Wahrscheinlichkeitstheorie II

Matthias Löwe

1 Große Abweichungen

Schon in der Wahrscheinlichkeitstheorie I haben wir die Frage nach der Konvergenzgeschwindigkeit in den Gesetzen der großen Zahlen gestellt. Diese soll in diesem Kapitel unter geeigneten Voraussetzungen beantwortet werden. Dies führt zu dem sogenannten Satz von Cramér, den dieser 1938 bewies. Es ist der erste (mathematische) Fall eines Prinzips der großen Abweichungen (physikalisch kann man das Boltzmannsche Gesetz $S = k \log W$ als ein Prinzip der großen Abweichungen ansehen).

Wir werden von nun an annehmen, dass die vorgelegten Zufallsvariablen $X_1, X_2, \ldots$ reellwertige, i.i.d. Zufallsvariablen sind mit einer endlichen momenterzeugenden Funktion

$$\varphi(t) := \mathbb{E}e^{tX_1} < \infty \qquad \text{für alle } t \in \mathbb{R} \tag{1}$$

(dass diese wichtig ist, wenn man exponentielle Konvergenz im Gesetz der großen Zahlen zeigen will, liegt auf der Hand, wenn man sich vor Augen führt, dass für so eine Konvergenzgeschwindigkeit in der Herleitung des Schwachen Gesetzes der großen Zahlen am besten die gewöhnliche (quadratische) Chebyshev–Ungleichung durch eine exponentielle ersetzt wird; bei dieser taucht dann automatisch die momenterzeugende Funktion (1) auf). Wir werden uns in diesem Kapitel vor allem damit beschäftigen, den Zusammenhang zwischen $\varphi(a)$ und

$$-I(a) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na)$$

herzustellen. Hierbei schreiben wir – wie immer – $S_n := \sum_{i=1}^n X_i$. In einem ersten Schritt werden wir zeigen, dass dieser Limes überhaupt existiert. Dazu definieren wir

$$\pi_n := \mathbb{P}(S_n \geq na) \tag{2}$$

und bemerken, dass

$$\pi_{m+n} \geq \mathbb{P}(S_m \geq ma, S_{n+m} - S_m \geq na) = \mathbb{P}(S_m \geq ma)\mathbb{P}(S_{n+m} - S_m \geq na) = \pi_m \pi_n$$

aufgrund der Unabhängigkeit und identischen Verteilung der X_i. Definieren wir weiter

$$\gamma_n := \log \mathbb{P}(S_n \geq na), \tag{3}$$

so folgt

Lemma 1.1 *Es gilt*

$$\gamma_{m+n} \geq \gamma_m + \gamma_n \tag{4}$$

uns daraus folgt, dass

$$\frac{\gamma_n}{n} \to \sup_{m\geq 1} \frac{\gamma_m}{m} \qquad \textit{wenn } n \to \infty. \tag{5}$$

Beweis: Offensichtlich gilt $\limsup \frac{\gamma_n}{n} \le \sup_m \frac{\gamma_m}{m}$. Es genügt also zu zeigen, dass für jedes m gilt

$$\liminf \frac{\gamma_n}{n} \ge \frac{\gamma_m}{m}.$$

Halten wir also m fest und schreiben $n = km + l$ mit $0 \le l < m$, erhalten wir unter wiederholter Benutzung der Ungleichung (4),

$$\gamma_n \ge k\gamma_m + \gamma_l.$$

Division durch n ergibt

$$\frac{\gamma_n}{n} \ge \left(\frac{km}{km+l}\right)\frac{\gamma_m}{m} + \frac{\gamma_l}{n}.$$

Schickt man n gegen ∞ und erinnert sich, dass $0 \le l < m$ war, erhält man das Resultat. □

Dieses Lemma impliziert schon, dass

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \ge na) = -I(a)$$

existiert und (natürlich) nicht–positiv ist. Aus der Formel, die wir für den Limes der $\frac{\gamma_n}{n}$ gewonnen haben, folgt

$$\mathbb{P}(S_n \ge na) \le e^{-nI(a)}. \tag{6}$$

Übung 1.2 *Man zeige, dass die folgenden Aussagen äquivalent sind:*

1. $I(a) = \infty$.
2. $\mathbb{P}(X_1 \ge a) = 0$.
3. $\mathbb{P}(S_n \ge na) = 0$ *für alle* n.

Übung 1.3 *Man zeige, dass*

$$I\left(\frac{a+b}{2}\right) \le \frac{1}{2}(I(a) + I(b))$$

gilt, I also konvex ist.

Wir werden nun die oben angekündigte Abschätzung mit einer exponentiellen Chebyshev–Ungleichung durchführen. In der Tat gilt ja durch Anwendung der exponentiellen Chebyshev–Ungleichung (d.h. einer Markov–Ungleichung mit der wachsenden Funktion $g(t) = e^t$) für jedes $t > 0$:

$$\mathbb{P}(S_n \ge na) \le e^{-nta}\varphi(t)^n$$

oder mit $\psi(t) := \log \varphi(t)$

$$\mathbb{P}(S_n \ge na) \le e^{-n(ta-\psi(t))}.$$

Die Abschätzung ist natürlich nur dann gut, wenn die rechte Seite wenigstens kleiner ist als 1, der Exponent also negativ.

Lemma 1.4 *Wenn $a > \mathbb{E}X_1$ gilt und t klein genug ist, gilt $at - \psi(t) > 0$.*

Bemerkung 1.5 *Die Existenz des Erwartungswertes $\mathbb{E}X_1$ folgt aus der Annahme (1).*

Beweis:(von Lemma 1.4) Bemerke, dass $\psi(0) = \log \varphi(0) = 0$. Somit ist in erster Näherung

$$\psi(t) = \psi'(t) + o(t^2)$$

füt $t \to 0$ wie man sich mit einer Taylor–Entwicklung veranschaulicht. Können wir also zeigen, dass

$$\psi'(t) = \frac{\varphi'(t)}{\varphi(t)} \to \mu := \mathbb{E}X_1$$

konvergiert, wenn t gegen 0 geht, so sind wir fertig, denn dann ist

$$at - \psi(t) \sim (a - \mu)t$$

positiv, wenn t klein genug ist. Zunächst zeigen wir, dass die Ableitungen existieren. Sei $F(x) := \mathbb{P}(X_1 \leq x)$. Dann ist

$$\mathbb{E}e^{tX_1} = \int e^{tx}\mathbb{P}_X(dx) = \int e^{tx} dF(x).$$

Da e^{tx} für alle t als integrierbar vorausgesetzt ist, können wir nach dem Satz über dominierte Konvergenz φ' berechnen, indem wir die Ableitung unter das Integral ziehen:

$$\varphi'(t) = \int xe^{tx} dF(x) \qquad \text{für } t \in (0, t_0).$$

Nimmt man den Limes $t \to 0$ und wendet für die $x < 0$ den Satz von der monotonen Konvergenz and für die $x > 0$ den Satz von der dominierten Konvergenz, sieht man, dass $\varphi'(t) \to \mu$ für $t \to 0$. Da nun andererseits $\varphi(t) \to 1$, wenn $t \to 0$, haben wir somit gezeigt, dass $\psi'(t) \to \mu$ gilt, wenn $t \to 0$, was nach der Eingangsbemerkung die Behauptung beweist. $\square$

Nun, da wir eine Schranke für $\mathbb{P}(S_n \geq na)$ gefunden haben, liegt es nahe, diese Schranke zu optimieren, also das Minimum von $-ta + \psi(t)$ zu finden. Dazu bilden wir

$$\frac{d}{dt}\left[ta - \psi(t)\right] = a - \frac{\varphi'(t)}{\varphi(t)}$$

und somit sollte das Minimum (wenn alles gut geht) bei $a = \frac{\varphi'(t)}{\varphi(t)}$ angenommen werden. Um sicherzustellen, dass wirklich alles gut geht, definieren wir

$$F_t(x) = \frac{1}{\varphi(t)} \int_{-\infty}^{x} e^{ty} dF(y).$$

Man beachte, dass $F_t(x)$ eine Verteilungsfunktion ist. Ihr Mittwelwert berechnet sich als

$$\int x dF_t(x) = \frac{1}{\varphi(t)} \int_{-\infty}^{\infty} xe^{tx} dF(x) = \frac{\varphi'(t)}{\varphi(t)}.$$

Differenziert man noch einmal, erhält man

$$\frac{d}{dt}\frac{\varphi'(t)}{\varphi(t)} = \frac{\varphi''(t)}{\varphi(t)} - \left(\frac{\varphi'(t)}{\varphi(t)}\right)^2 = \int x^2 dF_t(x) - \left(\int x dF_t(x)\right)^2 \geq 0.$$

Diese Ungleichung ist sogar strikt, wenn wir annehmen

$$\text{F ist nicht die Dirac-Verteilung in } \mu. \tag{7}$$

Unter (7) ist $\frac{\varphi'(t)}{\varphi(t)}$ strikt wachsend. Da $\frac{\varphi'(0)}{\varphi(0)} = \mu$, zeigt dies, dass für $a > \mu$ höchstens ein t_a existiert, das die Gleichung $a = \frac{\varphi'(t_a)}{\varphi(t_a)}$ löst.

Ein solches t_a ist für uns der wesentliche Punkt, um die korrekte Rate für γ_n zu bestimmen. In der Tat gilt:

Theorem 1.6 *Es sei (X_i) eine Folge von i.i.d. Zufallsvariablen , die (1) und (7) erfüllt. Sei $S_n := \sum_{i=1}^n X_i$. Dann gilt für alle $a > \mathbb{E}X_1$ die folgende Gleichheit:*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -I(a), \tag{8}$$

wobei

$$I(a) := \sup_{t\in\mathbb{R}} [ta - \psi(t)] \tag{9}$$

gilt.

Beweis: Ohne Beschränkung der Allgemeinheit nehmen wir an, dass $a = 0$ und $\mathbb{E}X_1 < 0$ gilt (substituiert man nämlich $X_1 \to X_1 + a$, so ersetzt man auch $\varphi(t)$ durch $e^{at}\varphi(t)$. Mit $I(\cdot)$ – definiert wie in (9) – verschiebt sich dann auch $I(a)$ zu $I(0)$). Wir schreiben in der Folge

$$g := \inf_{t\in\mathbb{R}} \varphi(t)$$

und bemerken, dass

$$I(0) = -\log g \quad \text{mit } I(0) = \infty \quad \text{falls } g = 0$$

gilt.

Nun haben wir oben schon gesehen, dass mithilfe der exponentiellen Chebyschev–Ungleichung für alle positiven t

$$\mathbb{P}(S_n \geq na) \leq e^{-n(ta-\psi(t))} \tag{10}$$

folgt und somit

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \leq -\sup_{t\in\mathbb{R}^+} [ta - \psi(t)]\,. \tag{11}$$

Um das Supremum über die ganze reelle Achse auszudehnen, erinnern wir uns daran, dass nach den Vorüberlegungen φ eine strikt konvexe Funktion war. Es ist offenbar $\varphi'(0) = \mathbb{E}X_1 < 0$ (nach Annahme). Wir unterscheiden drei Fälle, je nachdem, wo $\mathbb{P}$ seine Masse hat.

- $\mathbb{P}(X_1 < 0) = 1$.

  Dann ist $\varphi'(t) = \int xe^{tx}dF(x) < 0$ (wobei F die zu $\mathbb{P}$ gehörige Verteilungsfunktion $F(x) := \mathbb{P}(X_1 \le x)$ ist) für alle $t \in \mathbb{R}$. Somit ist φ strikt fallend. Es ist somit
  $$g = \lim_{t\to\infty} \varphi(t) = \mathbb{P}(X_1 = 0) = 0.$$
  Da auch
  $$\mathbb{P}(S_n \ge 0) = 0$$
  gilt, haben wir in diesem Fall schon (8).

- $\mathbb{P}(X_1 \le 0) = 1$ und $1 \ne \mathbb{P}(X_1 = 0) > 0$.

  Wie oben zeigt man, dass φ strikt fallend ist und
  $$\lim_{t\to\infty} \varphi(t) = g = \mathbb{P}(X_1 = 0) > 0.$$
  Da in diesem Falle
  $$\mathbb{P}(S_n \ge 0) = \mathbb{P}(X_1 = \ldots = X_n = 0) = g^n$$
  gilt, folgt auch hier (8)

- $\mathbb{P}(X_1 < 0) > 1$ und $\mathbb{P}(X_1 > 0) > 0$.

  Dann gilt offenbar $\lim_{t\to\pm\infty} \varphi(t) = \infty$ und da φ wie oben bemerkt strikt konvex ist, gibt es ein eindeutiges τ, so dass φ in τ minimal wird. Für diese τ gilt natürlich $\varphi'(\tau) = 0$ und $\tau > 0$, denn die Ableitung von φ ist in 0 negativ. Somit gehört τ zu den in (10) zulässigen t und es gilt daher
  $$\mathbb{P}(S_n \ge 0) \le \mathbb{E}e^{\tau S_n} = (\varphi(\tau))^n = g^n,$$
  also
  $$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \ge 0) \le \log g.$$

Um zu zeigen, dass $\log g$ auch eine untere Schranke ist, verwenden wir eine Technik, die als Tilten oder exponentielle Maßtransformation bekannt ist. Die Idee hierbei ist es, die zugrunde liegende Verteilung der X_i so zu verschieben, dass der Erwartungswert 0 (also unser a) ist. Dann wissen wir aus den Gesetzen der großen Zahlen, dass sich S_n so wie na verhalten wird. Wir kassieren aber einen "Strafterm" dafür, dass wir die Verteilung geändert haben.

Genauer führen wir eine neue Folge (Y_i) von i.i.d. Zufallsvariablen ein, die die Verteilung
$$G(x) = \frac{1}{g} \int_{-\infty}^{x} e^{\tau y} dF(y),$$
d.h.
$$\frac{dG}{dF}(x) = \frac{1}{g} e^{\tau x}$$
besitzen. G heißt auch die Cramér–Transformierte von F. Bemerke, dass
$$g = \varphi(\tau) = \int_{-\infty}^{\infty} e^{\tau y} dF(y).$$

Wir benötigen nun die folgenden drei Lemmata.

Lemma 1.7 *Es gilt* $\mathbb{E}Y = 0$ *und* $\mathbb{V}Y \in (0, \infty)$.

Beweis: Wir bezeichnen mit $\hat{\varphi}(t) = \mathbb{E}e^{tY}$. Dann erhalten wir für alle $t \in \mathbb{R}$

$$\hat{\varphi}(t) = \int_{\mathbb{R}} e^{tx} dG(x) = \frac{1}{g} \int_{\mathbb{R}} e^{tx} e^{\tau x} dF(x) = \frac{1}{g} \varphi(t + \tau) < \infty.$$

Dies impliziert, dass mit φ auch $\hat{\varphi}$ eine C^∞–Funktion ist. Damit ergibt sich

$$\begin{aligned} \mathbb{E}Y &= \hat{\varphi}'(0) = \frac{1}{g}\varphi'(\tau) = 0 \text{ und} \\ \mathbb{V}Y &= \hat{\varphi}''(0) = \frac{1}{g}\varphi''(\tau) \in (0, \infty). \end{aligned}$$

□

Lemma 1.8 *Es sei* $T_n = \sum_{i=1}^n Y_i$. *Dann gilt*

$$\mathbb{P}(S_n \geq 0) = g^n \mathbb{E}(e^{-\tau T_n} 1_{\{T_n \geq 0\}}).$$

Beweis: Beachtet man, dass

$$\begin{aligned} \mathbb{P}(S_n \geq 0) &= \int_{\sum_{i=1}^n x_i \geq 0} dF(x_1) \dots dF(x_n) \\ &= \int_{\sum_{i=1}^n x_i \geq 0} [g e^{-\tau x_1} dG(x_1)] \dots [g e^{-\tau x_n} dG(x_n)], \end{aligned}$$

so folgt die Behauptung. □

Lemma 1.9 *Es gilt*

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}(e^{-\tau T_n} 1_{\{T_n \geq 0\}}) \geq 0.$$

Beweis: Aufgrund von Lemma 1.7 kann man den Zentralen Grenzwertsatz auf T_n anwenden. Wählen wir nun eine Zahl $C > 0$ so, dass

$$\frac{1}{\sqrt{2\pi}} \int_0^C e^{\frac{-x^2}{2}} dx > \frac{1}{4}$$

gilt, erhalten wir die folgende Schranke

$$\mathbb{E}(e^{-\tau T_n} 1_{\{T_n \geq 0\}}) \geq e^{-\tau C \sqrt{\mathbb{V}Y_1} \sqrt{n}} \mathbb{P}\left(\frac{T_n}{\sqrt{\mathbb{V}Y_1} \sqrt{n}} \in [0, C) \right).$$

Da die Wahrscheinlichkeit rechts für n gegen unendlich gegen eine Zahl $\geq \frac{1}{4}$ konvergiert, folgt die Behauptung. □

Der Beweis des Theorems ergibt sich nun, da aus Lemma 1.8 zusammen mit Lemma 1.9 folgt, dass

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{P}(S_n \geq 0) = \log g + \liminf_{n\to\infty} \frac{1}{n}\mathbb{E}(e^{-\tau T_n} 1_{\{T_n \geq 0\}}) \geq \log g.$$

Dies ist die Aussage des Theorems.

□

Bemerkung 1.10 *Das obige Theorem nennt man auch* **Prinzip der großen Abweichungen**. *Genauer sagt man, die Folge* (S_n) *genügt einem Prinzip der großen Abweichungen mit Geschwindigkeit* n *und Rate (oder Ratenfunktion)* I.

Beispiel 1.11 *Ist* X_1 *normalverteilt zu den Parametern* $\mu = 0$ *und* $\sigma^2 = 1$, *so ist*

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int e^{tx} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x+t)^2} dx\, e^{\frac{1}{2}t^2} = e^{\frac{1}{2}t^2},$$

also

$$I(a) = \sup[ta - t^2/2] = a^2/2.$$

Übung 1.12 *Berechnen Sie die Ratenfunktion, für* X_i *die* $\mathcal{N}(\mu, \sigma^2)$*–verteilt sind.*

Übung 1.13 *Berechnen Sie die folgenden Ratenfunktionen:*

1. *Berechnen Sie die Ratenfunktion für* X_i, *die Poisson–verteilt sind zum Parameter* $\lambda > 0$.
2. *Berechnen Sie die Ratenfunktion für* X_i, *die Bernoulli–verteilt sind zum Parameter* $p = \mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = 0)$.

Die Ratenfunktion hat die folgenden Eigenschaften:

Lemma 1.14 *Unter den Bedingungen von Theorem 1.6 gilt*

1. I *ist von unten–halbstetig und konvex auf* $\mathbb{R}$.
2. I *hat für alle* $L \geq 0$ *kompakte Niveaumengen* $N_L := \{z \in \mathbb{R} : I(z) \leq L\}$.
3. I *is stetig und strikt konvex auf dem Inneren von* $D_I := \{z \in \mathbb{R} : I(z) < \infty\}$
4. $I(z) \geq 0$ *mit* $I(z) = 0$ *genau dann, wenn* $z = \mathbb{E}X_1$.

Beweis:

1. Die Konvexität von I folgt aus der Definition: Für alle $0 \le t \le 1$ und x, y gilt

$$\begin{aligned} tI(x) + (1-t)I(y) & \\ &= \sup_\lambda \{t\lambda x - t\psi(\lambda)\} + \sup_\lambda \{(1-t)\lambda y - (1-t)\psi(\lambda)\} \\ &\ge \sup_\lambda \{(tx + (1-t)y)\lambda - \psi(\lambda)\} \\ &= I(tx + (1-t)y). \end{aligned}$$

   Da
$$\psi(0) = \log \mathbb{E}(1) = 0,$$
   folgt
$$I(x) \ge 0x - \psi(0) = 0$$
   für alle x. Zur Halbstetigkeit bemerken wir, dass für $x \in \mathbb{R}$ und $x_n \to x$ und jedes $\lambda \in \mathbb{R}$ gilt:
$$\liminf_{x_n \to x} I(x_n) \ge \liminf_{x_n \to x} [\lambda x_n - \psi(\lambda)] = \lambda x - \psi(\lambda).$$
   Daraus ergibt sich
$$\liminf_{x_n \to x} I(x_n) \ge \sup_\lambda [\lambda x - \psi(\lambda)] = I(x).$$

2. Die Behauptung ergibt sich aus der Stetigkeit in 3.

3. Die Stetigkeit folgt wiederum aus der Konvexität in 1, die strikte Konvexität der Rate ist eine Folge der strikten Konvexität von φ.

4. Die Nicht-Negativität haben wir schon unter 1. gezeigt. Für die fehlende Aussage beachte, dass aus der Jensenschen Ungleichung
$$\psi(\lambda) \ge E(\log e^{\lambda X_1}) = \lambda \mathbb{E} X_1$$
   folgt, also
$$I(\mathbb{E}X_1) \le \lambda \mathbb{E}X_1 - \lambda \mathbb{E}X_1 = 0.$$

□

Bemerkung 1.15 *Die untere Halbstetigkeit von I ist äquivalent dazu, dass die Niveaumengen abgeschlossen sind.*

Die Konvexität von I impliziert, dass D_I ein Intervall ist.

Bemerkung 1.16 *Die Aussagen von Theorem 1.6 bleibt natürlich wahr, wenn wir die Wahrscheinlichkeiten $\mathbb{P}(S_n \le an)$ für $a < \mathbb{E}X_1$ abschätzen. Dies sieht man leicht durch Übergang von X_1 zu $-X_1$.*

Zum Schluss bemerken wir, dass ein Prinzip der großen Abweichungen natürlich wieder das Gesetz der großen Zahlen zur Folge hat:

Korollar 1.17 *Unter den Bedingungen aus Theorem 1.6 gilt das Starke Gesetz der großen Zahlen für die Folge der* (S_n)

Beweis: Ohne Einschränkung können wir annehmen, dass $\mathbb{E}X_1 = 0$ ist. Man bemerke, dass für jedes $\delta > 0$

$$\mathbb{P}(|S_n| \geq \delta) = \mathbb{P}(S_n \geq \delta) + \mathbb{P}(S_n \leq -\delta)$$

gilt und dass aus Theorem 1.6 und der obigen Bemerkung über die Wahrscheinlichkeit einer unteren Abweichung folgt, dass für genügend große n

$$\mathbb{P}(S_n \geq \delta) \leq e^{-\frac{1}{2}nI(\delta)}$$

und

$$\mathbb{P}(S_n \leq -\delta) \leq e^{-\frac{1}{2}nI(\delta)}$$

gilt. Somit ist

$$\mathbb{P}(|S_n| \geq \delta) \leq 2e^{-\frac{1}{2}nI(\delta)}$$

und daher

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n| \geq \delta) \leq 2 \sum_{n=1}^{\infty} e^{-\frac{1}{2}nI(\delta)} < \infty$$

endlich. Aus dem Borel–Cantelli Lemma folgt daher, dass $|S_n|$ für jedes $\delta > 0$ mit Wahrscheinlichkeit eins nur für endlich viele n größer ist als δ. Dies aber heißt, dass S_n fast sicher gegen 0 – also seinen Erwartungswert – konvergiert. □

Bevor wir das Kapitel über große Abweichungen wieder verlassen, wollen wir noch eine interessante Folgerung aus dem Prinzip der großen Abweichungen für den Münzwurf betrachten, das sogenannte **Das Erdös-Renyi-Gesetz**.

In seinen Vorlesungen soll der ungarische Mathematiker A. Renyi das folgende Experiment durchgeführt haben: Er teilte seine Studenten in zwei Gruppen, von denen in der einen jeder 200 Mal eine faire Münze warf und das Ergebnis notierte, während in der anderen jeder einen 200-fachen Münzwurf "im Kopf" simulierte und notierte. Er zog dann aus den eingesammelten Zetteln einen willkürlich heraus und konnte mit großer Wahrscheinlichkeit sagen, ob die notierte Folge von einem echten oder einem vorgestellten Münzwurf stammte. Was steckt dahinter?

Sei R_m der längste 1-Run in einer Folge von 0en und 1en der Länge m, also

$$R_m := \max\{l - k + 1 : 0 \leq k < l \leq m \text{ und } \frac{S_l - S_k}{l - k} = 1 :\},$$

wobei $S_l = \sum_{i=1}^{l} X_i$ ist und X_i die Ausgänge des Münzwurfs beschreiben. Um die erwartete Größe von R_m zu berechnen setzen wir voraus, dass der längste 1-Run eindeutig ist. In diesem Falle gï¿½be es genau einen 1–Run der Länge R_m. Da es insgesamt (ungefähr) m Positionen gibt, wo dieser starten kann, eine Folge von R_m einsen aber Wahrscheinlichkeit p^{R_m} hat, wäre dann

$$1 = m \cdot p^{R_m}$$

und somit

$$R_m = \frac{\log m}{\log 1/p}.$$

Für $p = \frac{1}{2}$ ergibt dies für $m = 200,\ R_m \sim 7,64$. In der Praxis traut sich selten jemand in seinen "Simulationen" einen 1-Run der Länge sechs oder größer auftauchen zu lassen. Dies gibt ein handliches Kriterium zur Unterscheidung von echten und "gefakten" Münzwurfreihen. Dahinter steht der folgende Sachverhalt:

Satz 1.18 *Sei (X_i) eine Folge von i.i.d. Bernoulli-Zufallsvariablen mit*

$$\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0).$$

Dann gilt

$$\mathbb{P}(\lim \frac{R_m}{\log m} = \frac{1}{\log 1/p}) = 1.$$

Zum Beweis benötigen wir die folgende Version des LDP für Bernoulli-Folgen:

Korollar 1.19 *In der obigen Situation gilt*

$$\lim \frac{1}{n} \log \mathbb{P}(\frac{\sum_{i=1}^n X_i}{n} \in (a,b)) = - \inf_{x \in (a,b)} H(x|p).$$

Hierbei ist

$$H(x|p) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$$

und insbesondere $H(1|p) = \log \frac{1}{p}$.

Beweis des Satzes: Wir führen als Hilfsgröße die Wartezeit bis zum Auftreten des ersten Runs der Länge mindestens r ein:

$$T_r := \inf\{l : \frac{S_l - S_k}{l-k} = 1 \quad \text{für ein} \quad 0 \le k \le l - r\}.$$

Offenbar gilt

$$\{T_r \le m\} = \{R_m \ge r\}.$$

Sei für $l, k \in \mathbb{N},\ l > k$

$$C_{k,l} := \{\frac{S_l - S_k}{l-k} = 1\}.$$

Dann ist

$$\{T_r \le m\} \subseteq \bigcup_{k=0}^{m-r} \bigcup_{l=k+r}^{m} C_{k,l} \subseteq \bigcup_{k=0}^{m-1} \bigcup_{l=k+r}^{\infty} C_{k,l}.$$

Wegen $\mathbb{P}(C_{k,l}) = \mathbb{P}(\frac{S_l - S_k}{l-k} = 1)$ folgt

$$\begin{aligned}
\mathbb{P}(T_r \le m) &\le (m-1)\sum_{n=r}^{\infty} \mathbb{P}(\frac{S_n}{n} = 1) \\
&\le (m-1)\sum_{n=r}^{\infty} e^{-n\log 1/p} \\
&= c(m-1)e^{-r\log 1/p}
\end{aligned}$$

für eine Konstante $c > 0$.

Ist nun $m = \lfloor e^{r(\log 1/p - \varepsilon)} \rfloor$ für ein $\varepsilon > 0$, so folgt

$$\sum_{r=1}^{\infty} \mathbb{P}(T_r \le m) \le c \sum_{r=1}^{\infty} \lfloor e^{r(\log 1/p - \varepsilon)} \rfloor e^{-r\log 1/p} \le c \sum_{r=1}^{\infty} e^{-r\varepsilon} < +\infty.$$

Also ist nach Borel-Cantelli

$$\mathbb{P}(\limsup T_r \le e^{r(\log 1/p - \varepsilon)}) = 0$$

also $\liminf \frac{1}{r}\log T_r \ge \log \frac{1}{p}$ $\mathbb{P}$-f.s. und wegen $\{R_m \ge r\} = \{T_r \le m\}$ bedeutet dies

$$\frac{R_m}{\log m} \le \frac{1}{\log 1/p} \quad \text{für fast alle } m \ \mathbb{P}\text{-f.s.}$$

Für die andere Richtung sei

$$B_l := \{\frac{S_{l\cdot r} - S_{(l-1)\cdot r}}{r} = 1\}.$$

Die (B_l) sind unabhängig mit $\mathbb{P}(B_l) = \mathbb{P}(\frac{S_r}{r} = 1)$. Weiter gilt

$$\bigcup_{l=1}^{\lfloor \frac{m}{r} \rfloor} B_l \subseteq \{T_r \le m\}.$$

Also

$$\begin{aligned}
\mathbb{P}(T_r > m) &\le 1 - \mathbb{P}(\bigcup_{l=1}^{\lfloor m/r \rfloor} B_l) \\
&= \prod_{l=1}^{\lfloor m/r \rfloor} \mathbb{P}(B_l^c) \\
&= (1 - \mathbb{P}(B_1))^{\lfloor m/r \rfloor} \\
&\le \exp(-\frac{m}{r}\mathbb{P}(\frac{S_r}{r} = 1)).
\end{aligned}$$

Ist nun $m = \lfloor e^{r(\log 1/p+\varepsilon)} \rfloor$ für ein $\varepsilon > 0$, so ergibt sich

$$\begin{aligned} & \sum_{r=1}^{\infty} \mathbb{P}(T_r > e^{r(\log 1/p+\varepsilon)}) \\ \leq \quad & \sum_{r=1}^{\infty} \exp(-\frac{e^{r(\log 1/p+\varepsilon)}}{r} e^{-r \log 1/p}) \\ \leq \quad & \sum_{r=1}^{\infty} e^{-c_1 e^{c_2 r}} < +\infty. \end{aligned}$$

Also ist

$$T_r \leq e^{r(\log 1/p+\varepsilon)} \quad \text{für fast alle } r \ \mathbb{P}\text{-f.s.},$$

und somit auch

$$\frac{R_m}{\log m} \geq \frac{1}{\log 1/p}$$

für fast alle m $\mathbb{P}$-f.s.

2 Das Gesetz vom iterierten Logarithmus

Für die Zwecke dieses Kapitels betrachten wir eine Folge $X_1, X_2, \ldots$ unabhängiger, identisch verteilter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Es gelte

$$\mathbb{E}\, X_1 = 0 \quad \text{und} \quad 0 < \sigma^2 := \mathbb{V}\, X_1 < \infty.$$

Wenn wir $S_n := \sum_{i=1}^n X_i$ setzen, so wissen wir nach dem starken Gesetz der großen Zahlen, dass

$$\frac{1}{n} S_n \to 0 \ \mathbb{P}\text{-f.s.}$$

gilt. Aus der Theorie großer Abweichungen wissen wir zudem, dass die Wahrscheinlichkeit dafür, dass $|\frac{1}{n} S_n|$ größer ist als ein Wert $a \in \mathbb{R}^+$, klein ist von der Ordnung $e^{-nI(a)}$, wobei $I(a)$ die Legendre-Transformierte der logarithmischen Momentenerzeugenden Funktion von X_1 ist und somit $I(a) > 0$ für alle $a \neq 0$. Mit anderen Worten: Die Fluktuationen von S_n sind von deutlich kleinerer Ordnung als n. Andererseits besagt der zentrale Grenzwertsatz, dass $\frac{1}{\sqrt{n}} S_n$ schwach gegen die $\mathcal{N}(0, \sigma^2)$-Verteilung konvergiert. Also ist die Wahrscheinlichkeit $\mathbb{P}(\frac{1}{\sqrt{n}} S_n \geq a)$ für jedes $a \in \mathbb{R}^+$ größer als 0. Die Fluktuationen von S_n sind also von unten durch $\sqrt{n}$ beschränkt. Die Frage, die wir uns in diesem Abschnitt stellen wollen, ist die nach der präzisen Größenordnung der Fluktuationen von S_n. Gesucht ist somit eine isotone Folge $(a_n)_n$ mit $a_n \in \mathbb{R}$ und $\lim_{n\to\infty} a_n = \infty$, die das Fluktuationsverhalten von S_n beschreibt. Hierzu betrachten wir

$$Y_n := \frac{1}{a_n} \sum_{i=1}^n X_i$$

und für $a, b \in \mathbb{R}$

$$A = \{\limsup Y_n = a\} \cap \{\liminf Y_n = b\}.$$

$\limsup Y_n$ und $\liminf Y_n$ sind messbar bezüglich der terminalen σ-Algebra der X_i und somit $\mathbb{P}$-f.s. konstant. Somit können wir die gestellte Frage dahingehend präzisieren, dass wir auf der Suche nach einer isotonen Folge $(a_n)_n$ und konstanten $a, b \in \mathbb{R}$ sind mit

$$\mathbb{P}[\limsup Y_n = a] = \mathbb{P}[\liminf Y_n = -b] = 1. \tag{12}$$

Aufgrund des zentralen Grenzwertsatzes kann man darüber hinaus vermuten, dass a und b universell sind, d. h. nur von σ^2 abhängen, nicht aber von der Verteilung von X_1 (dies ist der typische Fall einer Vermutung, die man ausspricht, nachdem man das Ergebnis kennt; man könnte natürlich auch vermuten, dass es sich so verhält wie in der Theorie großer Abweichungen, wo die Rate sehr wohl verteilungsabhängig ist). In diesem Fall kann natürlich nur $a = -b$ gelten, denn die Aussage müsste für $(-X_n)_{n\in\mathbb{N}}$ ebenso gelten wie für die Folge $(X_n)_{n\in\mathbb{N}}$. Natürlich ist die Folge $(a_n)_{n\in\mathbb{N}}$ durch (12) keineswegs festgelegt: Falls für eine weitere Folge (a'_n) gilt $\frac{a'_n}{a_n} \to 1$, so hat $Y'_n := \frac{1}{a'_n} \sum_{i=1}^n X_i$ denselben Limes wie Y_n. Wir wollen also (a_n) und (a'_n) asymptotisch äquivalent nennen (in Zeichen $a_n \sim a'_n$), falls

$$\lim_{n\to\infty} \frac{a_n}{a'_n} = 1$$

gilt. Der folgende Satz beschreibt nun, wie die Folge $(a_n)_n$ und die Konstanten a und b gewählt werden müssen. Dass $a_n = n$ eine zu starke Skalierung ist, hatten wir ebenso schon eingesehen wie, dass $a_n = \sqrt{n}$ eine zu schwache Skalierung ist. Der folgende Satz (der in einer Vorform schon 1924 von Khinchine bewiesen wurde) besagt, dass die Skala $\sqrt{n}$ nur ein wenig zu klein ist und dass $a = -b = \sigma$ die richtigen Konstanten sind.

Satz 2.1 *(Hartmann-Wintner; Satz vom iterierten Logarithmus): Es sei $(X_n)_n$ eine Folge von i.i.d. Zufallsvariablen mit $\mathbb{E}\, X_1 = 0$ und $\mathbb{V}(X_1) =: \sigma^2 < \infty$. Dann gilt für den zugehörigen Partialsummenprozess $S_n := \sum_{i=1}^n X_i$:*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = +\sigma \quad \mathbb{P}\text{-}f.s. \tag{13}$$

und

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = -\sigma \quad \mathbb{P}\text{-}f.s. \tag{14}$$

Wir werden diesen Satz als ein Korollar eines viel allgemeineren Satzes ableiten, des Strassen'schen Satzes vom iterierten Logarithmus. Dieser besagt, dass $[-\sigma, +\sigma]$ sogar die Menge aller Häufungspunkte der Folge $\frac{S_n}{\sqrt{2n \log\log n}}$ ist.

Satz 2.2 *(Strassen; Satz vom iterierten Logarithmus): In der Situation von Satz 1 sei H die Menge aller Häufungspunkte der Folge $\frac{S_n}{\sqrt{2n \log\log n}}$. Dann gilt*

$$H = [-\sigma, +\sigma].$$

Der Beweis von Satz 2.2 besteht aus vielen kleinen Schritten. Er geht auf de Acosta zurück. Zunächst einige einfache analytische Vorbereitungen: Sei

$$L(x) := \begin{cases} 1, & \text{falls } \log x < e \\ \\ \log\log x, & \text{falls } \log x > e \end{cases} \quad (x > 0) \quad .$$

Dann gilt

Lemma 2.3 *Die auf $(0, \infty)$ definierte Funktion*

$$x \mapsto \frac{x}{L(x)}$$

ist strikt wachsend.

Beweis: Übung. □

Aufgrund des Mittelwertsatzes erhalten wir für ein $t \in (0,1)$:

$$\frac{\log(x+y) - \log(x)}{y} = \frac{1}{x+ty} < \frac{1}{x},$$

also

$$\log(x+y) < \log(x) + \frac{y}{x}, \tag{15}$$

also insbesondere für $n \in \mathbb{N}$

$$\log\log(n+1) < \log(n) + \frac{1}{n}.$$

Dies impliziert

$$\log\log(n+1) < \log(\log n + \frac{1}{n}),$$

also

$$\log\log(n+1) < \log\log n + \frac{1}{n\log n}$$

(wobei wir (15) verwendet haben). Somit erhalten wir

$$\frac{L(n+1)}{L(n)} < 1 + \frac{1}{n\,L(n)\log n} < 1 + \frac{1}{2n} \tag{16}$$

für $n \geq e^e$ (denn dann ist $\log n \geq e > 2$, also auch $L(n) \geq 1$). Hieraus leiten wir ab:

Lemma 2.4 *Sei $a_n := \sqrt{2n\,L(n)}$. Dann existiert ein $c > 0$ mit*

$$\sum_{i=1}^{n} \frac{1}{a_i} \leq c\frac{n}{a_n} \quad \textit{für alle} \quad n \in \mathbb{N}. \tag{17}$$

Beweis: Wegen (16) gilt:

$$\frac{a_{n+1}}{a_n} = \sqrt{(1+\frac{1}{n})}\sqrt{\frac{L(n+1)}{L(n)}} < (1+\frac{1}{2n})(1+\frac{1}{2n})^{1/2} = (1+\frac{1}{2n})^{3/2}$$

für alle $n \geq 16 > e^e$. Also

$$\frac{a_{n+1}}{a_n} < (1+\frac{1}{2n})(1+\frac{1}{4n}) < 1 + \frac{1}{n}(\frac{3}{4}+\frac{1}{8n}) < 1 + \frac{7}{8}\frac{1}{n}$$

für $n \geq 16$. Für $n \leq 16$ lässt sich nun ein $c \geq 8$ finden, so dass (17) mit diesem c gilt. Dann folgt aber per Induktion, dass (17) für alle n gilt. In der Tat: Gilt (17) für ein $n \geq 16$, so ergibt sich

$$\begin{aligned}
\sum_{i=1}^{n+1} \frac{1}{a_i} &\leq c\frac{n}{a_n} + \frac{1}{a_{n+1}} < \frac{cn}{a_{n+1}}(1+\frac{7}{8}\frac{1}{n}) + \frac{1}{a_{n+1}} \\
&\leq \frac{c_n}{a_{n+1}}(1+(1-\frac{1}{c})\frac{1}{n}) + \frac{1}{a_{n+1}} = \frac{c(n+1)}{a_{n+1}},
\end{aligned}$$

da $c \geq 8$ war. $\square$

Dies bedeutet somit, dass es eine (berechenbare) Zahl $c > 0$ gibt, so dass

$$\sum_{i=1}^{n} \sqrt{\frac{1}{iL(i)}} \leq c\sqrt{\frac{n}{L(n)}} \tag{18}$$

gilt.

Als weiteres Hilfsmittel benötigen wir

Lemma 2.5 *Für alle $\eta > 1$ gibt es eine isotone Folge $(k_n)_n$ natürlicher Zahlen mit*

(i) $k_n \uparrow \infty$;

(ii) $a_{k_{n+1}} < \eta a_{k_n}$ *für alle bis auf endlich viele n;*

(iii) $\sum_{n=n_0}^{\infty} (\log k_n)^{-\gamma} < \infty$ *für alle $\gamma > 1$ und ein $n_0 \in \mathbb{N}$ mit $k_{n_0} \geq 2$.*

Beweis: Übung $\square$

Als nächstes stellen wir einige probabilistische Hilfsmittel bereit.

Lemma 2.6 *Es sei $(Y_n)_n$ eine Folge unabhängiger Zufallsvariablen mit*

(i) $\mathbb{E}\, Y_n = 0$ *für alle $n \in \mathbb{N}$;*

(ii) $\sigma^2 := \sup \mathbb{E}\, Y_n^2 < \infty$;

(iii) $|Y_n| \leq \tau(\frac{n}{L(n)})^{1/2}$ *$\mathbb{P}$-fast sicher*

für ein $\tau > 0$ und alle $n \in \mathbb{N}$. Dann gilt für

$$T_n := \sum_{i=1}^{n} Y_i$$

die Abschätzung

$$\mathbb{P}(|\frac{T_n}{a_n}| > t) \leq 2\exp(-(\frac{t}{\alpha})^2(2 - e^{\sqrt{2}\alpha^{-2}\tau t})L(n))$$

für alle $\alpha \geq \sigma,\ t > 0$ und $n \in \mathbb{N}$.

Beweis: Diese kompliziert aussehende Abschätzung ist im Prinzip eine Abschätzung der Wahrscheinlichkeit für große (bzw. moderate) Abweichungen.

Aus (iii) folgt für $j = 1, \dots, n$

$$|\frac{Y_j}{a_n}| \leq \tau \left(\frac{j}{L(j)}\right)^{1/2} \frac{1}{a_n} \leq \tau \left(\frac{n}{L(n)}\right)^{1/2} \frac{1}{a_n} = \frac{\tau}{L(n)} \cdot \frac{1}{\sqrt{2}}, \tag{19}$$

wobei die letzte Ungleichung aus Lemma 2.3 folgt. Nun ist wegen der Markov-Ungleichung für alle $t > 0$ und $\lambda > 0$

$$\mathbb{P}[\frac{T_n}{a_n} > t] \leq e^{-\lambda t} \mathbb{E}\, e^{\lambda \frac{T_n}{a_n}}. \tag{20}$$

Den Erwartungswert rechts berechnen wir wie folgt:

$$\mathbb{E}\, e^{\lambda \frac{T_n}{a_n}} = \prod_{j=1}^{n} \mathbb{E}\, e^{\lambda \frac{Y_j}{a_n}} \leq \prod_{j=1}^{n} \mathbb{E}(1 + \frac{\lambda Y_j}{a_n} + \frac{\lambda^2 Y_j^2}{2a_n^2} e^{\frac{\lambda\tau}{\sqrt{2}} \frac{1}{L(n)}}),$$

wobei wir neben dem Multiplikationssatz die Beziehung $e^x \leq 1 + x + \frac{x^2}{2} e^{|x|}$ sowie (19) benutzt haben. Aufgrund von (i) und (ii) ergibt sich für jedes $\alpha \geq \sigma$

$$\begin{aligned}
\mathbb{E}\, e^{\lambda \frac{T_n}{a_n}} &\leq \prod_{j=1}^{n} (1 + \frac{\lambda^2 \alpha^2}{2a_n^2} e^{\frac{\lambda\tau}{\sqrt{2}} \frac{1}{L(n)}}) \\
&\leq \prod_{j=1}^{n} e^{\frac{\lambda^2\alpha^2}{2a_n^2} \exp(\frac{\lambda\tau}{\sqrt{2}} \frac{1}{L(n)})} \\
&= \prod_{j=1}^{n} e^{\frac{\lambda^2\alpha^2}{4n\, L(n)} \exp(\frac{\lambda\tau}{\sqrt{2}} \frac{1}{L(n)})} \\
&= e^{\frac{\lambda^2\alpha^2}{4L(n)} \exp(\frac{\lambda\tau}{\sqrt{2}} \frac{1}{L(n)})}.
\end{aligned}$$

Setzt man dies in (20) ein und wählt $\lambda := \frac{2t}{\alpha^2} L(n)$, ergibt sich

$$\mathbb{P}[\frac{T_n}{a_n} > t] \leq e^{-\lambda t + \frac{\lambda^2\alpha^2}{4L(n)} \exp(\frac{\lambda\tau}{\sqrt{2}} \frac{1}{L(n)})} = \exp[-(\frac{t}{\alpha})^2 (2 - e^{\sqrt{2}\alpha^{-2}\tau t}) L(n)].$$

Da mit Y_n auch $(-Y_n)$ den Voraussetzungen des Lemmas genügt, gilt auch

$$\mathbb{P}[-\frac{T_n}{a_n} > t] \leq \exp[-(\frac{t}{\alpha})^2 (2 - e^{\sqrt{2}\alpha^{-2}\tau t}) L(n)]$$

und somit folgt aus

$$\mathbb{P}[|\frac{T_n}{a_n}| > t] = \mathbb{P}[\frac{T_n}{a_n} > t] + \mathbb{P}[-\frac{T_n}{a_n} > t]$$

die Behauptung. □

Offenbar erfüllen die vorgelegten (X_n) nicht die Anforderungen, die wir an die (Y_n) im vorangegangenen Lemma gestellt haben. Vielmehr müssen wir die X_n stutzen, um zu solchen Y_n zu gelangen. Das folgende Lemma untersucht die Konsequenzen einer solchen Stutzung:

Lemma 2.7 *Für* (X_n) *wie in Satz 2.1 und* $\tau > 0$ *sei*

$$Z_j := X_j \; \mathbb{1}_{\{|X_j| > \tau \sqrt{\frac{j}{L(j)}}\}}$$

und

$$U_n := \sum_{j=1}^{n} Z_j.$$

Dann konvergiert $(\frac{1}{a_n} U_n)$ $\mathbb{P}$*-fast sicher gegen 0 und es gilt*

$$\sum_{j=1}^{\infty} \mathbb{E}(\frac{1}{a_n} |Z_n|) < \infty. \tag{21}$$

Beweis: Die Konvergenz von $(\frac{1}{a_n} U_n)$ gegen 0 erhalten wir, falls wir zeigen können, dass $\sum_{n=1}^{\infty} \frac{1}{a_n} Z_n$ fast sicher konvergiert. Dies folgt sofort aus einem Lemma, das auf Kronecker zurückgeht und welches wir im Anschluss zeigen werden. Die fast sichere Konvergenz von $\sum_{n=1}^{\infty} \frac{1}{a_n} Z_n$ folgt aber aus (21), da

$$\mathbb{E}(\sum_{n=1}^{\infty} \frac{1}{a_n} |Z_n|) = \sum_{n=1}^{\infty} \mathbb{E}(\frac{1}{a_n} |Z_n|)$$

gilt und wir mit (21) also sogar die fast sichere absolute Konvergenz der zu untersuchenden Reihe erhalten. Wir zeigen also nun (21). Wir setzen $b_n := \sqrt{\frac{n}{L(n)}}$ (und erinnern uns, dass $(b_n)_n$ strikt wachsend ist). Dann gilt:

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{E}(|\frac{Z_n}{a_n}|) &= \sum_{n=1}^{\infty} \frac{1}{a_n} \mathbb{E}(|X_n|;\ |X_n| > \tau b_n) \\ &= \sum_{n=1}^{\infty} \frac{1}{a_n} \sum_{j=n}^{\infty} \mathbb{E}(|X_n|;\ \tau b_j < |X_n| \leq \tau b_{j+1}) \\ &\leq \tau \sum_{n=1}^{\infty} \frac{1}{a_n} \sum_{j=n}^{\infty} b_{j+1} \mu_j, \end{aligned}$$

wobei

$$\mu_j := \mathbb{P}(\tau b_j < |X_1| \leq \tau b_{j+1})$$

(man erinnere sich, dass die (X_n) i.i.d. waren). Vertauscht man die Summationsreihenfolge, ergibt sich

$$\sum_{n=1}^{\infty} \mathbb{E}(|\frac{Z_n}{a_n}|) \leq \tau \sum_{j=1}^{\infty} b_{j+1} \mu_j \sum_{n=1}^{j} \frac{1}{a_n}.$$

Nun gibt es nach Lemma 2.4 ein $c > 0$ mit

$$\sum_{n=1}^{j} a_n \leq c \frac{j}{a_j} = \frac{c}{\sqrt{2}} b_j.$$

Also
$$\sum_{n=1}^{\infty} \mathbb{E}(|\frac{Z_n}{a_n}|) \leq \frac{1}{\sqrt{2}} c\tau \sum_{j=1}^{\infty} b_j b_{j+1} \mu_j.$$
Nun ist
$$\frac{b_{j+1}}{b_j} = \left(\frac{j+1}{j} \frac{L(j)}{L(j+1)}\right)^{1/2} \leq \sqrt{\frac{j+1}{j}} \leq \sqrt{2}.$$
Ergo
$$\sum_{n=1}^{\infty} \mathbb{E}(|\frac{Z_n}{a_n}|) \leq c\tau \sum_{j=1}^{\infty} b_j^2 \mu_j = \frac{c}{\tau} \sum_{j=1}^{\infty} (b_j \tau)^2 \mu_j.$$
Nun ist
$$(\tau b_j)^2 \mu_j \leq \mathbb{E}(X_1^2;\ \tau b_j < |X_1| \leq \tau b_{j+1}).$$
Dadurch erhalten wir
$$\sum_{n=1}^{\infty} \mathbb{E}(|\frac{Z_n}{a_n}|) \leq \frac{c}{\tau} \sum_{j=1}^{\infty} E(X_1^2; \tau b_j < |X_1| \leq \tau b_{j+1}) \leq \frac{c}{\tau} \mathbb{E}(X_1^2) < \infty.$$
Dies zeigt (21). □

Im Beweis haben wir das folgende Lemma verwandt:

Lemma 2.8 *(Kronecker-Lemma): Seien $(x_n)_n$ und (τ_n) zwei Folgen reeller Zahlen und sei (τ_n) isoton, $\tau_n > 0$ und $\tau_n \uparrow \infty$. Dann gilt*
$$\sum_{i=1}^{\infty} \frac{x_i}{\tau_i} \quad \textit{konvergent} \quad \Rightarrow \lim_{n \to \infty} \frac{1}{\tau_n} \sum_{i=1}^{n} x_i = 0.$$

Beweis: Dies ist eine rein analytische Aussage. Wir überlassen den Beweis dem Leser zur Übung. Man findet das Lemma aber z.B. auch im Buch von Bauer (Lemma 14.4). □

Wir wollen die Abschätzung aus Lemma 2.7 benutzen, um die Wahrscheinlichkeit, dass $\frac{S_n}{\sqrt{2nL(n)}} \geq \sigma$ bzw. $\leq -\sigma$ ist, für alle n zu kontrollieren. Dazu ist es handlich, diese Wahrscheinlichkeiten durch die Wahrscheinlichkeit abzuschätzen, dass das entsprechende Ereignis für ein festes n gilt. Das folgende Lemma liefert die benötigte Ungleichung:

Lemma 2.9 *(Ottaviani-Skohorod-Ungleichung): Es seien $X_1, \ldots, X_n$ i.i.d. Zufallsvariablen und $S_k = \sum_{j=1}^{k} X_j$ $(k = 1, \ldots, n)$. Dann gilt für alle $\varepsilon > 0$ und $\eta > 0$ und alle $m = 1, \ldots, n$:*
$$\mathbb{P}(\max_{m \leq k \leq n} |S_k| > \eta + \varepsilon)(\min_{m \leq k \leq n} \mathbb{P}(|S_n - S_k| \leq \varepsilon)) \leq \mathbb{P}(|S_n| > \eta). \tag{22}$$

Beweis: Sei $A := \{\max_{m \le k \le n} |S_k| > \eta + \varepsilon\}$. Ferner sei für $\omega \in A$

$$T(\omega) := \min\{k : m \le k \le n : |S_k(\omega)| > \eta + \varepsilon\}.$$

Für alle $k = m, \ldots, n$ gilt dann

$$A_k := \{\omega \in A : T(\omega) = k\} = \{\max_{m \le j \le k-1} |S_j| \le \eta + \varepsilon\} \cap \{|S_k| > \eta + \varepsilon\}.$$

Da die A_k disjunkt sind und sich zu A ergänzen, erhält man

$$\mathbb{P}(A) = \sum_{k=m}^{n} \mathbb{P}(A_k).$$

Führen wir zudem

$$A'_k := A_k \cap \{|S_n - S_k| \le \varepsilon\} \quad k = m, \ldots, n$$

ein, so gilt wegen der Unabhängigkeit der Ereignisse A_k und $\{|S_n - S_k| \ge \varepsilon\}$

$$\mathbb{P}(A'_k) = \mathbb{P}(A_k)\, \mathbb{P}(|S_n - S_k| \le \varepsilon).$$

Hieraus ergibt sich

$$\sum_{k=m}^{n} \mathbb{P}(A'_k) \ge \gamma \mathbb{P}(A)$$

mit

$$\gamma := \min_{m \le k \le n} \mathbb{P}(|S_n - S_k| \le \varepsilon).$$

Wegen $|S_n - S_k| \ge |S_k| - |S_n|$ folgt $|S_n(\omega)| > \eta$ für $\omega \in A'_k$. Also

$$\bigcup_{k=m}^{m} A'_k \subseteq \{|S_n| > \eta\}.$$

Dies ergibt insgesamt wegen der Disjunktheit der (A'_k)

$$\gamma \mathbb{P}(A) \le \sum_{k=m} \mathbb{P}(A'_k) \le \mathbb{P}(|S_n| > \eta).$$

□

Bemerkung 2.10 *Setzt man*

$$\lambda := \max_{k \le m \le n} \mathbb{P}(|S_n - S_k| > \varepsilon),$$

so schreibt sich (22) in der Form

$$(1 - \lambda)\mathbb{P}(\max_{m \le k \le n} |S_k| > \eta + \varepsilon) \le \mathbb{P}(|S_n| > \eta)$$

und für $\lambda < 1$:

$$\mathbb{P}(\max_{m \le k \le n} |S_k| > \eta + \varepsilon) \le \frac{1}{1 - \lambda} \mathbb{P}(|S_n| > \eta). \tag{23}$$

Wir wollen die soeben bewiesene Ungleichung nun in der Form (23) anwenden:

Lemma 2.11 *Sei $(Y_n)_n$ eine Folge unabhängiger Zufallsvariablen und $T_n := \sum_{i=1}^n Y_i$. Es gelte:*

a) $\frac{1}{a_n}T_n \to 0 \quad \mathbb{P}$*-stochastisch.*

b) Es gibt $\gamma > 1,\ \beta > 0,\ c > 0$ *mit*

$$\mathbb{P}(|\frac{T_n}{a_n}| > \beta) \leq c \cdot e^{-\gamma L(n)}$$

für alle genügend große n.

Dann gilt

$$\limsup_{n\to\infty} |\frac{T_n}{a_n}| \leq \beta \quad \mathbb{P}\text{-}f.s. \tag{24}$$

Beweis: Da sich jedes $x > \beta$ als $x = \beta\eta + \varepsilon,\ \eta > 1,\ \varepsilon > 0$ schreiben lässt, besagt (24) mit anderen Worten

$$\mathbb{P}(|\frac{T_n}{a_n}| > \eta\beta + \varepsilon \text{ unendlich oft}) \ = 0 \tag{25}$$

für alle $\eta > 1,\ \varepsilon > 0$. Seien nun $\eta > 1,\ \varepsilon > 0$ fest gewählt. Zu η wählen wir eine Folge (k_n) wie in Lemma 5. Dann ist (25) bewiesen, falls wir

$$\mathbb{P}(\max_{k_n \leq m \leq k_{n+1}} |\frac{T_m}{a_m}| > \eta\beta + \varepsilon \text{ u.o.}) = 0$$

zeigen können. Dies ist gezeigt (nach dem Borel-Cantelli-Lemma), falls wir

$$\sum_{n=1}^{\infty} \mathbb{P}(\max_{k_n \leq m \leq k_{n+1}} |\frac{T_m}{a_m}| > \eta\beta + \varepsilon) < +\infty$$

zeigen können. Dies wollen wir mit Hilfe der Ottoviani-Skohorod-Ungleichung herleiten. Nach a) ist für genügend große m

$$\mathbb{P}(|\frac{T_m}{a_m}| \leq \frac{\varepsilon}{2\eta}) \geq \frac{3}{4}.$$

Nun gilt wegen der Eigenschaft (ii) der Folge (a_{k_n}) und ihrer Isotonie

$$|\frac{T_m}{a_{k_n}}| \leq \eta|\frac{T_m}{a_{k_{n+1}}}| \leq \eta|\frac{T_m}{a_m}|,$$

also

$$\mathbb{P}(|\frac{T_m}{a_{k_n}}| \leq \frac{\varepsilon}{2}) \geq \mathbb{P}(|\frac{T_m}{a_m}| \leq \frac{\varepsilon}{2\eta}) \geq \frac{3}{4},$$

insbesondere auch

$$\mathbb{P}(|\frac{T_{k_{n+1}}}{a_{k_n}}| \leq \frac{\varepsilon}{2}) \geq \frac{3}{4}$$

für alle n genügend groß und alle $k_n \leq m \leq k_{n+1}$. Nach der Dreiecksungleichung ist

$$\{|\frac{T_m}{a_{k_n}}| \leq \frac{\varepsilon}{2}\} \cap \{|\frac{T_{k_{n+1}}}{a_{k_n}}| \leq \frac{\varepsilon}{2}\} \subseteq \{|\frac{T_{k_{n+1}}}{a_{k_n}} - \frac{T_m}{a_{k_n}}| \leq \varepsilon\}$$

und wegen

$$\mathbb{P}(A \cap B) \geq P(A) + P(B) - 1$$

folgt für beliebige Ereignisse A und B

$$\mathbb{P}(|\frac{T_{k_{n+1}}}{a_{k_n}} - \frac{T_m}{a_{k_n}}| \leq \varepsilon) \geq \frac{3}{4} + \frac{3}{4} - 1 = \frac{1}{2}.$$

Lemma 2.9 in der Form (23) liefert mit $X_i := \frac{Y_i}{a_{k_n}}$

$$\mathbb{P}(\max_{k_n \leq m \leq k_{n+1}} |\frac{T_m}{a_{k_n}}| > \eta\beta + \varepsilon) \leq 2\mathbb{P}(|\frac{T_{k_{n+1}}}{a_{k_n}}| > \eta\beta).$$

Nutzt man noch einmal Eigenschaft (ii) aus, so erhält man hieraus auch

$$\mathbb{P}(\max_{k_n \leq m \leq k_{n+1}} |\frac{T_m}{a_{k_n}}| > \eta\beta + \varepsilon) \leq 2\mathbb{P}(|\frac{T_{k_{n+1}}}{a_{k_{n+1}}}| > \beta)$$

für alle n genügend groß (wegen $|\frac{T_{k_{n+1}}}{a_{k_n}}| \leq \eta|\frac{T_{k_{n+1}}}{a_{k_{n+1}}}|$). Wenden wir nun b) an (und wieder die Isotonie der Folge (a_n)), so ergibt sich

$$\mathbb{P}(\max_{k_n \leq m \leq k_{n+1}} |\frac{T_m}{a_m}| > \eta\beta + \varepsilon) \leq 2ce^{-\gamma L(k_{n+1})}$$

für alle n genügend groß. Nun ist $\gamma > 1$ und

$$e^{-\gamma L(k_{n+1})} = (\log k_n)^{-\gamma}$$

für alle n mit $\log k_n \geq e$. Gemäß (iii) aus Lemma 2.5 ist $\sum_n (\log k_n)^{-\gamma} < \infty$, dies beweist die Behauptung. □

Hiermit können wir nun zeigen, dass die Häufungspunkte von $\frac{S_n}{a_n}$ zwischen $-\sigma$ und σ liegen:

Satz 2.12 *Unter den Voraussetzungen und mit den Bezeichnungen von Satz 2.1 gilt:*

(i) $\limsup_{n\to\infty} \frac{S_n}{a_n} \leq \sigma$ $\mathbb{P}$-*f.s.* *und*

(ii) $\liminf_{n\to\infty} \frac{S_n}{a_n} \geq -\sigma$ $\mathbb{P}$-*f.s.*

Beweis: Um die vorherigen Ergebnisse anwenden zu können, müssen die (X_n) gestutzt werden. Hierzu seien zu $0 < \delta < 1$ und $\tau > 0$ so gewählt, dass

$$\frac{1}{(1+\delta)^2} < 2 - e^{2\sqrt{2}\tau/\sigma}$$

gilt. Wir setzen

$$X_n' := X_n \cdot \mathbb{1}_{\{|X_n| \leq \frac{\tau}{2} b_n\}}$$

und

$$Y_n := X_n' - \mathbb{E}\, X_n'.$$

Hierbei sei wieder $b_n = \sqrt{\frac{n}{L(n)}}$. Die Y_n sind unabhängig mit $\mathbb{E} Y_n = 0$ und $\mathbb{E}(Y_n^2) = \mathbb{V}(X_n') \leq \mathbb{E}(X_n')^2 \leq \mathbb{E} X_n^2 = \sigma^2$. Da $|X_n'| \leq \frac{\tau}{2} b_n$ gilt, folgt $|\mathbb{E} X_n'| \leq \frac{\tau}{2} b_n$, also

$$|Y_n| \leq \tau\, b_n.$$

Wir können also Lemma 2.6 anwenden und für $T_n := \sum_{k=1} Y_k$ und $\alpha = \sigma$ ableiten:

$$\mathbb{P}(|\frac{T_n}{a_n}| > t) \leq 2e^{-(\frac{t}{\sigma})^2 (2 - e^{\sqrt{2}\sigma^{-2}\tau t}) L(n)}$$

für jedes $t > 0$. Wählt man speziell $t := (1+\delta)\sigma$, gilt wegen $0 < \delta < 1$

$$\mathbb{P}(|\frac{T_n}{a_n}| > (1+\delta)\sigma) \leq 2e^{-\gamma L(n)}$$

mit

$$\gamma := (1+\delta)^2 (2 - e^{2\sqrt{2}\tau/\sigma}),$$

wobei wir hier im Exponenten δ durch 1 abgeschï¿½tzt haben. Nach Wahl von τ ist $\gamma > 1$. Außerdem erhalten wir mit der Tschebyscheff-Ungleichung

$$\mathbb{P}(|\frac{T_n}{a_n}| > \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{V}(\frac{T_n}{a_n}) \leq n(\frac{\sigma}{\varepsilon a_n})^2 \to 0$$

(da $\frac{n}{a_n^2} = \frac{1}{2L(n)} \to 0$), also konvergiert $\frac{T_n}{a_n}$ stochastisch gegen 0.

Lemma 2.11 liefert somit

$$\limsup_{n\to\infty} \frac{T_n}{a_n} \leq (1+\delta)\sigma \quad \mathbb{P}\text{-fast sicher.}$$

Nun heben wir die Stutzung der X_n wieder auf; sei

$$Z_j := X_j - X_j'.$$

Nach Lemma 2.7 gilt

$$\sum_{n=1}^{\infty} \mathbb{E}(|\frac{Z_n}{a_n}|) < \infty,$$

also wegen $\mathbb{E}\, Z_j = -\mathbb{E}\, X_j'$

$$\sum_{n=1}^{\infty} \left| \mathbb{E}\left(\frac{X_n'}{a_n}\right) \right| = \sum_{n=1}^{\infty} \left| \mathbb{E}\left(\frac{Z_n}{a_n}\right) \right| \leq \sum_{n=1}^{\infty} \mathbb{E}\left(\left| \frac{Z_n}{a_n} \right| \right) < \infty.$$

Für $S_n' = \sum_{i=1}^n X_i'$ gilt also nach dem Kronecker-Lemma

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{S_n'}{a_n}\right) = 0.$$

Setzen wir schließlich $U_n = \sum_{i=1}^n Z_i$, so konvergiert die Folge $(\frac{U_n}{a_n})$ nach Lemma 2.7 fast sicher gegen 0. Nun ist

$$S_n = S_n' + U_n = T_n + \mathbb{E}(S_n') + U_n.$$

Somit folgt

$$\limsup_{n\to\infty} \frac{S_n}{a_n} \le \limsup_{n\to\infty} \frac{T_n}{a_n} + \limsup_{n\to\infty} \frac{\mathbb{E}(S_n')}{a_n} + \limsup_{n\to\infty} \frac{U_n}{a_n} = \limsup_{n\to\infty} \frac{T_n}{a_n} \quad \mathbb{P}\text{-f.s.}$$

Also

$$\limsup_{n\to\infty} \frac{S_n}{a_n} \le (1+\delta)\sigma$$

für alle $\delta > 0$. Hieraus folgt (i).

(ii) erhält man, wenn man die Folge X_n durch die Folge $(-X_n)$ ersetzt. □

Bevor wir nun zeigen, dass $+\sigma$ und $-\sigma$ auch in der Tat die größten bzw. kleinsten Häufungspunkte der Folge $\frac{S_n}{a_n}$ sind, müssen wir noch ein technisches Hilfsmittel bereitstellen. Der Beweis ist leider aufwändiger als die Aussage vermuten lässt. Jedoch benötigen wir dieses Lemma, um von der Folge (a_n) zu einer asymptotisch äquivalenten Folge wechseln zu können.

Lemma 2.13 *Es sei $(T_n)_n$ eine Folge von Zufallsvariablen und $(t_n)_n$ eine Folge reeller Zahlen mit $t_n \underset{n\to\infty}{\longrightarrow} \infty$. Konvergiert dann $(\frac{T_n}{t_n})_n$ in Verteilung gegen ein Wahrscheinlichkeitsmaß ν auf $\mathbb{R}$, so konvergiert auch $(\frac{T_n}{s_n})_n$ in Verteilung gegen ν für jede Folge $(s_n)_n$, die zu $(t_n)_n$ asymptotisch äquivalent ist.*

Beweis: Wir wissen, dass

$$\lim_{n\to\infty} \int f\left(\frac{T_n}{t_n}\right) d\mathbb{P} = \int f\, d\nu \tag{26}$$

für alle stetigen, beschränkten Funktionen $f : \mathbb{R} \to \mathbb{R}$ gilt. Eine Übung zeigt, dass die Aussage (26) für alle $f \in C^b(\mathbb{R})$ mit kompaktem Träger (wir schreiben fortan $C_c(\mathbb{R})$ für diese Funktionenklasse) schon die Verteilungskonvergenz von $(\frac{T_n}{t_n})_n$ gegen ν impliziert. Wir erinnern daran, dass jedes $f \in C_c(\mathbb{R})$ auch gleichmäßig stetig ist, zu vorgelegtem $\varepsilon > 0$ existiert also ein $\delta > 0$, so dass

$$|x-y| \le \delta \;\Rightarrow\; |f(x) - f(y)| \le \varepsilon$$

für alle $x, y \in \mathbb{R}$. Wir wählen o. B. d. A. $\delta < 1$.

Ferner gibt es definitionsgemäß ein $K_0 > 0$, so dass für den Träger $\mathrm{supp}(f) := \{x : f(x) > 0\}$ gilt

$$\mathrm{supp}(f) \subseteq [-K_0, K_0].$$

Wir schreiben

$$K := 1 + K_0 \quad \text{und} \quad \eta := \frac{\delta}{K}.$$

Da (s_n) und (t_n) asymptotisch äquivalent sind, gibt es ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$ gilt

$$|\frac{t_n}{s_n} - 1| < \eta.$$

Wir werden nun zeigen, dass dann für alle $n \geq n_0$ auch

$$\left| f\left(\frac{T_n(\omega)}{t_n}\right) - f\left(\frac{T_n(\omega)}{s_n}\right) \right| \leq \varepsilon$$

für alle $\omega \in \Omega$ gilt. In der Tat:

- Ist $|\frac{T_n(\omega)}{t_n}| \leq K$, so gilt

$$\left| \frac{T_n(\omega)}{t_n} - \frac{T_n(\omega)}{s_n} \right| = \left| \frac{T_n(\omega)}{t_n} \right| \left| 1 - \frac{t_n}{s_n} \right| \leq K\eta = \delta.$$

- Ist $\frac{T_n(\omega)}{t_n} > K$, dann gilt

$$\frac{T_n(\omega)}{s_n} = \frac{T_n(\omega)}{t_n} \cdot \frac{t_n}{s_n} > K(1 - \eta) > (1 - \frac{1}{K}) \cdot K = K_0.$$

  Somit folgt

$$f\left(\frac{T_n}{s_n}\right) = f\left(\frac{T_n}{s_n}\right) = 0.$$

- Ist $\frac{T_n(\omega)}{t_n} < -K$, folgt die Behauptung analog.

Somit gilt

$$f\left(\frac{T_n}{t_n}\right) - f\left(\frac{T_n}{s_n}\right) \to 0.$$

Da zudem

$$\left| f\left(\frac{T_n}{t_n}\right) - f\left(\frac{T_n}{s_n}\right) \right| \leq 2\|f\|$$

gilt, folgt die Behauptung des Lemmas aus dem Satz über majorisierte Konvergenz, also

$$\lim_{n\to\infty} \int \left[f\left(\frac{T_n}{t_n}\right) - f\left(\frac{T_n}{s_n}\right) \right] d\mathbb{P} = 0,$$

was

$$\lim_{n\to\infty} \int f\left(\frac{T_n}{s_n}\right) d\mathbb{P} = \int f \, d\nu$$

impliziert. □

Übung 2.14 *Man zeige, dass eine Folge von Zufallsvariablen X_n auf einem Wahrscheinlichkeitszeitraum schon dann in Verteilung gegen ein Wahrscheinlichkeitsmaß ν konvergiert, wenn*

$$\int f(X_n) d\mathbb{P} \to \int f \, d\nu$$

für alle $f \in C_c(\mathbb{R})$ gilt.

Wir sind nun in der Lage, den Strassenschen Satz vom iterierten Logarithmus zu beweisen.

Beweis von Satz 2.2: Zu zeigen ist also mit den oben eingeführten Bezeichnungen, dass

$$H = [-\sigma, +\sigma]$$

gilt. Nun ist stets

$$H \subseteq [\liminf \frac{S_n}{a_n}, \limsup \frac{S_n}{a_n}],$$

also nach Satz 2.12

$$H \subseteq [-\sigma, +\sigma] \quad \mathbb{P}\text{-f.s.}$$

Zu zeigen ist also nur, dass tatsächlich jeder Punkt in $[-\sigma, +\sigma]$ Häufungspunkt von $\frac{S_n}{a_n}$ $\mathbb{P}$-f.s. ist. Es genügt, dies für das Innere, also für $(-\sigma, +\sigma)$ zu zeigen. Zu zeigen ist somit: Für jedes $x \in (-\sigma, +\sigma)$ gibt es eine Nullmenge $N_x \in \mathcal{F}$, $\mathbb{P}(\mathbb{N}_x) = 0$, so dass für alle $\omega \in N_x^c$ gilt, dass x ein Häufungspunkt von $\frac{S_n}{a_n}$ ist. Dies funktioniert, d. h. wir bekommen kein Problem mit den überabzählbar vielen Nullmengen $(N_x)_{x\in\mathbb{R}}$, denn mit der Behauptung haben wir insbesondere sichergestellt, dass jedes $q \in (-\sigma, +\sigma) \cap \mathbb{Q}$ Häufungspunkt der Folge $\frac{S_n(\omega)}{a_n}$ für alle $\omega \in N_q^c$ ist. Die abzählbar vielen Nullmengen N_q vereinigen sich aber wieder zu einer Nullmenge

$$N := \bigcup_{q \in \mathbb{Q} \cap (-\sigma, +\sigma)} N_q.$$

Für $\omega \in N^c$ ist somit jedes $q \in \mathbb{Q} \cap (-\sigma, +\sigma)$ Häufungspunkt von $\frac{S_n}{a_n}$. Da aber $\mathbb{Q}$ dicht liegt in $\mathbb{R}$, ist dann jedes $x \in [-\sigma, +\sigma]$ Häufungspunkt der Folge $\frac{s_n}{a_n}$ und dies wollen wir zeigen.

Es sei also $x \in (-\sigma, +\sigma)$. Wir definieren für $k \in \mathbb{N}$

$$n_k := k^k.$$

Aufgrund der Dreiecksungleichung gilt dann

$$\left| \frac{S_{n_{k+1}}}{a_{n_{k+1}}} - x \right| \leq \left| \frac{S_{n_k}}{a_{n_{k+1}}} \right| - \left| \frac{1}{a_{n_{k+1}}} (S_{n_{k+1}} - S_{n_k}) - x \right|. \tag{27}$$

Weiter ist

$$\frac{a_{n_k}}{a_{n_{k+1}}} \underset{k\to\infty}{\longrightarrow} 0,$$

denn

$$\left(\frac{a_{n_k}}{a_{n_{k+1}}}\right)^2 = \left(\frac{1}{1+\frac{1}{k}}\right)^k \frac{1}{k+1} \frac{\log(k \, \log \, k)}{\log[(k+1)\log(k+1)]},$$

was offensichtlich gegen 0 konvergiert. Somit folgt

$$\lim_{k\to\infty} \left|\frac{S_{n_k}}{a_{n_{k+1}}}\right| = \lim_{k\to\infty} \left|\frac{S_{n_k}}{a_{n_k}}\right| \left|\frac{a_{n_k}}{a_{n_{k+1}}}\right| = 0 \quad \mathbb{P}\text{-f.s.},$$

denn nach Satz 2.12 ist die Folge $\left|\frac{S_{n_k}}{a_{n_k}}\right|$ $\mathbb{P}$-f.s. beschränkt. Der relevante Summand auf der rechten Seite von (27) ist also der zweite. Wir wollen zeigen, dass auch dieser beliebig klein wird. Dazu definieren wir zu $\varepsilon > 0$

$$A_k := \left\{ \left| \frac{1}{a_{n_{k+1}}}(S_{n_{k+1}} - S_{n_k}) - x \right| < \varepsilon \right\}.$$

Zur Abkürzung sei noch

$$m_k := n_{k+1} - n_k, \quad k \in \mathbb{N},$$

eingeführt. Nun sind die $(X_i)_{i\in\mathbb{N}}$ als eine i.i.d. Folge gewählt. Dies impliziert einerseits, dass die Ereignisse A_k unabhängig sind, denn sie hängen von unabhängigen Zufallsvariablen ab. Andererseits ist die Verteilung der $(S_{n_{k+1}} - S_{n_k})_k$ offenbar dieselbe wie die Verteilung der $(S_{m_k})_k$. Also ist

$$\mathbb{P}(A_k) = \mathbb{P}\left(\left| \frac{S_{m_k}}{a_{n_{k+1}}} - x \right| < \varepsilon \right).$$

Sei $\alpha := |\frac{x}{\sigma}$. Man beachte, dass $\alpha < 1$ ist. Zu α wählen wir $0 < \delta < \frac{1}{4}\,(1-\alpha)$. Dann gilt

$$\mathbb{P}(A_k) \geq e^{-(\alpha+4\delta)L(m_{k+1})}. \tag{28}$$

Die Herleitung von (28) ist noch ein wenig aufwändig. Wenn wir aber (28) für den Moment voraussetzen, so können wir zeigen, dass Satz 2.2 in der Tat wahr ist: (28) impliziert nämlich

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty,$$

da $\alpha + 4\delta < 1$ ist und daher

$$e^{-(\alpha+4\delta)L(n_{k+1})} > e^{-L(n_{k+1})} = \frac{1}{(k+1)\log(k+1)}$$

gilt. Die Reihe $\sum_{k\geq 2} \frac{1}{k \, \log k}$ ist aber bekanntlich divergent.

Wenden wir also den zweiten Teil des Borel-Cantelli-Lemmas auf die (unabhängigen!) Ereignisse (A_k) an, so folgt

$$\mathbb{P}(\limsup_{k\to\infty} A_k) = 1,$$

d. h.

$$\mathbb{P}\left(\left|\frac{1}{a_{n_{k+1}}}(S_{n_{k+1}} - S_{n_k}) - x\right| < \varepsilon \text{ u.o.}\right) = 1.$$

Aus (27) ergibt sich damit

$$\mathbb{P}\left(\left|\frac{S_{n_{k+1}}}{a_{n_{k+1}}} - x\right| < 2\varepsilon \text{ u.o.}\right) = 1$$

und damit

$$\mathbb{P}\left(\liminf_{n\to\infty}\left|\frac{S_{n_{k+1}}}{a_{n_{k+1}}} - x\right| < 2\varepsilon\right) = 1.$$

Da $\varepsilon > 0$ beliebig war, können wir für $\varepsilon = \varepsilon_n = \frac{1}{n}$ wählen (um wieder abzählbar viele Nullmengen zu erhalten); somit ergibt sich

$$\mathbb{P}\left(\liminf_{k\to\infty}\left|\frac{S_{n_{k+1}}}{a_{n_{k+1}}} - x\right| = 0\right) = 1.$$

Also ist x $\mathbb{P}$-f.s. Häufungspunkt der Folge $(\frac{S_n}{a_n})$. Der Beweis ist somit modulo der Abschätzung (28) vollständig. □

Abschließend leiten wir (28) her.

Lemma 2.15 *(28) gilt, d. h. mit den Bezeichnungen des Beweises von Satz 2.2 gilt*

$$\mathbb{P}(A_k) \geq \exp(-(\alpha + 4\delta)L(n_k)).$$

Beweis: Sei wieder $m_k = n_{k+1} - n_k$ und

$$\alpha_k := a_{n_{k+1}}.$$

Dann gilt

$$\lim_{k\to\infty}\frac{\alpha_k}{m_k} = 0 \quad \text{und} \quad \lim_{k\to\infty}\frac{\alpha_k^2}{m_k} = +\infty.$$

Dies folgt, weil – wie man leicht nachrechnet – die Folgen (m_k) und (n_{k+1}) asymptotisch äquivalent sind und

$$\frac{\alpha_k}{m_k} = \frac{a_{n_{k+1}}}{n_{k+1}} \cdot \frac{n_{k+1}}{m_k} \quad \text{und} \quad \frac{\alpha_k^2}{m_k} = \frac{a_{n_{k+1}}^2}{n_{k+1}} \cdot \frac{n_{k+1}}{m_k}$$

sowie

$$\frac{a_n}{n} = \sqrt{(\frac{2}{n}L(n))} \quad \text{und} \quad \frac{a_n^2}{n} = \alpha L(n)$$

gilt.

Die Abschätzung (28) folgt somit aus der typischen Abschätzung über moderate Abweichungen für Folgen von i.i.d. Zufallsvariablen. Genauer werden wir zeigen, dass

$$\frac{m_k}{\alpha_k^2}\log\mathbb{P}(A_k) > -\frac{1}{2}(\frac{x}{\sigma})^2 - \delta \tag{29}$$

für $\delta > 0$ und alle genügend großen k gilt. Beachtet man, dass $\alpha := |\frac{x}{\sigma}| < 1$ und $\delta > 0$ war und dass dies

$$\alpha^2 < \alpha < \alpha + \delta \quad \text{und somit} \quad \left(\frac{x}{\sigma}\right)^2 < \alpha + \delta \tag{30}$$

impliziert, so liefert (29) das Verlangte. In der Tat: Wegen $m_k \sim n_{k+1}$ bekommen wir aus (30)

$$\frac{n_{k+1}}{m_k}\left(\frac{x}{\sigma}\right)^2 < \alpha + \delta$$

für k genügend groß. Somit folgt wegen $\alpha_k^2 = 2n_{k+1}L(n_{k+1})$

$$\frac{\alpha_k^2}{2m_k}\left(\frac{x}{\sigma}\right)^2 < (\alpha + \delta)L(n_{k+1}) \tag{31}$$

wiederum für k genügend groß. Aus (29) und (31) folgt somit

$$\log \mathbb{P}(A_k) > -\frac{1}{2}\left(\frac{x}{\sigma}\right)^2 \frac{\alpha_k^2}{m_k} - \delta\frac{\alpha_k^2}{m_k} > -(\alpha+\delta)L(n_{k+1}) - \delta\frac{\alpha_k^2}{m_k}$$

für hinreichend große k. Da (m_k) und (n_{k+1}) asymptotisch äquivalent sind, ist aber

$$\frac{\alpha_k^2}{m_k} = 2L(n_{k+1})\frac{n_{k+1}}{m_k} < 3L(n_{k+1})$$

für k genügend groß. Für solche k ergibt sich also

$$\log \mathbb{P}(A_k) \geq -(\alpha+\delta)L(n_{k+1}) - 3\delta L(n_{k+1}) = -(\alpha+4\delta)L(n_{k+1}),$$

also (28).

Es bleibt (29) zu beweisen (was noch ein wenig Arbeit ist). Es sei hierzu ν_{0,σ^2} die $\mathcal{N}(0,\sigma^2)$-Verteilung. Für ein offenes Intervall

$$J = (c, d)$$

und $t > 0$ wollen wir zunächst die Ungleichung

$$\liminf \frac{m_k}{\alpha_k^2} \log \mathbb{P}\left(\frac{S_{m_k}}{\alpha_k} \in J\right) \geq \frac{1}{t^2} \log \nu_{0,\sigma^2}(tJ) \tag{32}$$

herleiten. Hierzu sei für $\delta > 0$ genügend klein

$$J_\delta := (c+\delta, d-\delta) \quad \text{und} \quad U_\delta = (-\delta, \delta).$$

Ferner definieren wir die Zahlenfolgen

$$p_k := \left[\frac{t^2 m_k^2}{\alpha_k^2}\right], \; q_k := \left[\frac{\alpha_k^2}{t^2 m_k}\right] \quad \text{und} \quad r_k := \frac{\alpha_k}{t q_k},$$

wobei $[\cdot]$ die Gauß-Klammer bezeichne. Beachte, dass $p_k \geq 1$ und $q_k \geq 1$ für hinreichend große k gilt (nur solche seien in der Folge betrachtet). Wir werden (32) aus drei Hilfsbehauptungen ableiten:

$$[\mathbb{P}(S_{p_k} \in r_k \cdot t\, J_\delta)]^{q_k} \leq \mathbb{P}\left(\frac{S_{p_k q_k}}{\alpha_k} \in J_\delta\right), \tag{33}$$

$$\lim_{k\to\infty} \lambda_k = 0, \tag{34}$$

wobei

$$\lambda_k := \mathbb{P}(|S_{m_k} - S_{p_k q_k}| \geq \delta_{\alpha_k})$$

bezeichnet, und

$$(1-\lambda_k)\mathbb{P}(\frac{S_{p_k q_k}}{\alpha_k} \in J_\delta) \geq \mathbb{P}(\frac{S_{m_k}}{\alpha_k} \in J). \tag{35}$$

Um (33) einzusehen, erinnern wir uns, dass

$$S_{p_k q_k} = \sum_{i=1}^{q_k} Y_i$$

ist, wenn wir unter Y_i wieder Summen von p_k aufeinander folgenden Gliedern der Folge (X_n) verstehen. Somit ist die Verteilung eines jeden der Y_i gleich der Verteilung von S_{p_k} und wir bekommen

$$\mathbb{P}\left(\bigcap_{i=1}^{q_k}\{Y_i \in r_k t J_\delta\}\right) = \prod_{i=1}^{q_k} \mathbb{P}(Y_i \in r_k t J_\delta) = (\mathbb{P}(S_{p_k} \in r_k t\, J_\delta))^{q_k}. \tag{36}$$

Außerdem ist auch $S_{p_k q_k} \in t r_k q_k J_\delta$, wenn alle $Y_i \in t r_k J_\delta$ sind, also

$$\bigcap_{i=1}^{q_k}\{Y_i \in t r_k J_\delta\} \leq \{S_{p_k q_k} \in t r_k q_k J_\delta\} \tag{37}$$

(da die Summe von q_k Zahlen aus einem Intervall in dem q_k-fachen des Intervalls liegt). Aus (36) und (37) folgt (33).

(34) erhält man aus der Tschebyscheffschen Ungleichung: Es ist

$$\begin{aligned}
\lambda_k &= \mathbb{P}(|S_{m_k} - S_{p_k q_k}| \geq \delta\alpha_k) \leq \frac{\mathbb{V}(S_{m_k} - S_{p_k q_k})}{\delta^2 \alpha_k^2} \\
&= \frac{(m_k - p_k q_k)\sigma^2}{\delta^2\alpha_k^2} \leq \frac{m_k}{\alpha_k^2}\frac{\sigma^2}{\delta^2} \underset{k\to\infty}{\longrightarrow} 0.
\end{aligned}$$

(35) schließlich sieht man folgendermaßen ein: Es ist

$$S_{p_k q_k} + (S_{m_k} - S_{p_k q_k}) = S_{m_k}$$

(das ist zugegebenermaßen nicht tiefsinnig) und die Summanden links sind unabhängig. Dies ergibt (35), denn

$$1 - \lambda_k = \mathbb{P}(\frac{1}{\alpha_k}(S_{m_k} - S_{p_k q_k}) \in U_\delta),$$

und $x \in U_\delta,\ y \in J_\delta$ impliziert $x + y \in J$.

Aus (33) - (35) leiten wir nun (32) ab. Es gilt

$$\begin{aligned}
& \liminf_{k\to\infty} \frac{m_k}{\alpha_k^2} \log \mathbb{P}\left(\frac{S_{m_k}}{\alpha_k} \in J\right) \\
\geq \;& \liminf \frac{m_k}{\alpha_k^2} \log(1-\lambda_k) + \liminf_{k\to\infty} \frac{m_k}{\alpha_k^2} q_k \log \mathbb{P}(S_{p_k} \in r_k t J_\delta) \\
= \;& \liminf_{k\to\infty} \frac{m_k q_k}{\alpha_k^2} \log \mathbb{P}\left(\frac{S_{p_k}}{r_k} \in t J_\delta\right) \\
= \;& t^{-2} \liminf_{k\to\infty} \log \mathbb{P}\left(\frac{S_{p_k}}{r_k} \in t J_\delta\right),
\end{aligned}$$

wobei die letzte Gleichung aus $q_k \sim \frac{\alpha_k^2}{t^2 m_k}$ und somit

$$\lim_{k\to\infty} \frac{m_k q_k}{\alpha_k^2} = \frac{1}{t^2}$$

folgt. Nach dem Satz von de-Moivre-Laplace (oder dem Zentralen Grenzwertsatz) folgt

$$\lim_{k\to\infty} \frac{S_{p_k}}{\sqrt{p_k}} = \nu_{0,\sigma^2} \quad \text{in Verteilung}$$

und somit mit Lemma 2.13 auch

$$\lim_{k\to\infty} \frac{S_{p_k}}{r_k} = \nu_{0,\sigma^2} \quad \text{in Verteilung,}$$

denn die Folgen $\sqrt{p_k}$ und r_k sind asymptotisch äquivalent. In der Tat gilt ja

$$\sqrt{p_k} \sim \frac{m_k t}{\alpha_k} = \frac{\alpha_k}{t} \cdot \frac{m_k \cdot t^2}{\alpha_k^2} \sim \frac{\alpha_k}{t q_k} = r_k.$$

Aus der Verteilungskonvergenz schließen wir nun mit Hilfe des Portmanteau-Theorems

$$\liminf_{k\to\infty} \mathbb{P}\left(\frac{S_{p_k}}{r_k} \in t J_\delta\right) \geq \nu_{0,\sigma^2}(t J_\delta).$$

Somit ergibt sich aus der obigen Ungleichungskette

$$\liminf_{k\to\infty} \frac{m_k}{\alpha_k^2} \log \mathbb{P}\left(\frac{S_{m_k}}{\alpha_k} \in J\right) \geq \frac{1}{t^2} \log \nu(t J_\delta).$$

Da $\delta > 0$ beliebig war und $J_\delta \uparrow J$ konvergiert (wobei wir δ wieder nur durch die rationalen Zahlen laufen lassen können), folgt (32).

Um nun hieraus (28) zu gewinnen, wählen wir als $J = (c, d)$ mit $c := x - \varepsilon$ und $d := x + \varepsilon$ für ein $\varepsilon > 0$. Dann ist offenbar

$$\mathbb{P}(A_k) = \mathbb{P}(|\frac{S_{m_k}}{\alpha_k} - x| < \varepsilon) = \mathbb{P}(\frac{S_{m_k}}{\alpha_k} \in J).$$

Erinnert man sich nun an die Dichte

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-J^2/2\sigma^2}$$

der ν_{0,σ^2}-Verteilung (bezüglich des Lebesgue-Maßes), so erhält man vermöge des Übergangs

$$y \mapsto y - xt$$

die Gleichungskette

$$\begin{aligned}
\nu_{0,\sigma^2}(tJ) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{tU_\varepsilon} e^{-\frac{y^2+2xty+x^2t^2}{2\sigma^2}} d\lambda(y) \\
&= e^{-\frac{x^2t^2}{2\sigma^2}} \int_{tU_\varepsilon} e^{-\frac{xt}{\sigma^2}y} \nu_{0,\sigma^2}(dy) \\
&= e^{-\frac{x^2t^2}{2\sigma^2}} \nu_{0,\sigma^2}(tU_\varepsilon) \int e^{-\frac{xt}{\sigma^2}y} \bar{\nu}(dy),
\end{aligned}$$

wobei $\bar{\nu}$ das Wahrscheinlichkeitsmaß mit Dichte

$$\frac{1}{\nu_{0,\sigma^2}(tU_\varepsilon)} 1_{tU_\varepsilon}(\cdot)$$

bzgl. des Maßes ν_{0,σ^2} ist. Da die Exponentialfunktion konvex ist, liefert die Jensensche Ungleichung

$$\nu_{0,\sigma^2}(tJ) \geq e^{-\frac{x^2t^2}{2\sigma^2}} \nu_{0,\sigma^2}(tU_\varepsilon) e^{-\int xty/\sigma^2} \bar{\nu}(dy).$$

Da $\bar{\nu}$ invariant ist unter der Transformation $y \mapsto -y$, folgt

$$\int y \, d\bar{\nu}(y) = 0,$$

also

$$\nu_{0,\sigma^2}(tJ) \geq \exp\left(-\frac{x^2t^2}{2\sigma^2}\right) \nu_{0,\sigma^2}(tU_\varepsilon).$$

Logarithmieren ergibt

$$\frac{1}{t^2} \log \nu(tJ) \geq -\frac{1}{2}\left(\frac{x}{\sigma}\right)^2 + \frac{1}{t^2} \log \nu_{0,\sigma^2}(tU_\varepsilon)$$

für alle $t > 0$. Bei festgehaltenem $\varepsilon > 0$ gilt $tU_\varepsilon \uparrow \mathbb{R}$, wenn $t \uparrow \infty$, also

$$\nu_{0,\sigma^2}(tU_\varepsilon) \to 1$$

und somit

$$\lim_{t\to\infty} \frac{1}{t^2} \log \nu_{0,\sigma^2}(tU_\varepsilon) = 0.$$

Dies ergibt

$$\lim_{t\to\infty} \frac{1}{t^2} \nu_{0,\sigma^2}(tJ) \geq -\frac{1}{2}\left(\frac{x}{\sigma}\right)^2$$

und wegen (32) also die Behauptung (29), also (28) und den Beweis des Lemmas. □

Bemerkung 2.16 *a) Lemma 2.15 und insbesondere (29) dort ist ein Spezialfall eines viel allgemeineren Sachverhalts, des sogenannten moderaten Abweichungsprinzips. Dieses sagt, dass für eine Folge von i.i.d. Zufallsvariablen* (X_i) *mit*

$$\mathbb{E}X_1 = 0 \quad \textit{und} \quad \mathbb{V}X_1 = \sigma^2$$

unter einer zusätzlichen Bedingung, die beispielsweise erfüllt ist, wenn

$$\mathbb{E}e^{tX_1} < \infty \quad \textit{für ein} \quad t > 0$$

gilt, für jedes Intervall das Folgende wahr ist: Für jede Folge $(b_n)_n$ *reeller Zahlen mit* $b_n > 0$*,*

$$\frac{b_n}{n} \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{und} \quad \frac{b_n^2}{n} \underset{n\to\infty}{\longrightarrow} \infty$$

gilt

$$\lim_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{b_n} \in I\right) = -\inf_{x\in I} \frac{x^2}{2\sigma^2}.$$

b) Man kann auch zeigen, dass die Voraussetzungen des Satzes vom iterierten Logarithmus nicht abgeschwächt werden können. Es gilt nämlich (siehe Feller [1])

Satz 2.17 *Es sei* (X_n) *eine Folge von i.i.d. Zufallsvariablen. Gilt dann*

$$\mathbb{P}\left[\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2nL(n)}} < \infty\right] > 0$$

für die Summenvariable $S_n = \sum_{i=1}^n X_i$*, so sind die* X_i *quadratisch integrierbar mit* $\mathbb{E}X_i = 0$ *für alle* i*.*

c) So interessant und ästhetisch ansprechend der Satz vom iterierten Logarithmus ist, so gering ist leider sein praktischer Nutzen. Die Funktion

$$(L(n))^{1/2} = \sqrt{\log \log n}$$

ist „für alle praktischen Zwecke beschränkt, sogar konstant“. In der Tat ist $\sqrt{L(10)} \sim 0,91$*, während man für eine Zahl wie* $n = 10^{88}$ *(die ungefähre Zahl von Teilchen im Universum inklusive der Photonen)*

$$\sqrt{L(10^{88})} \approx 2,3$$

erhält.

3 Bedingte Erwartungen

Um das Konzept der bedingten Erwartung zu verstehen, beginnen wir mit einem kleinen Beispiel.

Beispiel 3.1 *Es sei Ω eine endliche Menge, z.B. die Mitglieder einer endlichen Population (von Menschen). Die Zufallsvariable $X(\omega)$ bezeichne das Einkommen von Person ω. Sind wir also nur am Einkommen interessiert, so enthält X die vollständige Information unserer Umfrage. Nun stellen wir uns vor, wir seien Soziologen und wollten den Einfluss der Religion bzw. Konfession eines Menschen auf sein Einkommen messen. Wir interessieren uns also nicht mehr für die volle in X enthaltenen Information, sondern nur noch dafür, wie sich X auf den Teilmengen*

$$\{katholisch\}, \{protestantisch\}, \{islamisch\}, \{jüdisch\}, \{atheistisch\},$$

etc. von Ω verhält. Diese eingeschränkte Betrachtung von X führt zu einer neuen Zufallsvariable, die wir die bedingte Erwartung von X gegeben die Ereignisse

$$\{katholisch\}, \{protestantisch\}, \{islamisch\}, \{jüdisch\}, \{atheistisch\}$$

nennen werden.

Für die Definition der bedingten Erwartung benötigen wir eine Zufallsvariable

$$X : (\Omega, \mathcal{F}) \to \mathbb{R}$$

und eine Sub - σ - Algebra $\mathcal{A}$ von $\mathcal{F}$. Hieraus konstruieren wir eine neue $Zufallsvariable$, die wir mit $\mathbb{E}[X \mid \mathcal{A}] =: X_0$ bezeichnen. Die Eigenschaft von X_0 soll sein, dass

$$\int_C X_0 d\mathbb{P} = \int_C X d\mathbb{P}$$

für alle $C \in \mathcal{A}$ gilt. Somit enthält X_0 alle notwendigen Informationen, wenn wir uns nur auf Ereignisse aus $\mathcal{A}$ beschränken. Zunächst müssen wir einsehen, dass es so ein X_0 gibt und dass es sogar eindeutig ist.

Theorem 3.2 *Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine integrierbare Zufallsvariable . Ferner sei $\mathcal{C} \subseteq \mathcal{F}$ eine Sub-σ-Algebra. Dann gibt es eine (bis auf $\mathbb{P} - f.s.$ Gleichheit) eindeutige Zufallsvariable X_0, die $\mathcal{C}$–messbar ist und der Gleichheit*

$$\int_C X_0 d\mathbb{P} = \int_C X d\mathbb{P} \quad \text{für alle } C \in \mathcal{C} \tag{38}$$

genügt. Ist $X \geq 0$, dann ist auch $X_0 \geq 0 \quad \mathbb{P} - f.s.$

Beweis: Wir beginnen mit dem Fall $X \geq 0$. Wir bezeichnen mit $P_0 := \mathbb{P} \,|\mathcal{C}$ und $Q = X\mathbb{P} \,|\mathcal{C}$. Wir bemerken, dass sowohl P_0 als auch Q Maße auf $\mathcal{C}$ sind, P_0 ist sogar ein Wahrscheinlichkeitsmaß. Definitionsgemäß gilt

$$Q(C) = \int_C X d\mathbb{P}.$$

Also ist $Q(C) = 0$ für alle C mit $\mathbb{P}(C) = 0 = \mathbb{P}_0(C)$. Mit anderen Worten gilt $Q \ll P_0$. Nun bringen wir den Satz von Radon–Nikodym in Stellung. Danach gibt es eine $\mathcal{C}$–messbare Fukction $X_0 \geq 0$ auf Ω, so dass $Q = X_0 P_0$ gilt. Wir erhalten somit

$$\int_C X_0 dP_0 = \int_C X d\mathbb{P} \text{ für alle } C \in \mathcal{C}.$$

Also folgt

$$\int_C X_0 d\mathbb{P} = \int_C X d\mathbb{P} \text{ für alle } C \in \mathcal{C}.$$

Also genügt X_0 der Gleichung (38). Da für jedes $\bar{X}_0$, das $\mathcal{C}$–messbar ist, die Menge $\left\{\bar{X}_0 = X_0\right\}$ in $\mathcal{C}$ liegt, folgt, dass $\bar{X}_0 \geq 0$ $\mathbb{P} - f.s.$ gleich X_0 ist.

Für Zufallsvariablen X, die nicht notwendig positiv sind, wendet man die übliche Zerlegung von X in Positiv- und Negativteil an. □

Übung 3.3 *Man beweise den obigen Satz für integrierbare $X : \Omega \to \mathbb{R}$, die nicht notwendig positiv sind.*

Definition 3.4 *Unter den Bedingungen von Satz 3.2 heißt die dort auftretende (und $\mathbb{P}$-f.s. eindeutige) Zufallsvariable X_0 die bedingte Erwartung von X gegeben $\mathcal{C}$. Sie wird mit*

$$X_0 =: \mathbb{E}[X \mid \mathcal{C}] =: \mathbb{E}^{\mathcal{C}}[X]$$

bezeichnet.

Falls $\mathcal{C}$ von einer Folgen von Zufallsvariablen $(Y_i)_{i \in I}$ erzeugt wird (d.h., ist $\mathcal{C} = \sigma(Y_i, i \in I)$), schreiben wir auch

$$\mathbb{E}\left[X \mid (Y_i)_{i \in I}\right] = \mathbb{E}[X \mid \mathcal{C}].$$

Ist $I = \{1, ..., n\}$, schreiben wir auch $\mathbb{E}[X \mid Y_1, ..., Y_n]$.

Man bemerke, dass man zur Überprüfung, ob Y eine (Version der) bedingte(n) Erwartung einer Zufallsvariablen X bezüglich Teil-σ-Algebra $\mathcal{C}$ ist, die Identität

$$\int_C Y d\mathbb{P} = \int_C X d\mathbb{P}$$

für alle $C \in \mathcal{C}$ überprüfen muss. Dies legt $\mathbb{E}[X \mid \mathcal{C}]$ nur $\mathbb{P} - f.s.$ auf Mengen $C \in \mathcal{C}$ fest. Wir sprechen daher auch von verschiedenen Versionen der bedingten Erwartung.

Beispiel 3.5 *1. Falls $\mathcal{C} = \{\emptyset, \Omega\}$, dann ist die konstante Zufallsvariable $\mathbb{E}X$ eine Version von $\mathbb{E}[X \mid \mathcal{C}]$. In der Tat: für $C = \emptyset$ erfüllt jede beliebige Zufallsvariable die gewünschte Identität. Ist $C = \Omega$, so folgt*

$$\int_C X = \mathbb{E}X = \int \mathbb{E}X d\mathbb{P}.$$

2. *Natürlich erfüllt X persönlich die gewünschte Identität. Das hilft aber i.a. nicht, denn X ist in der Regel nicht $\mathcal{C}$–messbar und die ganze Idee der bedingten Erwartung ist, eine „einfachere“, d.h. $\mathcal{C}$– messbare Variante von X zu finden.*

3. *Ist $\mathcal{C}$ durch eine Menge von Ereignissen $(B_i)_{i\in I}$ erzeugt, wobei I eine abzählbare Menge ist und die $B_i \in \mathcal{A}$ erfüllen (wobei $(\Omega, \mathcal{A}, \mathbb{P})$ der zu Grunde liegende Raum ist) und*

$$\Omega = \dot{\bigcup_{i\in I}} B_i$$

(d.h. insbesondere sind die B_i paarweise disjunkt), dann gilt

$$\mathbb{E}\left[X \mid \mathcal{C}\right] = \sum_{i\in I} 1_{B_i} \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X d\mathbb{P} \qquad \mathbb{P}-f.s. \tag{39}$$

Dies ist Inhalt der folgenden Übung.

Übung 3.6 *Man zeige, dass* (39) *gilt.*

Nun sammeln wir einige wesentliche Eigenschaften der bedingten Erwartung.

Proposition 3.7 *Für Zufallsvariablen*

$$Y, X : (\Omega, \mathcal{A}) \to \left(\mathbb{R}, B^1\right) \qquad .$$

und eine σ-Algebra $\mathcal{C} \subseteq \mathcal{A}$ gilt das Folgende:

1. $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{C}\right]\right] = \mathbb{E}X$
2. *Falls X $\mathcal{C}$–messbar ist, so gilt* $\mathbb{E}\left[X \mid \mathcal{C}\right] = X \quad \mathbb{P}-f.s.$
3. *Ist $X = Y$ $\mathbb{P}-f.s.$, so gilt* $\mathbb{E}\left[X \mid \mathcal{C}\right] = \mathbb{E}\left[Y \mid \mathcal{C}\right] \quad \mathbb{P}-a.s.$
4. *Ist $X \equiv \alpha$, so ist* $\mathbb{E}\left[X \mid \mathcal{C}\right] = \alpha \quad \mathbb{P}-f.s.$
5. $\mathbb{E}\left[\alpha X + \beta Y \mid \mathcal{C}\right] = \alpha\mathbb{E}\left[X \mid \mathcal{C}\right] + \beta\mathbb{E}\left[X \mid \mathcal{C}\right] \quad \mathbb{P}-f.s.$
   Hierbei sind $\alpha, \beta \in \mathbb{R}$.
6. $X \leq Y \quad \mathbb{P}-f.s.$ *impliziert* $\mathbb{E}\left[X \mid \mathcal{C}\right] \leq \mathbb{E}\left[Y \mid \mathcal{C}\right] \quad \mathbb{P}-f.s.$
7. *Es gilt*
$$|\mathbb{E}[X|\mathcal{C}]| \leq \mathbb{E}[|X|\,|\mathcal{C}]|.$$

Beweis: 1., 2., 3., 4. und 5. sind offensichtlich (1. folgt zum Beispiel, da Ω in jeder σ–Algebra liegt).

Für 6. kann man wegen 3. annehmen, dass $X \leq Y$ auf ganz Ω gilt. Dann aber gibt es offenbar eine Zufallsvariable $Z \geq 0$, so dass

$$Y(\omega) = X(\omega) + Z(\omega) \qquad \forall \omega \in \Omega$$

gilt und die Behauptung folgt wegen 5. und $\mathbb{E}[Z|\mathcal{C}] \geq 0$ $\mathbb{P}$-f.s.

7. ist für nicht-negative X evident. Für allgemeine X folgt dies aus $X^{\pm} \leq |X|$, sowie aus der Zerlegung $X = X^+ - X^-$ und 5. und 6. □

Die folgenden beiden Sätze haben Beweise, die beinahe identisch sind mit den Beweisen, die man für die entsprechenden Theoreme für Erwartungswerte anstelle bedingter Erwartungen gibt. (Hierbei sollte allerdings stets im Hinterkopf behalten werden, dass bedingte Erwartungen Zufallsvariablen sind, während es sich bei Erwartungswerten um Zahlen handelt).

Theorem 3.8 *(monotone Konvergenz) Es sei (X_n) eine wachsende Folge von nicht-negativen Zufallsvariablen und $X = \sup X_n$. Dann gilt*

$$\sup_n \mathbb{E}\left[X_n \mid \mathcal{C}\right] = \lim_{n\to\infty} \mathbb{E}\left[X_n \mid \mathcal{C}\right] = \mathbb{E}\left[X \mid \mathcal{C}\right].$$

Beweis: Wegen 3. und 6. aus dem vorhergehenden Satz kann die Folge der $\mathbb{E}\left[X_n \mid \mathcal{C}\right]$ als wachsend vorausgesetzt werden. Die Behauptung folgt nun durch Limesübergang in (38), den man mittels des Satzes von der monotonen Konvergenz vollzieht. □

Theorem 3.9 *(Lebesguescher Konvergenzsatz für bedingte Erwartungen) Es sei (X_n) eine Folge von Zufallsvariablen, die punktweise gegen eine integrierbare Zufallsvariable X konvergiere, so dass es eine integrierbare Zufallsvariable Y gibt mit $|Y| \geq X$. Dann gilt*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n \mid \mathcal{C}\right] = \mathbb{E}\left[X \mid \mathcal{C}\right].$$

Beweis: O.B.d.A nimmt Y und damit auch die X_n nur reelle Werte an (also nicht $\pm\infty$). Setze

$$X_n' := \sup_{k \geq n} X_k$$

und

$$X_n'' := \inf_{k \geq n} X_k.$$

Dann gilt

$$-Y \leq X_n'' \leq X_n \leq X_n' \leq Y \qquad \forall n \in \mathbb{N}.$$

Ferner sind $(Y - X_n')$ und $(Y + X_n'')$ monoton wachsende Folgen integrierbarer, nicht-negativer Zufallsvariablen mit Supremum $Y - \limsup X_n$ bzw. $Y + \liminf X_n$. Nach

Voraussetzung konvergiert die Folge der X_n fast sicher gegen X. Wegen der Linearität der bedingten Erwartung und dem Satz von der monotonen Konvergenz konvergieren somit auch die Folgen $(\mathbb{E}[X'_n|\mathcal{C}])$ und $(\mathbb{E}[X''_n|\mathcal{C}])$ fast sicher gegen $(\mathbb{E}[X|\mathcal{C}])$. Aus

$$X''_n \leq X_n \leq X'_n$$

und 6. aus dem obigen Satz folgt daher die fast sichere Konvergenz von $(\mathbb{E}[X_n|\mathcal{C}])$ gegen $(\mathbb{E}[X|\mathcal{C}])$. □

Auch die Jensensche Ungleichung hat eine Version für bedingte Erwartungswerte.

Theorem 3.10 *(Jensensche Ungleichung)*

Es sei X eine integrierbare Zufallsvariable mit Werten in einem offenen Intervall $I \subseteq \mathbb{R}$ und es sei

$$q : I \to \mathbb{R}$$

eine konvexe Funktion. Dann gilt für jede σ–Algebra $\mathcal{C} \subseteq \mathcal{A}$

$$\mathbb{E}\left[X \mid \mathcal{C}\right] \in I$$

und

$$q\left(\mathbb{E}\left[X \mid \mathcal{C}\right]\right) \leq \mathbb{E}\left[q \circ X \mid \mathcal{C}\right].$$

Beweis: Der Beweis ist nur eine geringfügige Modifikation des Beweises der Jensenschen Ungleichung für gewöhnliche Erwartungswerte. Wir verweisen auf das Wahrscheinlichkeitstheoriebuch von Bauer. □

Eine direkte Konsequenz von 3.10 ist, dass für $1 \leq p < \infty$ gilt:

$$\left|\mathbb{E}\left[X \mid \mathcal{C}\right]\right|^p \leq \mathbb{E}\left[|X|^p \mid \mathcal{C}\right],$$

woraus

$$\mathbb{E}\left(\left|\mathbb{E}\left[X \mid \mathcal{C}\right]\right|^p\right) \leq \mathbb{E}\left(|X|^p\right)$$

folgt. Bezeichnen wir mit

$$N_p\left(f\right) = \left(\int |f|^p \, d\mathbb{P}\right)^{1/p},$$

so ergibt sich

$$N_p\left(\mathbb{E}\left[X \mid \mathcal{C}\right]\right) \leq N_p\left(X\right), \quad X \in \mathcal{L}^p\left(\mathbb{P}\right).$$

Dies gilt für $1 \leq p < \infty$. Der Fall $p = \infty$, der bedeutet, dass mit X auch $\mathbb{E}\left[X \mid \mathcal{C}\right]$ $\mathbb{P}$-f.s. beschränkt ist, folgt aus 4. und 7. des obigen Satzes.

Wir formulieren nun die Definition der bedingten Erwartung leicht um, um weitere ihrer Eigenschaften besser erkennen zu können.

Lemma 3.11 *Eine $\mathcal{C}$–messbare integrierbare Zufallsvariable $X_0 : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}^1)$ ist eine Version von $\mathbb{E}[X \mid \mathcal{C}]$ (wobei wie immer X als integrierbar vorausgesetzt sei), genau dann wenn*

$$\int ZX_0 d\mathbb{P} = \int ZX d\mathbb{P} \tag{40}$$

für alle $\mathcal{C}$ - messbaren, positiven oder beschränkten Zufallsvariablen Z gilt.

Beweis: (40) impliziert (38), wenn man $Z = 1_C$ für $C \in \mathcal{C}$ wählt.

Gilt andererseits (38), so auch (40) für Treppenfunktionen. Für positive Zufallsvariablen X folgt (40) mittels monotoner Konvergenz. Schließlich folgt (40) für beschränkte Zufallsvariablen durch die Aufteilung

$$Z = Z^+ - Z^-.$$

□

Nun sind wir in der Lage, weitere Eigenschaften der bedingten Erwartung zu untersuchen. Die ersten fassen wir unter dem Namen "Glättungseigenschaften" zusammen.

Theorem 3.12 *(Glättungseigenschaften der bedingten Erwartung)*

1. *Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X, Y \geq 0$ (oder $X \in \mathcal{L}^p(\mathbb{P})$ und $Y \in \mathcal{L}^q(\mathbb{P})$, $1 \leq p \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$) seien Zufallsvariablen . Falls $\mathcal{C} \subseteq \mathcal{F}$ und X $\mathcal{C}$ - messbar ist, dann gilt*

$$\mathbb{E}[XY \mid \mathcal{C}] = X\mathbb{E}[Y \mid \mathcal{C}]. \tag{41}$$

2. *Unter den Voraussetzungen von Teil 1 gilt*

$$\mathbb{E}[Y \cdot \mathbb{E}[X \mid \mathcal{C}] \mid \mathcal{C}] = \mathbb{E}[Y \mid \mathcal{C}] \mathbb{E}[X \mid \mathcal{C}].$$

3. *Sind $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathcal{F}$ σ–Algebren mit $\mathcal{C}_1 \subseteq \mathcal{C}_2$, so gilt*

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{C}_2] \mid \mathcal{C}_1] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{C}_1] \mid \mathcal{C}_2] = \mathbb{E}[X \mid \mathcal{C}_1].$$

Beweis:

1. Es sei zunächst $X, Y \geq 0$ angenommen. Ist $Z \geq 0$ und $\mathcal{C}$ - messbar, so gilt

$$\int ZXY d\mathbb{P} = \int ZX\mathbb{E}[Y \mid \mathcal{C}] d\mathbb{P}.$$

Dies folgt in der Tat aus dem vorangegangenen Lemma, da ZX $\mathcal{C}$-messbar ist. Da aber auch Z messbar ist, gilt auch

$$\int ZXY d\mathbb{P} = \int Z\mathbb{E}\left[XY \mid \mathcal{C}\right] d\mathbb{P}.$$

Da $X\mathbb{E}\left[Y \mid \mathcal{C}\right]$ $\mathcal{C}$-messbar ist, erhalten wir

$$\mathbb{E}\left[XY \mid \mathcal{C}\right] = X\mathbb{E}\left[Y \mid \mathcal{C}\right] \qquad \mathbb{P} - f.s.$$

Im Falle, dass $X \in \mathcal{L}^p\left(\mathbb{P}\right), Y \in \mathcal{L}^q\left(\mathbb{P}\right)$, gilt $XY \in \mathcal{L}^1\left(\mathbb{P}\right)$ und wir können wie oben schließen.

2. Für diesen Teil sei daran erinnert, dass $X \in \mathcal{L}^p\left(\mathbb{P}\right)$ impliziert, dass $\mathbb{E}\left[X \mid \mathcal{C}\right] \in \mathcal{L}^p\left(\mathbb{P}\right)$. Also kann $\mathbb{E}\left[X \mid \mathcal{C}\right]$ die Rolle von X in Teil 1 übernehmen. Dies ergibt

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{C}\right] Y] \mathcal{C}\right] = \mathbb{E}\left[X \mid \mathcal{C}\right] \mathbb{E}\left[Y \mid \mathcal{C}\right],$$

also die Behauptung.

3. Zunächst beachte man, dass natürlich $\mathbb{E}\left[X \mid \mathcal{C}_1\right]$ $\mathcal{C}_1$–messbar und, da $\mathcal{C}_1 \subseteq \mathcal{C}_2$, somit auch $\mathcal{C}_2$–messbar ist. Die Eigenschaften der bedingten Erwartung implizieren daher

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{C}_1\right] \mid \mathcal{C}_2\right] = \mathbb{E}\left[X \mid \mathcal{C}_1\right], \quad \mathbb{P} - f.s.$$

Weiter gilt für alle $C \in \mathcal{C}_1$

$$\int_C \mathbb{E}\left[X \mid \mathcal{C}_1\right] d\mathbb{P} = \int_C X d\mathbb{P}.$$

Also folgt für alle $C \in \mathcal{C}_1$

$$\int_C \mathbb{E}\left[X \mid \mathcal{C}_1\right] d\mathbb{P} = \int_C \mathbb{E}\left[X \mid \mathcal{C}_2\right] d\mathbb{P}.$$

Dies aber heißt

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{C}_2\right] \mid \mathcal{C}_1\right] = \mathbb{E}\left[X \mid \mathcal{C}_1\right] \qquad \mathbb{P} - f.s.$$

□

Bevor wir diesen Satz benutzen, um noch eine andere Charakterisierung der bedingten Erwartung herzuleiten, geben wir eine der wesentlichen Eigenschaften bedingter Erwartungen wieder.

Theorem 3.13 *Seien $\mathcal{C}_1$ und $\mathcal{C}_2$ Unter-σ-Algebren der zu Grunde liegenden σ–Algebra $\mathcal{A}$. Sei weiter $\mathcal{C} := \sigma(\mathcal{C}_1, \mathcal{C}_2)$ und X eine integrierbare Zufallsvariable . Ist dann die von X und $\mathcal{C}_1$ erzeugte σ–Algebra $\sigma(X, \mathcal{C}_1)$ unabhängig von $\mathcal{C}$, so gilt*

$$\mathbb{E}\left[X \mid \mathcal{C}\right] = \mathbb{E}\left[X \mid \mathcal{C}_1\right].$$

Beweis: Sei X_0 eine Version von $\mathbb{E}[X \mid \mathcal{C}_1]$. Diese ist $\mathcal{C}$–messbar und wir wollen zeigen, dass X_0 auch eine Version von $\mathbb{E}[X \mid \mathcal{C}]$ ist.

Sei dazu $C \in \mathcal{C}$. Wegen der Integrierbarkeit von X (und damit auch von X_0) ist das System $\mathcal{D}$ aller $C \in \mathcal{C}$ mit

$$\int_C X d\mathbb{P} = \int_C X_0 d\mathbb{P}$$

ein Dynkin-System (dies folgt direkt aus den Eigenschaften des Integrals). Also genügt es, diese Gleichheit für alle Mengen C eines durchschnittstabilen Erzeugers von $\mathcal{C}$ zu beweisen. Eine solcher ist z.B.

$$\mathcal{E} := \{C_1 \cap C_2 : C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2\}.$$

Für jede Menge $C_1 \cap C_2 \in \mathcal{E}$ gilt nun auf Grund der Unabhängigkeitsvoraussetzungen

$$\int_{C_1 \cap C_2} X_0 d\mathbb{P} = \mathbb{E}(1_{C_1} 1_{C_2} X_0) = \mathbb{E}(1_{C_2})\mathbb{E}(1_{C_1} X_0).$$

Da X_0 eine Version der bedingten Erwartung von X bezï¿½glich $\mathcal{C}_1$ ist gilt weiter

$$\mathbb{E}(1_{C_1} X_0) = \mathbb{E}(1_{C_1} X)$$

und daher

$$\int_{C_1 \cap C_2} X_0 d\mathbb{P} = \mathbb{E}(1_{C_2})\mathbb{E}(1_{C_1} X).$$

Da 1_{C_2} und $1_{C_1} X$ unabhängig sind, ergibt dies

$$\int_{C_1 \cap C_2} X_0 d\mathbb{P} = \int_{C_1 \cap C_2} X d\mathbb{P},$$

was zu zeigen war. □

Korollar 3.14 *Für jede von einer σ–Algebra $\mathcal{C} \subseteq \mathcal{A}$ unabhängige, integrierbare Zufallsvariable X gilt*

$$\mathbb{E}[X|\mathcal{C}] = \mathbb{E}(X).$$

Beweis: Man wähle im vorangegangenen Theorem einfach $\mathcal{C}_1 := \{\emptyset, \Omega\}$. □

Wir werden nun noch eine weitere Charakterisierung der bedingten Erwartung kennenlernen. Dazu sei $X \in \mathcal{L}^2(\mathbb{P})$ und $X_0 := \mathbb{E}[X \mid \mathcal{C}]$ für eine σ–Algebra $\mathcal{C} \subseteq \mathcal{F}$. Dann ist $X_0 \in \mathcal{L}^2(\mathbb{P})$ und nach (41)

$$\mathbb{E}[X X_0 \mid \mathcal{C}] = X_0 \mathbb{E}[X \mid \mathcal{C}] = X_0^2.$$

Aus den Eigenschaften der bedingten Erwartung folgt somit

$$\mathbb{E}[X X_0] = \mathbb{E} X_0^2: \tag{42}$$

Damit erhalten wir

$$\mathbb{E}\left[(X - X_0)^2\right] = \mathbb{E}\left(X^2\right) - \mathbb{E}\left(X_0^2\right).$$

Theorem 3.15 *Für alle $X \in \mathcal{L}^2(\mathbb{P})$ und jede σ–Algebra $\mathcal{C} \subseteq \mathcal{F}$ gilt für die bedingte Erwartung $\mathbb{E}[X \mid \mathcal{C}]$, dass sie (bis auf fast sichere Gleichheit) die eindeutige $\mathcal{C}$–messbare Zufallsvariable $X_0 \in \mathcal{L}^2(\mathbb{P})$ mit*

$$\mathbb{E}\left[(X - X_0)^2\right] = \min \quad \mathbb{E}\left[(X-Y)^2, Y \in \mathcal{L}^2(\mathbb{P}), Y\ \mathcal{C}\text{–}\textit{messbar}\right]$$

ist.

Beweis: Es sei $Y \in \mathcal{L}^2(\mathbb{P})$ $\mathcal{C}$-messbar. Setze $X_0 := \mathbb{E}[X \mid \mathcal{C}]$. Wie oben (in (42)) zeigt man, dass

$$\mathbb{E}[XY] = \mathbb{E}[X_0 Y]$$

gilt.

Zusammen mit (42) ergibt dies

$$\mathbb{E}\left[(X-Y)^2\right] - \mathbb{E}\left[(X-X_0)^2\right] = \mathbb{E}\left[(X_0 - Y)^2\right]. \tag{43}$$

Da Quadrate nicht-negativ sind, folgt

$$\mathbb{E}\left[(X-X_0)^2\right] \leq \mathbb{E}\left[(X-Y)^2\right].$$

Ist andererseits

$$\mathbb{E}\left[(X-X_0)^2\right] = \mathbb{E}\left[(X-Y)^2\right],$$

so folgt

$$\mathbb{E}\left[(X_0-Y)^2\right] = 0$$

und somit $Y = X_0 = \mathbb{E}[X \mid \mathcal{C}] \quad \mathbb{P} - f.s.$ □

Der letzte Satz besagt, dass $\mathbb{E}[X \mid \mathcal{C}]$ für $X \in \mathcal{L}^2(\mathbb{P})$ die beste Näherung von X durch $\mathcal{C}$–messbare Funktionen im Sinne der L^2-Distanz ist. Es ist die L^2-Projektion von X auf die Menge der quadratisch integrierbaren $\mathcal{C}$–messbaren Funktionen.

Mithilfe der bedingten Erwartung können wir auch die bedingten Wahrscheinlichkeiten neu definieren.

Definition 3.16 *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathcal{C} \subseteq \mathcal{F}$ eine Unter-σ–Algebra. Für $A \in \mathcal{F}$ heißt*

$$\mathbb{P}[A \mid \mathcal{C}] := \mathbb{E}[1_A \mid \mathcal{C}]$$

die bedingte Wahrscheinlichkeit von A gegeben $\mathcal{C}$.

Beispiel 3.17 *In der Situation von Beispiel 3.5.3. ist die bedingte Wahrscheinlichkeit von $A \in \mathcal{F}$ durch*

$$\mathbb{P}(A \mid \mathcal{C}) = \sum_{i \in I} \mathbb{P}(A \mid B_i)\, 1_{B_i} := \sum_{i \in I} \frac{\mathbb{P}(A \cap B_i)\, 1_{B_i}}{\mathbb{P}(B_i)}$$

gegeben.

In einem letzten Schritt werden wir bedingte Erwartungen bezüglich Ereignissen, die Wahrscheinlichkeit null haben einführen (alleredings werden wir nicht beweisen, dass es solche bedingten Erwartungen gibt). Natürlich wird das im allgemeinen zu Unsinn führen, allerdings können wir im Falle von bedingten Erwartungen $\mathbb{E}\left[X \mid Y=y\right]$, wobei X, Y Zufallsvariablen sind, so dass der Vektor (X, Y) eine (zweidimensionale) Lebesguedichte hat, können wir diesem Ausruck einen Sinn geben.

Theorem 3.18 *Es seien X, Y reellwertigeZufallsvariablen , so dass (X, Y) die Dichte $f : \mathbb{R} \to \mathbb{R}^{+}_{\{0\}}$ bzgl. des zweidimensionalen Lebesguemaßes λ^2 hat. Weiter sei X integrierbar und es gelte*

$$f_0\left(y\right) := \int f\left(x, y\right) dx > 0 \qquad \textit{für alle } y \in \mathbb{R}.$$

Dann gilt für jede Funktion

$$y \mapsto \mathbb{E}\left(X \mid Y = y\right)$$

die Identität

$$\mathbb{E}\left(X \mid Y = y\right) = \frac{1}{f_0\left(y\right)} \int x f\left(x, y\right) dx \qquad f\ddot{\imath}\text{¿}\frac{1}{2}r\ \mathbb{P}_y - f.a. \quad y \in \mathbb{R}.$$

Weiter ist

$$\mathbb{E}\left(X \mid Y\right) = \frac{1}{f_0\left(Y\right)} \int x f\left(x, Y\right) dx \qquad \mathbb{P} - f.s.$$

4 Markoff-Ketten

Bisher haben wir uns hauptsächlich mit unabhängigen Ereignissen und unabhängigen Zufallsgrößen beschäftigt. *Andrej Andrejewitsch Markoff* (1856–1922) hat erstmalig in einer Arbeit 1906 Zufallsexperimente analysiert, bei denen die einfachste Verallgemeinerung der unabhängigen Versuchsfolge betrachtet wurde. Man spricht bei diesen Versuchsfolgen heute von Markoff-Ketten. Wir werden sehen, dass sehr viele Modelle Markoff-Ketten sind. Man kann sie anschaulich wie folgt beschreiben: Ein Teilchen bewegt sich in diskreter Zeit auf einer höchstens abzählbaren Menge I. Befindet es sich auf einem Platz $i \in I$, so wechselt es mit gewissen Wahrscheinlichkeiten (die von i abhängen) zu einem anderen Platz $j \in I$. Diese Übergangswahrscheinlichkeiten hängen aber nicht weiter von der ,,Vorgeschichte" ab, das heißt von dem Weg, auf dem das Teilchen zum Platz i gekommen ist.

Definition 4.1 *Es sei I eine nichtleere, höchstens abzählbare Menge. Eine Matrix $\mathbb{P} = (p_{ij})_{i,j\in I}$ heißt* stochastische Matrix (stochastic matrix), *wenn $p_{ij} \in [0,1]$ für alle $i, j \in I$ und $\sum_{j\in I} p_{ij} = 1$ für alle $i \in I$ gelten. Die Komponenten p_{ij} heißen* Übergangswahrscheinlichkeiten (transition probabilities). *Eine stochastische Matrix wird im Zusammenhang mit Markoff-Ketten auch* Übergangsmatrix (transition matrix) *genannt. Eine auf einem Grundraum $(\Omega, \mathcal{F}, P)$ definierte Zufallsgröße $X : \Omega \to I$ nennt man* I-wertige Zufallsgröße.

Definition 4.2 *Eine endlich oder unendlich lange Folge $X_0, X_1, X_2, \ldots$ I-wertiger Zufallsgrößen heißt (zeitlich homogene, time homogeneous)* Markoff-Kette (Markov chain) *mit stochastischer Matrix $\mathbb{P}$, wenn für alle $n \geq 0$ und alle $i_0, i_1, \ldots, i_n, i_{n+1} \in I$ mit $P(X_0 = i_0, \ldots, X_n = i_n) > 0$*

$$P(\, X_{n+1} = i_{n+1} \mid X_0 = i_0,\, X_1 = i_1, \ldots,\, X_n = i_n \,) = p_{i_n i_{n+1}}$$

gilt.

Die Startverteilung (initial distribution) *ν einer Markoff-Kette ist definiert durch $\nu(i) = P(X_0 = i)$ für alle $i \in I$. Oft schreibt man P_ν, um die Startverteilung zu betonen. Ist die Startverteilung auf einen Punkt konzentriert, d. h. gilt $\nu(i) = 1$ für ein $i \in I$, so schreiben wir meist P_i anstelle von P_ν.*

Satz 4.3 *Sei $\{X_n\}_{n\in\mathbb{N}_0}$ eine Markoff-Kette mit Startverteilung ν.*

a) Für alle $n \in \mathbb{N}_0$ und $i_0, i_1, \ldots, i_n \in I$ gilt

$$P(X_0 = i_0,\, X_1 = i_1, \ldots, X_n = i_n) = \nu(i_0) p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.$$

b) Es seien $n < m$ und $i_n \in I$ sowie $A \subset I^{\{0,1,\ldots,n-1\}}$ und $B \subset I^{\{n+1,\ldots,m\}}$. Falls $P((X_0, X_1, \ldots, X_{n-1}) \in A,\, X_n = i_n) > 0$ ist, so gilt

$$\begin{aligned} & P(\,(X_{n+1}, \ldots, X_m) \in B \mid (X_0, \ldots, X_{n-1}) \in A,\, X_n = i_n\,) \\ = \;& P(\,(X_{n+1}, \ldots, X_m) \in B \mid X_n = i_n\,). \end{aligned}$$

Beweis. (a) folgt durch Induktion nach n: Definitionsgemäß gilt die Behauptung für $n = 0$. Gelte die Behauptung für ein $n \in \mathbb{N}_0$ und seien $i_0, i_1, \ldots, i_{n+1} \in I$. Ist $P(X_0 = i_0, \ldots, X_n = i_n) = 0$, so gilt die behauptete Formel ebenfalls für $n+1$: Ist $P(X_0 = i_0, \ldots, X_n = i_n) > 0$, so folgt aus Definition 4.2

$$\begin{aligned} P(X_0 = i_0, \ldots, X_n = i_n,\, X_{n+1} = i_{n+1}) &= P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) \\ &\quad \times P(X_0 = i_0, \ldots, X_n = i_n) \\ &= \nu(i_0) p_{i_0 i_1} \ldots p_{i_{n-1} i_n} p_{i_n i_{n+1}}. \end{aligned}$$

(b) Sei $P((X_0, X_1, \ldots, X_{n-1}) \in A,\, X_n = i_n) > 0$. Mit der Definition der bedingten Wahrscheinlichkeit und Teil (a) folgt

$$\begin{aligned} & P(\,(X_{n+1}, \ldots, X_m) \in B \mid (X_0, \ldots, X_{n-1}) \in A,\ X_n = i_n\,) \\ = {} & \frac{P(\,(X_{n+1}, \ldots, X_m) \in B,\ X_n = i_n,\ (X_0, \ldots, X_{n-1}) \in A\,)}{P(\,(X_0, \ldots, X_{n-1}) \in A,\ X_n = i_n\,)} \\ = {} & \frac{\sum_{(i_{n+1}, \ldots, i_m) \in B} \sum_{(i_0, \ldots, i_{n-1}) \in A} \nu(i_0) p_{i_0 i_1} \ldots p_{i_{m-1} i_m}}{\sum_{(i_0, \ldots, i_{n-1}) \in A} \nu(i_0) p_{i_0 i_1} \ldots p_{i_{n-1} i_n}} \\ = {} & \sum_{(i_{n+1}, \ldots, i_m) \in B} p_{i_n i_{n+1}} p_{i_{n+1} i_{n+2}} \ldots p_{i_{m-1} i_m}. \end{aligned}$$

Dieser Ausdruck hängt nicht von A ab, insbesondere führt also die obige Rechnung für $A = I^{\{0,1,\ldots,n-1\}}$ zum gleichen Resultat. Aber für $A = I^{\{0,1,\ldots,n-1\}}$ gilt die in (b) behauptete Formel. □

Bemerkung 4.4 Die Aussage von (b) heißt *Markoff-Eigenschaft (Markov property)*. Sie spiegelt genau die eingangs erwähnte Eigenschaft wieder, daß in einer Markoff-Kette die Wahrscheinlichkeit, zur Zeit $n+1$ in einen beliebigen Zustand zu gelangen, nur vom Zustand zur Zeit n abhängt, aber nicht davon, in welchem Zustand die Kette früher war. Nicht jede Folge von I-wertigen Zufallsgrößen mit dieser Eigenschaft ist eine homogene Markoff-Kette in unserem Sinn: Die Übergangswahrscheinlichkeiten können nämlich noch von der Zeit abhängen. Genauer: Sei $X_0, X_1, \ldots$ eine Folge I-wertiger Zufallsgrößen, die die Eigenschaft aus Satz 4.3 b) hat. Dann existiert eine Folge $\{\mathbb{P}_n\}_{n \in \mathbb{N}_0}$ von stochastischen Matrizen $\mathbb{P}_n = (p_n(i,j))_{i,j \in I}$ mit

$$P(X_0 = i_0, \ldots, X_n = i_n) = \nu(i_0) p_0(i_0, i_1) \ldots p_{n-1}(i_{n-1}, i_n)$$

für alle $n \in \mathbb{N}_0$ und $i_0, \ldots, i_n \in I$. Der Beweis sei dem Leser überlassen. Man spricht dann von einer (zeitlich) inhomogenen Markoff-Kette. Wir werden jedoch nur (zeitlich) homogene Ketten betrachten, ohne dies jedesmal besonders zu betonen.

Satz 4.5 *Es seien $\mathbb{P} = (p_{ij})_{i,j \in I}$ eine stochastische Matrix, ν eine Verteilung auf I und $N \in \mathbb{N}_0$. Dann gibt es eine abzählbare Menge Ω, eine Wahrscheinlichkeitsverteilung p auf Ω und Abbildungen $X_i : \Omega \to I$ für alle $i \in \{0, 1, \ldots, N\}$, so dass $X_0, \ldots, X_N$ eine homogene Markoff-Kette mit Startverteilung ν und Übergangsmatrix $\mathbb{P}$ ist.*

Beweis. Es sei $\Omega := I^{\{0,\dots,N\}}$ und $p(i_0,\dots,i_N) := \nu(i_0)p_{i_0i_1}\dots p_{i_{N-1}i_N}$ sowie $X_n(i_0,\dots,i_N) = i_n$ für alle $n \in \{0,1,\dots,N\}$ und $(i_0,\dots,i_N) \in \Omega$. Da die Summe der Komponenten der stochastischen Matrix $\mathbb{P}$ in jeder Zeile gleich eins ist, gilt für alle $n \in \{0,1,\dots,N\}$ und $(i_0,\dots,i_n) \in I^{\{0,\dots,n\}}$

$$\begin{aligned} P(X_0 = i_0,\dots,X_n = i_n) &= \sum_{(i_{n+1},\dots,i_N)\in I^{\{n+1,\dots,N\}}} P(X_0 = i_0,\dots,X_N = i_N) \\ &= \sum_{(i_{n+1},\dots,i_N)\in I^{\{n+1,\dots,N\}}} \nu(i_0)p_{i_0i_1}\dots p_{i_{N-1}i_N} \\ &= \nu(i_0)p_{i_0i_1}\dots p_{i_{n-1}i_n}. \end{aligned}$$

Dieses Produkt ist größer als Null genau dann, wenn jeder Faktor größer als Null ist. Ist dies der Fall, so ist offenbar

$$P(\,X_{n+1} = i_{n+1} \mid X_0 = i_0,\dots,X_n = i_n\,) = p_{i_ni_{n+1}}.$$

□

Bemerkung. Nachfolgend soll stets von einer unendlich langen Markoff-Kette ausgegangen werden, dies jedoch nur wegen einer bequemeren Notation. Alle nachfolgenden Überlegungen benötigen die Konstruktion einer unendlichen Markoff-Kette nicht, sondern kommen damit aus, dass für jedes N eine Kette gemäß Satz 4.5 konstruiert werden kann.

Beispiel 4.6 a) Sei $p_{ij} = q_j$ für alle $i, j \in I$, wobei $\sum_{j\in I} q_j = 1$ ist. Dann gilt

$$P(X_0 = i_0,\ X_1 = i_1,\dots,X_n = i_n) = \nu(i_0)q_{i_1}\dots q_{i_n}.$$

Man sieht leicht, dass $q_j = P(X_m = j)$ für $m \geq 1$ ist. Somit gilt

$$P(X_0 = i_0,\dots,X_n = i_n) = P(X_0 = i_0)P(X_1 = i_1)\dots P(X_n = i_n),$$

d. h., die $X_0, X_1,\dots,X_n$ sind unabhängig. Satz 4.5 liefert also als Spezialfall die Konstruktion von unabhängigen, I-wertigen Zufallsgrößen.

b) *Irrfahrt auf $\mathbb{Z}$:* Es sei $Y_1, Y_2,\dots$ eine Folge unabhängiger, $\{1,-1\}$-wertiger Zufallsgrößen mit $P(Y_j = 1) = p$ und $P(Y_j = -1) = 1-p$, wobei $p \in [0,1]$ ist. Sei $X_0 := 0$ und $X_n := \sum_{j=1}^n Y_j$ für $n \geq 1$. Dann ist $X_0, X_1,\dots$ eine Markoff-Kette auf $\mathbb{Z}$. Die Übergangsmatrix $\mathbb{P} = (p_{ij})_{i,j\in\mathbb{Z}}$ ist durch $p_{i,i+1} = p$ und $p_{i,i-1} = 1-p$ eindeutig festgelegt, und die Startverteilung ist in 0 konzentriert.

c) *Symmetrische Irrfahrt auf $\mathbb{Z}^d$:* Hier ist $I = \mathbb{Z}^d$ und $p_{(i_1,\dots,i_d),(j_1,\dots,j_d)} = 1/(2d)$, falls $i_k = j_k$ für alle bis auf genau ein $k \in \{1,2,\dots,d\}$, für das $|i_k - j_k| = 1$ ist. Alle anderen Übergangswahrscheinlichkeiten müssen dann gleich Null sein.

d) *Ehrenfests Modell der Wärmebewegung:* Es seien n Kugeln auf zwei Schachteln verteilt. Zu einem bestimmten Zeitpunkt seien r Kugeln in der rechten Schachtel und $l := n - r$ in der linken. Mit Wahrscheinlichkeit $1/2$ tun wir nun überhaupt nichts (dass diese auf den ersten Blick unsinnige Annahme begründet ist, werden wir zu einem späteren erkennen). Im anderen Fall wird mit Wahrscheinlichkeit $1/2$ eine der n Kugeln nun zufällig ausgewählt, wobei jede dieselbe Chance hat, und in die andere Schachtel gelegt. Wir können für I die Anzahl der Kugeln in der rechten Schachtel nehmen, also $I = \{0, \dots, n\}$. Die Übergangswahrscheinlichkeiten sind gegeben durch

$$\begin{aligned} p_{r,r-1} &= r/2n, \quad r \in \{1, 2, \dots, n\}, \\ p_{r,r+1} &= 1/2 - r/2n, \quad r \in \{0, 1, \dots, n-1\}. \end{aligned}$$

e) *Polyas Urnenschema:* In einer Urne liegen rote und schwarze Kugeln. Eine wird zufällig gezogen und zusammen mit einer neuen gleicher Farbe zurückgelegt. Hier ist $I = \{\, (r, s) \mid r, s \in \mathbb{N} \,\}$ sowie $p_{(r,s),(r+1,s)} = r/(r + s)$ und $p_{(r,s),(r,s+1)} = s/(r + s)$ für alle $r, s \in \mathbb{N}$.

f) *Galton-Watson-Prozess:* Sei $(q_j)_{j \in \mathbb{N}_0}$ die Verteilung der Anzahl der Nachkommen eines Individuums. I ist gleich $\mathbb{N}_0$, und für jedes $i \in \mathbb{N}$ ist der i-te Zeilenvektor $(p_{ij})_{j \in \mathbb{N}_0}$ der stochastischen Matrix $\mathbb{P}$ gerade die i-fache Faltung der Verteilung $(q_j)_{j \in \mathbb{N}_0}$. Für $i = 0$ gilt $p_{0j} = 1$, falls $j = 0$ ist, und $p_{0j} = 0$, falls $j \geq 1$ ist.

g) *Irrfahrt auf $I = \{0, \dots, n\}$ mit Absorption (random walk with absorbing barriers):* 0 und n seien absorbierend, also $p_{00} = 1$ und $p_{nn} = 1$. Für $i \in \{1, 2, \dots, n-1\}$ geschehe ein Schritt nach rechts mit Wahrscheinlichkeit $p \in (0, 1)$ und ein Schritt nach links mit Wahrscheinlichkeit $q := 1-p$, also $p_{i,i+1} = p$ und $p_{i,i-1} = q$. Die stochastische Matrix hat somit die Form

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 & & \\ q & 0 & p & & \\ & \ddots & \ddots & \ddots & \\ & & q & 0 & p \\ & & 0 & 0 & 1 \end{pmatrix}.$$

h) *Irrfahrt mit Reflexion (reflecting barriers):* Das gleiche Modell wie in Beispiel (e) mit der Änderung, dass $p_{01} = p_{n,n-1} = 1$ sein soll.

i) *Wettervorhersage:* Wenn wir annehmen, dass die Wahrscheinlichkeit für Regen am folgenden Tag nur von Bedingungen von heute abhängt und unbeeinflusst ist vom Wetter der vergangenen Tage, so liefert dies eine ganz einfache Markoff-Kette. Ist α die Wahrscheinlichkeit, dass es morgen regnet, wenn es heute geregnet hat, und β die Wahrscheinlichkeit, dass es morgen regnet, wenn es heute nicht geregnet hat, so hat die stochastische Matrix die Form

$$\mathbb{P} = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Auf Grund der Vielzahl von Beispielen für Markoff-Ketten könnte man vermuten, dass Markoff selbst aus angewandten Fragestellungen heraus die Ketten analysiert hat. Markoff hatte jedoch bei seinen Untersuchungen primär im Sinn, Gesetze der großen Zahlen und zentrale Grenzwertsätze für die Ketten zu studieren. Er hatte nur ein Beispiel vor Augen: er analysierte die möglichen Zustände ,,Konsonant" und ,,Vokal" bei der Buchstabenfolge des Romans ,,Eugen Onegin" von Puschkin. Die Zufallsgröße X_n soll hier den n-ten Buchstaben des Textes angeben.

Eine stochastische Matrix $\mathbb{P} = (p_{ij})_{i,j \in I}$ kann man stets ohne Probleme potenzieren: Für $n \in \mathbb{N}_0$ definiert man die n-te Potenz $\mathbb{P}^n = (p_{ij}^{(n)})_{i,j \in I}$ rekursiv durch $p_{ij}^{(0)} = \delta_{ij}$ und

$$p_{ij}^{(n+1)} = \sum_{k \in I} p_{ik}^{(n)} p_{kj}$$

für alle $i, j \in I$, das heißt, $\mathbb{P}^n$ ist das n-fache Matrixprodukt von $\mathbb{P}$ mit sich selbst. Aus der rekursiven Definition folgt, dass $\mathbb{P}^n$ selbst eine stochastische Matrix ist. Es gelten die aus der linearen Algebra bekannten Rechenregeln für Matrizen, insbesondere gilt $\mathbb{P}^m \mathbb{P}^n = \mathbb{P}^{m+n}$, das heißt

$$\sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)} = p_{ij}^{(m+n)}, \quad i, j \in I.$$

Diese Gleichungen nennt man auch *Chapman-Kolmogoroff-Gleichungen.*

Definition 4.7 *Die Komponenten $p_{ij}^{(n)}$ der Übergangsmatrix $\mathbb{P}^n = (p_{ij}^{(n)})_{i,j \in I}$ heißen n-stufige Übergangswahrscheinlichkeiten (n th order transition probabilities).*

Bemerkung 4.8 *Sei $X_0, X_1, X_2, \ldots$ eine Markoff-Kette mit stochastischer Matrix $\mathbb{P} = (p_{ij})_{i,j \in I}$. Sind $m, n \in \mathbb{N}_0$ und $i, j \in I$ mit $P(X_m = i) > 0$, so gilt*

$$P(X_{m+n} = j \mid X_m = i) = p_{ij}^{(n)}.$$

Beweis. Es gilt

$$\begin{aligned} & P(\, X_{m+n} = j \mid X_m = i\,) \\ = & \sum_{i_{m+1}, \ldots, i_{m+n-1} \in I} P(\, X_{m+1} = i_{m+1}, \ldots, \\ & X_{m+n-1} = i_{m+n-1},\ X_{m+n} = j \mid X_m = i\,) \end{aligned}$$

und mit der Definition 4.2 folgt

$$\begin{aligned} & P(\, X_{m+1} = i_{m+1}, \ldots, X_{m+n-1} = i_{m+n-1}, X_{m+n} = j \mid X_m = i\,) \\ = \; & P(\, X_{m+n} = j \mid X_m = i,\ X_{m+1} = i_{m+1}, \ldots, X_{m+n-1} = i_{m+n-1}\,) \\ & \times \prod_{k=1}^{n-1} P(X_{m+k} = i_{m+k} \mid X_m = i,\, X_{m+1} = i_{m+1}, \ldots, X_{m+k-1} = i_{m+k-1}) \\ = \; & p_{i i_{m+1}} p_{i_{m+1} i_{m+2}} \cdots p_{i_{m+n-1} j}. \end{aligned}$$

Somit gilt

$$P(X_{m+n} = j \mid X_m = i) = \sum_{i_{m+1},\dots,i_{m+n-1} \in I} p_{ii_{m+1}} \dots p_{i_{m+n-1}j} = p_{ij}^{(n)}.$$

$\square$

Lemma 4.9 *Für alle $m, n \in \mathbb{N}_0$ und $i, j, k \in I$ gilt $p_{ij}^{(m+n)} \geq p_{ik}^{(m)} p_{kj}^{(n)}$.*

Beweis. Dies ergibt sich sofort aus den Chapman-Kolmogoroff-Gleichungen. $\square$

Lemma 4.10 *Es sei $X_0, X_1, X_2, \dots$ eine Markoff-Kette mit Startverteilung ν und Übergangsmatrix $\mathbb{P}$. Dann gilt*

$$P_\nu(X_n = j) = \sum_{i \in I} \nu(i) p_{ij}^{(n)}$$

für alle $n \in \mathbb{N}_0$ und $j \in I$. Ist die Startverteilung ν auf $i \in I$ konzentriert, so gilt $P_i(X_n = j) = p_{ij}^{(n)}$.

Beweis. Aus Satz 4.3 a) folgt

$$\begin{aligned} P_\nu(X_n = j) &= \sum_{i_0,\dots,i_{n-1} \in I} P_\nu(X_0 = i_0, \dots, X_{n-1} = i_{n-1},\ X_n = j) \\ &= \sum_{i_0,\dots,i_{n-1} \in I} \nu(i_0) p_{i_0 i_1} \dots p_{i_{n-1}j} = \sum_{i \in I} \nu(i) p_{ij}^{(n)}. \end{aligned}$$

$\square$

Definition 4.11 *Es sei $\mathbb{P} = (p_{ij})_{i,j \in I}$ eine stochastische Matrix. Man sagt, $j \in I$ sei von $i \in I$ aus erreichbar (can be reached from), wenn ein $n \in \mathbb{N}_0$ existiert mit $p_{ij}^{(n)} > 0$. Notation: $i \rightsquigarrow j$.*

Die in Definition 4.11 definierte Relation auf I ist reflexiv und transitiv. Wegen $p_{ii}^{(0)} = 1 > 0$ gilt $i \rightsquigarrow i$ für alle $i \in I$. Falls $i \rightsquigarrow j$ und $j \rightsquigarrow k$ gelten, so gibt es $m, n \in \mathbb{N}_0$ mit $p_{ij}^{(m)} > 0$ und $p_{jk}^{(n)} > 0$, und dann ist $p_{ik}^{(m+n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0$ nach Lemma 4.9.

Die durch $i \sim j \Leftrightarrow (i \rightsquigarrow j$ und $j \rightsquigarrow i)$ für alle $i, j \in I$ definierte Relation ist offenbar eine Äquivalenzrelation auf I. Wir werden $i \sim j$ für den Rest dieses Kapitels stets in diesem Sinne verwenden.

Sind $A, B \subset I$ zwei Äquivalenzklassen der obigen Äquivalenzrelation, so sagen wir, B ist von A aus erreichbar und schreiben $A \rightsquigarrow B$, wenn $i \in A$ und $j \in B$ existieren mit $i \rightsquigarrow j$. Offensichtlich hängt dies nicht von den gewählten Repräsentanten in A und B ab.

Definition 4.12 *Es sei $\mathbb{P}$ eine stochastische Matrix.*

a) Eine Teilmenge I' von I heißt abgeschlossen (closed)*, wenn keine $i \in I'$ und $j \in I \setminus I'$ existieren mit $i \rightsquigarrow j$.*

b) Die Matrix $\mathbb{P}$ und auch eine Markoff-Kette mit Übergangsmatrix $\mathbb{P}$ heißen irreduzibel (irreducible)*, wenn je zwei Elemente aus I äquivalent sind.*

Bemerkung 4.13 Es sei $\mathbb{P} = (p_{ij})_{i,j \in I}$ eine stochastische Matrix.

a) Ist $I' \subset I$ abgeschlossen, so ist die zu I' gehörige Untermatrix $\mathbb{P}' := (p_{ij})_{i,j \in I'}$ eine stochastische Matrix für I'.

b) Ist $\mathbb{P}$ irreduzibel, so existieren keine abgeschlossenen echten Teilmengen von I.

Beispiel 4.14 a) Die symmetrische Irrfahrt auf $\mathbb{Z}^d$ ist irreduzibel.

b) Polyas Urnenschema: Keine zwei Elemente von $I = \{ (r,s) \mid r,s \in \mathbb{N} \}$ sind äquivalent. Es gibt aber sehr viele abgeschlossene Teilmengen von I, zum Beispiel ist für jede Wahl von $r_0, s_0 \in \mathbb{N}$ die Menge $\{ (r,s) \mid r \geq r_0,\ s \geq s_0 \}$ abgeschlossen.

c) Bei der Irrfahrt auf $\{0, \ldots, n\}$ mit absorbierenden Rändern gibt es drei Äquivalenzklassen, nämlich $\{0\}$, $\{1, \ldots, n-1\}$ und $\{n\}$. Die Mengen $\{0\}$ und $\{n\}$ sind abgeschlossen, und es gelten $\{1, \ldots, n-1\} \rightsquigarrow \{n\}$ und $\{1, \ldots, n-1\} \rightsquigarrow \{0\}$.

d) Eine symmetrische Irrfahrt auf einem Graphen $\mathcal{G}$ ist offenbar genau dann irreduzibel, wenn der Graph zusammenhängend ist. (Ein Graph heißt zusammenhängend, wenn je zwei Knoten über einen endlichen Zug verbunden werden können.)

e) Es sei $I = \{0, 1, 2\}$ und die stochastische Matrix gegeben durch

$$\mathbb{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}.$$

Dann ist die Markoff-Kette irreduzibel.

f) Es sei $I = \{0, 1, 2, 3\}$ und die stochastische Matrix gegeben durch

$$\mathbb{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Dann gibt es drei Äquivalenzklassen: $\{0, 1\}, \{2\}$ und $\{3\}$. Der Wert 0 ist von 2 aus erreichbar, aber nicht umgekehrt. Der Wert 3 hat absorbierendes Verhalten; kein anderer Wert ist von 3 aus erreichbar.

Es sei $X_0, X_1, X_2, \dots$ eine Markoff-Kette mit Übergangsmatrix $\mathbb{P} = (p_{ij})_{i,j \in I}$ und Startverteilung ν. Die wichtigste Frage, die uns für einen Großteils des Kapitels beschäftigen wird, ist die Diskussion der Verteilung von X_n für große n, also

$$P_\nu(X_n = j) = \sum_{i \in I} \nu(i) p_{ij}^{(n)}, \quad j \in I.$$

Zu diesem Zwecke werden wir annehmen, dass der Zustandsraum I *endlich* ist. Aus obigen Überlegungen erhält man dann, dass die Frage der asymptotischen Verteilung von X_n äquivalent ist zur Frage, wie sich große Potenzen von stochastischen Matrizen verhalten. Im dem Falle, in dem I nur aus zwei Elementen besteht, kann man sich das noch recht leicht überlegen.

Beispiel 4.15 Sei $|I| = 2$ und

$$\mathbb{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Dann ist für $\alpha = \beta = 0$ $\mathbb{P}^n = Id$ für jedes n (wobei Id bei uns immer die Identität bezeichnet, egal auf welchem Raum sie lebt). Im Falle von $\alpha = \beta = 1$ ist offenbar $\mathbb{P}^n = \mathbb{P}$ für jedes ungerade n und $\mathbb{P}^n = Id$ für alle geraden n.

Im Falle von $0 < \alpha + \beta < 2$ (dem interessanten Fall) diagonalisieren wir $\mathbb{P}$, um seine Potenzen zu berechnen. Es ist

$$\mathbb{P} = RDR^{-1},$$

wobei

$$R = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}$$

und

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{pmatrix}$$

ist. Daher ist

$$\mathbb{P}^n = RD^nR^{-1}.$$

Nun konvergiert aber

$$D^n = \begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} \underset{n\to\infty}{\longrightarrow} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Eingesetzt ergibt das

$$\lim_{n\to\infty} \mathbb{P}^n = R \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} R^{-1} = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix},$$

mit

$$\pi_1 = \frac{\beta}{\alpha+\beta} \qquad \pi_2 = \frac{\alpha}{\alpha+\beta}.$$

Im allgemeinen, d.h. für $|I| > 2$ sind wir leider ziemlich schnell am Ende unserer Weisheit, wenn es um die Berechnung der Eigenwerte von $\mathbb{P}$ und damit um das Diagonalisieren von $\mathbb{P}$ geht. Die obige Methode taugt also nicht, um allgemein Erkenntnisse über das Langzeitverhalten von Markoff-Ketten zu gewinnen. Der Effekt, den wir aber im Beispiel 4.15 gesehen haben, dass nämlich die Limesmatrix aus lauter identischen Zeilen besteht – und das bedeutet, dass die Markoff-Kette asymptotisch ihren Startort “vergißt” – werden wir in dem allgemeinen Limesresultat wiederfinden. Um dieses zu beweisen, müssen wir zunächst den Begriff der Entropie, den wir schon in Kapitel 4 und 6 für zweielementige Grundräume kennengelernt haben, auf größere Räume übertragen.

Definition 4.16 *Es sei I eine endliche, mindestens zweielementige Menge und ν, ϱ seinen Wahrscheinlichkeiten auf I mit $\varrho(i) > 0$ für alle $\imath \in I$. Dann heißt*

$$H(\nu|\varrho) := \sum_{i \in I} \nu(i) \log\left(\frac{\nu(i)}{\varrho(i)}\right)$$

die relative Entropie (relative entropy) *von ν bezüglich ϱ. Hierbei setzen wir $0 \log 0 = 0$.*

Wir sammeln ein paar Eigenschaften der Entropiefunktion

Proposition 4.17 *In der Situation von Definition 4.15 ist $H(\cdot|\varrho)$ positiv und strikt konvex und es ist $H(\nu|\varrho) = 0 \Leftrightarrow \nu = \varrho$.*

Beweis. Sei die nicht–negative, strikt–konvexe Funktion $\psi(t)$ gegeben durch $\psi(t) = t \log t - t + 1$ (und wieder ist $\psi(t) = 0 \Leftrightarrow t = 1$). Dann ist

$$\begin{aligned} H(\nu|\varrho) &= \sum_{i \in I} \varrho(i) \left(\frac{\nu(i)}{\varrho(i)} \log\left(\frac{\nu(i)}{\varrho(i)}\right) - \frac{\nu(i)}{\varrho(i)} + 1 \right) \\ &= \sum_{i \in I} \varrho(i) \psi\left(\frac{\nu(i)}{\varrho(i)}\right), \end{aligned}$$

woraus die Behauptungen folgen. □

Wir kommen nun zu einem Satz, der das aymptotische Verhalten einer großen Gruppe von Markoff-Ketten klärt. Dieser Satz ist gewissermassen ein Gesetz der großen Zahlen für Markoff-Ketten; er wird in der Literatur häufig auch als *Ergodensatz* für Markoff-Ketten bezeichnet.

Satz 4.18 *Ergodensatz (ergodic theorem) Sei* $\mathbb{P}$ *eine stochastische Matrix über einem endlichen Zustandsraum* I *und* ν *irgendeine Anfangsverteilung. Weiter existiere ein* N*, so dass* $\mathbb{P}^N$ *nur strikt positive Einträge hat. Dann konvergiert*

$$\nu\mathbb{P}^n \to_{n\to\infty} \varrho,$$

wobei ϱ *eine Wahrscheinlichkeit auf* I *ist, die der Gleichung*

$$\varrho\mathbb{P} = \varrho$$

genügt.

Bemerkung 4.19 Die Bedingung "es existiere ein N, so dass $\mathbb{P}^N$ nur strikt positive Einträge hat" impliziert natürlich, dass $\mathbb{P}$ irreduzibel ist (man kann nach spätestens N Schritten jeden Punkt von jedem anderen aus erreichen). Umgekehrt ist die Bedingung aber nicht äquivalent zur Irreduzibilität von $\mathbb{P}$. Beispielsweise ist die Matrix

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

irreduzibel, aber natürlich ist keine ihrer Potenzen strikt positiv. Man kann sich überlegen, dass obige Bedingung äquivalent ist zur Irreduzibilität von $\mathbb{P}$ plus einer weiteren Bedingung, die Aperiodizität von $\mathbb{P}$ heisst. Unter letzterem wollen wir verstehen, dass der ggT über sämtliche Zeiten, zu denen man mit positiver Wahrscheinlichkeit in den Punkt i zurückkehren kann, wenn man in i gestartet ist, und über sämtliche Startpunkte i eins ist. Wir werden diese Äquivalenz hier nicht beweisen und nur bemerken, dass irreduzible und aperiodische Markoff-Ketten manchmal auch *ergodisch (ergodic)* heißen.

Satz 4.18 enthält offenbar unter anderem eine unbewiesene Existenzaussage. Diese werden wir getrennt beweisen. Wir zeigen also zunächst, dass es eine Wahrscheinlichkeit ϱ mit

$$\varrho\mathbb{P} = \varrho$$

gibt. Die Existenz eines beliebigen ϱ, das obiger Gleichung genügt, ist ziemlich offensichtlich, denn offenbar ist 1 Eigenwert jeder stochastischen Matrix (die konstanten Funktionen sind rechte Eigenvektoren) – also muss es auch linke Eigenvektoren zum Eigenwert 1 geben; ein solcher ist ϱ. Auch ist es nicht schwierig, ein solches ϱ so zu normieren, dass die Summe seiner Einträge 1 ist. Was aber a priori überhaupt nicht klar ist, ist, warum ein solches ϱ eigentlich nicht-negativ sein sollte. Wer in der linearen Algebra ein wenig Perron-Froebenius Theorie betrieben hat, wird dies schon wissen. Wir werden es hier mit Hilfe eines anderen, mehr stochastischen Arguments herleiten.

Satz 4.20 *Sei* Q *eine stochastische* $r \times r$ *Matrix. Dann existiert*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} Q^j =: H$$

und es gilt

$$HQ = QH = H \qquad H^2 = H.$$

Beweis. Zunächst bemerken wir, dass mit Q auch Q^n stochastisch ist (es ist z.B.

$$\sum_{f=1}^{r} Q^2(e,f) = \sum_{f=1}^{r}\sum_{d=1}^{r} Q(e,d)Q(d,f) = 1;$$

für beliebiges n geht das analog.) Damit ist dann auch

$$P_k := \frac{1}{k}\sum_{j=1}^{k} Q^j$$

stochastisch. Darüber sind die $P_k \in \mathbb{R}^{r^2}$ und als solche beschränkt. Nach dem Satz von Bolzano–Weierstraß besitzt somit die Folge der P_k einen Häufungspunkt H. Wir wollen im folgenden sehen, dass es genau einen Häufungspunkt dieser Folge gibt. Dazu betrachten wir eine Teilfolge (H_l) der Folge (P_k), die gegen H konvergiert. Damit erhalten wir

$$\begin{aligned} QH_l = H_lQ &= \frac{1}{l}\sum_{j=1}^{l} Q^{j+1} \\ &= H_l - \frac{1}{l}Q + \frac{1}{l}Q^{l+1}. \end{aligned}$$

Da die letzten beiden Terme für $l \to \infty$ verschwinden, ergibt sich

$$QH = HQ = H. \tag{44}$$

Ist nun H' ein weiterer Häufungspunkt und (H_m) eine Folge die gegen H' konvergiert, dann erhalten wir aus (44) einerseits

$$H'H = HH' = H.$$

Andererseits folgert man analog zu oben

$$H'P_k = P_kH' = H'$$

für alle k und somit

$$H'H = HH' = H'.$$

Daher ist $H' = H$ und $H^2 = H$. □

Was haben wir nun damit gewonnen? Nun, die Gleichung $HQ = H$ impliziert doch, dass für jede Zeile ϱ von H gilt, dass

$$\varrho Q = \varrho,$$

jede Zeile (und jede konvexe Kombination von Zeilen) von H ist also ein linker Eigenvektor von H zum Eigenwert eins. Darüber hinaus ist die Menge der stochastischen Matrizen abgeschlossen in $\mathbb{R}^{r^2}$. Das sieht man, indem man einerseits die Abgeschlossenheit aller nicht-negativen Matrizen erkennt (das ist nicht schwer) und andererseits sieht, dass die Menge aller Matrizen mit Zeilensumme eins für alle Zeilen abgeschlossen ist (die Menge der stochastischen Matrizen ist dann der Durchschnitt dieser beiden abgeschlossenen Mengen). Letzteres ist wahr, denn die Funktionen f_i, die die i'te Zeilensumme bilden sind stetig, und die Menge der Matrizen mit Zeilensumme 1 ist dann das Urbild der (abgeschlossenen) Menge $(1, \dots, 1)$ unter der stetigen Abbildung $f = (f_1, \dots, f_r)$.

Somit ist H als Limes stochastischer Matrizen wieder stochastisch, seine Zeilen sind also Wahrscheinlichkeiten auf dem Grundraum. Dies beweist die Existenz einer Wahrscheinlichkeit ϱ mit

$$\varrho Q = \varrho.$$

Solche Wahrscheinlichkeiten heißen auch *stationär (stationary)* bzgl. Q. Nun sind wir in der Lage Satz 4.17 zu beweisen.

Beweis von Satz 4.17 Wie wir eben gesehen haben, existiert eine stationäre Verteilung ϱ bzgl. $\mathbb{P}$, nämlich beispielsweise eine Zeile des entsprechend Satz 4.19 gebildeten Cesaro-Limes der Potenzen von $\mathbb{P}$. Ein solches ϱ besitzt nur strikt positive Einträge. Wäre z.B. $\varrho(i) = 0$, so ergäbe das

$$0 = \varrho(i) = \sum_{j \in I} \varrho(j) \mathbb{P}^N(j, i)$$

im Widerspruch dazu, dass $\mathbb{P}^N$ strikt positiv ist und $\sum \varrho(j) = 1$ ist.

Darüber hinaus gibt es nur eine Verteilung ϱ, die stationär zu $\mathbb{P}$ ist (insbesondere besteht H aus lauter identischen Zeilen). Gäbe es nämlich ϱ, ϱ', die beide stationär bzgl. $\mathbb{P}$ wären, so gälte für jedes $a \in \mathbb{R}$ und $n \in \mathbb{N}$

$$\varrho - a\varrho' = (\varrho - a\varrho')\mathbb{P}^n.$$

Wir wählen

$$a = \min_{i \in I} \frac{\varrho(i)}{\varrho'(i)} =: \frac{\varrho(i_0)}{\varrho'(i_0)}.$$

Damit ist

$$0 = (\varrho - a\varrho')(i_0) = \sum_{j \in I} (\varrho - a\varrho')(j) \mathbb{P}^N(j, i_0).$$

Aus der strikten Positivität von $\mathbb{P}^N$ folgt somit, dass $\varrho(j) = a\varrho'(j)$ für alle $j \in I$ gelten muss. Da ϱ und ϱ' Wahrscheinlichkeiten sind, impliziert das, dass $a = 1$ ist und folglich $\varrho = \varrho'$. Die im Satz behauptete Konvergenz ist also die Konvergenz gegen *einen* Punkt im klassichen Sinne.

Um diese Konvergenz schließlich zu zeigen, verwenden wir die Entropiefunktion aus Definition 4.15 in der Schreibweise

$$H(\nu|\varrho) = \sum_{i\in I} \varrho(i)\psi\left(\frac{\nu(i)}{\varrho(i)}\right),$$

wobei ψ wieder die strikt konvexe Funktion

$$\psi(t) = t\log t - t + 1$$

ist. Daher ist

$$\begin{aligned}
H(\nu\mathbb{P}|\varrho) &= \sum_{i\in I} \varrho(i)\psi\left(\frac{\nu\mathbb{P}(i)}{\varrho(i)}\right) \\
&= \sum_{i\in I} \varrho(i)\psi\left(\frac{\sum_{j\in I}\nu(j)\mathbb{P}(j,i)}{\varrho(i)}\right) \\
&= \sum_{i\in I} \varrho(i)\psi\left(\frac{\sum_{j\in I}\varrho(j)\mathbb{P}(j,i)}{\varrho(i)}\frac{\nu(j)}{\varrho(j)}\right) \\
&\leq \sum_{i\in I}\sum_{j\in I} \varrho(j)\mathbb{P}(j,i)\psi\left(\frac{\nu(j)}{\varrho(j)}\right) \\
&= \sum_{j\in I} \varrho(j)\psi\left(\frac{\nu(j)}{\varrho(j)}\right) \\
&= H(\nu|\varrho),
\end{aligned}$$

wobei das "$\leq$"-Zeichen aus der Tatsache, dass $\frac{\sum_{j\in I}\varrho(j)\mathbb{P}(j,i)}{\varrho(i)}\frac{\nu(j)}{\varrho(j)}$ eine konvexe Kombination der $\frac{\nu(j)}{\varrho(j)}$ ist, folgt, zusammen mit der Konvexität von ψ und das vorletzte Gleichheitszeichen eine Konsequenz der Stochastizität von $\mathbb{P}$ ist. Somit ist

$$H(\nu\mathbb{P}|\varrho) \leq H(\nu|\varrho)$$

mit Gleichheit genau dann, wenn $\nu\mathbb{P} = \nu$, also $\nu = \varrho$ ist. Anwenden von $\mathbb{P}$ verkleinert also die Entropie und damit eine Art Distanz zum invarianten Maß.

Somit ist insbesondere die Folge $\big(H(\nu\mathbb{P}^n|\varrho)\big)_n$ monoton fallend und zwar strikt, solange $\nu\mathbb{P}^n \neq \varrho$ ist.

Wir wollen abschließend sehen, dass dies schon impliziert, dass die Folge $\varrho_n := \nu\mathbb{P}^n$ gegen ϱ konvergiert. Da ϱ_n beschränkt ist, besitzt die Folge zumindest im $\mathbb{R}^{|I|}$ einen Häufungspunkt ϱ' und es existiert eine Teilfolge $(\varrho_{n_l})_l$, die gegen ϱ' konvergiert. Wir zeigen, dass $\varrho' = \varrho$ ist (und sind dann fertig, da die Argumentation für jeden Häufungspunkt gilt und die Folge ϱ_n damit gegen ϱ konvergiert).

Nun ist einerseits

$$H(\varrho'|\varrho) \geq H(\varrho'\mathbb{P}|\varrho).$$

Andererseits haben wir

$$
\begin{aligned}
H(\varrho'\mathbb{P}|\varrho) &= \sum_{j\in I} \varrho(j)\psi\left(\frac{(\varrho'\mathbb{P})(j)}{\varrho(j)}\right) \\
&= \lim_{l\to\infty} \sum_{j\in I} \varrho(j)\psi\left(\frac{(\nu\mathbb{P}^{n_l})\mathbb{P}(j)}{\varrho(j)}\right) \\
&= \lim_{l\to\infty} \sum_{j\in I} \varrho(j)\psi\left(\frac{(\nu\mathbb{P}^{n_l+1})(j)}{\varrho(j)}\right).
\end{aligned}
$$

Nun ist $(n_l)_l$ eine Teilfolge und daher $n_l + 1 \leq n_{l+1}$. Dies ergibt mit der vorher gezeigten Monotonie

$$
\begin{aligned}
& \lim_{l\to\infty} \sum_{j\in I} \varrho(j)\psi\left(\frac{(\nu\mathbb{P}^{n_l+1})(j)}{\varrho(j)}\right) \\
\geq\ & \lim_{l\to\infty} \sum_{j\in I} \varrho(j)\psi\left(\frac{(\nu\mathbb{P}^{n_{l+1}})(j)}{\varrho(j)}\right) = H(\varrho'|\varrho).
\end{aligned}
$$

Insgesamt ist also

$$H(\varrho'|\varrho) = H(\varrho'\mathbb{P}|\varrho)$$

und daher

$$\varrho' = \varrho.$$

□

Beispiel 4.21 1. *Irrfahrt auf dem Kreis*

Für $n \in \mathbb{N}$ sei C_n der n-Kreis, d.h. der Graph, der entsteht, wenn man n Punkte durchnummeriert und den Punkt k mit den Punkten $k-1$ und $k+1$ verbindet (Punkt 1 wird mit 2 und n verbunden). Auf C_n definiert man eine Markoff-Kette vermöge der Übergangsvorschrift $p_{ii} = 1/2$ und $p_{i,i+1} = p_{i,i-1} = 1/4$ (dabei ist die Addition modulo n zu verstehen). Offenbar ist für die zugehörige stochastische Matrix $\mathbb{P}$ und jedes $r > n/2+1$, $\mathbb{P}^r$ strikt positiv. Also sind die Voraussetzungen des Ergodensatzes erfüllt und für jede beliebige Startverteilung ν konvergiert $\nu\mathbb{P}^n$ gegen das invariante Maß der Kette, was offensichtlich die Gleichverteilung auf allen Zuständen ist.

2. *Ehrenfests Urnenmodell*

In der Situation von Beispiel 4.6 d) rechnet man wieder nach, dass die Bedingungen des Ergodensatzes erfüllt sind. Die Kette konvergiert daher gegen ihre Gleichgewichtsverteilung, d.h. die Binomialverteilung.

Wir werden uns im folgenden auf eine besondere Markoff-Kette konzentrieren. Dazu bemerken wir zunächst, dass – hat man eine Folge (X_i) von unabhängigen, identisch verteilten Zufallsvariablen mit endlich vielen Werten gegeben (dass es so eine Folge gibt, können wir allerdings hier nicht zeigen) – man daraus eine Markoffkette S_n bilden kann, indem man

$$S_n = \sum_{i=1}^{n} X_i$$

und $S_0 = 0$ setzt. In der Tat rechnet man schnell nach, dass für jedes Ereignis $\{S_{n-1} = a_{n-1}, \ldots, S_1 = a_1, S_0 = a_0\}$ mit $P(\{S_{n-1} = a_{n-1}, \ldots, S_1 = a_1, S_0 = a_0\}) > 0$ gilt

$$P(S_n = a_n | S_{n-1} = a_{n-1}, \ldots, S_1 = a_1, S_0 = a_0) = P(X_n = a_n - a_{n-1}),$$

also die Markoff-Eigenschaft erfüllt ist. Wir werden im folgenden genau eine solche Markoff-Kette betrachten, wobei die X_i unabhängige Zufallsvariablen mit Werten in $\{-1, 1\}$ und $P(X_i = 1) = P(X_i = -1) = 1/2$ sind. Anschaulich entpricht das einer Art Pfad, der in der 0 startet und in jedem Punkt $n \in \mathbb{N}$ entscheidet, ob er einen Schritt nach oben oder einen Schritt nach unten geht. Die Menge aller solcher Pfade der Länge n sei Ω_n. Aus naheliegenden Gründen bezeichnet man die Folge $S_0 = 0, S_1, \ldots, S_n$ auch als *Irrfahrt (random walk)* auf $\mathbb{Z}$. Den Index dieser Zufallsgrößen bezeichnet man meist als die ,,Zeit“. Wir sagen also etwa ,,die Wahrscheinlichkeit, dass zum Zeitpunkt 100 die Irrfahrt erstmals in 20 ist, ist...“ und meinen damit die Wahrscheinlichkeit des Ereignisses

$$A = \{S_1 \neq 20, S_2 \neq 20, \ldots, S_{99} \neq 20, S_{100} = 20\}.$$

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. Eine genauere theoretische Diskussion des richtigen Streckungsmaßstabs kann hier nicht gegeben werden, dies geschieht in Kapitel 7.

Hier sollen zunächst "Pfadeigenschaften" der S_n studiert werden. Hierzu wollen wir $(S_n)_n$ nicht nur in einer Dimension betrachten, sondern in d Dimensionen. Es sei also $(S_n)_n$ die d-dimensionale Irrfahrt ohne Drift, die wir schon in den Beispielen kennengelernt haben, also

$$S_n = \sum_{i=1} X_i$$

und X_i nehmen iid. und gleichverteilt die Werte $\pm e_i,\ i = 1, \ldots, d$ mit Wahrscheinlichkeit $\frac{1}{2d}$ an, wobei die e_i die Einheitsvektoren in $\mathbb{Z}^d$ sind.

Wir fragen uns zunächst, ob eine Irrfahrt (S_n), die definitionsgemäß im Urpsrung 0 beginnt, wieder nach 0 zurückkehrt. Dazu sei T der Zeitpunkt der ersten Rückkehr (T ist die Zufallsvariable). Der folgende Satz geht auf Georg Polya zurück. Er zeigt einen "Phasenübergang" des Verhaltens in der Dimension.

Satz 4.22 *Es gilt*

a) $\mathbb{P}(T < \infty) = 1$, *falls* $d \leq 2$,

b) $\mathbb{P}(T < \infty) < 1$, *falls* $d \geq 3$.

Bemerkung 4.23 *a) nennt man Rekurrenz, b) heißt Transienz der Irrfahrt.*

Beweis: Sei $N := \sum_n \mathbb{1}_{\{S_n=0\}} = \sum_n \mathbb{1}_{\{S_{2n}=0\}}$ die Anzahl der Besuche der Irrfahrt im Ursprung und

$$L := \sup\{2n : S_{2n} = 0\}$$

der letzte Besuch dort (wobei möglicherweise $L = \infty$) ist. Es ist

$$EN = \sum_{n=0}^{\infty} \mathbb{P}(S_{2n} = 0).$$

Die Translationsinvarianz unter Zeitshifts ergibt:

$$\begin{aligned}\mathbb{P}(L = 2n) &= \mathbb{P}(S_{2n} = 0)\mathbb{P}(S_{2n+2j} \neq 0 \quad \forall\, j \geq 0 | S_{2n} = 0)\\ &= \mathbb{P}(S_{2n} = 0) \cdot \mathbb{P}(S_{2j} \neq 0 \quad \forall j)\\ &= \mathbb{P}(S_{2n} = 0) \cdot \mathbb{P}(T = \infty).\end{aligned}$$

Summation über n ergibt

$$\mathbb{P}(L < \infty) = \mathbb{P}(N < \infty) = \mathbb{E}[N] \cdot \mathbb{P}(T = \infty).$$

Ist $\mathbb{E}N = \infty$, folgt $\mathbb{P}(N < \infty) = 0$ und somit $\mathbb{P}(T = \infty) = 0$. Die Irrfahrt ist also rekurrent. Ist $0 < \mathbb{E}N < \infty$, so ist $\mathbb{P}(N < \infty) = 1$, also

$$\mathbb{P}(T = \infty) = \frac{1}{EN} > 0,$$

also ist die Irrfahrt transient. Man rechnet dann

$$\begin{aligned}\mathbb{P}(T < \infty) = 1 - \mathbb{P}(T = \infty) = 1 - \frac{1}{EN} &= \frac{\mathbb{E}N - 1}{\mathbb{E}N}\\ &= \frac{\sum_{n=1}^{\infty} \mathbb{P}(S_{2n} = 0)}{\sum_{n=0}^{\infty} \mathbb{P}(S_{2n} = 0)}.\end{aligned}$$

Konvergiert also $\sum_{n=1}^{\infty} \mathbb{P}(S_{2n} = 0)$, so ist die Irrfahrt transient; definiert man "$\frac{\infty}{1+\infty} = 1$", so ergibt die Formel auch für divergente Reihen $\sum \mathbb{P}(S_{2n} = 0)$ Sinn, dann ist die Irrfahrt nämlich rekurrent.

Ist nun $d = 1$, so ist

$$\mathbb{P}(S_{2n} = 0) = \binom{2n}{n} 2^{-2n} \sim \frac{1}{\sqrt{\pi n}}$$

nach der Stirlingschen Formel. Da

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty$$

folgt die Behauptung für $d = 1$.

Für $d = 2$ muss die Irrfahrt zur Zeit $2n$ je k Schritte nach oben und unten gegangen sein und jeweils $n-k$ Schritte nach links und rechts, um zur 0 zurückzukehren. Also

$$\begin{aligned}
\mathbb{P}(S_{2n} = 0) &= \frac{1}{4^{2n}} \sum_{k=0}^{n} \binom{2n}{k} \binom{2n-k}{k} \binom{2n-2k}{n-k} \\
&= \frac{1}{4^{2n}} \binom{2n}{n} \sum_{k=0}^{n} \binom{n}{n-k} \\
&= \frac{1}{4^{2n}} \binom{2n}{n}^2 = \left(\frac{1}{2^{2n}} \binom{2n}{n} \right)^2 \\
&\sim \frac{1}{\pi n} \quad \text{für} \quad n \to \infty.
\end{aligned}$$

Da auch die harmonische Reihe divergiert, folgt die Behauptung für $d = 2$.

Für $d \geq 3$ impliziert $S_{2n} = 0$, dass man in den ersten $2n$ Schritten jeweils k_i Schritte in Richtung von $\pm e_i$ gemacht haben muss. Sei

$$C_n = \{0 \leq k_i \leq n : \sum_{i=1}^{d} k_i = n\}.$$

Also gilt

$$\begin{aligned}
\mathbb{P}(S_{2n} = 0) &= \frac{1}{(2d)^{2n}} \sum_{k=(k_1,\ldots,k_d) \in C_n} \frac{2n!}{(k_1!)^2 \ldots (k_d!)^2} \\
&= \frac{1}{2^{2n}} \binom{2n}{n} \sum_{k \in C_n} \left[d^{-n} \frac{n!}{k_1! \ldots k_d!} \right]^2 \\
&\leq \frac{1}{2^{2n}} \binom{2n}{n} \max_{k \in C_n} \left\{ d^{-n} \frac{n!}{k_1! \ldots k_d!} \right\} \times \sum_{k \in C_n} d^{-n} \frac{n!}{k_1! \ldots k_d!}.
\end{aligned}$$

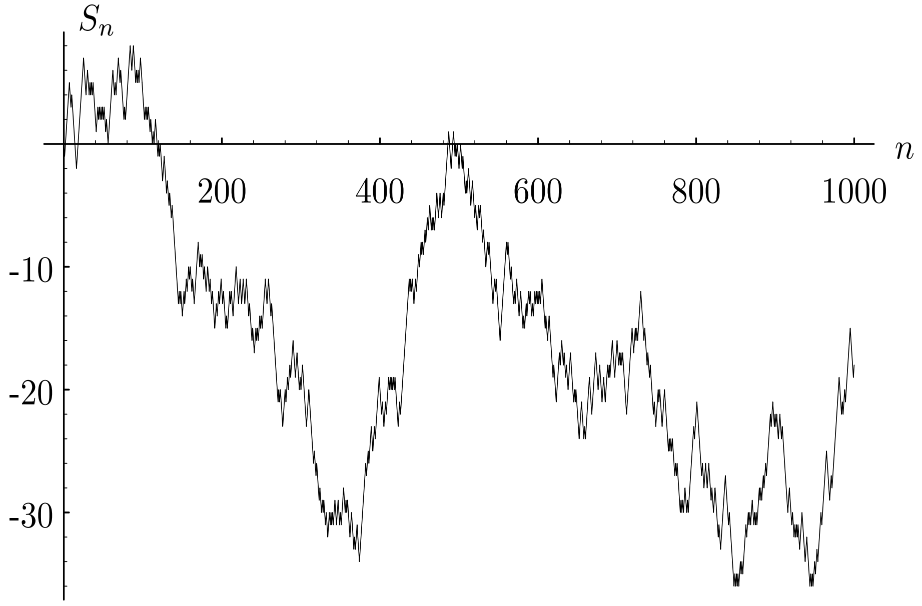
Die letzte Summe ist als Summe über die Wahrscheinlichkeiten einer Multinomialverteilung 1. Das Maximum wird bei $|k_j - \frac{n}{d}| \leq 1$ angenommen. Die Stirlingformel liefert daher

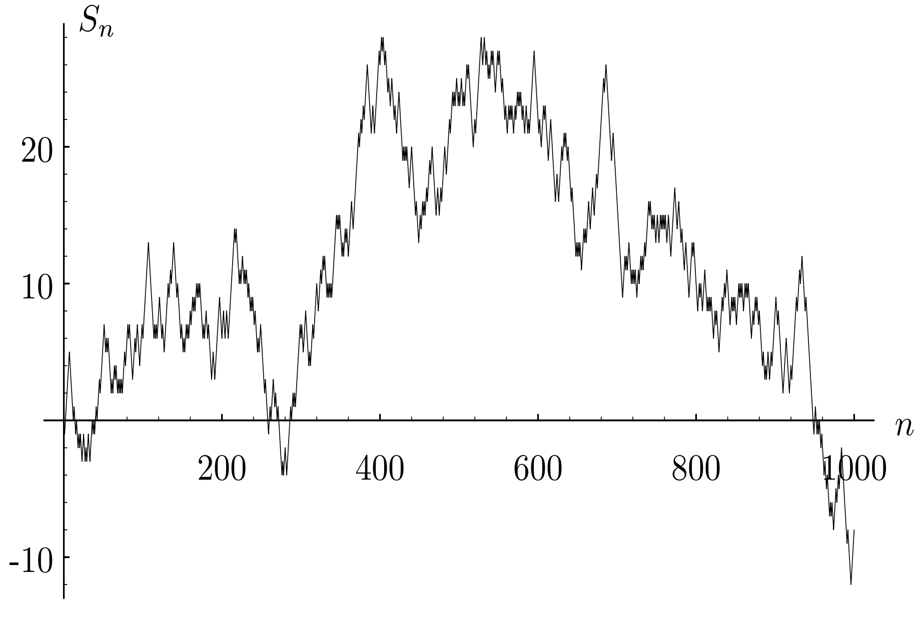
$$\begin{aligned}
\mathbb{P}(S_{2n} = 0) &\leq \frac{1}{\sqrt{\pi n}} d^{-n} \frac{(\frac{n}{e})^n \sqrt{2\pi n}}{(\frac{n}{d})^{n/2d} \left(\sqrt{2\pi \frac{n}{d}}\right)^d} \\
&= \text{Const.}\ n^{-d/2}.
\end{aligned}$$

Dies ist für $d \geq 3$ summierbar. □

Wir werden uns nun mit dem Verhalten von S_n in $d = 1$ befassen.

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}, $2k \leq n$. Offensichtlich


S_n
n
200
400
600
800
1000
-10
-20
-30


S_n
n
200
400
600
800
1000
20
10
-10

gilt

$$P(A_{2k}) = \binom{2k}{k} 2^{-2k} = b(k; 2k, 1/2).$$

Wir kürzen diese Größe auch mit u_{2k} ab ($u_0 = 1$). Wir bemerken zunächst, dass $P(A_{2k})$ nicht von n, der Gesamtlänge des Experiments, abhängt, sofern nur $n \geq 2k$ gilt. Dies ist nicht weiter erstaunlich, denn die X_i sind ja unabhängig.

Wir werden diesem Phänomen noch mehrmals begegnen und wollen es deshalb genau ausformulieren: Sei $k < n$ und A ein Ereignis in Ω_k. Wir können ihm das Ereignis

$$\bar{A} = \{ \omega = (s_0, \dots, s_n) \in \Omega_n : (s_0, \dots, s_k) \in A \}$$

in Ω_n zuordnen. Dann gilt

$$P^{(k)}(A) = P^{(n)}(\bar{A}),$$

wobei $P^{(n)}$ die durch die Gleichverteilung auf den Teilmengen von Ω_n definierte Wahrscheinlichkeit ist. Der Leser möge dies selbst verifizieren. Für ein derartiges Ereignis ist es deshalb gleichgültig, in welchem Pfadraum Ω_n die Wahrscheinlichkeit berechnet wird, sofern nur $n \geq k$ ist. Wir werden im weiteren stillschweigend auch endlich viele Ereignisse miteinander kombinieren (z.B$\dot{\text{D}}$urchschnitte bilden), die zunächst für Pfade unterschiedlicher Länge definiert sind. Dies bedeutet einfach, dass diese Ereignisse im obigen Sinne als Ereignisse in einem gemeinsamen Raum Ω_n interpretiert werden, wobei nur n genügend groß gewählt werden muss.

Um die Größenordnung von $u_{2k} = P(A_{2k})$ für große k zu bestimmen, erinnern wir uns an den lokalen Grenzwertsatz. Dieser liefert sofort:

Satz 4.24

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}},$$

d.h.

$$\lim_{k \to \infty} u_{2k} \sqrt{\pi k} = 1.$$

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

$$f_{2k} = P(S_1 \neq 0, \, S_2 \neq 0, \dots, S_{2k-1} \neq 0, \, S_{2k} = 0).$$

Dann gilt

Satz 4.25 *1.* $f_{2k} = \frac{1}{2k} u_{2k-2} = P(S_1 \geq 0, \, S_2 \geq 0, \dots, S_{2k-2} \geq 0, \, S_{2k-1} < 0)$
$= u_{2k-2} - u_{2k}$.

2. $u_{2k} = P(S_1 \neq 0,\, S_2 \neq 0, \ldots, S_{2k} \neq 0) = P(S_1 \geq 0,\, S_2 \geq 0, \ldots, S_{2k} \geq 0)$.

3. $u_{2k} = \sum_{j=1}^{k} f_{2j} u_{2k-2j}$.

Zum Beweis dieses Satzes müssen wir ein wenig ausholen. Insbesondere stellen wir einen eleganten Trick vor, mit dem sich die Mächtigkeit gewisser Pfadmengen bestimmen lässt. Dieser beruht auf einer teilweisen Spiegelung der Pfade an der x-Achse.

Wir sagen, dass ein Pfad $(s_i, s_{i+1}, \ldots, s_j)$ *die x-Achse berührt*, falls ein k mit $i \leq k \leq j$ existiert, für das $s_k = 0$ ist.

Lemma 4.26 *(Reflektionsprinzip, reflection principle) 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).*

Beweis. Wir geben eine bijektive Abbildung an, die die Menge der Pfade von $(i, -a)$ nach (j, b) auf die Menge der Pfade von (i, a) nach (j, b), welche die x-Achse berühren, abbildet. Sei

$$(s_i = -a, s_{i+1}, \ldots, s_{j-1},\, s_j = b)$$

ein Pfad von $(i, -a)$ nach (j, b). Dieser Pfad muss notwendigerweise die x-Achse berühren. Sei τ die kleinste Zahl $> i$, für welche $s_\tau = 0$ gilt. Offensichtlich ist dann

$$(-s_i, -s_{i+1}, \ldots, -s_{\tau-1},\, s_\tau = 0,\, s_{\tau+1}, \ldots, s_j = b)$$

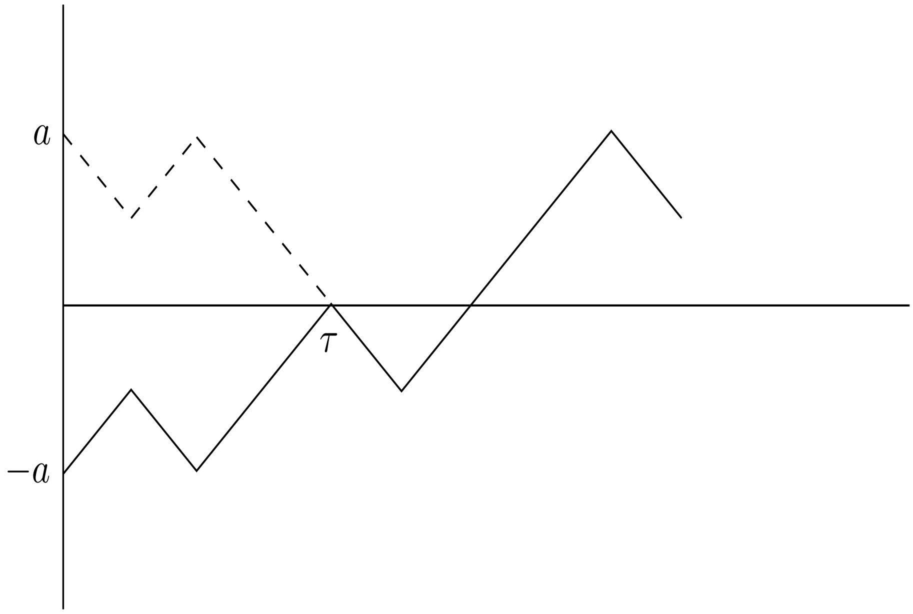
ein Pfad von (i, a) nach (j, b), der die x-Achse berührt, und die Zuordnung ist bijektiv. □

Das Spiegelungsprinzip werden wir nun verwenden, um die Menge der Pfade, die nach $2k$ Schritten zum ersten Mal wieder die x-Achse berühren abzuzählen.

Satz 4.27 *1. Es gibt $\frac{1}{p}\binom{2p-2}{p-1}$ Pfade von $(0,0)$ nach $(2p, 0)$ mit $s_1 > 0,\, s_2 > 0, \ldots, s_{2p-1} > 0$.*

2. 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. (1) Es ist notwendigerweise $s_1 = 1$ und $s_{2p-1} = 1$. Wir suchen somit nach der Anzahl der Pfade von $(1, 1)$ nach $(2p-1, 1)$ mit $s_1 > 0,\, s_2 > 0, \ldots, s_{2p-1} = 1$. Diese ist gleich der Anzahl aller Pfade von $(1, 1)$ nach $(2p-1, 1)$ minus der Anzahl der Pfade, die die x-Achse berühren. Dies ist nach dem Spiegelungsprinzip gleich



der Anzahl aller Pfade von $(1,1)$ nach $(2p-1,1)$ minus der Anzahl der Pfade von $(-1,1)$ nach $(2p-1,1)$. Nach ein bisschen elementarer Kombinatorik erhält man daher

$$\binom{2p-2}{p-1} - \binom{2p-2}{p} = \frac{1}{2p-1}\binom{2p-1}{p} = \frac{1}{p}\binom{2p-2}{p-1}$$

als die gesuchte Anzahl der Pfade.

(2) 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.

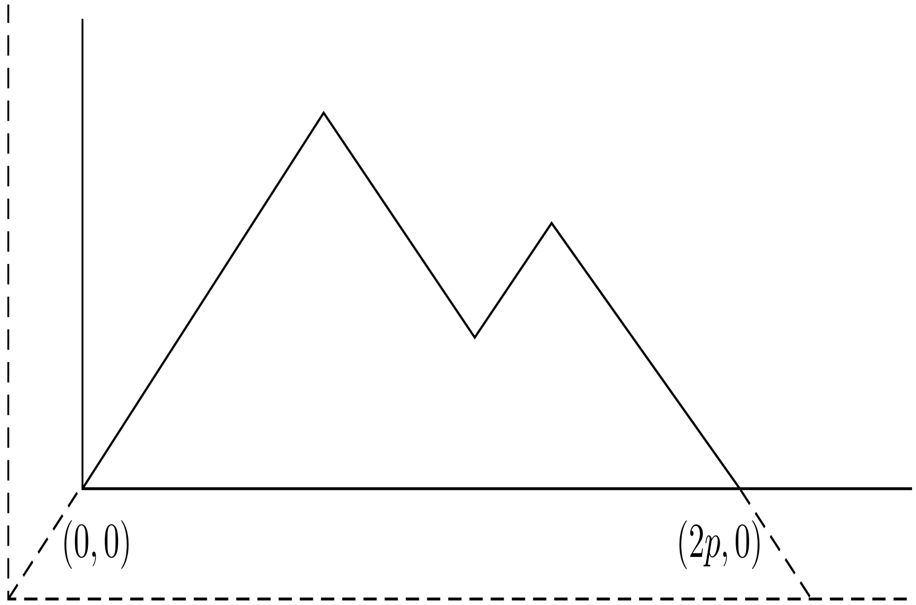
Auf diese Weise wird eine bijektive Abbildung von der gesuchten Menge von Pfaden auf die Menge der Pfade von $(-1,-1)$ nach $(2p+1,-1)$, welche die Bedingung $s_0 > -1$, $s_1 > -1, \ldots, s_{2p} > -1$ erfüllen, hergestellt. Die Anzahl der Pfade in dieser Menge ist gleich der Anzahl der Pfade von $(0,0)$ nach $(2p+2,0)$ mit $s_1 > 0$, $s_2 > 0, \ldots, s_{2p+1} > 0$ (Verschiebung des Ursprungs). (2) folgt dann aus (1). $\square$

Nun sind wir in der Lage Satz 4.25 zu beweisen:

Beweis von Satz 4.25. (1) Nach Satz 9.23 (1) 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 ≥ 0. Die zweite Gleichung in (1) folgt dann mit Hilfe von Satz 4.27 (2). Die dritte ergibt sich aus

$$u_{2k} = \binom{2k}{k} 2^{-2k} = \frac{2k(2k-1)}{k \cdot k} \binom{2k-2}{k-1} \cdot \frac{1}{4} \cdot 2^{-2k+2} = \left(1 - \frac{1}{2k}\right) u_{2k-2}.$$

(2) 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 $u_0 = 1$

$$P(S_1 \neq 0,\ S_2 \neq 0, \ldots, S_{2k} \neq 0) = 1 - P\left(\bigcup_{j=1}^{k} C_{2j}\right) = 1 - \sum_{j=1}^{k} (u_{2j-2} - u_{2j}) = u_{2k}.$$

Die zweite Gleichung folgt analog aus der dritten Identität in (1).

(3) 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

$$u_{2k} = \sum_{j=1}^{k} P(B_j) = \sum_{j=1}^{k} f_{2j} u_{2k-2j}.$$

□

Eine interessante Folgerung ergibt sich aus der ersten Gleichung in (2). Da $\lim_{k\to\infty} u_{2k} = 0$ 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 (1) und $u_{2k} \to 0$ folgt

$$\begin{aligned}\sum_{k=1}^{\infty} f_{2k} &= \lim_{N\to\infty} \sum_{k=1}^{N} f_{2k} \\ &= \lim_{N\to\infty} \sum_{k=1}^{N} (u_{2k-2} - u_{2k}) \\ &= \lim_{N\to\infty} (u_0 - u_{2N}) = 1.\end{aligned}$$

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:

$$ET = \sum_{k=1}^{\infty} 2k f_{2k} = \sum_{k=1}^{\infty} u_{2k-2},$$

wobei wir die Gleichung wieder Satz 4.25 (1) anwenden. Diese Reihe divergiert jedoch! Man kann auch sagen, dass ET gleich ∞ 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 gerade 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 Satz 4.25 (2) 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.

Satz 4.28 *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 haben 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

$$G_{l,m}^+ \subset F_{l+m}^{(n)}, \; G_{l,m}^- \subset F_m^{(n)}.$$

Man beachte, dass für $1 \leq j \leq n-1$ jeder Pfad aus $F_j^{(n)}$ 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

$$P(F_j^{(n)}) = \sum_{l=1}^{j} P(G_{l,j-l}^+) + \sum_{l=1}^{n-j} P(G_{l,j}^-).$$

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 $s_1 > 0$, $s_1 > 0, \ldots, s_{2l-1} > 0$ multipliziert mit der Anzahl der Pfade der Länge $2n-2l$ mit Start in $(2l,0)$ und $2m$ positiven Strecken, das heißt

$$\begin{aligned} |G_{l,m}^+| &= |G_{l,m}^-| = \frac{1}{2} f_{2l} 2^{2l} P(F_m^{(n-l)}) 2^{2n-2l}, \\ P(G_{l,m}^+) &= P(G_{l,m}^-) = \frac{1}{2} f_{2l} P(F_m^{(n-l)}). \end{aligned}$$

Nach der weiter oben stehenden Gleichung ist also

$$P(F_j^{(n)}) = \frac{1}{2}\sum_{l=1}^{j} f_{2l} P(F_{j-l}^{(n-l)}) + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l} P(F_j^{(n-l)}).$$

Nach Induktionsvoraussetzung ist das

$$= \frac{1}{2}\sum_{l=1}^{j} f_{2l}\, u_{2j-2l}\, u_{2n-2j} + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l}\, u_{2n-2j-2l}\, u_{2j} = u_{2j}\, u_{2n-2j}.$$

□

Um das Verhalten von $P(F_j^{(k)})$ für festes k als Funktion von j zu untersuchen, betrachten wir für $1 \le j \le k-1$ die Quotienten

$$\begin{aligned} \frac{P(F_j^{(k)})}{P(F_{j+1}^{(k)})} &= \frac{\binom{2j}{j}\binom{2k-2j}{k-j}}{\binom{2j+2}{j+1}\binom{2k-2j-2}{k-j-1}} = \frac{(2j)!(2k-2j)!((j+1)!)^2((k-j-1)!)^2}{(j!)^2((k-j)!)^2(2j+2)!(2k-2j-2)!} \\ &= \frac{(2k-2j-1)(j+1)}{(2j+1)(k-j)}. \end{aligned}$$

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 genauer diskutieren zu können, sei aber daran erinnert, dass die Rückkehrzeit T nach 0 keinen endlichen Erwartungswert hat, wie wir oben gezeigt haben.

Mit Hilfe des Vorangegangenen lässt sich nun eine einfach Approximation für $P(F_j^{(k)})$ für große j und $k-j$ gewinnen:

Satz 4.29 *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*

$$\lim_{\substack{j\to\infty \\ k-j\to\infty}} \sqrt{j(k-j)}\,P(F_j^{(k)}) = \frac{1}{\pi}.$$

□

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 Satz 4.29 ist gewissermaßen auch ein lokaler 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 α und β liegt. Genauer: T_k sei (die auf Ω_{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 wollen nun aus Satz 4.29 für $k \to \infty$ folgern:

$$\gamma_k(\alpha, \beta) \sim \frac{1}{\pi} \sum_{j: \alpha \leq \frac{j}{k} \leq \beta} \frac{1}{k} \frac{1}{\sqrt{\frac{j}{k}(1 - \frac{j}{k})}}. \tag{45}$$

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 damit:

Satz 4.30 *(Arcus-Sinus-Gesetz)*

$$\lim_{k \to \infty} \gamma_k(\alpha, \beta) = \frac{2}{\pi}(\arcsin \sqrt{\beta} - \arcsin \sqrt{\alpha}).$$

Beweis. 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

$$P(F_j^{(k)}) = \binom{2j}{j}\binom{2k-2j}{k-j}\frac{1}{2^{2k}} = \frac{1}{\pi} \frac{1}{\sqrt{(\frac{j}{k})(1 - (\frac{j}{k}))}} \frac{1}{k} \frac{F(2j)\, F(2(k-j))}{F(j)\, F(j)\, F(k-j)\, F(k-j)}.$$

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 $F(j), F(k-j), F(2j)$ 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

$$\sum_{\alpha \leq \frac{j}{k} \leq \beta} P(F_j^{(k)}) = \left(\frac{1}{k} \sum_{\alpha \leq \frac{j}{k} \leq \beta} \frac{1}{\pi} \frac{1}{\sqrt{(j/k)(1-(j/k))}} \right) G_{\alpha,\beta}(k).$$

Nun folgt die Behauptung gleichmäßig für $\delta \leq \alpha < \beta \leq 1 - \delta$, wie auch immer $0 < \delta < 1/2$ gewählt war. Damit folgt die Behauptung. □

Bemerkung 4.31 *Die Aussage von Satz 4.30 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, $\leq \beta$ ist — gegen $\frac{2}{\pi} \arcsin \sqrt{\beta}$ konvergiert.*

Beweis Offensichtlich gilt $\lim_{k\to\infty} \gamma_k(0, \frac{1}{2}) = 1/2$. Ist $\beta \in (0, 1/2)$, so folgt

$$\lim_{k\to\infty} \gamma_k(0, \beta) = \lim_{k\to\infty} (\gamma_k(0, 1/2) - \gamma_k(\beta, 1/2)) = \frac{2}{\pi} \arcsin \sqrt{\beta},$$

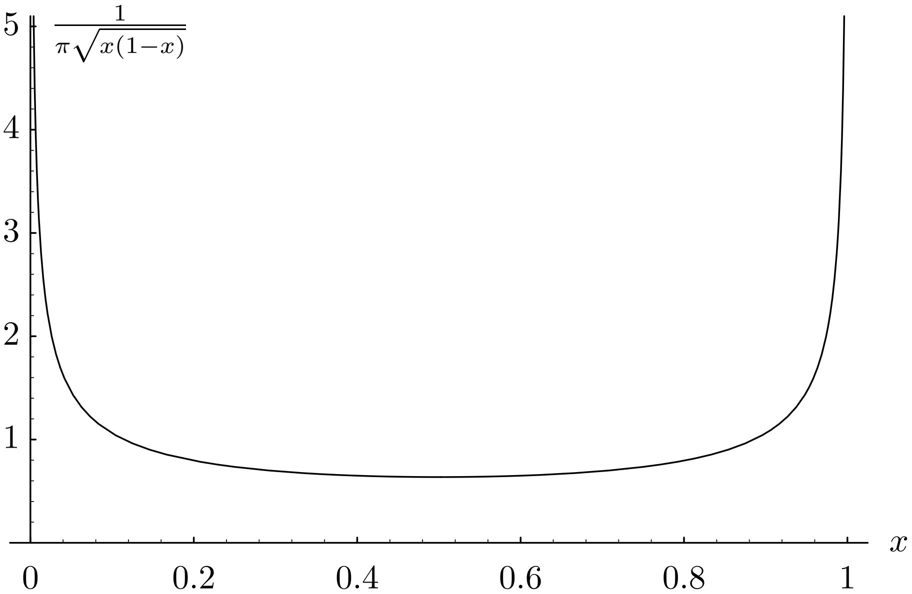
für $\beta > 1/2$

$$\lim_{k\to\infty} \gamma_k(0, \beta) = \lim_{k\to\infty} (\gamma_k(0, 1/2) + \gamma_k(1/2, \beta)) = \frac{2}{\pi} \arcsin \sqrt{\beta}.$$

Für $\gamma_k(\alpha, 1)$ führt dasselbe Argument zum Ziel. □

Der Beweis des Arcus-Sinus-Gesetzes wurde in einer allgemeineren Form zuerst von *Paul Pierre Lévy* (1886-1971) im Jahre 1939 gegeben.

Die Funktion $\frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$ hat das folgende Aussehen:



$\frac{1}{\pi\sqrt{x(1-x)}}$
5
4
3
2
1
0
0.2
0.4
0.6
0.8
1
x

5 Ergodensätze

Schon im Kapitel über die Gesetze der großen Zahlen hatten wir uns mit dem Grenzverhalten des Mittelwertes $\frac{1}{n}\sum_{i=1}^{n} X_i$ einer Folge $X_1, X_2, \ldots$ von Zufallsvariablen befasst. Diese waren dort unabhängig und identisch verteilt. Wir wollen die Frage nun für eine Folge, die allgemeineren Bedingungen genügt (die insbesondere die Abhängigkeit der Zufallsvariablen zulassen), wieder aufnehmen. Natürlich kann die Verteilung der $X_1, X_2, \ldots$ nicht völlig willkürlich sein. Insbesondere wären X_i mit zu großem Gewicht natürlich schädlich, denn sie würden allein das Verhalten des Mittelwerts bestimmen. Wir werden uns deshalb auf sogenannte stationäre Zufallsvariablen konzentrieren.

Definition 5.1 *Eine Folge $X_0, X_1, \ldots$ von Zufallsvariablen heißt stationär, falls für jedes $k \in \mathbb{N}$ die Folge $X_k, X_{k+1}, \ldots$ dieselbe Verteilung hat wie die Originalfolge $X_0, X_1, \ldots$ Dies ist genau dann der Fall, wenn für jedes $n \in \mathbb{N}$ die Vektoren $(X_0, \ldots, X_n)$ und $(X_k, \ldots, X_{k+n})$ für jedes $k \in \mathbb{N}$ die gleiche Verteilung haben.*

Die folgenden Beispiele werden uns durch dieses Kapitel begleiten.

Beispiel 5.2 *Die Folge $X_0, X_1, \ldots$ ist i.i.d. Offenbar haben dann die Zufallsvektoren $(X_0, \ldots, X_n)$ und $(X_k, \ldots, X_{k+n})$ dieselbe (Produkt-)Verteilung; die Folge ist also stationär.*

Beispiel 5.3 *Es sei Ω eine endliche Menge und $X_0, X_1, \ldots$ eine Markov-Kette auf Ω Wir nehmen an, dass diese Markov-Kette homogen ist, d.h. dass die Übergangsmatrix Q nicht vom Index n abhängt. Q habe die stationäre Verteilung π, d. h. es gelte*

$$\pi Q = \pi$$

oder mit anderen Worten

$$\pi(A) = \int \pi(x) Q(x, A).$$

Aus der Definition der Stationarität von π folgt unmittelbar, dass auch die Folge $X_0, X_1, \ldots$ stationär ist. Eine Markov-Kette, die für Gegenbeispiele interessant sein wird, ist die folgende: $\Omega = \{0, 1\}$,

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

und $\pi(0) = \pi(1) = \frac{1}{2}$. Es gibt hierbei genau zwei stationäre Folgen, nämlich

$$\begin{aligned} (X_0, X_1, \ldots) &= (1, 0, 1, 0 \ldots) \quad \textit{und} \\ (X_0, X_1, \ldots) &= (0, 1, 0, 1 \ldots). \end{aligned}$$

Beispiel 5.4 *(Drehung auf dem Kreis) Es sei hierfür* $\Omega = [0,1)$ *(das wir mit der Einheitssphäre* S^1 *im* $\mathbb{R}^2$ *identifizieren),* $\mathcal{F} = \mathcal{B}^1$ *und* $\mathbb{P} = \lambda^1$*. Für* $\vartheta \in (0,1)$ *und* $n \geq 0$ *definieren wir*

$$X_n(\omega) := (\omega + n\vartheta) \text{ mod } 1$$

was wir mit der obigen Interpretation von Ω *als die Drehung von* ω *und den Winkel* $2\pi\vartheta$ *verstehen können. Da* λ^1 *die Gleichverteilung auf* $[0,1)$ *ist, ist die Folge stationär, denn somit ist auch* $(\omega + n\vartheta)$ *mod 1 gleichverteilt. Man kann Beispiel 2.3 als eine kontinuierliche Markov-Kette mit Übergangskern*

$$Q(x,y) = \begin{cases} 1 & y = x + \vartheta \ (\text{mod } 1) \\ 0 & \text{sonst} \end{cases}$$

auffassen.

Um weitere Beispiele zu konstruieren, benutzen wir das folgende Theorem.

Theorem 5.5 *Ist die Folge* $X_0, X_1, \ldots$ *stationär und*

$$g : \mathbb{R}^{\mathbb{N}_0} \to \mathbb{R}$$

messbar, dann ist auch die Folge der

$$Y_k = g(X_k, X_{k+1}, \ldots)$$

eine stationäre Folge.

Beweis: Für $x \in \mathbb{R}^{\mathbb{N}_0}$ bezeichnen wir mit $g_k(x)$ die Abbildung

$$g_k(x) = g(x_k, x_{k+1}, \ldots).$$

Für $B \in \mathcal{B}^{\mathbb{N}_0}$ sei weiter

$$A = \{x : (g_0(x), g_1(x), g_2(x), \ldots) \in B\}.$$

Um die behauptete Stationarität zu beweisen, beobachten wir, dass

$$\begin{aligned} \mathbb{P}(\omega : (Y_0, Y_1, \ldots) \in B) &= \mathbb{P}(\omega : (X_0, X_1, \ldots) \in A) \\ &= \mathbb{P}(\omega : (X_k, X_{k+1}, \ldots) \in A) \\ &= \mathbb{P}(\omega : (Y_k, Y_{k+1}, \ldots) \in B) \end{aligned}$$

gilt. Dies heißt $(Y_n)_n$ ist stationär. □

Beispiel 5.6 *(Bernoulli-Shift) Es sei hierfür* $\Omega = [0,1)$*,* $\mathcal{F} = \mathcal{B}^1$ *und* $\mathbb{P} = \lambda^1|_{[0,1)}$*. Weiter sei* $Y_0(\omega) = \omega$ *und für* $n \geq 1$ *definieren wir*

$$Y_n(\omega) = 2 \cdot Y_{n-1}(\omega) \text{ mod } 1.$$

Dies ist offenbar ein Spezialfall von Theorem 5.5, wenn man die folgende Darstellung von $X \in (0,1)$ wählt: Wähle $X_0, X_1, \dots$ i.i.d. mit

$$\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = \frac{1}{2}$$

und schreibe für

$$g(x) = \sum_{i=0}^{\infty} x_i 2^{-(i+1)}.$$

Der Name Bernoulli-Shift kommt daher, dass eine Multiplikation mit 2 die x_i nach links schiebt. Man kann Beispiel 5.6 auch als Spezialfall von Beispiel 5.3 auffassen.

Schließlich sind Beispiel 5.4 und 5.6 Spezialfälle des folgenden Beispiels:

Beispiel 5.7 *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und*

$$\varphi : \Omega \to \Omega$$

eine messbare Abbildung. Wir wollen φ maßtreu nennen, wenn

$$\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A)$$

für alle $A \in \mathcal{F}$ gilt. Für eine messbare Funktion X ist dann

$$X_n(\omega) = X(\varphi^n(\omega))$$

eine stationäre Folge. In der Tat: Sei $B \in \mathcal{B}^{n+1}$ und

$$A = \{\omega : (X_0(\omega), X_1(\omega), \dots, X_n(\omega)) \in B\}.$$

Dann gilt für jedes k

$$\mathbb{P}((X_k, \dots, X_{k+n}) \in B) = \mathbb{P}(\varphi^k(\omega) \in A) = \mathbb{P}(\omega \in A) = \mathbb{P}((X_0, \dots, X_n) \in B).$$

Dieses Beispiel ist nicht nur ein sehr wichtiges Beispiel, sondern in gewissem Sinne das einzige Beispiel. In der Tat folgt aus dem Komlogorovschen Erweiterungssatz (siehe Anhang), dass man zu jeder stationären Folge $Y_0, Y_1, \dots$ ein Maß $\mathbb{P}$ auf $\bigotimes_{i=0}^{\infty}(S, \mathcal{S})$ konstruieren kann (wobei $(S, \mathcal{S})$ der zugrunde liegende Wahrscheinlichkeitsraum ist), so dass $X_n(\omega) = \omega_n$ die gewünschte Verteilung hat. Wenn wir φ nun als Shift-Operator definieren, d.h.

$$\varphi(\omega_0, \omega_1, \omega_2, \dots) = (\omega_1, \omega_2, \dots)$$

und

$$X(\omega) = \omega_0$$

wählen, dann ist φ maßtreu, also $X_n(\omega) = X(\varphi^n(\omega))$.

Derselbe Kolmogorovsche Erweiterungssatz erlaubt es uns auch, statt einseitiger stationärer Folgen $X_0, X_1, \dots$ zweiseitige $(X_n)_{n \in \mathbb{Z}}$ zu betrachten.

Theorem 5.8 *Jede stationäre Folge $(X_n)_{n\in\mathbb{N}_0}$ kann in eine zweiseitige stationäre Folge $(Y_n)_{n\in\mathbb{Z}}$ eingebettet werden.*

Beweis: Man beachte nun, dass die Folge

$$\mathbb{P}(Y_{-m} \in A_0, \ldots, Y_n \in A_{m+n}) = \mathbb{P}(X_0 \in A_0, \ldots, X_{m+n} \in A_{m+n})$$

eine konsistente Familie endlich-dimensionaler Verteilung ist. Nach dem Kolmogorovschen Erweiterungssatz gibt es somit ein Maß $\mathbb{P}$ auf $(S, \mathcal{S})$, so dass $Y_n(\omega) = \omega_n$ die gewünschte Verteilung hat. □

Aufgrund dieser Beobachtungen werden wir die folgende Theorie stationärer Folgen im Kontext von Beispiel 5.7 entwickeln. Wir nennen dabei eine messbare Menge A invariant, wenn $\varphi^{-1}A = A$ gilt. Zwei Mengen werden dabei als gleich angesehen, wenn ihre symmetrische Differenz Maß Null hat.

Eine maßtreue Abbildung auf $(\Omega, \mathcal{F}, \mathbb{P})$ wollen wir ergodisch nennen, wenn die σ-Algebra $\mathcal{J}$ der invarianten Mengen

$$\mathcal{J} := \{A \in \mathcal{F} : \varphi^{-1}A = A\}$$

trivial ist, d. h. wenn

$$\mathcal{J} = \{\emptyset, \Omega\}$$

gilt oder mit anderen Worten, wenn für alle $A \in \mathcal{J}$

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

gilt.

Die Bedeutung dieser Definition liegt darin, dass man im nicht-ergodischen Fall den Raum in zwei Teile A und A^c positiven Maßes zerlegen kann, so dass

$$\varphi(A) = A \quad \text{und} \quad \varphi(A^c) = A^c$$

gilt, also φ nicht "irreduzibel" ist. Somit lässt sich gewissermaßen jedes interessante Beispiel auf den ergodischen Fall zurückspielen.

Um noch mehr über die Bedeutung von Ergodizität zu erfahren, kehren wir zu unseren Beispielen zurück. Falls $\Omega = \mathbb{R}^{\mathbb{N}_0}$ ist und φ der Shift-Operator, dann ist eine invariante Menge von der Gestalt

$$A = \{\omega : \omega \in A\} = \{\omega : \varphi\omega \in A\}.$$

Iteriert man das Anwenden von φ, sieht man, dass für eine invariante Menge A gilt

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots) \in \mathcal{T}_\infty.$$

Also ist $\mathcal{J} \subseteq \mathcal{T}_\infty$.

Ist $(X_n)_n$ eine i.i.d. Folge, so impliziert das Kolmogorovsche 0-1-Gesetz, dass $\mathcal{T}_\infty$ trivial ist, also auch das $\mathcal{J}$ trivial ist. Also ist die Folge ergodisch, d.h., wenn man das Produktmaß auf den Folgenraum liftet, ist der Shift ergodisch.

Wenden wir uns Beispiel 5.3, den Markov-Ketten, zu: Wir nehmen an, dass die invariante Verteilung der Markov-Kette π der Bedingung $\pi(x) > 0$ für alle $x \in \Omega$ genügt. Im vorigen Kapitel haben wir gesehen, dass dies unter den Voraussetzungen des Ergodensatzes für Markov-Ketten der Fall ist, also falls es eine Potenz von Q mit nur positiven Einträgen gibt (d.h. ein $N \in \mathbb{N}$ mit $Q^N \gg 0$). Dies ist auch der Zusammenhang zwischen den Ergodensätzen, die in der Folge studiert werden sollen, und dem Ergodensatz für Markov-Ketten.

Unter diesen Voraussetzungen ist jeder Zustand der Kette rekurrent, d.h.

$$\mathbb{P}(\text{„Die Kette startet in } y \text{ und kommt irgendwann nach } y \text{ zurück"}) = 1$$

für alle $y \in \Omega$. Das hat damit zu tun, dass

$$\sum_{x \in \Omega} \pi(x) \sum_{n=1}^{\infty} Q^n(x, y) = \sum_{n=1}^{\infty} \pi(y) = \infty$$

gilt und ferner daraus

$$\infty = \sum_x \frac{\pi(x)\rho_{xy}}{1 - \rho_{yy}} \leq \frac{1}{1 - \rho_{yy}}$$

folgt, wobei ρ_{xy} die Wahrscheinlichkeit bezeichnet, dass die Irrfahrt irgendwann von x nach y läuft. Also können wir

$$\Omega = \bigcup R_i$$

schreiben, wobei R_i disjunkte, irreduzible Teilmengen von Ω sind, d. h. solche, fï¿½r die für alle $x, y \in R_i$ ein n mit $Q(x, y) > 0$ existiert. Mit anderen Worten, ist $X_0 \in R_i$, so ist mit Wahrscheinlichkeit 1 auch $X_n \in R_i$ für alle $n \geq 1$. Also ist

$$\{\omega : X_0(\omega) \in R_i\} \in \mathcal{J}.$$

Diese Beobachtung zeigt, dass die Folge nur ergodisch sein kann, wenn die Markov-Kette ergodisch ist, d. h. wenn $\Omega = R_1$ ist.

Um auch die Umkehrung zu beweisen, beachte man, dass für $A \in \mathcal{J}$ gilt $1_A \circ \vartheta_n = 1_A$, wobei

$$\vartheta_n(\omega_0, \omega_1, \omega_2, \ldots) = (\omega_n, \omega_{n+1}, \ldots)$$

der Shift-Operator ist. Schreiben wir also

$$\mathcal{F}_n = \sigma(X_0, \ldots, X_n),$$

so implizieren die Shift-Invarianz von 1_A und die Markov-Eigenschaft

$$\mathbb{E}_\pi[1_A | \mathcal{F}_n] = \mathbb{E}_\pi[1_A \vartheta_n | \mathcal{F}_n] = h(X_n),$$

wobei $h(x) = \mathbb{E}_x \mathbb{1}_A$ gesetzt werde. Das 0-1-Gesetz von Levy besagt nun, dass für ein $A \in \mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \geq 1)$

$$\mathbb{E}[1_A | \mathcal{F}_n] \to 1_A \text{ f.s.}$$

gilt, also $h(x) \to 1_A$ $\mathbb{P}$-f.s. mit $n \to \infty$.

Falls X_n irreduzibel und rekurrent ist, so ist für jedes $y \in \Omega$ die rechte Seite unendlich oft $h(y)$. Also ist $h(x)$ entweder identisch 0 oder 1, also $\mathbb{P}_\pi(A) \in \{0,1\}$. An diesem Beispiel lässt sich auch veranschaulichen, dass $\mathcal{J}$ und $\mathcal{T}_\infty$ verschieden sein können. Ist nämlich Q irreduzibel, aber gilt, dass jeder Zustand $x \in \Omega$ eine Periode $d > 1$ hat für ein $N \in \mathbb{N}$, dann ist $\mathcal{J}$ trivial, aber $\mathcal{T}$ nicht. Dies macht man sich leicht am Beispiel einer Markov-Kette auf einem zweielementigen Zustandsraum mit Übergangsmatrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

klar.

Betrachten wir das Beispiel der Drehung des Kreises (Beispiel 5.4), so ist klar, dass diese nicht ergodisch ist, wenn der Winkel $\Theta = \frac{m}{n}$ für ein $0 \leq m \leq n$ erfüllt. In der Tat: Ist $B \in [0, \frac{1}{n})$ eine Borelmenge und setzen wir

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

so ist A offenbar invariant. Ist umgekehrt Θ irrational, so ist φ ergodisch. Um dies zu beweisen ziehen wir ein Faktum aus der Fourier-Analysis heran: Ist $f : [0,1) \to \mathbb{R}$ messbar mit

$$\int f^2(x)dx < \infty,$$

dann kann f als die folgende Fourierreihe geschrieben werden:

$$f(x) = \sum_k c_k e^{2\pi ikx}$$

(wobei man, wie üblich, dies als

$$\sum_{k=-N}^{N} c_k e^{2\pi ikx} \underset{N\to\infty}{\longrightarrow} f(x)$$

in $L^2[0,1)$ auffassen sollte). Man kann auch die Koeffizienten c_k bestimmen. Diese haben die Gestalt

$$c_k = \int_0^1 f(x)e^{-2\pi ikx}dx.$$

Mit der Wahl von φ als Drehung des Kreises ist also

$$f(\varphi(x)) = \sum_k c_k e^{2\pi ik(x+\Theta)} = \sum_k (c_k e^{2\pi ik\Theta})e^{2\pi ikx}.$$

Da die c_k eindeutig bestimmt sind, kann also $f(x) = f(\varphi(x))$ nur dann gelten, wenn

$$c_k(e^{2\pi ik\Theta} - 1) = 0$$

gilt. Ist Θ irrational, so kann dies nur für

$$c_k = 0 \quad (k \neq 0)$$

der Fall sein. Somit ist f konstant. Wendet man dies auf $f = 1_A$ für ein $A \in \mathcal{J}$ an, folgt daraus, dass

$$A \in \{\emptyset, [0,1)\} \ \mathbb{P}\text{-f.s.}$$

folgt.

Schließlich zeigen wir, dass auch der Bernoulli-Shift aus Beispiel 5.6 ergodisch ist. Um dies zu beweisen, sei daran erinnert, dass die stationäre Folge $Y_n(\omega) = \varphi^n(\omega)$ sich als

$$Y_n = \sum_{m=0}^{\infty} 2^{-(m+1)} X_{n+m}$$

schreiben lässt, wobei $X_0, X_1, \ldots$ eine i.i.d. Folge mit

$$\mathbb{P}(X_k = 0) = \mathbb{P}(X_k = 1) = \frac{1}{2}$$

ist. Weiter benötigen wir

Theorem 5.9 *Falls $X_0, X_1, \ldots$ eine ergodische, stationäre Folge ist und*

$$g : \mathbb{R}^{\mathbb{N}_0} \to \mathbb{R}$$

messbar, dann ist auch die Folge der

$$Y_k := g(X_k, X_{k+1}, \ldots)$$

ergodisch.

Beweis: Aufgrund der Bemerkung nach Beispiel 5.7 genügt es, die Situation zu betrachten, in der für $\omega \in \Omega^{\mathbb{N}_0}$ die Folge der X_i durch $X_n(\omega) = \omega_n$ definiert ist. Falls B der Bedingung

$$\{\omega : (Y_0, Y_1, \ldots) \in B\} = \{\omega : (Y_1, Y_2, \ldots) \in B\}$$

genügt, dann ist die Menge

$$A = \{\omega : (Y_0, Y_1, \ldots) \in B\}$$

Shift-invariant. Von diesen wissen wir schon, dass sie Maß 0 oder 1 haben.

Die Ergodizität des Bernoulli-Shifts folgt nun aus der Ergodizität der i.i.d. Folge $(X_k)_k$. □

Wir wollen nun eine Art "Gesetz der großen Zahlen" für stationäre Folgen beweisen. Dieses ist unter dem Namen Birkhoffscher Ergodensatz bekannt. Es besagt, das das zeitliche Mittel einer Folge von stationären Beobachtungen einer integrierbaren Größe (physikalisch: Observable) gegen den bedingten Erwartungswert dieser Größe bzgl. der σ-Algebra $\mathcal{J}$ der invarianten Mengen konvergiert. Insbesondere ist der Limes im Falle ergodischer Beobachtungen gleich dem Erwartungswert der Größe.

Theorem 5.10 *Es sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $X \in \mathcal{L}^1(\mathbb{R})$*. Ferner sei*

$$\varphi : \Omega \to \Omega$$

eine maßtreue Abbildung. Dann gilt

$$\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m \omega) \to \mathbb{E}(X|\mathcal{J})$$

$\mathbb{P}$*-f.s. und in* $\mathcal{L}^1(\mathbb{P})$*. Hierbei ist wie oben* $\mathcal{J}$ *die* σ*-Algebra der invarianten Mengen.*

Dieser Ergodensatz geht auf Birkhoff (1931) zurück. Die Bezeichnung stammt aus der Physik, in welcher die Ergodenhypothese besagt, dass für eine Observable ihr räumliches und ihr zeitliches Mittel übereinstimmen.

Unser Beweis beruht auf dem sogenannten maximalen Ergodenlemma.

Lemma 5.11 *(Maximales Ergodenlemma) Sei* $X_j(\omega) = X(\varphi^j \omega)$ *und*

$$S_k(\omega) = \sum_{m=0}^{k-1} X_m(\omega)$$

und schließlich

$$M_k(\omega) = \max(0, S_1(\omega), \ldots, S_k(\omega)).$$

Dann gilt

$$\mathbb{E}(X, M_k > 0) \geq 0,$$

wobei

$$\mathbb{E}(X, M_k > 0) := \int_{\{M_k > 0\}} X d\mathbb{P}.$$

Beweis: Die Aussage des obigen Lemmas ist nicht besonders einsichtig. Der folgende Beweis von Garsia (1965) folgt diesem Beispiel.

Allerdings ist keiner der angeführten Schritte schwierig: Für $j \leq k$ ist per definitionem $M_k(\varphi\omega) \geq S_j(\varphi\omega)$, also

$$X(\omega) + M_k(\varphi\omega) \geq X(\omega) + S_j(\varphi\omega) = S_{j+1}(\omega).$$

Mit anderen Worten gilt

$$X(\omega) \geq S_{j+1}(\omega) - M_k(\varphi\omega)$$

für alle $j = 1 \ldots k$. Trivialerweise gilt auch für $j = 0$

$$X(\omega) \geq S_1(\omega) - M_k(\varphi\omega),$$

denn $S_1(\omega) = X(\omega)$ und M_k ist definitionsgemäß nicht negativ. Integrieren wir, ergibt sich

$$\begin{aligned}\mathbb{E}(X(\omega), M_k > 0) &\geq \int_{\{M_k>0\}} \max(S_1(\omega, \ldots, S_k(\omega)) - M_k(\varphi\omega) d\mathbb{P} \\ &= \int_{\{M_k>0\}} M_k(\omega) - M_k(\varphi\omega) d\mathbb{P}.\end{aligned}$$

Nun ist auf der Menge $\{M_k > 0\}^c$ die Zufallsgröße $M_k(\omega) = 0$ und (wie immer) gilt $M_k(\varphi\omega) \geq 0$. Somit gilt

$$\mathbb{E}(X(\omega), M_k > 0) \geq \int_{\{M_k>0\}} M_k(\omega) - M_k(\varphi\omega) d\mathbb{P} \geq \int_\Omega M_k(\omega) - M_k(\varphi\omega) d\mathbb{P} = 0,$$

wobei die letzte Gleichheit folgt, da φ als maßtreu vorausgesetzt war. □

Der Beweis des eigentlichen Ergodensatzes beginnt mit einer kleinen Übung:

Übung 5.12 *Man zeige, dass eine Abbildung X messbar ist bezüglich $\mathcal{J}$ genau dann, wenn*

$$X \circ \varphi = X$$

$\mathbb{P}$-*f.s. gilt.*

Beweis von Theorem 5.10. Da definitionsgemäß $\mathbb{E}(X|\mathcal{J})$ messbar ist bezüglich $\mathcal{J}$, folgt

$$\mathbb{E}(X|\mathcal{J}) \circ \varphi = \mathbb{E}(X|\mathcal{J}).$$

Also kann man, notfalls durch Übergang auf

$$X' = X - \mathbb{E}(X|\mathcal{J}),$$

annehmen, dass $\mathbb{E}(X|\mathcal{J}) = 0$ ist (dazu ist die Shiftinvarianz von $\mathbb{E}(X|\mathcal{J})$ offenbar notwendig). Wir setzen

$$\bar{X} := \limsup \frac{S_n}{n}$$

und wollen also zeigen, dass $\bar{X}$ gegen 0 konvergiert. Sei also $\varepsilon > 0$ gegeben. Wir bezeichnen mit

$$D := \{\omega : \bar{X}(\omega) > \varepsilon\}$$

und wollen also $\mathbb{P}(D) = 0$ zeigen. Nun ist der Unterschied zwischen $S_n^{(\omega)}$ und $S_n(\varphi(\omega))$ gerade $X_0(\omega) - X_n(\omega)$, diese Terme entfallen bei der Mittelwertbildung im Limes $n \to \infty$. Also gilt $\bar{X}(\varphi(\omega)) = \bar{X}(\omega)$ und somit $D \subset \mathcal{J}$. Wir führen die folgenden Bezeichnungen ein:

$$\begin{aligned}&X^*(\omega) := (X(\omega) - \varepsilon)\mathbb{1}_D(\omega), \\ &S_n^*(\omega) := X^*(\omega) + \ldots + X^*(\varphi^{n-1}(\omega)), \\ &M_n^*(\omega) := \max(0, S_1^*(\omega), \ldots, S_n^*(\omega)), \\ &F_n := \{M_n^* > 0\} \quad \text{und} \\ &F := \bigcup_n F_n = \{\sup_{k \geq 1} \frac{S_k^*}{k} > 0\}.\end{aligned}$$

Da nun X^* gleich $X(\omega) - \varepsilon$ auf der Menge

$$D = \{\limsup \frac{S_k}{k} > \varepsilon\}$$

und 0 sonst ist, bekommen wir

$$F = \{\sup_k \frac{S_k}{k} > \varepsilon\} \cap D = D.$$

Das maximale Ergodenlemma (Lemma 5.11) besagt nun, dass

$$\mathbb{E}(X^*, F_n) \geq 0$$

gilt. Da nun

$$\mathbb{E}|X^*| \leq E|X| + \varepsilon < \infty$$

gilt, können wir den Satz von der majorisierten Konvergenz anwenden und

$$\mathbb{E}(X^*, F_n) \to \mathbb{E}(X^*, F)$$

folgern. Somit ist auch $\mathbb{E}[X^*; F] \geq 0$. Diese unschuldig aussehende Behauptung bekommt ihre Bedeutung daher, dass wir nach obiger Überlegung auch D für F schreiben dürfen; somit folgt

$$0 \leq \mathbb{E}(X^*; D) = \mathbb{E}(X - \varepsilon; D) = \mathbb{E}(\mathbb{E}(X|\mathcal{J}); D) - \varepsilon\mathbb{P}(D) = -\varepsilon\mathbb{P}(D).$$

Hierbei folgt die vorletzte Gleichung aus der Definition der bedingten Erwartung und die letzte Gleichheit aus der Tatsache, dass wir $\mathbb{E}(X|\mathcal{J}) = 0$ vorausgesetzt hatten. Also bekommen wir

$$0 = \mathbb{P}(D) = \mathbb{P}(\limsup \frac{S_n}{n} > \varepsilon).$$

Somit folgt $\limsup \frac{S_n}{n} \leq 0$ $\mathbb{P}$-f.s. Ersetzt man X durch $-X$, so erhält man zusammen

$$\frac{S_n}{n} \to 0 \quad \mathbb{P}\text{-f.s.}$$

Um zu beweisen, dass die Konvergenz auch in $\mathcal{L}^1(\mathbb{P})$ ist, setzen wir

$$X'(\omega) = X(\omega)\mathbb{1}_{\{|X(\omega)| \leq \omega\}}$$

und

$$X''(\omega) = X(\omega) - X'(\omega).$$

Nun besagt der erste Teil des Satzes, den wir soeben bewiesen haben, dass

$$\frac{1}{n}\sum_{m=0}^{n-1} X'(\varphi^m \omega) \to \mathbb{E}(X'|\mathcal{J}) \quad \mathbb{P}\text{-f.s.}$$

gilt. Da X' beschränkt ist, bekommen wir aus dem Satz über marjorisierte Konvergenz auch, dass auch

$$\mathbb{E}(|\frac{1}{n}\sum_{m=0}^{n-1} X'(\varphi^m \omega) - \mathbb{E}(X'|\mathcal{J})| \to 0 \tag{46}$$

gilt, wenn $n \to \infty$ gilt. Um auch X'' zu behandeln, benutzen wir, dass

$$\mathbb{E}|\frac{1}{n}\sum_{m=0}^{n-1} X''(\varphi^m(\omega))| \leq \frac{1}{n}\sum_{m=0}^{n-1} \mathbb{E}|X''(\varphi^m(\omega))| = \mathbb{E}|X''|$$

gilt und dass

$$\mathbb{E}|\mathbb{E}(X''|\mathcal{J})| \leq \mathbb{E}\,\mathbb{E}(|X''|\,|\mathcal{J}) = \mathbb{E}|X''|$$

aus der Glättungseigenschaft der bedingten Erwartung folgt. Also ergibt sich zusammen

$$\begin{aligned}
& \mathbb{E}|\frac{1}{n}\sum_{m=0}^{n-1} X''(\varphi^m(\omega)) - \mathbb{E}(X''|\mathcal{J})| \\
\leq\ & \mathbb{E}|\frac{1}{n}\sum_{m=0}^{n-1} X''(\varphi^m(\omega))| + \mathbb{E}|\mathbb{E}(X''|\mathcal{J})| \\
\leq\ & 2\mathbb{E}|X''|.
\end{aligned}$$

Somit gilt auch

$$\limsup_{n\to\infty} \mathbb{E}|\frac{1}{n}\sum_{m=0}^{n-1} X''(\varphi^m(\omega)) - \mathbb{E}(X''|\mathcal{J})| \leq 2\mathbb{E}(X'').$$

Da nun X als integrierbar vorausgesetzt war, folgt

$$\mathbb{E}|X''| \to 0 \quad \text{wenn} \quad M \to \infty$$

aus dem Satz über majorisierte Konvergenz. Zusammen mit (46) ergibt dies die behauptete $\mathcal{L}^1$-Konvergenz von X gegen $\mathbb{E}(X|\mathcal{J})$. Somit ist der Ergodensatz bewiesen. □

Bevor wir uns die Konsequenzen des Ergodensatzes anhand unserer Beispiele betrachten, leiten wir zunächt eine für Zwecke nützliche Ungleichung aus dem maximalen Ergodenlemma her:

Proposition 5.13 *(Wieners Maximal-Ungleichung) Wie im Ergodensatz sei $\varphi : \Omega \to \Omega$ eine maßtreue Abbildung und $X \in \mathcal{L}^1(\mathbb{P})$. Ferner setzen wir $X_j(\omega) = X(\varphi^j(\omega))$ und*

$$S_k(\omega) = X_0(\omega) + X_1(\omega) + \ldots + X_{k-1}(\omega)$$

und $A_k(\omega) = \frac{S_k(\omega)}{k}$. Schließlich sei $D_k = \max(A_1, \ldots, A_k)$. Dann gilt für jedes $\alpha > 0$

$$\mathbb{P}(D_k > \alpha) \leq \frac{1}{\alpha}\mathbb{E}|X|.$$

Beweis: Setze $B := \{D_k > \alpha\}$. Wenden wir das Maximale Ergodenlemma auf

$$X' := X - \alpha 1_B$$

an, wobei wir

$$S'_j(\omega) = X'(\varphi)(\omega)) \quad \text{und} \quad S'_k = \sum_{j=0}^{k-1} X'_j(\omega)$$

und schließlich

$$M'_k = \max(0, S'_1, \ldots, S'_k)$$

setzen, so ergibt sich

$$\mathbb{E}(X', M'_k > 0) \geq 0.$$

Da nun

$$\{M'_k > 0\} = \{D_k > \alpha\} \equiv B$$

gilt, folgt

$$\mathbb{E}|X| \geq \int_B |X| d\mathbb{P} \geq \alpha \int_B 1_B d\mathbb{P} = \alpha \mathbb{P}(B).$$

Teilt man durch $\alpha > 0$, ergibt sich die Behauptung. □

Nun betrachten wir die Auswirkungen des Birkhoffschen Ergodensatzes für unsere Beispiele.

Beispiel 5.14 *(I.I.D. Folgen) Da wir schon festgestellt haben, dass i.i.d. Folgen ergodisch sind, d.h. dass $\mathcal{J}$ trivial ist, ist*

$$\mathbb{E}(X|\mathcal{J}) = \mathbb{E}X \text{ } \mathbb{P}\text{-f.s.}$$

für alle integrierbaren X. Der Ergodensatz (Theorem 5.10) behauptet somit für i.i.d. Folgen $X_0, X_1, \ldots$

$$\frac{1}{n} \sum_{m=0}^{n-1} X_m \underset{n \to \infty}{\longrightarrow} \mathbb{E}X_0,$$

wobei die Konvergenz sowohl $\mathbb{P}$-f.s. als auch in $\mathcal{L}^1(\mathbb{P})$ ist. Die Aussage der $\mathbb{P}$-f.s.-Konvergenz ist auch als das starke Gesetz der großen Zahlen bekannt.

Beispiel 5.15 *(Markov-Ketten) Es sei $(X_n)_n$ eine irreduzible Markov-Kette auf einem endlichen Zustandsraum Ω und stationärer Verteilung π. Es sei $f : \Omega \to \mathbb{R}$ eine bezüglich π integrierbare Funktion, d. h.*

$$\sum |f(x)|\pi(x) < \infty.$$

In der Diskussion des Begriffs "ergodisch" haben wir gesehen, dass $\mathcal{J}$ in dieser Situation trivial ist, d.h. für alle $A \in \mathcal{J}$ gilt $\mathbb{P}(A) \in \{0, 1\}$. Somit ist wie in Beispiel 5.14

$$\mathbb{E}(X|\mathcal{J}) = \mathbb{E}X \quad \mathbb{P}\text{-f.s.}$$

Wendet man daher den Ergodensatz auf $f(X_0(\omega))$ an, so ergibt sich die Konvergenzaussage

$$\frac{1}{n} \sum_{m=0}^{n-1} f(X_m) \to \mathbb{E}_\pi(f) = \sum_x f(x)\pi(x)$$

$\mathbb{P}$-f.s. als auch in $\mathcal{L}^1(\mathbb{P})$.

Beispiel 5.16 *(Drehung des Kreises) Hierfür sei wieder* $\Omega = [0,1)$, $\mathcal{F} = B|_{[0,1)}$ *und* $\mathbb{P} = \lambda^1|_{[0,1)}$. φ *sei die maßtreue Abbildung*

$$\varphi(\omega) = (\omega + \vartheta) \; mod \; 1$$

für ein irrationales $\vartheta \in (0,1)$. *Wie oben diskutiert ist auch hier die* σ*-Algebra* $\mathcal{J}$ *trivial, somit für jedes* $X \in \mathcal{L}^1(\mathbb{P})$

$$\mathbb{E}(X|\mathcal{J}) = \mathbb{E}(X) \quad \mathbb{P}\text{-}f.s.$$

Setzen wir insbesondere für $X = 1_A$ *mit einer messbaren Menge* A, *so impliziert der Birkhoffsche Ergodensatz*

$$\frac{1}{n} \sum_{m=0}^{m-1} \mathbb{1}_{(\varphi^m(\omega) \in A)} \to \lambda^1(A) \quad \mathbb{P}\text{-}f.s. \text{ und in } \mathcal{L}^1(\lambda^1). \tag{47}$$

Wendet man dieses Resultat für $\omega = 0$ *an (wobei der Ergodensatz nicht sagt, dass das Resultat für* $\omega = 0$ *stimmen muss; das ist vielmehr ein Ergebnis aus der Zahlentheorie) und ein Intervall* A *an, so erhält man*

$$\frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_{((m \cdot \vartheta) mod\, 1 \in A)} \to \lambda^1(A).$$

Dies ist der sogenannte Weylsche Gleichverteilungssatz. Einen nicht-probabilistischen Beweis findet man im Buch von Hardy und Wright [2], S. 390 – 393.

Wie gesagt bekommen wir dieses Resultat nicht direkt aus dem Birkhoffschen Ergodensatz. Wir wollen es hier probabilistisch herleiten. Dazu zeigen wir, dass für $A = [a,b)$ *die Ausnahmemenge in Gleichung* (47) *(also die Menge, für die* (47) *nicht gilt) die leere Menge ist. Wir schreiben hierfür*

$$A_k := [a + \frac{1}{k}, b - \frac{1}{k}).$$

Für hinreichend großes k *ist* $A_k \neq \emptyset$ *und* (47) *impliziert*

$$\frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_{A_k}(\varphi^m(\omega)) = b - a - \frac{2}{k}$$

für alle $\omega \in \Omega_k$, *wobei* Ω_k *eine Menge mit* $\mathbb{P}(\Omega_k) = 1$ *ist. Setzen wir*

$$\Omega_\infty = \bigcap_k \Omega_k,$$

wobei nur solche Ω_k *am Durchschnitt teilnehmen, für die* A_k *nicht leer ist, so folgt auch*

$$\mathbb{P}(\Omega_\infty) = 1.$$

Also ist Ω_∞ *dicht in* $[0,1)$. *Ist nun* $x \in [0,1)$ *und* $\omega \in \Omega_\infty$ *mit* $|\omega - x| < \frac{1}{k}$, *so folgt aus* $\varphi^m(\omega) \in A_k$, *dass* $\varphi^m(x) \in A$ *gilt. Also erhalten wir für jedes* $x \in [0,1)$

$$\liminf \frac{1}{n} \sum_{m=0}^{n-1} 1_A(\varphi^m(x)) \geq b - a - \frac{2}{k}$$

für alle hinreichend großen k. Wendet man die gleichen Überlegungen auf A^c (das sich als Vereinigung zweier Intervalle schreiben lässt) an, ergibt sich

$$\frac{1}{n}\sum_{m=0}^{n-1} 1_A(\varphi^m(x)) \to b - a$$

für alle $x \in [0,1)$, also der Weylsche Gleichverteilungssatz. Dieser Satz hat interessante Konsequenzen für die Zweierpotenzen 2^m: Sei

$$\vartheta = \log_{10} 2,$$

und für $1 \leq k \leq 9$

$$A = [\log_{10} k, \log_{10}(k+1)]$$

(wobei $\log_{10}$ den Logarithmus zur Basis 10 bezeichnet). Setzt man nun $x = 0$, betrachtet also den eigentlichen Welyschen Gleichverteilungssatz, so ergibt sich

$$\frac{1}{n}\sum_{m=0}^{n-1} 1_A(\varphi^m(0)) \to \log_{10}(\frac{k+1}{k}).$$

Nun ist die erste Ziffer der Zahl 2^m gleich k genau dann, wenn

$$m\vartheta \; mod \; 1 \in A$$

ist. Somit haben wir beispielsweise für $k = 1$ gezeigt, dass der asymptotische Anteil von 2er Potenzen, deren Dezimalentwicklung mit einer 1 beginnt, $\log_{10} 2 = 0,3010\ldots$ ist.

Die Limesverteilung auf den Ziffern $\{1,\ldots,9\}$ heißt oft auch Benford-Verteilung. Raimi hat 1976 Tabellen analysiert und in vielen von ihnen die Benford-Verteilung für die Verteilung der ersten Ziffer beobachtet. Als Beispiel nennt er u. a. die Zweierpotenzen, aber auch die Hausnummern der ersten 342 Menschen in „American Man in Science” oder die Kilowattstunden von 1243 Elektrizitätsrechnungen in Honiara auf den Britischen Salomoninseln.

Beispiel 5.17 *(Bernoulli-Shift) Hier sei $(\Omega, \mathcal{F}, \mathbb{P})$ wie im vorigen Beispiel und die maßtreue Abbildung*

$$\varphi : \Omega \to \Omega$$

gegeben durch

$$\varphi(\omega) = (2\omega) \; mod \; 1.$$

Sei ferner $i_1, \ldots, i_k \in \{0,1\}$ und

$$r = \sum_{m=1}^{k} i_m 2^{-m}$$

das „Muster“, das durch die Ziffernfolge $(i_1, \ldots, i_k)$ dargestellt wird. Schließlich sei

$$X(\omega) = \begin{cases} 1, & \textit{falls}\; r \leq \omega < 1 + 2^{-k} \\ 0, & \textit{sonst} \end{cases} ,$$

d.h. $X(\omega)$ ist genau dann 1, wenn die ersten k Ziffern von ω in Binärdarstellung genau $i_1, \ldots, i_k$ sind. Der Ergodensatz behauptet nun wegen der Trivialität von $\mathcal{J}$, dass

$$\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m(\omega)) \to 2^{-k} \ \mathbb{P}\text{-}f.s.$$

(und in $\mathcal{L}^1$). Mit anderen Worten: Das Muster $(i_1, \ldots, i_k)$ taucht in fast jeder Zahl ω genau so häufig auf, wie man es erwarten würde (und dies gilt für alle Muster endlicher Länge). Dies ist eine verallgemeinerte Fassung der binären Version des Borelschen Gesetzes der normalen Zahlen, die wir im Abschnitt über Gesetze der großen Zahlen kennengelernt haben.

Man mag sich nun fragen, wann man sehen kann, dass eine Abbildung $\varphi : \Omega \to \Omega$ ergodisch ist, bzw. wann die Folge ergodisch ist. Ein inhaltlich leicht fassbares Konzept ist das des „Mischens“.

Definition 5.18 *Eine maßtreue Abbildung*

$$\varphi : \Omega \to \Omega$$

auf einem Maßraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt mischend, falls

$$\lim_{n\to\infty} \mathbb{P}(A \cap \varphi^{-n} B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \tag{48}$$

für je zwei Mengen $A, B \in \mathcal{F}$ gilt. Eine stationäre Folge $(X_n)_n$ von Zufallsvariablen ist mischend, wenn der Shift auf dem Folgenraum mischend ist, mit anderen Worten, falls für je zwei messbare Mengen C und D und $m \in \mathbb{N}$ gilt

$$\lim_{n\to\infty} \mathbb{P}(X_m \in C, X_n \in D) = \mathbb{P}(X_1 \in C)\mathbb{P}(X_1 \in D).$$

Es ist leicht zu sehen, dass eine Abbildung ergodisch ist, wenn sie mischend ist.

Proposition 5.19 *Ist φ mischend, so ist φ auch ergodisch.*

Beweis: Wir müssen zeigen, dass $\mathcal{J}$ trivial ist, wenn φ mischend ist. Sei $A \in \mathcal{J}$. Dann ist $\varphi^{-n}(A) = A$ für alle $n \in \mathbb{N}$. Also

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \lim_{n\to\infty} \mathbb{P}(A \cap \varphi^{-n}(A)) = \mathbb{P}(A)^2,$$

d. h. $\mathbb{P}(A) \in \{0, 1\}$. $\square$

Umgekehrt ist „mischend“ nicht zu weit von Ergodizität entfernt. In der Tat impliziert Ergodizität von φ ja über den Ergodensatz

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}_B(\varphi^m(\omega)) \to \mathbb{P}(B) \ \mathbb{P}\text{-f.s.}$$

Integriert man diese Konvergenz über A, d.h. benutzt man den Satz über majorisierte Konvergenz, erhält man

$$\frac{1}{n}\sum_{m=0}^{n-1}\mathbb{P}(A\cap\varphi^{-m}B)\to\mathbb{P}(A)\mathbb{P}(B),$$

also gilt (2.3) zumindest im Cesaro-Mittel.

Wir wollen das neue Konzept nun anhand von Beispielen noch etwas genauer studieren. Dazu benötigen wir das folgende Theorem. Hierzu sei φ der Shift-Operator auf dem Folgenraum, d. h. $\Omega = \{(\omega_0,\omega_1,\ldots)\}$ und $\varphi(\omega) = (\omega_1,\omega_2,\ldots)$ und

$$X_n(\omega) = \omega_n.$$

Ferner sei

$$\mathcal{F}'_n = \sigma(X_n, X_{n+1},\ldots)$$

und

$$\mathcal{T}_\infty = \bigcap_n \mathcal{F}'_n.$$

Theorem 5.20 *Falls $\mathcal{T}_\infty$ trivial ist, d. h. falls für alle $T\in\mathcal{T}_\infty$, $\mathbb{P}(T)\in\{0,1\}$ gilt, so ist φ mischend und es gilt für alle $A\in\mathcal{F}$*

$$\lim_{n\to\infty}\sup_{B\in\mathcal{F}}|\mathbb{P}(A\cap\varphi^{-n}B)-\mathbb{P}(A)\mathbb{P}(B)| = 0.$$

Beweis: Sei $C = \varphi^{-n}B\in\mathcal{F}'_n$. Dann gilt für $A\in\mathcal{F}_n$

$$\begin{aligned}|\mathbb{P}(A\cap C)-\mathbb{P}(A)\mathbb{P}(C)| &= |\int_C 1\!\!1_A-\mathbb{P}(A)d\mathbb{P}|\\ &= |\int_C \mathbb{P}(A|\mathcal{F}'_n)-\mathbb{P}(A)d\mathbb{P}|\\ &\le \int \mathbb{P}(A|\mathcal{F}'_n)-\mathbb{P}(A)|d\mathbb{P}\to 0.\end{aligned}$$

Hierbei haben wir im zweiten Schritt benutzt, dass $C\in\mathcal{F}'_n$ vorausgesetzt war. Die Konvergenz benutzt einen Satz, den wir erst im Kapitel über Rückwärtsmartingale kennenlernen werden. □

Gilt umgekehrt für $A\in\mathcal{T}_\infty$, dass $\mathbb{P}(A) = 0$ oder $\mathbb{P}(A) = 1$ ist, und setzen wir $A := \varphi^{-n}B_n$, dann folgt

$$|\mathbb{P}(A\cap\varphi^{-n}B_n)-P(A)\mathbb{P}(B)| = \mathbb{P}(A)-\mathbb{P}^2(A).$$

Somit erhalten wir, dass $\mathcal{T}_\infty$ trivial ist genau dann, wenn

$$\lim_{n\to\infty}\sup_{B\in\mathcal{F}}|\mathbb{P}(A\cap\varphi^{-n}B)-\mathbb{P}(A)\mathbb{P}(B)| = 0$$

für alle A gilt. Definitionsgemäß ist φ mischend genau dann, wenn

$$\lim_{n\to\infty} |\mathbb{P}(A \cap \varphi^{-n}(B)) - \mathbb{P}(A)\mathbb{P}(B)| = 0$$

für alle A und B gilt. Schließlich ist φ genau dann ergodisch, wenn

$$\lim_{n\to\infty} |\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{P}(A \cap \varphi^{-m}B) - \mathbb{P}(A)\mathbb{P}(B)| = 0$$

für alle messbaren A und B gilt. Dies zeigt der Trivialität von T_∞ und den Begriffen „mischend“ und „ergodisch“.

Übung 5.21 *Man zeige die unbewiesene Richtung der letzten Behauptung.*

Beispiel 5.22 *(I.I.D. Folgen) Folgen von i.i.d. Zufallsvariablen sind mischend. Dies prüft man entweder direkt nach oder entnimmt es der Tatsache, dass $\mathcal{T}_\infty$ trivial ist.*

Beispiel 5.23 *(Markov-Ketten) Es sei $X_0, X_1, \ldots$ eine irreduzible Markov-Kette über einem endlichen Zustandsraum mit invarianter Verteilung π. Ist die Kette zudem aperiodisch, so ist $\mathcal{T}_\infty$ trivial (das ist ein wenig zu aufwendig, um es hier zu zeigen) und die Folge ist mischend.*

Beispiel 5.24 *(Drehung auf dem Kreis) Sei wieder $\Omega = [0,1)$, $\mathcal{F} = B^1|_{[0,1)}$ und $\mathbb{P} = \lambda^1|_{[0,1)}$. Die maßtreue Abbildung*

$$\varphi : \Omega \to \Omega$$

sei gegeben durch

$$\varphi(\omega) = (\omega + \varphi) \; mod \; 1,$$

wobei $\vartheta \in (0,1)$ irrational ist. Diese Abbildung ist nicht mischend. Um dies einzusehen, beachte man, dass die Menge $T = \{(n\vartheta) \; mod \; 1 | n \in \mathbb{N}\}$ dicht ist in $[0,1)$. Dies zeigt man ähnlich wie in Beispiel 2.16. Da T dicht ist in $[0,1)$, gibt es eine Folge (n_k) mit $n_k \to \infty$, wenn $k \to \infty$ und

$$(n_k\vartheta) \; mod \; 1 \to \frac{1}{2}.$$

Es sei $A = B = [0, \frac{1}{3})$. Ist k hinreichend groß, so gilt

$$A \cap \varphi^{-n_k} B = \emptyset.$$

Somit gilt

$$0 = \lim_{k\to\infty} \mathbb{P}(A \cap \varphi^{-n_k}(B)) \neq \frac{1}{9}$$

und φ ist nicht mischend.

Wir beschließen diesen Abschnitt mit einem Satz, der ebenfalls den Namen „Ergodensatz" trägt, obschon er technisch auf viel einfacheren Ideen beruht. Trotzdem ist er in einer Vielzahl von Situationen sehr nützlich.

Theorem 5.25 *(Subadditiver Ergodensatz) Es sei* $(X_{m,n})$, $0 \leq m < n$ *und* $n \in \mathbb{N}$ *ein Dreiecksschema von Zufallsvariablen, das den folgenden vier Bedingungen genügt:*

a) $X_{0,m} + X_{m,n} \geq X_{0,n}$.

b) Für jedes k *ist die Folge der Zufallsvariablen* $(X_{nk,(n+1)k})_k$ *stationär.*

c) Die Verteilung der $(X_{m,m+k})_k$ *hängt nicht von* m *ab.*

d) $\mathbb{E}X_{0,1}^+ < \infty$ *und für jedes* n *gilt*

$$\mathbb{E}X_{0,n} \geq \gamma_0 n \quad \text{für ein} \quad \gamma_0 > -\infty.$$

Dann gilt:

1. $\lim_{n\to\infty} \frac{\mathbb{E}X_{0,n}}{n} = \inf_m \frac{EX_{0,m}}{m}$ *existiert und ist gleich einem Wert* γ.
2. $X = \lim_{n\to\infty} \frac{X_{0,n}}{n}$ *existiert* $\mathbb{P}$*-f.s. und in* $\mathcal{L}^1$ *und es gilt* $\mathbb{E}X = \gamma$.
3. *Wenn alle stationären Folgen aus Voraussetzung b) ergodisch sind, gilt sogar*

$$X \equiv \gamma \quad \mathbb{P}\text{-f.s.}$$

Wir werden den Subadditiven Ergodensatz zunächst beweisen und dann ein paar Beispiele geben. Der Beweis zerfällt in vier Schritte. Interessanterweise gibt es bei allen existierenden Beweisen in den Schritten 1, 2 und 4 wenig Variation, während sie sich in Schritt 3 unterscheiden.

Beweis: SCHRITT 1: Wir zeigen zunächst 1.

Da es sich bei dem $\mathbb{E}X_{0,n}$ um reelle Zahlen handelt, ist dies im wesentlichen eine Fragestellung über subadditive Folgen. Sei also

$$a_n := \mathbb{E}X_{0,n}.$$

Voraussetzungen a) und c) implizieren

$$a_m + a_{n-m} \geq a_n. \tag{49}$$

Setze $\gamma := \inf_{m\geq 1} \frac{a_m}{m}$. Dann ist offensichtlich

$$\liminf_{n\to\infty} \frac{a_n}{n} \geq \inf_{m\geq 1} \frac{a_m}{m} = \gamma.$$

Wir zeigen nun noch, dass auch $\limsup_{n\to\infty} \frac{a_n}{n} \geq \frac{a_m}{m}$ für jedes $m \in \mathbb{N}$ gilt. Schreiben wir dazu $n = km + \ell$ für ein $0 \leq \ell < m$, so ergibt wiederholte Anwendung von (2.4)

$$a_n \leq k a_m + a_\ell.$$

Division durch n ergibt

$$\frac{a_n}{n} \leq \frac{k_m}{k_{m+\ell}} \frac{a_m}{m} + \frac{a_\ell}{n}.$$

Im Limes $n \to \infty$ verschwindet der zweite Summand der rechten Seite und der erste konvergiert gegen $\frac{a_m}{m}$. Dies zeigt

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_m}{m}$$

für festes m.

Schritt 2. Wir untersuchen nun die Konvergenz der $X_{0,n}$ persönlich. Wiederholte Anwendung von a) ergibt

$$\begin{aligned} X_{0,n} &\leq X_{0,km} + X_{km,n} \\ X_{0,n} &\leq X_{0,(k-1)m} + X_{(k-1)m,km} + X_{km,n} \end{aligned}$$

usf. bis der erste Term der rechten Seite $X_{0,m}$ ist:

$$X_{0,n} \leq X_{0,m_1} + \ldots + X_{(k-1)m,k_m} + X_{km,n}.$$

Division durch $n = k\,m + \ell$ ergibt

$$\frac{X_{0,n}}{n} \leq \left(\frac{k}{km+\ell}\right) \frac{X_{0,m} + \ldots + X_{(k-1)m,km}}{k} + \frac{X_{km,n}}{n}.$$

Voraussetzung b) erlaubt es uns, den Birkhoffschen Ergodensatz anzuwenden, nach dem

$$\frac{X_{0,m} + \ldots + X_{(k-1)m,km}}{k} \underset{k\to\infty}{\longrightarrow} A_m \tag{50}$$

in $\mathcal{L}^1$ und $\mathbb{P}$-f.s. gilt.

Hält man andererseits ℓ fest und wählt $\varepsilon > 0$, so folgt mit Voraussetzung c):

$$\sum_{k=1}^{\infty} \mathbb{P}(X_{km,km+\ell} > n\varepsilon) \leq \sum_{k=1}^{\infty} \mathbb{P}(X_{0,\ell} > k\varepsilon) < \infty, \tag{51}$$

da wir $\mathbb{E}\, X_{0,\ell} < \infty$ vorausgesetzt hatten (dies ergibt sich aus unseren Voraussetzungen, denn nach a) ist

$$X_{0,m}^+ + X_{m,n}^+ \geq X_{0,n}^+$$

und daher auch

$$\mathbb{E}(X_{0,n}) \leq C_n < \infty.)$$

(2.5) und (2.6) implizieren, dass

$$\bar{X} := \limsup_{n\to\infty} \frac{X_{0,n}}{n} \leq \frac{A_m}{m}$$

gilt. Somit erhalten wir für die Erwartungswerte

$$\mathbb{E}\bar{X} \leq \frac{\varepsilon X_{0,m}}{m}$$

für alle m, also im Infimum über m auch

$$\mathbb{E}\bar{X} \leq \gamma.$$

Sind die stationären Folgen unter b) sogar ergodisch, folgt sogar $\bar{X} \leq \gamma$ $\mathbb{P}$-f.s.

Schritt 3. In diesem Schritt geben wir die entsprechende untere Schranke; sei also

$$\underline{X} = \liminf_{n\to\infty} \frac{X_{0,n}}{n}.$$

Wir wollen $\mathbb{E}\bar{X} \geq \gamma$ zeigen. Da

$$\infty > \mathbb{E}X_{0,1} \geq \gamma \geq \gamma_0 > -\infty$$

gilt und wir im vorherigen Schritt $\mathbb{E}\bar{X} \leq \gamma$ gezeigt haben, folgt dann

$$\underline{X} = \bar{X}$$

und dies bedeutet, dass $\lim_{n\to\infty} \frac{X_{0,n}}{n}$ existiert ($\mathbb{P}$-f.s.). Sei nun für $m \in \mathbb{N}$

$$\underline{X}_m = \liminf_{n\to\infty} \frac{X_{m,m+n}}{n}.$$

Wegen a) folgt

$$X_{0,m+n} \leq X_{0,m} + X_{m,m+n}.$$

Dividiert man beide Seiten durch n und schickt $n \to \infty$, so erhält man $\underline{X} \leq \underline{X}_m$ $\mathbb{P}$-f.s. für alle $m \in \mathbb{N}$. Andererseits folgt aus c), dass $\underline{X}$ und $\underline{X}_m$ die gleiche Verteilung haben, also ist

$$\underline{X} = \underline{X}_m \quad \mathbb{P}\text{-f.s.}$$

Sei nun $\varepsilon > 0$ und

$$Z = \varepsilon + (\underline{X} \vee -M).$$

Da nun aus dem 2. Schritt bekannt ist, dass $\underline{X} \leq \bar{X}$ und $\mathbb{E}\bar{X} \leq \gamma < \infty$, so folgt $\varepsilon(|Z|) < \infty$. Wir setzen

$$Y_{m,n} = X_{m,n} - (n-m)Z.$$

Der Vektor der $Y_{m,n}$ genügt den Bedingungen a) – d), denn $X_{m,n}$ genügt diesen Bedingungen ebenso wie $Z_{m,n} = -(n-m)Z$ (Übung).

Konstruktionsgemäß gilt ferner

$$\underline{Y} = \liminf_{n\to\infty} \frac{Y_{0,n}}{n} \leq -\varepsilon \quad \mathbb{P}\text{-f.s.} \tag{52}$$

Es sei für jedes $n \in \mathbb{N}$ T_n die erste Zeit, bei der $Y_{m,m+n}$ nicht mehr positiv ist, also

$$T_m = \min\{n \geq 1 : Y_{m,m+n} \leq 0\}.$$

Nun folgt aus c), dass T_m und T_0 dieselbe Verteilung haben und auch

$$\mathbb{E}(Y_{m,m+1}; T_m > N) = \mathbb{E}(Y_{0,1}; T_0 > N)$$

für jedes N gilt. (2.7) impliziert, dass $\mathbb{P}(T_0 < \infty) = 1$ gilt, also können wir N groß wählen, dass

$$\mathbb{E}(Y_{0,1}; T_0 > N) \leq \varepsilon$$

vorausgesetzt werden kann. Sei

$$S_m = \begin{cases} T_m & \text{auf } \{T_m \leq N\} \\ 1 & \text{auf } \{T_m > N\} \end{cases}.$$

Ferner sei

$$\xi_m = \begin{cases} 0 & \text{auf } \{T_m \leq N\} \\ Y_{m,m+1} & \text{auf } \{T_m > N\} \end{cases}.$$

Da stets $Y_{m,m+T_m} \leq 0$ gilt und $S_m = 1$ und $Y_{m,m+1} > 0$ auf $\{T_m > N\}$ ist, folgt

$$Y_{m,m+S_m} \leq \xi_m \quad \text{und} \quad \xi_m \geq 0.$$

Sei $R_0 = 0$ und $R_k = R_{k-1} + S(R_{k-1})$ für $k \geq 1$. Schließlich sei

$$K = \max\{k : R_k \leq n\}.$$

Aus a) erhalten wir

$$Y_{0,n} \leq Y_{R_0,R_1} + \ldots + Y_{R_{K-1},R_K} + Y_{R_K,n}.$$

Da wir $\xi_m \geq 0$ und $n - R_K \leq N$ abschätzen können, folgt somit

$$Y_{0,n} \leq \sum_{m=0}^{n-1} \xi_m + \sum_{j=1}^{N} |Y_{n-j,n-j+1}|,$$

(wobei wir im letzten Schritt noch einmal a) auf $Y_{R_K,n}$ angewandt haben). Dividiert man beide Seiten durch n und bildet den Limes $n \to \infty$, so erhält man

$$\limsup_{n\to\infty} \frac{\mathbb{E}Y_{0,n}}{n} \leq \mathbb{E}\xi_0 \leq \mathbb{E}(Y_{0,1}, T_0 > N) \leq \varepsilon.$$

Aus der ersten (schon bewiesenen) Behauptung des subadditiven Ergodensatzes und der Definition von $Y_{0,n}$ ergibt sich somit

$$\gamma = \limsup_{n\to\infty} \frac{\mathbb{E}X_{0,n}}{n} \leq 2\varepsilon + \mathbb{E}(\underline{X} \vee -M).$$

Da $\varepsilon > 0$ und $M > 0$ beliebig waren, folgt

$$\mathbb{E}\underline{X} \geq \gamma,$$

was wir in diesem Schritt zeigen wollten.

Schritt 4. Es bleibt noch die $\mathcal{L}^1$-Konvergenz zu zeigen.

Sei dazu A_m wie in Schritt 2 gewählt und $T_m := \frac{A_m}{m}$. Dann ist

$$\mathbb{E}\Gamma_m = \frac{1}{m}\mathbb{E}(X_{0,m}).$$

Setze $\Gamma = \inf_m \Gamma_m$. Unter Ausnutzung des kleinen Tricks

$$|z| = 2z^+ - z$$

erhalten wir

$$\mathbb{E} \mid \frac{1}{n}X_{0,n} - \Gamma \mid = 2\mathbb{E}(\frac{1}{n}X_{0,n} - \Gamma)^+ - \mathbb{E}(\frac{1}{n}X_{0,n} - \Gamma) \leq 2\mathbb{E}(\frac{1}{n}X_{0,n} - \Gamma)^+.$$

Hierbei haben wir im letzten Schritt verwendet, dass

$$\mathbb{E}(X_{0,n}/n) \geq \gamma = \int_m \mathbb{E}\Gamma_m \geq \mathbb{E}\Gamma$$

gilt. Aus

$$(x+y)^+ \leq x^+ + y^+$$

folgt somit weiter

$$\mathbb{E}(\frac{1}{n}X_{0,n} - \Gamma)^+ \leq \mathbb{E}(\frac{1}{n}X_{0,n} - \Gamma_m)^+ + \mathbb{E}(\Gamma_m - \Gamma)^+. \tag{53}$$

Da $\Gamma_m \geq \Gamma$ ist, kann das $+$-Zeichen im letzten Summanden auch fallengelassen werden. Nun konvergiert $\mathbb{E}\Gamma_m$ gegen γ, wenn $m \to \infty$ geht und aus den Schritten 2 und 3 erhalten wir

$$\mathbb{E}\Gamma \geq \mathbb{E}\bar{X} \geq \mathbb{E}\underline{X} \geq \gamma.$$

Somit ist auch schon $\mathbb{E}\Gamma = \gamma$ und $\mathbb{E}(\Gamma_m - \Gamma)$ konvergiert gegen 0, wenn $m \to \infty$ geht. Um auch den ersten Summanden auf der rechten Seite in (2.8) zu beschränken, bemerken wir

$$\mathbb{E}(\frac{1}{n}X_{0,n} - \Gamma_m)^+ \leq \mathbb{E}\left(\frac{X_{0,m} + \ldots + X_{(k-1)m,km}}{k_{m+\ell}} - \Gamma_m\right)^+ + \mathbb{E}\left(\frac{X_{km,n}}{n}\right)^+, \tag{54}$$

wobei wir wiederum a) verwendet haben. Nun ist

$$\mathbb{E}\left(\frac{X_{km,n}}{n}\right)^+ = \mathbb{E}\left(\frac{X_{0,\ell}}{n}\right)^+ \underset{n\to\infty}{\longrightarrow} 0.$$

Auf den ersten Summanden in (2.9) wenden wir wieder den Birkhoffschen Ergodensatz an und erhalten

$$\mathbb{E} \mid \frac{X_{0,m} + \ldots + X_{(k-1)m,km}}{k} - \Gamma_m \mid \to 0.$$

Dies beschließt den Beweis des subadditiven Ergodensatzes. $\square$

Wir schließen das Kapitel mit ein paar Beispielen für Theorem 2.25.

Beispiel 5.26 *(Stationäre Folgen) Sei $\xi_1, \xi_2, \ldots$ eine stationäre Folge von Zufallsvariablen mit*

$$\mathbb{E} \mid \xi_k \mid \; < \infty.$$

Setze

$$X_{m,n} = \xi_{m+1} + \ldots + \xi_n.$$

Dann gilt sogar Additivität, d. h.

$$X_{0,n} = X_{0,m} + X_{m,n}$$

und Theorem 2.25 ist anwendbar.

Beispiel 5.27 *(Perkolation) Wir betrachten das Gitter $\mathbb{Z}^2$ und nennen x und y in $\mathbb{Z}^2$ benachbart, wenn ihr L^1-Abstand = 1 ist, d. h. wenn sie sich in genau einer Koordinate und genau 1 unterscheiden. Sei*

$$\mathcal{N} = \{(x,y) : x, y \in \mathbb{Z}^2, \ x \text{ und } y \text{ sind benachbart}\}.$$

Für (x,y) wählt man eine Folge von positiven, i.i.d. Zufallsvariablen $(T_{x,y})_{(x,y)\in\mathcal{N}}$. Diese stellen die Reisezeiten entlang der Kante (x,y) dar. Zu $u, v \in \mathbb{Z}^2$ sei

$$Z(u,v) = \min \sum_{(x,y)\in\mathcal{P}(u,v)} T_{x,y},$$

wobei $\mathcal{P}(u,v)$ die Menge aller zusammenhängenden Pfade von u nach v ist. Z ist also die kürzeste Reisezeit von u nach v. Man rechnet nach, dass

$$X_{n,m} = Z((0,n),(0,m))$$

den Voraussetzungen des subadditiven Ergodensatzes genügt (Übung). Also existiert der Limes $\frac{X_{0,n}}{n}$.

Beispiel 5.28 *(Die besuchten Punkte einer Irrfahrt) Es sei $\xi_1, \xi_2, \ldots$ eine stationäre Folge von Zufallsvariablen (z. B. eine i.i.d. Folge) und*

$$S_n = \xi_1 + \ldots + \xi_n.$$

Wir setzen $X_{m,n}$ als die Anzahl der zwischen den Zeitpunkten m und n besuchten Punkte, also

$$X_{m,n} = \#\{S_{m+1}, \ldots, S_n\}.$$

Klarerweise gilt

$$X_{0,m} + X_{m,n} \geq X_{0,n}.$$

Da darüberhinaus auch

$$0 \leq X_{0,n} \leq n$$

offensichtlich ist, ist auch Voraussetzung 4 des subadditiven Ergodensatzes erfüllt. Es folgt somit, dass die Anzahl der besuchten Punkte bis zur Zeit n, $X_{0,n}$,

$$\frac{X_{0,n}}{n} \to X$$

$\mathbb{P}$-f.s. und in $\mathcal{L}^1$ genügt. Allerdings weiï¿½ man nicht, was X ist, die Aussage ist daher nur mäßig interessant. In dem Falle, dass die $(\xi_i)_i$ i.i.d. sind, kann man allerdings unter Zuhilfenahme von 1. und 3. in Theorem 2.25 zeigen, dass

$$\frac{X_{0,n}}{n} = \frac{\#\{S_1, \ldots, X_n\}}{n} \to \mathbb{P}(S_n \textit{ kehrt nicht nach 0 zurück})$$

gilt (Übung).

6 Martingale

Im vorherigen Kapitel über den Ergodensatz sind wir erstmals von unseren Unabhängigkeitsvoraussetzungen, die für große Teile der Wahrscheinlichkeitstheorie kennzeichnend sind, abgewichen. In diesem Kapitel wollen wir mit den Martingalen eine besondere Klasse (möglicherweise) abhängiger Prozesse kennenlernen. Ihre Idee geht auf die Modellierung eines fairen Spiels zurück, wie es im einführenden Beispiel vorgestellt werden soll.

Beispiel 6.1 *Es sei $(X_n)_n$ eine Folge von i.i.d. integrierbaren Zufallsvariablen. Sei*

$$S_n = \sum_{i=1}^{n} X_i.$$

Sei ferner

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n) = \sigma(S_1, \ldots, S_n).$$

X_{n+1} ist also von $\mathcal{F}_n$ unabhängig und daher gilt

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

Andererseits ist offenbar

$$\mathbb{E}(X_i|\mathcal{F}_n) = X_i \quad \mathbb{P}\text{-f.s.}$$

für alle $i = 1, \ldots, n$. Gilt dann $\mathbb{E}X_n = 0$ für alle n, so folgt

$$\mathbb{E}(S_{n+1}|\mathcal{F}_n) = S_n.$$

Ist umgekehrt $\mathbb{E}X_n \leq 0$ bzw. $\mathbb{E}X_n \geq 0$, ergibt sich analog

$$\begin{aligned} \mathbb{E}(S_{n+1}|\mathcal{F}_n) &\leq S_n \quad \mathbb{P}\text{-f.s.} \qquad \text{bzw.} \\ \mathbb{E}(S_{n+1}|\mathcal{F}_n) &\geq S_n \quad \mathbb{P}\text{-f.s.} \end{aligned}$$

Sind die X_i z. B. $\pm$ 1-wertig, so lässt sich X_i als die Auszahlung eines Spiels betrachten, bei dem man eine Münze wirft und bei „Kopf" eine Geldeinheit gewinnt und bei "Zahl" eine Geldeinheit verliert. S_n ist dann der Kontostand zur Zeit n. Je nachdem, ob $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1 < 0$ oder $\mathbb{E}X_1 > 0$ ist, ist das Spiel fair, nachteilhaft oder vorteilhaft. Dies ist das Beispiel einer eindimensionalen Irrfahrt. Solche Zufallsvariablen werden in der Theorie stochastischer Prozesse untersucht. Dort untersucht man Fragen dergestalt, ob der Prozess S_n unendlich oft in die Null zurückkehrt oder nicht und was die durchschnittliche Zeit bis zu einer solchen Rückkehr ist. Dies soll hier nun abe nicht analysiert werden – wir verweisen auf Kapitel 4.

Wir werden nun eine Definition geben, die den zentralen Begriff dieses Abschnitts klärt. Beispiel 3.1 ist ein Spezialfall dieses Begriffs. Bevor wir diese Definition geben können, muss noch vorab ein anderer Begriff geklärt werden.

Definition 6.2 *Es sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* I *eine bezüglich „$\leq$" vollständig geordnete Menge. Ferner sei* $(\mathcal{F}_t)_{t\in I}$ *eine Folge von Teil-*σ*-Algebren mit* $\mathcal{F}_t \subseteq \mathcal{F}$ *für alle* $t \in I$. $(\mathcal{F}_t)_{t\in I}$ *heißt Filtration, wenn aus* $s \leq t$ *folgt*

$$\mathcal{F}_s \subseteq \mathcal{F}_t.$$

Definition 6.3 *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $(\mathcal{F}_t)_{t\in I}$ *eine Filtration. Eine Familie* $(X_t)_{t\in I}$ *heißt adaptiert bezüglich* $(\mathcal{F}_t)_{t\in I}$*, wenn* X_t *messbar ist bezüglich* $\mathcal{F}_t$ *für alle* $t \in I$.

Ist $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X_t)_{t\in I}$ (für eine geordnete Menge I) eine Familie von Zufallsvariablen darauf, so gibt es eine natürliche Familie von σ-Algebren $(\mathcal{F}_t)_{t\in I}$, bezüglich derer die Folge $(X_t)_{t\in I}$ adaptiert ist. Man wählt einfach

$$\mathcal{F}_t = \sigma(X_s,\, s \leq t)$$

als die von den $(X_s)_{s\leq t}$ erzeugte σ-Algebra. Diese Familie $(\mathcal{F}_t)_t$ ist offensichtlich eine Filtration. Man nennt sie auch die *natürliche Filtration* (oder kanonische Filtration) für den Prozess $(X_t)_{t\in I}$.

Wir kommen nun zum zentralen Begriff dieses Kapitels.

Definition 6.4 *Gegeben seien ein Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbb{P})$ *und eine Filtration* $(\mathcal{F}_t)_{t\in I}$ *über* $(\Omega, \mathcal{F}, \mathbb{P})$*. Ferner sei* $(X_t)_{t\in I}$ *adaptiert bezüglich* $(\mathcal{F}_t)_{t\in I}$ *mit Werten in* $\mathbb{R}$*. Schließlich seien die* $X_t,\ t \in I$*, alle integrierbar. Man nennt* $(X_t)_{t\in I}$ *ein Supermartingal bezüglich der Filtration* $(\mathcal{F}_t)_{t\in I}$*, wenn für alle* $s, t \in I$ *mit* $s \leq t$ *gilt:*

$$\mathbb{E}(X_t|\mathcal{F}_s) \leq X_s \quad \mathbb{P}\text{-f.s.} \tag{55}$$

d.h.

$$\int_C X_t\, d\mathbb{P} \leq \int_C X_s\, dP \quad \text{für alle } C \in \mathcal{F}_s. \tag{56}$$

(X_t) *heißt Submartingal bezüglich* $(\mathcal{F}_t)$*, wenn* $(-X_t)$ *ein Supermartingal ist. Schließlich nennt man* $(X_t)_t$ *ein Martingal bezüglich* $(\mathcal{F}_t)$*, wenn* (X_t) *sowohl ein Super- als auch ein Submartingal ist.*

Bemerkung 6.5 *1.* (55) *und* (56) *sind in der Tat äquivalent. Da* $\mathcal{F}_s \subseteq \mathcal{F}_t$ *ist, gilt nämlich*

$$\int_C \mathbb{E}(X_t|\mathcal{F}_s) d\mathbb{P} = \int_C X_t\, d\mathbb{P},$$

also impliziert (3.1) Ungleichung (3.2). Andererseits folgt aus (3.2)

$$\int_C X_s - \mathbb{E}(X_t|\mathcal{F}_s) d\mathbb{P} \geq 0 \quad \text{für alle } C \in \mathcal{F}_s.$$

(Man beachte, dass der Integrand $\mathcal{F}_s$*-messbar ist.) Wählt man speziell*

$$C = C_0 := \{X_s - \mathbb{E}(X_t|\mathcal{F}_s) < 0\} \in \mathcal{F}_s,$$

so folgt $\mathbb{P}(C_0) = 0$*, also (3.1).*

2. *Handelt es sich bei $(X_t)_{t\in I}$ um ein Super-, Sub- bzw. Martingal bezüglich der kanonischen Filtration, so sprechen wir schlicht von einem Supermartingal, Submartingal oder Martingal ohne die Filtration besonders anzugeben.*

3. *Da* (55) *bzw.* (56) *für $s = t$ offensichtlich ist, genügt es, die Ungleichungen für $s < t$ zu überprüfen.*

4. *Aus der Glättungseigenschaft der bedingten Erwartung folgt, dass für Supermartingale mit Indexmenge $\mathbb{N}$ gilt*

$$\mathbb{E}(X_{n+p}|\mathcal{F}) \leq X_n \quad \mathbb{P}\text{-}f.s.$$

   für jedes $p \in \mathbb{N}$.

5. *Offenbar haben wir es in Beispiel 3.1 im Falle von $\mathbb{E}X_n = 0$ für alle n mit einem Martingal zu tun. Gilt*

$$\mathbb{E}[X_n] \leq 0 \quad \textit{für alle} \quad n \in \mathbb{N}$$

   so ist S_n ein Supermartingal, für $\mathbb{E}X_n \geq 0$ für alle n ist S_n ein Submartingal.

6. *Setzt man in* (56) *$C = \Omega$, so erhält man, dass für Supermartingale die Folge der Erwartungswerte $(\mathbb{E}X_t)_t$ fallend ist, also*

$$s \leq t \;\Rightarrow\; \mathbb{E}X_s \geq \mathbb{E}X_t,$$

   während dieselbe Folge für Submartingale wachsend ist, also

$$s \leq t \;\Rightarrow\; \mathbb{E}X_s \leq \mathbb{E}X_t$$

   gilt.

7. *Für einelementige Mengen Ω sind Supermartingale, Submartingale bzw. Martingale nichts anderes als monoton fallende, wachsende bzw. konstante Funktionen.*

Wir geben nun noch ein weiteres wichtiges Beispiel:

Beispiel 6.6 *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{F}_t)_{t\in F}$ eine Filtration. Ferner sei*

$$X : \Omega \to \mathbb{R}$$

messbar und integrierbar. Dann definiert

$$X_t := E[X_t|\mathcal{F}_t]$$

ein Martingal bezüglich $(\mathcal{F}_t)$. In der Tat bekommen wir die Adaptiertheit aufgrund der Konstruktion der bedingten Erwartung geschenkt. Andererseits folgt aus der Glättungseigenschaft der bedingten Erwartung für $s \leq t$

$$\mathbb{E}(X_t|\mathcal{F}_s) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_t)|\mathcal{F}_s) = \mathbb{E}(X|\mathcal{F}_s) = X_s.$$

Der folgende Satz erlaubt es aus einer vorgelegten Folge von Supermartingalen bzw. Martingalen neue zu konstruieren.

Satz 6.7 *a) Sind* $(X_t, \mathcal{F}_t)$ *und* $(Y_t, \mathcal{F}_t)$ *Supermartingale, so ist für* $\alpha, \beta \in \mathbb{R}^+$ *auch* $(\alpha X_t + \beta Y_t, \mathcal{F}_t)$ *ein Supermartingal.*

b) Sind $(X_t, \mathcal{F}_t)$ *und* $(Y_t, \mathcal{F}_t)$ *sogar Martingale, so ist auch* $(\alpha X_t + \beta Y_t, \mathcal{F}_t)$ *für alle* $\alpha, \beta \in \mathbb{R}$ *ein Martingal.*

c) Mit $(X_t, \mathcal{F}_t)$ *und* $(Y_t, \mathcal{F}_t)$ *ist auch* $(X_t \wedge Y_t, \mathcal{F}_t)$ *ein Supermartingal.*

d) Für jedes Submartingal $(X_t, \mathcal{F}_t)$ *ist* $(X_t^+, \mathcal{F}_t)$ *ein Submartingal (und somit ist für ein Supermartingal* $(Y_t, \mathcal{F}_t)$ *der Prozess* $(Y_t^-, \mathcal{F}_t)$ *ein Submartingal).*

e) Ist $(X_t, \mathcal{F}_t)$ *ein reellwertiges Submartingal (also gilt dies insbesondere für Martingale) und ist* φ *eine steigende konvexe Funktion, so ist* $(\varphi \circ X_t, \mathcal{F}_t)$ *ein Submartingal, falls* $\varphi(X_t)$ *für alle* t *integrierbar ist.*

Beweis: a) und b) folgen direkt aus (56) in der Definition 3.4 eines Supermartingals.

Für c) wissen wir schon, dass für $s \leq t \in I$ gilt:

$$\mathbb{E}(X_t|\mathcal{F}_s) \leq X_s \quad \text{und} \quad \mathbb{E}(Y_t|\mathcal{F}_s) \leq Y_s$$

$\mathbb{P}$-fast sicher. Dann gilt aber auch

$$\begin{aligned} \mathbb{E}(X_t \wedge Y_t \| f_s) &\leq \mathbb{E}(X_t|\mathcal{F}_s) \leq X_s \quad \text{und} \\ \mathbb{E}(X_t \wedge Y_t|\mathcal{F}_s) &\leq \mathbb{E}(Y_t|\mathcal{F}_s) \leq Y_s \quad \mathbb{P}\text{-f.s.} \end{aligned}$$

also zusammen

$$\mathbb{E}(X_t \wedge Y_t|\mathcal{F}_s) \leq X_s \wedge Y_s \quad \mathbb{P}\text{-f.s.}$$

d) folgt aus c), wenn man für den Prozess Y_t unter c) das Martingal $Y_t = 0$ für alle $t \in I$ wählt. Dann ist nämlich $X_t^- = -(X_t \wedge 0) = -(X_t \wedge Y_t)$.

Schließlich folgt e) aus der Jensenschen Ungleichung. Es gilt nämlich für $t \geq s$

$$\mathbb{E}(\varphi(X_t)|\mathcal{F}_s) \geq \varphi(\mathbb{E}(X_t|\mathcal{F}_s)) \geq \varphi(X_s).$$

Hierbei haben wir im letzten Schritt die Monotonie von φ benutzt. Man beachte, dass in dem Falle, dass $(X_t|\mathcal{F}_t)$ sogar ein Martingal ist, die Isotonie von φ nicht vorausgesetzt werden muss, da dann stets

$$\mathbb{E}(\varphi(X_t)|\mathcal{F}_s) \geq \varphi(\mathbb{E}(X_t|\mathcal{F}_s)) = \varphi(X_s)$$

folgt. □

Korollar 6.8 *Ist* $p \geq 1$ *und* $(X_t, \mathcal{F}_t)$ *ein Martingal mit* $\mathbb{E}|X_t|^p < \infty$ *für alle* $t \in I$, *dann ist* $(|X_t|^p, \mathcal{F}_t)$ *ein Martingal.*

Wir werden uns nun mit dem Zusammenhang zwischen Supermartingalen und Martingalen beschäftigen (bzw. deren Unterschied). Dafür werden wir in der Folge $I = \mathbb{N}$ annehmen und einen auf J. L. Doob zurückgehenden Zerlegungssatz herleiten. Dabei wollen wir eine Folge $(Z_n)_n$ *wachsend* oder einen wachsenden Prozess nennen, wenn für alle $n \in \mathbb{N}$

$$Z_1 = 0 \quad \text{und} \quad Z_n \leq Z_{n+1} \quad \mathbb{P}\text{-f.s.}$$

gilt. Ein Prozess (Z_n) heißt *fallend*, wenn $(-Z_n)_n$ wachsend ist.

Offenbar ist für jedes Martingal $(Y_n, \mathcal{F}_n)$ und jeden adaptierten, wachsenden Prozess $(Z_n, \mathcal{F}_n)$ die Summenfolge $X_n = Y_n + Z_n$ ein Submartingal, denn

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(Y_{n+1}|\mathcal{F}_n) + \mathbb{E}(Z_{n+1}|\mathcal{F}_n) = Y_n + \mathbb{E}(Z_{n+1}|\mathcal{F}_n) \geq X_n.$$

Analog weist man nach, dass es sich bei $X'_n = Y_n - Z_n$ um ein Supermartingal handelt. Insbesondere sind alle wachsenden Prozesse Submartingale, und alle fallenden Prozesse Supermartingale (man setze $Y_n \equiv 0$ für alle $n \in \mathbb{N}$). Der Doobsche Zerlegungssatz besagt nun, dass von der obigen Überlegung auch die Umkehrung gilt: Submartingale sind die Summe aus einem Martingal und einem wachsenden Prozess.

Satz 6.9 *(Doobscher Zerlegungssatz)*
Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ ein beliebiges Submartingal. Dann existieren ein Martingal $(Y_n, \mathcal{F}_n)_n$ und ein adaptierter, wachsender Prozess $(Z_n, \mathcal{F}_n)_n$ mit

$$X_n = Y_n + Z_n.$$

Bemerkung 6.10 *Der Prozess Z_n ist für $n \geq 2$ sogar $\mathcal{F}_{n-1}$-messbar. Solche Prozesse nennt man auch vorhersagbar, da man ihren Wert zum Zeitpunkt n allein aus der Kenntnis des Pfades bis zum Zeitpunkt $n-1$ ableiten kann.*

Beweis von Satz 9: Zunächst definieren wir

$$\begin{aligned} \bar{X}_1 &:= X_1 \quad \text{und} \quad \bar{X}_n := X_n - X_{n-1} \\ \bar{Y}_1 &:= \bar{X}_1 \quad \text{und} \quad \bar{Y}_n := \bar{X}_n - \mathbb{E}(\bar{X}_n|\mathcal{F}_{n-1}) \\ \bar{Z}_1 &:= 0 \quad \text{und} \quad \bar{Z}_n := \bar{X}_n - \bar{Y}_n, \end{aligned}$$

wobei für die rechten Definitionen jeweils $n \geq 2$ vorausgesetzt ist. Es gilt offenbar $X_n = \sum_{i=1}^n \bar{X}_i$ für alle $n \in \mathbb{N}$. Wir setzen

$$Y_n := \sum_{i=1}^n \bar{Y}_i \quad \text{und} \quad Z_n := \sum_{i=1}^n \bar{Z}_i.$$

Können wir nachweisen, dass (Y_n) ein Martingal ist und Z_n ein adaptierter, wachsender Prozess, so sind wir fertig, denn

$$X_n = \sum_{i=1}^n \bar{X}_i = \sum_{i=1}^n (\bar{Y}_i + \bar{Z}_i) = Y_n + Z_n.$$

Ferner ist

$$\bar{Z}_i = \bar{X}_i - \bar{Y}_i = \mathbb{E}(\bar{X}_i|\mathcal{F}_{i-1})$$

$\mathcal{F}_{i-1}$-messbar und somit $\mathcal{F}_{n-1}$-messbar und also $\mathcal{F}_n$-messbar. Daher ist auch Z_n $\mathcal{F}_{n-1}$- und $\mathcal{F}_n$-messbar.

Schließlich ist $(X_n)_n$ ein Submartingal und somit

$$\mathbb{E}(\bar{X}_n|\mathcal{F}_n) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} \geq 0 \quad \mathbb{P}\text{-f.s.}$$

Das bedeutet, dass $\bar{Z}_n = \mathbb{E}(\bar{X}_n|\mathcal{F}_{n-1})$ in der Tat ein wachsender Prozess ist. Schließlich gilt für $n \geq 2$

$$\begin{aligned}
\mathbb{E}(Y_n\|f_{n-1}) &= \sum_{i=1}^{n} \mathbb{E}(\bar{Y}_i|\mathcal{F}_{n-1}) \\
&= \sum_{i=1}^{n} \mathbb{E}(\bar{X}_i|\mathcal{F}_{n-1}) - \sum_{i=2}^{n} \mathbb{E}(\mathbb{E}(\bar{X}_i|\mathcal{F}_{n-1})|\mathcal{F}_{n-1}) \\
&= \sum_{i=i}^{n} \bar{X}_i + \mathbb{E}(\bar{X}_n|\mathcal{F}_{n-1}) - \sum_{i=2}^{n} \mathbb{E}(X_i|\mathcal{F}_{n-1}) \\
&= \sum_{i=1}^{n-1} Y_i + \mathbb{E}(\bar{X}_n|\mathcal{F}_{n-1}) - \mathbb{E}(\bar{X}_n|\mathcal{F}_{n-1}) \\
&= \sum_{i=1}^{n-1} Y_i = Y_{n-1},
\end{aligned}$$

wobei wir die Glättungsregel der bedingten Erwartung verwendet haben. Damit ist (Y_n) ein Martingal, denn definitionsgemäß ist Y_n für jedes n $\mathcal{F}_n$-messbar. □

Korollar 6.11 *Die Zerlegung aus Satz 3.9 ist* $\mathbb{P}$*-f.s. eindeutig.*

Beweis: Seien $X_n = Y_n + Z_n = Y'_n + Z'_n$ zwei Zerlegungen von X_n in Martingale Y_n, Y'_n und wachsende Prozesse Z_n und Z'_n. Dann ist

$$M_n := Y_n - Y'_n = Z'_n - Z_n$$

ein Martingal. Da $M_n = Z'_n - Z_n$ gilt, ist M_n sogar $\mathcal{F}_{n-1}$-messbar für $n \geq 2$. Somit folgt

$$M_n = \mathbb{E}(M_n|\mathcal{F}_{n-1}) = M_{n-1} \quad \mathbb{P}\text{-f.s.}$$

für alle $n \geq 2$. Also ist auch

$$M_n = M_1 = Z'_1 - Z_1 = 0 \quad \mathbb{P}\text{-f.s.}$$

Somit gilt

$$Y_n = Y'_n \quad \text{und} \quad Z_n = Z'_n \quad \mathbb{P}\text{-f.s.}$$

□

Wie das Wort "Martingal" genau in die Sprache der Wahrscheinlichkeitstheorie gekommen ist, ist nicht geklärt. Ursprünglich aus dem Provençalischen stammend bezeichnet es unter anderem einen Teil des Zaumzeugs beim Pferd, das zu starke Kopfbewegungen vermeiden soll oder ein die Takelage des Segelschiffes absicherndes Seil. Von dort muss sich auf eine Weise der Name Martingal für eine Spielstrategie beim Roulette eingebürgert haben, die in jeder Runde die Verdoppelung des zuvor verlorenen Einsatzes vorschreibt. Die genaue Regel wäre die Folgende: Sei (ξ_m) eine Folge von i.i.d. Zufallsvariablen, die die Werte +1 und -1 annehmen. Hierbei steht +1 für einen Gewinn im n-ten Spiel und -1 für einen Verlust im n-ten Spiel. Die Strategie "Martingal" schreibt nun die folgende Folge von Einsätzen vor:

$$H_1 = 1 \quad \text{und} \quad H_n = \begin{cases} 2H_{n-1} & \text{falls } \xi_{n-1} = -1 \\ 1 & \text{falls } \xi_n > 1 \end{cases}.$$

Mit anderen Worten verdoppeln wir unsere Wetten solange, bis wir gewinnen; verlieren wir beispielsweise k mal und gewinnen beim $(k+1)$-ten mal, so ist unser Gesamtgewinn

$$-1 - 2 - 4 - \ldots - 2^{k-1} + 2^k = 1.$$

Dies scheint uns mit einer sicheren Gewinnmöglichkeit auszustatten. Dennoch besagt der folgende Satz, dass es für unvorteilhafte Spiele keine Gewinnstrategien geben kann.

Satz 6.12 *Sei $(X_n)_{n\geq 0}$ ein Supermartingal. Wenn $H_n \geq 0$ vorhersagbar ist und jedes H_n beschränkt ist, so ist auch $(HX)_n := \sum_{m=1}^n H_m(X_m - X_{m-1})$ ein Supermartingal.*

Beweis: Es gilt

$$\begin{aligned} \mathbb{E}((H \cdot X)_{m+1}|\mathcal{F}_n) &= (H \cdot X)_n + \mathbb{E}(H_{n+1}(X_{m+1} - X_n)|\mathcal{F}_n) \\ &= (H \cdot X)_m + H_{m+1}\mathbb{E}(X_{m+1} + X_n|\mathcal{F}_n) \\ &\leq (H \cdot X)_n, \end{aligned}$$

da $\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) \leq 0$ und $H_{n+1} \geq 0$ ist. □

Bemerkung 6.13 *Ein ganz analoges Resultat gilt für Submartingale und Martingale.*

Um den Zusammenhang zwischen Satz 3.12 und der Situation eines Spiels besser zu illustrieren, benötigen wir den Begriff einer Stoppzeit.

Definition 6.14 *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $(\mathcal{F}_t)_{t\in\mathbb{R}}$ *eine Filtration. Eine Zufallsvariable* N *heißt Stoppzeit, falls*

$$\{N = t\} \in \mathcal{F}_t \quad \textit{bzw.} \quad \{N \leq t\} \in \mathcal{F}_t$$

für alle t *gilt.*

Wenn man sich eine Stoppzeit N als die (zufällige) Zeit vorstellt, zu der ein Spieler aufhört zu spielen, dann bedeutet die Bedingung "Stoppzeit" gerade, dass die Entscheidung, zur Zeit n zu stoppen, nur von der zu dieser Zeit verfügbaren Information, also nur von den ersten n Spielausgängen, abhängen darf. Setzen wir

$$H_n = \mathbb{1}_{\{N \geq n\}},$$

dann ist

$$\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1},$$

also ist H_n vorhersagbar. Mit Satz 3.12 folgt dann, dass mit $(X_n)_n$ auch $(H \cdot X)_n = X_{N\wedge n} - X_0$, also der Gewinn zum Ausstiegszeitpunkt, ein Supermartingal ist. Da $(HX)_0 = 0$ ist, ist das Spiel somit mit jeder Stoppzeit als Stoppstrategie unvorteilhaft.

Da die konstante Folge $Y_n = X_0$ ein Submartingal ist und die Summe zweier Submartingale wieder ein Submartingal ist, haben wir en passant auch das folgende gezeigt:

Korollar 6.15 *Wenn* N *eine Stoppzeit ist und* (X_n) *ein Submartingal, dann ist auch* $X_{N\wedge n}$ *ein Submartingal.*

In die gleiche Richtung weisen die folgenden Prinzipien, die sich mit der Frage befassen, was mit einem Martingal geschieht, das wir zu verschiedenen, aufsteigenden Zeiten stoppen. Es ist nicht überraschend, dass die Martingalstruktur erhalten bleibt:

Satz 6.16 *(Optional Sampling)*
Es sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ *ein Supermartingal. Ferner seien* $T_1, \ldots, T_p$ *endlich viele beschränkte Stoppzeiten bezüglich einer Filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *mit* $T_1 \leq T_2 \leq \cdots \leq T_p$. *Dann ist auch* $(X_{T_j}, \mathcal{F}_{F_j})_{j=1,\ldots,p}$ *ein Supermartingal. Ist* $(X_n, \mathcal{F}_n)$ *sogar ein Martingal, so ist auch* $(X_{T_j}, \mathcal{F}_{T_j})_{j=1,\ldots,p}$ *ein Martingal.*

Wir bereiten den Beweis des Optional-Sampling-Theorems mit einem Lemma vor; hierzu muss zunächst die Notation $\mathcal{F}_{T_j}$ aus Satz 3.16 geklärt werden:

Definition 6.17 *Für eine Stoppzeit* T *bezüglich einer Filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *ist die* σ*-Algebra* $\mathcal{F}_T$ *definiert als*

$$\begin{aligned} \mathcal{F}_T : &= \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n \quad \forall\, n \in \mathbb{N}\} \\ &= \{A \leq \Omega : \mathcal{A} \cap \{T \leq n\} \in \mathcal{F}_n \quad \forall\, n \in \mathbb{N}_\infty\}. \end{aligned}$$

Hierbei ist $\mathcal{F}_\infty := \sigma(\bigcup_{n=1}^\infty \mathcal{F}_n)$ *und* $\mathbb{N}_\infty = \mathbb{N} \cup \{+\infty\}$.

Die Gleichheit in Definition 3.17 folgt, da $\{T \leq \infty\} = \Omega$ ist. Nun folgt das angekündigte Lemma:

Lemma 6.18 *Sind S und T Stoppzeiten bezüglich derselben Filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, so gilt*

$$S \leq T \;\Rightarrow\; \mathcal{F}_S \leq \mathcal{F}_T.$$

Ferner setzen wir

$$X_T(\omega) := X_{T(\omega)}(\omega).$$

Nimmt dann T ausschließlich Werte in $\mathbb{N}$ an, so ist X_T $\mathcal{F}_T$-messbar. Gilt nun $\mathbb{P}(T < \infty) = 1$, so gibt es eine ($\mathbb{P}$-f.s.) eindeutige Zufallsvariable X^*, die $\mathcal{F}_T$-messbar ist und für die

$$X^*(\omega) = X_{T(\omega)}(\omega) \quad \forall\, \omega \in \{T < \infty\}$$

gilt.

Beweis: Für die erste Aussage bemerken wir, dass wegen $S \leq T$ auch $\{T \leq n\} \leq \{S \leq n\}$ gilt und damit

$$A \cap \{T \leq n\} = A \cap \{S \leq n\} \wedge \{T \leq n\}$$

für $n \in \mathbb{N}$ und jedes $A \subseteq \Omega$. Dann folgt aber aus $A \in \mathcal{F}_S$ auch $A \in \mathcal{F}_T$, da $A \in \mathcal{F}_\infty$, $A \cap \{S \leq n\} \in \mathcal{F}_n$ und $\{T \leq n\} \in \mathcal{F}_n$ gilt. $\mathcal{F}_T \leq \mathcal{F}_\infty$ ist evident.

Um die $\mathcal{F}_T$-Messbarkeit (einer Variante von) X_T zu beweisen, genügt es, die zweite Behauptung zu betrachten. Sei ω' ein beliebiges Element aus dem Bildraum von X. Wir setzen

$$X^*(\omega) := \begin{cases} X_{T(\omega)}(\omega), & \omega \in \{T < \infty\} \\ \omega', & \omega \in \{T = \infty\} \end{cases}.$$

Für A' aus der Bildmenge von X müssen wir zeigen, dass

$$A := \{X^* \in A'\} \in \mathcal{F}_T$$

gilt. Für jedes $n \in \mathbb{N}$ ist aber

$$A \cap \{T \leq n\} = \bigcup_{i=1}^{n} (A \cap \{T = i\}) = \bigcup_{i=1}^{n} (\{X_i \in A'\} \cap \{T = i\}).$$

Hieraus folgt $A \cap \{T \leq n\} \in \mathcal{F}_n$ für alle $n \in \mathbb{N}$, denn $\{X_i \in A'\} \in \mathcal{F}_i$ für alle $i \in \mathbb{N}$ (da X_i adaptiert ist). Es folgt aber auch

$$A \cap \{T < +\infty\} = \bigcup_{n\in\mathbb{N}} (A \cap \{T \leq n\}) \in \mathcal{F}_\infty$$

wegen $\mathcal{F}_n \leq \mathcal{F}_\infty$ für alle $n \in \mathbb{N}$. Die Menge $A \cap \{T = +\infty\}$ ist entweder leer oder gleich $\{T = \infty\}$, je nachdem, ob $\omega' \notin A'$ gilt oder $\omega' \in A'$. Also ist

$$A \cap \{T = +\infty\} \in \mathcal{F}_\infty.$$

Also ist wegen
$$A = (A \cap \{T < \infty\}) \cup (A \cap \{T = \infty\})$$
auch $A \in \mathcal{F}_T$. Dies beweist das Lemma. $\square$

Nun können wir auch das Optional Sampling Theorem beweisen.

Beweis von Satz 3.16: Nach dem vorhergehenden Lemma induzieren die Stoppzeiten eine Filtration $(\mathcal{F}_{T_j})_{j=1,\dots,p}$. Ebenfach nach dem vorhergehenden Lemma sind die (X_{T_j}) den $(\mathcal{F}_{T_j})$ adaptiert. Nun sind die Stoppzeiten $T_1, \dots, T_p$ nach Voraussetzung beschränkt. Sei also $T_p \le k \in \mathbb{N}$ und somit $T_j \le k$ für alle $j = 1, \dots, p$.

Zunächst zeigen wir, dass die X_{T_j} allesamt integrierbar sind. Es gilt nämlich
$$\mathbb{E}(|X_{T_j}|) = \sum_{i=1}^{k} \int_{\{T_j = i\}} |X_{T_j}| d\mathbb{P} \le \sum_{i<1}^{k} \mathbb{E}(|X_i|) < \infty.$$
Zu zeigen bleibt also noch die Supermartingal-Eigenschaft, d. h.
$$\int_A X_{T_{j+1}} d\mathbb{P} \le \int_A X_{T_j} d\mathbb{P}$$
für alle $j = 1, \dots, p-1$ und alle $A \in \mathcal{F}_{T_j}$. Für $k = 1$ ist nichts zu zeigen, denn dann stimmen alle T_j überein (und sind gleich 1), also sind auch die X_{T_j} und die $\mathcal{F}_{T_j}$ dieselben. Sei also von nun an $k \ge 2$ und – um eine übergroße Anzahl von Indizes zu vermeiden –
$$S := T_j \quad \text{und} \quad T := T_{j+1}$$
gesetzt. Zu zeigen ist somit
$$\int_A X_s d\mathbb{P} \le \int_A X_T d\mathbb{P}$$
für alle $A \in \mathcal{F}_s$.

Der Fall $T - S \le 1$ soll gesondert behandelt werden. Für $A \in \mathcal{F}_s$ gilt dann
$$\int_A (X_s - X_T) d\mathbb{P} = \sum_{i=1}^{k-1} \int_{A_i} (X_s - X_T) d\mathbb{P} = \sum_{i=1}^{k-1} \int_{A_i} (X_i - X_{i+1}) d\mathbb{P},$$
wobei wir hier
$$A_i > A \cap \{S = i\} \cap \{T > s\} = A \cap \{S = i\} \cap \{T > i\} = A \cap \{S = i\} \cap \{T =??\}$$
schreiben ($i = 1, \dots, k-1$). Wenn wir also zeigen können, dass $A_i \in \mathcal{F}_i$ für alle $i = 1, \dots, k-1$ gilt, dann können wir benutzen, dass die Folge (X_n) als Supermartingal vorausgesetzt war und so die gewünschte Ungleichung ableiten. Hierfür beobachten wir, dass $\{T > i\} = \{T \le i\}^c$ in $\mathcal{F}_i$ liegt und dass aus $A \in \mathcal{F}_S$ folgt
$$A \cap \{S = i\} = A \cap \{S \le i\} \backslash \bigcup_{j=1}^{i-1} \{A \cap \{S \le j\}) \in \mathcal{F}_i.$$

Also liegt auch $A_i \in \mathcal{F}_i$.

Der allgemeine Fall folgt aus diesem Spezialfall folgendermaßen: Für jedes $i = 1, \ldots, k$ ist

$$R_i := T \wedge (S + i)$$

auch eine Stoppzeit (nachrechnen!). Nun gilt

$$\int \leq R_1 \leq R_2 \leq \ldots \leq R_k = T$$

sowie $R_1 - S \leq 1$, $T - R_k \leq 1$ und $R_j - R_{j-1} \leq 1$ für alle $j = 2, \ldots, k$. Benutzt man nun

$$\mathcal{F}_S \leq \mathcal{F}_{R_1} \leq \ldots \leq \mathcal{F}_{R_k} \leq \mathcal{F}_T,$$

so folgt aus dem oben Gezeigten (wegen $A \in \mathcal{F}_S$)

$$\int_A X_T d\mathbb{P} = \int_A X_{R_k} d\mathbb{R} \leq \ldots \leq \int_A X_{R_1} d\mathbb{P} \leq \int_A X_S d\mathbb{P},$$

also genau, was wir zeigen wollten.

Ist $(X_n, \mathcal{F}_n)$ sogar ein Martingal, so sind sowohl $(X_n, \mathcal{F}_n)_n$ als auch $(-X_n, \mathcal{F}_n)$ Supermartingale, auf die wir die obigen Überlegungen anwenden können. Also ist auch $(X_{T_j}, \mathcal{F}_{T_j})_j$ ein Martingal. □

Korollar 6.19 *Unter den Voraussetzungen von Satz 3.16 gilt*

$$\mathbb{E}X_1 \geq \mathbb{E}X_{T_1} \geq \ldots \geq \mathbb{E}X_{T_p} \geq \mathbb{E}X_k, \tag{57}$$

wenn $T_p \leq k$ $\mathbb{P}$-f.s. gilt. Ist $(X_n, \mathcal{F}_n)$ ein Martingal, so gilt in (3.3) überall das Gleichheitszeichen.

Beweis: Dies folgt aus Satz 3.16, d. h. die Supermartingal- bzw. Martingaleigenschaft von $(X_{T_j}, \mathcal{F}_{T_j})$ zusammen mit Bemerkung 5.6, die die Monotonie der Erwartungswerte sicherstellt. □

Eine Möglichkeit, Korollar 3.19 zu interpretieren, ist die folgende: Man stelle sich ein Spiel vor, dessen Auszahlungen zur Zeit n gerade durch den Prozess $(X_n)_n$ gegeben sind. Wenn es sich bei $(X_n)_n$ um ein Martingal handelt, so ist die erwartete Auszahlung (im Falle beschränkter Auszahlungen) für alle Stoppzeiten dieselbe. Da man bei Stoppregeln nur solche verwenden darf, die sich auf die zum betreffenden Zeitpunkt vorhandene Information beziehen (und nicht auf die Zukunft), kann man Korollar 3.19 als (erneute) Bestätigung sehen, dass sich in eiem fairen Spiel (Martingal) der Gewinn allein durch Gebrauch der bis zur Zeit n verfügbaren Information nicht erhöhen lässt.

Eine weitere wichtige Konsequenz aus Satz 3.16 ist das folgende Resultat über optionales Stoppen:

Korollar 6.20 *(Optionales Stoppen)*
Es sei $(X_n, \mathcal{F}_n)$ ein Supermartingal bzw. Martingal und T eine Stoppzeit, dann ist auch $(X_{T\wedge n}, \mathcal{F}_{T\wedge n})$ ein Supermartingal bzw. Martingal.

Beweis: Das haben wir schon in Korollar 3.15 festgestellt. □

Interessanterweise kann auf die Beschränktheit der Stoppzeiten in Satz 3.16 nicht einfach verzichtet werden. Dies zeigt das folgende Beispiel.

Beispiel 6.21 *Es bezeichne $(S_n)_n$ das Martingal des Münzwurf-Spiels, genauer sei also $X_1, X_2, \dots$ eine Folge von i.i.d. Zufallsvariablen mit*

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.$$

Sei dann $S_n = \sum_{i=1}^n X_i$ und $\mathcal{F}_n$ die kanonische Filtration. Aus Satz ?? aus der Wahrscheinlichkeitstheorie wissen wir, dass

$$\liminf_{n\to\infty} S_n = -\infty \quad \textit{und} \quad \limsup_{n\to\infty} S_n = -\infty \tag{58}$$

gilt, bei der jeweils $\mathbb{P}$-f.s. Es bezeichne T_k die Ersteintrittszeit in die Menge $\{k\} \in \mathbb{Z}$, also

$$T_k = \inf_{j\in\mathbb{N}_0} \{S_j = k\}.$$

Man prüft leicht nach, dass T_k eine Stoppzeit ist, denn T_k hängt offenbar nur von dem Verhalten von S_j, $j \leq T_k$, ab. Nun gilt

$$\mathbb{P}(T_k < +\infty) = 1,$$

denn wegen (3.4) gibt es eine Menge $\tilde{\Omega} \leq \Omega$ mit vollem Maß, so dass auf $\tilde{\Omega}$

$$-\infty = \liminf S_n(\omega) < \limsup S_n(\omega) = +\infty$$

für alle $\omega \in \tilde{\Omega}$ gilt. Also gibt es zu $\omega \in \tilde{\Omega}$ zwei natürlichen Zahlen $n_1 < n_2$ mit

$$S_{n_1}(\omega) < k < S_{n_2}(\omega).$$

Lässt man nun n die natürlichen Zahlen von n_1 bis n_2 durchlaufen, so muss es offenbar ein $n \in [n_1, n_2] \cap \mathbb{N}$ geben mit $S_n(\omega) = \infty$. Folglich ist $T_k(\omega) \leq n < +\infty$. Definitionsgemäß ist S_{T_k} auf T_k konstant gleich k, also auch

$$\mathbb{E}[S_{T_k}] = k.$$

Ist also $k \neq 0$, so kann die Folge S_1, S_{T_k} kein Martingal sein, dess es gilt $\mathbb{E}S_1 = 0$.

Schließlich setzen wir $\tilde{S}_n := S_{T_k\wedge n}$. Nach dem Optional Stopping Theorem ist $\tilde{S}_k$ ein Martingal. Erinnern wir uns daran, dass Martingale einen konstanten Erwartungswert haben, so erhalten wir wegen $T_k \geq 1$

$$\mathbb{E}(\tilde{S}_n) = \mathbb{E}(\tilde{S}_1) = \mathbb{E}[S_{T_k\wedge 1}] = \mathbb{E}S_1 = 0.$$

Für $\omega \in \{T_k < +\infty\}$ und für $n \geq T_k(\omega)$ ist

$$\tilde{S}_n(\omega) = S_{T_k(\omega)\wedge n}(\omega) = S_{T_k(\omega)}(\omega)$$

konstant gleich k. Das Martingal $(\tilde{S}_n)$ konvergiert daher fast sicher gegen die Zufallsvariable S_{T_k}, d. h.

$$\lim_{n\to\infty} \tilde{S}_n = S_{T_k} = k \quad \mathbb{P}\text{-}f.s.$$

Dies motiviert, was wir in der Folge untersuchen wollen. Dort wollen wir nämlich zeigen, dass eine Folge von Submartingalen unter geeigneten Voraussetzungen konvergiert. Um dies vorzubereiten, starten wir mit einer Beobachtung. Hierzu sei $(X_n)_n$, $n \geq 0$, ein Submartingal. Es sei $a < b$ gegeben und $N_0 := -1$. Für $k \geq 1$ definiere

$$\begin{aligned} N_{2k-1} &:= \inf\{m > N_{2k-2} : X_m \leq a\} \quad \textit{und} \\ N_{2k} &:= \inf\{m > N_{2k-1} : X_m \geq b\}. \end{aligned}$$

Offensichtlich sind die N_j Stoppzeiten und es gilt:

$$\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k} < m-1\} \cap \{N_{2k} \leq m-1\}^c.$$

also ist die Folge der (H_m) definiert als

$$H_m = \begin{cases} 1, & \textit{falls } N_{2k-1} < m \leq N_{2k} \textit{ für ein } k \\ 0, & \textit{sonst} \end{cases}$$

eine vorhersagbare Folge.

Nun ist

$$X(N_{2k-1}) \leq a \quad \text{und} \quad X(N_{2k}) \geq b,$$

was bedeutet, dass X_m sich zwischen N_{2k-1} und N_{2k} von einem Wert kleiner als a zu einem Wert größer als b bewegt. (H_m) lässt sich als eine Spielstrategie interpretieren, die versucht, von dieser Bewegung zu profitieren. Stellt man sich die Werte der (X_n) als die Werte eines Aktienkurses zu verschiedenen Zeiten vor (wobei n beispielsweise Tage indiziert), so besagt die Strategie H_m, dass man eine Aktie kauft, wenn $X_m \leq a$ ist und hält, bis $X_m \geq b$ ist und sie dann verkauft. Also machen wir jedes Mal, wenn ein "Aufwärtslauf" von a nach b vollendet ist, einen Gewinn von mindestens $b - a$. Sei daher

$$U_n := \sup\{k : N_{2k} \leq n\} \tag{59}$$

die Anzahl der Aufwärtsläufe bis zur Zeit n. Als Hilfsschritt für den Martingalkonvergensatz beweisen wir eine Abschätzung für die erwartete Größe von U_n:

Lemma 6.22 *(Upcrossing Inequality)*
Sei (X_m), $m \geq 0$, ein Supermartingal. Dann gilt für (U_n) (definiert wie in (3.5))

$$(b-a)\mathbb{E}U_n \leq \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+.$$

Beweis: Aus der Jensen-Ungleichung folgt, dass auch $(\varphi_n)_n$ ein Submartingal ist. Offensichtlich hat Y_n genau dann einen Aufwärtslauf von 0 nach $b-a$, wenn X_n einen Aufwärtslauf von a nach b hat. Offenbar gilt auch

$$(b-a)U_n \leq (H \cdot Y)_n. \tag{60}$$

Hierzu erinnere man sich an die Definition von

$$(HY)_n := \sum_{m=1}^{n} H_m(Y_m - Y_{n-1}).$$

Dann folgt (3.6), da man bei jedem Aufwärtslauf einen Gewinn von mindestens $(b-a)$ macht und ein letzter unvollendeter Aufwärtslauf auf der linken Seite von (3.6) unberücksichtigt bleibt. Setze

$$K_m := 1 - H_m.$$

Offenbar gilt dann

$$Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n.$$

Aus Satz 3.12 folgt dann, dass

$$\mathbb{E}(K \cdot Y)_n \geq \mathbb{E}(K \cdot Y)_0 = 0,$$

also

$$\mathbb{E}(H \cdot Y)_n \leq \mathbb{E}[Y_n - Y_0],$$

was die Behauptung zeigt. □

Mit Hilfe der Upcrossing Inequality können wir nun einen Konvergenzsatz für Martingale zeigen:

Satz 6.23 *(Martingal-Konvergenzsatz)*
Ist $(X_n, \mathcal{F}_n)_n$ ein Submartingal mit $\mathbb{E}X_n^+ < +\infty$, dann gibt es eine Zufallsvariable X mit $\mathbb{E}|X| < +\infty$, so dass

$$X_n \to X \quad \textit{wenn} \quad n \to \infty \quad \mathbb{P}\text{-}\textit{f.s.}$$

Beweis: Da

$$(X-a)^+ \leq X^+ + |a|,$$

folgt aus der Upcrossing Inequality

$$\mathbb{E}U_n \leq (|a| + \mathbb{E}X_n^+)/(b-a).$$

Lassen wir $n \to \infty$ gehen, so konvergiert U_n gegen die Zufallsvariable U, die die Anzahl der Aufwärtsläufe der Gesamtfolge zählt. Ist also $\mathbb{E}X_n^+ < \infty$, so auch $\mathbb{E}U < +\infty$ und somit auch $U < +\infty$ $\mathbb{P}$-f.s. Da dieser Schluss für alle $a < b \in \mathbb{Q}$ wahr ist, hat das Ereignis

$$\bigcup_{a,b \in \mathbb{Q}} \{\liminf X_n < a < b < \limsup X_n\}$$

Wahrscheinlichkeit Null. Also gilt

$$\limsup X_n = \liminf X_n \quad \mathbb{P}\text{-f.s.},$$

d. h. $\lim X_n$ existiert $\mathbb{P}$-f.s. Nennen wir diesen Limes X. Aus dem Fatouschen Lemma erhalten wir, dass

$$\mathbb{E}X^+ \leq \liminf EX_n^+ < \infty.$$

Also gilt $X < \infty$ $\mathbb{P}$-f.s. Um zu sehen, dass auch $X > -\infty$ gilt, beobachten wir, dass

$$\mathbb{E}X_n^- = \mathbb{E}X_n^+ - \mathbb{E}X_n \leq \mathbb{E}X_n^+ - \mathbb{E}X_0$$

aus der Submartingaleigenschaft der (X_n) folgt. Also bekommen wir mit einer weiteren Anwendung des Fatouschen Lemmas

$$\mathbb{E}X^- \leq \liminf \mathbb{E}X_n^- < \infty,$$

was wir zeigen wollten. □

Als wichtige Konsequenz erhalten wir

Korollar 6.24 *Sei $(X_n)_{n\geq 0}$ ein Supermartingal mit $X_n \geq 0$. Dann gibt es eine Zufallsvariable X mit $\mathbb{E}X \leq \mathbb{E}X_0$, so dass*

$$X_n \to X, \quad \textit{wenn} \quad n \to \infty \quad \mathbb{P}\textit{-f.s.}$$

Beweis: Definiere

$$Y_n = -X_n \leq 0.$$

$(Y_n)_n$ ist ein Submartingal mit $\mathbb{E}Y_n^+ = 0$. Die Behauptung folgt dann aus dem vorangehenden Satz und der Ungleichung aus dem Satz von Fatou. □

Wir geben nun eine Reihe von Beispielen und Gegenbeispielen zum Martingalkovergenzsatz.

Zunächst zeigen wir, dass die Voraussetzungen von Satz 3.23 und Korollar 3.24 nicht ausreichen, um auch L^1-Konvergenz sicherzustellen.

Beispiel 6.25 *Sei S_n die symmetrische Irrfahrt auf $\mathbb{Z}$ mit Start in 1, d. h.*

$$S_0 = 1 \quad \textit{und} \quad S_n = 1 + \sum_{i=1}^{n} \xi_i \quad \textit{für} \quad n \geq 1,$$

wobei die $(\xi_i)_i$ i.i.d. Bernoulli-Zufallsvariablen mit

$$\mathbb{P}(\xi_i = +1) = \mathbb{P}(\xi_i = -1) = \frac{1}{2}$$

sind. Sei N die erste Treffzeit seit der Null, d. h.

$$N = \inf\{n \in \mathbb{N}_0 : S_n = 0\}.$$

Sei ferner

$$X_n := S_{N \wedge n}.$$

Korollar 3.15 (und 3.20) besagt, dass X_n ein (nicht-negatives) Martingal ist. Daher konvergiert X_n nach Satz 3.23 $\mathbb{P}$-f.s. gegen einen Limes X. Offensichtlich kann dieser nur $X \equiv 0$ sein, denn mit Wahrscheinlichkeit 1 wird der Punkt 0 besucht. Andererseits ist

$$\mathbb{E}X_n = \mathbb{E}X_0 = 1 \quad \textit{für alle} \quad n \geq 0.$$

Da aber $\mathbb{E}X = 0$ ist, kann die Konvergenz nicht in L^1 sein.

Als nächstes geben wir ein Beispiel eines Martingals, das zwar in Wahrscheinlichkeit gegen 0 konvergiert aber nicht fast sicher.

Beispiel 6.26 *Sei dazu $X_0 = 0$. Sei ferner*

$$X_k = \begin{cases} 1 & \textit{mit Wahrscheinlichkeit } \frac{1}{2k} \\ -1 & \textit{mit Wahrscheinlichkeit } \frac{1}{2k} \\ 0 & \textit{mit Wahrscheinlichkeit } 1 - \frac{1}{k} \end{cases},$$

falls $X_{k-1} = 0$ ist. Ist $X_{k-1} \neq 0$, so sei

$$X_k = \begin{cases} k \cdot X_{k-1} & \textit{mit Wahrscheinlichkeit } \frac{1}{k} \\ 0 & \textit{mit Wahrscheinlichkeit } 1 - \frac{1}{k} \end{cases}.$$

Wir sehen, dass

$$\mathbb{P}(X_k = 0) = 1 - \frac{1}{k}, \quad k \geq 1,$$

also $X_k \to 0$ in Wahrscheinlichkeit. Andererseits sind zwar nicht die X_k unabhängig (wie man aus der Konstruktion leicht ersieht), wohl aber die Ereignisse

$$A_k := \{X_k = 0\}.$$

Offenbar gilt $\mathbb{P}(A_k^c) = \frac{1}{k}$, also folgt aus dem zweiten Teil des Borel-Cantelli-Lemas, dass

$$\mathbb{P}(X_k = 0 \textit{ für schließlich alle } k) = 0.$$

Da aber X_k nur den Wert 0 annimmt und Werte in $\mathbb{Z}$, kann die Folge (X_k) nicht $\mathbb{P}$-f.s. gegen 0 konvergieren.

Ein Beispiel, das sich mit Hilfe des Martingalkonvergenzsatzes studieren lässt, ist die sogenannte Polya-Urne.

Beispiel 6.27 *In einer Urne seien r rote und g grüne Kugeln. Zu jedem Zeitpunkt $n \in \mathbb{N}$ ziehen wir eine Kugel aus der Urne und legen sie zusammen mit $c \in \mathbb{N}_0$ Kugeln derselben Farben zurück ($c = 0$ ist langweilig). Mit X_n bezeichnen wir den Anteil der grünen Kugeln nach dem n-ten Durchgang. X_n ist ein Martingal, denn für jedes n kann X_n nur endlich viele Werte annehmen und für ?? r_n und g_n gilt*

$$\int_{\{X_n=\frac{g_n}{r_n}\}} d\mathbb{P} = \frac{g_n(g_n+c)}{(r_n+g_n)(g_n+r_n+c)} + \frac{r_n \cdot g_n}{(r_n+g_n)(r_n+g_n+c)} = \frac{g_n}{(r_n+g_n)} = X_n \cdot \mathbb{1}_{\{X_n=\frac{r_n}{g_n}\}}.$$

Da die $X_n \geq 0$ sind, folgt aus Korollar 3.24, dass es eine Zufallsvariable X gibt mit

$$X_n \underset{n\to\infty}{\longrightarrow} X \quad \mathbb{P}\text{-f.s.}$$

Um die Verteilung von x zu berechnen, beobachten wir das Folgende:

a) Die Wahrscheinlichkeit, zunächst m grüne Kugeln zu ziehen und dann l rote ($l = n - m$), berechnet sich als

$$\frac{g}{g+r} \cdot \frac{g+c}{g+r+c} \cdot \ldots \cdot \frac{g+(m-1)c}{g+r+(m-1)c} \cdot \frac{r}{g+r+mc} \cdot \ldots \cdot \frac{r+(l-1)c}{g+r+(n-1)c}.$$

b) Die Wahrscheinlichkeit, in den ersten n Zügen m grüne und $l = n - m$ rote Kugeln zu ziehen, ist dieselbe wie die unter a) berechnete, denn die Nenner bleiben die gleichen und die Zähler werden permutiert.

Betrachten wir den Spezialfalls $g = r = c = 1$. Sei G_n die Anzahl der grünen Kugeln nach Vollendung des n-ten Schrittes. Aus a) und b) oben folgt

$$\mathbb{P}(G_n = m+1) = \binom{n}{m} \frac{m!(n-m)!}{(n+1)!} = \frac{1}{n+1}.$$

Also kann X_n nur gegen ein X konvergieren, das auf (0,1) gleichverteilt ist. Im allgemeinen hat X eine Dichte der Form

$$\frac{\Gamma(\frac{g+r}{c})}{\Gamma(\frac{g}{c})\Gamma(\frac{r}{c})}(1-x)^{g/c-1}x^{r/c-1}$$

auf (0,1). Diese ist auch als Dichte der β-Verteilung zu den Parametern $\frac{g}{c}$ und $\frac{r}{c}$ bekannt.

Eine weitere wichtige und interessante Anwendung des Martingalkonvergensatzes befasst sich mit Radon-Nikodym-Dichten. Sei dazu $(\Omega, \mathcal{F})$ ein messbarer Raum, μ ein endliches Maß und ν ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$. Seien $\mathcal{F}_n$ σ-Algebren mit $\mathcal{F}_n \uparrow \mathcal{F}$, d. h. $\sigma(U\mathcal{F}_n) = \mathcal{F}$. Es seien

$$\mu_n := \mu|\mathcal{F}_n \quad \text{und} \quad \nu_n := \nu|\mathcal{F}_n.$$

Wir nehmen an, dass μ_n stetig ist bzgl. ν_n (in Zeichen $\mu_n \ll \nu_n$) für alle n und schreiben

$$X_n := \frac{d\mu_n}{d\nu_n}$$

für die Dichte.

Lemma 6.28 *X_n definiert auf $(\Omega, \mathcal{F}, \nu)$ ist ein $\mathcal{F}_n$-Martingal.*

Beweis: Definitionsgemäß ist X_n als Dichte von μ_n bzgl. ν_n $\mathcal{F}_n$-messbar. Sei nun $A \in \mathcal{F}_n$. Da X_n $\mathcal{F}_n$-messbar ist und ν_n die Einschränkung von ν auf $\mathcal{F}_n$, ist offenbar

$$\int_A X_n d\nu_n = \int_A X_n d\nu$$

für alle $n \in \mathbb{N}$. Nach Definition von X_n gilt:

$$\int_A X_n d\nu_n = \int_A d\mu_n = \mu_n(A) = \mu(A).$$

Für ein $A \in \mathcal{F}_{m-1} \subseteq \mathcal{F}_m$ können wir dies mit $n = m$ und $n = m - 1$ benutzen, um

$$\int_A X_m d\mu = \mu(A) = \int_A X_{m-1} d\mu.$$

Dies aber sagt nichts anderes als

$$\mathbb{E}(X_m | \mathcal{F}_{m-1}) = X_{m-1}.$$

□

Beispiel 6.29 *Es sei $\Omega = [0, 1]$ und ν das Lebesguemaß auf Ω. Sei*

$$I_{k,n} = [k \cdot 2^{-n}, (k+1)2^{-n})$$

für $0 \leq k < 2^n$ und $n \in \mathbb{N}$. Es sei ferner

$$\mathcal{F}_n := \sigma(I_{k,n} : 0 \leq k < K_n).$$

Dann ist das Martingal

$$X_n = \frac{\mu(I_{k,n})}{\nu(I_{k,n})} = \frac{\mu(I_{k,n})}{\alpha^{-n}}$$

eine Approximation der Lebesgue-Dichte von μ.

Da $X_n = \frac{d\mu_n}{d\nu_n}$ sogar ein nicht-negatives Martingal ist, folgt aus Korollar 3.24, dass es eine Zufallsvariable X_∞ gibt, so dass $X_n \to X_\infty$ ν-f.s. gilt. Um den Limes auf dem ganzen Raum zur Verfügung zu haben, setzen wir

$$X := \limsup X_n.$$

Für diesen Limes X können wur nun nachweisen

Theorem 6.30 *Für μ, ν und X definiert wie oben gilt*

$$\mu(A) = \int_A X \, d\nu + \mu(A \cap \{X = +\infty\})$$

für alle $A \in \mathcal{F}$.

Bemerkung 6.31 *Definieren wir*

$$\bar{\mu}(A) = \int_A X d\nu, \quad A \in \mathcal{F},$$

so ist $\bar{\mu}$ ein Maß, das offensichtlich stetig ist bezüglich ν. Aus Korollar 3.24 folgt nun

$$\nu(X = \infty) = 0.$$

Definieren wir weiter

$$\hat{\mu}(A) = \mu(A \cap \{X = +\infty\}),$$

so ist dieses Maß singulär bezüglich ν. Der Lebesguesche Zerlegungssatz, den wir in der Wahrscheinlichkeitstheorie I schon einmal angesprochen hatten, besagt, dass sich jedes σ-additive Maß μ bezüglich eines jedes σ-endlichen Maßes ν eine Zerlegung

$$\mu = \bar{\mu} + \hat{\mu}$$

besitzt, wobei $\bar{\mu}$ ν-stetig ist und $\hat{\mu}$ singulär ist bzgl. ν. Unsere Wahl von $\bar{\mu}$ und $\hat{\mu}$ gibt also die (ν-f.s. eindeutige) Lebesgue-Zerlegung von μ. Daher ist auch

$$X_\infty := \frac{d\bar{\mu}}{d\nu} \quad \nu\text{-f.s.}$$

Beweis von Satz 3.30: Wir nehmen o.B.d.A. an, dass μ ein Wahrscheinlichkeitsmaß ist. Es sei

$$\rho = \frac{\mu + \nu}{2}$$

und

$$\rho_n = \frac{\mu_n + \nu_n}{2} = \rho_n | \mathcal{F}_n.$$

Setze ferner

$$Y_n = \frac{d\mu_n}{d\rho_n} \quad \text{und} \quad Z_n = \frac{d\nu_n}{d\rho_n}.$$

Offensichtlich sind Y_n und Z_n beide nicht-negativ und

$$Y_n + Z_n = \frac{d\mu_n + d\nu_n}{d\rho_n} = \frac{d(\mu_n + \nu_n)}{d\rho_n} = 2 \cdot \frac{d\rho_n}{d\rho_n} = 2.$$

Somit sind Y_n und Z_n beschränkte, nicht-negative Martingale. Ihre Limiten, die gemäß Korollar 3.24 existieren, nennen wir Y bzw. Z. Wie zu erwarten stand, gilt

$$Y = \frac{d\mu}{d\rho} \quad \text{und} \quad Z = \frac{d\nu}{d\rho}.$$

In der Tat: Erinnern wir uns an den Beweis von Lemma 3.28 und übernehmen die dortigen Bezeichnungen, so erhalten wir für $A \in \mathcal{F}_m \subseteq \mathcal{F}_n$

$$\mu(A) = \int_A Y_n d\rho \to \int_A Y d\rho$$

aufgrund des Satzes von der dominierten Konvergenz. Aus dieser Rechnung ergibt sich offenbar

$$\mu(A) = \int_A Y d\rho \quad \text{für alle} \quad A \in \mathcal{G} = \bigcup_m \mathcal{F}_m. \tag{61}$$

$\mathcal{G}$ ist ein $\cap$-stabiler Erzeuger von $\mathcal{F} = \sigma(\bigcup_m \mathcal{F}_m)$, somit erhalten wir, dass (3.7) für alle $A \in \mathcal{F}$ gilt. Analog zeigt man auch $Z = \frac{d\nu}{d\rho}$.

Nun folgt aber

$$X_n = \frac{d\mu_n}{d\nu_n} = \frac{d\mu_n}{d\rho_n} \cdot \frac{d\rho_n}{d\nu_n} = \frac{Y_n}{Z_n}, \quad \rho\text{-fast sicher},$$

aber auch $X = \frac{Y}{Z}$, ρ-fast sicher. Sei schließlich

$$W = \frac{1}{Z} \mathbb{1}_{\{Z>0\}}.$$

(3.7) impliziert

$$\mu(A) = \int_A Y d\rho = \int_A YW \cdot Z d\rho + \int_A \mathbb{1}_{\{Z=0\}} Y d\rho = \int_A X d\nu + \int_A \mathbb{1}_{\{X=\infty\}} d\mu.$$

Hierbei haben wir benutzt, dass

$$d\nu = Z d\rho,\ Y \cdot W = X\ \nu\text{-f.s.},\ d\mu = Y d\rho \quad \text{und} \quad \{X = \infty\} = \{Z = 0\}\ \mu\text{-f.s.}$$

gilt. Dies zeigt die Behauptung. □

Um die folgende sehr interessante Dichotonie herzuleiten, benötigen wir noch:

Übung 6.32 *Seien μ und ν zwei Wahrscheinlichkeitsmaße, die den Voraussetzungen von Satz 3.30 genügen, und sei*

$$X := \limsup_{n\to\infty} \frac{d\mu_n}{d\nu_n}.$$

Dann gilt:

$$\mu \ll \nu \ \Leftrightarrow\ \mu(X < \infty) = 1 \ \Leftrightarrow\ \mathbb{E}_\nu X = 1$$
$$\mu \perp \nu \ \Leftrightarrow\ \mu(X = +\infty) = 1 \ \Leftrightarrow\ \mathbb{E}_\nu X = 0.$$

(Um zu begreifen, dass die rechten Äquivalenzen nicht ?? sind, bemerke man, dass wir es dort mit dem Erwartungswert bzgl. ν zu tun haben.)

Seien für das Folgende μ und ν nun Wahrscheinlichkeitsmaße auf $(\mathbb{R}^{\mathbb{N}}, B^{\mathbb{N}})$, so dass die Koordinatenabbildungen $Y_n(\omega) = \omega_n$ unabhängig sind. Es seien F_n und G_n die Verteilungsfunktionen von Y_n unter μ bzw. ν, also

$$F_n(x) = \mu(Y_n \leq x) \quad \text{und} \quad G_n(x) = \nu(Y_n \leq x).$$

Wir nehmen an, dass $F_n \ll G_n$ gilt, und wir setzen q_n als die Radon-Nikodym-Dichte $q_n = \frac{dF_n}{dG_n}$. Sei ferner

$$\mathcal{F}_n := \sigma(X_m,\, m \leq n)$$

und

$$\mu_n := \mu|\mathcal{F}_n \quad \text{und} \quad \nu_n := \nu|\mathcal{F}_n.$$

Aufgrund der vorausgesetzten Unabhängigeit der Y_n gilt für die Dichten $X = \frac{d\mu_n}{d\nu_n}$

$$X_n(\omega) = \prod_{m=1}^{n} q_m(\omega_m).$$

Aufgrund der Beziehung zwischen unendlichen Produkten und Reihen, die man schon in der Analysis I lernt, erhalten wir

$$\{X < \infty\} = \{\sum_{m=1}^{\infty} \log q_m \text{ konvergiert}\}.$$

Die rechte Seite dieser Identität unterliegt aber dem Kolmogorovschen 0-1-Gesetz, also gilt

$$\mu(X < \infty) \in \{0, 1\}.$$

Dies impliziert mit Übung 3.32 aber sofort die sogenannte Kahntani-Dichotonie:

Satz 6.33 *In der obigen Situation gilt entweder*

$$\mu \ll \nu \quad \textit{oder} \quad \mu \perp \nu.$$

Wir wollen schließlich Martingale, insbesondere den Martingalkonvergenzsatz, auch verwenden, um für Anwendungszwecke sehr interessante Prozesse zu studieren, sogenannte Verzweigungsprozesse. Die Idee hierbei ist die folgende: Ausgehend von einem Individuum betrachten wir die Populationsgröße einer Bevölkerung, wenn pro Generation jedes Individuum eine zufällige Anzahl Kinder bekommt und dabei selbst stirbt. Das mathematische Modell dahinter sieht so aus: Seien $\xi_i^n,\ i, n \in \mathbb{N}$ i.i.d. Zufallsgrößen mit Werten in $\mathbb{N}_0$. Wir definieren die Folge (Z_n) von Zufallsgrößen vermöge $Z_0 = 1$ und

$$Z_{n+1} = \begin{cases} \xi_1^n + \ldots + \xi_{Z_n}^n, & \text{falls } Z_n > 0 \\ 0 & \text{sonst} \end{cases}.$$

Definition 6.34 *Der so definierte Prozess $(Z_n)_n)$ heißt Galton-Watson-Prozess mit Geburtswahrscheinlichkeiten*

$$p_k := \mathbb{P}(\xi_i^n = k).$$

Wir beweisen nun, dass geeignet normierte Galton-Watson-Prozesse Martingale sind.

Satz 6.35 *Sei*

$$\mathcal{F}_n := \sigma(\xi_i^m : 1 \leq m \leq n,\, i \in \mathbb{N})$$

und $\mu := \mathbb{E}\xi_i^m$. *Dann ist* $\frac{Z_n}{\mu^n}$ *ein Martingal bezüglich* $\mathcal{F}_n$.

Beweis: Offensichtlich ist Z_n messbar bezüglich $\mathcal{F}_n$. Aufgrund der Additivität bedingter Erwartungswerte gilt

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = \sum_{k=1}^{\infty} \mathbb{E}(Z_{n+1}\mathbb{1}_{\{Z_n=k\}}|\mathcal{F}_n). \tag{62}$$

Ist aber $Z_n = k$, so ist dort

$$Z_{n+1} = \xi_1^{m+1} + \ldots + \xi_k^{n+1}.$$

Also folgt für (3.8)

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = \sum_{k=1}^{\infty} \mathbb{E}((\sum_{j=1}^{k} \xi_j^{n+1})\mathbb{1}_{\{Z_n=k\}}|\mathcal{F}_n) = \sum_{k=1}^{\infty} \mathbb{1}_{\{Z_n=k\}}\mathbb{E}(\sum_{j=1}^{k} \xi_j^{n+1}|\mathcal{F}_n)$$

wegen der $\mathcal{F}_n$-Messbarkeit von Z_n. Nun sind aber die ξ_j^{n+1} unabhängig von $\mathcal{F}_n$. Wir erhalten somit

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = \sum_{k=1}^{\infty} \mathbb{1}_{\{Z_n=k\}} \cdot k \cdot \mu = \mu \cdot Z_n.$$

Division durch μ^{n+1} ergibt

$$\mathbb{E}(Z_{n+1}|\mu^{n+1}|\mathcal{F}_n) = \frac{Z_n}{\mu^n},$$

was zu beweisen war. □

Definitionsgemäß ist also $\frac{Z_n}{\mu^n}$ ein nicht-negatives Martingal. Also folgt aus Korollar 3.24, dass $\frac{Z_n}{\mu^n}$ $\mathbb{P}$-f.s. gegen einen Limes konvergiert, wenn $n \to \infty$ geht. Wir unterscheiden drei Fälle.

Satz 6.36 *Ist* $\mu < 1$, *dann ist* $Z_n = 0$ *für hinreichend große* n; *also gilt*

$$\frac{Z_n}{\mu_n} \to 0 \quad \mathbb{P}\text{-}f.s.$$

Bemerkung 6.37 *Satz 3.36 ist intuitiv klar: Wenn eine Population im Durchschnitt weniger als ein Kind pro Individuum gebiert, stirbt sie aus.*

Beweis: Es gilt:

$$\mathbb{E}(\frac{Z_n}{\mu^n}) = \mathbb{E}(Z_0) = 1$$

also $\mathbb{E}(Z_n) = \mu^n$. Nun ist $Z_N > 0$ und $Z_n \geq 1$ gleichbedeutend, also folgt

$$\mathbb{P}(Z_n > 0) \leq \mathbb{E}(Z_n, Z_n > 0) = \mathbb{E}Z_n = \mu^n \to 0$$

exponentiell schnell, da $\mu < 1$ vorausgesetzt war. $\square$

Das nächste Resultat besagt, dass die Schlussfolgerung aus Satz 3.36 für $\mu = 1$ wahr bleibt, wenn wir den Fall einer Geburtenrate $p_1 = 1$ und $p_k = 0\ \forall\, k \neq 1$ ausschließen.

Satz 6.38 *Es sei $\mu = 1$ und $\mathbb{P}(\xi_i^m = 1) < 1$. dann gilt*

$$Z_n = 0 \quad \textit{für alle hinreichend großen } n \quad \mathbb{P}\textit{-f.s.}$$

Beweis: Ist $\mu = 1$, so stimmen $\frac{Z_n}{\mu^n}$ und Z_n überein, also ist auch Z_n ein nicht-negatives Martingal. Da Z_n nur Werte in $\mathbb{N}_0$ annimmt und $Z_n \to Z_\infty$ $\mathbb{P}$-f.s. gilt (für einen geeigneten Limes Z_∞), muss schon $Z_n = Z_\infty$ für hinreichend große n gelten. Ist $\mathbb{P}(\xi_i^m = 1) < 1$ und $k > 0$, dann folgt

$$\mathbb{P}(Z_n = k \text{ für alle } n \geq N) = 0$$

für jedes $N \in \mathbb{N}$. Also muss $Z_n \equiv 0$ sein. $\square$

Somit wissen wir schon, dass für $\mu \leq 1$ (und $\mathbb{P}(\xi_i^m = 1) < 1$) der Prozess existiert. Nun untersuchen wir den Fall $\mu > 1$:

Satz 6.39 *Ist $\mu > 1$, so folgt*

$$\mathbb{P}(Z_n > 0 \textit{ für alle } n \in \mathbb{N}) > 0.$$

Beweis: Für $s \in [0, 1]$ definiere

$$\varphi(s) = \sum_{k=1}^{\infty} p_k s^k,$$

wobei stets noch $p_k = \mathbb{P}(\xi_i^n = k)$ ist. φ heißt auch die *erzeugende Funktion*. Differenzieren unter der Summe ergibt

$$\begin{aligned} \varphi'(s) &= \sum_{k=1}^{\infty} k p_k s^{k-1} \geq 0, \quad s \in [0,1] \quad \text{und} \\ \varphi^n(s) &= \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2} \geq 0, \quad s \in [0,1]. \end{aligned}$$

Also ist φ eine wachsende, konvexe Funktion, und es gilt

$$\lim_{s \uparrow 1} \sum_{k=1}^{\infty} k \cdot p_k = \mu$$

wegen des Abelschen Satzes. φ enthält nun Nützliches über den Prozess $(Z_n)_n$:

a) Ist $\vartheta_m = \mathbb{P}(Z_m = 0)$, dann gilt

$$\vartheta_m = \sum_{k=p}^{\infty} p_k \vartheta_{m-1}^k,$$

denn die rechte Seite beschreibt eine disjunkte Zerlegung des Ereignisses $\{Z_m = 0\}$ in $\{Z_a = k\}$ und alle k Familien sterben in den verbleibtenden $(m-1)$ Schritten.

b) Falls $\varphi'(1) = \mu > 1$, dann gibt es ein eindeutiges $X_0 < 1$ mit

$$\varphi(X_0) = X_0$$

(denn $\varphi(0) = p_0 \neq 0$ und φ ist konvex und steigend (wobei wir $p_0 \neq 0$ o.B.d.A. voraussetzen dürfen, denn sonst ist die Aussage von Satz 3.39 trivial)). Um die Eindeutigkeit des Fixpunkts zu beweisen, bemerken wir, dass aus $\mu > 1$ die Existenz eines $k \geq 2$ mit $p_l \neq 0$ folgt. Also ist sogar

$$\varphi^n(X) > 0 \quad \text{für alle} \quad x < 0,$$

und somit ist φ strikt konvex. Damit folgt, dass für alle $x \in (x_0, 1)$ die Funktion $\varphi(x)$ nicht gleich x sein kann (hierbei ist x_0 ein Fixpunkt).

c) Mit $m \to \infty$ konvergiert ϑ_m (definiert wie unter a) von unten gegen x_0. In der Tat gilt ja $\vartheta_0 = 0$ und $\varphi(x_0) = x_0$. Weiter ist φ wachsend, also ist auch ϑ_m per Induktion wachsend und $\vartheta_m \leq x_0$. Damit existiert

$$\vartheta_\infty := \lim_{m \to \infty} \vartheta_m.$$

Nimmt man in der Gleichung

$$\vartheta_m = \varphi(\vartheta_{m-1}) = \sum_{k=0}^{\infty} p_k \vartheta_{m-1}^k$$

den Limes $m \to \infty$, so ergibt sich

$$\vartheta_\infty = \varphi(\vartheta_\infty),$$

also ist ϑ_∞ ein Fixpunkt von φ. Dieser ist aber nach b) x_0, also ist $\varphi_\infty = x_0$.

Kombiniert man nun a) - c), so sieht man, dass

$$\mathbb{P}(Z_n = 0 \text{ für ein } n) = \lim \vartheta_n = \vartheta_\infty = x_0 < 1$$

gilt, was die Behauptung zeigt. □

Satz 3.39 zeigt mithin, dass $\frac{Z_n}{\mu^n}$ für $\mu > 1$ eine Chance hat, einen Limes, der ungleich null ist, zu besitzen. Ein tieferliegendes Resultat von Kesten und Stigum gibt dafür notwendige und hinreichende Bedingungen:

Satz 6.40 *Definiere*

$$W := \lim_{n\to\infty} \frac{Z_n}{\mu^n}.$$

W ist nicht identisch gleich null genau dann, wenn

$$\sum_{k=1}^{\infty} p_k k \log k < \infty$$

gilt.

Der Beweis dieses Satzes ist für diese Vorlesung zu umfangreich.

Wir werden nun noch eine weitere wichtige Ungleichung für Submartingale herleiten. Wir beginnen mit einer Konsequenz aus dem Optional Sampling Theorem:

Satz 6.41 *Ist X_n ein Submartingal und N eine beschränkte Stoppzeit mit*

$$\mathbb{P}(N \leq k) = 1,$$

so folgt

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_N] \leq \mathbb{E}[X_k]. \tag{63}$$

Beweis: Korollar 3.15 oder 3.20 impliziert, dass die Folge $(X_{N\wedge n})$ ein Submartingal ist. Hieraus ergibt sich:

$$\mathbb{E}[X_0] = \mathbb{E}[X_{N\wedge 0}] \leq \mathbb{E}[X_{N\wedge k}] = \mathbb{E}[X_N].$$

Für die zweite Ungleichung in (3.9) definiere

$$K_n = 1_{\{N<n\}} = \mathbb{1}_{\{N\leq n-1\}}.$$

Offenbar ist k_n $\mathcal{F}_{n-1}$-messbar, also vorhersagbar. Wir können also Satz 3.12 anwenden. Aus diesem erhalten wir, dass $(K \cdot X)_n = X_n - X_{N\wedge n}$ ein Submartingal ist. Daher folgt, dass

$$\mathbb{E}[X_k] - \mathbb{E}[X_N] = \mathbb{E}(K \cdot X)_k \geq \mathbb{E}(K \cdot X)_0 = 0$$

gilt. □

Satz 3.41 ist schon an sich sehr nützlich, besonders aber, um die folgende Ungleichung herzuleiten.

Satz 6.42 *(Doobsche Ungleichung)*
Ist (X_m) ein Submartingal und für $\lambda \in \mathbb{R}$

$$A_\lambda := A := \{\max_{0\leq m\leq n} X_m \geq \lambda\},$$

dann gilt

$$\lambda\mathbb{P}(A) \leq \mathbb{E}X_n 1_A \leq \mathbb{E}X_n^+. \tag{64}$$

Beweis: Wir definieren die Stoppzeit

$$N := \inf\{m : X_m \geq \lambda \text{ oder } m \geq n\}.$$

Auf A gilt offenbar $X_N \geq \lambda$. Also erhalten wir

$$\lambda \mathbb{P}(A) \leq \mathbb{E} X_N|_A \leq \mathbb{E} X_n 1_A.$$

Hierbei folgt die zweite Ungleichung aus Satz 3.41. Dieser impliziert nämlich, dass $\mathbb{E}X_N \leq \mathbb{E}X_n$ gilt. Andererseits sind definitionsgemäß X_N und X_n auf A^c gleich.

Die zweite Ungleichung in (3.10) folgt, da

$$\mathbb{E} X_n 1_A = \mathbb{E} X_n^+ 1_A - \mathbb{E} X_n^- \mathbb{1}_A \leq \mathbb{E} X_n^+ 1_A \leq \mathbb{E} X_n^+$$

gilt. □

Die Doobsche Ungleichung wird uns noch von Nutzen sein, wenn (falls) wir uns der stochastischen Analysis zuwenden. Mit ihrer Hilfe kann man darüber hinaus u. a. auch die Konvergenz von Martingalen in L^p untersuchen. Damit wollen wir uns nicht weiter beschäftigen. Näheres findet man in vielen Büchern, z. B. in dem Buch von Durrett [**?**].

Wir wenden uns abschließend einem Thema zu, dessen Nützlichkeit wir schon im Kapitel über Ergodensätze kennengelernt haben (als wir Satz 3.47 (unten) zitiert haben), den Rückwärtsmartingalen. Wir beginnen mit einer Definition.

Definition 6.43 *$(X_n)_n$ heißt Rückwärtsmartingal, falls die Indexmenge die negativen ganzen Zahlen durchläuft, d. h. falls*

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \quad \textit{für} \quad n \leq -1$$

gilt. Hierbei ist $(\mathcal{F}_n)_{n\in\mathbb{Z}}$ eine wachsende Folge von σ-Algebren.

Die letzte Bedingung besagt mit anderen Worten, dass die Folge der $(\mathcal{F}_n)_n$ fallend ist, wenn $n \to \infty$ geht. Dies macht die Konvergenz von Rückwärtsmartingalen besonders gut zugänglich.

Satz 6.44 *Es sei $(X_n)_{n\in\mathbb{Z}}$ ein Rückwärtsmartingal. Dann existiert $\lim_{n\to\infty} X_n$ $\mathbb{P}$-f.s.*

Beweis: Es sei U_n die Anzahl der Aufwärtsläufe von $X_{-n}, \dots, X_0$. Die Upcrossing Inequality impliziert dann

$$(b-a)\mathbb{E}U_n \leq \mathbb{E}(X_0 - a)^+.$$

Definiert man $U_\infty = \lim_{n\to-\infty} U_n$, so existiert aslo $\mathbb{E}U_\infty$, d. h. $\mathbb{E}U_\infty < \infty$ (hier benötigen wir, dass die σ-Algebren immer größer werden). Nun schließt man wie im Beweis des Martingal-Konvergenzsatzes. Wieder hat das Ereignis

$$\bigcup_{c,b\in\mathbb{Q}} \{\liminf_{n\to-\infty} X_n < a < b < \limsup X_n\}$$

Wahrscheinlichkeit Null und damit ist

$$\liminf_{n\to\infty} X_n = \limsup_{n\to-\infty} X_n \quad \mathbb{P}\text{-f.s.},$$

also existiert $\lim_{n\to\infty} X_n$ $\mathbb{P}$-f.s. □

Bemerkung 6.45 *Auch wenn wir das hier icht zeigen können gilt*

$$\lim_{n\to-\infty} X_n \quad \textit{existiert auch in } L^p.$$

Ist $X_0 \in L^p$, so existiert der Limes auch in L^p.

Satz 6.46 *Definiert man*

$$X_{-\infty} = \lim_{n\to-\infty} X_n \quad \textit{und} \quad \mathcal{F}_{-\infty} = \bigcap_{n\le 0} \mathcal{F}_n,$$

so gilt

$$X_{-\infty} = \mathbb{E}(X_0|\mathcal{F}_{-\infty}).$$

Beweis: Offenbar ist $X_{-\infty}$ messbar bezüglich $\mathcal{F}_{-\infty}$. Ist nun

$$A \in \mathcal{F}_{-\infty} \subseteq \mathcal{F}_n,$$

dann gilt nach Definition eines Rückwärtsmartingals

$$\int_A X_n d\mathbb{P} = \int_A X_0 d\mathbb{P},$$

denn $X_n = \mathbb{E}(X_0|\mathcal{F}_n)$. Aus Bemerkung 3.45 folgt auch, dass

$$X_n 1_A \underset{n\to\infty}{\longrightarrow} X_{-\infty} 1_A$$

gilt. Insgesamt erhält man somit

$$\int_A X_0 d\mathbb{P} = \int_A X_n d\mathbb{P} = \int_A X_{-\infty} d\mathbb{P}.$$

Dies ist aber die Behauptung. □

Satz 3.44 und 3.46 ergeben nun zusammen den schon im Kapitel über Ergodensätze zitierten

Satz 6.47 *Fällt $\mathcal{F}_n$ gegen $\mathcal{F}_{-\infty}$ wenn $n \to -\infty$, d. h. gilt $\mathcal{F}_{-\infty} = \bigcap_{n\le 0} \mathcal{F}_n$, dann folgt*

$$\mathbb{E}[Y|\mathcal{F}_n] \underset{n\to-\infty}{\longrightarrow} \mathbb{E}[Y|\mathcal{F}_{-\infty}]$$

$\mathbb{P}$-f.s. und in L^p.

Beweis: Setzt man $X_n := \mathbb{E}[Y|\mathcal{F}_n]$, so ist die Folge $(X_n)_n$ ein Rückwärtsmartingal. Nach Satz 3.44, Bemerkung 3.45 und Satz 3.46 konvergiert X_n, wenn $n \to -\infty$ geht, $\mathbb{P}$-fast sicher und in L^p. Für den Limes $X_{-\infty}$ gilt

$$X_{-\infty} = \mathbb{E}[X_0|\mathcal{F}_{-\infty}] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_0]|\mathcal{F}_{-\infty}] = \mathbb{E}[Y|\mathcal{F}_{-\infty}].$$

Dies war die Behauptung. □

Auch wenn die Konvergenzsätze für Rückwärtsmartingale nicht sonderlich anspruchsvoll zu beweisen sind, haben sie nette Anwendungen.

Beispiel 6.48 *(Das starke Gesetz der großen Zahlen)*
Seien $\xi_1, \xi_2, \ldots$ i.i.d. Zufallsvariablen mit $\mathbb{E}|\xi_i| < +\infty$. Wir setzen

$$S_n := \xi_1 + \ldots + \xi_n \quad \textit{und} \quad X_{-n} = \frac{S_n}{n}.$$

Schließlich sei

$$\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \ldots) = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \ldots).$$

Um $\mathbb{E}(X_{-n}|\mathcal{F}_{-n-n})$ zu berechnen, was wir zwangsläufig müssen, um die Rückwärtsmartingaleigenschaft von X_{-n} nachzuweisen, bemerke, dass für $j, k \leq m+1$ aus Symmetriegründen

$$\mathbb{E}(\xi_i|\mathcal{F}_{-n-1}) = \mathbb{E}(\xi_k|\mathcal{F}_{-n-1})$$

gelten muss. Also ist

$$\mathbb{E}[\xi_{n+1}|\mathcal{F}_{-n-1}] = \frac{1}{n+1}\sum_{k=1}^{n+1}\mathbb{E}[\xi_k|\mathcal{F}_{-n-1}] = \frac{1}{n+1}\mathbb{E}[S_{n+1}|\mathcal{F}_{-n-1}] = \frac{1}{n+1}S_{n+1}.$$

Hieraus ergibt sich

$$\mathbb{E}[X_{-n}|\mathcal{F}_{-n-1}] = \mathbb{E}\left[\frac{S_{n+1}}{n}|\mathcal{F}_{-n-1}\right] - \mathbb{E}\left[\frac{\xi_{n+1}}{n}|\mathcal{F}_{-n-1}\right] = \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = \frac{S_{n+1}}{n+1} = X_{-n-1}.$$

Wir haben somit nachgewiesen, dass X_{-n} ein Rückwärtsmartingal bildet. Satz 3.44 und 3.46 besagen also, dass

$$\lim_{n\to\infty}\frac{S_n}{n} = \mathbb{E}[X_1|\mathcal{F}_{-\infty}] \quad \mathbb{P}\text{-}f.s.$$

gilt. Nun sind definitionsgemäß die Ereignisse in $\mathcal{F}_{-\infty}$ nicht sensibel für Permutationen endlich vieler der ξ_i, somit permutierbar; also haben $A \in \mathcal{F}_{-\infty}$ entweder Wahrscheinlichkeit 0 oder 1. Daraus folgt

$$\mathbb{E}[X_1|\mathcal{F}_{-\infty}] = \mathbb{E}X_1 \quad \mathbb{P}\text{-}f.s.$$

Dies ist das Gesetz der großen Zahlen.

Beispiel 6.49 *(Ballot Theorem)*
Es sei $X_1, X_2, \ldots, X_n$ eine Folge von i.i.d. Zufallsvariablen mit

$$\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 2) = \frac{1}{2}.$$

Hierbei stellen wir uns vgor, dass wir nach einer Wahl die n Stimmen in der Urne auszählen und interpretieren eine 0 als Stimme für Kandidat A und eine 2 als Stimme für Kandidat B. Das Ereignis

$$G = \{S_j < j \text{ für } 1 \leq j \leq n\}$$

ist somit identisch mit $\{A$ führt vor B während der gesamten Auszählung$\}$. Wir wollen$\mathbb{P}(G)$ berechnen und behaupten

$$\mathbb{P}(G|S_n) = (1 - \frac{S_n}{n})^+, \tag{65}$$

wobei (schon oben)

$$S_k = X_1 + \ldots + X_k$$

gesetzt wurde. Dies sehen wir folgendermaßen ein: $S_n \geq n$ ist nicht zu zeigen (da Kandidat B schlussendlich mehr Stimmen erhält als Kandidat A, muss er auch irgendwann die Führung übernehmen). Sei also $S_n < n$. Die gleichen Rechnungen wie in Beispiel 3.40 zeigen, dass

$$X_{-j} := \frac{S_j}{j}$$

ein Martingal bezüglich

$$\mathcal{F}_{-j} := \sigma(S_j, \ldots, S_n)$$

ist. Sei

$$T := \inf\{k \geq -n : X_k \geq 1\} =: \inf \Gamma$$

und setzen $T = -1$, falls $\Gamma = \emptyset$. Sei nun wie oben

$$G = \{S_j < j \text{ für alle } 1 \leq j \leq n\} \leq \{T = -1\}.$$

Nun ist $X_T = 0$ auf G, denn $S_1 < 1$ impliziert, dass $S_1 = 0$ war und $X_T \geq 1$ auf G^c. Somit folgt

$$\mathbb{P}(G|S_n) \leq \mathbb{E}[X_Z|\mathcal{F}_{-n}] = X_{-n} = \frac{S_n}{n}.$$

Genaueres Hinsehen ergibt sogar, dass $X_T = 1$ auf G^c gilt, also

$$\mathbb{P}(G|S_n) = \mathbb{E}[X_T|\mathcal{F}_{-n}] = \frac{S_n}{n}.$$

Mit anderen Worten haben wir gezeigt

$$\mathbb{P}[G|A \text{ bekommt } r \text{ Stimmen}] = (1 - \frac{2r}{n})^+.$$

Dies ist ein klassisches Resultat.

7 Der Satz von Donsker

Im Kapitel über den Zentralen Grenzwertsatz haben wir die Normalverteilung als eines der zentralen Objekte der Wahrscheinlickeitstheorie kennengelernt. Ihre Rolle stammt daher, dass unter gewissen (relativ milden) Bedingungen geeignet skalierte Mittelwerte beliebiger unabhängiger Zufallsvariablen eine Standardnormalverteilung als Verteilungslimes haben. In diesem Kapitel wollen wir ein Objekt untersuchen, das eine ähnlich zentrale Role spielt, wenn wir statt zufälliger (eindimensionaler) Variablen zufällige Funktionen betrachten, die Brownsche Bewegung. Wir wollen ein Prinzip beweisen, dass uns sagt, dass Folgen geeignet normierter (und gewählter) zufälliger Funktionen gegen die Brownsche Bewegung konvergieren. Dies ist das Analogon zum Zentralen Grenzwertsatz auf Funktionenebene. Ähnlich wie beim Satz von de Moivre-Laplace, der ja ein Spezialfall des zentralen Grenzwertsatzes ist, werden wir dabei einen speziellen Prozess besonders im Blick haben. Der folgende Prozess ist dabei gewissermaßen das Analogon zum Münzwurf:

Definition 7.1 *Es sei $X_1, X_2, \dots$ eine Folge von i.i.d. Zufallsvariablen mit*

$$\mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.$$

Der Prozess $S_n = \sum_{i=1}^n X_i$ (in Abhängigkeit von n) heißt die (symmetrische) eindimensionale Irrfahrt auf $\mathbb{Z}$.

Unsere Hauptfrage in diesem Kapitel wird sich damit beschäftigen, wie man S_n geeignet skalieren kann, so dass man einen nicht-trivialen Limes erhält und welche interessanten Konsequenzen man daraus auch für den Limesprozess ableiten kann. Bevor wir dies tun wollen wir noch ein interessantes Resultat für $(S_n)_n$ herleiten, das wir für die Zwecke dieses Abschnitts eigentlich nicht benötigen, das aber interessant genug ist, um hier betrachtet zu werden. Es beschäftigt sich mit der Frage, ob und mit welcher Wahrscheilichkeit $(S_n)_n$ an seinen Ausgangspunkt 0 zurückkehrt. Für $y \in \mathbb{Z}$ sei dazu $T_y^0 = 0$ ("die nullte Rückkehrzeit") und

$$T_y^k := \inf\{n > T_y^{k-1} : X_n = y\}.$$

T_y^k ist also die k-te "Rückkehrzeit" zum Punkt y, wobei der Startpunkt (die Zeit 0) nicht mitgezählt wird. Wir setzen $T_y := T_j^1$ und

$$\rho_{xy} := \mathbb{P}_x(T_y < +\infty),$$

also die Wahrscheilichkeit, bei Start in x jemals nach y zu gelangen. Dann gilt

Satz 7.2 *Es gilt*

$$\mathbb{P}(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}. \tag{66}$$

Bemerkung 7.3 *(4.1) sollte intuitiv völlig klar sein; es bedeutet nur, dass man bei Start in x für k Besuche in y zunächst von x nach y laufen muss und dann $(k-1)$ mal nach y zurückkehren. Der formale Beweis ist etwas aufwendiger, folgt aber dieser Idee.*

Beweis: Für $k = 1$ ist nichts zu beweisen. Sei $k \geq 2$. Wir beweisen (4.1) induktiv. Auf $\{T_y^k < +\infty\}$ ist auch $T_y^{k-1} < +\infty$. Also gibt es ein endliches $n \in \mathbb{N}$ mit $G_y^{k-1} = n$. Die Prozesse $(S_k)_{k \leq n}$ und $(S_k)_{k>n}$ sind aber unabhängig. Somit folgt:

$$\begin{aligned}
\mathbb{P}_x(T_y^k < +\infty) &= \sum_n \mathbb{P}_x(T_y^{k-1} = n,\ ((S_k)_{k \geq n} \text{ läuft von } y \text{ nach } y) \\
&= \sum_n \mathbb{P}_x(T_y^{k-1} = n)\mathbb{P}_y((S_k)_{k \geq n} \text{ läuft von } y \text{ nach } y) \\
&= \mathbb{P}_x(T_y^{k-1} < +\infty) \cdot \mathbb{P}_y((S_k)_{k \geq 0} \text{ läuft nach } y) \\
&\underset{IV}{=} fxy\, fyy.
\end{aligned}$$

□

Wichtig ist nun die folgende Definition:

Definition 7.4 *Ein Zustand $y \in \mathbb{Z}$ heißt rekurrent, falls $\rho_{yy} = 1$ gilt. Er heißt transient, falls $\rho_{yy} < 1$ ist.*

Mit anderen Worten heißt ein Zustand rekurrent, wenn er mit Wahrscheinlichkeit 1 wieder besucht wird, wenn man in ihm startet (dann wird er auch unendlich oft besucht) und sonst heißt er transient. Wir wollen nun herausfinden, ob und welche Zustände in $\mathbb{Z}$ rekurrent bzw. transient sind. Dabei ist die folgende Größe von Wichtigkeit: Für $y \in \mathbb{Z}$ definiere

$$N(y) := \sum_{n=1}^{\infty} \mathbb{1}_{\{S_n = y\}}.$$

$N(\cdot)$ zählt also die Anzahl der Besuche in einem Punkt. Die entscheidende Hilfestellung bietet uns nun

Satz 7.5 *Ein Zustand $y \in \mathbb{Z}$ ist rekurrent genau dann, wenn*

$$\mathbb{E}_y N(y) = +\infty \tag{67}$$

gilt.

Beweis: Es gilt

$$\mathbb{E}_y N(y) = \sum_{k=1}^{\infty} \mathbb{P}_y(N(y) \geq k) = \sum_{k=1}^{\infty} \mathbb{P}_y(T_y^k < +\infty) = \sum_{k=1}^{\infty} \rho_{yy}^k = \frac{\rho_{yy}}{1 - \rho_{yy}}.$$

Dies ist genau dann eine endliche Größe, wenn $\rho_{yy} < 1$ ist, also wenn y transient ist.

□

Bevor wir nun mit Hilfe von Satz 4.5 die Rekurrenz von $y \in \mathbb{Z}$ untersuchen, beweisen wir noch einen Satz, der besagt, dass wir die Rekurrenz nur für die 0 untersuchen müssen, weil Rekurrenz ansteckend ist.

Satz 7.6 *Wenn $x \in \mathbb{Z}$ rekurrent ist, dann ist auch $y \in \mathbb{Z}$ rekurrent und es gilt $\rho_{yx} = 1$.*

Beweis: Wir zeigen zuerst, dass $\rho_{yx} = 1$ ist, da sonst $\rho_{xx} < 1$ gelte. Sei

$$K := \inf\{k : p^k(x,y) > 0\},$$

wobei $p^k(x,y)$ die k-Schritt-Übergangswahrscheinlichkeit von x nach y bezeichnet. Also gibt es eine Folge $y_1, \dots, y_{K-1}$, so dass

$$p(x,y)p(y_1,y_2)\dots p(y_{K-1},y) > 0$$

gilt. Da K minimal ist, sind die y_i alle verschieden und insbesondere verschieden von y. Wäre nun $\rho_{yx} < 1$, so folgte

$$\mathbb{P}_X(T_X = +\infty) \geq p(x,y_1)p(y_1,y_2)\dots p(y_{k-1},y)(1-\rho_{yx}) > 0,$$

also ein Widerspruch. Somit ist $\rho_{yx} = 1$. Um die Rekurrenz von y zu beweisen, bemerken wir, dass aus $\rho_{yx} > 0$ die Existenz eines L mit $p^L(y,x) > 0$ folgt. Nun gilt

$$p^{L+n+K}(y,y) \geq p^L(y,x)p^n(x,x)p^K(x,y), \tag{68}$$

da die rechte Seite nur die Möglichkeit darstellt, in $L+n+K$ Schritten von y nach y zu laufen; es könnte noch andere geben. Summiert man (4.3) über n, ergibt sich

$$\sum_{n=1}^{\infty} p^{L+n+K}(y,y) \geq p^L(y,x)p^K(x,y)\sum_{n=1}^{\infty} p^n(x,x) = \infty,$$

also ist y rekurrent. □

Wir werden nun die Rekurrenz der symmetrischen eindimensionalen Irrfahrt beweisen.

Satz 7.7 *Für (S_n) ist jeder Zustand $y \in \mathbb{Z}$ rekurrent.*

Bemerkung 7.8 *Betrachten wir an Stelle der symmetrischen Irrfahrt die asymmetrische Irrfahrt auf $\mathbb{Z}$, also einen Prozess $S_n = \sum_{i=1}^n X_i$ mit i.i.d. X_i mit*

$$\mathbb{P}(X_i = +1) = 1 - \mathbb{P}(X_i = -1) = p \neq \frac{1}{2},$$

so folgt aus dem starken Gesetz der großen Zahlen, dass

$$\frac{S_n}{n} \to \mathbb{E}X_1 = 2p - 1 \neq 0 \quad \mathbb{P}\text{-f.s.}$$

gilt, also S_n entweder fast sicher jede Grenze überschreitet oder fast sicher unter jede Grenze fällt. Also ist in diesem Fall jeder Zustand transient.

Beweis von Satz 4.7: Offenbar müssen wir nur die Rekkurenz des Zustands 0 beweisen (wegen Satz 4.6). Nun ist

$$\mathbb{E}_0 N(0) = \sum_{n=1}^{\infty} p^n(0,0)$$

und $p^{2n+1}(0,0) = 0$. Weiter ist

$$p^{2n}(0,0) = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \sim \frac{1}{\sqrt{\pi n}}$$

(letzteres folgt mit Hilfe der Stirlingschen Formel; wir haben dies schon bei der Herleitung des lokalen Grenzwertsatzes in der Stochastikvorlesung gesehen). Somit gilt

$$p^{2n}(0,0) \leq C \cdot \frac{1}{\sqrt{\pi n}}$$

für ein $C > 1$ und alle hinreichend großen n. Da bekanntlich

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} < \infty$$

konvergent ist, folgt die Behauptung aus Satz 4.5. $\square$

Ohne Beweis sei noch bemerkt, dass das eigentlich Spannende an Satz 4.7 ist, dass man die Rekurrenz ebenfalls für die 2-dimensionale symmetrische Irrfahrt beweisen kan, also jenen Prozess, der pro Zeiteinheit stets zu einem seiner 4 Nachbarn springt und dies mit gleicher Wahrscheinlichkeit. Andererseits ist die Rekurrenz falsch für die symmetrische Irrfahrt im $\mathbb{Z}^d$, $d \geq 3$, diese Prozesse sind transient, laufen also nach ∞ davon. Diese Dichotonie steht im unmittelbaren Zusammenhang mit der Divergenz bzw. Konvergenz der Reihen $\sum(\frac{1}{\sqrt{\pi n}})^d$ für $d \leq 2$ bzw. $d \geq 3$.

Wir wollen nun für ein richtig skaliertes S_n (und viele andere Summenprozesse) einen Grenzwertsatz herleiten, der dem zentralen Grenzwertsatz für i.i.d.-Folgen von Zufallsvariablen entspricht. Die erste Frage dabei ist, wie wir S_n dabei skalieren müssen. Dabei ist zweierlei zu beachten: Zum einen sollte die räumliche Skala durch $\sqrt{n}$ gestaucht werden – wir sollten also $\frac{S_n}{\sqrt{n}}$ betrachten – dies legt schon der zentrale Grenzwertsatz nahe. Andererseits (das ist vielleicht weniger offensichtlich) sollten wir auch etwas an der "Zeitskala", also dem unteren Index n, verändern. Anderenfalls haben wir für jedes n eine andere "Zeitebene" (nämlich das Intervall $[0, n]$), auf der es definiert ist und die einzige gemeinsame Ebene, das Intervall $[0, \infty)$ nicht kompakt ist. Dies macht es schwerer, Limespunkte zu finden. Wir wollen stattdessen S_{nt}, $0 \leq t \leq 1$ betrachten. Dies hat allerdings den Nachteil, dass $n \cdot t$ für die allermeisten t keine ganze Zahl ist und dadurch S_{nt} nicht definiert ist. Um dies zu überwinden, betrachten wir den Prozess

$$Y_n := \frac{S_{[nt]}}{\sqrt{n}} + \frac{nt - [nt]}{\sqrt{n}} X_{[nt]+1}, \quad 0 \leq t \leq 1,$$

wobei $[x]$ die Gaußklammer von x bezeichnet. Y_n ist offenbar für jedes n eine (zufällige) stetige Funktion auf $[0,1]$, also ein Element in $C([0,1])$. Diesen Raum wollen wir fortan mit der Sipremumsmetrik

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

versehen. Um den gewünschten Konvergenzsatz beweisen zu können, müssen wir eine ganze Menge Hilfsmittel bereitstellen. Das erste ist

Satz 7.9 *$C([0,1], d)$ ist ein vollständiger, separabler metrischer Raum.*

Beweis: Die Vollständigkeit wird meist in der Analysis bewiesen. Sie soll hier nicht gezeigt werden. Der interessierte Leser findet sie beispielsweise im Analysisbuch von Ahmann/Escher [**?**] oder im [**?**]. Die Separabilität folgt sofort aus dem Weierstraßschen Approximationssatz: Die Polynome liegen dircht in $C([0,1], d)$, die Polynome mit rationalen ?? liegen wiederum dicht in den Polynomen. Die Polynome mit rationalen Koeffizienten sind aber nach dem Cantor-Verfahren abzählbar. □

Da wir auf $C([0,1], d)$ Wahrscheinlichkeitstheorie betreiben wollen, benötigen wir eine σ-Algebra dort. Wir wählen (kanonisch) die Borelsche σ-Algebra, also die, die von den offenen Mengen erzeugt wird. Diese nennen wir $\mathcal{B}_C$. Interessanterweise wird $\mathcal{B}_C$ schon von den endlich-dimensionalen Projektionen erzeugt. Genauer sei für $m \in \mathbb{N}$ und $0 \le t_1 < \ldots < t_m \le 1$

$$\begin{aligned} \pi_{t_1,\ldots,t_m} : C([0,1]) &\to \mathbb{R}^m \\ f &\mapsto (f(t_1), \ldots, f(t_m)). \end{aligned}$$

Dann gilt:

Lemma 7.10 *Es gilt*

$$\mathcal{B}_C := \sigma(\pi_t^{-1}(\mathcal{B}),\, t \in [0,1]).$$

Beweis: Mit $\mathcal{B}' := \sigma(\pi_t^{-1}(\mathcal{B}),\, t \in [0,1])$ wollen wir $\mathcal{B}_C = \mathcal{B}'$ zeigen. Da π_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,

$$\mathcal{B}_\varepsilon(f) := \bigcap_{t\in[0,1]\cap\mathbb{Q}} \{g \in C[0,1] : |g(t) - f(t)| \le \varepsilon\} = \bigcap_{t\in[0,1]\cap\mathbb{Q}} \pi_t^{-1}(\mathcal{B}_\varepsilon(f(t))) \in \mathcal{B}'.$$

Da $C[0,1]$ separabel ist, ist jede offene Menge abzählbare Vereinigung von derartigen Kugeln, also in $\mathcal{B}'$. □

Wir wollen uns nun daran machen, das Limesobjekt zu beschreiben. Letztlich werden wir die Konvergenz der Verteilung von $Y_n(\omega, t)$ und ähnlicher Prozesse untersuchen und hoffen, dass es zu diesen Verteilungen ein schwaches Limesmaß gibt. Um zu wissen, welchen Funktionen dieses Limesmaß Masse geben sollte, ist es aber sicherlich hilfreich, die Y_n noch einmal zu betrachten. Offensichtlich sind alle $Y_n(t)$ stetige Funktionen in t. Es ist daher nicht unvernünftig zu vermuten, dass dies der Limes auch ist. Darüber hinaus konvergiert für jedes $0 \leq s < t \leq 1$ die Folge

$$(Y_n(\omega, t) - Y_n(\omega, s))_n$$

in Verteilung gegen die Normalverteilung $\mathcal{N}(0, t - s)$. Für $0 \leq t_1 < t_2 < t_3 \leq 1$ sind die $(Y_n|t_3) - Y_n(t_2))$ und $(Y_n(t_2) - Y_n(t_1))$ sogar unabhängig (und dies lässt sich auf mehrere Zeitpunkte verallgemeinern). Wir definieren daher

Definition 7.11 *Ein Maß auf* $(C[0, 1], \mathcal{B}_C)$ *mit*

- $\mu(C[0, 1]) = 1$,
- $\mu\pi^{-1}_{t_1,\dots,t_m}$ *ist die* m*-dimensionale Normalverteilung mit Erwartungswert 0 in Kovarianzmatrix* $(min(t_i, t_j))_{i,j}$ *(für alle* $m \in \mathbb{N}$*)*

heißt Wiener Maß auf $(C[0, 1], \mathcal{B}_C)$*.*

Bemerkung 7.12 *Der zweite Punkt entspricht gerade den unabhängigen normalverteilten Zuwächsen: In der Tat ist* $X \sim \mathcal{N}(0, s)$*- und* $Y \sim \mathcal{N}(0, t)$*-verteilt und sind* X *und* Y *unabhängig, so ist* $X + Y \sim \mathcal{N}(0, s + t)$*-verteilt und* $Cov(X, X + Y) = (X) = s$*.*

Unter μ hat ein Pfad $\mathcal{B}_t$ offenbar die Eigenschaften:

1. $\mu(B_t$ ist stetig $\forall\ t) = 1$.
2. $\mu(B_t \leq \alpha) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{-\frac{s^2}{2t}} ds$.
3. Für $0 = t_0 < t_1 < \ldots < t_m \leq 1$ und $\alpha_1, \ldots, \alpha_m \in \mathbb{R},\ m \in \mathbb{N}$, gilt $\mu(B_{t_i} - B_{t_1} \leq \alpha_i,\ i = 1, \ldots, m) = \prod_{i=1}^m \mu(B_{t_i} - B_{t_{i-1}} \leq \alpha_i)$.

Definition 7.13 *Eine Funktion* B *mit den obigen Eigenschaften heißt eindimensionale Brownsche Bewegung in 0. Genauer muss hierbei stets der Grundraum und das Maß spezifiziert werden. Ist dieser* $C[0, 1]$ *und* μ *das Wienermaß, so spricht man auch von der Standardbrownschen Bewegung. Nun haben wir schon eine ganze Menge über das Wiener-Maß und die Brownsche Bewegung gesammelt. Freilich bedeutet dies nicht, dass es diese Prozesse auch geben muss.*

Satz 7.14 *Es gibt genau ein Wahrscheinlichkeitsmaß μ auf $C([0,1],\mathcal{B}_C)$, das den Anforderungen an ein Wienermaß genügt.*

Die Eindeutigkeit ist hierbei wenig problematisch: Es gibt für jedes m die m-dimensionale Normalverteilung, und diese ist eindeutig. Da die endlich-dimensionalen Projektionen die σ-Algebra $\mathcal{B}_C$ erzeugen, kann es nur höchstens ein Wahrscheinlichkeitsmaß geben, das den Ansprüchen an ein Wiener-Maß genügt. Die Existenz ist da schon problematischer. Es gibt verschiedene Möglichkeiten, diese zu beweisen. Eine ist der Gebrauch des Kolmogorovschen Fortsetzungssatzes im Anhang. Man zeigt die Konsistenz der Familien $(\pi_{t_1,\ldots,t_m})$ für alle $t_1,\ldots,t_m$ und alle $m \in \mathbb{N}$ und ist fertig. Eine andere, konkrete Möglichkeit besteht darin, einfach eine Brownsche Bewegung zu konstruieren, indem man die Stützstellen der richtigen Größe interpoliert. Diese Konstruktion geht auf Paul Levy zurück und findet sich z. B. im Buch von Karatzas und Shreve "Stochastic Calculus" [**?**] (in der gleichnamigen Vorlesung folgen wir diesem Beweis).

Wir folgen einem dritten Weg, indem wir das Wienermaß μ einfach als Limesmaß einer geeigneten Folge von Verteilungen nachweisen. Dies ist der Satz von Donsker. Bevor wir ihn endgültig formulieren und beweisen wollen, m+ssen wir uns noch ein paar Werkzeuge verschaffen. Das erste ist dazu geeignet, sich kompakte Teilmengen der Menge der Maße auf $(C[0,1],\mathcal{B}_C)$ zu verschaffen. Dazu sei noch einmal an den Begriff der Straffheit erinnert:

Definition 7.15 *(Erinnerung) Eine Folge $(\mu_n)_n$ von Wahrscheinlichkeitsmaßen auf einem vollständigen, separablen, metrischen Raum Ω (der in der früheren Definition $\mathbb{R}$ war) heißt straff, falls es für jedes $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subseteq \Omega$ gibt mit*

$$\mu_n(K_\varepsilon^c) \le \varepsilon$$

für alle $n \in \mathbb{N}$. Interessanterweise benutzt dieser Begriff nicht nur den Begriff der Kompaktheit, sondern er sagt auch etwas über ihn aus.

Satz 7.16 *(Prohorov) Es sei S ein separabler, metrischer Raum und $(\mu_n)_{n\in\mathbb{N}}$ eine straffe Folge von Wahrscheinlichkeitsmaßen auf $(S,\mathcal{B}_S)$ (wobei $\mathcal{B}_S$ die Borelsche σ-Algebra über S bezeichnet). Dann hat $(\mu_n)_n$ eine schwach konvergente Teilfolge. Ist S vollständig, so gibt es ein $\mu \in S$ und eine Teilfolge $(\mu_{n_j})_j$ von $(\mu_n)_n$, so dass*

$$\mu_{n_j} \ \Rightarrow \ \mu$$

gilt (wobei "$\Rightarrow$" schwache Konvergenz anzeigt). Darüber hinaus gilt auch die Umkehrung: Ist S vollständig und separabel und konvergiert $(\mu_n)_n$ schwach, so ist $(\mu_n)_n$ auch straff.

Der Beweis von Satz 4.16 bedarf einiger Vorbereitung. Zunächst verwenden wir eine Variante des Rieszschen Darstellungssatzes, den wir in anderer Form schom im Beweis des Satzes von Radon und Nikodym kennengelernt hatten. Dazu bringen wir zunächst die folgende

Definition 7.17 *Es sei S ein metrischer Raum. 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. Hierbei bezeichnet $C(S)$ die Menge der stetigen Funktionen auf S.*

Die von uns benötigte Variante des Rieszschen Darstellungssatzes lässt sich nun wie folgt formulieren:

Satz 7.18 *Darstellungssatz von Riesz)*
Es sei S ein kompakter metrischer Raum. Dann existiert zu jedem normierten, nicht-negativen linearen Funktional $\Lambda : C(S) \to \mathbb{R}$ ein eindeutig bestimmtes Wahrscheinlichkeitsmaß μ auf $(S, \mathcal{B}_S)$ mit

$$\Lambda(f) = \int f \, d\mu \quad \forall\, f \in C(S). \tag{69}$$

Jedes Wahrscheinlichkeitsmaß auf $(S, \mathcal{B}_S)$ bestimmt vermöge (4.4) ein normiertes, nicht-negatives lineares Funktional auf $C(S)$.

Der Beweis von Satz 4.18 verläuft im wesentlichen analog zum Beweis des Rieszschen Darstellungssatzes, den wir schon im Kapitel über den Satz von Radon-Nikodym kennengelernt haben. Wir werden ihn daher hier weglassen.

Mit Hilfe von Satz 4.18 können wir nun den einen (wesentlichen) Teil des Satzes von Prohorov für kompakte Grundmengen herleiten. Dabei können wir auf die Straffheit verzichten. Hierzu bezeichnen wir für einen metrischen Raum S mit

$$\mathcal{M}^1(S) := \{\mu : \mu \text{ ist ein Wahrscheinlichkeitsmaß auf}; (S, \mathcal{B}_S)\} \tag{70}$$

die Menge aller Wahrscheinlichkeitsmaße auf S. Auf $\mathcal{M}^1(S)$ haben wir durch die schwache Konvergenz einen Konvergenzbegriff eingeführt und damit eine Topologie induziert (diese ist sogar metrisierbar). Der folgende Satz stellt nun fest, dass $\mathcal{M}^1(S)$ die Kompaktheit von S in dieser Topologie erbt.

Satz 7.19 *Ist S ein kompakter, metrischer Raum, so ist $\mathcal{M}^1(S)$ schwach folgenkompakt.*

Bemerkung 7.20 *Da $\mathcal{M}^1(S)$ wie schon bemerkt schwach metrisierbar ist, ist $\mathcal{M}^1(S)$ auch schwach kompakt.*

Beweis: 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 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

$$|\int f\, d\mu_{n_k} - \int f\, d\mu_{n_m} \le \int f_j\, d\mu_{n_k} - \int f_j\, d\mu_{n_m}| + \int \underbrace{|f - f_j|}_{\le\varepsilon}\, d\mu_{n_k} + \int \underbrace{|f - f_j|}_{\le\varepsilon}\, d\mu_{n_m}.$$

Der erste Summand konvergiert gegen Null für $k, m \to \infty$, also

$$\lim_{k,m\to\infty} |\int f\, d\mu_{n_k} - \int f\, d\mu_{n_m}| = 0,$$

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 Λ ein nicht-negatives lineares Funktional auf $C(S)$ mit $\Lambda(1) = 1$, also existiert nach dem Rieszschen Darstellungssatz ein $\mu \in M^1(S)$ mit

$$\Lambda(f) = \int f\, d\mu \quad \forall\, f \in C(S),$$

womit die schwache Konvergenz von $(\mu_{n_k})_k$ gegen μ folgt. □

Das Hauptproblem ist nun, dass wir im Satz von Prohorov nicht die Kompaktheit des Raumes sondern nur die Straffheit der Folge vorausgesetzt haben. Aus der Definition der Straffheit ist relativ klar, dass alle Folgeglieder μ_n "bis auf ein $\varepsilon > 0$" die gleiche kompakte Menge ?? Um diese Information für uns nutzbar zu machen benötigen wir aber noch zwei topologische Aussagen:

Satz 7.21 *(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

Satz 7.22 *(Tychonov)*
$(K^{\mathbb{N}}, \bar{d})$ ist kompakt.

Bemerkung 7.23 *Wie man aus der Topologie weiß, gilt sogar noch mehr: Beliebige Produkte kompakter Mengen sind wieder kompakt. Die obige Form des Satzes von Tychonov ist einfacher zu zeigen. Der Beweis geht wieder auf das Diagonalverfahren zurück.*

Beweis von Satz 4.22: Es sei $(x_n)_n$ eine Folge in $\mathbb{K}^{\mathbb{N}}$,

$$x_n = (x_n^{(i)})_{i\in\mathbb{N}}.$$

Aufgrund der Kompaktheit der Menge K lässt sich eine Teilfolge $(x_{n_1,m})$ finden, so dass die erste Koordinate $(x^{(1)}_{n_1,m})_m$ konvergent ist; von dieser Folge gibt es wieder eine Teilfolge $(x_{n_2,m})$, so dass $(x^{(2)}_{n_2,m})$ konvergiert etc. Man überlegt sich schnell, dass dann die Diagonalfolge $(x_{n_m,m})$ insgesamt konvergiert, denn die ersten N Koordinaten können für jedes N beliebig klein gemacht werden und die hinteren bekommt man mit der Konvergenz der geometrischen Reihe und der Beschränktheit der Metrik klein. $\square$

Der Erfolg, den wir mit den Sätzen 4.21 und 4.22 verbuchen können ist der, dass wir nun wissen, dass ein separabler metrischer Raum homöomorph ist zu einer Teilmenge eines kompakten, metrischen Raumes, die damit selbst Präkompakt ist. Somit können wir Satz 4.19 ins Spiel bringen.

Beweis von Satz 4.16: 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 Satz 4.19 hat $(\tilde{\mu}_n)_n$ eine konvergente Teilfolge $(\tilde{\mu}_{n_k})_k$, die schwach gegen ein Wahrscheinlichkeitsmaß ν 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 Portmanteau-Theorem gilt

$$\limsup_{n\to\infty} \tilde{\mu}_{n_k}(K_r) \leq \nu(K_r),\ r \in \mathbb{N}.$$

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 $A_i =_i \cap S_i,\ i \in \mathbb{N}$, 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

$$\begin{aligned}\mu(\bigcup_i A_i) &= \nu(\bigcup B_i) = \nu(\bigcup_i (B_i \cap E_0)) \\ &= \sum_i \nu(B_i \cap E_0) = \sum_i \nu(B_i) = \sum_i \mu(A_i).\end{aligned}$$

Also ist μ ein Wahrscheinlichkeitsmaß 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 $\tilde{\mu}_{n_k} \xrightarrow{w} \tilde{\mu}$ gilt (wobei $\tilde{\mu}_{n_k} \xrightarrow{w} \tilde{\mu}$ und $\tilde{\mu}_{n_k} \Rightarrow \tilde{\mu}$ gleichbedeutend sind), folgt

$$\limsup_{k\to\infty} \mu_{n_k}(C) = \limsup_{k\to\infty} \tilde{\mu}_{n_k}(D) \leq \tilde{\mu}(D) = \mu(C).$$

Das Portmanteau-Theorem 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 $BV_{n1}, B_{n2}, \ldots$ 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

$$\mu_i(\bigcup_{j=1}^{k_n} B_{nj}) > 1 - \delta, \quad i \in \mathbb{N}.$$

Angenommen, dies stimmt nicht. Also existiert ein $\delta_0 > 0$ und folgen $i_1 < i_2 < \ldots$ und $k_1 y k_2 \ldots$ mit

$$\mu_{i_m}(\bigcup_{j=1}^{k_m} B_{nj}) \leq 1 - \delta_0 \quad \text{für} \quad m = 1, 2, \ldots.$$

Es gilt $\bigcup_{j=1}^{k_r} B_{nj} \subset \bigcup_{j=1}^{k_m} B_{nj}$ für $m \geq r$, also

$$\mu_{i_m}(\bigcup_{j=1}^{k_r} B_{nj}) \leq \mu_{i_m}(\bigcup_{j=1}^{k_m} B_{nj}) \leq 1 - \delta_0$$

für $m \geq r$.

Da $\mu_{i_m} \xrightarrow{w} \mu$ und $\bigcup_{j=1}^{k_r} B_{nj}$ offen, sagt das Portmanteau-Theorem

$$\mu(\bigcup_{j=1}^{k_r} B_{nj}) \leq \liminf_{m\to\infty} \mu_{i_m}(\bigcup_{j=1}^{k_r} B_{nj}) \leq 1 - \delta_0.$$

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

$$\mu_i(\bigcup_{j=1}^{k_n} B_{nj}) > 1 - \frac{\varepsilon}{2^n}, \quad n \in \mathbb{N}.$$

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. □

Damit ist der Satz von Prohorov bewiesen. Er ist für unsere Zwecke sehr nützlich, denn schließlich wollen wir ja die Verteilungskonvergenz einer (richtig skalierten) Folge von Irrfahrten nachweisen (dies wird auch für viele, viele andere Prozesse gelten). Der Satz von Prohorov sagt uns nun, dass die Folge von Verteilungen zumindest konvergente Teilfolgen hat, wenn wir ihre Straffheit nachweisen können. Dies ist allerdings zunächst ein etwas unhandliches Kriterium, da wir hierzu große kompakte Mengen des Grundraums, also in unserem Fall des $C([0,1], d)$, wobei d die durch die Supremumsnorm induzierte Metrik ist, kennen müssen. Was also sind die kompakten Teilmengen von $C([0,1], d)$? Auskunft darüber gibt der Satz von Arcela und Asodi; hierzu zunächst folgende Definition:

Definition 7.24 *Es sei*

$$f : [0,1] \to \mathbb{R}$$

eine stetige Funktion und $\delta > 0$. Das Stetigkeitsmodul $\omega_\delta(f)$ ist definiert als

$$\omega_\delta(f) := \sup\{|f(s) - f(t)|,\ s, t \in [0,1], |s-t| \leq \delta\}.$$

Bemerkung 7.25 *Es gilt natürlich*

$$|\omega_\delta(f) - \omega_\delta(g)| \leq 2d(f,g),$$

also ist $\omega_\delta(\cdot)$ für jedes $\delta > 0$ stetig. Ferner folgt aus der gleichmäpigen Stetigkeit einer Funktion $f \in C[0,1]$, dass

$$\lim_{\delta \downarrow 0} \omega_\delta(f) = 0$$

gilt.

Nun der angekündigte

Satz 7.26 *(Satz von Arzela-Ascoli)*
Eine Teilmenge $A \subset C[0,1]$ hat genau dann kompakten Abschluss, wenn

(i) $\sup\{|f(0)|,\ f \in A\} < \infty$ *ist und*

(ii) $\lim_{\delta\to 0} \sup_{f\in A} \omega_\delta(f) = 0$

gelten.

Wir bereiten den Beweis durch ein Kriterium für Kompaktheit von Mengen in metrischen Räumen vor.

Satz 7.27 *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 $x_0, \ldots, x_m \in K$ 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 Raumes ist genau dann kompakt, wenn sie folgenkompakt ist, Analysis I), also bestitz $(x_j)_j$ 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 $(s_j)_j$, die ganz in einem Ball mit

Radius $1/2$ enthalten ist, etc. Also gibt es zu jedem $n \in \mathbb{N}$ eine Teilfolge $(x_{n+1,j})_j$, 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 *Diagonalforgenprinzip* verwendet. Wir wählen aus einer Folge gemäß einer Vorschrift sukzessive Teilfolgen aus und bilden dann die Diagonalfolge, indem wir von der n-ten Teilfolge das n-te Glied auswählen. 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 Limes-Eigenschaften wie jede der Teilfolgen $(x_{n,j})_j$ besitzt.

Da $A \subset X$ totalbeschränkt ist genau dann, 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 Arzela-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 (i) zur Folge hat.

Seien nun (i) und (ii) gegeben. Wähle k groß genug, so dass $\sup_{f\in A} \omega_{1/k}(f)$ endlich ist. Da

$$|f(t)| \leq |f(0)| + \sum_{i=1}^{k} |f\left(\frac{i}{k}t\right) - f\left(\frac{i-1}{k}t\right)|,$$

folgt mit (i)

$$\sup_{t\in[0,1]} \sup_{f\in A} |f(t)| < \infty. \tag{71}$$

Wir zeigen nun, dass aus (ii) und (4.6) 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 $H := \{\frac{u}{v}\alpha, u = 0, \pm1, \pm2, \ldots, \pm v, v \in \mathbb{N}\}$ 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 2ε-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

$$|f(\frac{i}{k}) - g(\frac{i}{k})| < \varepsilon, \quad i = 0, \dots, k. \tag{72}$$

Da $\omega_{1/k}(f) < \varepsilon$ und g linear in jedem Teilintervall $[\frac{i-1}{k}, \frac{i}{k}]$ ist, folgt aus (4.7) $d(f,g) < 2\varepsilon$. Dies war zu zeigen. □

Satz 4.26, der Satz von Arzela-Ascoli, lässt sich nun schnell in ein Kriterium für die Straffheit einer Folge von Wahrscheinlichkeitsmaßen auf $C[0,1]$ übersetzen:

Satz 7.28 *Eine Folge $(\nu_n)_n$ von Wahrscheinlichkeitsmaßen auf $(C, \mathcal{B}_C)$ ist genau dann straff, wenn*

$$\begin{aligned} \lim_{a \nearrow \infty} \sup_n \nu_n(\{f : |f(0)| > a\}) &= 0 \quad \textit{und} && (73)\\ \lim_{a \searrow 0} \limsup_{n\to\infty} \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) &= 0 \quad \textit{für alle} \quad \varepsilon > 0 && (74) \end{aligned}$$

gelten.

Nach obiger Bemerkung ist $\{f : \omega_\delta(f) \geq \varepsilon\} \in \mathcal{B}_C$. Die Bedingungen (4.8) und (4.9) in Satz 4.28 können wie folgt übersetzt werden:

$$\begin{aligned} \forall\, \eta > 0\, \exists\, a > 0\, \forall\, n \in \mathbb{N} : &\quad \nu_n(\{f(0)| > a\}) \leq \eta, && (75)\\ \forall\, \varepsilon > 0,\ \eta > 0\, \exists\, \delta > 0\, \exists\, n_0 \in \mathbb{N}\, \forall\, n \geq n_0 : &\quad \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta. && (76) \end{aligned}$$

Diese Bedingungen werden wir zu einem späteren Zeitpunkt für die von uns untersuchte Folge untersuchen.

Wir haben nun die wesentlichen Hilfsmittel bereitgestellt, um den Satz von Donsker zu beweisen. Wir wollen uns nun an die Formulierung machen. Hierbei sei $X_1, X_2, \dots$ eine Folge von i.i.d. Zufallsvariablen mit Werten in $\mathbb{R}$, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Wir nehmen an, dass

$$\mathbb{E}X_1 = 0 \quad \text{und} \quad \mathbb{V}X_1 =: \sigma^2 \in (0, \infty)$$

gilt. Wir werden fortan sogar ohne Einschränkung $\mathbb{V}X_1 = 1$ annehmen. Mit diesen allgemeineren Zufallsvariablen $(X_n)_n$ lässt sich nun auch eine "verallgemeinerte Irrfahrt" beschreiben:

$$S_0 = 0 \quad \text{und} \quad S_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N}.$$

Wir führen wieder die schon eingangs erwähnte Stauchung von Raum und Zeit durch, indem wir den Raum jeweils auf das Intervall [0,1] zurückstauchen und die Zeit so skalieren, dass der zentrale Grenzwertsatz anwendbar ist. Wir definieren also

$$Y_n(\omega, t) := \frac{S_{[nt]}(\omega)}{\sqrt{n}} + \frac{nt - [nt]}{\sqrt{n}} X_{[nt]+1}(\omega) \tag{77}$$

für $0 \leq t \leq 1$ (wieder sei $[x]$ die Gaußklammer von x). Für jedes $\omega \in \Omega$ erhalten wir mit $Y_n(\omega, \cdot)$ also eine Funktion aus $C[0,1]$, also eine zufällige stetige Funktion. Nun ist konstruktionsgemäß für jedes feste $t \in [0,1]$ die Abbildung $Y_n(\cdot, t)$ eine $\mathcal{F} - B$-messbare Abbildung. Lemma 4.10 liefert dann, dass $Y_n(\cdot) := (Y_n(\cdot, t))_t$ auch $\mathcal{F} - \mathcal{B}_C$-messbar, also eine $\mathcal{F} - \mathcal{B}_C$-messbare Zufallsvariable ist. Wie schon mehrfach bemerkt, wollen wir uns im Satz von Donsker mit der Verteilungskonvergenz der Zufallsvariablen Y_n und somit um die schwache Konvergenz der Maße

$$\mu_n := \mathbb{P}^{Y_n}$$

kümmern. Da nun die Abbildung

$$\begin{array}{rcl} \pi_{t_1,\ldots,t_m} : C & \to & \mathbb{R}^m \\ f & \mapsto & (f(t_1), \ldots, f(t_m)) \end{array}$$

für jedes m-Tupel $0 \leq t_1 < t_2 < \ldots < t_m \leq 1$ stetig ist, ist für die Verteilungskonvergenz der Y_n zumindest die schwache Konvergenz der endlich-dimensionalen Verteilungen $\mu_n \circ \pi^{-1}_{t_1,\ldots,t_m}$ notwendig. Dies klärt der folgende

Satz 7.29 *Für jedes $m \in \mathbb{N}$ und $0 \leq t_1 < \ldots < t_m \leq 1$ konvergiert die Folge $\mu_n \circ \pi^{-1}_{t_1,\ldots,t_m}$ schwach auf $(\mathbb{R}^m, \mathcal{B}^m)$ gegen die m-dimensionale Normalverteilung mit Erwartungswert 0 und Kovariantmatrix* $(\min(t_i, t_j))_{i,j}$.

Für $m = 1$ folgt die Verteilungskonvergenz von $\mu_n \circ \pi_t^{-1}$ für jedes $t \neq 0$ gegen die $\mathcal{N}(0,t)$-Verteilung sofort aus dem zentralen Grenzwertsatz. Für $t = 0$ ist die Konvergenz von $\mu_n \circ \pi_0^{-1} = \mathcal{L}(Y_n(0))$ gegen δ_0 offensichtlich (so ist das Dirac-Maß mit Masse in 0). Für $m \geq 2$ benötigen wir noch

Lemma 7.30 *Sei $d \in \mathbb{N}$ und für $j = 1, \ldots, d$ sei $(\mu_n^{(j)})_n$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$ mit $\mu_n^{(j)} \xrightarrow{w} \mu^{(j)} \in \mathcal{M}_1(\mathbb{R})$. Dann gilt*

$$\mu_n^{(1)} \otimes \ldots \otimes \mu_n^{(d)} \xrightarrow{w} \mu^{(1)} \otimes \ldots \otimes \mu^{(d)}$$

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

$$\mathcal{U} := \{(a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_d, b_d) : a_j, b_j \in B_j \text{ für } j = 1, \ldots, d\}$$

eine Basis der Topologie von $\mathbb{R}^d$. $\mathcal{U}$ ist durchschnittsstabil und für $(a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_d, b_d) \in \mathcal{U}$ gilt wegen dem Portmanteau-Theorem

$$\begin{aligned} & \mu_n^{(1)} \otimes \ldots \otimes \mu_n^{(d)}(a_1, b_1) \times \ldots \times (a_d, b_d) = \prod_{j=1}^d \mu_n^{(j)}((a_j, b_j)) \xrightarrow{n \to \infty} \prod_{j=1}^d \mu^{(j)}((a_j, b_j)) \\ = \; & \mu^{(1)} \otimes \ldots \otimes \mu^{(d)}((a_1, b_1) \times \ldots \times (a_d, b_d)). \end{aligned}$$

Das Lemma folgt nun, da sich jede offene Menge als abzählbare Vereinigung von Mengen aus $\mathcal{U}$ schreiben lässt. $\square$

Beweis von Satz 4.29: Wir können annehmen, dass $t_1 > 0$ gilt. Setze $\sum_{i=1}^{0} := 0$ und

$$Z_1^{(n)} := \sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}},\ Z_2^{(n)} := \sum_{i=[nt_1]+1}^{[nt_2]} \frac{X_i}{\sqrt{n}}, \dots, Z_m^{(n)} := \sum_{i=[nt_{m-1}]+1}^{[nt_m]} \frac{X_i}{\sqrt{n}}.$$

$Z_1^{(n)}, \dots, Z_m^{(n)}$ sind für jedes $n \in \mathbb{N}$ unabhängig. Mit Lemma 4.30 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

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le s\Big) = \Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-x^2/2} dx.$$

Nun gilt $\lim_{n\to\infty} \frac{k(n)}{n} = t_j - t_{j-1}$. Für $\varepsilon > 0$ und $s \in \mathbb{R}$ folgt

$$\begin{aligned}
\limsup_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &\le \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le \frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big) \\
&= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big) \qquad \text{und} \\
\liminf_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &\ge \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le \frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big) \\
&= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big), \\
\text{also} \quad \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}}\Big).
\end{aligned}$$

Dies ist die Verteilungsfunktion der eindimensionalen Normalverteilung mit Erwartungswert 0 und Varianz $t_j - t_{j-1}$. Nach Lemma 4.30 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 ν 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 4.?? konvergiert die Verteilung von

$$f(Z_1^{(n)}, \dots, Z_m^{(n)}) = \Big(\sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}}, \sum_{i=1}^{[nt_2]} \frac{X_i}{\sqrt{n}}, \dots, \sum_{i=1}^{[nt_n]} \frac{X_i}{\sqrt{n}}\Big)$$

gegen νf^{-1}. Sei $(U_1, \dots, U_m)$ eine Zufallsgröße mit Verteilung ν, dann besitzt die Normalverteilung νf^{-1} den Erwartungswert 0 und die Kovarianzmatrix mit Komponenten

$$\mathbb{E}\Big(\sum_{k=1}^{i} U_k \sum_{s=1}^{j} U_s\Big) = \sum_{k=1}^{\min\{i,j\}} \mathbb{E}(U_k^2) + \sum_{k=1, k\neq s}^{i} \sum_{s=1}^{j} \mathbb{E}(U_k U_s) = \sum_{k=1}^{\min\{i,j\}} (t_k - t_{k-1}) = \min\{t_i, t_j\}.$$

Sei nun $W_j^{(n)} := \sum_{i=1}^{[nt_j]} \frac{X_i}{\sqrt{n}} - Y_n(t_j)$. Dann gilt $|W_j^{(n)}| \leq \frac{|X_{[nt_j+1]}|}{\sqrt{n}}$, falls $t_j < 1$ und $W_j^{(n)} = 0$ sonst. Damit ist für $\varepsilon > 0$

$$\begin{aligned} P(|(W_1^{(n)}, \ldots, W_m^{(n)})| \geq \varepsilon) &\leq P\Big(\bigcup_{j=1}^m \{|W_j^{(n)}| \geq \varepsilon/m\}\Big) \\ \leq \sum_{j=1}^m P(|X_{[nt_j]+1}| \geq \sqrt{n}\varepsilon(m) &= mP(|X_1| \geq \sqrt{n}\varepsilon/m) \to 0 \end{aligned}$$

für $n \to \infty$, also konvergiert $(W_1^{(n)}, \ldots, W_m^{(n)})$ in Wahrscheinlichkeit geben 0. Nach Lemma 4.31 (unten) konvergiert dann auch $\mathcal{L}(Y_n(t_1), \ldots, Y_n(t_m))$ gegen νf^{-1}. □

Lemma 7.31 *Sei S ein separabler metrischer Raum mit Metrik d und es seien $(X_n)_n$ und $(Y_n)_n$ zwei Folgen von $(S, \mathcal{B}_S)$-wertigen Zufallsgrößen. Konvergiert $(X_n)_n$ in Verteilung gegen μ und $d(X_n, Y_n)$ stochastisch gegen 0, so konvergiert auch Y_n in Vrteilung gegen μ.*

Beweis: Sei $F \subseteq S$ abgeschlossen und für $\varepsilon > 0$

$$F^\varepsilon := \{x \in S : d(x, F) \leq \varepsilon\}.$$

Dann gilt

$$\limsup_{n\to\infty} \mathbb{P}(Y_n \in F) \leq \limsup_{n\to\infty} \mathbb{P}(x_n \in F^\varepsilon) + \limsup \mathbb{P}(x_n \in F^\varepsilon) \leq \mu(F^\varepsilon).$$

Da μ stetig ist, folgt $\mu(F^\varepsilon) \downarrow \mu(T)$, wenn $\varepsilon \downarrow 0$ konvergiert und damit die Behauptung. □

Nun formulieren wir das zentrale Resultat dieses Kapitels:

Satz 7.32 *(Satz von Donsker)*
Es gibt genau ein Wahrscheinlichkeitsmaß μ auf $(C[0,1], \mathcal{B}_C)$, so dass für alle $m \in \mathbb{N}$ und alle

$$0 \leq t1 < t_2 < \ldots < t_m \leq 1$$

das Maß $\mu \circ \pi_{t_1,\ldots,t_m}^{-1}$ die m-dimensionale Normalverteilung mit Erwartung 0 und Kovarianzmatrix $(\min(t_i, t_j))_{i,j}$ ist. Under den obigen Voraussetzungen an die X_i gilt

$$\mu_n \Rightarrow \mu. \tag{78}$$

Wie schon oben erwähnt ist die Eindeutigkeit von μ (mehr oder weniger) klar. Die Existenz von μ leiten wir aus der Konvergenzaussage ab. (4.13) ist eine unmittelbare Konsequenz aus dem folgenden Satz und dem darauf folgenden Lemma.

Satz 7.33 *Die Folge $(\mu_n)_n$ ist straff.*

Lemma 7.34 *$(\mu_n)_n$ ist genau dann schwach konvergent gegen μ, wenn jede Teilfolge $(\mu_{n_k})_k$ von $(\mu_n)_n$ eine gegen μ konvergente Teilfolge $(\mu_{n_{k_e}})_e$ besitzt.*

Beweis: Dies folgt aus der Definition von schwacher Konvergenz und der Tatsache, dass die entsprechende Aussage für reelle Zahlenfolgen wahr ist. □

?? Satz 4.33 können wir nun Satz 4.32 beweisen.

Beweis von Satz 4.32: Aus der Straffheit von $(\mu_n)_n$ (Satz 4.33) folgt, dass jede Teilfolge von $(\mu_n)_n$ eine konvergente Teilfolge hat (hierzu bemüht man den Satz von Prohorov). Der Limes dieser (Teil-)Teilfolge kann aber nur μ sein, denn die endlich dimensionalen Verteilungen konvergieren gegen die von μ. Also folgt die Behauptung des Satzes von Donsker aus Lemma 4.34. □

Es bleibt also Satz 4.33 zu beweisen. Dies ist (leider) noch ein ganzes Stück Arbeit. Ausgangspunkt dabei ist Satz 4.28.

Bemerkung 7.35 *$C([0,1]$ ist vollständig und separabel, also ist jedes Wahrscheinlichkeitsmaß ν 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 (4.11) äquivalent zu*

$$\forall\, \varepsilon > 0,\ \eta > 0\, \exists\, \delta > 0 \quad \forall\, n \in \mathbb{N} : \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta. \tag{79}$$

Beweis: (Von Satz 4.28) 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 Arzela-Ascoli die Aussagen (4.10) und (4.14), denn $K \subset \{f : |f(0)| \leq a\}$ für a groß genug und $K \subset \{f : \omega_\delta(f) < \varepsilon\}$ für δ klein genug. Für die Umkehrung sei $(\nu_n)_n$ eine Folge, die (4.10) und (4.14) erfüllt. Sei $\eta > 0$ vorgegeben. Nach (4.10) 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 δ_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 Arzela-Ascoli hat

$$K := A \cap \bigcup_{k=1}^{\infty} \{f : \omega_{\delta_k}(f) < 1/k\}$$

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. □

Bemerkung 7.36 *Hinreichend für (4.10) ist* $\nu_n(\{f : f(0) = 0\}) = 1$*, was für die* μ_n *im Satz von Donsker erfüllt ist.*

Lemma 7.37 *Hinreichend für (4.11) ist:*

$$\begin{aligned} \forall\, \varepsilon, \eta > 0\, \exists\, \delta \in (0,1),\ \exists\, n_0 \in \mathbb{N} : \forall\, n \quad &\geq \quad n_0,\ \forall\, t \in [0, 1-\delta] : \\ \frac{1}{\delta}\nu_n(\{f : \sup_{t \leq s \leq t+\delta} |f(x) - f(t)| \geq \varepsilon\}) \quad &\leq \quad \eta. \end{aligned} \tag{80}$$

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 (4.15). $m \in \mathbb{N}$ sei die kleinste natürliche Zahl mit $1/n < \delta_0$. 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 $|f(t) - f(\frac{k}{2m})| \geq \varepsilon/2$ oder $|f(s) - f(\frac{k}{2m})| \geq \varepsilon/2$. Also ist

$$\{f : \omega_\delta(f) \geq \varepsilon\} \subset \bigcup_{k=0}^{2m-2} \{f : \sup_{\frac{k}{2m} \leq s \leq \frac{k}{2m} + \delta_0} |f(s) - f(\frac{k}{2m})| \geq \varepsilon/2\},$$

und somit gilt für alle $n \geq n_0$:

$$\begin{aligned} \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \quad &\leq \quad \sum_{k=0}^{2m-2} \nu_n(\{f : \sup_{\frac{k}{2m} \leq s \leq \frac{k}{2m} + \delta_0} |f(s) - f(\frac{k}{2m})| \geq \varepsilon/2\}) \\ &\leq \quad (2m-1)\delta_0 \frac{\eta}{3} \leq (2 + \delta_0)\frac{\eta}{3} \leq \eta. \end{aligned}$$

Damit ist (4.11) gezeigt. □

Bemerkung 7.38 *Die Bedingung in Lemma 4.37 folgt aus der folgenden Aussage: Für alle* $\varepsilon > 0$ *gilt*

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} \sup_{t \in [0, 1-\delta]} \frac{1}{\delta} \nu_n(\{f : \sup_{t \leq s \leq t+\delta} |f(s) - f(t)| \geq \varepsilon\}) = 0.$$

Die Bedingung aus Bemerkung 4.38 soll nun für $\mu_n = P^{Y_n}$ untersucht werden: Für $\delta \in (0,1)$ und $t \in [0, 1-\delta]$ ist

$$\mu_n(\{f : \sup_{t \leq s \leq t+\delta} |f(s) - f(t)| \geq \varepsilon\}) = P(\sup_{t \leq s \leq t+\delta} |Y_n(s) - Y_n(t)| \geq \varepsilon).$$

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 $t+\delta \le 1$ kann man so abschätzen: Es existieren $j,k \in \{0,1,\dots,n\}$ mit $k<j$ und $\frac{k}{n} \le t < \frac{k+1}{n}$ sowie $\frac{j-1}{n} < t+\delta \le \frac{j}{n}$. Dann gilt für jedes $s \in [t,t+\delta]$:

$$\begin{aligned}|Y_n(s)-Y_n(t)| &\le |Y_n(t)-Y_n(\frac{k}{n}| + \max_{1\le i\le j-k} |Y_n(\frac{k+i}{n}) - Y_n(\frac{k}{n})| \\ &\le 2\max_{1\le i\le j-k} |Y_n(\frac{k+i}{n}) - Y_n(\frac{k}{n})|,\end{aligned}$$

also

$$\begin{aligned}\sup_{t\le s\le t+\delta} |Y_n(s)-Y_n(t)| &\le 2\max_{1\le i\le j-k} |Y_n(\frac{k+i}{n}) - Y_n(\frac{k}{n})| \\ &= 2\max_{1\le i\le j-k} |\sum_{r=k+1}^{k+i} X_r|/\sqrt{n}.\end{aligned}$$

Es ist $\frac{j-k-2}{n} \le \delta$. Für $n \ge \frac{1}{\delta}$ folgt $j-k \le 2n\delta$. Somit ist die rechte Seite der letzten Ungleichung nicht größer als $2\max_{1\le i\le 3n\delta} |\sum_{r=k+1}^{k+i} X_r|/\sqrt{n}$. Die Verteilung dieser Zufallsvariablen hängt nicht von k ab. Für $n \ge \frac{1}{\delta}$ gilt somit

$$\sup_{t\in[0,1-\delta]} P(\sup_{t\le s\le t+\delta} |Y_n(s)-Y_n(t)| \ge \varepsilon) \le P(\max_{1\le i\le 3n\delta} \frac{|S_i|}{\sqrt{n}} \ge \frac{\varepsilon}{2}).$$

Sei $m := [3n\delta]$, so ist $\sqrt{n} \ge \sqrt{m/3\delta}$ und somit

$$P(\max_{1\le i\le 3n\delta} \frac{|S_i|}{\sqrt{n}} \ge \frac{\varepsilon}{2}) \le P(\max_{1\le i\le m} \frac{|S_i|}{\sqrt{m}} \ge \frac{\varepsilon}{\sqrt{12\delta}}).$$

Für jedes feste $\delta > 0$ geht $m \to \infty$ für $n \to \infty$. Nach Bemerkung 4.38 müssen wir für jedes $\varepsilon > 0$ zeigen, dass

$$\lim_{\delta\searrow 0} \limsup_{m\to\infty} \frac{1}{\delta} P(\max_{1\le i\le m} \frac{|S_i|}{\sqrt{m}} \ge \frac{\varepsilon}{\sqrt{\delta}}) = 0 \tag{81}$$

gilt. Leider hilft die Abschätzung

$$P(\max_{1\le i\le m} \frac{|S_i|}{\sqrt{m}} \ge \frac{\varepsilon}{\sqrt{\delta}}) \le \sum_{i=1}^{m} P\Big(\frac{|S_i|}{\sqrt{m}} \ge \frac{\varepsilon}{\sqrt{\delta}})$$

analog zum Beweis von Lemma 4.37 nicht. Wir müssen diese Wahrscheilichkeit wesentlich genauer abschätzen:

Lemma 7.39 *Für alle $\lambda > 0$ und $m \in \mathbb{N}$ gilt*

$$P(\max_{1\le i\le m} |S_i| \ge \lambda\sqrt{m}) \le 2P(|S_m| \ge (\lambda-\sqrt{2})\sqrt{m}).$$

Beweis: Für $\lambda \le \sqrt{2}$ ist nichts zu zeigen. Sei $\lambda > \sqrt{2}$.

$$A_i := \bigcap_{j=1}^{i-1} \{|S_j| < \lambda\sqrt{m}\} \cap \{|S_i| \ge \lambda\sqrt{m}\},\ i = 1, \ldots, m.$$

Die A_i sind disjunkt und $A = \{\max_{1\le i\le m} |S_i| \ge \lambda\sqrt{m}\} = \bigcup_{i=1}^m A_i$. Also

$$\begin{aligned} P(A) &= P(A \cap \{|S_m| \ge (\lambda - \sqrt{2})\sqrt{m}\}) + P(A \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}) \\ &\le P(|S_m| \ge (\lambda - \sqrt{2})\sqrt{m}) + \sum_{j=1}^{m-1} P(A_j \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}), \end{aligned}$$

denn $A_m \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\} = \emptyset$. Weiter gilt

$$A_j \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\} \subset A_j \cap \{|S_m - S_j| \ge \sqrt{2m}\},\ j = 1, \ldots, m-1.$$

Die Ereignisse A_j und $\{|S_m - S_j| \ge \sqrt{2}\sqrt{m}\}$ sind unabhängig, also haben wir

$$P(A) \le P(|S_m| \ge (\lambda - \sqrt{2})\sqrt{m}) + \sum_{j=1}^{m-1} P(A_j) P(|S_m - S_j| \ge \sqrt{2m}).$$

Wegen

$$P(|S_m - S_j| \ge \sqrt{2m}) \le \frac{1}{2m} \mathbb{E}((\sum_{k=j+1}^m X_k)^2) = \frac{1}{2m} \sum_{k=j+1}^m \mathbb{E}(X_k^2) \le \frac{1}{2}$$

folgt

$$P(A) \le P(|S_m| \ge (\lambda - \sqrt{2})\sqrt{m}) + \frac{1}{2}\sum_{j=1}^m P(A_j) = P(|S_m| \ge (\lambda - \sqrt{2})\sqrt{m}) + \frac{1}{2} P(A),$$

also folgt die Behauptung. □

Wir schließen mit dem Beweis von (4.16) ab: Mit Lemma 4.39 und dem zentralen Grenzwertsatz folgt

$$\limsup_{m\to\infty} \frac{1}{\delta} P(\max_{1\le i\le m} \frac{|S_i|}{\sqrt{m}} \ge \frac{\varepsilon}{\sqrt{\delta}}) \le \limsup_{m\to\infty} \frac{2}{\delta} P(\frac{|S_m|}{\sqrt{m}} \ge \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}) = \frac{2}{\delta} P(|N| \ge \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}),$$

wenn n eine $N(0,1)$-verteilte Zufallsgröße bezeichnet. Die Markov-Ungleichung liefert

$$P(|N| \ge \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}) \le \frac{\mathbb{E}(|B|^3)}{(\frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2})^3}.$$

Dies führt zu (4.16). Somit ist die Straffheit der Folge $(\mu_n)_n$ bewiesen und somit Satz 4.33, also auch Satz 4.32. □

Bevor wir uns mit den Folgen von Satz 4.32 beschäftigen ein paar historische Bemerkungen zur Brownschen Bewegung: Die Brownsche Bewegung beschreibt die Bewegung eines Pollers in einer Flüssigkeit (natürlich unter dem Mikroskop, anderenfalls bewegt sich da wenig). 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.

Wir wollen uns nun mit Konsequenzen aus dem Satz von Donsker befassen. Hierzu sei zunächst bemerkt, dass aus der schwachen Konvergenz von μ_n gegen μ natürlich für jede stetige Funktion

$$h : C[0, 1] \to \mathbb{R}^d$$

folgt, dass $\mu_n \circ h^{-1}$ gegen $\mu \circ h^{-1}$ konvergiert. Da μ-Nullmengen von μ nicht gesehen werden gilt sogar für

$$D_h := \{x \in C[0, 1] : h \text{ ist unstetig in } x\}$$

der folgende

Satz 7.40 *Ist $h : C[0, 1] \to \mathbb{R}$ eine Borel-messbare Abbildung mit $\mu(D_h) = 0$ 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 oben definierte $(C, \mathcal{B}_C)$-wertige Zufallsvariable sei.*

Wir wollen dieses sogenannte *Invarianzprinzip* anhand zweier Beispiele ausführlich diskutieren. Zunächst diskutieren wir die Verteilung des Maximums einer Brownschen Bewegung. Die Technik, Satz 4.40 anzuwenden ist nun die, eine geeignete Folge von i.i.d. Zufallsvariablen zu finden, für die sich die Grenzverteilung von $\max_{0 \le t \le 1} Y_n(t)$ "leicht" bestimmen lässt. Satz 4.40 sagt uns dann, dass sich $\max Y_n(t)$ für alle anderen Wahlen von $(X_i)_i$, die im Satz von Donsker erlaubt sind, auch so verhält. Sei also

$$\begin{aligned} h : C[0, 1] &\to \mathbb{R} \\ f &\mapsto \sup_{0 \le t \le 1} f(t). \end{aligned} \tag{82}$$

Wir bemerken, dass das Supremum in der Definition 4.17 in der Tat ein Maximum ist und dass h natürlich in der Supremumsnorm auf $C[0, 1]$ stetig ist. Wir wählen die Zufallsgrößen $(X_i)_i$ i.i.d. mit

$$\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = +1) = \frac{1}{2}$$

(es ist somit $\mathbb{E}X_i = 0$ und $\mathbb{V}X_i = 1$ für alle i) und setzen

$$S_0 = 0 \quad \text{und} \quad S_n = \sum_{i=1}^{n} X_i;$$

$(S_n)_n$ ist also die eindimensionale Irrfahrt. Weiter setze

$$M_n := \max_{1 \le i \le n} S_i, \quad n \in \mathbb{N}.$$

Wir wollen die Verteilung von M_n analysieren. Hierbei beachte man

$$\sup_{0 \le t \le 1} Y_n(t) = \max_{0 \le i \le n} \frac{S_i}{\sqrt{n}} = \frac{M_n}{\sqrt{n}}.$$

Wir wollen die Folge $(M_n)_n$ auch als Folge der Maximalgewinne beim Münzwurfspiel bezeichnen. Es gilt:

Satz 7.41 *Für die Folge $(M_n)_n)$ der Maximalgewinne beim Münzwurfspiel gilt für alle $t \ge 0$*

$$\lim_{n \to \infty} P(\frac{M_n}{\sqrt{n}} \le t) = 2\Phi(t) - 1.$$

Für $t < 0$ gilt

$$P(\frac{M_n}{\sqrt{n}} \le t) = 0.$$

Hierbei bezeichnet Φ die Verteilungsfunktion der Standardnormalverteilung.

Als unmittelbare Konsequenz aus Satz 4.41 und Satz 4.40 ergibt sich

Satz 7.42 *Erfüllen die $(X_i)_i$ die Voraussetzungen des Satzes 4.40, so gilt für alle $t \in \mathbb{R}$*

$$\lim_{n \to \infty} P(\max_{0 \le i \le n} \frac{S_i}{\sqrt{n}} \le t) = \max\{2\Phi(t) - 1, 0\}.$$

Für den Beweis von Satz 4.41 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.

Lemma 7.43 *(Reflexionsprinzip)*

(i) 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).

(ii) 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: (i) Sei $(s_i = -a, s_{i+1}, \ldots, s_{j-1}, s_j = b)$. Dieser Pfad muss die x-Achse berühren. τ 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 (ii) ist

τ ist das erstmalige Erreichen des Wertes a. $\square$

Beweis von Satz 4.41: 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 (ii) von Lemma 4.43

$$P(M_n \geq a, S_n = b) = P(M_n \geq a, S_n = 2a - b)$$

für jedes $b \in \mathbb{Z}$. Also ist

$$P(S_n = l + k) = P(M_n \geq k, S_n = k - l).$$

Damit ist

$$\begin{aligned}
P(M_n \geq k) &= \sum_{l=-\infty}^{\infty} P(M_n \geq k, S_n = l+k) \\
&= \sum_{l=-\infty}^{\infty} P(M_n \geq k, S_n = l+k) + \sum_{l=1}^{\infty} P(S_n = l+) + P(S_n = k) \\
&= 2P(S_n > k) + P(S_n = k) \\
&= 2P(S_n \geq k) - P(S_n = k).
\end{aligned}$$

Sei $t \in \mathbb{R}_+$. Für $n \in \mathbb{N}$ bezeichne k_n die kleinste ganze Zahl größer-gleich $t\sqrt{n}$. Es gilt

$$P^{S_n/\sqrt{n}} \xrightarrow{w} \mathcal{N}(0,1).$$

Da $\{S_n/\sqrt{n} \geq t\} = \{S_n \geq k_n\}$, folgt

$$\lim_{n\to\infty} P(S_n \geq k_n) = \nu_{0,1}([t,\infty)).$$

Wegen $t\sqrt{n} \leq k_n < t\sqrt{n}+1$ gilt weiter für jedes $\varepsilon > 0$ und alle $n \in \mathbb{N}$ mit $1/\sqrt{n} \leq \varepsilon$

$$\{S_n = k_n\} = \{\frac{S_n}{\sqrt{n}} = \frac{k_n}{\sqrt{n}}\} \subset \{t \leq \frac{S_n}{\sqrt{n}} < t+\varepsilon\},$$

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

$$\lim_{n\to\infty} P(\frac{M_n}{\sqrt{n}} \geq t) = 2\nu_{0,1}([t,\infty)) = (\frac{2}{\pi})^{1/2} \int_t^{\infty} e^{-x^2/2}dx = 2(1-\Phi(t)),$$

womit die Behauptung des Satzes folgt. □

In einer zweiten Anwendung interessieren wir uns für den relativen Zeitanteil, den die Brownsche Bewegung oberhalb der x-Achse verbringt. Formal bekommen wir diesen mittels der folgenden Abbildung

$$g(f) := \lambda(\{t \in [0,1] : f(t) \geq 0\}),$$

wobei λ das Lebesgue-Maß bezeichnet. Eine direkte Anwendung von Satz 4.40 hat das Problem, dass die Abbildung g nicht in allen Punkten $f \in C[0,1]$ stetig ist. Beispielsweise haben die Funktionen

$$f_0 \equiv 0 \quad \text{und} \quad f_1 \equiv -\delta < 0$$

einen Supremumsabstand von $\delta > 0$ (wobei wir $\delta > 0$ beliebig klein wählen dürfen), aber es gilt

$$g(f_0) - g(f_1) = 1.$$

Es gilt aber

Lemma 7.44 *g ist $\mathcal{B}_C/\mathcal{B}$-messbar und $\mu(D_g) = 0$, wobei μ das Wiener-Maß bezeichnet und $D_g = \{x : g$ ist unstetig in $x\}$.*

Beweis: Es sei $\psi : C[0,1] \times [0,1] \to \mathbb{R}$ definiert durch $\psi(f,t) = f(t)$. ψ 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 aus Lemma 4.45 unten

$$\mathcal{B}_{C\times[0,1]} = \mathcal{B}_C \otimes \mathcal{B}_{[0,1]}.$$

Also ist ψ $\mathcal{B}_C \otimes \mathcal{B}_{[0,1]}/\mathcal{B}$-messbar.

Sei nun $A = \{(f,t) : f(t) \geq 0\} = \psi^{-1}([0,\infty)) \in \mathcal{B}_C \otimes \mathcal{B}_{[0,1]}$. Für $f \in C$ ist $g(f) = \lambda(\{t : (f,t) \in A\})$. Also ist $f \mapsto g(f)$ $\mathcal{B}_C/\mathcal{B}$-messbar (dies ist der Satz von Fubini). Es gilt

$$g(f) = \int_0^1 1_{[0,\infty]}(f(t))dt.$$

Ist $f \in C$ mit $\lambda(\{t : f(t) = 0\}) = 0$, und ist $(f_n)_n$ eine Folge in C mit $d(f_n, f) \to 0$, so gilt $1_{[0,\infty]}(f_n(t)) \to 1_{[0,\infty]}(f(t))$ für λ-fast alle $t \in [0,1]$. Nach dem Satz von der dominierten Konvergenz folgt

$$g(f_n) \to g(f).$$

Also ist $D_g \subset \{f : \lambda(\{t : f(t) = 0\}) > 0\}$ gezeigt.

Wir zeigen

$$\mu(\{f : \lambda(\{t : f(t) = 0\}) > 0\}) = 0.$$

Dazu müssen wir zeigen, dass $f \mapsto \lambda(\{t : f(t)00\})$ messbar ist. Dies geht analog zur Messbarkeit von g. Es ist zu zeigen:

$$0 = \int_C \lambda(\{t : f(t) = 0\})\mu(df) = \int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f,t)dt\,\mu(df).$$

Nach dem Satz von Fubini gilt

$$\begin{aligned}
\int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f,t)dt\,\mu(df) &= \int_{[0,1]} \int_C (1_{\{0\}} \circ \psi)(f,t)\mu(df)dt \\
&= \int_{[0,1]} \mu(\{f : f(t) = 0\})dt \\
&= \int_{[0,1]} \mu\pi_t^{-1}(\{0\})dt.
\end{aligned}$$

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. □

Hierbei haben wir von dem folgenden Lemma Gebrauch gemacht:

Lemma 7.45 *Sind S, S' separable topologische Räume mit Borelschen σ-Algebren $\mathcal{B}_S$ und $\mathcal{B}_{S'}$, dann gilt*

$$\mathcal{B}_{S\times S'} = \mathcal{B}_S \otimes \mathcal{B}_{S'}.$$

Beweis: Sind $A \subseteq S,\ B \subseteq S'$ offen, so ist $A \times B$ offen in $S \times S'$, also $A \times B \in \mathcal{B}_{S\times S'}$. Da die Mengen der Form $A \times B,\ A$ und B offen, die σ-Algebra $\mathcal{B}_S \otimes \mathcal{B}_{S'}$ erzeugen, ergibt sich

$$\mathcal{B}_S \otimes \mathcal{B}_{S'} \subseteq \mathcal{B}_{S-S'}.$$

Da $S,\ S'$ als separabel vorausgesetzt sind, gibt es abzählbare Basen

$$\begin{aligned} &\{U_i,\ i \in \mathbb{N}\} \quad \text{von } S \quad \text{und} \\ &\{U_i,\ i \in \mathbb{N}\} \quad \text{von } S'. \end{aligned}$$

$\{U_i \times U_j,\ i,j \in \mathbb{N}\}$ ist dann eine abzählbare Basis von $S \times S'$. Also ist jede offene Teilmenge von $S \times S'$ in $\mathcal{B}_S \otimes \mathcal{B}_{S'}$ enthalten, also gilt auch

$$\mathcal{B}_{S\times S'} \subseteq \mathcal{B}_S \otimes \mathcal{B}_{S'}.$$

□

Die Abbildung g erfüllt also die Voraussetzung des Invarianzprinzips. Es folgt nun die Berechnung von $\mathcal{L}(g(Y_n))$ im Spezialfall $P(X_i = \pm 1) = 1/2$. Dies ist eine elementare und schöne Auseinandersetzung mit der eindimensionalen, symmetrischen Irrfahrt und hebt die Bedeutung des Reflexionsprinzips eindrücklich hervor. Es gilt:

Satz 7.46 *Sind die $(S_i)_i$ unabhängig und $P(X_i = \pm 1) = 1/2$, so gilt für $t \in [0,1]$*

$$\lim_{n\to\infty} P(g(Y_n) \leq t) = \frac{2}{\pi}\ \arcsin\sqrt{t}.$$

Dies liefert somit die Verteilungsfunktion von μg^{-1}, wenn μ das Wiener-Maß ist. Es folgt mit dem Invarianzprinzip

Satz 7.47 *(Arcussinus-Gesetz)*
Die auf $(C, \mathcal{B}_C, \mu)$ definierte Zufallsgröße $f \mapsto \lambda(\{t : f(t) \geq 0\})$ 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 4.40, so gilt für $t \in [0,1]$:

$$\lim_{n\to\infty} P(g(Y_n) \leq t) = \frac{2}{\pi}\ \arcsin\sqrt{t}.$$

Bemerkung 7.48 *Es ist nicht sehr schwer zu zeigen, dass*

$$g(Y_n) - \frac{1}{n}|\{m \leq n : S_m > 0\}|$$

in Wahrscheinlichkeit gegen 0 konvergiert. Also folgt, dass auch

$$\mathcal{L}(\frac{1}{n}|\{m \leq n : S_m > 0\})|$$

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. 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. $p := |\{i \in \{0, \dots, n-1\} : s_{i+1} = s_i + 1\}|$), 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

$$|\Omega_{(n,s)}| = \binom{n}{(n+2)/2} = \binom{p+q}{p}. \tag{83}$$

(II) *Freier Endpunkt:* Ω_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 und etwa vorstellen, das 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 stetis vorn bei der Auszählung? Diese Wahrscheinlichkeit ist gleich $|A|/|\Omega_{(p+q,p-q)}| = |A|/\binom{p+q}{p}$, wobei

$$A = \{\omega = (0, s_1, \dots, s_{p+q}) \in \Omega_{(p+q,p-q)} : s_k > 0 \text{ für } 1 \le k \le p+1\}$$

ist. Zum Abzählen der Pfade in A verwenden wir Lemma 4.43. Für $\omega = (0, s_1, \dots, s_n) \in A$ 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 4.43 gleich der Anzahl aller Pfade von (1,-1) nach $(p+q, p-q)$. Wenden wir (4.18) an, so ergibt sich also

$$|A| = \binom{p+q-1}{p-1} - \binom{p+q-1}{p} = \frac{p-q}{p+q}\binom{p+q}{p}. \tag{84}$$

(Wir haben hier natürlich $p > q$ vorausgesetzt.) Die Anzahl aller Elemente in $\Omega_{(p+q,p-q)}$ ist nach (4.18) $\binom{p+q}{p}$. Somit ergibt sich das folgende Resultat, das wir schon im Kapitel über Martingale kennengelernt haben.

Satz 7.49 *(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:

(i) Kandidat K_1 führt während der gesamten Auszählung, erst am Schluss tritt Gleichstand ein.

(ii) 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:

Satz 7.50 *(i) Es gibt* $\frac{1}{p}\binom{2p-2}{p-1}$ *Pfade von* $(0,0)$ *nach* $(2p,0)$ *mit* $s_1 > 0$, $s_2 > 0, \ldots, s_{2p-1} > 0$.

(ii) 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:

(i) 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 (4.19) mit $q = p - 1$ ist dies gleich

$$\frac{1}{2p-1}\binom{2p-1}{p} = \frac{1}{p}\binom{2p-2}{p-1}.$$

(ii) 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). (ii) folgt dann aus (i).

□

Aus (ii) 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.

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

$$|A_{2k}| = \binom{2k}{k} 2^{n-2k}.$$

Ω_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

$$\lim_{n\to\infty} n!/(\sqrt{2\pi} n^{n+1/2} e^{-n}) = 1. \tag{85}$$

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.

Setzen wir die Stirling-Approximation ein, so erhalten wir (siehe auch den lokalen Grenzwertsatz aus der Stochastik)

Satz 7.51 *Es gilt*

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}}.$$

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

$$f_{2k} = P(S_1 \neq 0,\ S_2 \neq 0, \dots, S_{2k-1} \neq 0,\ S_{2k} = 0).$$

Lemma 7.52 *(i)* $f_{2k} = \frac{1}{2k} u_{2k-2} = P(S_1 \geq 0,\ S_2 \geq 0, \dots, S_{2k-2} \geq 0,\ S_{2k-1} < 0) = u_{2k-2} - u_{2k}$.

(ii) $u_{2k} = P(S_1 \neq 0,\ S_2 \neq 0, \dots, S_{2k} \neq 0) = P(S_1 \geq 0,\ S_2 \geq 0, \dots, S_{2k} \geq 0)$.

(iii) $u_{2k} = \sum_{j=1}^{k} f_{2j} u_{2k-2j}$.

Beweis:

(i) Nach Satz 4.50 (i) gibt es $\frac{1}{k}\binom{2k-2}{k-1}$ Pfade von (0,0) nach $(2k, 0)$ mit $s_1 > 0, \dots, s_{2k-1} > 0$ und natürlich genauso viele mit $s_1 < 0, \dots, s_{2k-1} < 0$. 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 $s_1 \geq 0, \dots, s_{2k-3} \geq 0,\ s_{2k-2} = 0$ ist gleich der Anzahl der Pfade von (0,0) nach $(2k-2, 0)$ mit allen y-Koordinaten ≥ 0. Die zweite Gleichung in (i) folgt dann mit Hilfe von Satz 4.50. Die dritte ergibt sich aus

$$u_{2k} = \binom{2k}{k} 21{-}2k = \frac{2k(2k-1)}{k \cdot k}\binom{2k-2}{k-1} \cdot \frac{1}{4} \cdot 2^{-2k+2} = (1 - \frac{1}{2k}) u_{2k-2}. \quad (86)$$

(ii) C_{2j} sei das Ereignis $\{S_1 \neq 0,\ S_2 \neq 0, \dots, S_{2j-1} \neq 0,\ S_{2j} = 0\}$. Diese Ereignisse schließen sich gegenseitig aus und haben Wahrscheinlichkeiten $f_{2j} = u_{2j-2} - u_{2j}$. Somit ist mit $u_0 = 1$

$$P(S_1 \neq 0, S_2 \neq 0, \dots, S_{2k} \neq 0) = 1 - P(\bigcup_{j=1}^{k} C_{2j}) = 1 - \sum_{j=1}^{k} (u_{2j-2} - u_{2j}) = u_{2k}.$$

Die zweite Gleichung folgt analog aus der dritten Identität von (i).

(iii) Für $1 \leq j \leq k$ sei $B_j = \{S_1 \neq 0, S_2 \neq 0, \dots, S_{2j-1} \neq 0, S_{2j} = 0, S_{2k} = 0\}$. Diese Ereignisse sind parweise 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

$$u_{2k} = \sum_{j=1}^{k} P(B_j) = \sum_{j=1}^{k} f_{2j} u_{2k-2j}.$$

□

Eine interessante Konsequenz ergibt sich aus (4.21). Nach Satz 4.7 ist jeder Zustand der eindimensionalen Irrfahrt rekurrent, insbesondere kehrt sie unendlich oft in den Punkt 0 zurück. Sei T der Zeitpunkt der ersten Rückkehr, also

$$T = \inf\{n \geq 1 : S_n = 0\}.$$

Offenbar ist T gerade und es gilt

$$\mathbb{P}(T = 2k) = f_{2k}.$$

Aus (i) 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 (i) in Lemma 4.52 anwenden. Nach Satz 4.?? divergiert jedoch diese Reihe! Man kann auch sagen, dass $\mathbb{E}T$ gleich ∞ 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 gerade 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 4.52 (ii) 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. Andererseits 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.

Satz 7.53 *(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 4.52 (ii) 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 Menge von Pfaden ein.

Für $1 \leq l \leq n,\ 0 \leq m \leq n-l$ sei $g_{l,m}^+$ die Menge der Pfade der Lände $2n$ 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

$$G_{l,m}^+ \subset F_{l+m}^{(n)}, G_{l,m}^- \subset F_m^{(n)}.$$

Man beachte, dass für $1 \leq j \leq n-1$ jeder Pfad aus $F_j^{(n)}$ zu genau einer der Mengen $G_{l,m}^+, G_{l,m}^-$ gehört. Dies foglt 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 von $2l$ positiv, und zu $G_{l,j}^-$, falls er vor $2l$ negativ ist. Demzufolge ist

$$P(F_j^{(n)}) = \sum_{l=1}^{j} P(G_{l,j-l}^+) + \sum_{l=1}^{n-j} P(G_{l,j}^-).$$

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. $|G_{lm,m}^+|$ 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

$$|G_{l,m}^+| = |G_{l,m}^-| = \frac{1}{2} f_{2l} 2^{2l} P(F_m^{(n-l)}) 2^{2n-2l},$$

und

$$P(G_{l,m}^+) = P(G_{l,m}^-) = \frac{1}{2} f_{2l} P(F_m^{(n-l)}).$$

Nach der weiter oben stehenden Gleichung ist also

$$P(F_j^{(n)}) = \frac{1}{2}\sum_{l=1}^{j} f_{2l} P(F_{j-l}^{(n-l)}) + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l} P(F_j^{(n-l)}).$$

Nach der Induktionsvoraussetzung ist das

$$= \frac{1}{2}\sum_{l=1}^{j} f_{2l} u_{2j-2l} u_{2n-2j} + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l} u_{2n-2j-2l} u_{2j} = u_{2j} u_{2n-2j}$$

nach Lemma 4.52 (iii). □

Um das Verhalten von $P(F_j^{(k)})$ für festes k als Funktion von j zu untersuchen, betrachten wir für $1 \le j \le k-1$ die Quotienten

$$\frac{P(F_j^{(k)})}{P(F_{j+1}^{(k)})} = \frac{\binom{2j}{j}\binom{2k-2j}{k-j}}{\binom{2j+2}{j+1}\binom{2k-2j-2}{k-j-1}} = \frac{(2k-2j-1)(j+1)}{(2j+1)(k-j)}.$$

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 schein 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.

Ein zweiter Gedanke zeigt, dass dieser Widerspruch in der Tat nur scheinbar ist: Im ersten Schritt muss die Irrfahrt notwendig positiv oder negativ werden; danach muss sie schon einmal mehr negativ (bz. positiv) werden, um den "Vorsprung ins Positive" wieder auszugleichen. Das ist um so unwahrscheinlicher je größer der "Anfangsvorsprung" ist.

Mit Hilfe von Satz 4.51 lässt sich eine einfache Approximation für $P(F_j^{(k)})$ für große j und $k-j$ gewinnen:

Satz 7.54 *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 4.54 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 glein 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 α und β liegt. Genauer: T_k sei (die auf Ω_{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

$$\gamma_k(\alpha, \beta) := P(\alpha \leq \frac{T_k}{2k} \leq \beta) = \sum_{j : \alpha \leq \frac{j}{k} \leq \beta} P(F_j^{(k)}).$$

Wir sind übrigens nun bei der in Satz 4.46 diskutierten Abbildung $g(Y_n)$ angekommen, denn $T_k = 2k\, g(Y_{2k})$. Wir wollen nun aus Satz 4.54 für $k \to \infty$ folgern:

$$\gamma_k(\alpha, \beta) \sim \frac{1}{\pi} \sum_{j : \alpha \leq \frac{j}{k} \leq \beta} \frac{1}{k} \frac{1}{\sqrt{\frac{j}{k}(1 - \frac{j}{k})}}. \tag{87}$$

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 4.46):

Satz 7.55 *(Arcussinus-Gesetz)*

$$\lim_{k \to \infty} \gamma_k(\alpha, \beta) = \frac{2}{\pi} (\arcsin \sqrt{\beta} - \arcsin \sqrt{\alpha}).$$

Beweis: Wir müssen (4.22) zeigen. Wir schreiben die Stirling-Approximation als $n! = \sqrt{2\pi n}(\frac{n}{e})^n F(n)$ mit $\lim_{n\to\infty} F(n) = 1$. Es folgt

$$P(F_j^{(k)}) = \binom{2j}{j}\binom{2k-2j}{k-j}\frac{1}{2^{2k}} = \frac{1}{\pi}\frac{1}{\sqrt{(\frac{j}{k}(1-(\frac{j}{k}))}}\frac{1}{k}\frac{F(2j)F(2(k-j))}{F(j)F(j)F(k-j)F(k-j)}.$$

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

$$\sum_{\alpha \leq \frac{j}{k} \leq \beta} P(F_j^{(k)}) = \Big(\frac{1}{k}\sum_{\alpha \leq \frac{j}{k} \leq \beta}\frac{1}{\pi}\frac{1}{\sqrt{(j/k)(1-(j/k))}}\Big)G_{\alpha,\beta}(k).$$

Nun folgt die Behauptung gleichmäßig für $\delta \leq \alpha < \beta \leq 1-\delta$, wie auch immer $0 < \delta < 1/2$ gewählt war. Damit folgt die Behauptung. $\square$

Bemerkung 7.56 *Die Aussage von Satz 4.55 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, $\leq \beta$ ist – gegen $\frac{2}{\pi}\arcsin\sqrt{\beta}$ konvergiert.*

Beweis: Offensichtlich gilt $\lim_{k\to\infty}\gamma_k(0,\frac{1}{2}) = 1/2$. Ist $\beta \in (0, 1/2)$, so folgt

$$\begin{aligned}\lim_{k\to\infty}\gamma_k(0,\beta) &= \lim_{k\to\infty}(\gamma_k(0,1/2) - \gamma_k(\beta,1/2)) = \frac{2}{\pi}\arcsin\sqrt{\beta} \quad \text{für} \quad \beta > 1/2\\ \lim_{k\to\infty}\gamma_k(0,\beta) &= \lim_{k\to\infty}(\gamma_k(0,1/2) + \gamma_k(1/2,\beta)) = \frac{2}{\pi}\arcsin\sqrt{\beta}.\end{aligned}$$

Für $\gamma_k(\alpha,1)$ führt dasselbe Argument zum Ziel. $\square$

Der Beweis des Arcus-Sinus-Gesetzes wurde zuerst von P. Levy 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 diese 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.

$$A(x) = \frac{2}{\pi} \arcsin \sqrt{x}$$

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	0.495
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		

Literatur

[1] Feller, :

[2] Hardy / Wright: 1959