann muss nach Lemma 11.3 und Kolmogorovs 0-1-Gesetz (Satz 10.9), Wahrscheinlichkeitstheorie,

$$P\Big(\limsup_{n\to\infty} \frac{S_n}{n} \geq c\Big) = 1$$

für alle $c \in \mathbb{R}$ gelten, also

$$P\Big(\limsup_{n\to\infty} \frac{S_n}{n} = \infty) = 1 \,.$$

Analog zeigt man $P(\liminf_{n\to\infty} S_n/n = -\infty) = 1$.

$S_n/n$ konvergiert nicht fast sicher, was dem Gesetz der großen Zahlen aber nicht widerspricht.

Sind $(X_n)_n$ unabhängig und standard-normalverteilt, so folgt mit Beispiel 6.6 (b), dass $S_n/\sqrt{n}$ $N(0,1)$-verteilt ist. Wie oben folgt dann, dass

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty \quad \text{und} \quad \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty$$

fast sicher, was wir mit Korollar 12.8, Wahrscheinlichkeitstheorie, auch allgemein kennen.

Wir untersuchen den Begriff der Konvergenz in Verteilung, siehe Definition 9.14, Wahrscheinlichkeitstheorie, für allgemeinere Wertebereiche. Es sei $(S, d)$ ein metrischer Raum.

**6.7 Definition**

(a) Die Borel-$\sigma$-Algebra $\mathcal{B}_S$ sei die kleinste $\sigma$-Algebra auf $S$, die die offenen Mengen enthält. ($\mathcal{B}_S$ wird auch von der Familie der abgeschlossenen Mengen erzeugt!)

(b) Mit $\mathcal{M}_1(S)$ sei die Menge der W-Maße auf $(S, \mathcal{B}_S)$ bezeichnet.

(c) $C(S)$ bezeichne die Menge der beschränkten stetigen Funktionen von $S$ nach $\mathbb{R}$.

**6.8 Bemerkung** Es seien $\mu, \nu \in \mathcal{M}_1(S)$. Gilt $\int f\, d\mu = \int f\, d\nu$ für alle $f \in C(S)$, so gilt $\mu = \nu$.

*Beweis:* Sei $F \subset C$ abgeschlossen und für $n \in \mathbb{N}$ sei $f_n \in C(S)$ definiert durch

$$f_n(x) = \max\big\{\big(1 - n\, d(x, F)\big), 0\big\} \,.$$

Dann gilt $f_n \downarrow 1_F$ für $n \to \infty$ und aus dem Satz von der dominierten Konvergenz folgt

$$\mu(F) = \lim_{n\to\infty} \int f_n\, d\mu = \lim_{n\to\infty} \int f_n\, d\nu = \nu(F) \,.$$

Da die abgeschlossenen Mengen ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_S$ bilden, folgt die Behauptung $\mu = \nu$ wie gewohnt, also mit Satz 2.9, Wahrscheinlichkeitstheorie. $\square$

Nun können wir Definition 9.14, Wahrscheinlichkeitstheorie, übernehmen:

**6.9 Definition**

(a) Es seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. $(\mu_n)_n$ *konvergiert schwach gegen* $\mu$ (in Zeichen $\mu_n \xrightarrow{w} \mu$), wenn
$$\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu \quad \forall f \in C(S)$$
gilt.

(b) Es seien $X_n$ und $X$ $(S, \mathcal{B}_S)$-wertige Zufallsvariablen für $n \in \mathbb{N}$, die auf einem W-Raum $(\Omega, \mathcal{A}, P)$ definiert sind. Falls $P^{X_n} \xrightarrow{w} P^X$ gilt, so sagt man, die Folge $(X_n)_n$ *konvergiert in Verteilung* gegen $X$ (und schreibt oft $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$).

Die Wahl dieser Definition wurde in Bemerkung 9.15, Wahrscheinlichkeitstheorie, diskutiert. Formal ist sie bequem. Oft möchte man jedoch lieber wissen, für welche $A \in \mathcal{B}_S$ gilt: $\mu(A) = \lim_{n\to\infty} \mu_n(A)$.

Es gilt:

**6.10 Satz (Portmanteau)** *Es seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. Die folgenden Aussagen sind äquivalent:*

(a) $\mu_n \xrightarrow{w} \mu$ .

(b) $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ *für jede abgeschlossene Menge* $F \subset S$ .

(c) $\liminf_{n\to\infty} \mu_n(U) \ge \mu(U)$ *für jede offene Menge* $U \subset S$.

(d) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *für alle* $A \in \mathcal{B}_S$ *mit* $\mu(\partial A) = 0$ .

*Beweis:* (a) $\Rightarrow$ (b) Sei $F$ abgeschlossen, $\varepsilon > 0$, und $f_\varepsilon(x) := \max\{0, 1 - d(x, F)/\varepsilon\}$. $f_\varepsilon$ ist beschränkt und stetig mit $1_F \le f_\varepsilon$, also
$$\limsup_{n\to\infty} \mu_n(F) \le \lim_{n\to\infty} \int f_\varepsilon \, d\mu_n = \int f_\varepsilon \, d\mu \, .$$
Mit $\varepsilon \downarrow 0$ folgt $f_\varepsilon \downarrow 1_F$. Der Satz von der dominierten Konvergenz liefert
$$\int f_\varepsilon \, d\mu \downarrow \mu(F)$$
für $\varepsilon \downarrow 0$, also $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$.

(b) $\Leftrightarrow$ (c) folgt aus der Tatsache, dass die offenen Mengen die Komplemente der abgeschlossenen sind.

((b) und (c)) $\Rightarrow$ (d): Sei $A \in \mathcal{B}_S$ mit $\mu(\partial A) = 0$. Es gilt

$$\limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A})$$

und

$$\liminf_{n\to\infty} \mu_n(A) \geq \liminf_{n\to\infty} \mu_n(\mathring{A}) \geq \mu(\mathring{A})\,.$$

Aus $\mu(\partial A) = 0$ folgt $\mu(\bar{A}) = \mu(\mathring{A}) = \mu(A)$, also folgt (d).

(d) $\Rightarrow$ (b): $F \subset S$ sei abgeschlossen. Für $\delta \geq 0$ sei

$$F^\delta := \big\{x : d(x,\, F) \leq \delta\big\}\,.$$

Dann ist $\partial(F^\delta) \subset \{x : d(x,\, F) = \delta\}$. Die Mengen $\partial(F^\delta)$ mit $\delta > 0$ sind also paarweise disjunkt. Die Menge

$$\big\{\delta > 0 : \mu\big(\partial(F^\delta)\big) > 0\big\} = \bigcup_{m=1}^{\infty} \Big\{\delta > 0 : \mu\big(\partial(F^\delta)\big) \geq \frac{1}{m}\Big\}$$

ist höchstens abzählbar. Also existiert eine fallende Nullfolge $(\delta_k)_k$ mit $\mu\big(\partial(F^{\delta_k})\big) = 0$ für alle $k \in \mathbb{N}$. Dann gilt

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F^{\delta_k}) = \mu(F^{\delta_k})$$

für alle $k$. Wegen $F^{\delta_k} \downarrow F$ haben wir $\mu(F^{\delta_k}) \downarrow \mu(F)$ für $k \to \infty$, also folgt (b).

(c) $\Rightarrow$ (a): Sei $f \geq 0$ und stetig. Dann folgt aus Satz 8.16 und dem Lemma von Fatou aus der Wahrscheinlichkeitstheorie

$$\begin{aligned}\int f\,d\mu = \int_0^\infty \mu(f > t)\,dt &\leq \int_0^\infty \liminf_{n\to\infty} \mu_n(f > t)\,dt \\ &\leq \liminf_{n\to\infty} \int_0^\infty \mu_n(f > t)\,dt = \liminf_{n\to\infty} \int f\,d\mu_n\,.\end{aligned} \tag{6.1}$$

Sei nun $f$ stetig mit $|f| \leq c < \infty$. Wende (6.1) auf $c \pm f$ an. Dies liefert $\lim_{n\to\infty} \int f\,d\mu_n = \int f\,d\mu$. □

„Portmanteau“ ist nicht der Name eines Mathematikers, sonder eine englische Vokabel für „Handkoffer“. Es soll zum Ausdruck bringen, dass man den Satz wie einen Koffer mit sich tragen sollte, wenn man in der Welt der schwachen Konvergenz spaziert. Das folgende Lemma gibt eine hinreichende (aber nicht notwendige) Bedingung für schwache Konvergenz:

**6.11 Lemma** *Seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. Es sei $\mathcal{U}$ eine durchschnittstabile Teilfamilie von $\mathcal{B}_S$, die die Eigenschaft hat, dass jede offene Teilmenge von $S$ als endliche oder abzählbare Vereinigung von Mengen aus $\mathcal{U}$ dargestellt werden kann. Gilt $\lim_{n\to\infty} \mu_n(U) = \mu(U)$ für alle $U \in \mathcal{U}$, so gilt $\mu_n \xrightarrow{w} \mu$.*

*Beweis:* Für $m \in \mathbb{N}$ und $A_1, \ldots, A_m \in \mathcal{U}$ gilt

$$\begin{aligned}\mu_n\Big(\bigcup_{j=1}^m A_j\Big) &= \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<\cdots<j_k} \mu_n(A_{j_1} \cap \cdots \cap A_{j_k}) \\ &\underset{n\to\infty}{\longrightarrow} \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<\cdots<j_k} \mu(A_{j_1} \cap \cdots \cap A_{j_k}) = \mu\Big(\bigcup_{j=1}^m A_j\Big).\end{aligned}$$

Sei $G \subset S$ offen. Dann gilt $G = \bigcup_i A_i$ für eine Folge in $\mathcal{U}$. Dann existieren für jedes $\varepsilon > 0$ ein $m \in \mathbb{N}$ und $A_1, \ldots, A_m \in \mathcal{U}$ mit $\mu(G) \le \mu\big(\bigcup_{j=1}^m A_j\big) + \varepsilon$. Also gilt

$$\liminf_{n\to\infty} \mu_n(G) \ge \liminf_{n\to\infty} \mu_n\Big(\bigcup_{j=1}^m A_j\Big) = \mu\Big(\bigcup_{j=1}^m A_j\Big) \ge \mu(G) - \varepsilon\,.$$

Da $\varepsilon > 0$ beliebig war, folgt die Aussage (c) von Satz 6.10. □

Wir wollen nun das Verhalten induzierter W-Maße untersuchen, wenn die Ursprungsmaße schwach konvergieren. Ist $h$ eine messbare Abbildung auf $(S, \mathcal{B}_S)$ in einen zweiten metrischen Raum, so braucht aus $\mu_n \xrightarrow{w} \mu$ nicht $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$ zu folgen:

**6.12 Beispiel** Sei $(x_n)_n$ eine Folge in $S\backslash\{x\}$, die gegen $x \in S$ konvergiert. Dann gilt $\delta_{x_n} \xrightarrow{w} \delta_x$, siehe 9.15 (b), Wahrscheinlichkeitstheorie. Ist $h : S \to \mathbb{R}$ durch $h(y) = 1_{\{x\}}(y)$ definiert, so gelten

$$\delta_{x_n} h^{-1} = \delta_0 \quad \text{und} \quad \delta_x h^{-1} = \delta_1\,,$$

also konvergiert $\delta_{x_n} h^{-1}$ nicht schwach gegen $\delta_x h^{-1}$.

**6.13 Lemma** *Seien $(S, d)$ und $(S', d')$ zwei metrische Räume, und $h : S \to S'$ sei stetig. Es seien $\mu_n$ und $\mu$ W-Maße auf $(S, \mathcal{B}_S)$, $n \in \mathbb{N}$, mit $\mu_n \xrightarrow{w} \mu$. Dann gilt $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$ auf $(S', \mathcal{B}_{S'})$.*

*Beweis:* Ist $f \in C(S')$, so ist $f \circ h \in C(S)$, also

$$\begin{aligned}\lim_{n\to\infty} \int f\, d(\mu_n\, h^{-1}) &= \lim_{n\to\infty} \int (f \circ h)\, d\mu_n \\ &= \int (f \circ h)\, d\mu = \int f\, d(\mu\, h^{-1})\,.\end{aligned}$$ □

Die Forderung nach Stetigkeit ist stark einschränkend. Wir wollen sie abschwächen und bereiten dies vor:

**6.14 Lemma** *Es seien $(S, d)$ und $(S', d')$ zwei metrische Räume, und $h : S \to S'$ sei $\mathcal{B}_S/\mathcal{B}_{S'}$-messbar. Dann ist*

$$D_h := \{x \in S : h \textit{ ist nicht stetig in x}\} \in \mathcal{B}_S\,.$$

*Beweis:* Für $m, n \in \mathbb{N}$ sei

$$A_{m,n} = \Big\{x \in S : \text{es gibt } y, z \in S \text{ mit } d(x,\, y) < \frac{1}{m} \\ \text{und } d(x,\, z) < \frac{1}{m} \text{ sowie } d'\big(h(y),\, h(z)\big) \geq \frac{1}{n}\Big\}$$

Die Menge $A_{m,n}$ ist offen. Daraus folgt

$$D_h = \bigcup_n \bigcap_m A_{m,n} \in \mathcal{B}_S\,.$$ □

**6.15 Satz** *Es liege dieselbe Situation vor wie in Lemma 6.13, h sei jedoch nur als $\mathcal{B}_S/\mathcal{B}_{S'}$-messbar vorausgesetzt. Gilt $\mu(D_h) = 0$, so folgt $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$.*

*Beweis:* Sei $F \subset S'$ abgeschlossen. Dann gilt

$$\limsup_{n\to\infty} \mu_n\big(h^{-1}(F)\big) \leq \limsup_{n\to\infty} \mu_n\left(\overline{h^{-1}(F)}\right) \leq \mu\left(\overline{h^{-1}(F)}\right)\,.$$

Es ist $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$. Wegen $\mu(D_h) = 0$ folgt

$$\mu\left(\overline{h^{-1}(F)}\right) = \mu\big(h^{-1}(F)\big)$$

und aus Kriterium (b) von Satz 6.10 folgt die Behauptung. □

Häufig gibt es in Bezug auf schwache Konvergenz verhältnismäßig große kompakte bzw. relativ kompakte Mengen in $\mathcal{M}_1(S)$.

**6.16 Definition**

(a) Eine Teilmenge $\Gamma \subset \mathcal{M}_1(S)$ heißt *relativ kompakt*, wenn jede Folge $(\mu_n)_n$ in $\Gamma$ eine schwach konvergente Teilfolge hat. (Der Grenzwert muss nicht in $\Gamma$ liegen.)

(b) Eine Teilmenge $\Gamma \subset \mathcal{M}_1(S)$ heißt *straff*, wenn für jedes $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subset S$ existiert, so dass $\mu(K_\varepsilon) \geq 1 - \varepsilon$ für jedes $\mu \in \Gamma$.

**6.17 Bemerkungen**

(a) Ist $S$ kompakt, so ist offenbar $\mathcal{M}_1(S)$ straff.

(b) $\mathcal{M}_1(\mathbb{R})$ ist nicht straff, weil schon $\{\delta_x : x \in \mathbb{R}\}$ nicht straff ist.

(c) Ein einzelnes W-Maß $\mu \in \mathcal{M}_1(S)$ heißt straff, wenn $\{\mu\}$ straff ist, das heißt, wenn für jedes $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon$ existiert mit $\mu(K_\varepsilon) \geq 1-\varepsilon$. Ist $S$ $\sigma$-kompakt, das heißt, existiert eine Folge $(K_n)_n$ von kompakten Mengen in $S$ mit $K_n \uparrow S$, so ist jedes W-Maß straff, denn $\mu(K_n) \uparrow \mu(S) = 1$. Dies ist für $S = \mathbb{R}$ oder $S = \mathbb{R}^d$ der Fall.

Es gibt eine erstaunlich große Klasse von metrischen Räumen, die nicht unbedingt $\sigma$-kompakt sind und in denen jedes W-Maß straff ist, nämlich vollständig separabel[1]. Diese Klasse umfasst separable Hilbert- und Banachräume, wie etwa den Folgenraum $\ell_2$ oder den Raum $C[0, 1]$ der stetigen Funktionen $[0, 1] \to \mathbb{R}$, versehen mit der Supremumsmetrik. Unendlichdimensionale Banachräume sind nie $\sigma$-kompakt.

Die Aussage, dass jedes W-Maß auf einem vollständigen, separablen metrischen Raum straff ist, ist ein Spezialfall des Satzes von Prohorov, der im Anhang A bewiesen wird:

**6.18 Satz (von Prohorov, 1956)** *Es sei $S$ vollständig und separabel. Dann ist jede Teilmenge von $\mathcal{M}_1(S)$ genau dann relativ kompakt, wenn sie straff ist.*

Wir wollen diesen Satz anwenden. Dazu eine Vorbereitung:

**6.19 Lemma** *Seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. Dann gilt $\mu_n \xrightarrow{w} \mu$ genau dann, wenn jede Teilfolge $(\mu'_n)_n$ von $(\mu_n)_n$ ihrerseits eine Teilfolge $(\mu''_n)_n$ hat mit $\mu''_n \xrightarrow{w} \mu$.*

*Beweis:* Folgt aus Definition 6.9 und der Tatsache, dass reelle Zahlenfolgen sich so verhalten. □

Wir leiten ein sehr nützliches Kriterium für schwache Konvergenz auf $\mathbb{R}^d$ her. Für $x \in \mathbb{R}^d$ sei $\pi_x : \mathbb{R}^d \to \mathbb{R}$ definiert durch $\pi_x(y) := \langle x, y\rangle$, Euklids Skalarprodukt.

**6.20 Satz (von Cramér-Wold, 1936)** *Es seien $\mu_n$ und $\mu$ W-Maße auf $(\mathbb{R}^d, \mathcal{B}^d)$ für $n \in \mathbb{N}$. Dann gilt $\mu_n \xrightarrow{w} \mu$ genau dann, wenn*

$$\mu_n \pi_x^{-1} \xrightarrow{w} \mu \pi_x^{-1} \text{ in } (\mathbb{R}, \mathcal{B})$$

*für alle $x \in \mathbb{R}^d$ gilt.*

[1] *Separabel*: $S$ enthält eine abzählbare, dichte Teilmenge

*Beweis:* Da $\pi_x$ stetig ist, folgt aus $\mu_n \xrightarrow{w} \mu$ und Lemma 6.13 die Behauptung $\mu_n \pi_x^{-1} \xrightarrow{w} \mu \pi_x^{-1}$. Zum Beweis der Umkehrung betrachten wir zunächst die Projektionen $\pi_i := \pi_{e_i}$, $1 \leq i \leq d$, auf die $d$ Einheitsvektoren $e_i \in \mathbb{R}^d$. $\mu_n \pi_i^{-1}$ konvergiert schwach, also ist $\{\mu_n \pi_i^{-1}, n \in \mathbb{N}\}$ relativ kompakt. Somit existiert für jedes $\varepsilon > 0$ eine kompakte Menge $K_i \subset \mathbb{R}$ mit

$$\mu_n\big(\pi_i^{-1}(K_i)\big) \geq 1 - \frac{\varepsilon}{d}$$

für alle $n \in \mathbb{N}$ und alle $i \in \{1, \dots, d\}$. Die Menge

$$K := \bigcap_{i=1}^{d} \pi_i^{-1}(K_i) \subset \mathbb{R}^d$$

ist abgeschlossen und beschränkt in $\mathbb{R}^d$, also kompakt. Für alle $n \in \mathbb{N}$ gilt:

$$\mu_n(K^c) = \mu_n\Big(\bigcup_{i=1}^{d} \pi_i^{-1}(K_i)^c\Big) \leq \sum_{i=1}^{d} \mu_n\big(\pi_i^{-1}(K_i^c)\big) \leq \varepsilon\,.$$

Aus Satz 6.18 folgt, dass $\{\mu_n : n \in \mathbb{N}\}$ relativ kompakt ist. Sei $(\mu'_n)_n$ eine beliebige Teilfolge von $(\mu_n)_n$. Diese hat eine konvergente Teilfolge $(\mu''_n)_n$ mit $\mu''_n \xrightarrow{w} \mu''$ für ein $\mu'' \in \mathcal{M}_1(\mathbb{R}^d)$. Für alle $x \in \mathbb{R}^d$ folgt dann

$$\mu''_n \pi_x^{-1} \xrightarrow{w} \mu'' \pi_x^{-1}\,.$$

Wegen $\mu_n \pi_x^{-1} \xrightarrow{w} \mu \pi_x^{-1}$ folgt $\mu \pi_x^{-1} = \mu'' \pi_x^{-1}$ für alle $x \in \mathbb{R}^d$. Damit stimmen auch die charakteristischen Funktionen von $\mu \pi_x^{-1}$ und $\mu'' \pi_x^{-1}$ überein, insbesondere im Punkt 1. Somit gilt

$$\hat{\mu}(x) = \int e^{i\langle x, y\rangle}\, \mu(dy) = \int e^{it}\, (\mu \pi_x^{-1})(dt) = \int e^{it} (\mu'' \pi_x^{-1})(dt) = \widehat{\mu''}(x)\,.$$

Aus Satz 6.2 folgt $\mu = \mu''$ und Lemma 6.19 führt zur Behauptung. $\square$

Mit Satz 6.20 erhalten wir recht leicht eine mehrdimensionale Version des zentralen Grenzwertsatzes, $|\cdot|$ bezeichne die Euklidische Norm im $\mathbb{R}^d$.

**6.21 Satz** *Es sei $(X_n)_n$ eine Folge unabhängiger, identisch verteilter $d$-dimensionaler Zufallsvektoren. Es gelte $\mathbb{E}|X_i|^2 < \infty$. Seien $a := \mathbb{E}X_1$ und $\Sigma$ die Kovarianzmatrix von $X_1$. Dann gilt*

$$P^{\sum_{i=1}^n (X_i - a)/\sqrt{n}} \xrightarrow{w} \mu\,,$$

*wobei $\mu$ die $d$-dimensionale Normalverteilung mit Mittelwert 0 und Kovarianzmatrix $\Sigma$ ist.*

*Beweis:* Sei $T_n := \sum_{i=1}^n (X_i - a)/\sqrt{n}$. Nach Satz 6.20 genügt es zu zeigen, dass für jedes $x \in \mathbb{R}^d$ gilt

$$P^{\langle x,\, T_n\rangle} \xrightarrow{w} \mu\pi_x^{-1}\,.$$

Es ist

$$\langle x, T_n\rangle = \sum_{i=1}^n \big(\langle x,\, X_i\rangle - \langle x,\, a\rangle\big)/\sqrt{n}\,.$$

Die $\langle x,\, X_i\rangle$, $i \in \mathbb{N}$, sind unabhängige, identisch verteilte, eindimensionale Zufallsgrößen mit Erwartungswert $\langle x,\, a\rangle$ und Varianz $\sigma_x^2 = \mathbb{E}(\langle x,\, X_i - a\rangle^2) = x^t\Sigma x$, wenn $x$ als Spaltenvektor geschrieben wird.

Ist $\sigma_x^2 > 0$, so konvergiert $P^{\langle x,\, T_n/\sigma_x\rangle}$ nach Korollar 12.7, Wahrscheinlichkeitstheorie, gegen die Standard-Normalverteilung, also konvergiert $P^{\langle x,\, T_n\rangle}$ nach Lemma 6.13 mit $h : \mathbb{R} \to \mathbb{R}$, $h(y) = \sigma_x y$, gegen die Normalverteilung mit Mittelwert 0 und Varianz $\sigma_x^2$.

Gilt $\sigma_x^2 = 0$, so ist $\langle x,\, T_n\rangle = 0$ fast sicher (siehe Satz 7.16, Wahrscheinlichkeitstheorie), und somit gilt $P^{\langle x,\, T_n\rangle} \xrightarrow{w} \delta_0$ (Normalverteilung mit Erwartungswert 0 und Varianz 0).

Nun ist aber $\mu\pi_x^{-1}$ die Normalverteilung mit Erwartungswert 0 und Varianz $\sigma_x^2$. Damit ist der Satz bewiesen. $\square$

Abschließend soll definiert werden, was es heißt, dass eine Folge $(X_n)_n$ von $(S,\, \mathcal{B}_S)$-wertigen Zufallsvariablen „in Wahrscheinlichkeit" gegen eine Zufallsvariable $X$ konvergiert. Es liegt nahe, $P\big(d(X_n,\, X) \geq \varepsilon\big)$ zu verwenden. Aber $d(X_n,\, X)$ ist nicht immer eine Zufallsgröße! Man muss voraussetzen, dass $S$ separabel ist. Ist $(S,\, d)$ ein metrischer Raum, so betrachten wir $(S \times S,\, d')$ mit

$$d'\big((x_1,\, x_2),\, (y_1,\, y_2)\big) := \big(d(x_1,\, y_1)^2 + d(x_2,\, y_2)^2\big)^{1/2}\,.$$

**6.22 Lemma** *Ist $S$ separabel, so ist $\mathcal{B}_{S\times S} = \mathcal{B}_S \otimes \mathcal{B}_S$.*

*Beweis:* Jede Produktmenge $A \times B$ mit $A \subset S$ und $B \subset S$ offen, ist offen in $S \times S$, liegt also in $\mathcal{B}_{S\times S}$. Da diese Mengen $\mathcal{B}_S \otimes \mathcal{B}_S$ erzeugen, folgt $\mathcal{B}_S \otimes \mathcal{B}_S \subset \mathcal{B}_{S\times S}$. Hier wurde die Separabilität nicht benutzt.

Ist $S$ separabel, so existiert eine abzählbare Basis $\{U_i,\, i \in \mathbb{N}\}$ der Topologie von $S$ und $\{U_i \times U_j,\, i, j \in \mathbb{N}\}$ ist dann eine abzählbare Basis der Topologie von $S \times S$. Somit ist jede offene Teilmenge von $S \times S$ in $\mathcal{B}_S \otimes \mathcal{B}_S$ enthalten, also ist $\mathcal{B}_{S\times S} \subset \mathcal{B}_S \otimes \mathcal{B}_S$. $\square$

Sind nun $X$ und $Y$ zwei $(S,\, \mathcal{B}_S)$-wertige Zufallsvariablen, definiert auf einem gemeinsamen W-Raum $(\Omega,\, \mathcal{A},\, P)$, so ist $(X,\, Y)$ eine $(S \times S,\, \mathcal{B}_S \otimes \mathcal{B}_S)$-wertige Zufallsvariable. $d : S \times S \to \mathbb{R}$ ist eine $d'$-stetige Abbildung, also $\mathcal{B}_{S\times S}/\mathcal{B}$-messbar, also ist $d(X,\, Y)$ mit Hilfe von Lemma 6.22 eine Zufallsgröße:

**6.23 Definition** Es sei $S$ separabel und $X$ sowie $X_n$ für $n \in \mathbb{N}$ Zufallsvariablen mit Werten in $(S, \mathcal{B}_S)$. Die Folge $(X_n)_n$ *konvergiert in Wahrscheinlichkeit* gegen $X$, falls $d(X_n, X)$ in Wahrscheinlichkeit gegen 0 konvergiert, also für alle $\varepsilon > 0$

$$\lim_{n\to\infty} P\big(d(X_n, X) \geq \varepsilon\big) = 0$$

gilt.

**6.24 Satz** *Konvergiert $(X_n)_n$ in W-keit gegen $X$, so gilt $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$.* [2]

*Beweis:* Sei $A \in \mathcal{B}_S$ mit $P(X \in \partial A) = 0$. Für $\varepsilon > 0$ gilt

$$P(X_n \in A, X \notin A) \leq P\big(d(X_n, X) \geq \varepsilon\big) + P\big(d(X, A) < \varepsilon, X \notin A\big).$$

Dies und dasselbe mit $A^c$ ergibt auf Grund $\lim_{n\to\infty} P\big(d(X_n, X) \geq \varepsilon\big) = 0$ die Abschätzung

$$\limsup_{n\to\infty} P(\{X_n \in A\}\Delta\{X \in A\}) \leq P\big(d(X, A) < \varepsilon, X \notin A\big) + P\big(d(X, A^c) < \varepsilon, X \in A\big).$$

Dabei sei an die Definition $A\Delta B := (A \setminus B) \cup (B \setminus A)$ erinnert. Für $\varepsilon \downarrow 0$ gilt

$$\{d(X, A) < \varepsilon, X \notin A\} \downarrow \{X \in \partial A \cap A^c\}$$

und

$$\{d(X, A^c) < \varepsilon, X \in A\} \downarrow \{X \in A \cap \partial A\}.$$

Wegen $P(X \in \partial A) = 0$ folgt

$$\lim_{n\to\infty} P(\{X_n \in A\}\Delta\{X \in A\}) = 0,$$

also

$$\lim_{n\to\infty} P(X_n \in A) = P(X \in A).$$

Nach Satz 6.10 (d) folgt $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$. □

**6.25 Lemma** *Es sei $S$ separabel, und $(X_n)_n$ und $(X_n')_n$ seien zwei Folgen von $(S, \mathcal{B}_S)$-wertigen Zufallsvariablen. Gelten $\mathcal{L}(X_n) \xrightarrow{w} \mu$ und $d(X_n, X_n') \to 0$ in Wahrscheinlichkeit, so gilt $\mathcal{L}(X_n') \xrightarrow{w} \mu$.*

*Beweis:* Seien $F \subset S$ abgeschlossen, $\varepsilon > 0$ und $F^\varepsilon = \{x : d(x, F) \leq \varepsilon\}$. Dann gilt

$$\begin{aligned}\limsup_{n\to\infty} P(X_n' \in F) &\leq \limsup_{n\to\infty} P\big(d(X_n, X_n') \geq \varepsilon\big) + \limsup_{n\to\infty} P(X_n \in F^\varepsilon)\\ &\leq \mu(F^\varepsilon).\end{aligned}$$

Für $\varepsilon \to 0$ gilt $\mu(F^\varepsilon) \to \mu(F)$. □

---
[2] Vgl. Satz 9.16, Wahrscheinlichkeitstheorie

KAPITEL 7

# Der Satz von Donsker

Fortan betrachten wir $S = C[0,1]$ die Menge der stetigen Funktionen $[0,1] \to \mathbb{R}$. Für $f, g \in C[0,1]$ sei $d(f,g) := \sup_{t\in[0,1]} |f(t) - g(t)|$.

**7.1 Satz** $(C[0,1], d)$ *ist ein vollständiger und separabler metrischer Raum.*

*Beweis:* Die Banachraum-Eigenschaft kennen wir aus der Analysis-Vorlesung. Die Polynome mit rationalen Koeffizienten bilden eine abzählbare, dichte Teilmenge in $C[0,1]$ nach dem Approximationssatz von Weierstrass (der dies für jede kompakte Teilmenge in $\mathbb{R}^k$ liefern würde). $\square$

Für $m \in \mathbb{N}$ und $0 \le t_1 < t_2 < \cdots < t_m \le 1$ sei $\pi_{t_1,\dots,t_m} : C[0,1] \to \mathbb{R}^m$ die *endlichdimensionale Projektion* $f \mapsto (f(t_1), \dots, f(t_m))$.

**7.2 Lemma** *Es gilt*

$$\mathcal{B}_C = \sigma(\pi_t^{-1}(\mathcal{B}), t \in [0,1])$$

*mit der Notation* $C := C[0,1]$.

*Beweis:* Mit $\mathcal{B}' := \sigma(\pi_t^{-1}(\mathcal{B}), t \in [0,1])$ wollen wir $\mathcal{B}_C = \mathcal{B}'$ zeigen. Da $\pi_t$ stetig ist, ist für $U \subset \mathbb{R}$ offen auch $\pi_t^{-1}(U)$ offen, liegt also in $\mathcal{B}_C$. Daraus folgt $\mathcal{B}' \subset \mathcal{B}_C$. Für $f \in C[0,1]$ und $\varepsilon > 0$ sei

$$\mathcal{B}_\varepsilon(f) := \{g \in C[0,1] : d(f,g) \le \varepsilon\} .$$

Dann ist, da $f$ stetig,

$$\mathcal{B}_\varepsilon(f) = \bigcap_{t\in[0,1]\cap\mathbb{Q}} \{g \in C[0,1] : |g(t) - f(t)| \le \varepsilon\} = \bigcap_{t\in[0,1]\cap\mathbb{Q}} \pi_t^{-1}\big(\mathcal{B}_\varepsilon(f(t))\big) \in \mathcal{B}' .$$

Da $C[0,1]$ separabel ist, ist jede offene Menge abzählbare Vereinigung von derartigen Kugeln, also in $\mathcal{B}'$. $\square$

Der Satz von Donsker ist eine Verallgemeinerung des zentralen Grenzwertsatzes, indem nicht nur die Asymptotik der Verteilung von $S_n/\sqrt{n}$ ($S_n = \sum_{i=1}^n X_i$, unabhängig, identisch verteilte $X_i$) untersucht wird, sondern die Verteilung des „gesamten Pfades“.

Es sei $(X_n)_n$ eine Folge unabhängiger, identisch verteilter, $\mathbb{R}$-wertiger Zufallsvariablen, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Es sei

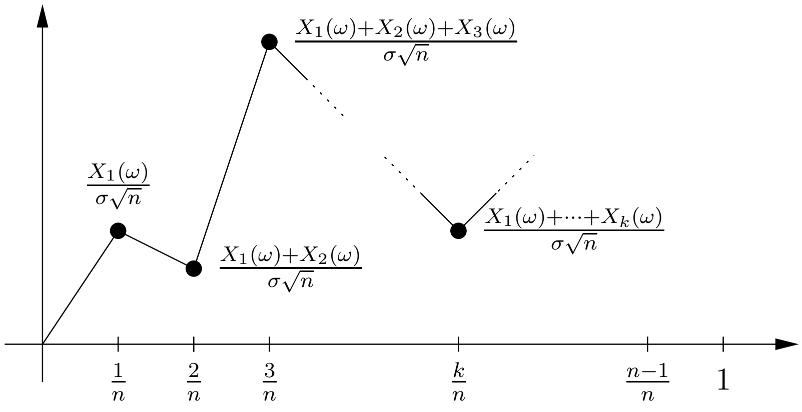



Abb. 7.1: Broken-Line-Prozess.

$\mathbb{E}X_i = 0$ (falls nicht, ersetzen wir $X_i$ durch $X_i - \mathbb{E}X_i$) und $\sigma^2 := \mathbb{E}X_i^2 \in (0,\infty)$. Wir setzen $S_0 = 0$ und $S_n = \sum_{i=1}^n X_i$, $n \in \mathbb{N}$. Für $n \in \mathbb{N}$ und $\omega \in \Omega$ definieren wir die Abbildung $Y_n(\omega,\cdot) : [0,1] \to \mathbb{R}$ durch

$$Y_n\big(\omega, \frac{k}{n}\big) := \frac{S_k(\omega)}{\sigma\sqrt{n}} \ , \ k = 0,1,\dots,n$$

und lineare Interpolation („Broken-Line-Prozess“).

(Bild einer Irrfahrt: Teilchen schneller springen lassen und die Sprünge immer kleiner werden lassen, „Zeit- und Ortsskala simultan ändern“.)

Eine äquivalente Definition ist

$$Y_n(\omega,t) = \frac{S_{[nt]}(\omega)}{\sigma\sqrt{n}} + \frac{nt-[nt]}{\sigma\sqrt{n}} X_{[nt]+1}(\omega) \ , 0 \le t \le 1 \ . \tag{7.1}$$

($[x]$ sei der ganzzahlige Anteil der reellen Zahl $x$).

$Y_n$ kann als Abbildung von $\Omega$ nach $C[0,1]$ aufgefasst werden. Für ein festes $t \in [0,1]$ ist $Y_n(\cdot,t)$ offenbar $\mathcal{A}$-$\mathcal{B}$-messbar. Nach Lemma 7.2 folgt, dass $Y_n : \Omega \to C[0,1]$ eine $(C,\mathcal{B}_C)$-wertige Zufallsvariable ist. Ohne Einschränkung ist $\sigma^2 = 1$, sonst ersetze $X_i$ durch $X_i/\sigma$. Der Satz von Donsker wird liefern, dass $\mathcal{L}(Y_n)$ schwach gegen ein W-Maß auf $(C,\mathcal{B}_C)$ konvergiert. Da $\pi_{t_1,\dots,t_m}$ stetig ist, ist nach Lemma 6.13 für die Konvergenz von

$$\mu_n := P^{Y_n}$$

notwendig, dass $\mu_n\pi_{t_1,\dots,t_m}^{-1}$ für $n \to \infty$ auf $(\mathbb{R}^m,\mathcal{B}^m)$ konvergiert.

**7.3 Satz** *Für jedes $m \in \mathbb{N}$ und $0 \le t_1 < t_2 < \dots < t_m \le 1$ konvergiert $\mu_n\pi_{t_1,\dots,t_m}^{-1}$ schwach auf $(\mathbb{R}^m,\mathcal{B}^m)$ gegen die $m$-dimensionale Normalverteilung mit Erwartungswert 0 und Kovarianzmatrix $(\min\{t_i,t_j\})_{i,j}$.*

$m = 1$, $t_1 = 1$: $\mu_n \pi_1^{-1} = \mathcal{L}(Y_n(1)) = \mathcal{L}(\frac{S_n}{\sqrt{n}})$ konvergiert gegen die Standard-Normalverteilung. Für $m = 1$ und $t_1 = 0$ ist $\mu_n \pi_0^{-1} = \mathcal{L}(Y_n(0)) = \delta_0$.

Wir müssen noch etwas vorbereiten:

**7.4 Lemma** *Sei $d \in \mathbb{N}$ und für $j = 1, \ldots, d$ sei $(\mu_n^{(j)})_n$ eine Folge von W-Maßen auf $(\mathbb{R}, \mathcal{B})$ mit $\mu_n^{(j)} \xrightarrow{w} \mu^{(j)} \in \mathcal{M}_1(\mathbb{R})$. Dann gilt*

$$\mu_n^{(1)} \otimes \cdots \otimes \mu_n^{(d)} \xrightarrow{w} \mu^{(1)} \otimes \cdots \otimes \mu^{(d)}$$

*auf $(\mathbb{R}^d, \mathcal{B}^d)$.*

*Beweis:* Es sei $A_j := \{x \in \mathbb{R} : \mu^{(j)}(\{x\}) = 0\}$. $A_j^c$ ist abzählbar und somit ist $A_j$ dicht. Sei $B_j \subset A_j$ eine abzählbare dichte Teilmenge von $A_j$. Dann ist $\{(a_j, b_j) : a_j, b_j \in B_j\}$ eine abzählbare Basis der Topologie von $\mathbb{R}$, also ist

$$\mathcal{U} := \{(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d) : a_j, b_j \in B_j \text{ für } j = 1, \ldots, d\}$$

eine Basis der Topologie von $\mathbb{R}^d$. $\mathcal{U}$ ist durchschnittstabil und für $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d) \in \mathcal{U}$ gilt wegen Satz 6.10 (d).

$$\mu_n^{(1)} \otimes \cdots \otimes \mu_n^{(d)}((a_1, b_1) \times \cdots \times (a_d, b_d)) = \prod_{j=1}^{d} \mu_n^{(j)}((a_j, b_j))$$

$$\xrightarrow{n\to\infty} \prod_{j=1}^{d} \mu^{(j)}((a_j, b_j)) = \mu^{(1)} \otimes \cdots \otimes \mu^{(d)}((a_1, b_1) \times \cdots \times (a_d, b_d)) \,.$$

Das Lemma folgt nun aus Lemma 6.11. $\square$

*Beweis von Satz 7.3:* Wir können annehmen, dass $t_1 > 0$ gilt. Setze $\sum_{i=1}^{0} := 0$ und

$$Z_1^{(n)} := \sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}},\ Z_2^{(n)} := \sum_{i=[nt_1]+1}^{[nt_2]} \frac{X_i}{\sqrt{n}},\ \ldots,\ Z_m^{(n)} := \sum_{i=[nt_{m-1}]+1}^{[nt_m]} \frac{X_i}{\sqrt{n}} \,.$$

$Z_1^{(n)}, \ldots, Z_m^{(n)}$ sind für jedes $n \in \mathbb{N}$ unabhängig. Mit Lemma 7.4 untersuchen wir das Konvergenzverhalten von $(Z_j^{(n)})_n$ für festes $j$:

$\mathcal{L}(Z_j^{(n)}) = \mathcal{L}(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}})$, wo wir $t_0 := 0$ und $k(n) := [nt_j] - [nt_{j-1}]$ setzen. Der zentrale Grenzwertsatz liefert:

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \leq s\Big) = \Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-x^2/2}\, dx \,.$$

Nun gilt $\lim_{n\to\infty} \frac{k(n)}{n} = t_j - t_{j-1}$. Für $\varepsilon > 0$ und $s \in \mathbb{R}$ folgt

$$\begin{aligned}
\limsup_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &\le \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le \frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big) \\
&= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big) \quad \text{und} \\
\liminf_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &\ge \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le \frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big) \\
&= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big)\ , \\
\text{also } \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}}\Big)
\end{aligned}$$

Dies ist die Verteilungsfunktion der eindimensionalen Normalverteilung mit Erwartungswert 0 und Varianz $t_j - t_{j-1}$. Nach Lemma 7.4 folgt, dass $\mathcal{L}(Z_1^{(n)}, \dots, Z_m^{(n)})$ für $n \to \infty$ gegen die Produktverteilung konvergiert, und dies ist die $m$-dimensionale Normalverteilung $\nu$ mit Erwartungswert 0 und Kovarianzmatrix $(\delta_{ij}(t_j - t_{j-1}))_{i,j}$.

Sei nun $f : \mathbb{R}^m \to \mathbb{R}^m$ durch $f(x_1, \dots, x_m) := (x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + \dots + x_m)$ definiert. Nach Lemma 6.13 konvergiert die Verteilung von

$$f(Z_1^{(n)}, \dots, Z_m^{(n)}) = \left(\sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}}, \sum_{i=1}^{[nt_2]} \frac{X_i}{\sqrt{n}}, \dots, \sum_{i=1}^{[nt_m]} \frac{X_i}{\sqrt{n}}\right)$$

gegen $\nu f^{-1}$. Sei $(U_1, \dots, U_m)$ eine Zufallsgröße mit Verteilung $\nu$, dann besitzt die Normalverteilung $\nu f^{-1}$ den Erwartungswert 0 und die Kovarianzmatrix mit Komponenten

$$\begin{aligned}
\mathbb{E}(\sum_{k=1}^{i} U_k \sum_{s=1}^{j} U_s) &= \sum_{k=1}^{\min\{i,j\}} \mathbb{E}(U_k^2) + \sum_{s=1}^{j} \sum_{k=1, k\neq s}^{i} \mathbb{E}(U_k U_s) = \sum_{k=1}^{\min\{i,j\}} (t_k - t_{k-1}) \\
&= \min\{t_i, t_j\}\ .
\end{aligned}$$

Sei nun $W_j^{(n)} := \sum_{i=1}^{[nt_j]} \frac{X_i}{\sqrt{n}} - Y_n(t_j)$. Dann gilt $|W_j^{(n)}| \le \frac{|X_{[nt_j]+1}|}{\sqrt{n}}$, falls $t_j < 1$ und $W_j^{(n)} = 0$ sonst. Damit ist für $\varepsilon > 0$

$$\begin{aligned}
&P(|(W_1^{(n)}, \dots, W_m^{(n)})| \ge \varepsilon) \le P\Big(\bigcup_{j=1}^{m} \{|W_j^{(n)}| \ge \varepsilon/m\}\Big) \\
&\le \sum_{j=1}^{m} P\Big(|X_{[nt_j]+1}| \ge \sqrt{n}\varepsilon/m\Big) = mP(|X_1| \ge \sqrt{n}\varepsilon/m) \longrightarrow 0
\end{aligned}$$

für $n \to \infty$, also konvergiert $(W_1^{(n)}, \dots, W_m^{(n)})$ in Wahrscheinlichkeit gegen 0. Nach Lemma 6.25 konvergiert dann auch $\mathcal{L}(Y_n(t_1), \dots, Y_n(t_m))$ gegen $\nu f^{-1}$. $\square$

Konvergiert, wie behauptet wird, $\mu_n = \mathcal{L}(Y_n)$ gegen ein Wahrscheinlichkeitsmaß $\mu$ auf $(C, \mathcal{B}_C)$, so konvergiert für alle $m \in \mathbb{N}$ und $0 \leq t_1 < t_2 < \cdots < t_m \leq 1$ auch $\mu_n \pi^{-1}_{t_1,\dots,t_m}$ gegen $\mu \pi^{-1}_{t_1,\dots,t_m}$. Dieses Wahrscheinlichkeitsmaß muss dann nach Satz 7.3 das dort angegebene Wahrscheinlichkeitsmaß auf $(\mathbb{R}^m, \mathcal{B}^m)$ sein. Gibt es ein solches Maß $\mu$?

**7.5 Satz** *Es gibt genau ein Wahrscheinlichkeitsmaß $\mu$ auf $(C[0,1], \mathcal{B}_C)$ derart, dass für alle $m \in \mathbb{N}$ und $0 \leq t_1 < t_2 < \cdots < t_m \leq 1$ das Maß $\mu \pi^{-1}_{t_1,\dots,t_m}$ die $m$-dimensionale Normalverteilung mit Erwartungswert $0$ und Kovarianzmatrix $(\min\{t_i, t_j\})_{i,j}$ ist.*

**7.6 Definition** Das Maß $\mu$ aus Satz 7.5 ist das sogenannte WIENER-*Maß* oder die Verteilung der eindimensionalen BROWN*schen Bewegung*.

Die Eindeutigkeit in Satz 7.5 folgt aus der Tatsache, dass die Mengen

$$\{\pi^{-1}_{t_1,\dots,t_m}(A), m \in \mathbb{N}, 0 \leq t_1 < t_2 < \cdots < t_m \leq 1, A \in \mathcal{B}^m\}$$

ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_C$ bilden (vgl. Beweis von Satz 6.2). Die Existenz des WIENER-Maßes $\mu$ wird nun simultan mit dem folgenden Satz bewiesen:

**7.7 Satz (von Donsker, 1951)** *Es gilt $\mu_n \xrightarrow{w} \mu$ auf $(C, \mathcal{B}_C)$.*

Die Existenz in Satz 7.7 und die in Satz 7.5 folgen aus der folgenden Aussage:

**7.8 Satz** *Die Folge $\{\mu_n, n \in \mathbb{N}\}$ ist straff.*

Aus Satz 7.8 folgt mit Satz 6.18 (Satz von PROHOROV), dass $(\mu_n)_n$ konvergente Teilfolgen enthält. Jedes Grenzelement $\mu$ einer derartigen Teilfolge hat aber nach Satz 7.3 die richtigen endlichdimensionalen Randverteilungen $\mu \pi^{-1}_{t_1,\dots,t_m}$. Damit ist die Existenz in Satz 7.5 gezeigt.

Aus Satz 7.8 folgt weiter, dass jede Teilfolge von $(\mu_n)_n$ wieder eine konvergente Teilfolge hat. Deren Grenzelement stimmt mit Satz 7.3 mit dem WIENER-Maß überein. Aus Lemma 6.19 folgt dann $\mu_n \xrightarrow{w} \mu$. Damit ist der Satz von DONSKER bewiesen.

Der Beweis von Satz 7.8 wird noch etwas verschoben. Zunächst folgt aus Satz 7.7

**7.9 Satz** *Ist $h : C[0,1] \to \mathbb{R}$ eine* BOREL*-messbare Abbildung mit $\mu(D_h) = 0$* [1] *und ist $(X_i)_i$ eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit $\mathbb{E}X_i = 0$ und $\mathbb{E}X_i^2 = 1$, so gilt $\mathcal{L}(h(Y_n)) \xrightarrow{w} \mu h^{-1}$, wobei $Y_n$ die durch* (7.1) *definierte $(C, \mathcal{B}_C)$-wertige Zufallsvariable sei.*

*Beweis:* Der Satz folgt sofort aus Satz 6.15.

Der Satz liefert die asymptotische Verteilung von $h(Y_n)$, wenn man $\mu h^{-1}$ kennt. Der Grenzwert hängt aber gar nicht von der speziellen Gestalt der Verteilung der $X_i$ ab. Daher kann der Satz auch zur Berechnung von $\mu h^{-1}$ dienen, wenn man die Verteilung der $h(Y_n)$ kennt. Man kann dazu die Verteilung der $X_i$ beliebig wählen, solange $\mathbb{E}X_i = 0$ und $\mathbb{E}X_i^2 = 1$ erfüllt sind. Meist ist die Berechnung von $\mathcal{L}(h(Y_n))$ am einfachsten, wenn $P(X_i = \pm 1) = 1/2$ ist. Die für diesen Spezialfall gewonnene Grenzverteilung gilt dann für jede Verteilung der $X_i$. Man nennt dies das *Invarianzprinzip* von DONSKER. Wir betrachten im folgenden Kapitel Anwendungen dazu!

Sind $(X_i)_i$ und $(X_i')_i$ zwei Folgen unabhängig, identisch verteilter Zufallsgrößen mit $\mathbb{E}X_i = \mathbb{E}X_i' = 0$ und $\mathbb{E}X_i^2 = \mathbb{E}(X_i')^2 = 1$, und sind $Y_n$ und $Y_n'$ die dazugehörigen interpolierten Irrfahrten, so gilt für jede messbare Funktion $h : C[0,1] \to \mathbb{R}$ mit $\mu(D_h) = 0$

$$\lim_{n\to\infty} \mathcal{L}(h(Y_n)) = \lim_{n\to\infty} \mathcal{L}(h(Y_n')) \ . \tag{7.2}$$

Aus Satz 7.3 wissen wir, dass für $0 \leq t_1 < \cdots < t_m \leq 1$ gilt

$$\lim_{n\to\infty} \mathcal{L}(Y_n(t_1), \ldots, Y_n(t_m)) = \lim_{n\to\infty} \mathcal{L}(Y_n'(t_1), \ldots, Y_n'(t_m)),$$

und somit für jede stetige Funktion $h : \mathbb{R}^m \to \mathbb{R}$

$$\lim_{n\to\infty} \mathcal{L}\big(h(Y_n(t_1), \ldots, Y_n(t_m))\big) = \lim_{n\to\infty} \mathcal{L}\big(h(Y_n'(t_1), \ldots, Y_n'(t_m))\big). \tag{7.3}$$

Wir betrachten die spezielle Abbildung $h : C[0,1] \to \mathbb{R}$ mit $h(f) := \max_{0\leq t\leq 1} f(t)$, bzw. $h : \mathbb{R}^m \to \mathbb{R}$ mit $h(x_1, \ldots, x_m) := \max_{i=1,\ldots,m} x_i$. Wir wollen aus (7.3) nun (7.2) herleiten. Dies ist eine Limesvertauschung von $m \to \infty$ und $n \to \infty$. Genauer: Sei $(t_0^{(m)}, \ldots, t_m^{(m)})_m$ eine Folge von Einteilungen $0 = t_0^{(m)} \leq t_1^{(m)} < \cdots < t_m^{(m)} \leq 1$ des Einheitsintervalls, wobei wir einfach annehmen, dass die $m+1$-te Einteilung durch Hinzunahme eines Punktes aus der $m$-ten entsteht. Es gelte $\max_{1\leq i\leq m}(t_i^{(m)} - t_{i-1}^{(m)}) \to 0$ für $m \to \infty$. Es gilt für jedes $n \in \mathbb{N}$

$$\lim_{m\to\infty} \max_{1\leq i\leq m} Y_n(t_i^{(m)}) = \max_{0\leq t\leq 1} Y_n(t),$$

[1]siehe Notation Lemma 6.14

also

$$\mathcal{L}\Big(\max_{0\le t\le 1} Y_n(t)\Big) = \lim_{m\to\infty} \mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big) .$$

Somit folgt (7.2) aus (7.3), sofern man

$$\lim_{n\to\infty}\lim_{m\to\infty} \mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big) = \lim_{m\to\infty}\lim_{n\to\infty} \mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big) \tag{7.4}$$

zeigen kann (falls Limites existieren).

Erinnerung an Doppelfolgen $(a_{nm})_{n,m\in\mathbb{N}}$: Falls $b_n := \lim_{m\to\infty} a_{nm}$ und $c_m := \lim_{n\to\infty} a_{nm}$ sowie $b := \lim_{n\to\infty} b_n$ existieren, wann gilt $b = \lim_{m\to\infty} c_m$? Hinreichend dafür ist die Konvergenz von $a_{nm} \overset{m\to\infty}{\longrightarrow} b_n$ gleichmäßig in $n$, bzw. $(a_{nm})_m$ ist eine in $n$ gleichmäßige Cauchyfolge: $\sup_n \lim_{m,m'\to\infty} |a_{nm} - a_{nm'}| = 0$.

Zeige also: $\mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big)$ liegt für große $m, m'$ nahe bei $\mathcal{L}\Big(\max_{1\le i\le m'} Y_n(t_i^{(m')})\Big)$, gleichmäßig in $n$.

Für $f \in C[0,1]$ und $\delta > 0$ sei

$$\omega_\delta(f) := \sup\{|f(s) - f(t)| : s, t \in [0,1] \text{ mit } |s-t| \le \delta\}$$

Es gilt

$$|\max_{1\le i\le m} Y_n(t_i^{(m)}) - \max_{1\le i\le m'} Y_n(t_i^{(m')})| \le \omega_\delta(Y_n)$$

für $m' \ge m$, falls $m$ so groß ist, dass $\max_i(t_i^{(m)} - t_{i-1}^{(m)}) \le \delta$. Nun gilt (Übung): Wenn

$$\sup_n P\big(\omega_\delta(Y_n) \ge \varepsilon\big) \overset{\delta\to 0}{\longrightarrow} 0 \text{ für alle } \varepsilon > 0 , \tag{7.5}$$

so folgt (7.4).

Es ist erstaunlich, dass der sehr viel allgemeinere Satz von Donsker sich ebenfalls aus (7.5) ergibt, diese Aussage also die wirklich entscheidende für den Beweis von Satz 7.8 sein wird: Es geht um eine Charakterisierung relativ kompakter Teilmengen in $C[0,1]$.

Das Stetigkeitsmodul $\omega_\delta(f)$ sei wie oben definiert. Es gilt $|\omega_\delta(f) - \omega_\delta(g)| \le 2d(f,g)$, also ist für $\delta > 0$ fest $\omega_\delta$ stetig. Da ein $f \in C[0,1]$ gleichmäßig stetig ist, gilt $\lim_{\delta\to 0} \omega_\delta(f) = 0$.

**7.10 Satz (von Arzelà-Ascoli)** *Eine Teilmenge $A \subset C[0,1]$ hat genau dann kompakten Abschluss, wenn*

(a) $\sup\{|f(0)|, f \in A\} < \infty$ *ist und*

(b) $\lim_{\delta\to 0} \sup_{f\in A} \omega_\delta(f) = 0$.

*gelten. (Wir geben im Anhang dieses Kapitels einen Beweis.)*

Dieser Satz kann in ein Kriterium für die Straffheit einer Folge von W-Maßen auf $C$ übersetzt werden:

**7.11 Satz** *Eine Folge $(\nu_n)_n$ von W-Maßen auf $(C, \mathcal{B}_C)$ ist genau dann straff, wenn*

$$\lim_{a \nearrow \infty} \sup_n \nu_n(\{f : |f(0)| > a\}) = 0 \text{ und} \tag{7.6}$$

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) = 0 \text{ für alle } \varepsilon > 0 \tag{7.7}$$

*gelten.*

Nach obiger Bemerkung ist $\{f : \omega_\delta(f) \geq \varepsilon\} \in \mathcal{B}_C$. Die Bedingungen (7.6) und (7.7) in Satz 7.11 können wie folgt übersetzt werden:

$$\forall \eta > 0 \; \exists a > 0 \; \forall n \in \mathbb{N} : \quad \nu_n(\{f : |f(0)| > a\}) \leq \eta \,, \tag{7.8}$$

$$\forall \varepsilon > 0, \eta > 0 \; \exists \delta > 0 \; \exists n_0 \in \mathbb{N} \; \forall n \geq n_0 :$$
$$\nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta. \tag{7.9}$$

**7.12 Bemerkung** $C[0,1]$ ist vollständig und separabel, also ist jedes Wahrscheinlichkeitsmaß $\nu$ auf $C$ straff: $\forall \eta > 0$ existiert eine kompakte Menge $K$ mit $\nu(K) \geq 1 - \eta$. Insbesondere folgt, dass für $\varepsilon > 0$ ein $\delta > 0$ existiert mit $\nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta$. Somit ist (7.9) äquivalent zu

$$\forall \varepsilon > 0, \eta > 0 \; \exists \delta > 0 \; \forall n \in \mathbb{N} : \quad \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta \,. \tag{7.10}$$

*Beweis von Satz 7.11:* Sei $\{\nu_n, n \in \mathbb{N}\}$ straff. Für $\eta > 0$ sei $K$ eine kompakte Menge mit $\nu_n(K) \geq 1 - \eta$ für alle $n$. Daraus folgen mit dem Satz von Arzelà-Ascoli die Aussagen (7.8) und (7.10), denn $K \subset \{f : |f(0)| \leq a\}$ für $a$ groß genug und $K \subset \{f : \omega_\delta(f) < \varepsilon\}$ für $\delta$ klein genug. Für die Umkehrung sei $(\nu_n)_n$ eine Folge, die (7.8) und (7.10) erfüllt. Sei $\eta > 0$ vorgegeben. Nach (7.8) existiert ein $a \in \mathbb{R}$, so dass $A := \{f : |f(0)| \leq a\}$ erfüllt: $\nu_n(A) \geq 1 - \eta/2$ für alle $n \in \mathbb{N}$. Für $k \in \mathbb{N}$ sei $\delta_k$ so gewählt, dass $\nu_n(\{f : \omega_{\delta_k}(f) < 1/k\}) \geq 1 - \eta/2^{k+1}$ für alle $n$ gilt. Nach dem Satz von Arzelà-Ascoli hat

$$K := A \cap \bigcap_{k=1}^{\infty} \{f : \omega_{\delta_k}(f) < 1/k\}$$

kompakten Abschluss und es gilt

$$\nu_n(\bar{K}^c) \leq \nu(K^c) \leq \eta/2 + \sum_{k=1}^{\infty} \eta/2^{k+1} = \eta$$

für alle $n \in \mathbb{N}$, was zu zeigen war. $\square$

**7.13 Bemerkung** Hinreichend für (7.8) ist $\nu_n(\{f : f(0) = 0\}) = 1$, was für die $\mu_n$ im Satz von DONSKER erfüllt ist.

**7.14 Lemma** *Hinreichend für* (7.9) *ist:*

$$\begin{aligned}&\forall \varepsilon, \eta > 0 \ \exists \delta \in (0,1), \ \exists n_0 \in \mathbb{N} : \quad \forall n \geq n_0, \ \forall t \in [0, 1-\delta] : \\ &\frac{1}{\delta}\nu_n\Big(\Big\{f : \sup_{t \leq s \leq t+\delta} |f(s) - f(t)| \geq \varepsilon\Big\}\Big) \leq \eta \ . \end{aligned} \tag{7.11}$$

*Beweis:* Seien $\varepsilon, \eta > 0$. Zu $\varepsilon/2$ und $\eta/3$ wählen wir $\delta_0 \in (0,1)$ und $n_0 \in \mathbb{N}$ wie in (7.11). $m \in \mathbb{N}$ sei die kleinste natürliche Zahl mit $1/m < \delta_0$. Setze $\delta := \frac{1}{2m}$. Ist $f \in C[0,1]$ mit $\omega_\delta(f) \geq \varepsilon$, so existieren $t < s$ mit $|f(t) - f(s)| \geq \varepsilon$ und $|t-s| \leq \delta$. Zu $t, s$ existiert ein $k \in \mathbb{N}_0$ mit $k \leq 2m-2$ und $\frac{k}{2m} \leq t < s \leq \frac{k}{2m} + \frac{1}{m}$. Dann ist $\big|f(t) - f\big(\frac{k}{2m}\big)\big| \geq \varepsilon/2$ oder $\big|f(s) - f\big(\frac{k}{2m}\big)\big| \geq \varepsilon/2$. Also ist

$$\{f : \omega_\delta(f) \geq \varepsilon\} \subset \bigcup_{k=0}^{2m-2} \Big\{f : \sup_{\frac{k}{2m} \leq s \leq \frac{k}{2m} + \delta_0} \big|f(s) - f\big(\frac{k}{2m}\big)\big| \geq \varepsilon/2\Big\},$$

und somit gilt für alle $n \geq n_0$:

$$\begin{aligned}\nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) &\leq \sum_{k=0}^{2m-2} \nu_n\Big(\Big\{f : \sup_{\frac{k}{2m} \leq s \leq \frac{k}{2m} + \delta_0} \big|f(s) - f\big(\frac{k}{2m}\big)\big| \geq \varepsilon/2\Big\}\Big) \\ &\leq (2m-1)\delta_0 \frac{\eta}{3} \leq (2 + \delta_0)\frac{\eta}{3} \leq \eta \ .\end{aligned}$$

Damit ist (7.9) gezeigt. □

**7.15 Bemerkung** Die Bedingung in Lemma 7.14 folgt aus der folgenden Aussage: Für alle $\varepsilon > 0$ gilt:

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} \sup_{t \in [0, 1-\delta]} \frac{1}{\delta} \nu_n\Big(\Big\{f : \sup_{t \leq s \leq t+\delta} |f(s) - f(t)| \geq \varepsilon\Big\}\Big) = 0 \ .$$

Die Bedingung aus Bemerkung 7.15 soll nun für $\mu_n = P^{Y_n}$ untersucht werden: Für $\delta \in (0,1)$ und $t \in [0, 1-\delta]$ ist

$$\mu_n\Big(\Big\{f : \sup_{t \leq s \leq t+\delta} |f(s) - f(t)| \geq \varepsilon\Big\}\Big) = P\Big(\sup_{t \leq s \leq t+\delta} |Y_n(s) - Y_n(t)| \geq \varepsilon\Big).$$

Für $t = k/n$ und $t + \delta = j/n$ $(k < j)$ ist

$$\sup_{t \leq s \leq t+\delta} |Y_n(s) - Y_n(t)| = \max_{1 \leq i \leq n\delta} \frac{|S_{k+i} - S_k|}{\sqrt{n}}.$$

Für allgemeine $t \in [0,1]$ und $\delta \in (0,1)$ mit $t+\delta \leq 1$ kann man so abschätzen: Es existieren $j,k \in \{0,1,\dots,n\}$ mit $k < j$ und $\frac{k}{n} \leq t < \frac{k+1}{n}$ sowie $\frac{j-1}{n} < t+\delta \leq \frac{j}{n}$. Dann gilt für jedes $s \in [t, t+\delta]$:

$$\begin{aligned}|Y_n(s) - Y_n(t)| &\leq \Big|Y_n(t) - Y_n\Big(\frac{k}{n}\Big)\Big| + \max_{1\leq i\leq j-k}\Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big| \\ &\leq 2\max_{1\leq i\leq j-k}\Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big|,\end{aligned}$$

also

$$\begin{aligned}\sup_{t\leq s\leq t+\delta}|Y_n(s) - Y_n(t)| &\leq 2\max_{1\leq i\leq j-k}\Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big| \\ &= 2\max_{1\leq i\leq j-k}\Big|\sum_{r=k+1}^{k+i} X_r\Big|/\sqrt{n}\ .\end{aligned}$$

Es ist $\frac{j-k-2}{n} \leq \delta$. Für $n \geq \frac{1}{\delta}$ folgt $j-k \leq 3n\delta$. Somit ist die rechte Seite der letzten Ungleichung nicht größer als $2\max_{1\leq i\leq 3n\delta}\Big|\sum_{r=k+1}^{k+i} X_r\Big|/\sqrt{n}$. Die Verteilung dieser Zufallsvariablen hängt nicht von $k$ ab. Für $n \geq \frac{1}{\delta}$ gilt somit

$$\sup_{t\in[0,1-\delta]} P\Big(\sup_{t\leq s\leq t+\delta}|Y_n(s) - Y_n(t)| \geq \varepsilon\Big) \leq P\Big(\max_{1\leq i\leq 3n\delta}\frac{|S_i|}{\sqrt{n}} \geq \frac{\varepsilon}{2}\Big)\ .$$

Sei $m := [3\,n\,\delta]$, so ist $\sqrt{n} \geq \sqrt{m/3\delta}$ und somit

$$P\Big(\max_{1\leq i\leq 3n\delta}\frac{|S_i|}{\sqrt{n}} \geq \frac{\varepsilon}{2}\Big) \leq P\Big(\max_{1\leq i\leq m}\frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{12\delta}}\Big)\ .$$

Für jedes feste $\delta > 0$ geht $m \to \infty$ für $n \to \infty$. Nach Bemerkung 7.15 müssen wir für jedes $\varepsilon > 0$ zeigen, dass:

$$\lim_{\delta\searrow 0}\limsup_{m\to\infty}\frac{1}{\delta}P\Big(\max_{1\leq i\leq m}\frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big) = 0 \tag{7.12}$$

gilt. Leider hilft die Abschätzung

$$P\Big(\max_{1\leq i\leq m}\frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big) \leq \sum_{i=1}^{m} P\Big(\frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big)$$

analog zum Beweis von Lemma 7.14 nicht. Wir müssen diese Wahrscheinlichkeit wesentlich genauer abschätzen:

**7.16 Lemma** *Für alle* $\lambda > 0$ *und* $m \in \mathbb{N}$ *gilt*

$$P\Big(\max_{1\leq i\leq m}|S_i| \geq \lambda\sqrt{m}\Big) \leq 2P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big)\ .^{2}$$

[2] vgl. Kolmogorov-Ungleichung, Satz 11.7, Wahrscheinlichkeitstheorie

*Beweis:* Für $\lambda \leq \sqrt{2}$ ist nichts zu zeigen. Sei $\lambda > \sqrt{2}$.

$$A_i := \bigcap_{j=1}^{i-1} \{|S_j| < \lambda\sqrt{m}\} \cap \{|S_i| \geq \lambda\sqrt{m}\}\ ,\ i = 1, \ldots, m.$$

Die $A_i$ sind disjunkt und $A = \{\max_{1\leq i\leq m} |S_i| \geq \lambda\sqrt{m}\} = \bigcup_{i=1}^m A_i$. Also

$$\begin{aligned} P(A) &= P\Big(A \cap \{|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\}\Big) + P\Big(A \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}\Big) \\ &\leq P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) \\ &\quad + \sum_{j=1}^{m-1} P\Big(A_j \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}\Big)\ , \end{aligned}$$

denn $A_m \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\} = \emptyset$. Weiter gilt

$$A_j \cap \Big\{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\Big\} \subset A_j \cap \Big\{|S_m - S_j| \geq \sqrt{2m}\Big\}\ ,\ j = 1, \ldots, m-1\ .$$

Die Ereignisse $A_j$ und $\{|S_m - S_j| \geq \sqrt{2m}\}$ sind unabhängig, also haben wir

$$P(A) \leq P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) + \sum_{j=1}^{m-1} P(A_j) P\Big(|S_m - S_j| \geq \sqrt{2m}\Big).$$

Wegen

$$P\Big(|S_m - S_j| \geq \sqrt{2m}\Big) \leq \frac{1}{2m}\mathbb{E}\Big(\big(\sum_{k=j+1}^m X_k\big)^2\Big) = \frac{1}{2m}\sum_{k=j+1}^m \mathbb{E}(X_k^2) \leq \frac{1}{2}$$

folgt

$$\begin{aligned} P(A) &\leq P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) + \frac{1}{2}\sum_{j=1}^m P(A_j) \\ &= P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) + \frac{1}{2}P(A)\ , \end{aligned}$$

also folgt die Behauptung. □

Wir schließen mit dem Beweis von (7.12) ab:

Mit Lemma 7.16 und dem zentralen Grenzwertsatz folgt:

$$\begin{aligned} \limsup_{m\to\infty} \frac{1}{\delta} P\Big(\max_{1\leq i\leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big) &\leq \limsup_{m\to\infty} \frac{2}{\delta} P\Big(\frac{|S_m|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big) \\ &= \frac{2}{\delta} P\Big(|N| \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big), \end{aligned}$$

wenn $N$ eine $N(0,1)$-verteilte Zufallsgröße bezeichnet. Die MARKOV-Ungleichung liefert

$$P\Big(|N| \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big) \leq \frac{\mathbb{E}(|N|^3)}{\big(\frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\big)^3}.$$

Dies führt zu (7.12). Somit ist die Straffheit der Folge $(\mu_n)_n$ bewiesen und somit Satz 7.5. $\square$

Wir sammeln noch ein paar Eigenschaften des WIENER-Maßes $\mu$. Natürlich gilt

(a) $\mu(C[0,1]) = 1$

(b) Die Familie der Projektionen $(\pi_t)_{t\in[0,1]}$ erfüllt:

$$\mu(\pi_t \leq \alpha) = \lim_{n\to\infty} \mu_n(\pi_t \leq \alpha) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \exp\big(-\frac{s^2}{2t}\big)\, ds$$

für $t > 0$ und $\mu(\pi_0 = 0) = \lim_{n\to\infty} \mu_n(\pi_0 = 0) = 1$.

(c) Die Familie $(\pi_t)_{t\in[0,1]}$ hat *unabhängige Zuwächse*, d.h. für $0 = t_0 \leq \cdots \leq t_m \leq 1$, $\alpha_1, \dots, \alpha_m \in \mathbb{R}$, $m \in \mathbb{N}$, gilt

$$\mu(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i, i = 1, \dots, m) = \prod_{i=1}^{m} \mu(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i).$$

Den Punkt (c) sieht man mit Hilfe von Satz 7.3 so:

$$\begin{aligned}
&\mu(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i, i = 1, \dots, m) = \lim_{n\to\infty} \mu_n(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i, i = 1, \dots, m)\\
&= \mu_{t_1} \otimes \mu_{t_2 - t_1} \otimes \cdots \otimes \mu_{t_m - t_{m-1}} \Big( \times_{i=1}^{m} (-\infty, \alpha_i] \Big)\\
&= \prod_{i=1}^{m} \lim_{n\to\infty} \mu_n(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i)\\
&= \prod_{i=1}^{m} \mu(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i).
\end{aligned}$$

**7.17 Definition** Die Familie $(\pi_t)_{t\in[0,1]}$ heißt BROWN*sche Bewegung*. Präziser nennt man das Tupel $(C, \mathcal{B}_C, \mu, (\pi_t)_{t\in[0,1]})$ so.

Damit ist uns schon der wohl wichtigste stochastische Prozess, gemeint ist die Familie $(\pi_t)_{t\in[0,1]}$, begegnet, die BROWNsche Bewegung eines Teilchens (z.B. Pollen- oder Staubkorns, eines markierten Moleküls) in einer Flüssigkeit oder einem Gas. Der Ort $x_t \in \mathbb{R}^3$ (bei uns zunächst in $\mathbb{R}$) wird durch eine Zufallsvariable $\pi_t$ beschrieben. BROWN entdeckte 1828 das Phänomen dieser Bewegung. EINSTEIN entwickelte 1905 die physikalische Theorie, unabhängig davon 1906 SMOLUCKOWSKI. EINSTEIN beschreibt die Bewegung eines Teilchens unter Berücksichtigung von Kollisionen mit vielen Teilchen und nimmt unabhängige Zuwächse und zeitlich stationäre Zuwächse an. Er bestimmt die Verteilung des Zuwachses in $[0,t]$ als Normalverteilung $N(0,\sigma^2)$ mit $\sigma^2 = 2t$. BACHELIER untersuchte 1900 in seiner bei POINCARÉ geschriebenen Dissertation ökonomische Agenten zur Beschreibung von Kursschwankungen an der Pariser Börse. Dabei nahm er für Fluktuationen in $[0,t]$ eine Normalverteilung $N(0,2t)$ an! Der mathematische Begriff der BROWNschen Bewegung wurde 1920 von N.WIENER geprägt.

ANHANG: Beweis des Satzes von ARZELÀ-ASCOLI:

Wir bereiten den Beweis durch ein Kriterium für Kompaktheit von Mengen in metrischen Räumen vor.

**7.18 Satz** *Eine Teilmenge eines metrischen Raumes $(X,d)$ ist genau dann kompakt, wenn sie vollständig und totalbeschränkt ist. Dabei heißt $K \subset X$* totalbeschränkt, *wenn es zu jedem $r > 0$ ein $m \in \mathbb{N}$ und $x_0, \dots, x_m \in K$ gibt mit $K \subset \bigcup_{k=0}^m B(x_k, r)$ (womit jede totalbeschränkte Menge beschränkt ist).*

*Beweis:* Es sei $K \subset X$ kompakt, $(x_j)_j$ sei eine Cauchyfolge in $K$. $K$ ist folgenkompakt (denn eine Teilmenge eines metrischen Raums ist genau dann kompakt, wenn sie folgenkompakt ist, Analysis I), also besitzt $(x_j)_j$ eine in $K$ konvergente Teilfolge. Damit konvergiert die Folge (denn besitzt eine Cauchyfolge eine konvergente Teilfolge, so ist sie selbst konvergent, Analysis I) in $K$, also ist $K$ vollständig. Für jedes $r > 0$ ist $\{B(x,r), x \in K\}$ eine offene Überdeckung von $K$. Da $K$ kompakt, gibt es eine endliche Teilüberdeckung, also ist $K$ auch totalbeschränkt.

Sei nun $K$ vollständig und totalbeschränkt. $(x_j)_j$ sei eine Folge in $K$. Zu jedem $n \in \mathbb{N}$ existieren endlich viele Bälle mit Mittelpunkten in $K$ und Radius $1/n$, die $K$ überdecken. Es existiert also eine Teilfolge $(x_{1,j})_j$ von $(x_j)_j$, die ganz in einem Ball mit Radius 1 enthalten ist. Dann gibt es eine Teilfolge $(x_{2,j})_j$ von $(x_{1,j})_j$, die ganz in einem Ball mit Radius $1/2$ enthalten ist, etc. Also gibt es zu jedem $n \in \mathbb{N}$ eine Teilfolge $(x_{n+1,j})_j$ von $(x_{n,j})_j$, die ganz in einem Ball mit Radius $1/(n+1)$ enthalten ist. Sei $y_n := x_{n,n}$, $n \in \mathbb{N}$ (Diagonalfolge). Dann ist $(y_n)_n$ offensichtlich eine Cauchyfolge in $K$, also konvergiert $(y_n)_n$ in $K$, da $K$ vollständig. $(x_j)_j$ hat also eine in $K$ konvergente Teilfolge: $(y_n)_n$, also ist $K$ folgenkompakt, also kompakt. $\square$

Im zweiten Teil des Beweises haben wir das *Diagonalfolgenprinzip* verwendet. Wir wählen aus einer Folge gemäß einer Vorschrift sukzessive Teilfolgen aus und bilden dann die Diagonalfolge, indem man von der $n$-ten Teilfolge das $n$-te Glied auswählt. Hier ist $(x_{n+1,j})_j$ für jedes $n \in \mathbb{N}$ eine Teilfolge von $(x_{n,j})_j$. Die Diagonalfolge $(y_n)_n$ hat dann die Eigenschaft, dass $(y_n)_{n \geq N}$ für jedes $N \in \mathbb{N}$ eine Teilfolge von $(x_{N,j})_j$ ist, also dieselben infinitären Eigenschaften wie jede der Teilfolgen $(x_{n,j})_j$ besitzt.

Da $A \subset X$ genau dann totalbeschränkt ist, wenn $\bar{A}$ totalbeschränkt ist, besagt der obige Satz, dass für eine Teilmenge $A \subset X$ gilt: $\bar{A}$ ist genau dann kompakt, wenn $A$ totalbeschränkt und $\bar{A}$ vollständig ist.

*Beweis des Satzes von* Arzelà-Ascoli*:* Sei $\bar{A} \subset C[0,1]$ kompakt. Dann ist $A$ totalbeschränkt: Zu $\varepsilon > 0$ existieren $f_1, \ldots, f_n \in A$ mit $d(f, f_j) < \varepsilon/3$ für ein $j \in \{1, \ldots, n\}$ für alle $f \in A$. Jedes $f_j$ in $C[0,1]$ ist gleichmäßig stetig, also gilt für die endliche Menge $\{f_1, \ldots, f_n\}$: Wähle $\delta > 0$, so dass $|x - y| < \delta$ $|f_j(x) - f_j(y)| < \varepsilon/3$ für alle $j = 1, \ldots, n$ und $x, y \in [0,1]$ zur Folge hat. Also ist $|f(x) - f(y)| < \varepsilon$ für alle $f \in A$, somit gilt $\lim_{\delta \to 0} \sup_{f \in A} \omega_\delta(f) = 0$. $A$ ist auch beschränkt bezüglich $d$, was (a) zur Folge hat.

Seien nun (a) und (b) gegeben. Wähle $k$ groß genug, so dass $\sup_{f \in A} \omega_{1/k}(f)$ endlich ist. Da

$$|f(t)| \leq |f(0)| + \sum_{i=1}^{k} \left| f\left(\frac{i}{k}t\right) - f\left(\frac{i-1}{k}t\right) \right| ,$$

folgt mit (a)

$$\sup_{t \in [0,1]} \sup_{f \in A} |f(t)| < \infty . \tag{7.13}$$

Wir zeigen nun, dass aus (b) und (7.13) folgt, dass $A$ totalbeschränkt ist, also auch $\bar{A}$. Nun ist $C[0,1]$ vollständig, also auch $\bar{A}$, damit ist $\bar{A}$ dann kompakt. Sei $\varepsilon > 0$ und

$$\alpha := \sup_{t \in [0,1]} \sup_{f \in A} |f(t)| .$$

Ferner sei $H := \{\frac{u}{v}\alpha, u = 0, \pm 1, \pm 2, \ldots, \pm v, v \in \mathbb{N}\}$ mit $v \in \mathbb{N}$ so, dass $\frac{\alpha}{v} < \varepsilon$. $H$ hat dann die Eigenschaft, dass zu jedem $t \in [-\alpha, \alpha]$ ein $t_k \in H$ existiert mit $|t - t_k| < \varepsilon$. Nun wähle $k$ groß genug, so dass $\omega_{1/k}(f) < \varepsilon$ für alle $f \in A$. $B$ sei die Teilmenge in $C[0,1]$ derjenigen Funktionen, die in jedem Intervall $[\frac{i-1}{k}, \frac{i}{k}]$, $i = 1, \ldots, k$, linear sind und Werte aus $H$ an den Endpunkten $\frac{i}{k}$, $i = 0, \ldots, k$, annehmen. $B$ ist endlich (besteht aus $(2\nu + 1)^{k+1}$ Punkten). Wir zeigen nun, dass jedes $f \in A$ in einem $2\varepsilon$-Ball um ein Element aus $B$ liegt: Sei $f \in A$, also $|f(\frac{i}{k})| \leq \alpha$. Dann existiert ein $g \in B$ mit

$$\left| f\left(\frac{i}{k}\right) - g\left(\frac{i}{k}\right) \right| < \varepsilon , i = 0, \ldots, k . \tag{7.14}$$

Da $\omega_{1/k}(f) < \varepsilon$ und $g$ linear in jedem Teilintervall $[\frac{i-1}{k}, \frac{i}{k}]$ ist, folgt aus (7.14) $d(f,g) < 2\varepsilon$. Dies war zu zeigen. $\square$

KAPITEL 8

# Anwendungen des Invarianzprinzips, die eindimensionale Irrfahrt

Im ersten Beispiel soll Satz 7.9, das Invarianzprinzip von Donsker, auf $h : C[0,1] \to \mathbb{R}$, $h(f) := \sup_{0 \le t \le 1} f(t)$ angewendet werden. Es ist leicht einzusehen, dass $h$ stetig ist.

Nun seien die Zufallsgrößen $(X_i)_i$ unabhängig mit

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}\,,$$

also $\mathbb{E}X_i = 0$ und $\operatorname{Var} X_i = \mathbb{E}X_i^2 = 1$. Wieder sei $S_0 = 0$ und $S_n = \sum_{i=1}^n X_i$, sowie $M_n := \max_{1 \le i \le n} S_i$ für $n \in \mathbb{N}$. Wir interessieren uns für die Verteilung von $M_n$. Man beachte:

$$\sup_{0 \le t \le 1} Y_n(t) = \max_{0 \le i \le n} \frac{S_i}{\sqrt{n}} = \frac{M_n}{\sqrt{n}}\,.$$

$(M_n)_n$ heißt auch die Folge der *Maximalgewinne* beim Münzwurfspiel.

**8.1 Satz** *Für die Folge $(M_n)_n$ der Maximalgewinne beim Münzwurfspiel gilt für alle $t \ge 0$*

$$\lim_{n \to \infty} P\Big(\frac{M_n}{\sqrt{n}} \le t\Big) = 2\Phi(t) - 1\,.$$

*Für $t < 0$ gilt*

$$P\Big(\frac{M_n}{\sqrt{n}} \le t\Big) = 0\,.$$

**8.2 Satz** *Erfüllen die $(X_i)_i$ die Voraussetzungen des Satzes 7.9, so gilt für alle $t \in \mathbb{R}$*

$$\lim_{n \to \infty} P\Big(\max_{0 \le i \le n} \frac{S_i}{\sqrt{n}} \le t\Big) = \max\{2\Phi(t) - 1,\, 0\}\,.$$

Für den Beweis von Satz 8.1 bereiten wir das sogenannte *Spiegelungsprinzip (Reflexionsprinzip)* vor.

Für $i, j \in \mathbb{Z}$, $i < j$, nennen wir eine Folge $(i, s_i), \ldots, (j, s_j)$ mit $s_k \in \mathbb{Z}$, $i \le k \le j$, und $|s_{k+1} - s_k| = 1$ für $i \le k \le j-1$ einen *Pfad* von $(i, s_i)$ nach $(j, s_j)$. Oft schreibt man einfach $(s_i, s_{i+1}, \ldots, s_j)$. $j - i$ ist die *Länge* des Pfades. Wir sagen, dass ein Pfad $(s_i, s_{i+1}, \ldots, s_j)$ die $x$-Achse berührt, falls ein $k$ mit $i \le k \le j$ existiert, für das $s_k = 0$ ist.

**8.3 Lemma (Spiegelungsprinzip / Reflexionsprinzip)**

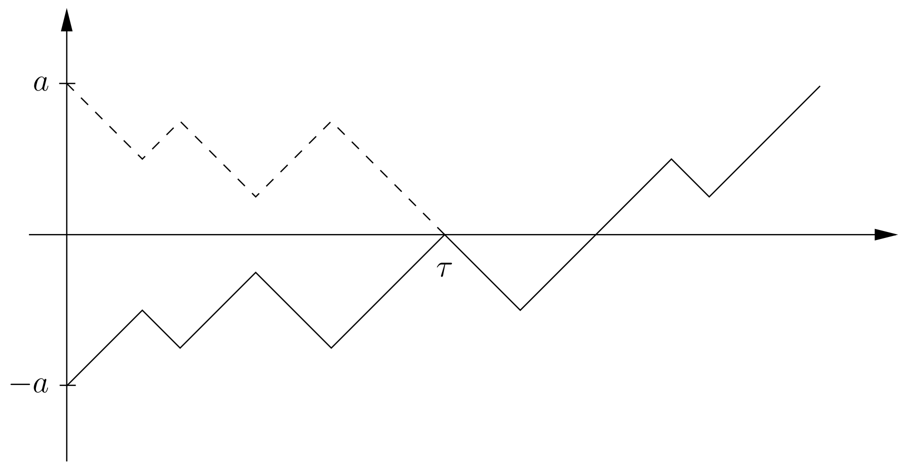
(a) *Es seien $a, b \in \mathbb{N}$ und $i, j \in \mathbb{Z}$ mit $i < j$. Die Anzahl der Pfade von $(i,\, a)$ nach $(j,\, b)$, welche die $x$-Achse berühren, ist gleich der Anzahl der Pfade von $(i,\, -a)$ nach $(j,\, b)$.*

(b) *Sei $a \in \mathbb{N}$, $b \in \mathbb{Z}$ und $n \in \mathbb{N}$. Die Anzahl der Pfade von $(0,0)$ nach $(n, b)$, die $s_j = a$ für ein $j \in \{1, \ldots, n\}$ erfüllen, ist gleich der Anzahl der Pfade von $(0,0)$ nach $(n,\, 2a - b)$, die $s_j = a$ für ein $j \in \{1, \ldots, n\}$ erfüllen.*

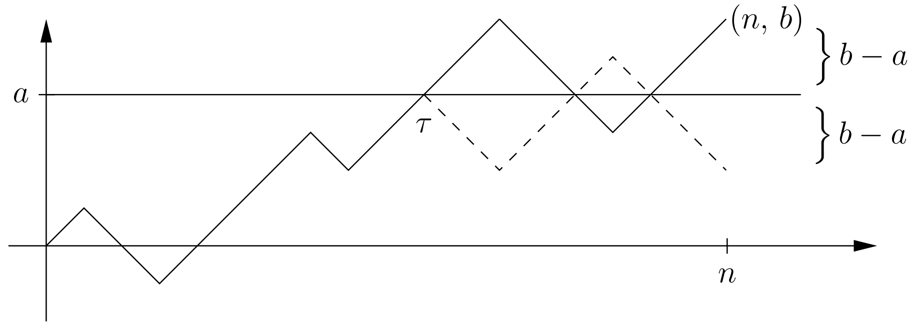
*Beweis:* (a) Sei $(s_i = -a,\ s_{i+1}, \ldots, s_{j-1}, s_j = b)$. Dieser Pfad muss die $x$-Achse berühren. $\tau$ sei die kleinste Zahl größer als $i$, für welche $s_\tau = 0$ gilt. Dann ist

$$(-s_i,\ -s_{i+1}, \ldots, -s_{\tau-1},\ s_\tau = 0, s_{\tau+1}, \ldots, s_j = b)$$

ein Pfad von $(i,\, a)$ nach $(j,\, b)$, der die $x$-Achse berührt, und die Zuordnung ist bijektiv.




Das Bild für den Beweis von (b) ist




$\tau$ ist das erstmalige Erreichen des Wertes $a$. $\square$

*Beweis von Satz 8.1:* Für $l, k \in \mathbb{N}_0$ gilt

$$P(S_n = l + k) = P(S_n = l + k,\ M_n \geq k)\,.$$

Nun ist nach Teil (b) von Lemma 8.3

$$P(M_n \geq a,\ S_n = b) = P(M_n \geq a,\ S_n = 2a - b)$$

für jedes $b \in \mathbb{Z}$. Also ist

$$P(S_n = l + k) = P(M_n \geq k,\ S_n = k - l)\,.$$

Damit ist

$$\begin{aligned} P(M_n \geq k) &= \sum_{l=-\infty}^{\infty} P(M_n \geq k,\ S_n = l + k) \\ &= \sum_{l=-\infty}^{-1} P(M_n \geq k,\ S_n = l + k) + \sum_{l=1}^{\infty} P(S_n = l + k) + P(S_n = k) \\ &= 2P(S_n > k) + P(S_n = k) \\ &= 2P(S_n \geq k) - P(S_n = k)\,. \end{aligned}$$

Sei $t \in \mathbb{R}_+$. Für $n \in \mathbb{N}$ bezeichne $k_n$ die kleinste natürliche Zahl größer-gleich $t\sqrt{n}$. Es gilt

$$P^{S_n/\sqrt{n}} \xrightarrow{w} N(0,1)\,.$$

Da $\{S_n/\sqrt{n} \geq t\} = \{S_n \geq k_n\}$, folgt

$$\lim_{n\to\infty} P(S_n \geq k_n) = \nu_{0,1}\big([t,\,\infty)\big)\,.$$

Wegen $t\sqrt{n} \leq k_n < t\sqrt{n} + 1$ gilt weiter für jedes $\varepsilon > 0$ und alle $n \in \mathbb{N}$ mit $1/\sqrt{n} \leq \varepsilon$

$$\{S_n = k_n\} = \Big\{\frac{S_n}{\sqrt{n}} = \frac{k_n}{\sqrt{n}}\Big\} \subset \Big\{t \leq \frac{S_n}{\sqrt{n}} < t + \varepsilon\Big\}\,,$$

und daraus folgt

$$\limsup_{n\to\infty} P(S_n = k_n) \leq \int_t^{t+\varepsilon} g_{0,1}(x)\,dx \quad \forall \varepsilon \geq 0\,,$$

also

$$\lim_{n\to\infty} P(S_n = k_n) = 0\,.$$

Zusammen erhalten wir

$$\begin{aligned} \lim_{n\to\infty} P\Big(\frac{M_n}{\sqrt{n}} \geq t\Big) &= 2\nu_{0,1}\big([t,\,\infty)\big) \\ &= \Big(\frac{2}{\pi}\Big)^{1/2} \int_t^{\infty} e^{-x^2/2}\,dx = 2\big(1 - \Phi(t)\big)\,, \end{aligned}$$

womit die Behauptung des Satzes folgt. $\square$

Im zweiten Beispiel betrachten wir die Abbildung $g : C[0,1] \to \mathbb{R}$ mit

$$g(f) := \lambda\big(\{t \in [0,1] : f(t) \geq 0\}\big)\,,$$

wobei $\lambda$ das LEBESGUE-Maß sei. Die Funktion $g$ ist nicht überall stetig, z.B. ist sie unstetig in $f \equiv 0$. Es gilt jedoch

**8.4 Lemma** *$g$ ist $\mathcal{B}_C/\mathcal{B}$-messbar und $\mu(D_g) = 0$, wobei $\mu$ das* WIENER-*Maß bezeichnet.*

*Beweis:* Es sei $\psi : C[0,1] \times [0,1] \to \mathbb{R}$ definiert durch $\psi(f, t) = f(t)$. $\psi$ ist stetig (Übung!), also $\mathcal{B}_{C\times[0,1]}/\mathcal{B}$-messbar, wobei wir wieder kurz $C := C[0,1]$ schreiben. Da $C$ und $[0,1]$ separabel sind, folgt analog zu Lemma 6.22

$$\mathcal{B}_{C\times[0,1]} = \mathcal{B}_C \otimes \mathcal{B}_{[0,1]}\,.$$

Also ist $\psi$ $\mathcal{B}_C \otimes \mathcal{B}_{[0,1]}/\mathcal{B}$-messbar.

Sei nun $A = \{(f, t) : f(t) \geq 0\} = \psi^{-1}\big([0,\infty)\big) \in \mathcal{B}_C \otimes \mathcal{B}_{[0,1]}$. Für $f \in C$ ist $g(f) = \lambda\big(\{t : (f, t) \in A\}\big)$. Also ist $f \mapsto g(f)$ $\mathcal{B}_C/\mathcal{B}$-messbar, siehe Satz 8.12 (Satz von Fubini), Wahrscheinlichkeitstheorie.

Es gilt

$$g(f) = \int_0^1 1_{[0,\infty]}\big(f(t)\big)\,dt\,.$$

Ist $f \in C$ mit $\lambda\big(\{t : f(t) = 0\}\big) = 0$, und ist $(f_n)_n$ eine Folge in $C$ mit $d(f_n, f) \to 0$, so gilt $1_{[0,\infty]}\big(f_n(t)\big) \to 1_{[0,\infty]}\big(f(t)\big)$ für $\lambda$-fast alle $t \in [0,1]$. Nach dem Satz von der dominierten Konvergenz folgt

$$g(f_n) \to g(f)\,.$$

Also ist $D_g \subset \big\{f : \lambda\big(\{t : f(t) = 0\}\big) > 0\big\}$ gezeigt.

Wir zeigen

$$\mu\Big(\big\{f : \lambda\big(\{t : f(t) = 0\}\big) > 0\big\}\Big) = 0.$$

Dazu müssen wir zeigen, dass $f \mapsto \lambda\big(\{t : f(t) = 0\}\big)$ messbar ist. Dies geht analog zur Messbarkeit von $g$. Es ist zu zeigen:

$$0 = \int_C \lambda\big(\{t : f(t) = 0\}\big)\,\mu(df) = \int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f, t)\,dt\,\mu(df)\,.$$

Nach dem Satz von FUBINI gilt

$$\begin{aligned}
\int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f, t)\,dt\,\mu(df) &= \int_{[0,1]} \int_C (1_{\{0\}} \circ \psi)(f, t)\,\mu(df)\,dt \\
&= \int_{[0,1]} \mu\big(\{f : f(t) = 0\}\big)\,dt \\
&= \int_{[0,1]} \mu\pi_t^{-1}\big(\{0\}\big)\,dt\,.
\end{aligned}$$

Das letzte Integral ist tatsächlich gleich Null, denn $\mu\pi_t^{-1}$ ist für $t > 0$ die Normalverteilung mit Erwartungswert 0 und Varianz $t$. Damit ist das Lemma bewiesen. $\square$

Die Abbildung $g$ erfüllt also die Voraussetzung des Invarianzprinzips, Satz 7.9. Es folgt nun die Berechnung von $\mathcal{L}\big(g(Y_n)\big)$ im Spezialfall, wo $P(X_i = \pm 1) = 1/2$ ist. Dies ist eine elementare und schöne Auseinandersetzung mit der eindimensionalen, symmetrischen Irrfahrt und hebt die Bedeutung des Reflexionsprinzips eindrücklich hervor. Es gilt:

**8.5 Satz** *Sind die $(X_i)_i$ unabhängig und $P(X_i = \pm 1) = 1/2$, so gilt für $t \in [0,1]$*

$$\lim_{n\to\infty} P\big(g(Y_n) \le t\big) = \frac{2}{\pi} \arcsin \sqrt{t}\,.$$

Dies liefert somit die Verteilungsfunktion von $\mu\, g^{-1}$, wenn $\mu$ das Wiener-Maß ist. Es folgt mit Satz 7.9:

**8.6 Satz (Arcussinus-Gesetz)** *Die auf $(C, \mathcal{B}_C, \mu)$ definierte Zufallsgröße $f \mapsto \lambda\big(\{t : f(t) \ge 0\}\big)$ hat die Verteilungsfunktion*

$$t \mapsto \frac{2}{\pi} \arcsin \sqrt{t}\,, \quad t \in [0,1]\,.$$

*Erfüllen die $(X_i)_i$ die Voraussetzungen von Satz 7.9, so gilt für $t \in [0,1]$:*

$$\lim_{n\to\infty} P\big(g(Y_n) \le t\big) = \frac{2}{\pi} \arcsin \sqrt{t}\,.$$

**8.7 Bemerkung** Es ist nicht sehr schwer zu zeigen, dass

$$g(Y_n) - \frac{1}{n}\big|\{m \le n : S_m > 0\}\big|$$

in Wahrscheinlichkeit gegen 0 konvergiert. Nach Lemma 6.25 folgt dann, dass auch

$$\mathcal{L}\Big(\frac{1}{n}\big|\{m \le n : S_m > 0\}\big|\Big)$$

asymptotisch nach der Arcussinus-Verteilung verteilt ist. Wir zeigen dies hier nicht.

Zunächst betrachten wir einige kombinatorische Resultate zu Pfaden, so wie sie von unserem Münzwurfspiel der $(X_i)_i$ erzeugt werden, siehe Seite 101. Wir betrachten zwei verschiedene Zufallsexperimente:

(I) *Der Endpunkt liegt fest:* Ist $n \in \mathbb{N}$ und hat $s$, $s \in \mathbb{Z}$, dieselbe Parität wie $n$, so bezeichne $\Omega_{(n,s)}$ die Menge der Pfade von $(0,0)$ nach $(n,s)$. Auf dieser Menge betrachten wir die Gleichverteilung. Wir müssen zunächst die Anzahl der Pfade zählen: Hat ein Pfad $\omega \in \Omega_{(n,s)}$ $p$ ansteigende Verbindungen und $q$ absteigende (d.h. $p := |\{i \in \{0,\ldots,n-1\} : s_{i+1} = s_i + 1\}|$), so gelten $p+q = n$, $p - q = s$, das heißt $p = (n+s)/2$, $q = (n-s)/2$. $p$ und $q$ sind also durch $n$ und $s$ vollständig festgelegt.

$|\Omega_{(n,s)}|$ ist die Anzahl der Möglichkeiten, die $p$ aufsteigenden Verbindungen in der Gesamtzahl von $n$ Schritten zu plazieren, das heißt, es gilt

$$|\Omega_{(n,s)}| = \binom{n}{(n+s)/2} = \binom{p+q}{p}. \tag{8.1}$$

(II) *Freier Endpunkt:* $\Omega_n$ bezeichne die Menge aller Pfade der Länge $n$ mit Startpunkt $(0,0)$. $|\Omega_n|$ ist hier offenbar $2^n$.

Wir betrachten zunächst den Fall (I), das heißt das Zufallsexperiment, das durch die Gleichverteilung auf $\Omega_{(n,s)} = \Omega_{(p+q,p-q)}$ beschrieben wird.

Wir können uns etwa vorstellen, dass eine Wahl zwischen zwei Kandidaten $K_1$, $K_2$ stattgefunden hat, wobei nun $p$ Stimmen für $K_1$ und $q$ Stimmen für $K_2$ in einer Wahlurne liegen. Diese Stimmen werden nun eine um die andere ausgezählt. Wir wollen zunächst das folgende Ereignis betrachten. Sei $p > q$ (d.h. $K_1$ hat gewonnen). Mit welcher Wahrscheinlichkeit liegt $K_1$ stets vorn bei der Auszählung? Diese Wahrscheinlichkeit ist gleich $|A|/|\Omega_{(p+q,p-q)}| = |A|/\binom{p+q}{p}$, wobei

$$A = \{\, \omega = (0, s_1, \ldots, s_{p+q}) \in \Omega_{(p+q,p-q)} : s_k > 0 \text{ für } 1 \le k \le p+q \,\}$$

ist. Zum Abzählen der Pfade in $A$ verwenden wir Lemma 8.3. Für $\omega = (0, s_1, \ldots, s_n) \in A$ gilt notwendigerweise $s_1 = 1$. $|A|$ ist somit die Anzahl der Pfade von $(1,1)$ nach $(p+q, p-q)$, die die $x$-Achse nicht berühren. Dies ist gleich der Anzahl aller Pfade von $(1,1)$ nach $(p+q, p-q)$, minus der Anzahl derjenigen, die die $x$-Achse berühren. Letztere ist nach Lemma 8.3 gleich der Anzahl aller Pfade von $(1,-1)$ nach $(p+q, p-q)$. Wenden wir (8.1) an, so ergibt sich also

$$|A| = \binom{p+q-1}{p-1} - \binom{p+q-1}{p} = \frac{p-q}{p+q}\binom{p+q}{p}. \tag{8.2}$$

(Wir haben hier natürlich $p > q$ vorausgesetzt.) Die Anzahl aller Elemente in $\Omega_{(p+q,p-q)}$ ist nach (8.1) $\binom{p+q}{p}$. Somit ergibt sich das folgende Resultat:

**8.8 Satz (Ballot-Theorem)** (von ballot (engl.) = geheime Abstimmung) *Die Wahrscheinlichkeit dafür, dass der Kandidat mit der größeren Anzahl $p$ der*

*Stimmen während des gesamten Verlaufs der Auszählung führt, ist* $(p-q)/(p+q)$*, wobei* $q$ *die Anzahl der Stimmen des Unterlegenen bezeichnet.*

Eine kleine Modifikation des obigen Arguments gestattet auch die Diskussion des Falles $p = q$. Natürlich kann dann keiner der Kandidaten dauernd führen, da nach der Auszählung Gleichstand herrscht. Wir können aber die beiden folgenden Ereignisse betrachten:

(a) Kandidat $K_1$ führt während der gesamten Auszählung, erst am Schluss tritt Gleichstand ein.

(b) Kandidat $K_2$ führt nie.

Da der zugrunde liegende Wahrscheinlichkeitsraum $\binom{2p}{p}$ Elementarereignisse hat, die alle die gleiche Wahrscheinlichkeit haben, ergeben sich aus dem folgenden Satz die Wahrscheinlichkeiten für diese beiden Ereignisse:

**8.9 Satz**

(a) *Es gibt* $\frac{1}{p}\binom{2p-2}{p-1}$ *Pfade von* $(0,0)$ *nach* $(2p,0)$ *mit* $s_1 > 0, s_2 > 0, \ldots, s_{2p-1} > 0$.

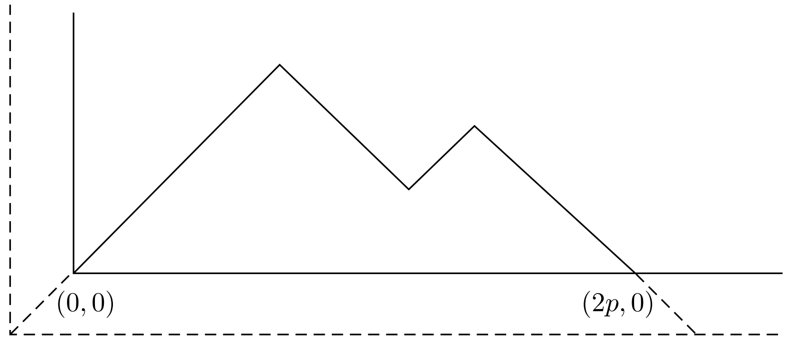
(b) *Es gibt* $\frac{1}{p+1}\binom{2p}{p}$ *Pfade von* $(0,0)$ *nach* $(2p,0)$ *mit* $s_1 \geq 0, s_2 \geq 0, \ldots, s_{2p-1} \geq 0$.

*Beweis:*

(a) Natürlich ist notwendigerweise $s_{2p-1} = 1$. Wir suchen somit nach der Anzahl der Pfade von $(0,0)$ nach $(2p-1,1)$ mit $s_1 > 0, s_2 > 0, \ldots, s_{2p-1} = 1$. Nach der Formel (8.2) mit $q = p-1$ ist dies gleich

$$\frac{1}{2p-1}\binom{2p-1}{p} = \frac{1}{p}\binom{2p-2}{p-1}.$$

(b) Wir verlängern jeden Pfad, der die Bedingung erfüllt, indem wir noch die beiden Punkte $(-1,-1)$ und $(2p+1,-1)$ anfügen und mit $(0,0)$ bzw. $(2p,0)$ verbinden.

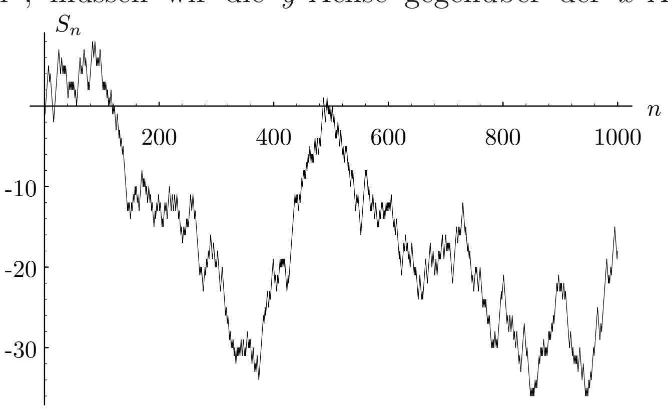


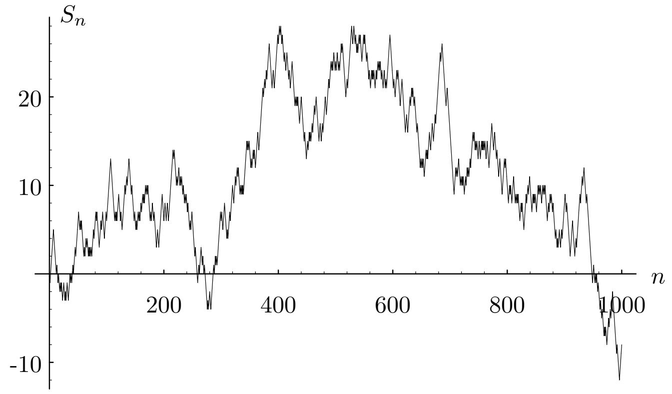
Auf diese Weise wird eine bijektive Abbildung von der gesuchten Menge von Pfaden auf die Menge der Pfade von $(-1,-1)$ nach $(2p+1,-1)$, welche die Bedingung $s_0 > -1$, $s_1 > -1, \ldots, s_{2p} > -1$ erfüllen, hergestellt. Die Anzahl der Pfade in dieser Menge ist gleich der Anzahl der Pfade von $(0,0)$ nach $(2p+2,0)$ mit $s_1 > 0$, $s_2 > 0, \ldots, s_{2p+1} > 0$ (Verschiebung des Ursprungs). (b) folgt dann aus (a). $\square$

Aus (b) des obigen Satzes folgt, dass bei Gleichstand der Stimmen mit Wahrscheinlichkeit $1/(p+1)$ der Kandidat $K_2$ zu keinem Zeitpunkt der Auszählung führt. Das Gleiche gilt auch für den Kandidaten $K_1$. Mit Wahrscheinlichkeit $2/(p+1)$ wechselt somit die Führung nie.

Wir wenden uns nun der Situation (II) zu, das heißt dem Zufallsexperiment, das durch die Gleichverteilung auf $\Omega_n$ beschrieben wird. Dies ist nichts anderes als eine Umformulierung unseres Münzwurfexperiments mit Werten $-1$, $1$. Einem Element $(a_1, \ldots, a_n) \in \{-1,1\}^n$ können wir einen Pfad $(s_0 = 0, s_1, \ldots, s_n) \in \Omega_n$ durch $s_k = \sum_{j=1}^k a_j$, $1 \leq k \leq n$, zuordnen. Dies definiert eine bijektive Abbildung $\{-1,1\}^n \to \Omega_n$. Der Gleichverteilung auf $\{-1,1\}^n$ entspricht dann via dieser bijektiven Abbildung die Gleichverteilung auf $\Omega_n$.

Nachfolgend sind zwei Simulationen einer derartigen Irrfahrt mit $n = 1000$ abgebildet. Aus dem Gesetz der großen Zahlen folgt, dass zum Beispiel $S_{1000}/1000$ mit großer Wahrscheinlichkeit nahe bei 0 liegt. Um etwas zu „sehen“, müssen wir die $y$-Achse gegenüber der $x$-Achse strecken.






Zunächst betrachten wir für $k \leq n$ das Ereignis $A_k = \{S_k = 0\}$. $A_k$ ist das unmögliche Ereignis, falls $k$ ungerade ist. Wir betrachten also $A_{2k}$, $2k \leq n$. Um die Anzahl der Pfade der Länge $n$ zu bestimmen, die zu $A_{2k}$ gehören, multiplizieren wir die Anzahl der Pfade der Länge $2k$ von $(0,0)$ nach $(2k,0)$ mit der Anzahl der Pfade der Länge $n-2k$, die in $(2k,0)$ starten (bei freiem Ende). Somit ist

$$|A_{2k}| = \binom{2k}{k} 2^{n-2k}.$$

$\Omega_n$ enthält $2^n$ Elemente. Also gilt

$$P(A_{2k}) = \binom{2k}{k} 2^{-2k}.$$

Wir kürzen diese Größe auch mit $u_{2k}$ ab ($u_0 = 1$). Man sieht zunächst nicht, von welcher Größenordnung $u_{2k} = P(A_{2k})$ für große $k$ ist. Da

$$u_{2k} = \frac{(2k)!}{(k!)^2} 2^{-2k}$$

ist, benötigen wir eine genauere Kenntnis des Verhaltens der Fakultätsfunktion für große Argumente. Diese erhält man über die Stirling-Approximation:

$$\lim_{n\to\infty} n!/(\sqrt{2\pi}n^{n+1/2}e^{-n}) = 1. \tag{8.3}$$

Für zwei reelle Zahlenfolgen $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, mit $a_n, b_n > 0$ schreiben wir $a_n \sim b_n$, sofern

$$\lim_{n\to\infty} a_n/b_n = 1$$

gilt. Dies bedeutet keineswegs, dass $|a_n - b_n|$ gegen 0 konvergiert. So gilt etwa

$$\lim_{n\to\infty} |n! - \sqrt{2\pi}n^{n+1/2}e^{-n}| = \infty.$$

Setzen wir die Stirling-Approximation ein, so erhalten wir

**8.10 Satz** *Es gilt*

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}}.$$

Satz 8.10 ist eine recht gute Näherung für $u_{2k}$. Um dies genauer zu diskutieren, brauchte man gute Abschätzungen für die Differenz $n! - \sqrt{2\pi} n^{n+1/2} e^{-n}$. Wir wollen diesen Punkt jedoch nicht weiter verfolgen.

Interessanterweise lassen sich die Wahrscheinlichkeiten einer Reihe anderer Ereignisse in Beziehung zu $u_{2k}$ setzen. Es sei zunächst für $k \in \mathbb{N}$ $f_{2k}$ die Wahrscheinlichkeit, dass die erste Nullstelle der Irrfahrt nach dem Zeitpunkt 0 die Zeitkoordinate $2k$ hat, das heißt

$$f_{2k} = P(S_1 \neq 0,\, S_2 \neq 0, \ldots, S_{2k-1} \neq 0,\, S_{2k} = 0).$$

**8.11 Lemma**

(a) $f_{2k} = \frac{1}{2k} u_{2k-2} = P(S_1 \geq 0,\, S_2 \geq 0, \ldots, S_{2k-2} \geq 0,\, S_{2k-1} < 0) = u_{2k-2} - u_{2k}$.

(b) $u_{2k} = P(S_1 \neq 0,\, S_2 \neq 0, \ldots, S_{2k} \neq 0) = P(S_1 \geq 0,\, S_2 \geq 0, \ldots, S_{2k} \geq 0)$.

(c) $u_{2k} = \sum_{j=1}^{k} f_{2j} u_{2k-2j}$.

*Beweis:*

(a) Nach Satz 8.9 (a) gibt es $\frac{1}{k}\binom{2k-2}{k-1}$ Pfade von $(0,0)$ nach $(2k,0)$ mit $s_1 > 0, \ldots, s_{2k-1} > 0$ und natürlich genauso viele mit $s_1 < 0, \ldots, s_{2k-1} < 0$. Es folgt

$$f_{2k} = \frac{2}{k}\binom{2k-2}{k-1} 2^{-2k} = \frac{1}{2k}\binom{2k-2}{k-1} 2^{-2(k-1)} = \frac{1}{2k} u_{2k-2}.$$

Wir beweisen die nächste Gleichung: Falls $s_{2k-2} \geq 0$ und $s_{2k-1} < 0$ sind, so gelten $s_{2k-2} = 0$ und $s_{2k-1} = -1$. Die Anzahl der Pfade von $(0,0)$ nach $(2k-1,-1)$ mit $s_1 \geq 0, \ldots, s_{2k-3} \geq 0$, $s_{2k-2} = 0$ ist gleich der Anzahl der Pfade von $(0,0)$ nach $(2k-2,0)$ mit allen $y$-Koordinaten $\geq 0$. Die zweite Gleichung in (a) folgt dann mit Hilfe von Satz 8.9, (b). Die dritte ergibt sich aus

$$u_{2k} = \binom{2k}{k} 2^{-2k} = \frac{2k(2k-1)}{k \cdot k}\binom{2k-2}{k-1} \cdot \frac{1}{4} \cdot 2^{-2k+2} = \left(1 - \frac{1}{2k}\right) u_{2k-2}.$$

(b) $C_{2j}$ sei das Ereignis $\{S_1 \neq 0,\, S_2 \neq 0, \ldots, S_{2j-1} \neq 0,\, S_{2j} = 0\}$. Diese Ereignisse schließen sich gegenseitig aus und haben Wahrscheinlichkeiten

$f_{2j} = u_{2j-2} - u_{2j}$. Somit ist mit $u_0 = 1$

$$\begin{aligned} P(S_1 \neq 0,\, S_2 \neq 0, \ldots, S_{2k} \neq 0) &= 1 - P\Big(\bigcup_{j=1}^{k} C_{2j}\Big) \\ &= 1 - \sum_{j=1}^{k} (u_{2j-2} - u_{2j}) = u_{2k}. \end{aligned}$$

Die zweite Gleichung folgt analog aus der dritten Identität in (a).

(c) Für $1 \leq j \leq k$ sei $B_j = \{S_1 \neq 0,\, S_2 \neq 0, \ldots, S_{2j-1} \neq 0,\, S_{2j} = 0,\, S_{2k} = 0\}$. Diese Ereignisse sind paarweise disjunkt, und ihre Vereinigung ist $\{S_{2k} = 0\}$. $|B_j|$ ist offenbar gleich der Anzahl der Pfade von $(0,0)$ nach $(2j,0)$, die die $x$-Achse dazwischen nicht berühren, multipliziert mit der Anzahl aller Pfade von $(2j,0)$ nach $(2k,0)$, das heißt $|B_j| = 2^{2j} f_{2j} 2^{2k-2j} u_{2k-2j}$. Somit gilt $P(B_j) = f_{2j} u_{2k-2j}$, das heißt

$$u_{2k} = \sum_{j=1}^{k} P(B_j) = \sum_{j=1}^{k} f_{2j} u_{2k-2j}.$$

$\square$

Eine interessante Folgerung ergibt sich aus der ersten Gleichung in (b). Da nach Satz 8.10 $\lim_{k\to\infty} u_{2k} = 0$ gilt, folgt, dass die Wahrscheinlichkeit für keine Rückkehr der Irrfahrt bis zum Zeitpunkt $2k$ mit $k \to \infty$ gegen 0 konvergiert. Man kann das folgendermaßen ausdrücken: „Mit Wahrscheinlichkeit 1 findet irgendwann eine Rückkehr statt.“ Man sagt auch, die Irrfahrt sei *rekurrent*. Wir wollen das noch etwas genauer anschauen und bezeichnen mit $T$ den Zeitpunkt der ersten Nullstelle nach dem Zeitpunkt 0. $T$ muss gerade sein, und es gilt $P(T = 2k) = f_{2k}$. Aus (a) und $u_{2k} \to 0$ folgt

$$\sum_{k=1}^{\infty} f_{2k} = \lim_{N\to\infty} \sum_{k=1}^{N} f_{2k} = \lim_{N\to\infty} \sum_{k=1}^{N} (u_{2k-2} - u_{2k}) = \lim_{N\to\infty} (u_0 - u_{2N}) = 1.$$

Wir sehen also, dass $(f_{2k})_{k\in\mathbb{N}}$ eine Wahrscheinlichkeitsverteilung auf den geraden natürlichen Zahlen definiert, die Verteilung von $T$. Daraus lässt sich der Erwartungswert von $T$ berechnen:

$$\mathbb{E}T = \sum_{k=1}^{\infty} 2k f_{2k} = \sum_{k=1}^{\infty} u_{2k-2},$$

wobei wir die Gleichung (a) in Lemma 8.11 anwenden. Nach Satz 8.10 divergiert jedoch diese Reihe! Man kann auch sagen, dass $\mathbb{E}T$ gleich $\infty$ ist.

Mit Wahrscheinlichkeit 1 findet also ein Ausgleich statt; man muss jedoch im Schnitt unendlich lange darauf warten.

Obgleich $P(S_1 \neq 0, \ldots, S_{2k} \neq 0) = P(S_1 \geq 0, \ldots, S_{2k} \geq 0) \sim 1/\sqrt{\pi k}$ gegen 0 konvergiert, ist diese Wahrscheinlichkeit erstaunlich groß. Wieso erstaunlich? Wir betrachten das Ereignis $F_j^{(k)}$, dass die Irrfahrt während genau $2j$ Zeiteinheiten bis $2k$ positiv ist. Aus formalen Gründen präzisieren wir „positiv sein" wie folgt: Die Irrfahrt ist positiv im Zeitintervall von $l$ bis $l+1$, falls $S_l$ *oder* $S_{l+1} > 0$ ist. Es kann also auch $S_l = 0,\ S_{l+1} > 0$ oder $S_l > 0,\ S_{l+1} = 0$ sein. Man überzeugt sich leicht davon, dass die Anzahl der Intervalle, wo dieses der Fall ist, gerade ist. $F_k^{(k)}$ ist natürlich das Ereignis $\{S_1 \geq 0,\ S_2 \geq 0, \ldots, S_{2k} \geq 0\}$. Aus Gründen der Symmetrie ist $P(F_0^{(k)}) = P(F_k^{(k)})$, was nach Lemma 8.11 (b) gleich $u_{2k} \sim 1/\sqrt{\pi k}$ ist.

Die $F_j^{(k)}$ sind für $0 \leq j \leq k$ paarweise disjunkt, und es gilt

$$\sum_{j=0}^{k} P(F_j^{(k)}) = 1.$$

Mithin können nicht allzuviele der $P(F_j^{(k)})$ von derselben Größenordnung wie $P(F_k^{(k)})$ sein, denn sonst müsste die obige Summe $> 1$ werden. Anderseits ist wenig plausibel, dass unter diesen Wahrscheinlichkeiten gerade $P(F_k^{(k)})$ und $P(F_0^{(k)})$ besonders groß sind. Genau dies ist jedoch der Fall, wie aus dem folgenden bemerkenswerten Resultat hervorgehen wird.

**8.12 Satz (von Chung und Feller)** *Für* $0 \leq j \leq k$ *gilt*

$$P(F_j^{(k)}) = u_{2j} u_{2k-2j}.$$

*Beweis:* Wir führen einen Induktionsschluss nach $k$. Für $k = 1$ gilt

$$P(F_0^{(1)}) = P(F_1^{(1)}) = \frac{1}{2} = u_2.$$

Wir nehmen nun an, die Aussage des Satzes sei bewiesen für alle $k \leq n - 1$, und beweisen sie für $k = n$.

Wir hatten in Lemma 8.11 (b), schon gesehen, dass $P(F_0^{(n)}) = P(F_n^{(n)}) = u_{2n}$ ist ($u_0$ ist $= 1$). Wir brauchen deshalb nur noch $1 \leq j \leq n-1$ zu betrachten. Zunächst führen wir einige spezielle Mengen von Pfaden ein.

Für $1 \leq l \leq n,\ 0 \leq m \leq n - l$ sei $G_{l,m}^{+}$ die Menge der Pfade der Länge $2n$ mit: $s_0 = 0,\ s_1 > 0,\ s_2 > 0, \ldots, s_{2l-1} > 0,\ s_{2l} = 0$ und $2m$ Strecken des Pfades zwischen den $x$-Koordinaten $2l$ und $2n$ sind positiv.

Analog bezeichne $G_{l,m}^{-}$ für $1 \leq l \leq n,\ 0 \leq m \leq n - l$, die Menge der Pfade mit: $s_0 = 0,\ s_1 < 0,\ s_2 < 0, \ldots, s_{2l-1} < 0,\ s_{2l} = 0$ und $2m$ Strecken des

Pfades zwischen den $x$-Koordinaten $2l$ und $2n$ sind positiv. Die $G^+_{l,m}, G^-_{l,m}$ sind offensichtlich alle paarweise disjunkt. Ferner gilt

$$G^+_{l,m} \subset F^{(n)}_{l+m},\ G^-_{l,m} \subset F^{(n)}_m.$$

Man beachte, dass für $1 \le j \le n-1$ jeder Pfad aus $F^{(n)}_j$ zu genau einer der Mengen $G^+_{l,m}$, $G^-_{l,m}$ gehört. Dies folgt daraus, dass ein solcher Pfad mindestens einmal das Vorzeichen wechseln, also auch die 0 passieren muss. Ist $2l$ die $x$-Koordinate der kleinsten Nullstelle $> 0$, so gehört der Pfad zu $G^+_{l,j-l}$, falls der Pfad vor $2l$ positiv, und zu $G^-_{l,j}$, falls er vor $2l$ negativ ist. Demzufolge ist

$$P(F^{(n)}_j) = \sum_{l=1}^{j} P(G^+_{l,j-l}) + \sum_{l=1}^{n-j} P(G^-_{l,j}).$$

Es bleibt noch die Aufgabe, die Summanden auf der rechten Seite dieser Gleichung zu berechnen.

Offensichtlich enthalten $G^+_{l,m}$ und $G^-_{l,m}$ gleich viele Pfade. $|G^+_{l,m}|$ ist gleich der Anzahl der Pfade von $(0,0)$ nach $(2l,0)$ mit $s_1 > 0$, $s_2 > 0, \ldots, s_{2l-1} > 0$ multipliziert mit der Anzahl der Pfade der Länge $2n-2l$ mit Start in $(2l,0)$ und $2m$ positiven Strecken, das heißt

$$|G^+_{l,m}| = |G^-_{l,m}| = \frac{1}{2} f_{2l} 2^{2l} P(F^{(n-l)}_m) 2^{2n-2l},$$

und

$$P(G^+_{l,m}) = P(G^-_{l,m}) = \frac{1}{2} f_{2l} P(F^{(n-l)}_m).$$

Nach der weiter oben stehenden Gleichung ist also

$$P(F^{(n)}_j) = \frac{1}{2}\sum_{l=1}^{j} f_{2l} P(F^{(n-l)}_{j-l}) + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l} P(F^{(n-l)}_j).$$

Nach Induktionsvoraussetzung ist das

$$P(F^{(n)}_j) = \frac{1}{2}\sum_{l=1}^{j} f_{2l}\, u_{2j-2l}\, u_{2n-2j} + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l}\, u_{2n-2j-2l}\, u_{2j} = u_{2j}\, u_{2n-2j}$$

nach Lemma 8.11 (c). □

Um das Verhalten von $P(F^{(k)}_j)$ für festes $k$ als Funktion von $j$ zu untersuchen, betrachten wir für $1 \le j \le k-1$ die Quotienten

$$\frac{P(F^{(k)}_j)}{P(F^{(k)}_{j+1})} = \frac{\binom{2j}{j}\binom{2k-2j}{k-j}}{\binom{2j+2}{j+1}\binom{2k-2j-2}{k-j-1}} = \frac{(2k-2j-1)(j+1)}{(2j+1)(k-j)}.$$

Dieser Quotient ist $> 1$, $= 1$ oder $< 1$, je nachdem, ob $j < \frac{k-1}{2}$, $j = \frac{k-1}{2}$ oder $j > \frac{k-1}{2}$ ist.

Als Funktion von $j$ fällt also $P(F_j^{(k)})$ für $j < \frac{k-1}{2}$ und steigt an für $j > \frac{k-1}{2}$.

$P(F_0^{(k)}) = P(F_k^{(k)})$ ist also der größte vorkommende Wert und $P(F_{\lceil \frac{k-1}{2} \rceil})$ der kleinste. Es ist bedeutend wahrscheinlicher, dass die Irrfahrt über das ganze betrachtete Zeitintervall positiv ist, als dass sich positive und negative Zahlen ausgleichen. Dies scheint im Widerspruch zum Gesetz der großen Zahlen zu stehen. Ohne dies hier genauer zu diskutieren, sei aber daran erinnert, dass die Rückkehrzeit $T$ nach 0 keinen endlichen Erwartungswert hat, wie wir oben gezeigt haben.

Mit Hilfe von Satz 8.10 lässt sich eine einfach Approximation für $P(F_j^{(k)})$ für große $j$ und $k - j$ gewinnen:

**8.13 Satz** *Für $j \to \infty$, $k - j \to \infty$ gilt $P(F_j^{(k)}) \sim \frac{1}{\pi} \frac{1}{\sqrt{j(k-j)}}$, das heißt*

$$\lim_{j \to \infty, k-j \to \infty} \sqrt{j(k-j)} P(F_j^{(k)}) = \frac{1}{\pi}.$$

Betrachten wir speziell $x \in (0,1)$ so gilt für $j, k \to \infty$ mit $j/k \sim x$

$$P(F_j^{(k)}) \sim \frac{1}{\pi k} \frac{1}{\sqrt{x(1-x)}}.$$

Diese Wahrscheinlichkeiten sind also von der Größenordnung $1/k$, das heißt asymptotisch viel kleiner als

$$P(F_0^{(k)}) = P(F_k^{(k)}) \sim \frac{1}{\sqrt{\pi k}}.$$

Die Funktion $(x(1-x))^{-1/2}$ hat für $x = 0$ und 1 Pole. Das steht in Übereinstimmung damit, dass für $j/k \sim 0$ und $j/k \sim 1$ die Wahrscheinlichkeiten $P(F_j^{(k)})$ von einer anderen Größenordnung als $1/k$ sind.

Eine Aussage wie die in Satz 8.13 nennt man einen *lokalen Grenzwertsatz*, da wir damit Informationen über die Wahrscheinlichkeit, dass der Zeitraum der Führung exakt $= 2j$ ist, erhalten. Da diese Wahrscheinlichkeiten jedoch alle für große $k$ klein werden, interessiert man sich eher zum Beispiel für die Wahrscheinlichkeit, dass der relative Anteil der Zeit, wo die Irrfahrt positiv ist, $\geq \alpha$ ist.

Es seien $0 < \alpha < \beta < 1$. $\gamma_k(\alpha, \beta)$ sei die Wahrscheinlichkeit, dass dieser relative Anteil der Zeit zwischen $\alpha$ und $\beta$ liegt. Genauer: $T_k$ sei (die auf $\Omega_{2k}$

definierte) Zufallsgröße, die die Dauer der Führung zählt:

$$T_k := \sum_{j=1}^{2k} 1_{\{S_{j-1}\geq 0,\, S_j \geq 0\}}.$$

Dann ist

$$\gamma_k(\alpha,\beta) := P\Big(\alpha \leq \frac{T_k}{2k} \leq \beta\Big) = \sum_{j:\alpha\leq\frac{j}{k}\leq\beta} P(F_j^{(k)}).$$

Wir sind übrigens nun bei der in Satz 8.5 diskutierten Abbildung $g(Y_n)$ angekommen, denn $T_k = 2k\, g(Y_{2k})$. Wir wollen nun aus Satz 8.13 für $k \to \infty$ folgern:

$$\gamma_k(\alpha,\beta) \sim \frac{1}{\pi} \sum_{j:\alpha\leq\frac{j}{k}\leq\beta} \frac{1}{k} \frac{1}{\sqrt{\frac{j}{k}\big(1-\frac{j}{k}\big)}}. \tag{8.4}$$

Die rechte Seite ist nichts anderes als die RIEMANN-Approximation für

$$\int_\alpha^\beta \frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}}dx = \frac{2}{\pi}(\arcsin\sqrt{\beta} - \arcsin\sqrt{\alpha}).$$

Es folgt nun (und damit Satz 8.5):

**8.14 Satz (Arcus-Sinus-Gesetz)**

$$\lim_{k\to\infty}\gamma_k(\alpha,\beta) = \frac{2}{\pi}(\arcsin\sqrt{\beta} - \arcsin\sqrt{\alpha}).$$

*Beweis:* Wir müssen (8.4) zeigen. Wir schreiben die STIRLING-Approximation als $n! = \sqrt{2\pi n}\big(\frac{n}{e}\big)^n F(n)$ mit $\lim_{n\to\infty} F(n) = 1$. Es folgt

$$\begin{aligned} P(F_j^{(k)}) &= \binom{2j}{j}\binom{2k-2j}{k-j}\frac{1}{2^{2k}} \\ &= \frac{1}{\pi}\frac{1}{\sqrt{(\frac{j}{k})(1-(\frac{j}{k}))}}\frac{1}{k}\frac{F(2j)\,F(2(k-j))}{F(j)\,F(j)\,F(k-j)\,F(k-j)}. \end{aligned}$$

Wir wählen nun ein $\delta > 0$ mit $0 < \delta < 1/2$ und betrachten für jedes $k$ nur die Werte $j$ für die gilt

$$\delta \leq \frac{j}{k} \leq 1-\delta,$$

womit $k\delta \leq j$ und $k\delta \leq k-j$ folgt. Für $k \to \infty$ konvergiert nun jedes $F(j), F(k-j), F(2j), F(2(k-j))$ gleichmäßig für alle obigen Werte von $j$. Somit existiert für $\delta \leq \alpha < \beta \leq 1-\delta$ ein $G_{\alpha,\beta}(k)$ für jedes $k = 1, 2, \ldots$, so dass für jedes obige $\delta > 0$ gilt:

$$\lim_{k\to\infty} G_{\alpha,\beta}(k) = 1 \quad \text{gleichmäßig für} \quad \delta \leq \alpha < \beta \leq 1-\delta$$

und

$$\sum_{\alpha \le \frac{j}{k} \le \beta} P(F_j^{(k)}) = \left( \frac{1}{k} \sum_{\alpha \le \frac{j}{k} \le \beta} \frac{1}{\pi} \frac{1}{\sqrt{(j/k)(1-(j/k))}} \right) G_{\alpha,\beta}(k).$$

Nun folgt die Behauptung gleichmäßig für $\delta \le \alpha < \beta \le 1-\delta$, wie auch immer $0 < \delta < 1/2$ gewählt war. Damit folgt die Behauptung. □

**8.15 Bemerkung** Die Aussage von Satz 8.14 ist auch richtig für $\alpha = 0$ oder $\beta = 1$. Das heißt etwa, dass $\gamma_k(0,\beta)$ — die Wahrscheinlichkeit dafür, dass der relative Anteil der Zeit, in der $K_1$ führt, $\le \beta$ ist — gegen $\frac{2}{\pi} \arcsin \sqrt{\beta}$ konvergiert.

*Beweis:* Offensichtlich gilt $\lim_{k\to\infty} \gamma_k\big(0, \frac{1}{2}\big) = 1/2$. Ist $\beta \in (0, 1/2)$, so folgt

$$\lim_{k\to\infty} \gamma_k(0,\beta) = \lim_{k\to\infty} (\gamma_k(0,1/2) - \gamma_k(\beta,1/2)) = \frac{2}{\pi} \arcsin \sqrt{\beta},$$

für $\beta > 1/2$

$$\lim_{k\to\infty} \gamma_k(0,\beta) = \lim_{k\to\infty} (\gamma_k(0,1/2) + \gamma_k(1/2,\beta)) = \frac{2}{\pi} \arcsin \sqrt{\beta}.$$

Für $\gamma_k(\alpha, 1)$ führt dasselbe Argument zum Ziel. □

Der Beweis des Arcus-Sinus-Gesetzes wurde zuerst von P. Lévy im Jahre 1939 gegeben. Die Funktion $\frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$ hat das folgende Aussehen:


Zur Illustration des Arcus-Sinus-Gesetzes diene die folgende Tabelle der sogenannten Arcus-Sinus-Verteilungsfunktion $A(x) = \frac{2}{\pi} \arcsin \sqrt{x}$. Für $x \in (\frac{1}{2}, 1]$ kann $A(x)$ mit der Formel $A(x) = 1 - A(1-x)$ berechnet werden.

$$A(x) = \tfrac{2}{\pi} \arcsin \sqrt{x}$$

| $x$ | $A(x)$ | $x$ | $A(x)$ | $x$ | $A(x)$ |
|---|---|---|---|---|---|
| 0.00 | 0.000 | 0.20 | 0.295 | 0.40 | 0.436 |
| 0.01 | 0.064 | 0.21 | 0.303 | 0.41 | 0.442 |
| 0.02 | 0.090 | 0.22 | 0.311 | 0.42 | 0.449 |
| 0.03 | 0.111 | 0.23 | 0.318 | 0.43 | 0.455 |
| 0.04 | 0.128 | 0.24 | 0.326 | 0.44 | 0.462 |
| 0.05 | 0.144 | 0.25 | 0.333 | 0.45 | 0.468 |
| 0.06 | 0.158 | 0.26 | 0.341 | 0.46 | 0.474 |
| 0.07 | 0.171 | 0.27 | 0.348 | 0.47 | 0.481 |
| 0.08 | 0.183 | 0.28 | 0.355 | 0.48 | 0.487 |
| 0.09 | 0.194 | 0.29 | 0.362 | 0.49 | 0.494 |
| 0.10 | 0.205 | 0.30 | 0.369 | 0.50 | 0.500 |
| 0.11 | 0.215 | 0.31 | 0.376 | | |
| 0.12 | 0.225 | 0.32 | 0.383 | | |
| 0.13 | 0.235 | 0.33 | 0.390 | | |
| 0.14 | 0.244 | 0.34 | 0.396 | | |
| 0.15 | 0.253 | 0.35 | 0.403 | | |
| 0.16 | 0.262 | 0.36 | 0.410 | | |
| 0.17 | 0.271 | 0.37 | 0.416 | | |
| 0.18 | 0.279 | 0.38 | 0.423 | | |
| 0.19 | 0.287 | 0.39 | 0.429 | | |

KAPITEL A

# Beweis des Satzes von Prohorov

Inspiziert man den Beweis von Satz 6.20 (Cramér-Wold) und den von Satz 7.5 und 7.7 (Satz von Donsker), so haben wir nur die folgende einfachere Version des Satzes von Prohorov verwendet:

Ist die Folge $(\mu_n)_n$ von W-Maßen auf einem vollständigen, separablen metrischen Raum straff, so hat $(\mu_n)_n$ eine schwach konvergente Teilfolge, ist also insbesondere relativ kompakt. Ist die Folge $(\mu_n)_n$ schwach konvergent, so ist sie auch straff (Satz 6.20). Wir beweisen daher auch nur diese Variante. Für einen Beweis von Satz 6.20 in seiner vollen Allgemeinheit verweisen wir auf Billingsley, „Convergence of Probability measures", oder Parthasarathy, „Probability measures on metric spaces".

**A.1 Satz** *Es sei $S$ ein separabler metrischer Raum und $(\mu_n)_n$ eine Folge von W-Maßen auf $(S, \mathcal{B}_S)$, die straff ist. Dann hat $(\mu_n)_n$ eine schwach konvergente Teilfolge. Ist $S$ ein vollständiger, separabler metrischer Raum und konvergiert $(\mu_n)_n$ schwach, dann ist $\{\mu_n,\, n \in \mathbb{N}\}$ straff.*

Alle Beweise in der Literatur verwenden tiefere Resultate der Analysis, so auch der hier vorgestellte Beweis.

**A.2 Lemma** *Es sei $S$ ein kompakter metrischer Raum und $(\mu_n)_n$ eine Folge von W-Maßen auf $(S, \mathcal{B}_S)$. Dann hat $(\mu_n)_n$ eine schwach konvergente Teilfolge.*

Der Beweis des Lemmas verwendet die folgende Variante des Darstellungssatzes von Riesz:

Eine Abbildung $\Lambda : C(S) \to \mathbb{R}$ heißt ein *normiertes, nicht-negatives lineares Funktional*, wenn $\Lambda(1) = 1$, $\Lambda(f) \geq 0$ für $f \geq 0$ und $\Lambda(af + bg) = a\Lambda(f) + b\Lambda(g)$ für alle $a, b \in \mathbb{R}$ und $f, g \in C(S)$ gilt.

**A.3 Satz (Darstellungssatz von Riesz)** *Es sei $S$ ein kompakter metrischer Raum. Dann existiert zu jedem normierten, nicht-negativem linearem Funktional $\Lambda : C(S) \to \mathbb{R}$ ein eindeutig bestimmtes W-Maß $\mu$ auf $(S, \mathcal{B}_S)$ mit*

$$\Lambda(f) = \int f \, d\mu \quad \forall f \in C(S) \, .$$

*Jedes W-Maß auf $(S, \mathcal{B}_S)$ bestimmt ein normiertes, nicht-negatives lineares Funktional auf $C(S)$.*

*Beweis:* Z.B. in PARTHASARATHY, „Introduction to Probability and Measure“.

*Beweis von Lemma A.2:* Für $f \in C(S)$ sei

$$||f|| := \sup_{x \in S} |f(x)| \, .$$

Da $S$ kompakt ist, ist $C(S)$ ein separabler metrischer Raum; dies folgt aus dem Satz von WEIERSTRASS. Sei $(f_n)_n$ eine dichte Folge in $C(S)$. Mit Hilfe des Diagonalfolgenverfahrens, beschrieben im Anhang zu Kapitel 7, finden wir eine Teilfolge $(\mu_{n_k})_k$ von $(\mu_n)_n$, so dass

$$\lim_{k \to \infty} \int f_j \, d\mu_{n_k} = a_j$$

für alle $j \in \mathbb{N}$ existiert. Zu einem $f \in C(S)$ und $\varepsilon > 0$ sei $f_j$ so gewählt, dass $||f - f_j|| < \varepsilon$. Dann ist

$$\begin{aligned} \Big| \int f \, d\mu_{n_k} - \int f \, d\mu_{n_m} \Big| \le \Big| & \int f_j \, d\mu_{n_k} - \int f_j \, d\mu_{n_m} | \\ & + \int \underbrace{|f - f_j|}_{\le \varepsilon} \, d\mu_{n_k} + \int \underbrace{|f - f_j|}_{\le \varepsilon} \, d\mu_{n_m} \, . \end{aligned}$$

Der erste Summand konvergiert gegen Null für $k, m \to \infty$, also

$$\lim_{k,m \to \infty} \Big| \int f \, d\mu_{n_k} - \int f \, d\mu_{n_m} \Big| = 0 \, ,$$

und somit konvergiert $\int f \, d\mu_{n_k}$ für $k \to \infty$ für jedes $f \in C(S)$. Setzen wir

$$\Lambda(f) := \lim_{k \to \infty} \int f \, d\mu_{n_k} \, , \quad f \in C(S) \, ,$$

so ist $\Lambda$ ein nicht-negatives lineares Funktional auf $C(S)$ mit $\Lambda(1) = 1$, also existiert nach Satz A.3 ein $\mu \in \mathcal{M}_1(S)$ mit

$$\Lambda(f) = \int f \, d\mu \quad \forall f \in C(S) \, ,$$

womit die schwache Konvergenz von $(\mu_{n_k})_k$ gegen $\mu$ folgt. □

Für den Beweis von Satz A.1 bereiten wir vor:

**A.4 Lemma (von Urysohn)** *Ist $S$ ein separabler metrischer Raum, so ist er homöomorph zu einer Teilmenge in* $[0, 1]^{\mathbb{N}}$.

*Beweis:* $d$ bezeichne die Metrik auf $S$ und $(s_n)_n$ eine dichte, abzählbare Teilmenge von $S$. $h : S \to [0,1]^{\mathbb{N}}$ sei definiert durch die $n$-ten Koordinatenfunktionen

$$h_n(x) = \frac{d(x,\, s_n)}{1 + d(x,\, s_n)}\,, \quad x \in S\,,\, n \in \mathbb{N}\,.$$

Es ist eine schöne Übung zu sehen, dass dies ein Homöomorphismus ist. □

Nun ist $[0,1]$ kompakt. Tatsächlich ist auch $[0,1]^{\mathbb{N}}$ kompakt. Ist $(K,\, d)$ ein kompakter metrischer Raum, so ist die Metrik $d$ offenbar beschränkt:

$$\sup_{x,y \in K} d(x,\, y) < \infty\,.$$

Auf $K^{\mathbb{N}}$ definieren wir

$$\bar{d}(x,\, y) := \sum_{i=1}^{\infty} \frac{d(x_i,\, y_i)}{2^i}$$

für $x = (x_i)_i$ und $y = (y_i)_i$. Dann ist $\bar{d}$ eine Metrik und eine Folge in $(K^{\mathbb{N}},\, \bar{d})$ konvergiert genau dann, wenn alle ihre Komponenten konvergieren. Es gilt

**A.5 Satz (von Tychonov)** *$(K^{\mathbb{N}},\, \bar{d})$ ist kompakt.*

*Beweis(skizze):* (Diagonalfolgen-Verfahren) Sei $(x^n)_n = \big((x_i^n)_i\big)_n$ eine Folge in $K^{\mathbb{N}}$. Wir wählen eine Teilfolge $(x^{n_{1,m}})_m$, so dass $(x_1^{n_{1,m}})_m$ konvergiert, dann eine Teilfolge $(x^{n_{2,m}})_m$ dieser Folge, so dass $(x_2^{n_{2,m}})_m$ konvergiert, etc. Dann konvergiert $(x^{n_{m,m}})_m$. □

Also ist nach Lemma A.4 und Satz A.5 ein separabler metrischer Raum homöomorph zu einer Teilmenge eines kompakten metrischen Raumes. Wir gehen zu dieser Teilmenge über und wählen die Relativtopologie (eine Menge $O$ ist offen, wenn $O = [0,1]^{\mathbb{N}} \cap O'$ mit $O'$ offen in $[0,1]^{\mathbb{N}}$ gilt).

*Beweis von Satz A.1:* Wir fassen den separablen metrischen Raum entsprechend der Vorbetrachtung als Teilmenge eines kompakten metrischen Raumes $\tilde{S}$ auf. Für $\mu \in \mathcal{M}_1(S)$ definieren wir $\tilde{\mu} \in \mathcal{M}_1(\tilde{S})$ durch

$$\tilde{\mu}(A) := \mu(A \cap S)\,, \quad A \in \mathcal{B}_{\tilde{S}}\,.$$

Mit Lemma A.2 hat $(\tilde{\mu}_n)_n$ eine konvergente Teilfolge $(\tilde{\mu}_{n_k})_k$, die schwach gegen ein W-Maß $\nu$ auf $\tilde{S}$ konvergiert. Für $r \in \mathbb{N}$ wähle eine kompakte Menge $K_r \subset S$ mit

$$\mu_{n_k}(K_r) \geq 1 - \frac{1}{r} \quad \forall k\,.$$

Da $K_r$ kompakt in $S$ ist, ist $K_r$ kompakt in $\tilde{S}$, also auch in $\mathcal{B}_{\tilde{S}}$ und

$$\tilde{\mu}_{n_k}(K_r) = \mu_{n_k}(K_r) \quad \text{für } r, k \in \mathbb{N}\,.$$

Nach dem Satz von Portmanteau gilt

$$\limsup_{n\to\infty} \tilde{\mu}_{n_k}(K_r) \leq \nu(K_r)\,, \quad r \in \mathbb{N}\,.$$

Dann folgt auch $\nu(K_r) \geq 1 - 1/r$ für $r \in \mathbb{N}$.

Sei $E_0 := \bigcup_r K_r$, dann ist $E_0 \subset S$, $E_0 \in \mathcal{B}_{\tilde{S}}$ und $\nu(E_0) = 1$. Wir behaupten nun, dass es ein $\mu \in \mathcal{M}_1(S)$ gibt mit $\tilde{\mu} = \nu$.

Es ist $\mathcal{B}_S = \mathcal{B}_{\tilde{S}} \cap S$. Für jedes $A \in \mathcal{B}_S$ existiert ein $B_1 \in \mathcal{B}_{\tilde{S}}$ mit $A = B_1 \cap S$. Sei $\mu(A) := \nu(B_1)$. Wenn $B_2 \in \mathcal{B}_{\tilde{S}}$ und $A = B_2 \cap S$, dann ist $B_1 \Delta B_2 \subset S^c \subset E_0^c$ und $\nu(B_1 \Delta B_2) = 0$, also $\nu(B_1) = \nu(B_2)$, also ist $\mu(A)$ wohldefiniert.

Es sei nun $(A_i)_i$ mit $A_i = B_i \cap S$, $i \in \mathbb{N}$, eine Folge von disjunkten Mengen mit $B_i \in \mathcal{B}_{\tilde{S}}$, $i \in \mathbb{N}$. Da $B_i \cap E_0 \subset B_i \cap S$ für alle $i$, sind die $B_i \cap E_0$ auch disjunkt. Also

$$\begin{aligned}\mu\Big(\bigcup_i A_i\Big) &= \nu\Big(\bigcup_i B_i\Big) = \nu\Big(\bigcup_i (B_i \cap E_0)\Big)\\ &= \sum_i \nu(B_i \cap E_0) = \sum_i \nu(B_i) = \sum_i \mu(A_i)\,.\end{aligned}$$

Also ist $\mu$ ein W-Maß mit $\tilde{\mu} = \nu$.

Sei $C$ eine abgeschlossene Menge in $S$. Dann existiert ein $D$ abgeschlossen in $\tilde{S}$ mit $C = D \cap S$. Da $\tilde{\mu}_{n_k} \xrightarrow{w} \tilde{\mu}$, folgt

$$\limsup_{k\to\infty} \mu_{n_k}(C) = \limsup_{k\to\infty} \tilde{\mu}_{n_k}(D) \leq \tilde{\mu}(D) = \mu(C)\,.$$

Der Satz von Portmanteau liefert $\mu_{n_k} \xrightarrow{w} \mu$. Damit ist der erste Teil des Satzes bewiesen.

Sei nun $S$ vollständig und separabel und $\mu_n \xrightarrow{w} \mu$. Da $S$ separabel ist, existiert eine Folge offener Bälle $B_{n1}, B_{n2}, \ldots$ mit Radius $1/n$, so dass

$$S = \bigcup_{j=1}^{\infty} B_{nj}\,, \quad n \in \mathbb{N}\,.$$

Wir zeigen nun, dass für jedes $\delta > 0$ ein $k_n \in \mathbb{N}$ existiert mit

$$\mu_i\Big(\bigcup_{j=1}^{k_n} B_{nj}\Big) > 1 - \delta\,, \quad i \in \mathbb{N}\,.$$

Angenommen, dies stimmt nicht. Also existiert ein $\delta_0 > 0$ und Folgen $i_1 < i_2 < \cdots$ und $k_1 < k_2 < \cdots$ mit

$$\mu_{i_m}\Big(\bigcup_{j=1}^{k_m} B_{nj}\Big) \leq 1 - \delta_0 \quad \text{für } m = 1, 2, \ldots$$

Es gilt $\bigcup_{j=1}^{k_r} B_{nj} \subset \bigcup_{j=1}^{k_m} B_{nj}$ für $m \geq r$, also

$$\mu_{i_m}\Big(\bigcup_{j=1}^{k_r} B_{nj}\Big) \leq \mu_{i_m}\Big(\bigcup_{j=1}^{k_m} B_{nj}\Big) \leq 1 - \delta_0$$

für $m \geq r$.

Da $\mu_{i_m} \xrightarrow{w} \mu$ und $\bigcup_{j=1}^{k_r} B_{nj}$ offen, sagt der Satz von Portmanteau

$$\mu\Big(\bigcup_{j=1}^{k_r} B_{nj}\Big) \leq \liminf_{m\to\infty} \mu_{i_m}\Big(\bigcup_{j=1}^{k_r} B_{nj}\Big) \leq 1 - \delta_0\,.$$

Für $r \to \infty$ folgt $\mu(S) \leq 1 - \delta_0$. Ein Widerspruch!

Sei nun $n \in \mathbb{N}$ fest und $\delta = \varepsilon/2^n$ und $k_n$ so gewählt, dass

$$\mu_i\Big(\bigcup_{j=1}^{k_n} B_{nj}\Big) > 1 - \frac{\varepsilon}{2^n}\,, \quad i \in \mathbb{N}\,.$$

Sei $C_n := \bigcup_{j=1}^{k_n} \bar{B}_{nj}$ und $K := \bigcap_{n=1}^{\infty} C_n$. Dann folgt $\mu_i(K) > 1 - \varepsilon$ für alle $i \in \mathbb{N}$. Tatsächlich ist $K$ kompakt:

Da die $C_n$ abgeschlossen sind, ist auch $K$ abgeschlossen. $(x_n)_n$ sei eine Folge in $K$. Da $K \subset C_1$, existiert ein $n_1 \leq k_1$, so dass $K \cap \bar{B}_{1n_1} =: K_1$ unendlich viele der $x_i$ enthält. Da $K_1 \subset C_2$, existiert ein $n_2 \leq k_2$, so dass $K_1 \cap \bar{B}_{2n_2} =: K_2$ unendlich viele der $x_i$ enthält. Wir gelangen so zu einer Kette $K_1 \supset K_2 \supset \ldots$, und jedes $K_j$ enthält unendlich viele der $x_i$. Nun ist $K_j \subset \bar{B}_{jn_j}$, also ist der Durchmesser von $K_j$ kleiner-gleich $2/j$, $j \in \mathbb{N}$. Nun liefert die Vollständigkeit von $S$

$$\bigcap_{j=1}^{\infty} K_j = \{x_0\}\,, \quad x_0 \in S\,.$$

Nun enthält ein Ball um $x_0$ ein $K_j$ für $j$ hinreichend groß, also enthält der Ball unendlich viele der $x_i$. $x_0$ ist also Limespunkt der Folge $(x_n)_n$, also ist $K$ kompakt und der Satz ist bewiesen. $\square$

# Literaturverzeichnis


[1] G. Ahlsmeyer. *Stochastische Prozesse*. Skripten zur Mathematischen Statistik, Universität Münster, 2002.

[2] N. Alon, J.H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization, With an appendix on the life and work of Paul Erdős, Wiley-Interscience, John Wiley & Sons, New York, 2000. ISBN 0-471-37046-0.

[3] H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter Lehrbuch, Walter de Gruyter & Co., Berlin, fifth edition, 2002. ISBN 3-11-017236-4.

[4] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. ISBN 0-471-00710-2.

[5] P. Billingsley. *Ergodic theory and information*. Reprint of the 1965 original, Robert E. Krieger Publishing Co., Huntington, N.Y., 1978. ISBN 0-88275-666-4.

[6] B. Bollobás. *Random graphs*. Cambridge Studies in Advanced Mathematics, 73, Cambridge University Press, Cambridge, 2001. ISBN 0-521-80920-7; 0-521-79722-5.

[7] B. Bollobás. *Modern graph theory*. Graduate Texts in Mathematics, 184, Springer-Verlag, New York, 1998. ISBN 0-387-98488-7.

[8] L. Breiman. *Probability*. Addison-Wesley Publishing Company, Reading, Mass., 1968.

[9] R. Diestel. *Graphentheorie*. Springer-Lehrbuch, Springer-Verlag, Berlin, 1996. ISBN 3-540-60918-0.

[10] P. Eichelsbacher. *Wahrscheinlichkeitstheorie*. Vorlesungsskript, WS 2004/2005, 2005.

[11] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[12] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[13] H.-O. Georgii. *Stochastik.* Walter de Gruyter, Berlin, 2002.

[14] G. Grimmett. *Percolation.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1999. ISBN 3-540-64902-6.

[15] S. Janson, T. Łuczak and A. Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. ISBN 0-471-17541-2.

[16] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[17] A. Klenke. *Wahrscheinlichkeitstheorie.* Springer-Verlag, second edition, 2008. ISBN 978-3-540-76317-8.

[18] D. W. Stroock. *Probability theory, an analytic view.* Cambridge University Press, Cambridge, 1993. ISBN 0-521-43123-9.

[19] S. R. S. Varadhan. *Probability theory*, volume 7 of *Courant Lecture Notes in Mathematics.* New York University Courant Institute of Mathematical Sciences, New York, 2001. ISBN 0-8218-2852-2.

[20] D. Williams. *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991. ISBN 0-521-40455-X; 0-521-40605-6.

# Index