

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, \dots reellwertige, i.i.d. Zufallsvariablen sind mit einer endlichen momenterzeugenden Funktion

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

(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

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

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

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

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

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

so folgt

Lemma 1.1 *Es gilt*

$$\gamma_{m+n} \geq \gamma_m + \gamma_n \quad (4)$$

und daraus folgt, dass

$$\frac{\gamma_n}{n} \rightarrow \sup_{m \geq 1} \frac{\gamma_m}{m} \quad \text{wenn } n \rightarrow \infty. \quad (5)$$

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

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

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

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

Division durch n ergibt

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

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

Dieses Lemma impliziert schon, dass

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

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

$$\mathbb{P}(S_n \geq na) \leq e^{-nI(a)}. \quad (6)$$

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

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

Übung 1.3 Man zeige, dass

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

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 $t > 0$:

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

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

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

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 $\psi(0) = \log \varphi(0) = 0$. Wenn wir somit zeigen können, dass

$$\psi'(t) \rightarrow \mu := \mathbb{E}X_1 \quad \text{wenn} \quad t \rightarrow 0,$$

erhalten wir

$$at - \psi(t) = \int_0^t a - \psi'(\theta) d\theta > 0$$

wenn t klein genug ist, wie gewünscht.

Zunächst zeigen wir, dass die Ableitungen existieren. Es ist

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

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:

$$\varphi'(t) = \int x e^{tx} \mathbb{P}_X(dx) \quad \text{für } t \in (0, t_0).$$

Nimmt man den Limes $t \rightarrow 0$ 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 $\varphi'(t) \rightarrow \mu$ für $t \rightarrow 0$. Da nun andererseits $\varphi(t) \rightarrow 1$, wenn $t \rightarrow 0$, haben wir somit gezeigt, dass $\psi'(t) \rightarrow \mu$ gilt, wenn $t \rightarrow 0$, was nach der Eingangsbeobachtung 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

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

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 ein Maß Q_t vermöge der Dichte

$$\frac{dQ_t}{d\mathbb{P}_X}(y) := \frac{1}{\varphi(t)} e^{ty}.$$

Man beachte, dass Q_t ein W -Maß ist. Sein Mittelwert berechnet sich als

$$\int x dQ_t(x) = \frac{1}{\varphi(t)} \int_{-\infty}^{\infty} x e^{tx} d\mathbb{P}_X(x) = \frac{\varphi'(t)}{\varphi(t)}.$$

Differenziert man noch einmal, erhält man

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

Diese Ungleichung ist sogar strikt, wenn wir annehmen

$$X \text{ ist nicht die Dirac-Verteilung in } \mu. \quad (7)$$

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:*

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

wobei

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

gilt.

Beweis: Ohne Beschränkung der Allgemeinheit nehmen wir an, dass $a = 0$ und $\mathbb{E}X_1 < 0$ gilt (substituiert man nämlich $X_1 \rightarrow X_1 + a$, 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 folgt mithilfe der exponentiellen Chebyshev–Ungleichung für alle positiven t

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

und somit

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

Um das Supremum über die ganze reelle Achse auszudehnen, überlegen wir, dass φ eine strikt konvexe Funktion ist. 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 $\varphi'(t) = \int x e^{tx} d\mathbb{P}^X(x) < 0$ für alle $t \in \mathbb{R}$. Somit ist φ strikt fallend. Es ist somit

$$g = \lim_{t \rightarrow \infty} \varphi(t) = \mathbb{P}(X_1 = 0) = 0.$$

Da auch

$$\mathbb{P}(S_n \geq 0) = 0$$

gilt, haben wir in diesem Fall schon (8).

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

Wie oben zeigt man, dass φ strikt fallend ist und

$$\lim_{t \rightarrow \infty} \varphi(t) = g = \mathbb{P}(X_1 = 0) > 0.$$

Da in diesem Falle

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}(X_1 = \dots = X_n = 0) = g^n$$

gilt, folgt auch hier (8).

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

Dann gilt offenbar $\lim_{t \rightarrow \pm\infty} \varphi(t) = \infty$ 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

$$\mathbb{P}(S_n \geq 0) \leq \mathbb{E} e^{\tau S_n} = (\varphi(\tau))^n = g^n,$$

also

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \leq \log g.$$

Um zu zeigen, dass $\log g$ auch eine untere Schranke ist, verwenden wir eine Technik, die als Tiltten 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 gemäß Q verteilt sind, wobei

$$\frac{dQ}{d\mathbb{P}}(x) = \frac{1}{g} e^{\tau x}$$

besitzen. Q heißt auch die Cramér-Transformierte von \mathbb{P} . Bemerke, dass

$$g = \varphi(\tau) = \int_{-\infty}^{\infty} e^{\tau y} d\mathbb{P}^X(y).$$

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 $t \in \mathbb{R}$

$$\hat{\varphi}(t) = \int_{\mathbb{R}} e^{tx} dQ^Y(x) = \frac{1}{g} \int_{\mathbb{R}} e^{tx} e^{\tau x} d\mathbb{P}^X(x) = \frac{1}{g} \varphi(t + \tau) < \infty.$$

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

$$\begin{aligned} \mathbb{P}(S_n \geq 0) &= \int_{\sum_{i=1}^n x_i \geq 0} d\mathbb{P}^X(x_1) \dots d\mathbb{P}^X(x_n) \\ &= \int_{\sum_{i=1}^n x_i \geq 0} [ge^{-\tau x_1} dQ^Y(x_1)] \dots [ge^{-\tau x_n} dQ^Y(x_n)], \end{aligned}$$

so folgt die Behauptung.

□

Lemma 1.9 Es gilt

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

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

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

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 .

Der Satz von Cramér ist somit ein Spezialfall der folgenden Definition.

Definition 1.11 Es sei (X, \mathcal{X}) ein metrischer Raum mit seiner Borelschen σ -Algebra. Wir sagen, dass eine Folge von Wahrscheinlichkeitsmaßen $(\mathbb{P}_n)_n$ auf (X, \mathcal{X}) einem Prinzip der großen Abweichungen (kurz: LDP) mit Ratenfunktion $I : X \rightarrow [0, \infty]$ und Geschwindigkeit a_n genügt, falls

1. Für jedes $L < \infty$ die Niveaumengen

$$N_L = \{x : I(x) \leq L\}$$

kompakt sind.

2. Für alle offenen Mengen $G \subseteq X$ gilt

$$\liminf_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}_n(G) \geq - \inf_{x \in G} I(x).$$

3. Für alle abgeschlossenen Mengen $A \subseteq X$ gilt

$$\limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}_n(A) \leq - \inf_{x \in A} I(x).$$

Wir sagen, eine Folge $(X_n)_n$ von Zufallsvariablen genügt einem LDP, falls (\mathbb{P}^{X_n}) dies tut.

Beispiel 1.12 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 1.13 Berechnen Sie die Ratenfunktion, für X_i die $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind.

Übung 1.14 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 im Satz von Cramér hat die folgenden Eigenschaften:

Lemma 1.15 *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 ist 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 $0 \leq t \leq 1$ und x, y gilt

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

Da

$$\psi(0) = \log \mathbb{E}(1) = 0,$$

folgt

$$I(x) \geq 0x - \psi(0) = 0$$

für alle x . Zur Halbstetigkeit bemerken wir, dass für $x \in \mathbb{R}$ und $x_n \rightarrow x$ und jedes $\lambda \in \mathbb{R}$ gilt:

$$\liminf_{x_n \rightarrow x} I(x_n) \geq \liminf_{x_n \rightarrow x} [\lambda x_n - \psi(\lambda)] = \lambda x - \psi(\lambda).$$

Daraus ergibt sich

$$\liminf_{x_n \rightarrow x} I(x_n) \geq \sup_{\lambda} [\lambda x - \psi(\lambda)] = I(x).$$

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

$$\psi(\lambda) \geq E(\log e^{\lambda X_1}) = \lambda \mathbb{E}X_1$$

folgt, also

$$I(\mathbb{E}X_1) \leq \lambda \mathbb{E}X_1 - \lambda \mathbb{E}X_1 = 0.$$

□

Bemerkung 1.16 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 1.17 Die Aussagen von Theorem 1.6 bleibt natürlich wahr, wenn wir die Wahrscheinlichkeiten $\mathbb{P}(S_n \leq na)$ für $a < \mathbb{E}X_1$ abschätzen. Dies sieht man leicht durch Übergang von X_1 zu $-X_1$. Eine Übung ist es, dies formal zu machen.

Bemerkung 1.18 Tatsächlich ist mit dieser Bemerkung der Satz von Cramér ein Prinzip der großen Abweichungen im Sinne der obigen Definition. Da nun wie oben gesehen $I(x) = 0 \Leftrightarrow x = \mathbb{E}X_1$, und außerdem I strikt konvex ist auf

$$D_I := \{x : I(x) < \infty\},$$

folgt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \leq na) = - \inf_{x \in (-\infty, a]} I(x)$$

für alle $a < \mathbb{E}X_1$ und

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = - \inf_{x \in [a, \infty)} I(x).$$

Dies impliziert aber auch

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \in (a, b)\right) = - \inf_{x \in (a, b)} I(x).$$

Z. B. überlegt man für $\mathbb{E}X_1 < a < b < \infty$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \in (a, b)\right) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log (\mathbb{P}(S_n > na, -\mathbb{P}(S_n \geq nb))) \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n > na) = -I(a) = \inf_{x \in (a, b)} I(x) \end{aligned}$$

und

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \in (a, b)\right) \geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n > na)(1 - \varepsilon) = -I(a) \quad \forall \varepsilon > 0.$$

Schließlich erhält man hieraus das LDP für $\frac{S_n}{n}$ mit Geschwindigkeit n und Rate I , indem man beliebige Mengen durch Intervalle approximiert.

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

Korollar 1.19 *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. \square

Wir wollen nun noch einige allgemeine Eigenschaften für Prinzipien der großen Abweichungen sammeln.

Es werden sich einige Eigenschaften herauskristallisieren, die ganz allgemein gelten. Desweiteren werden wir gewisse Übertragungsmechanismen aufzeigen, die uns erlauben, wieder neue Prinzipien der großen Abweichungen zu erhalten. Wir beginnen damit, die exponentielle Konvergenz im allgemeinen Rahmen zu studieren. Hier gibt es keinen Grund, dass eine Ratenfunktion genau eine Nullstelle besitzen sollte (dies ist im Allgemeinen auch falsch, wie wir später sehen werden). Dennoch bekommen wir eine exponentielle Konzentration auf „kleine“ Mengen:

Satz 1.20 *Sei $(\mathbb{P}_n)_n$ eine Folge von Maßen, die einem LDP mit Geschwindigkeit n und Ratenfunktion I genügt. Dann gibt es für jedes $\ell < \infty$ eine kompakte Menge D^ℓ , so dass*

$$\mathbb{P}_n(D^\ell) \geq 1 - e^{-n\ell} \quad \forall n.$$

Beweis: Nach Annahme sind die Niveaumengen

$$A_\ell = \{x : I(x) \leq \ell + 2\}$$

kompakt, sie können also für jedes k durch eine endliche Menge offener Kugeln vom Radius $\frac{1}{k}$ überdeckt werden. Ihre Vereinigung sei U_k . Es ist

$$I(x) \geq \ell + 2$$

auf der abgeschlossenen Menge U_k^c . Daher folgt aus der oberen Abschätzung des LDP

$$\overline{\lim} \frac{1}{n} \log \mathbb{P}_n(U_k^c) \leq -(\ell + 2)$$

und somit für n hinreichend groß ($n \geq n_0$)

$$\mathbb{P}_n(U_k^c) \leq e^{-n(\ell+1)}.$$

O.B.d.A. ist $n_0 = n_0(k) \geq k$ und daher für $n \geq n_0$

$$\mathbb{P}_n(U_k^c) \leq e^{-k} e^{-n\ell}.$$

Für die ersten Indizes $j = 1, \dots, n_0(k)$ können wir ohnehin kompakte Mengen $B_{k,1}, \dots, B_{k,n_0}$ finden, so dass

$$\mathbb{P}_j(B_{k,j}^c) \leq e^{-k} e^{-j\ell} \quad \forall j$$

gilt, denn einzelne Maße sind natürlich straff. Wir setzen

$$E = \bigcap_k [U_k \cup (\bigcup_j B_{k,j})].$$

E ist total beschränkt (Übung) und daher präkompakt, da polnische Räume vollständig sind. Weiter gilt

$$\mathbb{P}_n(E^c) \leq [\sum_{k \geq 1} e^{-k}] e^{-n\ell} \leq e^{-n\ell}$$

wir können also für unser D^ℓ den Abschluss von E wählen. □

Wir wenden uns nun Situationen zu, bei denen sich ein LDP von einer Situation auf eine verwandte Situation übertragen lässt.

Satz 1.21 *Seien $(\mathbb{P}_n)_n$ und $(\mathbb{Q}_n)_n$ zwei Folgen von Maßen, die einem LDP mit Geschwindigkeit a_n und Rate $I(\cdot)$ bzw. $J(\cdot)$ genügen. Dann genügt auch die Folge der Produktmaße $\mathbb{R}_n = \mathbb{P}_n \otimes \mathbb{Q}_n$ auf $X \times Y$ einem LDP mit Geschwindigkeit a_n . Die Rate ist gegeben durch*

$$K(x, y) = I(x) + J(y).$$

Da der Beweis typisch für einige Techniken in den großen Abweichungen ist, geben wir ihn schrittweise.

Beweis: 1. Schritt: Sei

$$(x, y) = z \in Z = X \times Y.$$

Sei $\varepsilon > 0$. Wir wollen zeigen, dass es eine offene Umgebung $N = N_{x,\varepsilon}$ von z gibt, so dass

$$\overline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{R}_n(N) \leq -K(z) + \varepsilon \quad (12)$$

gilt. Da I von unten halbstetig ist, gibt es eine offene Menge $U_1 \subseteq X$, so dass $I(x') \geq I(x) - \frac{\varepsilon}{2}$ für alle $x' \in U_1$ gilt. Wir wählen (da U_1 offen ist) eine offene Menge $U_2 \subseteq X$ mit

$$x \in U_2 \subseteq \overline{U_2} \subseteq U_1.$$

Wegen des LDP für (\mathbb{P}_n) gilt

$$\overline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{P}_n(U_2) \leq \overline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{P}_n(\overline{U_2}) \leq - \inf_{x' \in \overline{U_2}} I(x') \leq -I(x) + \frac{\varepsilon}{2}.$$

Analog findet man offene Mengen $V_1, V_2 \subseteq Y$ mit $y \in V_2 \subseteq \overline{V_2} \subseteq V_1$ und

$$\overline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{Q}_n(V_2) \leq \overline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{Q}_n(\overline{V_2}) \leq - \inf_{y' \in \overline{V_2}} J(y') \leq -J(y) + \frac{\varepsilon}{2}.$$

Wenn wir $N = U_2 \times V_2$ als offene Umgebung von z wählen, sind wir fertig mit (12).

2. Schritt: Sei $D \subseteq Z$ kompakt und $\varepsilon > 0$. Wir wollen nun zeigen, dass für eine ε -Umgebung $D_\varepsilon \supseteq D$ von D gilt

$$\overline{\lim}_n \frac{1}{n} \log \mathbb{R}_n(D_\varepsilon) \leq - \inf_{z \in D} K(z) + \varepsilon. \quad (13)$$

Aus Schritt 1 wissen wir, dass für alle $z \in D$ eine Umgebung N_z existiert, so dass

$$\overline{\lim}_n \frac{1}{n} \log \mathbb{R}_n(N_z) \leq -K(z) + \varepsilon \leq - \inf_{z \in D} K(z) + \varepsilon$$

gilt. Aufgrund der Kompaktheit von D enthält die offene Überdeckung $\{N_z\}_{z \in D}$ von D eine endliche Teilüberdeckung $\{N_j : 1 \leq j \leq k\}$. Für $D_\varepsilon = \bigcup_{j=1}^k N_j$ erhalten wird

$$\mathbb{R}_n(D_\varepsilon) \leq k \cdot \sup_{j=1, \dots, k} \mathbb{R}_n(N_j)$$

und daher

$$\overline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{R}_n(D_\varepsilon) \leq - \inf_{z \in D} K(z) + \varepsilon.$$

Da ε beliebig war, folgt (13). Insbesondere folgt auch

$$\overline{\lim}_n \frac{1}{n} \log \mathbb{R}_n(D) \leq - \inf_{z \in D} K(z). \quad (14)$$

3. Schritt: Aus der superexponentiellen Straffheit (Satz 1.20) wissen wir, dass es für jedes ℓ kompakte Teilmengen A_ℓ und B_ℓ von X bzw. Y gibt, $\mathbb{P}_n(A_\ell^c) \leq e^{-a_n \ell}$ und

$\mathbb{Q}_n(B_\ell^c) \leq e^{-a_n \ell}$. Setzen wir $C_\ell = A_\ell \times B_\ell$, so ist C_ℓ nach dem Satz von Tychonov kompakt und es gilt

$$\mathbb{R}_n(C_\ell^c) \leq 2e^{-a_n \ell}.$$

Nun schreiben wir eine beliebige abgeschlossene Menge $C \subseteq Z$ als

$$C = (C \cap C_\ell) \dot{\cup} (C \cap C_\ell^c).$$

Damit bekommen wir

$$\begin{aligned} \overline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{R}_n(C) &\leq \overline{\lim}_{a_n} \frac{1}{a_n} \log 2 \cdot \max(\mathbb{R}_n(C \cap C_\ell), \mathbb{R}_n(C \cap C_\ell^c)) \\ &= \overline{\lim}_{a_n} \frac{1}{a_n} \log \max(\mathbb{R}_n(C \cap C_\ell), e^{-a_n \ell}) \\ &\leq \max\left(-\inf_{z \in C_\ell \cap C} K(z), -\ell\right) \\ &\leq \max\left(-\inf_{z \in C} K(z), -\ell\right). \end{aligned}$$

Schickt man $\ell \rightarrow \infty$, erhält man die obere Schranke.

4. Schritt: Die untere Schranke benutzt ein lokales Argument (das ist typisch) und ist viel einfacher (das ist weniger typisch). Wir müssen zeigen, dass für jedes $z \in Z$ und eine Umgebung N von Z gilt:

$$\underline{\lim}_{a_n} \frac{1}{a_n} \log \mathbb{R}_n(N) \geq -K(z). \quad (15)$$

Ist N nun eine Umgebung von z . Dann enthält N eine Menge $U \times V$, wobei U und V Umgebungen von x bzw. y sind. Da \mathbb{P}_n und \mathbb{Q}_n ein LDP und somit die untere Schranke für U bzw. V erfüllt, folgt (15). \square

In der Folge wollen wir uns mit Möglichkeiten befassen, ein LDP zu zeigen. Wir beginnen mit einer Technik, die allein auf topologischen Eigenschaften beruht.

Satz 1.22 *Falls (\mathbb{P}_n) einem LDP mit Geschwindigkeit (a_n) und Rate I auf einem polnischen Raum X genügt underline*

$$F : X \rightarrow Y$$

stetig ist (dabei ist Y ein weiterer polnischer Raum), so genügt auch die Folge $\mathbb{Q} = \mathbb{P}_n F^{-1} = \mathbb{P}_n^F$ einem LDP mit Geschwindigkeit (a_n) und Rate

$$J(y) = \inf_{x: F(x)=y} I(x).$$

Beweis: Zunächst sei bemerkt, dass J klarerweise nicht-negativ ist. Ferner nimmt die unterhalb-stetige Funktion I auf jeder ihrer kompakten Niveaumengen ihr Minimum an, das Infimum in der Definition von J ist also ein Minimum. Somit gilt für die Niveaumengen von J :

$$\{y : J(y) \leq L\} = \{y : \inf_{x: F(x)=y} I(x) \leq L\} = \{F(x) : I(x) \leq L\} = F(N_L),$$

wobei N_L die Niveaumenge zum Niveau L der Ratenfunktion I ist. Da N_L kompakt ist, ist $\{y : J(y) \leq L\}$ dies auch.

Sei $X \subseteq Y$ abgeschlossen. Dann ist auch $F^{-1}[C] \subseteq X$ abgeschlossen und es gilt

$$\begin{aligned} \overline{\lim} \frac{1}{a_n} \log Q_n(C) &= \overline{\lim} \frac{1}{a_n} \log \mathbb{P}_n(F_n^{-1}[C]) \\ &\leq - \inf_{x \in F^{-1}[C]} I(x) \\ &= - \inf_{y \in C} \inf_{x: F(x)=y} I(x) \\ &= - \inf_{y \in C} J(y). \end{aligned}$$

Analog ist für $G \subseteq Y$ offen auch $F^{-1}[G] \subseteq X$ offen und es gilt

$$\underline{\lim} \frac{1}{a_n} \log \mathbb{Q}_n(G) \geq - \inf_{x \in F^{-1}[G]} I(x) = - \inf_{y \in G} J(y).$$

□

Bemerkung 1.23 *Obwohl das obige Prinzip sehr einfach ist, ist es dennoch sehr nützlich. Die Hauptschwierigkeit besteht darin, dass die Ratenfunktion oft durch die Form in Satz 1.22 sehr implizit gegeben ist.*

Der folgende Satz bildete eine der Hauptmotivation für das Studium der großen Abweichungen durch Varadhan in den 1960er und 1970er Jahren.

Satz 1.24 (Varadhan): *Es sei $(X_n)_n$ eine Folge von Zufallsvariablen in \mathbb{R}^d , die einem LDP mit Geschwindigkeit n und Rate I genügt und*

$$f : \mathbb{R}^d \rightarrow \mathbb{R}$$

sei stetig und beschränkt. Dann gilt

$$\lim \frac{1}{n} \log \mathbb{E} e^{nf(X_n)} = \sup_x [f(x) - I(x)]$$

(wobei die Existenz des Limes gleich mitbehauptet wird).

Der Beweis dieses Satzes ist recht instruktiv:

Beweis: Da f stetig und beschränkt ist, lässt sich eine Überdeckung des \mathbb{R}^d durch endlich viele abgeschlossene Mengen A_1, \dots, A_M finden, so dass f auf jede dieser Mengen höchstens um ein vorgegebenes $\delta > 0$ variiert, also

$$\sup_{x \in A_i} f(x) - \inf_{x \in A_i} f(x) \leq \delta \quad \forall i = 1, \dots, M.$$

Dann folgt

$$\begin{aligned}
\int_{\mathbb{R}^d} e^{nf(X_n)} d\mathbb{P} &\leq \sum_{j=1}^M \int_{A_j} e^{nf(X_n)} d\mathbb{P} \\
&\leq \sum_{j=1}^M \int_{A_j} e^{n \sup_{y \in A_j} f(y)} d\mathbb{P}_{X_n}(y) \\
&\leq \sum_{j=1}^M \int_{A_j} e^{n(\inf_{y \in A_j} f(y) + \delta)} d\mathbb{P}_{X_n}(y) \\
&= \sum_{j=1}^M e^{n(\inf_{y \in A_j} f(y) + \delta)} \mathbb{P}(X_n \in A_j).
\end{aligned}$$

Also

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} e^{nf(X_n)} &\leq \sup_{1 \leq j \leq M} \left[\inf_{y \in A_j} f(y) + \delta - \inf_{y \in A_j} I(y) \right] \\
&\leq \sup_{1 \leq j \leq M} \left[\sup_{y \in A_j} (f(y) - I(y)) + \delta \right] \\
&= \sup_{y \in \mathbb{R}^d} (f(y) - I(y)) + \delta.
\end{aligned}$$

Lässt man δ gegen 0 gehen, ergibt sich die obere Schranke. Für die untere Schranke argumentiert man, wie oft in der Theorie der großen Abweichungen, lokal. Da f stetig ist, gibt es zu jedem $y_0 \in \mathbb{R}^d$ und $\varepsilon > 0$ eine offene Umgebung $U_\varepsilon(y_0)$, so dass für alle $y \in U_\varepsilon(y_0)$ gilt

$$f(y) \geq f(y_0) - \varepsilon.$$

Somit folgt

$$\mathbb{E} e^{nf(X_n)} \geq \int_{U_\varepsilon(y_0)} e^{nf(X_n)} d\mathbb{P} \geq e^{n(f(y_0) - \varepsilon)} \mathbb{P}(X_n \in U_\varepsilon(y_0)).$$

Mit Hilfe des LDP ergibt sich

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} e^{nf(X_n)} \geq f(y_0) - \varepsilon - \inf_{y \in U_\varepsilon(y_0)} I(y) \geq f(y_0) - I(y_0) - \varepsilon.$$

Da dies für alle y_0 stimmt, folgt

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} e^{nf(X_n)} \geq \sup_y [f(y) - I(y)].$$

□

Ganz analog zu Varadhans Lemma lässt sich der folgende Satz über große Abweichungen beweisen:

Satz 1.25 Die Folge von Zufallsvariablen (X_n) im \mathbb{R}^d genüge einem LDP mit Geschwindigkeit n und Ratenfunktion I . Für eine stetige beschränkte Funktion $F : \mathbb{R}^d \rightarrow \mathbb{R}$ setze

$$J_n(S) = \int_S e^{nF(x)} d\mathbb{P}_{X_n}(x), \quad S \subseteq \mathbb{R}^d \text{ Borelsch.}$$

Definiere weiter

$$\mathbb{P}_n^F(S) = \frac{J_n(S)}{J_n(\mathbb{R}^d)}; \quad S \in \mathcal{B}^d$$

die \mathbb{P}_n^F sind Wahrscheinlichkeitsmaße. Dann genügt die Folge der $(\mathbb{P}_n^F)_n$ einem LDP mit Geschwindigkeit n und Ratenfunktion

$$I^F(x) = -[F(x) - I(x)] + \sup_{y \in \mathbb{R}^d} [F(y) - I(y)].$$

Beweis: Es ist

$$\limsup \frac{1}{n} \log \mathbb{P}_n^F(S) = \limsup \frac{1}{n} \log J_n(S) - \limsup \frac{1}{n} \log J_n(\mathbb{R}^d)$$

und Analoges für \liminf . Nun wissen wir aus dem vorhergehenden Satz schon, dass

$$\lim \frac{1}{n} \log J_n(\mathbb{R}^d) = \sup_y [F(y) - I(y)]$$

gilt. Somit bleibt zu zeigen, dass für alle offenen Mengen G und alle abgeschlossenen Mengen A gilt

$$\liminf \frac{1}{n} \log J_n(G) \geq - \inf_{x \in G} [F(x) - I(x)]$$

und

$$\limsup \frac{1}{n} \log J_n(A) \leq - \inf_{x \in A} [F(x) - I(x)].$$

Dies aber geht genau wie im Beweis des Varadhanschen Lemmas. \square

Eine Anwendung des Satzes von Varadhan befasst sich mit einem Modell für Ferromagnetismus aus der statistischen Mechanik. Naiv lässt sich eine magnetische Substanz so modellieren, dass sie aus vielen Atomen aufgebaut ist, die alle einen magnetischen Dipol besitzen, der entweder nach oben zeigt (+1) oder nach unten (−1). Wenn alle Atome rein zufällig ihren Dipol wählen und nicht interagieren, so wird die durchschnittliche Magnetisierung

$$m_N(\sigma) := \frac{\sum_{i=1}^N \sigma_i}{N}$$

nach dem Gesetz der großen Zahlen nahezu 0 sein, wir sehen kein magnetisches Verhalten. Hierbei bezeichne σ_i den "Spin" des i -ten Atoms. Die Interaktion, die für das magnetische Verhalten notwendig ist, kodiert man in einer Energie- oder Hamiltonfunktion $H_N(\sigma)$ und wählt die Spinkonfigurationen $\sigma = (\sigma_i)_{i=1}^N$ gemäß dem sogenannten Gibbs-Maß aus:

$$\mu_N(\sigma) = \exp(-\beta H_N(\sigma)) / Z_N.$$

Dabei ist $\beta > 0$ ein zusätzlicher Parameter, die inverse Temperatur $\beta = \frac{1}{T}$, Z_N die Zustandssumme, macht μ_N zu einem Wahrscheinlichkeitsmaß

$$Z_N = \sum_{\sigma'} e^{-\beta H_N(\sigma')}.$$

Das Minuszeichen entspricht der physikalischen Konvention, dass Systeme versuchen, ihre Energie zu minimieren. Eine Möglichkeit, die zu einer möglichst gleichen Ausrichtung von Spins sorgt, ist es, interagierende Spins σ_i und σ_j miteinander zu multiplizieren. Im einfachsten Fall, dem so genannten Curie-Weiss-Modell, interagieren einfach alle Spins miteinander mit gleicher Stärke.

Dies führt zu der Energie- bzw. Hamiltonfunktion

$$H_N(\sigma) = -\frac{1}{2N} \sum_{1 \leq i, j \leq N} \sigma_i \sigma_j - h \sum_{i=1}^N \sigma_i.$$

Hierbei modelliert $h > 0$ ein externes Magnetfeld, dem die Spins ausgesetzt sind. Der große Vorteil des Curie-Weiss-Modells ist, dass die Energiefunktion eine Funktion des Ordnungsparameters m_N ist

$$H_N(\sigma) = -\frac{N}{2} [m_N(\sigma)]^2 - hN m_N(\sigma).$$

Diese Hamiltonfunktion hängt also von der mittleren Magnetisierung ab, man spricht auch von einem meanfield-Modell, einem Mittelwertmodell. Das zugehörige GIBBS-MASS ist von der Form

$$\mu_{N,\beta,h}(\sigma) = \frac{e^{\frac{\beta N}{2} m_N^2(\sigma) + N\beta h m_N(\sigma)}}{Z_{N,\beta,h}}$$

mit

$$Z_{N,\beta,h} = \sum_{\sigma} e^{\frac{\beta N}{2} m_N^2(\sigma) + N\beta h m_N(\sigma)}.$$

Um dies zu behandeln, erinnern wir an das LDP für i.i.d. Zufallsvariablen (X_i) , die $Ber(p)$ -verteilt sind auf $\{0, 1\}$. Dann genügt $\frac{S_n}{n}$ einem LDP mit Geschwindigkeit n und Rate

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

Betrachtet man anstelle der (X_i) eine Folge (Y_i) mit

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = \frac{1}{2},$$

so entspricht dies einer Transformation

$$Y_i = 2X_i - 1.$$

Somit erhalten wir

$$\mathbb{P}\left(\frac{\sum_{i=1}^n Y_i}{n} = y\right) = \mathbb{P}\left(2\frac{S_n}{n} - 1 = y\right) = \mathbb{P}\left(\frac{S_n}{n} = \frac{y+1}{2}\right).$$

Wir sehen also, dass mit $\frac{S_n}{n}$ auch $\frac{\sum_{i=1}^n Y_i}{n}$ einem LDP genügt und zwar mit Rate

$$\begin{aligned} I(y) &= H\left(\frac{y+1}{2} \middle| \frac{1}{2}\right) = \frac{y+1}{2} \log \frac{y+1}{2} + \frac{1-y}{2} \log \frac{1-y}{2} + \log 2 \\ &= \frac{(1+y)}{2} \log \frac{(1+y)}{2} + \frac{(1-y)}{2} \log(1-y). \end{aligned}$$

Nun können wir Satz 1.24 und Satz 1.25 anwenden, um das Curie-Weiss-Modell zu studieren.

Satz 1.26 *Für jede inverse Temperatur $\beta > 0$ und jedes magnetische Feld h gilt*

$$\begin{aligned} f_{\beta,h} &:= \lim_{N \rightarrow \infty} -\frac{1}{\beta N} \log Z_{N,\beta,h} = \inf_{m \in [-1,+1]} \left[-\frac{m^2}{2} - hm - \frac{1}{\beta} (\log 2 - I(m)) \right] \\ &= - \sup_{m \in [-1,+1]} \left[\frac{m^2}{2} + hm + \frac{1}{\beta} (\log 2 - I(m)) \right]. \end{aligned}$$

Hierbei ist

$$I(m) = \frac{1+m}{2} \log(1+m) + \frac{1-m}{2} \log(1-m).$$

Beweis: Es ist

$$Z_{N,\beta,h} = \sum_{\substack{\sigma_i \in \{\pm 1\} \\ \forall i=1,\dots,N}} e^{\frac{\beta N}{2} \left(\frac{\sum_{i=1}^N \sigma_i}{N}\right)^2 + N\beta h \left(\frac{\sum_{i=1}^N \sigma_i}{N}\right)} = 2^N \frac{1}{2^N} \sum_{\substack{\sigma_i \in \{\pm 1\} \\ \forall i=1,\dots,N}} e^{\frac{\beta N}{2} \left(\frac{\sum_{i=1}^N \sigma_i}{N}\right)^2 + N\beta h \left(\frac{\sum_{i=1}^N \sigma_i}{N}\right)}.$$

(Dies ist der Grund für das Auftreten des $\log 2$ -Terms im Ausdruck für $f_{\beta,h}$). Nun haben wir schon in einem Beispiel geklärt, dass $\frac{\sum_{i=1}^N \sigma_i}{N}$ unter dem Produktmaß einem LDP mit Geschwindigkeit N und Ratenfunktion $I(\cdot)$ genügt. Nun ist

$$F(x) = \frac{\beta x^2}{2} + \beta h x$$

natürlich im allgemeinen zwar stetig, aber nicht beschränkt. Nun lebt $\frac{\sum_{i=1}^N \sigma_i}{N}$ aber auf $[-1, +1]$. Dort ist $F(\cdot)$ sehr wohl beschränkt. Eine Anwendung des Varadhanschen Lemmas ergibt somit

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log(2^{-N} Z_{N,\beta,h}) = \sup_{m \in [-1,+1]} \left[\frac{\beta m^2}{2} + \beta h m - I(m) \right].$$

Dies ist äquivalent zu unserer Behauptung. □

Ebenso lässt sich zeigen, dass die Magnetisierung

$$m_N := \frac{1}{N} \sum_{i=1}^N \sigma_i$$

unter dem Gibbs-Maß

$$\mu_{N,\beta,h}(\sigma) = \frac{e^{-\beta H_N(\sigma)}}{Z_{N,\beta,h}}$$

einem LDP genügt.

Satz 1.27 $m_N(\cdot)$ genügt unter dem Gibbs-Maß $\mu_{N,\beta,h}$ einem LDP mit Geschwindigkeit N und Ratenfunktion

$$J(x) = -\frac{\beta x^2}{2} - \beta h x + I(x) + \sup_{y \in [-1,1]} \left[\frac{\beta y^2}{2} + \beta h y - I(y) \right],$$

wobei wieder

$$I(x) = \frac{1+x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x)$$

ist.

Beweis: Dies folgt direkt aus den Sätzen über große Abweichungen, die wir zuvor bewiesen haben, da $\mu_{N,\beta,h}$ genau die dortige exponentielle Struktur besitzt. \square

Das Schöne an Prinzipien der großen Abweichungen ist, dass sie auch Gesetze der großen Zahlen implizieren. Dieses Faktum haben wir schon in Wahrscheinlichkeitstheorie II kennengelernt. Es soll hier kurz wiederholt werden.

Satz 1.28 Eine Folge von Zufallsvariablen $(X_n)_n$ in \mathbb{R}^d genüge einem LDP mit Rate $I(\cdot)$ und Geschwindigkeit n . Dann gilt für die Menge der Nullstellen

$$\mathcal{N} = \{x \in \mathbb{R}^d : I(x) = 0\}$$

der Ratenfunktion und jedes $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \in \mathcal{N}_\varepsilon^c) < +\infty.$$

Hierbei ist

$$\mathcal{N}_\varepsilon = \{x : \|x - \mathcal{N}\| < \varepsilon\}.$$

Ist insbesondere \mathcal{N} einelementig, so folgt für $\nu \in \mathcal{N}$

$$\mathbb{P}(X_n \rightarrow \nu) = 1.$$

Beweis: Da $\mathcal{N}_\varepsilon^c$ abgeschlossen ist, I von unten halbstetig und \mathcal{N} die Menge der globalen Minima von I ist, folgt

$$a := \inf_{x \in \mathcal{N}_\varepsilon^c} I(x) > 0.$$

Aus der oberen Abschätzung der großen Abweichungen folgt, dass für hinreichend große n

$$\mathbb{P}(X_n \in \mathcal{N}_\varepsilon^c) \leq e^{-na/2}$$

gilt. Somit ist

$$\sum_n \mathbb{P}(X_n \in \mathcal{N}_\varepsilon^c) < +\infty.$$

Die fast sichere Konvergenz ist eine unmittelbare Konsequenz dieser Summierbarkeit und des Borel-Cantelli-Lemmas. \square

Wir müssen uns somit, um die Minima von

$$J(x) = -\frac{\beta x^2}{2} - \beta h x + I(x) + \sup_{y \in [-1, +1]} \left[\frac{\beta y^2}{2} + \beta h y - I(y) \right]$$

mit

$$I(x) = \frac{1+x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x)$$

kümmern. Diese Minima erfüllen also

$$I'(x) = \frac{1}{2} \log \frac{1+x}{1-x} = \beta(x+h),$$

also

$$e^{2\beta(x+h)} = \frac{1+x}{1-x}.$$

Durch langes Hinschauen erkennt man, dass dies äquivalent ist zu

$$x = \frac{e^{2\beta(x+h)} - 1}{e^{2\beta(x+h)} + 1} = \tanh(\beta(x+h)).$$

Man unterscheidet nun verschiedene Fälle:

- Ist $h \neq 0$ und β klein ($\beta \leq 1$), so hat die Gleichung nur eine Lösung, diese ist ein Minimum von J
- Ist $h > 0$ und β groß ($\beta > 1$), so hat die Gleichung drei Lösungen, von denen die kleinste positive ein Minimum von J ergibt.
- Ist $h < 0$ und β groß ($\beta > 1$), so hat die Gleichung drei Lösungen, von denen die kleinste negative ein Minimum von J ergibt.
- Für $h = 0$ ist die Situation symmetrisch zum Ursprung. Für $\beta \leq 1$ ist $x = 0$ die einzige Lösung dieser Gleichung und liefert somit ein Minimum von $J(\cdot)$. Für $\beta > 1$ hat die Gleichung allerdings drei Lösungen, von denen die Lösung $x = 0$ diesmal ein Maximum von $J(\cdot)$ liefert, die Lösungen, die verschieden sind von 0 aber Minima.

Zusammen erhält man:

Satz 1.29 *Im Curie-Weiss-Modell gelten für die Magnetisierung m_N die folgenden Grenzwertsätze unter $\mu_{N,\beta,h}$.*

1. Ist $h > 0$, so konvergiert m_N exponentiell schnell gegen die kleinste positive Lösung von

$$x = \tanh(\beta(x+h)).$$

2. Ist $h < 0$, so konvergiert m_N exponentiell schnell gegen die größte negative Lösung von

$$x = \tanh(\beta(x + h)).$$

3. Ist $h = 0$ und $\beta \leq 1$, so konvergiert m_N exponentiell schnell gegen 0.

4. Ist $h = 0$ und $\beta > 1$, so konvergiert m_N exponentiell schnell gegen die beiden von Null verschiedenen Lösungen von

$$x = \tanh(\beta x).$$

5. Insbesondere erhält man

$$\lim_{h \downarrow 0} \lim_{N \rightarrow \infty} \mu_{N,\beta,h} \circ m_N^{-1} \Rightarrow \delta_{m^*(\beta)},$$

wobei $m^*(\beta)$ die größte Lösung von

$$x = \tanh(\beta x)$$

ist.

5. ist in der physikalischen Literatur auch als “spontane Magnetisierung” bekannt: Ein Material, das in ein Magnetfeld gehalten wird, merkt sich dies bei genügend tiefen Temperaturen und wird selbst magnetisch. Bei hohen Temperaturen bleibt dieses Phänomen aus.

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 Erdős-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 gäbe 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 eisen 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” Münzwurffreihen. Dahinter steht der folgende Sachverhalt:

Satz 1.30 *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

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

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

Korollar 1.31 *In der obigen Situation gilt*

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

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:

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

Offenbar gilt

$$\{T_r \leq m\} = \{R_m \geq r\}.$$

Sei für $l, k \in \mathbb{N}$, $l > k$

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

Dann ist

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

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

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

für eine Konstante $c > 0$.

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

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

Also ist nach Borel-Cantelli

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

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

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

Für die andere Richtung sei

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

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

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

Also

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

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

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

Also ist

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

und somit auch

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

für fast alle m \mathbb{P} -f.s.

Wir kehren zum Satz von Cramér zurück. Dieser galt für Folgen von iid Zufallsvariablen. Wir können versuchen, exponentielle Abschätzungen auch unter schwächeren Voraussetzungen zu erhalten und entweder auf die identische Verteilung oder die Unabhängigkeit zu verzichten. Wir bekommen allerdings zusätzliche Bedingungen. Wie üblich ist eine Kontrolle über die momenten-erzeugende Funktion wichtig:

Lemma 1.32 (Hoeffding) *Sei X eine Zufallsvariable im Intervall $[a, b]$. Dann gilt für alle $t > 0$*

$$\mathbb{E}e^{tX} \leq e^{t\mathbb{E}X} e^{t^2(b-a)^2/8}. \quad (16)$$

Beweis: Zunächst können wir $\mathbb{E}X = 0$ voraussetzen, sonst ersetzen wir X durch $X - \mathbb{E}X$. Nun ist

$$X \mapsto e^{tx}$$

für alle $t > 0$ konvex. Also gilt

$$e^{tx} \leq \frac{b-x}{b-a} e^{ba} + \frac{x-a}{b-a} e^{tb}$$

für alle $x \in [a, b]$ Daher:

$$\begin{aligned} \mathbb{E}e^{tX} & \leq \frac{b - \mathbb{E}X}{b-a} e^{ba} + \frac{\mathbb{E}X - a}{b-a} e^{tb} \\ & = \frac{b}{b-a} e^{ta} - \frac{a}{b-a} e^{tb} \\ & = \left(-\frac{a}{b-a}\right) e^{ta} \left(-\frac{b}{a} + e^{tb-ta}\right) \\ & = \left(-\frac{a}{b-a}\right) e^{ta} \left(-\frac{b-a+a}{a} + e^{t(b-a)}\right) \\ & = \left(-\frac{a}{b-a}\right) e^{ta} \left(-\frac{b-a}{a} - 1 + e^{t(b-a)}\right) \\ & = (1 - s + se^{t(b-a)}) e^{-ts(b-a)} \end{aligned}$$

mit $s = -\frac{a}{b-a} > 0$, da $a < 0$.

Mit $u := t(b-a)$ sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch

$$\varphi(u) = -su + \log(1 - s + se^u)$$

(das ist wohldefiniert, Übung). Somit

$$\mathbb{E}e^{tX} \leq e^{\varphi(u)}.$$

Wir Taylor-entwickeln φ :

$$\varphi(u) = \varphi(0) + u\varphi'(0) + \frac{u^2}{2}\varphi''(v)$$

für ein $0 \leq v \leq u$. Nun ist $\varphi(0) = 0$,

$$\varphi'(0) = -s + \frac{se^u}{1 - s + se^u}|_{u=0} = -s + s = 0,$$

und

$$\varphi''(v) = \frac{se^v}{1 - s + se^v} \left(1 - \frac{se^v}{1 - s + se^v}\right) = \Theta(1 - \Theta) \leq \frac{1}{4}$$

für $\Theta = \frac{se^v}{1 - s + se^v}$. Daher folgt

$$\varphi(u) \leq \frac{u^2}{8} = \frac{1}{8}t^2(b-a)^2.$$

Daher ist

$$\mathbb{E}e^{tX} \leq e^{\frac{1}{8}t^2(b-a)^2}$$

wie behauptet. □

Dies hat eine obere Schranke für die Wahrscheinlichkeiten nicht notwendig identisch verteilter, unabhängiger Zufallsvariablen zur Folge:

Satz 1.33 (Chernov, Hoeffding) Seien X_1, \dots, X_n unabhängige Zufallsvariablen, so dass $|X_i| \in [a_i, b_i]$ für alle i . und sei $S_n = \sum_{i=1}^n X_i$. Dann gilt

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq \lambda\sigma) \leq 2e^{-2\lambda^2}$$

mit $\sigma^2 = \sum_{i=1}^n (b_i - a_i)^2$. Mit anderen Worten: Für alle positiven λ gilt:

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq \lambda) \leq 2e^{-2\frac{\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

Beweis: Aus Symmetriegründen genügt es zu zeigen, dass für alle positiven λ

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq \lambda\sigma) \leq e^{-2\lambda^2}$$

bzw.

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq \lambda) \leq e^{-2 \frac{\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

gilt.

Mit der Markovschen Ungleichung erhalten wir für alle positiven $t > 0$:

$$\begin{aligned} \mathbb{P}(S_n - \mathbb{E}S_n \geq \lambda) &\leq e^{-t\lambda} \mathbb{E} \exp(t(S_n - \mathbb{E}S_n)) \\ &= e^{-t\lambda} \prod_{i=1}^n \mathbb{E} e^{tX_i - \mathbb{E}X_i}. \end{aligned}$$

Die Erwartungswerte in dem Produkt auf der rechten Seite verarzten wir mit dem Hoeffding-Lemma Lemma 1.32. Aus diesem folgt:

$$\mathbb{E} e^{tX_i - \mathbb{E}X_i} \leq e^{t^2 \frac{(b_i - a_i)^2}{8}}$$

für alle i . Also

$$\begin{aligned} \mathbb{P}(S_n - \mathbb{E}S_n \geq \lambda) &\leq e^{-t\lambda} \prod_{i=1}^n e^{t^2 \frac{(b_i - a_i)^2}{8}} \\ &= e^{-t\lambda + t^2 \sum_{i=1}^n \frac{(b_i - a_i)^2}{8}}. \end{aligned}$$

Wählt man nun $t = \frac{4\lambda}{\sigma^2}$, so ergibt sich die Behauptung.

□

Wir versuchen als nächstes, die Unabhängigkeit abzuswächen. Eine offensichtliche Art dies zu tun ist anzunehmen, dass S_n keine Summe von iid Zufallsvariablen, aber von Martingaldifferenzen ist.

Satz 1.34 (Azuma) Seien X_1, \dots, X_n Zufallsvariablen mit $|X_i| \leq 1$ für alle i . Wir nehmen an, dass diese die Martingaldifferenz-Eigenschaft haben:

$$\mathbb{E}[X_i | X_1, \dots, X_{i-1}] = 0.$$

Dann gilt für alle $\lambda > 0$ und $S_n = \sum X_i$

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq Ce^{-c\lambda^2}$$

für positive Konstanten $c, C > 0$.

Beweis: Aufgrund von Symmetrie genügt es,

$$\mathbb{P}(S_n \geq \lambda\sqrt{n}) \leq Ce^{-c\lambda^2}$$

zu zeigen. Nun ist offenbar $|S_n| \leq \sqrt{n}$, sodass die Einschränkung $\lambda \leq \sqrt{n}$ angenommen werden kann. Wieder betrachten wir die moment-erzeugende Funktion $\mathbb{E}e^{tS_n}$ für $t > 0$. Dann

$$\mathbb{E}e^{tS_n} = \mathbb{E}e^{tS_{n-1}} e^{tX_n},$$

was aufgrund fehlender Unabhängigkeit nicht faktorisiert. Aber es gilt

$$\mathbb{E}e^{tS_1} = \mathbb{E}\mathbb{E}[e^{tS_{n-1}}e^{tX_n}|X_1, \dots, X_{n-1}] = \mathbb{E}e^{tS_{n-1}}\mathbb{E}[e^{tX_n}|X_1, \dots, X_{n-1}].$$

Nun überlegt man sich (zum Beispiel, indem man es als Übungsaufgabe löst), dass Hoeffdings Abschätzung aus Lemma 1.32 auch für bedingte Erwartungen gilt. Somit

$$\mathbb{E}[e^{tX_n}|X_1, \dots, X_{n-1}] \leq e^{Kt^2}$$

für eine Konstante $K > 0$. Somit

$$\mathbb{E}e^{tS_n} = \mathbb{E}e^{tS_{n-1}}e^{Kt^2}.$$

Führt man das nun n -mal durch, um von $\mathbb{E}e^{tS_n}$ über $\mathbb{E}e^{tS_{n-1}}$ bis zu $\mathbb{E}e^{tX_1}$ zu kommen, erhalten wir

$$\mathbb{E}e^{tS_n} \leq e^{Knt^2}.$$

Somit folgt wieder mit der Markov-Ungleichung:

$$\mathbb{P}(S_n \geq \lambda\sqrt{n}) \leq e^{-t\lambda\sqrt{n}}\mathbb{E}e^{tS_n} \leq e^{-t\lambda\sqrt{n}}e^{Knt^2}.$$

Optimieren in t gibt das Gewünschte. □

Die folgenden beiden Sätze zeigen, dass man noch wesentlich weiter gehen kann: Man kann iid Zufallsvariablen auf eine ‐Lipschitz‐-Weise miteinander kombinieren.

Satz 1.35 *Es seien X_1, \dots, X_n unabhängige Zufallsvariablen so dass das Bild von X_i R_i ist. Sei*

$$F : R_1 \times \dots \times R_n \rightarrow \mathbb{R}$$

mit der Eigenschaft, dass Änderungen nur im i -ten Eintrag den Funktionswert um höchstens c_i ändern, also

$$|F(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - F(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| \leq c_i$$

für alle $x_j \in R_j$, $j = 1, \dots, n$, $x'_i \in R_i$. Dann gilt für alle $\lambda > 0$:

$$\mathbb{P}(|F(X) - \mathbb{E}[F(X)]| \geq \lambda\sigma) \leq 2e^{-2\lambda^2}$$

mit $\sigma^2 = \sum_{i=1}^n c_i^2$.

Beweis: Wieder genügt es aus Symmetriegründen, die einseitige Abschätzung zu betrachten und zu zeigen

$$\mathbb{P}(F(X) - \mathbb{E}F(X) \geq \lambda\sigma) \leq e^{-2\lambda^2/\sigma^2}$$

bzw.

$$\mathbb{P}(F(X) - \mathbb{E}F(X) \geq \lambda) \leq e^{-2\lambda^2/\sigma^2}.$$

Definiere

$$Z_i = \mathbb{E}[F(X)|(X_1, \dots, X_i)].$$

Dann ist $Z_0 = \mathbb{E}F(X)$ und $Z_m = F(X)$. Beachte, dass

$$\mathbb{E}[(Z_i - Z_{i-1})|(X_1, \dots, X_{i-1})] = 0.$$

Ferner variiert F bei festgehaltenem X_1, \dots, X_{i-1} und ausintegrierten X_{i+1}, \dots, X_n nur um c_i . Aus Hoeffdings Lemma haben wir daher

$$\mathbb{E}[e^{t(Z_i - Z_{i-1})}|(X_1, \dots, X_{i-1})] \leq e^{t^2 c_i^2 / 8}.$$

Also:

$$\begin{aligned} \mathbb{P}[F(X) - \mathbb{E}F(X) \geq \lambda] &\leq e^{-t\lambda} \mathbb{E}e^{t(F - \mathbb{E}F)} \\ &= e^{-t\lambda} \mathbb{E}e^{t \sum_{i=1}^n Z_i - Z_{i-1}} \\ &= e^{-t\lambda} \mathbb{E}\mathbb{E}[e^{t \sum_{i=1}^n Z_i - Z_{i-1}} | X_1, \dots, X_{n-1}] \\ &= e^{-t\lambda} \mathbb{E}[e^{t \sum_{i=1}^{n-1} Z_i - Z_{i-1}} \mathbb{E}[e^{t(Z_n - Z_{n-1})} | X_1, \dots, X_{n-1}]] \\ &\leq e^{-t\lambda} e^{t^2 c_n^2 / 8} \mathbb{E}e^{t \sum_{i=1}^{n-1} Z_i - Z_{i-1}}. \end{aligned}$$

Setzt man das fort, ergibt sich

$$\mathbb{P}[F(X) - \mathbb{E}F(X) \geq \lambda] \leq e^{-t\lambda} e^{\frac{t^2}{8} \sigma^2}.$$

Wähle $t = \frac{4\lambda}{\sigma^2}$, so ergibt sich das Ergebnis. \square

Das folgende Resultat ist die derzeit vermutilch allgemeinste Konzentrationsungleichung.

Satz 1.36 (Talagrand) *Seien X_1, \dots, X_n unabhängige Zufallsvariablen in $[-1, 1]$. Sei $F : \mathbb{R}^n \rightarrow \mathbb{R}$ eine 1-Lipschitz-Funktion. Dann gilt für den Median IMF und den Erwartungswert $\mathbb{E}F(X)$*

$$\mathbb{P}(|F(X) - \text{IMF}(X)| \geq \lambda) \leq Ce^{-c\lambda^2}$$

und

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq \lambda) \leq Ce^{-c\lambda^2}.$$

Beweis: Wir geben nur einen Beweis der zweiten Ungleichung und hier nur die Abschätzung der oberen Abweichung. Dieser Beweis hat aber den Vorteil, strukturell verständlicher zu sein als das Ursprungswerk von Talagrand. Wir beginnen mit einer so genannten Log-Sobolev-Ungleichung.

Lemma 1.37 (Log-Sobolev-Ungleichung) *Sei $F : \mathbb{R}^n \rightarrow \mathbb{R}$ eine glatte, konvexe Funktion. Dann gilt*

$$\mathbb{E}F(X)e^{F(X)} \leq (\mathbb{E}e^{F(X)})(\log \mathbb{E}e^{F(X)}) + C\mathbb{E}[e^{F(X)}|\nabla F(X)|^2]$$

für eine Konstante $C > 0$, die nicht von n abhängt.

Beweis: Wir beginnen eindimensional: Sei Y eine unabhängige Kopie von X . Dann ist

$$\mathbb{E}[F(X)e^{F(X)}] = \frac{1}{2}\mathbb{E}[(F(X) - F(Y))(e^{F(X)} - e^{F(Y)})] + (\mathbb{E}F(X))(\mathbb{E}e^{F(X)}).$$

Da F konvex ist folgt mit Jensen:

$$\mathbb{E}F(X) \leq \log \mathbb{E}e^{F(X)},$$

also genügt es zu zeigen, dass gilt:

$$\mathbb{E}[(F(X) - F(Y))(e^{F(X)} - e^{F(Y)})] \leq 2C\mathbb{E}[e^{F(X)}|\nabla F(X)|^2].$$

Nun ist F konvex und X und Y sind beschränkt, also ist

$$F(X) - F(Y) \leq C_1|\nabla F(X)|.$$

Mit F ist auch e^F konvex und daher ist

$$e^{F(X)} - e^{F(Y)} \leq C_2|\nabla F(X)|e^{F(X)}.$$

Ist $F(X) \geq F(Y)$ ergibt das

$$(F(X) - F(Y))(e^{F(X)} - e^{F(Y)}) \leq Ce^{F(X)}|\nabla F(X)|^2.$$

Ist $F(Y) \geq F(X)$ vertauschen wir die Rolle von X und Y .

Der allgemeine Fall folgt per Induktion. Dafür sei

$$X' = (X_1, \dots, X_{n-1}) \quad \text{und} \quad X = (X', X_n).$$

Aus der Induktionsvoraussetzung wissen wir, dass

$$\mathbb{E}[F(X)e^{F(X)}|X_n] \leq (X_n)e^{f(X_n)} + C\mathbb{E}[e^{F(X)}|\nabla' F(X)|^2|X_n],$$

wobei ∇' den $(n-1)$ -dim Gradienten darstellt und $f(X_n) = \log \mathbb{E}[e^{F(X)}|X_n]$ ist. Durch Erwartungswert-Bildung folgt:

$$\mathbb{E}F(X)e^{F(X)} \leq \mathbb{E}f(X_n)e^{f(X_n)} + C\mathbb{E}[e^{F(X)}|\nabla' F(X)|^2]. \quad (17)$$

Da F konvex ist, kann man mit der Hölderschen Ungleichung zeigen, dass auch f konvex ist und das gilt

$$\mathbb{E}e^{f(X_n)} = \mathbb{E}e^{F(X)}.$$

Aus dem eindimensionalen Fall erhalten wir:

$$\mathbb{E}[f(X_n)e^{f(X_n)}] \leq [\mathbb{E}e^{F(X)}][\log \mathbb{E}e^{F(X)}] + C \cdot \mathbb{E}[e^{f(X_n)}(f'(X_n))^2].$$

Nach Kettenregel ergibt sich

$$e^{f(X_n)}|f'(X_n)|^2 = e^{-f(X_n)}|Ee^{F(X)}F_X(X)|^2,$$

wobei F_{X_n} die Ableitung von F in x_n -Richtung ist. Aus der Cauchy-Schwarz-Ungleichung erhalten wir somit

$$e^{f(X_n)} |f'(X_n)|^2 \leq \mathbb{E} e^{F(X)} |F_{x_n}(X)|^2.$$

Daher:

$$\mathbb{E}[f(X_n) e^{f(X_n)}] \leq (\mathbb{E} e^{F(X)}) (\log \mathbb{E} e^{F(X)}) + C \mathbb{E} e^{F(X)} |F_{x_n}(X)|^2.$$

Setzt man das in (??) ein, folgt die Behauptung. \square

Jetzt können wir uns an den Beweis von Satz 1.36 begeben (bzw. das, was wir davon zeigen wollen):

Beweis von Satz 1.36:

Sei F und 1-Lipschitz-stetig. Damit ist es auch differenzierbar. Wir wenden das obige Lemma auf die Funktion tF an:

$$\mathbb{E}[etF(X)e^{tF(X)}] \leq (\mathbb{E} e^{tF(X)}) (\log \mathbb{E} e^{tF(X)}) + Ct^2 \mathbb{E} e^{tF(X)}.$$

Mit

$$H(t) := \mathbb{E} e^{tF(X)}$$

bedeutet dies:

$$tH'(t) \leq H(t) \log H(t) + Ct^2 H(t)$$

oder

$$\frac{d}{dt} \left(\frac{1}{t} \log H(t) \right) \leq C.$$

Nun ist

$$H(t) = \mathbb{E} e^{tF(X)} = \mathbb{E}[1 - tF(X) + 0(t^2)],$$

sodass

$$\frac{1}{t} \log H(t) \rightarrow \mathbb{E} F(X) \quad \text{für } t \rightarrow 0$$

gilt. Daher gilt

$$\frac{1}{t} \log H(t) = \mathbb{E} F(X) + Ct$$

für alle $t > 0$. Mit anderen Worten:

$$\mathbb{E} e^{tF(X)} \leq \exp(t\mathbb{E} F(X) + Ct^2).$$

Wenden wir wieder die exponentielle Markov-Ungleichung an, so ergibt sich

$$\begin{aligned} \mathbb{P}(F(X) - \mathbb{E} F(X) > \lambda) &= \mathbb{P}(F(X) > \lambda + \mathbb{E} F(X)) \\ &\leq e^{-t(\lambda + \mathbb{E} F(X))} \mathbb{E} e^{tF(X)} \\ &\leq e^{-t\lambda + Ct^2}. \end{aligned}$$

Wählt man t in der Ordnung von λ , also $t = \frac{\lambda}{2c}$, so folgt die Behauptung. \square

2 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 2.1 *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 (Ω, \mathcal{F}, P) definierte Zufallsgröße $X : \Omega \rightarrow I$ nennt man I -wertige Zufallsgröße.*

Definition 2.2 *Eine endlich oder unendlich lange Folge X_0, X_1, X_2, \dots 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 $i_0, i_1, \dots, i_n, i_{n+1} \in I$ mit $P(X_0 = i_0, \dots, X_n = i_n) > 0$*

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

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

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

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

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

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

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 $i_0, i_1, \dots, i_{n+1} \in I$. Ist $P(X_0 = i_0, \dots, X_n = i_n) = 0$, so gilt die behauptete Formel ebenfalls für $n + 1$: Ist $P(X_0 = i_0, \dots, X_n = i_n) > 0$, so folgt aus Definition 2.2

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

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

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

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

Bemerkung 2.4 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, \dots eine Folge I -wertiger Zufallsgrößen, die die Eigenschaft aus Satz 2.3 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

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

für alle $n \in \mathbb{N}_0$ und $i_0, \dots, i_n \in I$. 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 2.5 *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 $X_i : \Omega \rightarrow I$ für alle $i \in \{0, 1, \dots, N\}$, so dass X_0, \dots, X_N eine homogene Markoff-Kette mit Startverteilung ν und Übergangsmatrix \mathbb{P} ist.*

Beweis. Es sei $\Omega := I^{\{0, \dots, N\}}$ und $p(i_0, \dots, i_N) := \nu(i_0)p_{i_0 i_1} \dots p_{i_{N-1} i_N}$ 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 $(i_0, \dots, i_n) \in I^{\{0, \dots, n\}}$

$$\begin{aligned} P(X_0 = i_0, \dots, X_n = i_n) &= \sum_{(i_{n+1}, \dots, i_N) \in I^{\{n+1, \dots, N\}}} P(X_0 = i_0, \dots, X_N = i_N) \\ &= \sum_{(i_{n+1}, \dots, i_N) \in I^{\{n+1, \dots, N\}}} \nu(i_0)p_{i_0 i_1} \dots p_{i_{N-1} i_N} \\ &= \nu(i_0)p_{i_0 i_1} \dots p_{i_{n-1} i_n}. \end{aligned}$$

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

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

□

Bemerkung. Nachfolgend soll stets von einer unendlich langen Markoff-Kette ausgegangen werden, dies jedoch nur wegen einer bequemen Notation. Alle nachfolgenden Überlegungen benötigen die Konstruktion einer unendlichen Markoff-Kette nicht, sondern kommen damit aus, dass für jedes N eine Kette gemäß Satz 2.5 konstruiert werden kann.

Beispiel 2.6 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 2.5 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 2.7 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 2.8 Sei X_0, X_1, X_2, \dots 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

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

und mit der Definition 2.2 folgt

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

Somit gilt

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

□

Lemma 2.9 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. □

Lemma 2.10 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 2.3 a) folgt

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

□

Definition 2.11 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 2.11 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 2.9.

Die durch $i \sim j \Leftrightarrow (i \rightsquigarrow j \text{ und } j \rightsquigarrow i)$ 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 2.12 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 2.13 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 2.14 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 $\{0, \dots, n\}$ mit absorbierenden Rändern gibt es drei Äquivalenzklassen, nämlich $\{0\}$, $\{1, \dots, n-1\}$ und $\{n\}$. Die Mengen $\{0\}$ und $\{n\}$ sind abgeschlossen, und es gelten $\{1, \dots, n-1\} \rightsquigarrow \{n\}$ und $\{1, \dots, n-1\} \rightsquigarrow \{0\}$.
- 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 2.15 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

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

Eingesetzt ergibt das

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

mit

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

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 2.15 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 2.16 *Es sei I eine endliche, mindestens zweielementige Menge und ν, ϱ seien Wahrscheinlichkeiten auf I mit $\varrho(i) > 0$ für alle $i \in I$. 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 2.17 *In der Situation von Definition 2.15 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 asymptotische 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 2.18 *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

$$\nu \mathbb{P}^n \rightarrow_{n \rightarrow \infty} \varrho,$$

wobei ϱ eine Wahrscheinlichkeit auf I ist, die der Gleichung

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

genügt.

Bemerkung 2.19 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 2.18 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 2.20 Sei Q eine stochastische $r \times r$ Matrix. Dann existiert

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

und es gilt

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

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 $l \rightarrow \infty$ verschwinden, ergibt sich

$$QH = HQ = H. \tag{18}$$

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 2.17 zu beweisen.

Beweis von Satz 2.17 Wie wir eben gesehen haben, existiert eine stationäre Verteilung ϱ bzgl. \mathbb{P} , nämlich beispielsweise eine Zeile des entsprechend Satz 2.19 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 klassischen Sinne.

Um diese Konvergenz schließlich zu zeigen, verwenden wir die Entropiefunktion aus Definition 2.15 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 $(H(\nu\mathbb{P}^n|\varrho))_n$ 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

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

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

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

Insgesamt ist also

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

und daher

$$\varrho' = \varrho.$$

□

Beispiel 2.21 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 2.6 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 $\{S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0\}$ mit $P(\{S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0\}) > 0$ gilt

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

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 entspricht 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 $S_0 = 0, S_1, \dots, S_n$ 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

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

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

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

und X_i nehmen iid. und gleichverteilt die Werte $\pm e_i$, $i = 1, \dots, d$ 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 Ursprung 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 2.22 *Es gilt*

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

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

Bemerkung 2.23 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

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

also ist die Irrfahrt transient. Man rechnet dann

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

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

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

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

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

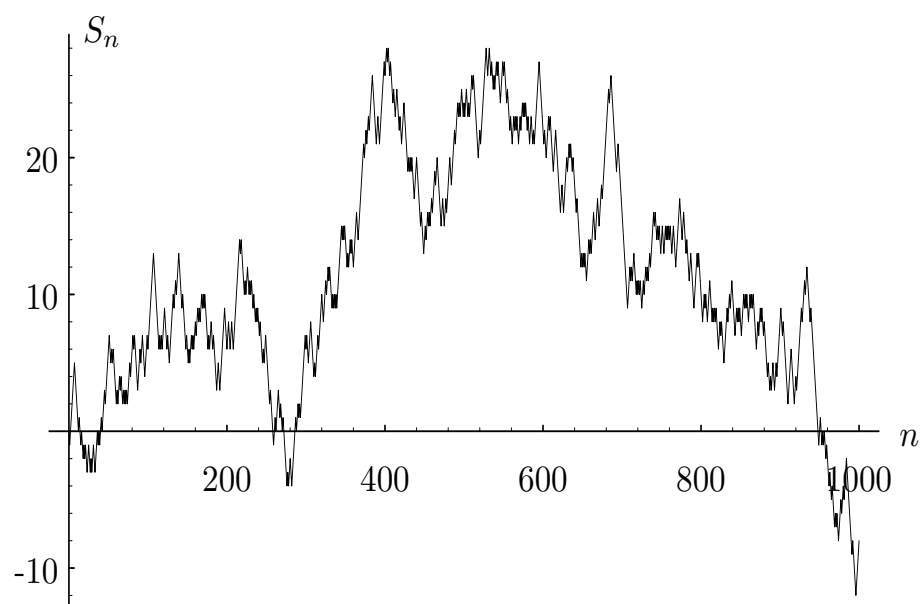
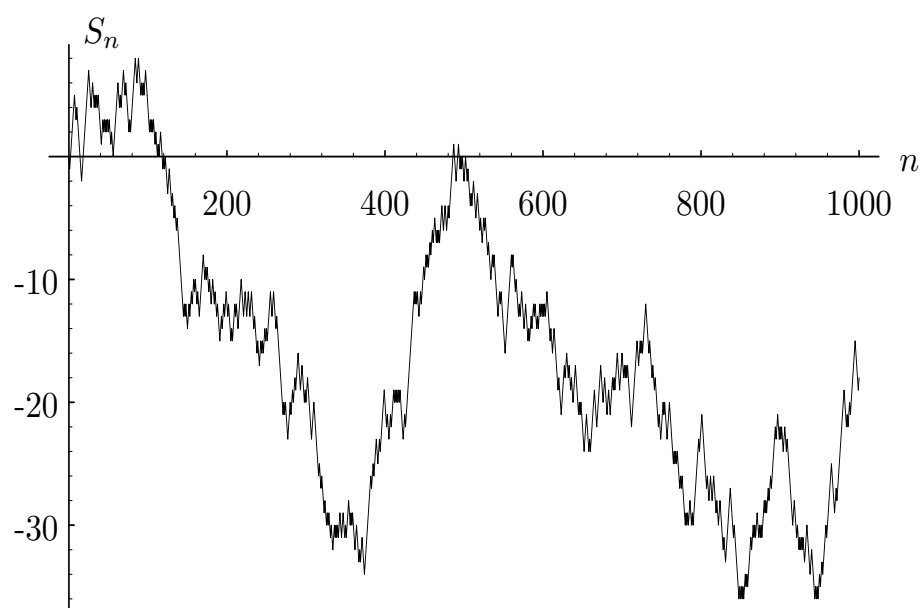
Die letzte Summe ist als Summe über die Wahrscheinlichkeiten einer Multinomialverteilung 1. Das Maximum wird bei $|k_j - \frac{n}{d}| \leq 1$ angenommen. Die Stirlingformel liefert daher

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

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



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. Durchschnitte 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 2.24

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

d.h.

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

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

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

$$3. u_{2k} = \sum_{j=1}^k f_{2j} u_{2k-2j}.$$

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 $(s_i, s_{i+1}, \dots, s_j)$ die x -Achse *berührt*, falls ein k mit $i \leq k \leq j$ existiert, für das $s_k = 0$ ist.

Lemma 2.26 (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

$$(s_i = -a, s_{i+1}, \dots, s_{j-1}, s_j = b)$$

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

$$(-s_i, -s_{i+1}, \dots, -s_{\tau-1}, s_\tau = 0, s_{\tau+1}, \dots, s_j = b)$$

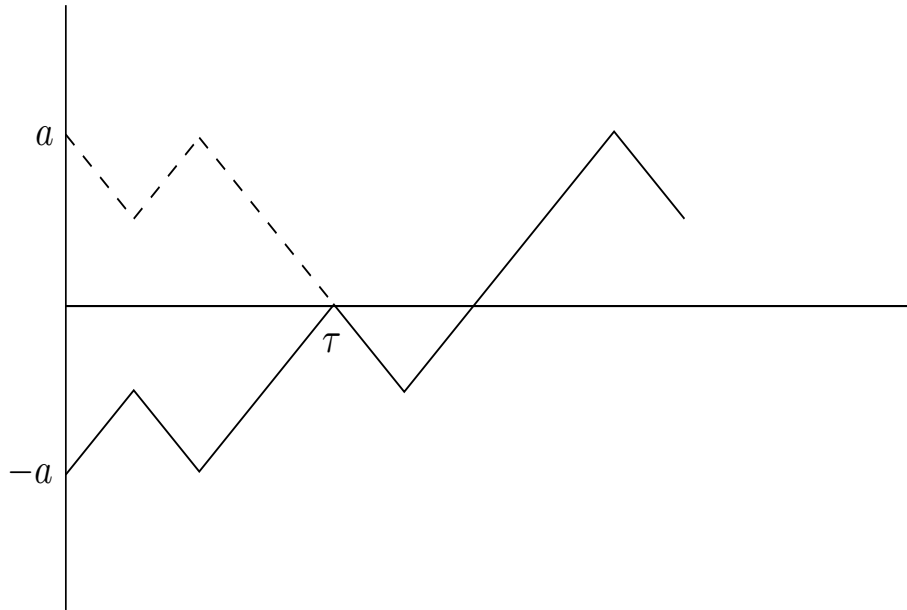
ein Pfad von (i, a) nach (j, b) , der die x -Achse berührt, und die Zuordnung ist bijektiv. \square

Das Spiegelungsprinzip werden wir nun verwenden, um die Menge der Pfade, die nach $2k$ Schritten zum ersten Mal wieder die x -Achse berühren abzzählen.

Satz 2.27 1. *Es gibt $\frac{1}{p} \binom{2p-2}{p-1}$ Pfade von $(0, 0)$ nach $(2p, 0)$ mit $s_1 > 0, s_2 > 0, \dots, s_{2p-1} > 0$.*

2. *Es gibt $\frac{1}{p+1} \binom{2p}{p}$ Pfade von $(0, 0)$ nach $(2p, 0)$ mit $s_1 \geq 0, s_2 \geq 0, \dots, s_{2p-1} \geq 0$.*

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 $s_1 > 0, s_2 > 0, \dots, s_{2p-1} = 1$. 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.

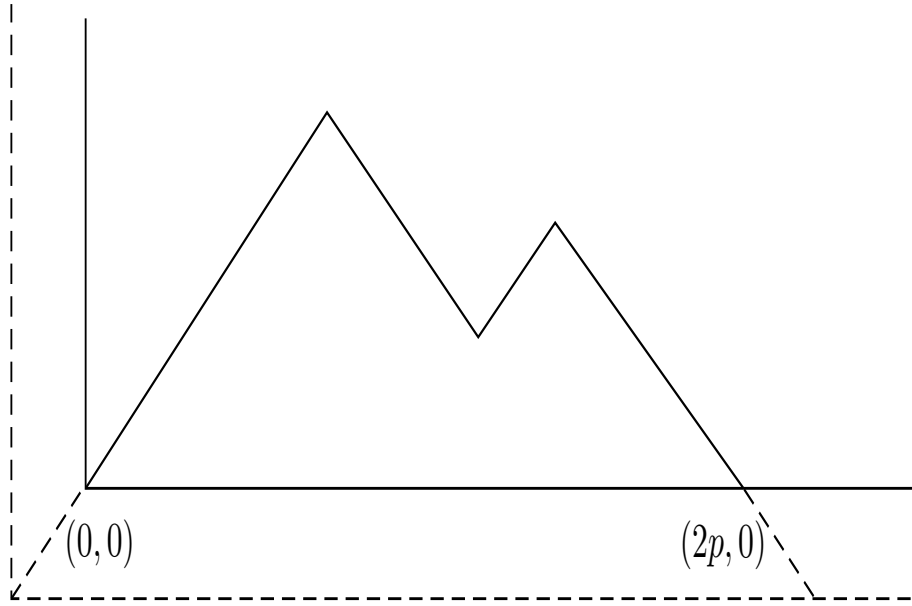
Auf diese Weise wird eine bijektive Abbildung von der gesuchten Menge von Pfaden auf die Menge der Pfade von $(-1, -1)$ nach $(2p + 1, -1)$, welche die Bedingung $s_0 > -1, s_1 > -1, \dots, s_{2p} > -1$ erfüllen, hergestellt. Die Anzahl der Pfade in dieser Menge ist gleich der Anzahl der Pfade von $(0, 0)$ nach $(2p + 2, 0)$ mit $s_1 > 0, s_2 > 0, \dots, s_{2p+1} > 0$ (Verschiebung des Ursprungs). (2) folgt dann aus (1). \square

Nun sind wir in der Lage Satz 2.25 zu beweisen:

Beweis von Satz 2.25. (1) Nach Satz 9.23 (1) gibt es $\frac{1}{k} \binom{2k-2}{k-1}$ Pfade von $(0, 0)$ nach $(2k, 0)$ mit $s_1 > 0, \dots, s_{2k-1} > 0$ 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 (1) folgt dann mit Hilfe von Satz 2.27 (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 $\{S_1 \neq 0, S_2 \neq 0, \dots, S_{2j-1} \neq 0, S_{2j} = 0\}$. Diese Ereignisse schließen sich gegenseitig aus und haben Wahrscheinlichkeiten $f_{2j} = u_{2j-2} - u_{2j}$. Somit ist mit $u_0 = 1$

$$P(S_1 \neq 0, S_2 \neq 0, \dots, S_{2k} \neq 0) = 1 - P\left(\bigcup_{j=1}^k C_{2j}\right) = 1 - \sum_{j=1}^k (u_{2j-2} - u_{2j}) = u_{2k}.$$

Die zweite Gleichung folgt analog aus der dritten Identität in (1).

(3) Für $1 \leq j \leq k$ sei $B_j = \{S_1 \neq 0, S_2 \neq 0, \dots, S_{2j-1} \neq 0, S_{2j} = 0, S_{2k} = 0\}$. 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 $\lim_{k \rightarrow \infty} u_{2k} = 0$ gilt, folgt, dass die Wahrscheinlichkeit für keine Rückkehr der Irrfahrt bis zum

Zeitpunkt $2k$ mit $k \rightarrow \infty$ 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 $u_{2k} \rightarrow 0$ folgt

$$\begin{aligned} \sum_{k=1}^{\infty} f_{2k} &= \lim_{N \rightarrow \infty} \sum_{k=1}^N f_{2k} \\ &= \lim_{N \rightarrow \infty} \sum_{k=1}^N (u_{2k-2} - u_{2k}) \\ &= \lim_{N \rightarrow \infty} (u_0 - u_{2N}) = 1. \end{aligned}$$

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 2.25 (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.

Obleich $P(S_1 \neq 0, \dots, S_{2k} \neq 0) = P(S_1 \geq 0, \dots, S_{2k} \geq 0) \sim 1/\sqrt{\pi k}$ 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 $\{S_1 \geq 0, S_2 \geq 0, \dots, S_{2k} \geq 0\}$. Aus Gründen der Symmetrie ist $P(F_0^{(k)}) = P(F_k^{(k)})$, was nach Satz 2.25 (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 allzu viele 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 2.28 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: $s_0 = 0$, $s_1 > 0$, $s_2 > 0, \dots, s_{2l-1} > 0$, $s_{2l} = 0$ 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: $s_0 = 0$, $s_1 < 0$, $s_2 < 0, \dots, s_{2l-1} < 0$, $s_{2l} = 0$ 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 $s_1 > 0$, $s_2 > 0, \dots, s_{2l-1} > 0$ 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 $1 \leq j \leq k-1$ 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 einfache Approximation für $P(F_j^{(k)})$ für große j und $k-j$ gewinnen:

Satz 2.29 Für $j \rightarrow \infty$, $k-j \rightarrow \infty$ gilt $P(F_j^{(k)}) \sim \frac{1}{\pi} \frac{1}{\sqrt{j(k-j)}}$, das heißt

$$\lim_{\substack{j \rightarrow \infty \\ k-j \rightarrow \infty}} \sqrt{j(k-j)} P(F_j^{(k)}) = \frac{1}{\pi}.$$

□

Betrachten wir speziell $x \in (0, 1)$ so gilt für $j, k \rightarrow \infty$ 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 2.29 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

$$\gamma_k(\alpha, \beta) := P\left(\alpha \leq \frac{T_k}{2k} \leq \beta\right) = \sum_{j: \alpha \leq \frac{j}{k} \leq \beta} P(F_j^{(k)}).$$

Wir wollen nun aus Satz 2.29 für $k \rightarrow \infty$ folgern:

$$\gamma_k(\alpha, \beta) \sim \frac{1}{\pi} \sum_{j: \alpha \leq \frac{j}{k} \leq \beta} \frac{1}{k} \frac{1}{\sqrt{\frac{j}{k}(1 - \frac{j}{k})}}. \quad (19)$$

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 2.30 (Arcus-Sinus-Gesetz)

$$\lim_{k \rightarrow \infty} \gamma_k(\alpha, \beta) = \frac{2}{\pi} (\arcsin \sqrt{\beta} - \arcsin \sqrt{\alpha}).$$

Beweis. Wir schreiben die Stirling-Approximation als $n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n F(n)$ mit $\lim_{n \rightarrow \infty} F(n) = 1$. Es folgt

$$P(F_j^{(k)}) = \binom{2j}{j} \binom{2k-2j}{k-j} \frac{1}{2^{2k}} = \frac{1}{\pi} \frac{1}{\sqrt{\left(\frac{j}{k}\right)\left(1 - \left(\frac{j}{k}\right)\right)}} \frac{1}{k} \frac{F(2j) F(2(k-j))}{F(j) F(j) F(k-j) F(k-j)}.$$

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 $k \rightarrow \infty$ 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 $k = 1, 2, \dots$, so dass für jedes obige $\delta > 0$ gilt:

$$\lim_{k \rightarrow \infty} G_{\alpha,\beta}(k) = 1 \quad \text{gleichmäßig für} \quad \delta \leq \alpha < \beta \leq 1 - \delta$$

und

$$\sum_{\alpha \leq \frac{j}{k} \leq \beta} P(F_j^{(k)}) = \left(\frac{1}{k} \sum_{\alpha \leq \frac{j}{k} \leq \beta} \frac{1}{\pi \sqrt{(j/k)(1 - (j/k))}} \right) G_{\alpha,\beta}(k).$$

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. \square

Bemerkung 2.31 Die Aussage von Satz 2.30 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 $\lim_{k \rightarrow \infty} \gamma_k(0, \frac{1}{2}) = 1/2$. Ist $\beta \in (0, 1/2)$, so folgt

$$\lim_{k \rightarrow \infty} \gamma_k(0, \beta) = \lim_{k \rightarrow \infty} (\gamma_k(0, 1/2) - \gamma_k(\beta, 1/2)) = \frac{2}{\pi} \arcsin \sqrt{\beta},$$

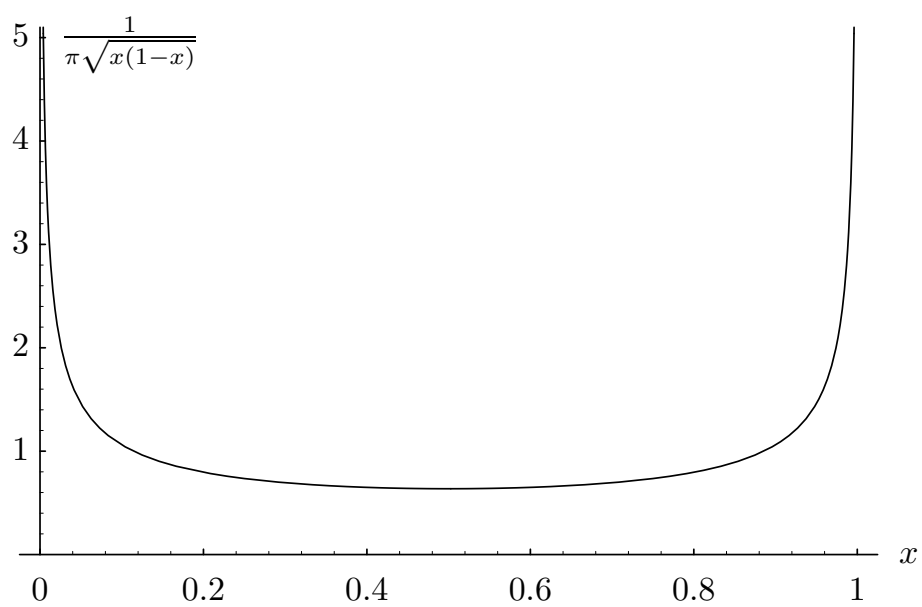
für $\beta > 1/2$

$$\lim_{k \rightarrow \infty} \gamma_k(0, \beta) = \lim_{k \rightarrow \infty} (\gamma_k(0, 1/2) + \gamma_k(1/2, \beta)) = \frac{2}{\pi} \arcsin \sqrt{\beta}.$$

Für $\gamma_k(\alpha, 1)$ führt dasselbe Argument zum Ziel. \square

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:



3 Ergodensätze

Schon im Kapitel über die Gesetze der großen Zahlen hatten wir uns mit dem Grenzwertverhalten des Mittelwertes $\frac{1}{n} \sum_{i=1}^n X_i$ einer Folge X_1, X_2, \dots 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, \dots 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 3.1 *Eine Folge X_0, X_1, \dots von Zufallsvariablen heißt stationär, falls für jedes $k \in \mathbb{N}$ die Folge X_k, X_{k+1}, \dots dieselbe Verteilung hat wie die Originalfolge X_0, X_1, \dots . Dies ist genau dann der Fall, wenn für jedes $n \in \mathbb{N}$ die Vektoren (X_0, \dots, X_n) und (X_k, \dots, X_{k+n}) für jedes $k \in \mathbb{N}$ die gleiche Verteilung haben.*

Die folgenden Beispiele werden uns durch dieses Kapitel begleiten.

Beispiel 3.2 *Die Folge X_0, X_1, \dots ist i.i.d. Offenbar haben dann die Zufallsvektoren (X_0, \dots, X_n) und (X_k, \dots, X_{k+n}) dieselbe (Produkt-)Verteilung; die Folge ist also stationär.*

Beispiel 3.3 *Es sei Ω eine endliche Menge und X_0, X_1, \dots 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, \dots 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

$$\begin{aligned} (X_0, X_1, \dots) &= (1, 0, 1, 0, \dots) \quad \text{und} \\ (X_0, X_1, \dots) &= (0, 1, 0, 1, \dots). \end{aligned}$$

Beispiel 3.4 (*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

$$X_n(\omega) := (\omega + n\vartheta) \bmod 1$$

was wir mit der obigen Interpretation von Ω als die Drehung von ω um den Winkel $2\pi\vartheta$ verstehen können. Da λ^1 die Gleichverteilung auf $[0, 1)$ ist, ist die Folge stationär, denn somit ist auch $(\omega + n\vartheta) \bmod 1$ gleichverteilt. Man kann Beispiel 3.3 als eine kontinuierliche Markov-Kette mit Übergangskern

$$Q(x, y) = \begin{cases} 1 & y = x + \vartheta \pmod{1} \\ 0 & \text{sonst} \end{cases}$$

auffassen.

Um weitere Beispiele zu konstruieren, benutzen wir das folgende Theorem.

Theorem 3.5 Ist die Folge X_0, X_1, \dots stationär und

$$g : \mathbb{R}^{\mathbb{N}_0} \rightarrow \mathbb{R}$$

messbar, dann ist auch die Folge der

$$Y_k = g(X_k, X_{k+1}, \dots)$$

eine stationäre Folge.

Beweis: Für $x \in \mathbb{R}^{\mathbb{N}_0}$ bezeichnen wir mit $g_k(x)$ die Abbildung

$$g_k(x) = g(x_k, x_{k+1}, \dots).$$

Für $B \in \mathcal{B}^{\mathbb{N}_0}$ sei weiter

$$A = \{x : (g_0(x), g_1(x), g_2(x), \dots) \in B\}.$$

Um die behauptete Stationarität zu beweisen, beobachten wir, dass

$$\begin{aligned} \mathbb{P}(\omega : (Y_0, Y_1, \dots) \in B) &= \mathbb{P}(\omega : (X_0, X_1, \dots) \in A) \\ &= \mathbb{P}(\omega : (X_k, X_{k+1}, \dots) \in A) \\ &= \mathbb{P}(\omega : (Y_k, Y_{k+1}, \dots) \in B) \end{aligned}$$

gilt. Dies heißt $(Y_n)_n$ ist stationär. □

Beispiel 3.6 (*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

$$Y_n(\omega) = 2 \cdot Y_{n-1}(\omega) \bmod 1.$$

Dies ist offenbar ein Spezialfall von Theorem 3.5, 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 3.6 auch als Spezialfall von Beispiel 3.3 auffassen.

Schließlich sind Beispiel 3.4 und 3.6 Spezialfälle des folgenden Beispiels:

Beispiel 3.7 Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und

$$\varphi : \Omega \rightarrow \Omega$$

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. Um dies einzusehen, benötigen wir den Kolmogorovschen Erweiterungssatz. Hierfür sei I eine beliebige Indexmenge und $\mathcal{E}(I)$ das System ihrer endlichen Teilmengen. Für jedes $1 \in I$ sei ein Messraum $(\Omega_i, \mathcal{A}_i)$. Für $K \in \mathcal{E}(I)$ schreiben wir

$$(\Omega^K, \mathcal{A}^K) := \left(\prod_{i \in K} \Omega_i, \bigotimes_{i \in K} \mathcal{A}_i \right).$$

Außerdem sei

$$(\Omega, \mathcal{A}) := \left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{A}_i \right).$$

Die Projektionen von Ω_L nach Ω_K , $K \subseteq L \in \mathcal{E}(I)$ heißen π_K^L , die Projektion von Ω auf Ω_L sei π_L . Eine Familie von Wahrscheinlichkeitsmaßen $\mathbb{P}_K, K \in \mathcal{E}(I)$ auf $(\Omega^K, \mathcal{A}^K)$ heißt *konsistent*, falls für alle $K \subseteq L \in \mathcal{E}(I)$ gilt:

$$\mathbb{P}_L = \mathbb{P}_K \circ (\pi_L^K)^{-1}.$$

Das Wichtige an konsistenten Familien ist, dass sie sich auf unendliche Produkte fortsetzen lassen. Diesen Satz, dessen Beweis im Wesentlichen ähnlich ist, zur Existenz unendlicher Produktmaße aus WT I, geben wir hier ohne Beweis an.

Satz 3.8 (Kolmogorov) Ist $((\Omega^K, \mathcal{A}^K, \mathbb{P}_K), K \in \mathcal{E}(I))$ eine projektive Familie von W -Räumen, so gibt es genau ein Wahrscheinlichkeitsmaß \mathbb{P} auf (Ω, \mathcal{A}) , das mit den \mathbb{P}_K kompatibel ist, in dem Sinne, dass gilt:

$$\mathbb{P}_K = \mathbb{P} \circ (\pi_K)^{-1}$$

für alle $K \in \mathcal{E}(I)$.

In der Tat folgt aus dem Kolmogorovschen Erweiterungssatz, 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 Bildraum der Y_i ist), so dass die Folge der $X_n(\omega) = \omega_n$ die gleiche Verteilung wie die Folge (Y_0, Y_1, \dots) 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 3.9 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

$$\mathbb{P}(Y_{-m} \in A_0, \dots, Y_n \in A_{m+n}) = \mathbb{P}(X_0 \in A_0, \dots, X_{m+n} \in A_{m+n})$$

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. \square

Aufgrund dieser Beobachtungen werden wir die folgende Theorie stationärer Folgen im Kontext von Beispiel 3.7 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.

Definition 3.10 Eine maßtreue Abbildung auf $(\Omega, \mathcal{F}, \mathbb{P})$ wollen wir ergodisch nennen, wenn die σ -Algebra \mathcal{I} der invarianten Mengen

$$\mathcal{I} := \{A \in \mathcal{F} : \varphi^{-1}A = A\}$$

trivial ist, d. h. wenn

$$\mathcal{I} = \{\emptyset, \Omega\} \quad \mathbb{P} - f.s$$

gilt oder mit anderen Worten, wenn für alle $A \in \mathcal{I}$

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

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

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots) \in \mathcal{T}_{\infty}.$$

Also ist $\mathcal{J} \subseteq \mathcal{T}_{\infty}$.

Ist $(X_n)_n$ eine i.i.d. Folge, so impliziert das Kolmogorowsche 0-1-Gesetz, dass \mathcal{T}_{∞} 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 3.3, 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, für 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

$$\vartheta_n(\omega_0, \omega_1, \omega_2, \dots) = (\omega_n, \omega_{n+1}, \dots)$$

der Shift-Operator ist. Schreiben wir also

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n),$$

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 $h(x) = \mathbb{E}_x 1_A$ gesetzt werde. Das 0-1-Gesetz von Levy besagt nun, dass für ein $A \in \mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \geq 1)$

$$\mathbb{E}[1_A | \mathcal{F}_n] \rightarrow 1_A \text{ f.s.}$$

gilt, also $h(x) \rightarrow 1_A$ \mathbb{P} -f.s. mit $n \rightarrow \infty$.

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}_∞ 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 3.4), 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-Analyse heran: Ist $f : [0, 1) \rightarrow \mathbb{R}$ messbar mit

$$\int f^2(x) dx < \infty,$$

dann kann f als die folgende Fourierreihe geschrieben werden:

$$f(x) = \sum_k c_k e^{2\pi i k x}$$

(wobei man, wie üblich, dies als

$$\sum_{k=-N}^N c_k e^{2\pi i k x} \xrightarrow{N \rightarrow \infty} f(x)$$

in $L^2[0, 1)$ auffassen sollte). Man kann auch die Koeffizienten c_k bestimmen. Diese haben die Gestalt

$$c_k = \int_0^1 f(x) e^{-2\pi i k x} dx.$$

Mit der Wahl von φ als Drehung des Kreises ist also

$$f(\varphi(x)) = \sum_k c_k e^{2\pi i k (x+\Theta)} = \sum_k (c_k e^{2\pi i k \Theta}) e^{2\pi i k x}.$$

Da die c_k eindeutig bestimmt sind, kann also $f(x) = f(\varphi(x))$ nur dann gelten, wenn

$$c_k (e^{2\pi i k \Theta} - 1) = 0$$

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

$$A \in \{\emptyset, [0, 1)\} \text{ } \mathbb{P}\text{-f.s.}$$

folgt.

Schließlich zeigen wir, dass auch der Bernoulli-Shift aus Beispiel 3.6 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, \dots eine i.i.d. Folge mit

$$\mathbb{P}(X_k = 0) = \mathbb{P}(X_k = 1) = \frac{1}{2}$$

ist. Weiter benötigen wir

Theorem 3.11 *Falls X_0, X_1, \dots eine ergodische, stationäre Folge ist und*

$$g : \mathbb{R}^{\mathbb{N}_0} \rightarrow \mathbb{R}$$

messbar, dann ist auch die Folge der

$$Y_k := g(X_k, X_{k+1}, \dots)$$

ergodisch.

Beweis: Aufgrund der Bemerkung nach Beispiel 3.7 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

$$\{\omega : (Y_0, Y_1, \dots) \in B\} = \{\omega : (Y_1, Y_2, \dots) \in B\}$$

genügt, dann ist die Menge

$$A = \{\omega : (Y_0, Y_1, \dots) \in B\}$$

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 $(X_k)_k$. \square

Wir wollen nun eine Art “Gesetz der großen Zahlen” für stationäre Folgen beweisen. Dieses ist unter dem Namen Birkhoffscher Ergodensatz bekannt. Es besagt, dass 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 3.12 *Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X \in \mathcal{L}^1(\mathbb{P})$. Ferner sei*

$$\varphi : \Omega \rightarrow \Omega$$

eine maßtreue Abbildung. Dann gilt

$$\frac{1}{n} \sum_{m=0}^{n-1} X(\varphi^m \omega) \rightarrow \mathbb{E}(X|\mathcal{J})$$

\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 3.13 (*Maximales Ergodenlemma*) *Sei $X_j(\omega) = X(\varphi^j \omega)$ und*

$$S_k(\omega) = \sum_{m=0}^{k-1} X_m(\omega)$$

und schließlich

$$M_k(\omega) = \max(0, S_1(\omega), \dots, S_k(\omega)).$$

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 $j = 1 \dots k$. 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

$$\begin{aligned} \mathbb{E}(X(\omega), M_k > 0) &\geq \int_{\{M_k > 0\}} \max(S_1(\omega), \dots, S_k(\omega)) - M_k(\varphi\omega) d\mathbb{P} \\ &= \int_{\{M_k > 0\}} M_k(\omega) - M_k(\varphi\omega) d\mathbb{P}. \end{aligned}$$

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. \square

Der Beweis des eigentlichen Ergodensatzes beginnt mit einer kleinen Übung:

Übung 3.14 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 3.10. 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 $\bar{X}(\omega)$ und $\bar{X}(\varphi(\omega))$ gerade, dass ersteres auch $X_0(\omega)$ berücksichtigt. Da man aber durch n teilt entfällt dieser Term bei der Mittelwertbildung im Limes $n \rightarrow \infty$. Also gilt $\bar{X}(\varphi(\omega)) = \bar{X}(\omega)$ und somit $D \subset \mathcal{J}$. Wir führen die folgenden Bezeichnungen ein:

$$\begin{aligned} X^*(\omega) &:= (X(\omega) - \varepsilon)\mathbb{1}_D(\omega), \\ S_n^*(\omega) &:= X^*(\omega) + \dots + X^*(\varphi^{n-1}(\omega)), \\ M_n^*(\omega) &:= \max(0, S_1^*(\omega), \dots, S_n^*(\omega)), \\ F_n &:= \{M_n^* > 0\} \quad \text{und} \\ F &:= \bigcup_n F_n = \left\{ \sup_{k \geq 1} \frac{S_k^*}{k} > 0 \right\}. \end{aligned}$$

Da nun X^* gleich $X(\omega) - \varepsilon$ auf der Menge

$$D = \left\{ \limsup \frac{S_k}{k} > \varepsilon \right\}$$

und 0 sonst ist, bekommen wir

$$F = \left\{ \sup_k \frac{S_k}{k} > \varepsilon \right\} \cap D = D.$$

Das maximale Ergodenlemma (Lemma 3.11) besagt nun, dass

$$\mathbb{E}(X^*, F_n) \geq 0$$

gilt. Da nun

$$\mathbb{E}|X^*| \leq \mathbb{E}|X| + \varepsilon < \infty$$

gilt, können wir den Satz von der majorisierten Konvergenz anwenden und

$$\mathbb{E}(X^*, F_n) \rightarrow \mathbb{E}(X^*, F)$$

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

$$\frac{S_n}{n} \rightarrow 0 \quad \mathbb{P}\text{-f.s.}$$

Um zu beweisen, dass die Konvergenz auch in $\mathcal{L}^1(\mathbb{P})$ ist, setzen wir für $M \in \mathbb{R}$

$$X'(\omega) = X'_M(\omega) = X(\omega) \mathbb{1}_{\{|X(\omega)| \leq M\}}$$

und

$$X''(\omega) = X''_M(\omega) = X(\omega) - X'(\omega).$$

Nun besagt der erste Teil des Satzes, den wir soeben bewiesen haben, dass

$$\frac{1}{n} \sum_{m=0}^{n-1} X'(\varphi^m \omega) \rightarrow \mathbb{E}(X'|\mathcal{J}) \quad \mathbb{P}\text{-f.s.}$$

gilt. Da X' beschränkt ist, bekommen wir aus dem Satz über majorisierte Konvergenz auch, dass auch

$$\mathbb{E} \left(\left| \frac{1}{n} \sum_{m=0}^{n-1} X'(\varphi^m \omega) - \mathbb{E}(X'|\mathcal{J}) \right| \right) \rightarrow 0 \quad (20)$$

gilt, wenn $n \rightarrow \infty$ gilt. Um auch X'' zu behandeln, benutzen wir, dass

$$\mathbb{E} \left| \frac{1}{n} \sum_{m=0}^{n-1} X''(\varphi^m(\omega)) \right| \leq \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{E}|X''(\varphi^m(\omega))| = \mathbb{E}|X''|$$

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

$$\begin{aligned} & \mathbb{E} \left| \frac{1}{n} \sum_{m=0}^{n-1} X''(\varphi^m(\omega)) - \mathbb{E}(X''|\mathcal{J}) \right| \\ & \leq \mathbb{E} \left| \frac{1}{n} \sum_{m=0}^{n-1} X''(\varphi^m(\omega)) \right| + \mathbb{E}|\mathbb{E}(X''|\mathcal{J})| \\ & \leq 2\mathbb{E}|X''|. \end{aligned}$$

Somit gilt auch

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left| \frac{1}{n} \sum_{m=0}^{n-1} X''(\varphi^m(\omega)) - \mathbb{E}(X''|\mathcal{J}) \right| \leq 2\mathbb{E}(X'').$$

Da nun X als integrierbar vorausgesetzt war, folgt

$$\mathbb{E}|X''| \rightarrow 0 \quad \text{wenn} \quad M \rightarrow \infty$$

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. \square

Bevor wir uns die Konsequenzen des Ergodensatzes anhand unserer Beispiele betrachten, leiten wir zunächst eine für Zwecke nützliche Ungleichung aus dem maximalen Ergodenlemma her:

Proposition 3.15 (*Wiener's Maximal-Ungleichung*) *Wie im Ergodensatz sei $\varphi : \Omega \rightarrow \Omega$ eine maßtreue Abbildung und $X \in \mathcal{L}^1(\mathbb{P})$. Ferner setzen wir $X_j(\omega) = X(\varphi^j(\omega))$ und*

$$S_k(\omega) = X_0(\omega) + X_1(\omega) + \dots + X_{k-1}(\omega)$$

und $A_k(\omega) = \frac{S_k(\omega)}{k}$. Schließlich sei $D_k = \max(A_1, \dots, A_k)$. 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

$$S'_j(\omega) = X'(\varphi^j(\omega)) \quad \text{und} \quad S'_k = \sum_{j=0}^{k-1} X'_j(\omega)$$

und schließlich

$$M'_k = \max(0, S'_1, \dots, S'_k)$$

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. \square

Nun betrachten wir die Auswirkungen des Birkhoffschen Ergodensatzes für unsere Beispiele.

Beispiel 3.16 (I.I.D. Folgen) Da wir schon festgestellt haben, dass i.i.d. Folgen ergodisch sind, d.h. dass \mathcal{J} trivial ist, ist

$$\mathbb{E}(X|\mathcal{J}) = \mathbb{E}X \quad \mathbb{P}\text{-f.s.}$$

für alle integrierbaren X . Der Ergodensatz (Theorem 3.10) behauptet somit für i.i.d. Folgen X_0, X_1, \dots

$$\frac{1}{n} \sum_{m=0}^{n-1} X_m \xrightarrow[n \rightarrow \infty]{} \mathbb{E}X_0,$$

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 3.17 (Markov-Ketten) Es sei $(X_n)_n$ eine irreduzible Markov-Kette auf einem endlichen Zustandsraum Ω und stationärer Verteilung π . Es sei $f : \Omega \rightarrow \mathbb{R}$ 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 3.14

$$\mathbb{E}(X|\mathcal{J}) = \mathbb{E}X \quad \mathbb{P}\text{-f.s.}$$

Wendet man daher den Ergodensatz auf $f(X_0(\omega))$ an, so ergibt sich die Konvergenzaussage

$$\frac{1}{n} \sum_{m=0}^{n-1} f(X_m) \rightarrow \mathbb{E}_\pi(f) = \sum_x f(x)\pi(x)$$

\mathbb{P} -f.s. als auch in $\mathcal{L}^1(\mathbb{P})$.

Beispiel 3.18 (Drehung des Kreises) Hierfür sei wieder $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}|_{[0,1]}$ und $\mathbb{P} = \lambda^1|_{[0,1]}$. φ sei die maßtreue Abbildung

$$\varphi(\omega) = (\omega + \vartheta) \bmod 1$$

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

$$\frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_{(\varphi^m(\omega) \in A)} \rightarrow \lambda^1(A) \quad \mathbb{P}\text{-f.s. und in } \mathcal{L}^1(\lambda^1). \quad (21)$$

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

$$\frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_{((m \cdot \vartheta) \bmod 1 \in A)} \rightarrow \lambda^1(A).$$

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 (21) (also die Menge, für die (21) 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 (21) impliziert

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_{A_k}(\varphi^m(\omega)) = b - a - \frac{2}{k}$$

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 $x \in [0, 1)$

$$\liminf \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_A(\varphi^m(x)) \geq b - a - \frac{2}{k}$$

für alle hinreichend großen k . Wendet man die gleichen Überlegungen auf A^c (das sich als Vereinigung zweier Intervalle schreiben lässt) an, ergibt sich

$$\frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_A(\varphi^m(x)) \rightarrow b - a$$

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

$$\frac{1}{n} \sum_{m=0}^{n-1} 1_A(\varphi^m(0)) \rightarrow \log_{10}\left(\frac{k+1}{k}\right).$$

Nun ist die erste Ziffer der Zahl 2^m gleich k genau dann, wenn

$$m\vartheta \bmod 1 \in A$$

ist. Somit haben wir beispielsweise für $k = 1$ gezeigt, dass der asymptotische Anteil von 2er Potenzen, deren Dezimalentwicklung mit einer 1 beginnt, $\log_{10} 2 = 0,3010 \dots$ ist.

Die Limesverteilung auf den Ziffern $\{1, \dots, 9\}$ 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 3.19 (Bernoulli-Shift) Hier sei $(\Omega, \mathcal{F}, \mathbb{P})$ wie im vorigen Beispiel und die maßtreue Abbildung

$$\varphi : \Omega \rightarrow \Omega$$

gegeben durch

$$\varphi(\omega) = (2\omega) \bmod 1.$$

Sei ferner $i_1, \dots, i_k \in \{0, 1\}$ und

$$r = \sum_{m=1}^k i_m 2^{-m}$$

das „Muster“, das durch die Ziffernfolge (i_1, \dots, i_k) dargestellt wird. Schließlich sei

$$X(\omega) = \begin{cases} 1, & \text{falls } r \leq \omega < 1 + 2^{-k} \\ 0, & \text{sonst} \end{cases},$$

d.h. $X(\omega)$ ist genau dann 1, wenn die ersten k Ziffern von ω in Binärdarstellung genau i_1, \dots, i_k sind. Der Ergodensatz behauptet nun wegen der Trivialität von \mathcal{J} , dass

$$\frac{1}{n} \sum_{m=0}^{n-1} X(\varphi^m(\omega)) \rightarrow 2^{-k} \mathbb{P}\text{-f.s.}$$

(und in \mathcal{L}^1). Mit anderen Worten: Das Muster (i_1, \dots, 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 $\varphi : \Omega \rightarrow \Omega$ ergodisch ist, bzw. wann die Folge ergodisch ist. Ein inhaltlich leicht fassbares Konzept ist das des „Mischens“.

Definition 3.20 *Eine maßtreue Abbildung*

$$\varphi : \Omega \rightarrow \Omega$$

auf einem Maßraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt mischend, falls

$$\lim_{n \rightarrow \infty} \mathbb{P}(A \cap \varphi^{-n}B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \quad (22)$$

für je zwei Mengen $A, B \in \mathcal{F}$ gilt. Eine stationäre Folge $X := (X_n)_{n \in \mathbb{Z}}$ von Zufallsvariablen ist mischend, wenn der Shift auf dem Folgenraum mischend ist.

Es ist leicht zu sehen, dass eine Abbildung ergodisch ist, wenn sie mischend ist.

Proposition 3.21 *Ist φ mischend, so ist φ auch ergodisch.*

Beweis: Wir müssen zeigen, dass \mathcal{I} trivial ist, wenn φ mischend ist. Sei $A \in \mathcal{I}$. Dann ist $\varphi^{-n}(A) = A$ für alle $n \in \mathbb{N}$. Also

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \lim_{n \rightarrow \infty} \mathbb{P}(A \cap \varphi^{-n}(A)) = \mathbb{P}(A)^2,$$

d. h. $\mathbb{P}(A) \in \{0, 1\}$. □

Umgekehrt ist „mischend“ nicht zu weit von Ergodizität entfernt. In der Tat impliziert Ergodizität von φ ja über den Ergodensatz

$$\frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}_B(\varphi^m(\omega)) \rightarrow \mathbb{P}(B) \text{ } \mathbb{P}\text{-f.s.}$$

Integriert man diese Konvergenz über A , d.h. benutzt man den Satz über majorisierte Konvergenz, erhält man

$$\frac{1}{n} \sum_{m=0}^{n-1} \mathbb{P}(A \cap \varphi^{-m}B) \rightarrow \mathbb{P}(A)\mathbb{P}(B),$$

also gilt (22) 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. $\Omega = \{(\omega_0, \omega_1, \dots)\}$ und $\varphi(\omega) = (\omega_1, \omega_2, \dots)$ und

$$X_n(\omega) = \omega_n.$$

Ferner sei

$$\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \dots)$$

und

$$\mathcal{T}_\infty = \bigcap_n \mathcal{F}'_n.$$

Theorem 3.22 Falls \mathcal{T}_∞ 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 $A \in \mathcal{F}$

$$\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{F}} |\mathbb{P}(A \cap \varphi^{-n}B) - \mathbb{P}(A)\mathbb{P}(B)| = 0.$$

Beweis: Sei $C = \varphi^{-n}B \in \mathcal{F}'_n$. Dann gilt für $A \in \mathcal{F}_n$

$$\begin{aligned} |\mathbb{P}(A \cap C) - \mathbb{P}(A)\mathbb{P}(C)| &= \left| \int_C \mathbb{1}_A - \mathbb{P}(A) d\mathbb{P} \right| \\ &= \left| \int_C \mathbb{P}(A|\mathcal{F}'_n) - \mathbb{P}(A) d\mathbb{P} \right| \\ &\leq \int |\mathbb{P}(A|\mathcal{F}'_n) - \mathbb{P}(A)| d\mathbb{P} \rightarrow 0. \end{aligned}$$

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. \square

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

$$|\mathbb{P}(A \cap \varphi^{-n}B_n) - \mathbb{P}(A)\mathbb{P}(B)| = \mathbb{P}(A) - \mathbb{P}^2(A).$$

Somit erhalten wir, dass \mathcal{T}_∞ trivial ist genau dann, wenn

$$\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{F}} |\mathbb{P}(A \cap \varphi^{-n}B) - \mathbb{P}(A)\mathbb{P}(B)| = 0$$

für alle A gilt. Definitionsgemäß ist φ mischend genau dann, wenn

$$\lim_{n \rightarrow \infty} |\mathbb{P}(A \cap \varphi^{-n}(B)) - \mathbb{P}(A)\mathbb{P}(B)| = 0$$

für alle A und B gilt. Schließlich ist φ genau dann ergodisch, wenn

$$\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{P}(A \cap \varphi^{-m}B) - \mathbb{P}(A)\mathbb{P}(B) \right| = 0$$

für alle messbaren A und B gilt.

Übung 3.23 Man zeige die unbewiesene Richtung der letzten Behauptung.

Beispiel 3.24 (I.I.D. Folgen) Folgen $X = (X_n)_{n \in \mathbb{Z}}$ von i.i.d. Zufallsvariablen sind mischend. Dies prüft man direkt, in dem man benutzt, dass es für jede $A, B \in \mathcal{B}^{\mathbb{Z}}$ und jedes $\varepsilon > 0$ Ereignisse A_ε und B_ε gibt, die nur von endlichen vielen Koordinaten abhängen und für die

$$\mathbb{P}^X(A \Delta A_\varepsilon) < \varepsilon \quad \text{und} \quad \mathbb{P}^X(B \Delta B_\varepsilon) < \varepsilon$$

gibt. Die Ereignisse A_ε und $\varphi^{-n}(B_\varepsilon)$ sind für hinreichend großes n sogar unabhängig, das zeigt die Behauptung.

Beispiel 3.25 (Markov-Ketten) Es sei X_0, X_1, \dots eine irreduzible Markov-Kette über einem endlichen Zustandsraum mit invarianter Verteilung π . Ist die Kette zudem aperiodisch, so ist \mathcal{T}_∞ trivial (das ist ein wenig zu aufwendig, um es hier zu zeigen) und die Folge ist mischend.

Beispiel 3.26 (Drehung auf dem Kreis) Sei wieder $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}^1|_{[0,1]}$ und $\mathbb{P} = \lambda^1|_{[0,1]}$. Die maßtreue Abbildung

$$\varphi : \Omega \rightarrow \Omega$$

sei gegeben durch

$$\varphi(\omega) = (\omega + \vartheta) \bmod 1,$$

wobei $\vartheta \in (0, 1)$ irrational ist. Diese Abbildung ist nicht mischend. Um dies einzusehen, beachte man, dass die Menge $T = \{(n\vartheta) \bmod 1 | n \in \mathbb{N}\}$ dicht ist in $[0, 1)$. Dies zeigt man ähnlich wie in Beispiel 3.16. Da T dicht ist in $[0, 1)$, gibt es eine Folge (n_k) mit $n_k \rightarrow \infty$, wenn $k \rightarrow \infty$ und

$$(n_k \vartheta) \bmod 1 \rightarrow \frac{1}{2}.$$

Es sei $A = B = [0, \frac{1}{3})$. Ist k hinreichend groß, so gilt

$$A \cap \varphi^{-n_k} B = \emptyset.$$

Somit gilt

$$0 = \lim_{k \rightarrow \infty} \mathbb{P}(A \cap \varphi^{-n_k}(B)) \neq \frac{1}{9}$$

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 3.27 (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

$$\mathbb{E}X_{0,n} \geq \gamma_0 n \quad \text{für ein } \gamma_0 > -\infty.$$

Dann gilt:

1. $\lim_{n \rightarrow \infty} \frac{\mathbb{E}X_{0,n}}{n} = \inf_m \frac{\mathbb{E}X_{0,m}}{m}$ existiert und ist gleich einem Wert γ .
2. $X = \lim_{n \rightarrow \infty} \frac{X_{0,n}}{n}$ 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

$$a_m + a_{n-m} \geq a_n. \quad (23)$$

Setze $\gamma := \inf_{m \geq 1} \frac{a_m}{m}$. Dann ist offensichtlich

$$\liminf_{n \rightarrow \infty} \frac{a_n}{n} \geq \inf_{m \geq 1} \frac{a_m}{m} = \gamma.$$

Wir zeigen nun noch, dass auch $\limsup_{n \rightarrow \infty} \frac{a_n}{n} \geq \frac{a_m}{m}$ für jedes $m \in \mathbb{N}$ gilt. Schreiben wir dazu $n = km + \ell$ für ein $0 \leq \ell < m$, so ergibt wiederholte Anwendung von (3.4)

$$a_n \leq ka_m + a_\ell.$$

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 $n \rightarrow \infty$ verschwindet der zweite Summand der rechten Seite und der erste konvergiert gegen $\frac{a_m}{m}$. Dies zeigt

$$\limsup_{n \rightarrow \infty} \frac{a_n}{n} \leq \frac{a_m}{m}$$

für festes m .

SCHRITT 3. 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:

$$X_{0,n} \leq X_{0,m_1} + \dots + X_{(k-1)m,km} + X_{km,n}.$$

Division durch $n = km + \ell$ ergibt

$$\frac{X_{0,n}}{n} \leq \left(\frac{k}{km + \ell} \right) \frac{X_{0,m} + \dots + X_{(k-1)m,km}}{k} + \frac{X_{km,n}}{n}.$$

Voraussetzung b) erlaubt es uns, den Birkhoffschen Ergodensatz anzuwenden, nach dem

$$\frac{X_{0,m} + \dots + X_{(k-1)m,km}}{k} \xrightarrow[k \rightarrow \infty]{} A_m \quad (24)$$

in \mathcal{L}^1 und \mathbb{P} -f.s. gilt.

Hält man andererseits ℓ fest und wählt $\varepsilon > 0$, so folgt mit Voraussetzung c):

$$\sum_{k=1}^{\infty} \mathbb{P}(X_{km,km+\ell} > n\varepsilon) \leq \sum_{k=1}^{\infty} \mathbb{P}(X_{0,\ell} > k\varepsilon) < \infty, \quad (25)$$

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

$$\mathbb{E}(X_{0,n}) \leq C_n < \infty.)$$

(3.5) und (3.6) implizieren, dass

$$\bar{X} := \limsup_{n \rightarrow \infty} \frac{X_{0,n}}{n} \leq \frac{A_m}{m}$$

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

$$\underline{X} = \liminf_{n \rightarrow \infty} \frac{X_{0,n}}{n}.$$

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 $\lim_{n \rightarrow \infty} \frac{X_{0,n}}{n}$ existiert (\mathbb{P} -f.s.). Sei nun für $m \in \mathbb{N}$

$$\underline{X}_m = \liminf_{n \rightarrow \infty} \frac{X_{m,m+n}}{n}.$$

Wegen a) folgt

$$X_{0,m+n} \leq X_{0,m} + X_{m,m+n}.$$

Dividiert man beide Seiten durch n und schickt $n \rightarrow \infty$, 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

$$\underline{Y} = \liminf_{n \rightarrow \infty} \frac{Y_{0,n}}{n} \leq -\varepsilon \quad \mathbb{P}\text{-f.s.} \quad (26)$$

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. (3.7) 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

$$Y_{0,n} \leq Y_{R_0,R_1} + \dots + Y_{R_{K-1},R_K} + Y_{R_K,n}.$$

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 $n \rightarrow \infty$, so erhält man

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}Y_{0,n}}{n} \leq \mathbb{E}\xi_0 \leq \mathbb{E}(Y_{0,1}, T_0 > N) \leq \varepsilon.$$

Aus der ersten (schon bewiesenen) Behauptung des subadditiven Ergodensatzes und der Definition von $Y_{0,n}$ ergibt sich somit

$$\gamma = \limsup_{n \rightarrow \infty} \frac{\mathbb{E}X_{0,n}}{n} \leq 2\varepsilon + \mathbb{E}(\underline{X} \vee -M).$$

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

$$\mathbb{E} \left| \frac{1}{n} X_{0,n} - \Gamma \right| = 2\mathbb{E}\left(\frac{1}{n} X_{0,n} - \Gamma\right)^+ - \mathbb{E}\left(\frac{1}{n} X_{0,n} - \Gamma\right) \leq 2\mathbb{E}\left(\frac{1}{n} X_{0,n} - \Gamma\right)^+.$$

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

$$\mathbb{E}\left(\frac{1}{n} X_{0,n} - \Gamma\right)^+ \leq \mathbb{E}\left(\frac{1}{n} X_{0,n} - \Gamma_m\right)^+ + \mathbb{E}(\Gamma_m - \Gamma)^+. \quad (27)$$

Da $\Gamma_m \geq \Gamma$ ist, kann das $+$ -Zeichen im letzten Summanden auch fallengelassen werden. Nun konvergiert $\mathbb{E}\Gamma_m$ gegen γ , wenn $m \rightarrow \infty$ geht und aus den Schritten 2 und 3 erhalten wir

$$\mathbb{E}\Gamma \geq \mathbb{E}\bar{X} \geq \mathbb{E}X \geq \gamma.$$

Somit ist auch schon $\mathbb{E}\Gamma = \gamma$ und $\mathbb{E}(\Gamma_m - \Gamma)$ konvergiert gegen 0, wenn $m \rightarrow \infty$ geht. Um auch den ersten Summanden auf der rechten Seite in (3.8) zu beschränken, bemerken wir

$$\mathbb{E}\left(\frac{1}{n}X_{0,n} - \Gamma_m\right)^+ \leq \mathbb{E}\left(\frac{X_{0,m} + \dots + X_{(k-1)m,km}}{k_{m+\ell}} - \Gamma_m\right)^+ + \mathbb{E}\left(\frac{X_{km,n}}{n}\right)^+, \quad (28)$$

wobei wir wiederum a) verwendet haben. Nun ist

$$\mathbb{E}\left(\frac{X_{km,n}}{n}\right)^+ = \mathbb{E}\left(\frac{X_{0,\ell}}{n}\right)^+ \xrightarrow{n \rightarrow \infty} 0.$$

Auf den ersten Summanden in (3.9) wenden wir wieder den Birkhoffschen Ergodensatz an und erhalten

$$\mathbb{E}\left|\frac{X_{0,m} + \dots + X_{(k-1)m,km}}{k} - \Gamma_m\right| \rightarrow 0.$$

Dies beschließt den Beweis des subadditiven Ergodensatzes. \square

Wir schließen das Kapitel mit ein paar Beispielen für Theorem 3.22.

Beispiel 3.28 (*Stationäre Folgen*) Sei ξ_1, ξ_2, \dots eine stationäre Folge von Zufallsvariablen mit

$$\mathbb{E}|\xi_k| < \infty.$$

Setze

$$X_{m,n} = \xi_{m+1} + \dots + \xi_n.$$

Dann gilt sogar Additivität, d. h.

$$X_{0,n} = X_{0,m} + X_{m,n}$$

und Theorem 3.25 ist anwendbar.

Beispiel 3.29 (*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

$$\mathcal{N} = \{(x, y) : x, y \in \mathbb{Z}^2, \text{ } x \text{ und } y \text{ sind benachbart}\}.$$

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 3.30 (Die besuchten Punkte einer Irrfahrt) Es sei ξ_1, ξ_2, \dots eine stationäre Folge von Zufallsvariablen (z. B. eine i.i.d. Folge) und

$$S_n = \xi_1 + \dots + \xi_n.$$

Wir setzen $X_{m,n}$ als die Anzahl der zwischen den Zeitpunkten m und n besuchten Punkte, also

$$X_{m,n} = \#\{S_{m+1}, \dots, S_n\}.$$

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 , $X_{0,n}$,

$$\frac{X_{0,n}}{n} \rightarrow X$$

\mathbb{P} -f.s. und in \mathcal{L}^1 genügt. Allerdings weißt 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 3.25 zeigen, dass

$$\frac{X_{0,n}}{n} = \frac{\#\{S_1, \dots, S_n\}}{n} \rightarrow \mathbb{P}(S_n \text{ kehrt nicht nach } 0 \text{ zurück})$$

gilt (Übung).

4 Die Brownsche Bewegung und der Satz von Donsker

In diesem Abschnitt werden wir den wichtigsten stochastischen Prozess in kontinuierlicher Zeit konstruieren und einige seiner Eigenschaften diskutieren, die Brownsche Bewegung. Die Brownsche Bewegung wurde zunächst von dem Botaniker Robert Brown als „mysteriöse“ Bewegung eines in Wasser schwimmenden Pollenteilchens unter dem Mikroskop beschrieben. 1905 gab Einstein eine Begründung dieser Bewegung aufgrund der damals nur vermuteten Atomhypothese. Demnach bewegen sich grob gesprochen die Moleküle aufgrund der Wassertemperatur (bzw. ist diese Bewegung gerade die Temperatur) und stoßen dabei in jeder Sekunde sehr oft von allen Seiten gegen das Pollenteilchen, das daraufhin die eratischen Bewegungen vollführt. Diese Vorstellung wurde in den folgenden Jahren durch die Experimente und Messungen von Jean Baptiste Perrin quantitativ bestätigt. Die erfolgreiche Erklärung der Brownschen Bewegung gilt als Meilenstein auf dem Weg zum wissenschaftlichen Nachweis der Existenz der Moleküle und damit der Atome. Zu erwähnen ist auch, dass der französische Mathematiker Bachelier schon fünf Jahre vor Einstein den gleichen stochastischen Prozess zur Modellierung von Aktienkursen beschrieb.

Der mathematische Begriff der Brownschen Bewegung wurde 1920 von N. Wiener geprägt.

Allen gemein ist die folgende Definition der Brownschen Bewegung.

Definition 4.1 *Eine Brownsche Bewegung $(B_t)_{t \in \mathbb{R}^+}$ ist ein stochastischer Prozess, d.h. für jedes t ist B_t eine Zufallsvariable, mit den folgenden Eigenschaften*

1. *Für $0 = t_0 \leq t_1 \leq \dots \leq t_n$ sind die Zufallsvariablen $B_{t_{j+1}} - B_{t_j}$ ($j = 0, \dots, n-1$) unabhängig;*
2. *B_t hat stationäre, normalverteilte Zuwächse, d.h., die gemeinsame Verteilung*

$$(B_{t_1+s} - B_{u_1+s}, B_{t_2+s} - B_{u_2+s}, \dots, B_{t_n+s} - B_{u_n+s})$$

hängt nicht von $s \geq 0$ ab, wobei die $t_i > u_i$ für $i = 1, 2, \dots, n$ beliebig sind. und für $s, t \geq 0$ ist

$$\mathbb{P}(B_{s+t} - B_s \in A) = \frac{1}{\sqrt{2\pi t}} \int_A e^{-\frac{x^2}{2t}} dx.$$

3. *$t \mapsto B_t$ ist fast sicher stetig*

Diese letzte Forderung entspringt dem Wunsch, einen eindeutigen Prozess zu definieren, mit den ersten beiden Bedingungen hätte man auch gewisse Sprungprozesse nicht ausgeschlossen.

Im Kapitel über den Zentralen Grenzwertsatz haben wir die Normalverteilung als eines der zentralen Objekte der Wahrscheinlichkeitstheorie kennengelernt. Ihre Rolle stammt daher, dass unter gewissen (relativ milden) Bedingungen geeignet skalierte Mittelwerte beliebiger unabhängiger Zufallsvariablen eine Standardnormalverteilung als Verteilungslimes haben.

Wir wollen nun einen ähnlichen Satz auf der Ebene zufälliger Funktionen herleiten. Zuvor aber wollen wir eine allgemeine Version des Zentralen Grenzwertsatzes kennenlernen und einen Beweis sehen, der vielleicht am ehesten deutlich macht, warum der Zentrale Grenzwertsatz gilt.

Theorem 4.2 (Zentraler Grenzwertsatz – CLT) *Es sei X_1, \dots, X_n, \dots eine Folge von i.i.d. Zufallsvariablen mit endlicher Varianz, d.h. mit $\mathbb{E}X_1^2 < \infty$. Dann gilt der zentrale Grenzwertsatz für sie, d.h.*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n (X_i - \mathbb{E}X_1)}{\sqrt{n \mathbb{V}X_1}} \leq a \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx. \quad (29)$$

gilt für alle $a \in \mathbb{R}$.

Bevor wir uns an den Beweis dieses Satzes machen, bemerken wir, dass der Satz auch unter den folgenden schwächeren Bedingungen gilt (die insbesondere auf die identische Verteilung der X_i verzichten).

In der Tat gilt der Zentrale Grenzwertsatz auch unter den folgenden schwächeren Bedingungen. Es sei X_i eine Folge unabhängiger Zufallsvariablen. Für $j = 1, \dots, n$ definiere

$$X_{nj} := \frac{1}{s_n} (X_j - \eta_j)$$

wobei

$$\eta_j = \mathbb{E}X_j$$

der Erwartungswert von X_j ist und

$$s_n := \sqrt{\sum_{i=1}^n \mathbb{V}X_i}$$

die Varianz der Summe.

Wir sagen, dass die Folge der (X_n) der **Lindeberg-Bedingung** genügt, falls

$$L_n(\varepsilon) \rightarrow 0 \text{ wenn } n \rightarrow \infty$$

für alle $\varepsilon > 0$ gilt. Hierbei ist

$$L_n(\varepsilon) = \frac{1}{s_n^2} \sum_{j=1}^n \mathbb{E} [(X_j - \eta_j)^2; |X_j - \eta_j| \geq \varepsilon s_n].$$

Intuitiv gesprochen fordert die Lindeberg-Bedingung, dass keine der Zufallsvariablen die gesamte Summe dominiert. Dies wird in der zweiten Bemerkung unten präziser gemacht.

Die verallgemeinerte Form des zentralen Grenzwertsatzes besagt nun, eine Folge von Zufallsvariablen X_n , die unabhängig ist und der **Lindeberg-Bedingung** genügt, auch dem zentralen Grenzwertsatz genügt, d.h. es gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n (X_i - \eta_i)}{\sqrt{\sum_{i=1}^n \mathbb{V} X_i}} \leq a \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx.$$

Der Beweis folgt im wesentlichen den gleichen Ideen, wie der Beweis, den wir unten angeben werden. Wir sparen uns die zusätzliche technische Arbeit, da sie keine weitere Einsicht liefert und geben nur einen kurzen Kommentar. Für einen Beweis entlang der Ideen von Stein verweisen wir auf die Skript von Alsmeyer.

Zunächst wollen wir die Lindeberg Bedingung noch etwas näher beleuchten.

Bemerkung 4.3 • Sind die Zufallsvariablen (X_n) i.i.d., so gilt die Lindeberg-Bedingung. In der Tat: Sei $\mu := \mathbb{P}_{X_n}$, $\eta := \mathbb{E}X_n$ und $\sigma^2 := \mathbb{V}X_n$. Dann ist $s_n^2 = n\sigma^2$ und somit divergiert s_n nach unendlich, wenn $n \rightarrow \infty$. Also folgt

$$L_n(\varepsilon) = \frac{1}{\sigma^2} \int_{\{|x-\eta| \geq \varepsilon s_n\}} (x - \eta)^2 \mu(dx) \rightarrow 0$$

für alle $\varepsilon > 0$, wenn n gegen ∞ strebt.

- Genügt die Folge der Zufallsvariablen (X_n) der sogenannten Lyapunov-Bedingung, d.h. existiert ein $\delta > 0$, für das gilt

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - \mathbb{E}X_j|^{2+\delta}) = 0$$

so genügt sie auch der Lindeberg-Bedingung. Das zu zeigen, ist eine Übung.

- Ist die Folge der (X_n) fast sicher gleichmäßig beschränkt und gilt zudem $s_n \rightarrow \infty$ mit $n \rightarrow \infty$, so gilt die Lindeberg Bedingung. Gemäß Voraussetzung gibt es nämlich ein $\alpha > 0$ mit $|X_n| \leq \alpha/2$ und somit auch mit

$$|\mathbb{E}X_n| \leq \mathbb{E}(|X_n|) \leq \frac{\alpha}{2}$$

für alle n fast sicher. Also ist auch

$$|X_n - \mathbb{E}X_n| \leq \alpha$$

fast sicher für alle n . Dann aber folgt auch für $\delta > 0$ die Lyapunov Bedingung

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - \mathbb{E}X_j|^{2+\delta}) \leq \frac{\alpha^\delta}{s_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - \mathbb{E}X_j|^2) = \left(\frac{\alpha}{s_n} \right)^\delta$$

und somit die Lindeberg-Bedingung.

- Umgekehrt folgt aus der Lindeberg-Bedingung die sogenannte Feller-Bedingung.

Lemma 4.4 Genügt eine Folge (X_n) von Zufallsvariablen der Lindeberg-Bedingung so genügt sie auch der Feller Bedingung

$$\lim_{n \rightarrow \infty} \left(\max_{1 \leq j \leq n} \frac{\sqrt{\mathbb{V}X_j}}{s_n} \right) = 0.$$

Beweis: Der Beweis ist eine Übung. □

Die Feller-Bedingung illustriert besonders deutlich, dass keine der Zufallsvariablen einen zu großen Einfluss auf die Summe haben darf.

Im Folgenden wollen wir den sehr eleganten Beweisweg von Satz ?? von LINDEBERG vorstellen: Gegeben sind also reellwertige Zufallsvariablen X_1, X_2, \dots , die unabhängig seien. Wir nehmen ohne Einschränkung an, dass die Erwartungswerte $\mathbb{E}(X_i) = \int X_i dP = 0$ sind (P bezeichnet das Wahrscheinlichkeitsmaß des Wahrscheinlichkeitsraumes, auf dem die Zufallsvariablen definiert sind). Weiter bezeichnen wir mit $\sigma_i^2 := \mathbb{V}(X_i) = \mathbb{E}((X_i - \mathbb{E}(X_i))^2)$ die Varianz der Zufallsvariable X_i . Dann ist auf Grund der angenommenen Unabhängigkeit der Zufallsvariablen $s_n^2 := \sum_{i=1}^n \sigma_i^2$ die Varianz der Summe $\sum_{i=1}^n X_i$.

Die Frage nach der Gültigkeit eines zentralen Grenzwertsatzes in dieser Situation ist die Frage nach den Bedingungen an die Zufallsvariablen X_i , so dass für jedes $x \in \mathbb{R}$ und für

$$W_n := \frac{1}{s_n} (X_1 + X_2 + \dots + X_n)$$

gilt:

$$\lim_{n \rightarrow \infty} P(W_n \leq x) = \Phi(x) \tag{30}$$

mit $\Phi(x) = \int_{-\infty}^x \varphi_{0,1}(t) dt$ und $\varphi_{0,1}(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$. Wir bemerken, dass $\mathbb{E}(W_n) = 0$ (da $\mathbb{E}(X_i) = 0$ für jedes i) und $\mathbb{V}(W_n) = \frac{1}{s_n^2} \mathbb{V}(\sum_{i=1}^n X_i) = 1$ gilt, also die Zufallsvariable W_n für jedes $n \geq 1$ in den ersten beiden Momenten $\mathbb{E}(W_n)$ und $\mathbb{E}(W_n^2)$ mit den Momenten der Gauß-Verteilung übereinstimmt. Wir kommen auf dieses *Momente-Matching* später zurück.

Um den Beweis nach LINDEBERG transparent zu halten, betrachten wir zunächst den Fall, in dem die Zufallsvariablen X_i die gleiche Verteilung besitzen ($P(X_i \leq t) = P(X_1 \leq t)$ für jedes i) und die Varianz $\mathbb{V}(X_1) = \mathbb{E}((X_1 - \mathbb{E}(X_1))^2) = 1$ ist. Es seien Z_1, Z_2, \dots unabhängige, identisch verteilte Zufallsvariablen, die $N(0, 1)$ -verteilt sind. Dies bedeutet $P(Z_i \leq x) = \Phi(x)$ für alle $x \in \mathbb{R}$. Dann gilt, dass

$$Z = \frac{1}{\sqrt{n}} (Z_1 + Z_2 + \dots + Z_n)$$

auch $N(0, 1)$ verteilt ist. Dies ist eine bekannte Tatsache, die sich leicht beweisen lässt und die wir später noch genauer kommentieren werden. Es wird zentral sein,

dass diese Beobachtung eine charakterisierende Eigenschaft der Normalverteilung ist.

Die wesentliche Idee von LINDBERG ist, in dem obigen W_n sukzessive (Teleskop-Summe) die Summanden X_i durch die normalverteilten Z_i zu ersetzen und somit W_n durch Z zu approximieren. Dieses Austauschverfahren (im Englischen *replacement trick* oder auch *swapping trick* genannt) führt schnell zum Beweis des zentralen Grenzwertsatzes. Dieser Ansatz führt darüberhinaus zu der Möglichkeit, den zentralen Grenzwertsatz für nicht notwendig identisch verteilte Zufallsvariablen zu beweisen, allerdings unter einer zusätzlichen Bedingung, die seit den Arbeiten von LINDBERG nach ihm benannt ist.

Eine kleine technische Vorbereitung ist die folgende: wir schreiben $P(W_n \leq x) = \mathbb{E}(1_{W_n \leq x})$ mit der Notation der Indikatorfunktion $1_{W_n \leq x}$, die den Wert 1 auf dem Ereignis $\{W_n \leq x\}$ annimmt, und sonst den Wert 0. Wir wollen zeigen, dass $\lim_{n \rightarrow \infty} \mathbb{E}(1_{W_n \leq x}) = \mathbb{E}(1_{Z \leq x})$ gilt. Dazu approximieren wir die Indikatorfunktion $1_{(-\infty, x]}$ durch eine hinreichend glatte Funktion f . Für unsere Ziele genügt es ein f zu wählen, welches dreimal differenzierbar ist und dessen Ableitungen stetig und beschränkt sind (diese *Testfunktionen* bilden eine Konvergenz-determinierende Klasse für die Konvergenz in (30), d.h. kann man die Konvergenz für diese Funktionenklasse beweisen, so bekommt sie auch in der geforderten Allgemeinheit).

Zu zeigen ist für jede solche Funktion f

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(W_n)) = \mathbb{E}(f(Z)).$$

Mittels TAYLOR-Entwicklung existiert nun ein η und ein $g(h)$ mit

$$g(h) \leq \eta \min(h^2, |h|^3) \quad (31)$$

und

$$|f(x + h_1) - f(x + h_2) - f'(x)(h_1 - h_2) - \frac{1}{2}f''(x)(h_1^2 - h_2^2)| \leq g(h_1) + g(h_2). \quad (32)$$

Hierbei ist η eine Konstante, die durch die Werte $M_i := \sup_{x \in \mathbb{R}} |f^{(i)}(x)| < \infty$ für $i = 0, 1, 2, 3$ bestimmt ist. Alternativ ist

$$g(h) \leq \eta |h|^3 \quad (33)$$

mit $\eta = M_3/6$ eine simplere Abschätzung. Nun definieren wir $T_k := X_1 + \dots + X_{k-1} + Z_{k+1} + \dots + Z_n$. Also ist $\frac{1}{\sqrt{n}}(T_n + X_n) = W_n$ und $\frac{1}{\sqrt{n}}(T_1 + Z_1) = Z$ und mittels *Teleskop-Summen* Bildung folgt

$$\begin{aligned} \mathbb{E}(f(W_n) - f(Z)) &= \mathbb{E}\left(f\left(\frac{1}{\sqrt{n}}(T_n + X_n)\right) - f\left(\frac{1}{\sqrt{n}}(T_1 + Z_1)\right)\right) \\ &= \sum_{k=1}^n \mathbb{E}\left(f\left(\frac{1}{\sqrt{n}}(T_k + X_k)\right) - f\left(\frac{1}{\sqrt{n}}(T_k + Z_k)\right)\right). \end{aligned}$$

Da $\mathbb{E}(X_k) = \mathbb{E}(Z_k) = 0$ und $\mathbb{E}(X_k^2) = \mathbb{E}(Z_k^2) (= 1)$ folgt

$$\mathbb{E}(X_k - Z_k) = \mathbb{E}(X_k^2 - Z_k^2) = 0$$

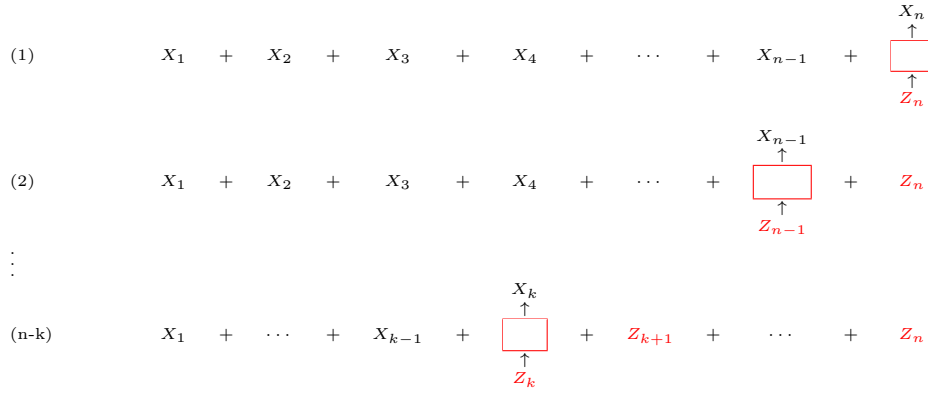


Abbildung 1: Schema der LINDBERG-Methode

und somit wegen der stochastischen Unabhängigkeit von T_k zu X_k und Z_k (für jedes k)

$$|\mathbb{E}(f(W_n) - f(Z))| \leq \sum_{k=1}^n \left| \mathbb{E} \left[f\left(\frac{1}{\sqrt{n}}(T_k + X_k)\right) - f\left(\frac{1}{\sqrt{n}}(T_k + Z_k)\right) - f'\left(\frac{T_k}{\sqrt{n}}\right) \frac{1}{\sqrt{n}}(X_k - Z_k) - \frac{1}{2} f''\left(\frac{T_k}{\sqrt{n}}\right) \frac{1}{n}(X_k^2 - Z_k^2) \right] \right|.$$

Die Unabhängigkeit impliziert etwa $\mathbb{E}(f'(T_k)(X_k - Z_k)) = \mathbb{E}(f'(T_k)) \mathbb{E}((X_k - Z_k))$.

Die vorgestellte Abschätzung basiert alleine auf der Annahme an die X_k , in den ersten beiden Momenten $\mathbb{E}(X_k)$ und $\mathbb{E}(X_k^2)$ mit den Momenten einer $N(0, 1)$ -verteilten Zufallsvariable übereinzustimmen. Es folgt mit der obigen TAYLOR-Vorbereitung (32) und der Annahme, dass alle X_i bzw. Z_i identisch verteilt sind:

$$|\mathbb{E}(f(W_n) - f(Z))| \leq n \mathbb{E}g\left(\frac{X_1}{\sqrt{n}}\right) + n \mathbb{E}g\left(\frac{Z_1}{\sqrt{n}}\right). \quad (34)$$

Nun wenden wir die Abschätzung (31) für g an und erhalten durch Zerlegung des Integrals

$$n \mathbb{E}g\left(\frac{X_1}{\sqrt{n}}\right) \leq n \left[\eta \int_{|X_1| \leq \varepsilon \sqrt{n}} \left| \frac{X_1}{\sqrt{n}} \right|^3 dP + \eta \int_{|X_1| > \varepsilon \sqrt{n}} \left| \frac{X_1}{\sqrt{n}} \right|^2 dP \right].$$

Der erste Summand kann durch $n \eta \varepsilon \int_{|X_1| \leq \varepsilon \sqrt{n}} \left| \frac{X_1}{\sqrt{n}} \right|^2 dP \leq \eta \varepsilon$ abgeschätzt werden, da $\int X_1^2 dP = 1$ nach Voraussetzung. Weiter steigt die Folge der Ereignisse $\{|X_1| > \varepsilon \sqrt{n}\}$ für wachsende n gegen die leere Menge ab, so dass der zweite Summand gegen Null für $n \rightarrow \infty$ konvergiert. Da die gleiche Überlegung für den Term $n \mathbb{E}g\left(\frac{Z_1}{\sqrt{n}}\right)$ verwendet werden kann, folgt bereits der Beweis des zentralen Grenzwertsatzes für Partialsummen unabhängiger, identisch verteilter Zufallsvariablen mit existierendem Erwartungswert und endlicher Varianz. Wir schauen auf den Beweis zurück:

Dieser Beweis, die Lindeberg-Methode hat – wie schon oben erwähnt – das Potenzial auf allgemeinere Situation ausgeweitet zu werden:

1. Eine Inspektion des obigen Beweises zeigt schnell, dass die Annahme *identisch* verteilter Zufallsvariablen nicht notwendig ist. Entlang der obigen Argumente entdeckt man die zweite fundamentale Beobachtung von LINDEBERG, die sogenannte LINDEBERG-Bedingung. Sind die Zufallsvariablen *nicht identisch verteilt*, so ergibt sich unmittelbar anstelle von (34):

$$|\mathbb{E}(f(W_n) - f(Z))| \leq \sum_{i=1}^n \left(\mathbb{E}g\left(\frac{X_i}{s_n}\right) + \mathbb{E}g\left(\frac{Z_i}{s_n}\right) \right). \quad (35)$$

Hierbei sind nun die Z_i unabhängige, normalverteilte Zufallsvariablen mit Erwartungswert 0 und Varianz $\sigma_i^2 = \mathbb{V}(X_i)$ und s_n^2 ist erneut die Varianz der Summe $\sum_i X_i$ bzw. $\sum_i Z_i$ (Unabhängigkeit). Die gleiche Zerlegung des Integrals führt zu der Abschätzung

$$\eta \sum_{i=1}^n \int_{|X_i| \leq \varepsilon s_n} \left| \frac{X_i}{s_n} \right|^3 dP + \eta \sum_{i=1}^n \int_{|X_i| > \varepsilon s_n} \left| \frac{X_i}{s_n} \right|^2 dP. \quad (36)$$

Der erste Summand kann nun analog durch $\eta \varepsilon$ abgeschätzt werden. Für den zweiten Summanden erhält man

$$\eta \frac{1}{s_n^2} \sum_{i=1}^n \int_{|X_i| > \varepsilon s_n} X_i^2 dP. \quad (37)$$

Man muss nun fordern, dass dieser Ausdruck gegen Null konvergiert für $n \rightarrow \infty$, und genau dies ist die LINDEBERG-Bedingung. Wenn man dann noch zeigt, dass die Gauß-Variablen Z_i diese Bedingung erfüllen, was wir hier nicht ausführen, so haben wir bewiesen:

Satz 4.5 (Zentraler Grenzwertsatz von LINDEBERG, 1922) *Gegeben seien unabhängige, reellwertige Zufallsvariablen X_1, X_2, \dots mit $\mathbb{E}(X_i) = 0$ und $\sigma_i^2 := \mathbb{V}(X_i) > 0$ für jedes i . Es sei $s_n^2 := \sum_{i=1}^n \sigma_i^2$. Gilt für jedes $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \int_{|X_i| > \varepsilon s_n} X_i^2 dP = 0, \quad (38)$$

dann folgt die Aussage (30).

2. Eine weitere Inspektion des obigen Beweises zeigt, dass die LINDEBERG-Methode eine Information über die *Konvergenzgeschwindigkeit* im zentralen Grenzwertsatz enthält. Dies betrachten wir nur im Sonderfall identisch verteilter Zufallsgrößen X_i . Tatsächlich hat LINDEBERG in seiner ersten Arbeit im Jahre 1920 eine stärkere Bedingung an die Zufallsvariablen X_i gestellt. Er forderte die Endlichkeit des dritten absoluten Momentes $\mathbb{E}|X_i|^3$ für jedes X_i . Erst in den Arbeiten aus dem Jahre 1922 konnte er diese Forderung zur Bedingung (38) abschwächen. Schauen wir auf den Spezialfall identisch verteilter Zufallsvariablen und nehmen wir $\mathbb{E}|X_1|^3 < \infty$ an, so kann jeder Summand in (34) mittels (33) wie folgt abgeschätzt werden:

$$\mathbb{E}g(X_1/\sqrt{n}) \leq \eta \mathbb{E}|X_1|^3 \frac{1}{\sqrt{n}}.$$

Wir erhalten also im identisch verteilten Fall für unsere Testfunktionen f mit $\|f'''\| \leq \eta$:

$$|\mathbb{E}(f(W_n) - f(Z))| = \eta \mathcal{O}\left(\frac{1}{\sqrt{n}} \mathbb{E}|X_1|^3\right).$$

Dies ist eine schwache Version einer *Konvergenzrate* im Sinne der Theorie von BERRY und ESSÉEN. *Schwach* bedeutet, dass dieses Resultat für die beschriebene Testfunktionen-Klasse nicht impliziert, dass $\sup_{x \in \mathbb{R}} |P(W_n \leq x) - \Phi(x)|$ (der sogenannte KOLMOGOROV-Abstand) ebenfalls in der Größenordnung $n^{-1/2}$ liegt. Es ist aber bekannt, dass die optimale Konvergenzrate für den KOLMOGOROV-Abstand in unserer Situation $\mathcal{O}(n^{-1/2})$ ist. Dies nennt man die BERRY-ESSÉEN-Rate. Wir kommen darauf in Kapitel 4 zurück.

Man kann darüberhinaus leicht erkennen, dass bei Berücksichtigung höherer Ordnungsterme der TAYLOR-Entwicklung von f und der Forderung der Endlichkeit und Übereinstimmung höherer Momente der X_i mit denen der Z_i die Konvergenzrate verbessert werden kann. Dieser Aspekt wird in unserem Kapitel zur Theorie der Zufallsmatrizen eine wichtige Rolle spielen.

3. Schliesslich mag man sich auch fragen, wieso im Zentralen Grenzwertsatz die Normalverteilung als Limes auftaucht und keine andere Grenzverteilung. Wesentlich ist dabei offensichtlich die Eigenschaft, dass die Summe unabhängiger, normal-verteilter Zufallsvariablen wieder normal-verteilt ist: Die Summe von n unabhängigen $N(0, 1)$ verteilten Zufallsvariablen ist $N(0, n)$ verteilt. Diese Eigenschaft ist in der Literatur als *unbegrenzte Teilbarkeit* der Normalverteilung bekannt. Es gibt allerdings eine ganze Klasse von Verteilungen, die diese Eigenschaft mit der Normalverteilung teilen. Was ist dann das Besondere an der Normalverteilung? Wichtig ist, dass mit den obigen Z_i die Zufallsvariable $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$ wieder $N(0, 1)$ -verteilt ist, d.h. auf der Verteilungsebene und der gewählten Skala $1/\sqrt{n}$ ist die $N(0, 1)$ -Verteilung ein Fixpunkt der Abbildung

$$g(X_1, \dots, X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i. \quad (39)$$

Ihre herausragende Rolle resultiert in gewisser Hinsicht daraus, dass sie der *einzigste* derartige Fixpunkt ist. Die Inspektion des LINDEBERG-Beweises zeigt aber gleichzeitig, dass der teleskopische Austauschprozess auch für andere *Vergleichssummen* $\frac{1}{s_n} \sum_{i=1}^n Z_i$ verwendet werden kann, man wird nur eine andere Limesverteilung beobachten (was im Jargon der Wahrscheinlichkeitstheorie ein *nichtzentraler* Grenzwertsatz genannt wird). Dieser Aspekt wird uns in Kapitel 4 und 5 begegnen.

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 Supremumsmetrik

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

versehen. Um den gewünschten Konvergenzsatz beweisen zu können, müssen wir eine ganze Menge Hilfsmittel bereitstellen. Das erste ist

Satz 4.6 $C([0, 1], d)$ 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 [?]. Die Separabilität folgt sofort aus dem Weierstraßschen Approximationssatz: Die Polynome liegen dicht in $C([0, 1], d)$, die Polynome mit rationalen Koeffizienten liegen wiederum dicht in den Polynomen. Die Polynome mit rationalen Koeffizienten sind aber nach dem Cantor-Verfahren abzählbar. \square

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 $0 \leq t_1 < \dots < t_m \leq 1$

$$\begin{aligned} \pi_{t_1, \dots, t_m} : C([0, 1]) &\rightarrow \mathbb{R}^m \\ f &\mapsto (f(t_1), \dots, f(t_m)). \end{aligned}$$

Dann gilt:

Lemma 4.7 *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

$$\mathcal{B}_\varepsilon(f) := \{g \in C[0, 1] : d(f, g) \leq \varepsilon\}.$$

Dann ist, da f stetig,

$$\mathcal{B}_\varepsilon(f) := \bigcap_{t \in [0, 1] \cap \mathbb{Q}} \{g \in C[0, 1] : |g(t) - f(t)| \leq \varepsilon\} = \bigcap_{t \in [0, 1] \cap \mathbb{Q}} \pi_t^{-1}(\mathcal{B}_\varepsilon(f(t))) \in \mathcal{B}'.$$

Da $C[0, 1]$ separabel ist, ist jede offene Menge abzählbare Vereinigung von derartigen Kugeln, also in \mathcal{B}' . \square

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 4.8 *Ein Maß auf $(C[0, 1], \mathcal{B}_C)$ mit*

- $\mu(C[0, 1]) = 1$,
- $\mu\pi_{t_1, \dots, t_m}^{-1}$ 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 4.9 *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 $\text{Cov}(X, X + Y) = \mathbb{V}(X) = s$.*

Unter μ hat ein Pfad \mathcal{B}_t offenbar die Eigenschaften:

1. $\mu(B_t \text{ ist stetig } \forall t) = 1$.
2. $\mu(B_t \leq \alpha) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{-\frac{s^2}{2t}} ds$.
3. Für $0 = t_0 < t_1 < \dots < t_m \leq 1$ und $\alpha_1, \dots, \alpha_m \in \mathbb{R}$, $m \in \mathbb{N}$, gilt $\mu(B_{t_i} - B_{t_{i-1}} \leq \alpha_i, i = 1, \dots, m) = \prod_{i=1}^m \mu(B_{t_i} - B_{t_{i-1}} \leq \alpha_i)$.

Definition 4.10 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. Dass es ihn gibt, haben wir allerdings schon zu Beginn des Kapitels gesehen.

Satz 4.11 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.

Wir werden die Existenz des Wienermaßes auf einem weiteren Weg zeigen, 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, müssen 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 4.12 (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

$$\mu_n(K_\varepsilon^c) \leq \varepsilon$$

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 4.13 (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 4.20 bedarf einiger Vorbereitung. Zunächst verwenden wir eine Variante des Rieszschen Darstellungssatzes, den wir in anderer Form schon im Beweis des Satzes von Radon und Nikodym kennengelernt hatten. Dazu bringen wir zunächst die folgende

Definition 4.14 *Es sei S ein metrischer Raum. Eine Abbildung $\Lambda : C(S) \rightarrow \mathbb{R}$ 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 4.15 *(Darstellungssatz von Riesz)*

Es sei S ein kompakter metrischer Raum. Dann existiert zu jedem normierten, nicht-negativen linearen Funktional $\Lambda : C(S) \rightarrow \mathbb{R}$ ein eindeutig bestimmtes Wahrscheinlichkeitsmaß μ auf (S, \mathcal{B}_S) mit

$$\Lambda(f) = \int f d\mu \quad \forall f \in C(S). \quad (40)$$

Jedes Wahrscheinlichkeitsmaß auf (S, \mathcal{B}_S) bestimmt vermöge (42) ein normiertes, nicht-negatives lineares Funktional auf $C(S)$.

Der Beweis von Satz 4.22 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 4.22 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

$$\mathcal{M}^1(S) := \{\mu : \mu \text{ ist ein Wahrscheinlichkeitsmaß auf } (S, \mathcal{B}_S)\} \quad (41)$$

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 4.16 *Ist S ein kompakter, metrischer Raum, so ist $\mathcal{M}^1(S)$ schwach folgenkompakt.*

Bemerkung 4.17 *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

$$\lim_{k \rightarrow \infty} \int f_j d\mu_{n_k} = a_j$$

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

$$\left| \int f d\mu_{n_k} - \int f d\mu_{n_m} \right| \leq \left| \int f_j d\mu_{n_k} - \int f_j d\mu_{n_m} \right| + \underbrace{\int |f - f_j| d\mu_{n_k}}_{\leq \varepsilon} + \underbrace{\int |f - f_j| d\mu_{n_m}}_{\leq \varepsilon}.$$

Der erste Summand konvergiert gegen Null für $k, m \rightarrow \infty$, also

$$\lim_{k, m \rightarrow \infty} \left| \int f d\mu_{n_k} - \int f d\mu_{n_m} \right| = 0,$$

und somit konvergiert $\int f d\mu_{n_k}$ für $k \rightarrow \infty$ für jedes $f \in C(S)$. Setzen wir

$$\Lambda(f) := \lim_{k \rightarrow \infty} \int f d\mu_{n_k}, \quad f \in C(S),$$

so ist Λ ein nicht-negatives lineares Funktional auf $C(S)$ mit $\Lambda(1) = 1$, also existiert nach dem Rieszschen Darstellungssatz ein $\mu \in \mathcal{M}^1(S)$ mit

$$\Lambda(f) = \int f d\mu \quad \forall f \in C(S),$$

womit die schwache Konvergenz von $(\mu_{n_k})_k$ gegen 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 Folgenglieder μ_n "bis auf ein $\varepsilon > 0$ " die gleiche kompakte Menge sehen. Um diese Information für uns nutzbar zu machen benötigen wir aber noch zwei topologische Aussagen:

Satz 4.18 (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 . $h : S \rightarrow [0, 1]^{\mathbb{N}}$ sei definiert durch die n -ten Koordinatenfunktionen

$$h_n(x) = \frac{d(x, s_n)}{1 + d(x, s_n)}, \quad x \in S, \quad n \in \mathbb{N}.$$

Es ist eine schöne Übung zu sehen, dass dies ein Homöomorphismus ist. \square

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 4.19 (*Tychonov*)
 $(K^{\mathbb{N}}, \bar{d})$ ist kompakt.

Bemerkung 4.20 *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 4.26: 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 $(x_{n_1,m}^{(1)})_m$ konvergent ist; von dieser Folge gibt es wieder eine Teilfolge $(x_{n_2,m})$, so dass $(x_{n_2,m}^{(2)})_m$ 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 4.25 und 4.26 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 4.23 ins Spiel bringen.

Beweis von Satz 4.20: 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 4.23 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

$$\limsup_{n \rightarrow \infty} \tilde{\mu}_{n_k}(K_r) \leq \nu(K_r), \quad r \in \mathbb{N}.$$

Dann folgt auch $\nu(K_r) \geq 1 - 1/r$ für $r \in \mathbb{N}$.

Sei $E_0 := \bigcup_r K_r$, dann ist $E_0 \subset S$, $E_0 \in \mathcal{B}_{\tilde{S}}$ und $\nu(E_0) = 1$. Wir behaupten nun, dass es ein $\mu \in \mathcal{M}_1(S)$ gibt mit $\tilde{\mu} = \nu$.

Es ist $\mathcal{B}_S = \mathcal{B}_{\tilde{S}} \cap S$. Für jedes $A \in \mathcal{B}_S$ existiert ein $B_1 \in \mathcal{B}_{\tilde{S}}$ mit $A = B_1 \cap S$. Sei $\mu(A) := \nu(B_1)$. Wenn $B_2 \in \mathcal{B}_{\tilde{S}}$ und $A = B_2 \cap S$, dann ist $B_1 \Delta B_2 \subset S^c \subset E_0^c$ und $\nu(B_1 \Delta B_2) = 0$, also $\nu(B_1) = \nu(B_2)$, also ist $\mu(A)$ wohldefiniert.

Es sei nun $(A_i)_i$ mit $A_i = b_i \cap S_i$, $i \in \mathbb{N}$, 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

$$\begin{aligned} \mu\left(\bigcup_i A_i\right) &= \nu\left(\bigcup_i B_i\right) = \nu\left(\bigcup_i (B_i \cap E_0)\right) \\ &= \sum_i \nu(B_i \cap E_0) = \sum_i \nu(B_i) = \sum_i \mu(A_i). \end{aligned}$$

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

$$\limsup_{k \rightarrow \infty} \mu_{n_k}(C) = \limsup_{k \rightarrow \infty} \tilde{\mu}_{n_k}(D) \leq \tilde{\mu}(D) = \mu(C).$$

Das Portmanteau-Theorem liefert $\mu_{n_k} \xrightarrow{w} \mu$. Damit ist der erste Teil des Satzes bewiesen.

Sei nun für die Umkehrung S vollständig und separabel und μ_n besitze eine schwach gegen ein W-Maß μ konvergente Teilfolge. Zunächst beginnen wir mit einer beliebigen offenen Überdeckung $(U_i)_{i \in \mathbb{N}}$. Dann gibt es zu jedem $\varepsilon > 0$ ein k , sodass

$$\mu_m\left(\bigcup_{i=1}^k U_i\right) > 1 - \varepsilon$$

für alle $m \in \mathbb{N}$ gilt. Wäre dies nämlich falsch, so gäbe es für jedes k ein μ_m , sodass

$$\mu_m\left(\bigcup_{i=1}^k U_i\right) \leq 1 - \varepsilon.$$

Nun hat (μ_m) eine gegen ein W-Maß μ konvergente Teilfolge, o.B.d.A. sei diese auch (μ_m) . Es ist für jedes k

$$\mu\left(\bigcup_{i=1}^k U_i\right) \leq \liminf_{m \rightarrow \infty} \mu_m\left(\bigcup_{i=1}^k U_i\right) \leq \liminf_{m \rightarrow \infty} \mu_m\left(\bigcup_{i=1}^m U_i\right) \leq 1 - \varepsilon.$$

Da aber $\bigcup_{i=1}^{\infty} U_i = S$, konvergiert $\mu(\bigcup_{i=1}^k U_i)$ für $k \rightarrow \infty$ gegen 1. Das ist ein Widerspruch.

Da S separabel ist, existiert eine Folge offener Bälle B_{n1}, B_{n2}, \dots mit Radius $1/n$, so dass

$$S = \bigcup_{j=1}^{\infty} B_{nj}, \quad n \in \mathbb{N}.$$

Sei nun $n \in \mathbb{N}$ fest. Wir können nun K

$$\mu_i\left(\bigcup_{j=1}^{k_n} B_{nj}\right) > 1 - \frac{\varepsilon}{2^n}, \quad n \in \mathbb{N}.$$

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 $i \in \mathbb{N}$, denn

$$\begin{aligned} \mu_i(S \setminus K) &= \mu_i\left(\bigcup_{n=1}^{\infty} C_n^c\right) \\ &\leq \sum_{n=1}^{\infty} \mu_i(C_n^c) \\ &\leq \sum_{n=1}^{\infty} \varepsilon 2^{-n} = \varepsilon. \end{aligned}$$

Tatsächlich ist K auch kompakt, denn es ist abgeschlossen und totalbeschränkt. Damit ist der Satz von Prohorov bewiesen. \square

Der Satz von Prohorov 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 Arzela und Ascoli; hierzu zunächst folgende Definition:

Definition 4.21 *Es sei*

$$f : [0, 1] \rightarrow \mathbb{R}$$

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 4.22 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 gleichmäßigen Stetigkeit einer Funktion $f \in C[0, 1]$, dass

$$\lim_{\delta \downarrow 0} \omega_\delta(f) = 0$$

gilt.

Nun der angekündigte

Satz 4.23 (Satz von Arzela-Ascoli)

Eine Teilmenge $A \subset C[0, 1]$ hat genau dann kompakten Abschluss, wenn

$$(i) \sup\{|f(0)|, f \in A\} < \infty \text{ ist und}$$

$$(ii) \lim_{\delta \rightarrow 0} \sup_{f \in A} \omega_\delta(f) = 0$$

gelten.

Wir bereiten den Beweis durch ein Kriterium für Kompaktheit von Mengen in metrischen Räumen vor.

Satz 4.24 Eine Teilmenge eines metrischen Raumes (X, d) ist genau dann kompakt, wenn sie vollständig und totalbeschränkt ist. Dabei heißt $K \subset X$ totalbeschränkt, wenn es zu jedem $r > 0$ ein $m \in \mathbb{N}$ und $x_0, \dots, x_m \in K$ gibt mit $K \subset \bigcup_{k=0}^m B(x_k, r)$ (womit jede totalbeschränkte Menge beschränkt ist).

Beweis: Es sei $K \subset X$ kompakt, $(x_j)_j$ sei eine Cauchyfolge in K . K ist folgenkompakt (denn eine Teilmenge eines metrischen Raumes ist genau dann kompakt, wenn sie folgenkompakt ist, Analysis I), also besitzt $(x_j)_j$ eine in K konvergente Teilfolge. Damit konvergiert die Folge (denn besitzt eine Cauchyfolge eine konvergente Teilfolge, so ist sie selbst konvergent, Analysis I) in K , also ist K vollständig. Für jedes $r > 0$ ist $\{B(x, r), x \in K\}$ eine offene Überdeckung von K . Da K kompakt, gibt es eine endliche Teilüberdeckung, also ist K auch totalbeschränkt.

Sei nun K vollständig und totalbeschränkt. $(x_j)_j$ sei eine Folge in K . Zu jedem $n \in \mathbb{N}$ existieren endlich viele Bälle mit Mittelpunkten in K und Radius $1/n$, die K überdecken. Es existiert also eine Teilfolge $(x_{1,j})_j$ von $(x_j)_j$, die ganz in einem Ball mit

Radius $1/2$ 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 *Diagonalforngenprinzip* 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 $f_1, \dots, f_n \in A$ mit $d(f, f_j) < \varepsilon/3$ für ein $j \in \{1, \dots, n\}$ für alle $f \in A$. Jedes f_j in $C[0,1]$ ist gleichmäßig stetig, also gilt für die endliche Menge $\{f_1, \dots, f_n\}$: Wähle $\delta > 0$, so dass aus $|x - y| < \delta$ folgt $|f_j(x) - f_j(y)| < \varepsilon/3$ für alle $j = 1, \dots, n$ und $x, y \in [0,1]$. Also ist $|f(x) - f(y)| < \varepsilon$ für alle $f \in A$, somit gilt $\lim_{\delta \rightarrow 0} \sup_{f \in A} \omega_\delta(f) = 0$. 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

$$|f(t)| \leq |f(0)| + \sum_{i=1}^k \left| f\left(\frac{i}{k}t\right) - f\left(\frac{i-1}{k}t\right) \right|,$$

folgt mit (i)

$$\sup_{t \in [0,1]} \sup_{f \in A} |f(t)| < \infty. \quad (42)$$

Wir zeigen nun, dass aus (ii) und (45) folgt, dass A totalbeschränkt ist, also auch \bar{A} . Nun ist $C[0,1]$ vollständig, also auch \bar{A} , damit ist \bar{A} dann kompakt.

Sei $\varepsilon > 0$ und

$$\alpha = \sup_{t \in [0,1]} \sup_{f \in A} |f(t)|.$$

Ferner sei $H := \{\frac{u}{v}\alpha, u = 0, \pm 1, \pm 2, \dots, \pm v, v \in \mathbb{N}\}$ 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 $[\frac{i-1}{k}, \frac{i}{k}]$, $i = 1, \dots, k$, linear sind und Werte aus H an den Endpunkten $\frac{i}{k}$, $i = 0, \dots, k$, annehmen. B ist endlich (besteht aus $(2v+1)^{k+1}$ 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

$$|f(\frac{i}{k}) - g(\frac{i}{k})| < \varepsilon, \quad i = 0, \dots, k. \quad (43)$$

Da $\omega_{1/k}(f) < \varepsilon$ und g linear in jedem Teilintervall $[\frac{i-1}{k}, \frac{i}{k}]$ ist, folgt aus (45) $d(f, g) < 2\varepsilon$. Dies war zu zeigen. \square

Satz 4.30, 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 4.25 *Eine Folge $(\nu_n)_n$ von Wahrscheinlichkeitsmaßen auf (C, \mathcal{B}_C) ist genau dann straff, wenn*

$$\lim_{a \nearrow \infty} \sup_n \nu_n(\{f : |f(0)| > a\}) = 0 \quad \text{und} \quad (44)$$

$$\lim_{\delta \searrow 0} \limsup_{n \rightarrow \infty} \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) = 0 \quad \text{für alle } \varepsilon > 0 \quad (45)$$

gelten.

Nach obiger Bemerkung ist $\{f : \omega_\delta(f) \geq \varepsilon\} \in \mathcal{B}_C$. Die Bedingungen (46) und (47) in Satz 4.32 können wie folgt übersetzt werden:

$$\forall \eta > 0 \exists a > 0 \forall n \in \mathbb{N} : \quad \nu_n(\{|f(0)| > a\}) \leq \eta, \quad (46)$$

$$\forall \varepsilon > 0, \eta > 0 \exists \delta > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \quad \nu_n(\{\omega_\delta(f) \geq \varepsilon\}) \leq \eta. \quad (47)$$

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

$$Y_n(\omega, t) := \frac{S_{[nt]}(\omega)}{\sqrt{n}} + \frac{nt - [nt]}{\sqrt{n}} X_{[nt]+1}(\omega) \quad (48)$$

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 $\mathcal{F} - \mathcal{B}$ -messbare Abbildung. Lemma 4.14 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

$$\begin{aligned} \pi_{t_1, \dots, t_m} : C &\rightarrow \mathbb{R}^m \\ f &\mapsto (f(t_1), \dots, f(t_m)) \end{aligned}$$

für jedes m -Tupel $0 \leq t_1 < t_2 < \dots < t_m \leq 1$ stetig ist, ist für die Verteilungskonvergenz der Y_n zumindest die schwache Konvergenz der endlich-dimensionalen Verteilungen $\mu_n \circ \pi_{t_1, \dots, t_m}^{-1}$ notwendig. Dies klärt der folgende

Satz 4.26 *Für jedes $m \in \mathbb{N}$ und $0 \leq t_1 < \dots < t_m \leq 1$ konvergiert die Folge $\mu_n \circ \pi_{t_1, \dots, t_m}^{-1}$ schwach auf $(\mathbb{R}^m, \mathcal{B}^m)$ gegen die m -dimensionale Normalverteilung mit Erwartungswert 0 und Kovarianzmatrix $(\min(t_i, t_j))_{i,j}$.*

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 4.27 *Sei $d \in \mathbb{N}$ und für $j = 1, \dots, d$ 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*

$$\mu_n^{(1)} \otimes \dots \otimes \mu_n^{(d)} \xrightarrow{w} \mu^{(1)} \otimes \dots \otimes \mu^{(d)}$$

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

$$\mathcal{U} := \{(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_d, b_d) : a_j, b_j \in B_j \text{ für } j = 1, \dots, d\}$$

eine Basis der Topologie von \mathbb{R}^d . \mathcal{U} ist durchschnittsstabil und für $(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_d, b_d) \in \mathcal{U}$ gilt wegen dem Portmanteau-Theorem

$$\begin{aligned} \mu_n^{(1)} \otimes \dots \otimes \mu_n^{(d)}(a_1, b_1) \times \dots \times (a_d, b_d) &= \prod_{j=1}^d \mu_n^{(j)}((a_j, b_j)) \xrightarrow{n \rightarrow \infty} \prod_{j=1}^d \mu^{(j)}((a_j, b_j)) \\ &= \mu^{(1)} \otimes \dots \otimes \mu^{(d)}((a_1, b_1) \times \dots \times (a_d, b_d)). \end{aligned}$$

Das Lemma folgt nun, da sich jede offene Menge als abzählbare Vereinigung von Mengen aus \mathcal{U} schreiben lässt. \square

Beweis von Satz 4.33: Wir können annehmen, dass $t_1 > 0$ gilt. Setze $\sum_{i=1}^0 := 0$ und

$$Z_1^{(n)} := \sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}}, \quad Z_2^{(n)} := \sum_{i=[nt_1]+1}^{[nt_2]} \frac{X_i}{\sqrt{n}}, \dots, Z_m^{(n)} := \sum_{i=[nt_{m-1}]+1}^{[nt_m]} \frac{X_i}{\sqrt{n}}.$$

$Z_1^{(n)}, \dots, Z_m^{(n)}$ sind für jedes $n \in \mathbb{N}$ unabhängig. Mit Lemma 4.34 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

$$\lim_{n \rightarrow \infty} P\left(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \leq s\right) = \Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^s e^{-x^2/2} dx.$$

Nun gilt $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = t_j - t_{j-1}$. Für $\varepsilon > 0$ und $s \in \mathbb{R}$ folgt

$$\begin{aligned} \limsup_{n \rightarrow \infty} P\left(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \leq s\right) &\leq \lim_{n \rightarrow \infty} P\left(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \leq \frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\right) \\ &= \Phi\left(\frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\right) \quad \text{und} \\ \liminf_{n \rightarrow \infty} P\left(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \leq s\right) &\geq \lim_{n \rightarrow \infty} P\left(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \leq \frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\right) \\ &= \Phi\left(\frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\right), \\ \text{also } \lim_{n \rightarrow \infty} P\left(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \leq s\right) &= \Phi\left(\frac{s}{\sqrt{t_j - t_{j-1}}}\right). \end{aligned}$$

Dies ist die Verteilungsfunktion der eindimensionalen Normalverteilung mit Erwartungswert 0 und Varianz $t_j - t_{j-1}$. Nach Lemma 4.34 folgt, dass $\mathcal{L}(Z_1^{(n)}, \dots, Z_m^{(n)})$ für $n \rightarrow \infty$ 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 $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$ 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 4.34 konvergiert die Verteilung von

$$f(Z_1^{(n)}, \dots, Z_m^{(n)}) = \left(\sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}}, \sum_{i=1}^{[nt_2]} \frac{X_i}{\sqrt{n}}, \dots, \sum_{i=1}^{[nt_m]} \frac{X_i}{\sqrt{n}}\right)$$

gegen 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

$$\mathbb{E}\left(\sum_{k=1}^i U_k \sum_{s=1}^j U_s\right) = \sum_{k=1}^{\min\{i,j\}} \mathbb{E}(U_k^2) + \sum_{k=1, k \neq s}^i \sum_{s=1}^j \mathbb{E}(U_k U_s) = \sum_{k=1}^{\min\{i,j\}} (t_k - t_{k-1}) = \min\{t_i, t_j\}.$$

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 $\varepsilon > 0$

$$\begin{aligned} P(|(W_1^{(n)}, \dots, W_m^{(n)})| \geq \varepsilon) &\leq P\left(\bigcup_{j=1}^m \{|W_j^{(n)}| \geq \varepsilon/m\}\right) \\ &\leq \sum_{j=1}^m P(|X_{[nt_j]+1}| \geq \sqrt{n}/\varepsilon(m)) = mP(|X_1| \geq \sqrt{n}\varepsilon/m) \rightarrow 0 \end{aligned}$$

für $n \rightarrow \infty$, also konvergiert $(W_1^{(n)}, \dots, W_m^{(n)})$ in Wahrscheinlichkeit gegen 0. Nach Lemma 4.35 (unten) konvergiert dann auch $\mathcal{L}(Y_n(t_1), \dots, Y_n(t_m))$ gegen νf^{-1} . \square

Lemma 4.28 *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 Verteilung gegen μ .*

Beweis: Sei $F \subseteq S$ abgeschlossen und für $\varepsilon > 0$

$$F^\varepsilon := \{x \in S : d(x, F) \leq \varepsilon\}.$$

Dann gilt

$$\limsup_{n \rightarrow \infty} \mathbb{P}(Y_n \in F) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F^\varepsilon) + \limsup_{n \rightarrow \infty} \mathbb{P}(d(X_n, Y_n) \geq \varepsilon) \leq \mu(F^\varepsilon).$$

Da μ stetig ist, folgt $\mu(F^\varepsilon) \downarrow \mu(F)$, wenn $\varepsilon \downarrow 0$ konvergiert und damit die Behauptung. \square

Nun formulieren wir das zentrale Resultat dieses Kapitels:

Satz 4.29 *(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

$$0 \leq t_1 < t_2 < \dots < t_m \leq 1$$

das Maß $\mu \circ \pi_{t_1, \dots, t_m}^{-1}$ 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

$$\mu_n \Rightarrow \mu. \quad (49)$$

Wie schon oben erwähnt ist die Eindeutigkeit von μ (mehr oder weniger) klar. Die Existenz von μ leiten wir aus der Konvergenzaussage ab. (51) ist eine unmittelbare Konsequenz aus dem folgenden Satz und dem darauf folgenden Lemma.

Satz 4.30 Die Folge $(\mu_n)_n$ ist straff.

Lemma 4.31 $(\mu_n)_n$ ist genau dann schwach konvergent gegen μ , wenn jede Teilfolge $(\mu_{n_k})_k$ von $(\mu_n)_n$ eine gegen μ konvergente Teilfolge $(\mu_{n_{k_l}})_{l \in \mathbb{N}}$ besitzt.

Beweis: Dies folgt aus der Definition von schwacher Konvergenz und der Tatsache, dass die entsprechende Aussage für reelle Zahlenfolgen wahr ist. \square

Mithilfe von Satz 4.37 können wir nun Satz 4.36 beweisen.

Beweis von Satz 4.36: Aus der Straffheit von $(\mu_n)_n$ (Satz 4.37) 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 dimensional Verteilungen konvergieren gegen die von μ . Also folgt die Behauptung des Satzes von Donsker aus Lemma 4.38. \square

Es bleibt also Satz 4.37 zu beweisen. Dies ist (leider) noch ein ganzes Stück Arbeit. Ausgangspunkt dabei ist Satz 4.32.

Bemerkung 4.32 $C([0, 1])$ 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 (49) äquivalent zu

$$\forall \varepsilon > 0, \eta > 0 \exists \delta > 0 \quad \forall n \in \mathbb{N} : \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta. \quad (50)$$

Beweis: (Von Satz 4.32) 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 (47) und (48), 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 (47) und (48) erfüllt. Sei $\eta > 0$ vorgegeben. Nach (47) 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

$$K := A \cap \bigcap_{k=1}^{\infty} \{f : \omega_{\delta_k}(f) < 1/k\}$$

kompakten Abschluss und es gilt

$$\nu_n(\bar{K}^c) \leq \nu_n(K^c) \leq \eta/2 + \sum_{k=1}^{\infty} \eta/2^{k+1} = \eta$$

für alle $n \in \mathbb{N}$, was zu zeigen war. \square

Bemerkung 4.33 *Hinreichend für (47) ist $\nu_n(\{f : f(0) = 0\}) = 1$, was für die μ_n im Satz von Donsker erfüllt ist.*

Lemma 4.34 *Hinreichend für (49) ist:*

$$\begin{aligned} \forall \varepsilon, \eta > 0 \exists \delta \in (0, 1), \exists n_0 \in \mathbb{N} : \forall n \geq n_0, \forall t \in [0, 1 - \delta] : \\ \frac{1}{\delta} \nu_n(\{f : \sup_{t \leq s \leq t + \delta} |f(s) - f(t)| \geq \varepsilon\}) \leq \eta. \end{aligned} \quad (51)$$

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 (54). $m \in \mathbb{N}$ sei die kleinste natürliche Zahl n 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 $n \geq n_0$:

$$\begin{aligned} \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) &\leq \sum_{k=0}^{2m-2} \nu_n(\{f : \sup_{\frac{k}{2m} \leq s \leq \frac{k}{2m} + \delta_0} |f(s) - f(\frac{k}{2m})| \geq \varepsilon/2\}) \\ &\leq (2m - 1) \delta_0 \frac{\eta}{3} \leq (2 + \delta_0) \frac{\eta}{3} \leq \eta. \end{aligned}$$

Damit ist (49) gezeigt. □

Bemerkung 4.35 *Die Bedingung in Lemma 4.41 folgt aus der folgenden Aussage: Für alle $\varepsilon > 0$ gilt*

$$\lim_{\delta \searrow 0} \limsup_{n \rightarrow \infty} \sup_{t \in [0, 1 - \delta]} \frac{1}{\delta} \nu_n(\{f : \sup_{t \leq s \leq t + \delta} |f(s) - f(t)| \geq \varepsilon\}) = 0.$$

Die Bedingung aus Bemerkung 4.42 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 $t + \delta = j/n$ (wobei $k < j$) ist

$$\sup_{t \leq s \leq t + \delta} |Y_n(s) - Y_n(t)| = \max_{1 \leq i \leq n\delta} \frac{|S_{k+i} - S_k|}{\sqrt{n}}.$$

Für allgemeine $t \in [0, 1]$ und $\delta \in (0, 1)$ mit $t + \delta \leq 1$ kann man so abschätzen: Es existieren $j, k \in \{0, 1, \dots, n\}$ mit $k < j$ und $\frac{k}{n} \leq t < \frac{k+1}{n}$ sowie $\frac{j-1}{n} < t + \delta \leq \frac{j}{n}$. Dann gilt für jedes $s \in [t, t + \delta]$:

$$\begin{aligned} |Y_n(s) - Y_n(t)| &\leq |Y_n(t) - Y_n(\frac{k}{n})| + \max_{1 \leq i \leq j-k} |Y_n(\frac{k+i}{n}) - Y_n(\frac{k}{n})| \\ &\leq 2 \max_{1 \leq i \leq j-k} |Y_n(\frac{k+i}{n}) - Y_n(\frac{k}{n})|, \end{aligned}$$

also

$$\begin{aligned} \sup_{t \leq s \leq t+\delta} |Y_n(s) - Y_n(t)| &\leq 2 \max_{1 \leq i \leq j-k} |Y_n(\frac{k+i}{n}) - Y_n(\frac{k}{n})| \\ &= 2 \max_{1 \leq i \leq j-k} \left| \sum_{r=k+1}^{k+i} X_r \right| / \sqrt{n}. \end{aligned}$$

Es ist $\frac{j-k-2}{n} \leq \delta$. Für $n \geq \frac{1}{\delta}$ folgt $j - k \leq 3n\delta$. Somit ist die rechte Seite der letzten Ungleichung nicht größer als $2 \max_{1 \leq i \leq 3n\delta} \left| \sum_{r=k+1}^{k+i} X_r \right| / \sqrt{n}$. Die Verteilung dieser Zufallsvariablen hängt nicht von k ab. Für $n \geq \frac{1}{\delta}$ gilt somit

$$\sup_{t \in [0, 1-\delta]} P\left(\sup_{t \leq s \leq t+\delta} |Y_n(s) - Y_n(t)| \geq \varepsilon\right) \leq P\left(\max_{1 \leq i \leq 3n\delta} \frac{|S_i|}{\sqrt{n}} \geq \frac{\varepsilon}{2}\right).$$

Sei $m := [3n\delta]$, so ist $\sqrt{n} \geq \sqrt{m/3\delta}$ und somit

$$P\left(\max_{1 \leq i \leq 3n\delta} \frac{|S_i|}{\sqrt{n}} \geq \frac{\varepsilon}{2}\right) \leq P\left(\max_{1 \leq i \leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{12\delta}}\right).$$

Für jedes feste $\delta > 0$ geht $m \rightarrow \infty$ für $n \rightarrow \infty$. Nach Bemerkung 4.42 müssen wir für jedes $\varepsilon > 0$ zeigen, dass

$$\lim_{\delta \searrow 0} \limsup_{m \rightarrow \infty} \frac{1}{\delta} P\left(\max_{1 \leq i \leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\right) = 0 \quad (52)$$

gilt. Leider hilft die Abschätzung

$$P\left(\max_{1 \leq i \leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\right) \leq \sum_{i=1}^m P\left(\frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\right)$$

analog zum Beweis von Lemma 4.41 nicht. Wir müssen diese Wahrscheinlichkeit wesentlich genauer abschätzen:

Lemma 4.36 *Für alle $\lambda > 0$ und $m \in \mathbb{N}$ gilt*

$$P\left(\max_{1 \leq i \leq m} |S_i| \geq \lambda \sqrt{m}\right) \leq 2P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}).$$

Beweis: Für $\lambda \leq \sqrt{2}$ ist nichts zu zeigen. Sei $\lambda > \sqrt{2}$.

$$A_i := \bigcap_{j=1}^{i-1} \{|S_j| < \lambda\sqrt{m}\} \cap \{|S_i| \geq \lambda\sqrt{m}\}, \quad i = 1, \dots, m.$$

Die A_i sind disjunkt und $A = \{\max_{1 \leq i \leq m} |S_i| \geq \lambda\sqrt{m}\} = \bigcup_{i=1}^m A_i$. Also

$$\begin{aligned} P(A) &= P(A \cap \{|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\}) + P(A \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}) \\ &\leq P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}) + \sum_{j=1}^{m-1} P(A_j \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}), \end{aligned}$$

denn $A_m \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\} = \emptyset$. Weiter gilt

$$A_j \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\} \subset A_j \cap \{|S_m - S_j| \geq \sqrt{2m}\}, \quad j = 1, \dots, m-1.$$

Die Ereignisse A_j und $\{|S_m - S_j| \geq \sqrt{2m}\}$ sind unabhängig, also haben wir

$$P(A) \leq P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}) + \sum_{j=1}^{m-1} P(A_j)P(|S_m - S_j| \geq \sqrt{2m}).$$

Wegen

$$P(|S_m - S_j| \geq \sqrt{2m}) \leq \frac{1}{2m} \mathbb{E}((\sum_{k=j+1}^m X_k)^2) = \frac{1}{2m} \sum_{k=j+1}^m \mathbb{E}(X_k^2) \leq \frac{1}{2}$$

folgt

$$P(A) \leq P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}) + \frac{1}{2} \sum_{j=1}^m P(A_j) = P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}) + \frac{1}{2}P(A),$$

also folgt die Behauptung. \square

Wir schließen mit dem Beweis von (55) ab: Mit Lemma 4.43 und dem zentralen Grenzwertsatz folgt

$$\limsup_{m \rightarrow \infty} \frac{1}{\delta} P\left(\max_{1 \leq i \leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\right) \leq \limsup_{m \rightarrow \infty} \frac{2}{\delta} P\left(\frac{|S_m|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\right) = \frac{2}{\delta} P(|N| \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}),$$

wenn n eine $N(0, 1)$ -verteilte Zufallsgröße bezeichnet. Die Markov-Ungleichung liefert

$$P(|N| \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}) \leq \frac{\mathbb{E}(|B|^3)}{(\frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2})^3}.$$

Dies führt zu (54). Somit ist die Straffheit der Folge $(\mu_n)_n$ bewiesen und somit Satz 4.37, also auch Satz 4.36. \square

Tatsächlich lässt sich die Doobsche Ungleichung aus dem Kapitel über Martingale in ähnoicher Weise verwenden wie Lemma 4.43.

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

$$h : C[0, 1] \rightarrow \mathbb{R}^d$$

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 4.37 *Ist $h : C[0, 1] \rightarrow \mathbb{R}$ 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 4.37 anzuwenden ist nun die, eine geeignete Folge von i.i.d. Zufallsvariablen zu finden, für die sich die Grenzverteilung von $\max_{0 \leq t \leq 1} Y_n(t)$ "leicht" bestimmen lässt. Satz 4.44 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

$$\begin{aligned} h : C[0, 1] &\rightarrow \mathbb{R} \\ f &\mapsto \sup_{0 \leq t \leq 1} f(t). \end{aligned} \tag{53}$$

Wir bemerken, dass das Supremum in (56) 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

$$M_n := \max_{1 \leq i \leq n} S_i, \quad n \in \mathbb{N}.$$

Wir wollen die Verteilung von M_n analysieren. Hierbei beachte man

$$\sup_{0 \leq t \leq 1} Y_n(t) = \max_{0 \leq i \leq n} \frac{S_i}{\sqrt{n}} = \frac{M_n}{\sqrt{n}}.$$

Wir wollen die Folge $(M_n)_n$ auch als Folge der Maximalgewinne beim Münzwurfspiel bezeichnen. Es gilt:

Satz 4.38 Für die Folge $(M_n)_n$ der Maximalgewinne beim Münzwurfspiel gilt für alle $t \geq 0$

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n}{\sqrt{n}} \leq t\right) = 2\Phi(t) - 1.$$

Für $t < 0$ gilt

$$P\left(\frac{M_n}{\sqrt{n}} \leq t\right) = 0.$$

Hierbei bezeichnet Φ die Verteilungsfunktion der Standardnormalverteilung.

Als unmittelbare Konsequenz aus Satz 4.45 und Satz 4.44 ergibt sich

Satz 4.39 Erfüllen die $(X_i)_i$ die Voraussetzungen des Satzes 4.44, so gilt für alle $t \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P\left(\max_{0 \leq i \leq n} \frac{S_i}{\sqrt{n}} \leq t\right) = \max\{2\Phi(t) - 1, 0\}.$$

Für den Beweis von Satz 4.45 bereiten wir das sogenannte *Spiegelungsprinzip/Reflexionsprinzip* vor.

Für $i, j \in \mathbb{Z}$, $i < j$, nennen wir eine Folge $(i, s_i), \dots, (j, s_j)$ mit $s_k \in \mathbb{Z}$, $i \leq k \leq j$, und $|s_{k+1} - s_k| = 1$ für $i \leq k \leq j-1$ einen *Pfad* von (i, s_i) nach (j, s_j) . Oft schreibt man einfach $(s_i, s_{i+1}, \dots, s_j)$. $j - i$ ist die *Länge* des Pfades. Wir sagen, dass ein Pfad $(s_i, s_{i+1}, \dots, s_j)$ die x -Achse berührt, falls ein k mit $i \leq k \leq j$ existiert, für das $s_k = 0$ ist.

Lemma 4.40 (*Reflexionsprinzip*)

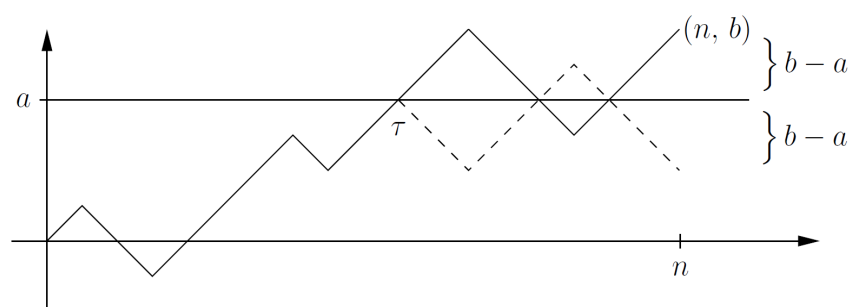
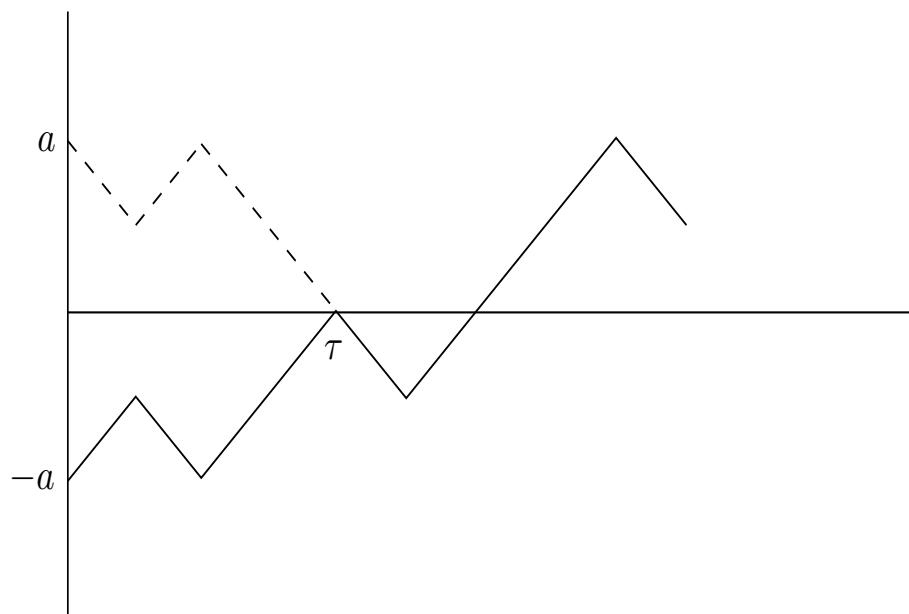
- (i) Es seien $a, b \in \mathbb{N}$ und $i, j \in \mathbb{Z}$ mit $i < j$. Die Anzahl der Pfade von (i, a) nach (j, b) , welche die x -Achse berühren, ist gleich der Anzahl der Pfade von $(i, -a)$ nach (j, b) .
- (ii) Sei $a \in \mathbb{N}$, $b \in \mathbb{Z}$ und $n \in \mathbb{N}$. Die Anzahl der Pfade von $(0, 0)$ nach (n, b) , die $s_j = a$ für ein $j \in \{1, \dots, n\}$ erfüllen, ist gleich der Anzahl der Pfade von $(0, 0)$ nach $(n, 2a - b)$, die $s_j = a$ für ein $j \in \{1, \dots, n\}$ erfüllen.

Beweis: (i) Sei $(s_i = -a, s_{i+1}, \dots, s_{j-1}, s_j = b)$. Dieser Pfad muss die x -Achse berühren. τ sei die kleinste Zahl größer als i , für welche $s_\tau = 0$ gilt. Dann ist

$$(-s_i, -s_{i+1}, \dots, -s_{\tau-1}, s_\tau = 0, s_{\tau+1}, \dots, s_j = b)$$

ein Pfad von (i, a) nach (j, b) , der die x -Achse berührt, und die Zuordnung ist bijektiv (siehe das obige Bild).

Das zweite Bild oben ist das für den Beweis von (ii).



τ ist das erstmalige Erreichen des Wertes a . □

Beweis von Satz 4.45: 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 4.46

$$P(M_n \geq a, S_n = b) = P(M_n \geq a, S_n = 2a - b)$$

für jedes $b \in \mathbb{Z}$. Also ist

$$P(S_n = l + k) = P(M_n \geq k, S_n = k - l).$$

Damit ist

$$\begin{aligned} P(M_n \geq k) &= \sum_{l=-\infty}^{\infty} P(M_n \geq k, S_n = l + k) \\ &= \sum_{l=-\infty}^{-1} P(M_n \geq k, S_n = l + k) + \sum_{l=1}^{\infty} P(S_n = l + k) + P(S_n = k) \\ &= \sum_{l=1}^{\infty} P(M_n \geq k, S_n = l - k) + \sum_{l=1}^{\infty} P(S_n = l + k) + P(S_n = k) \\ &= 2P(S_n > k) + P(S_n = k) \\ &= 2P(S_n \geq k) - P(S_n = k). \end{aligned}$$

Sei $t \in \mathbb{R}_+$. Für $n \in \mathbb{N}$ bezeichne k_n die kleinste 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

$$\lim_{n \rightarrow \infty} P(S_n \geq k_n) = \nu_{0,1}([t, \infty)).$$

Wegen $t\sqrt{n} \leq k_n < t\sqrt{n} + 1$ gilt weiter für jedes $\varepsilon > 0$ und alle $n \in \mathbb{N}$ mit $1/\sqrt{n} \leq \varepsilon$

$$\{S_n = k_n\} = \left\{ \frac{S_n}{\sqrt{n}} = \frac{k_n}{\sqrt{n}} \right\} \subset \left\{ t \leq \frac{S_n}{\sqrt{n}} < t + \varepsilon \right\},$$

und daraus folgt

$$\limsup_{n \rightarrow \infty} P(S_n = k_n) \leq \int_t^{t+\varepsilon} g_{0,1}(x) dx \quad \forall \varepsilon \geq 0,$$

also

$$\lim_{n \rightarrow \infty} P(S_n = k_n) = 0.$$

Zusammen erhalten wir

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n}{\sqrt{n}} \geq t\right) = 2\nu_{0,1}([t, \infty)) = \left(\frac{2}{\pi}\right)^{1/2} \int_t^{\infty} e^{-x^2/2} dx = 2(1 - \Phi(t)),$$

womit die Behauptung des Satzes folgt. \square

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 4.44 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 4.41 *g ist $\mathcal{B}_C/\mathcal{B}$ -messbar und $\mu(D_g) = 0$, wobei μ das Wiener-Maß bezeichnet und $D_g = \{x : g \text{ ist unstetig in } x\}$.*

Beweis: Es sei $\psi : C[0, 1] \times [0, 1] \rightarrow \mathbb{R}$ 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 4.49 unten

$$\mathcal{B}_{C \times [0, 1]} = \mathcal{B}_C \otimes \mathcal{B}_{[0, 1]}.$$

Also ist $\psi : \mathcal{B}_C \otimes \mathcal{B}_{[0, 1]}/\mathcal{B}$ -messbar.

Sei nun $A = \{(f, t) : f(t) \geq 0\} = \psi^{-1}([0, \infty)) \in \mathcal{B}_C \otimes \mathcal{B}_{[0, 1]}$. 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 $d(f_n, f) \rightarrow 0$, so gilt $1_{[0, \infty)}(f_n(t)) \rightarrow 1_{[0, \infty)}(f(t))$ für λ -fast alle $t \in [0, 1]$. Nach dem Satz von der dominierten Konvergenz folgt

$$g(f_n) \rightarrow g(f).$$

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 $f \mapsto \lambda(\{t : f(t) = 0\})$ 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. \square

Hierbei haben wir von dem folgenden Lemma Gebrauch gemacht:

Lemma 4.42 *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

$$\mathcal{B}_S \otimes \mathcal{B}_{S'} \subseteq \mathcal{B}_{S \times S'}.$$

Da S, S' als separabel vorausgesetzt sind, gibt es abzählbare Basen

$$\begin{aligned}
\{U_i, i \in \mathbb{N}\} &\text{ von } S \quad \text{und} \\
\{U_i, i \in \mathbb{N}\} &\text{ von } S'.
\end{aligned}$$

$\{U_i \times U_j, i, j \in \mathbb{N}\}$ 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

$$\mathcal{B}_{S \times S'} \subseteq \mathcal{B}_S \otimes \mathcal{B}_{S'}.$$

\square

Die Abbildung g erfüllt also die Voraussetzung des Invarianzprinzips. Nun haben wir schon im Kapitel über Markovketten $\mathcal{L}(g(Y_n))$ im Spezialfall $P(X_i = \pm 1) = 1/2$ berechnet:

Satz 4.43 *Sind die $(S_i)_i$ unabhängig und $P(X_i = \pm 1) = 1/2$, so gilt für $t \in [0, 1]$*

$$\lim_{n \rightarrow \infty} P(g(Y_n) \leq t) = \frac{2}{\pi} \arcsin \sqrt{t}.$$

Dies liefert somit die Verteilungsfunktion von μg^{-1} , wenn μ das Wiener-Maß ist. Es folgt mit dem Invarianzprinzip

Satz 4.44 (*Arcussinus-Gesetz*)

Die auf (C, \mathcal{B}_C, μ) 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 4.44, so gilt für $t \in [0, 1]$:

$$\lim_{n \rightarrow \infty} P(g(Y_n) \leq t) = \frac{2}{\pi} \arcsin \sqrt{t}.$$

Bemerkung 4.45 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

$$\mathcal{L}\left(\frac{1}{n} |\{m \leq n : S_m > 0\}|\right)$$

asymptotisch nach der Arcussinus-Verteilung verteilt ist. Wir zeigen dies hier nicht.

Als nächstes wollen wir einige Eigenschaften der Brownschen Bewegung studieren. Wir beginnen mit der Feststellung, dass die Brownsche Bewegung invariant unter einigen natürlichen Operationen ist. Diese transformieren den Raum der stetigen Funktionen, sodass jede einzelnen Brownsche Bewegung ein anderes Bild bekommt, die Verteilung aber gleich bleibt.

Theorem 4.46 [*Skaleninvarianz*] Sei $\{B(t) : t > 0\}$ eine standard Brownsche Bewegung und $a > 0$. Dann ist auch der Prozess $\{X(t) : t > 0\}$ definiert durch

$$X(t) = \frac{1}{a} B(a^2 t)$$

eine (standard) Brownsche Bewegung.

Beweis: Mit den Pfaden von $B(t)$ sind auch die von $X(t)$ stetig und da $B(t)$ unabhängige Zuwächse hat, gilt dies auch für $X(t)$. Nun ist aber

$$X(t) - X(s) = \frac{1}{a} (B(a^2 t) - B(a^2 s))$$

und nach Voraussetzung und den Rechenregeln für Varianzen ist die rechte Seite normalverteilt mit Erwartungswert 0 und Varianz $(1/a^2)(a^2 t - a^2 s) = t - s$. \square

Bemerkung 4.47 Die Skaleninvarianz hat viele nützliche Folgen. Sei, zum Beispiel, $a < 0 < b$ und

$$T(a, b) = \inf\{t > 0 : B(t) = a \text{ or } B(t) = b\}$$

die Zeit, zu der die Brownsche Bewegung erstmal das Intervall $[a, b]$ verlässt. Dann berechnen wir mit

$$X(t) = \frac{1}{a}B(a^2t),$$

dass

$$\mathbb{E}T(a, b) = a^2\mathbb{E}\inf\{t > 0 : X(t) = 1 \text{ oder } X(t) = b/a\} = a^2\mathbb{E}T(1, b/a),$$

woraus sich ergibt, dass $\mathbb{E}T(-b, b)$ ein Vielfaches von b^2 ist. Darüber hinaus gilt, dass die Wahrscheinlichkeit

$$\mathbb{P}(B(t) \text{ verlässt das Intervall } [a, b] \text{ in } a) = \mathbb{P}(X(t) \text{ verlässt das Intervall } [1, b/a] \text{ in } 1)$$

nur vom Verhältnis b/a abhängt.

Eine sehr weitreichende und tiefe Erweiterung der Skaleninvarianz, die konforme Invarianz, hat 2006 zur ersten Fields-Medaille für die Stochastik geführt.

Eine weitere nützliche Eigenschaft der Brownschen Bewegung ist ihre Invarianz unter Zeitumkehr.

Theorem 4.48 [Zeitumkehr] Sei $\{B(t) : t > 0\}$ eine standard Brownsche Bewegung. Dann ist der Prozess $\{X(t) : t > 0\}$ definiert vermöge

$$X(t) = \begin{cases} 0 & \text{für } t = 0, \\ tB(1/t) & \text{für } t > 0 \end{cases}$$

auch eine standard Brownsche Bewegung.

Beweis: Der Schlüssel ist die Beobachtung, dass die endlich-dimensionalen Verteilungen der Brownschen Bewegung $(B(t_1), \dots, B(t_n))$ Gauss'sche Zufallsvektoren sind und daher durch ihren Erwartungswert $\mathbb{E}[B(t_i)] = 0$ und ihre Kovarianzen

$$\text{Cov}(B(t_i), B(t_j)) = t_i \quad \text{für } 0 < t_i < t_j.$$

eindeutig charakterisiert sind. Unter Ausnutzung der Tatsache, dass $(B(t))$ eine Brownsche Bewegung ist, sieht man sofort, dass auch $\{X(t) : t > 0\}$ ein Gauss'scher Prozess ist und dass die Gauss'sche Zufallsvektoren $(X(t_1), \dots, X(t_n))$ Erwartungswert 0 haben. Die Kovarianzen berechnen sich für $t > 0$ und $h > 0$ zu

$$\text{Cov}(X(t+h), X(t)) = (t+h)t\text{Cov}(B(1/(t+h)), B(1/t)) = t(t+h)\frac{1}{t+h} = t.$$

Also sind die endlich-dimensionalen Verteilungen der Hence the law of all the finite-dimensional distributions

$$(X(t_1), X(t_2), \dots, X(t_n)) \quad \text{für } 0 \leq t_1 \leq \dots \leq t_n$$

dieselben wie die für die Brownsche Bewegung. Die Pfade der $t \mapsto X(t)$ sind offensichtlich stetig für alle $t > 0$ Für $t = 0$ brauchen wir ein kleines Argument, dass auf

zwei Tatsachen beruht: Erstens, dass die Menge der rationalen Zahlen \mathbb{Q} abzählbar ist und die Verteilung der $\{X(t) : t > 0, t \in \mathbb{Q}\}$ dieselbe ist wie die der Brownschen Bewegung. Daher folgt

$$\lim_{\substack{t \downarrow 0 \\ t \in \mathbb{Q}}} X(t) = 0 \quad \text{fast sicher.}$$

Andererseits ist $\mathbb{Q} \cap (0, 1)$ dicht in $(0, 1)$ und $\{X(t) : t > 0\}$ ist fast sicher stetig auf $(0, 1)$, sodass

$$0 = l \lim_{\substack{t \downarrow 0 \\ t \in \mathbb{Q}}} X(t) = \lim_{t \downarrow 0} X(t) \quad \text{fast sicher.}$$

Also hat $\{X(t) : t \geq 0\}$ fast sicher stetige Pfade und ist daher eine Brownsche Bewegung. \square

Dieser Satz hat eine interessante Konsequenz

Korollar 4.49 *[Gesetz der großen Zahlen] Fast sicher gilt*

$$\lim_{t \rightarrow \infty} \frac{B_t}{t} = 0.$$

Beweis: Sei $(X(t), t \geq 0)$ definiert wie in Satz 4.48. Benutzt man diesen Satz, so sieht man, dass

$$\lim_{t \rightarrow \infty} \frac{B_t}{t} = \lim_{t \rightarrow \infty} X(1/t) = 0$$

fast sicher gilt. \square

Übung 4.50 *Man beweise diese Aussage von Korollar 4.49 anders.*

Wir wollen uns nun der Frage zuwenden, wie glatt die Pfade der Brownschen Bewegung sind. Schon aus der Definition der Brownschen Bewegung wissen wir, dass diese fast sicher stetig ist. Wir werden nun zunächst zeigen, dass dies aber nicht bedeutet, dass sie „sehr glatt“ wäre. In der Tat werden wir das folgende zeigen:

Theorem 4.51 *Die Pfade der Brownschen Bewegung*

$$t \mapsto B_t$$

haben fast sicher unendliche Variation über jedem Zeitintervall.

Hierbei ist die Variation einer Funktion f auf einem Intervall $[a, b]$ wie folgt definiert:

$$|f| := |f|_{[a,b]} := \sup \left\{ \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| : n \in \mathbb{N}, a \leq t_0^{(n)} \leq t_1^{(n)} \leq \dots \leq t_n^{(n)} \leq b \right\}$$

Der erste Schritt in diese Richtung ist das folgende grundlegende Lemma.

Lemma 4.52 Sei $0 \leq t_0^{(n)} < t_1^{(n)} < \dots < t_n^{(n)} = T$ eine Familie von Partitionen des Intervalls $[0, T]$, die beliebig fein wird, sodass für $n \rightarrow \infty$ gilt

$$\lim_{n \rightarrow \infty} \max_{0 \leq j < n} |t_{j+1}^{(n)} - t_j^{(n)}| = 0.$$

Dann gilt

$$L^2 - \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} (B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}})^2 = T.$$

Beweis: Wir schreiben $\Delta B_j = B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}$ und $\Delta t_j = t_{j+1}^{(n)} - t_j^{(n)}$ und schließlich $\delta_n = \max_j \Delta t_j$. Dann folgt

$$\begin{aligned} \left\| \sum_j (\Delta B_j)^2 - T \right\|^2 &= \mathbb{E} \left(\left[\sum_j (\Delta B_j)^2 - T \right]^2 \right) \\ &= \mathbb{E} \left(\sum_{i,j} (\Delta B_i)^2 (\Delta B_j)^2 \right) - 2T \mathbb{E} \left(\sum_j (\Delta B_j)^2 \right) + T^2 \\ &= \sum_j \mathbb{E}((\Delta B_j)^4) + \sum_{i \neq j} \mathbb{E}((\Delta B_i)^2) \mathbb{E}((\Delta B_j)^2) - 2T \left(\sum_j \Delta t_j \right) + T^2 \\ &= \sum_j 3(\Delta t_j)^2 + \sum_{i \neq j} (\Delta t_i)(\Delta t_j) - T^2 \\ &= 2 \sum_j (\Delta t_j)^2 \\ &\leq 2\delta_n \sum_j \Delta t_j \\ &= 2\delta_n T, \end{aligned}$$

wobei wir die Tatsache ausgenutzt haben, dass das vierte Moment einer zentrierten Gauss-Variable ξ is durch

$$\mathbb{E}(\xi^4) = 3 \operatorname{Var}(\xi)^2$$

gegeben ist. Nimmt man den Limes $n \rightarrow \infty$, so erhält man die Behauptung. \square

Manchmal schreiben wir das Ergebnis von Lemma 4.52 symbolisch als

$$(dB_t)^2 = dt,$$

und sagen: „Die Brownsche Bewegung hat eine quadratische Variation, die linear mit der Zeit wächst.“ In diesem Sinne hat B_t relativ große Fluktuationen auf einer kleinen Skale, nämlich

$$dB_t \text{ ist von der Ordnung } \sqrt{dt} \gg dt.$$

Tatsächlich ist diese Beobachtung dafür verantwortlich, dass man bei der Integration bezüglich der Brownschen Bewegung als Integrator nicht die üblichen Rechenregeln verwenden kann und auf so etwas wie eine Itô-Formel stößt. Genauer lernt man in einem Kurs über stochastische Analysis. Um Theorem 4.51 zu beweisen, brauchen wir einen weiteren vorbereitenden Schritt, der für allgemeine Zufallsvariablen gilt.

Lemma 4.53 Sei X_1, X_2, X_3, \dots eine Folge reellwertiger Zufallsvariablen, sodass

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|^p) = 0$$

gilt. Dann gibt es eine Teilfolge $(X_{n_k})_{k=1}^\infty$, die fast sicher gegen 0 konvergiert.

Beweis: Der Beweis ist eine nette Anwendung des Borel-Cantelli-Lemmas. Wähle eine Teilfolge, so dass

$$\sum_{k=1}^{\infty} \mathbb{E}(|X_{n_k}|^p) < \infty.$$

Mithilfe der Chebyshev-Ungleichung folgern wir, dass für alle $m \in \mathbb{N}$

$$\mathbb{P}\left[|X_{n_k}| \geq \frac{1}{m}\right] \leq m^p \mathbb{E}(|X_{n_k}|^p)$$

gilt. Also erhalten wir für alle $m \in \mathbb{N}$

$$\sum_k \mathbb{P}\left[|X_{n_k}| \geq \frac{1}{m}\right] < \infty$$

gilt. Also impliziert das Borel-Cantelli-Lemma für jedes m ,

$$\mathbb{P}\left[|X_{n_k}| \geq \frac{1}{m} \text{ for finitely many } k\right] = 1.$$

Aufgrund der σ -Additivität hat auch der Schnitt der Ereignisse über m die Wahrscheinlichkeit 1:

$$\mathbb{P}\left[\bigcap_{m \in \mathbb{N}} \bigcup_{K \in \mathbb{N}} \bigcap_{k \geq K} |X_{n_k}| < \frac{1}{m}\right] = 1,$$

d.h., $X_{n_k} \rightarrow 0$ fast sicher, wenn $k \rightarrow \infty$. □

Wir sind nun in der Lage, zu zeigen, dass die Brownsche Bewegung von unendlicher Variation ist. **Beweis:** Nach Lemma 4.52 und Lemma 4.53 gibt es eine wachsende Folge (n_k) dergestalt, dass für fast alle $\omega \in \Omega$,

$$\lim_{k \rightarrow \infty} \sum_{j=0}^{n_k-1} \left(B_{t_{j+1}^{(n_k)}}(\omega) - B_{t_j^{(n_k)}}(\omega)\right)^2 = T.$$

Wir wählen ein $\omega \in \Omega$ für das dies wahr ist. Sei

$$\varepsilon_{n_k} := \max_j |\Delta B_j| := \max_{0 \leq j < n_k} \left| B_{t_{j+1}^{(n_k)}}(\omega) - B_{t_j^{(n_k)}}(\omega) \right|.$$

Dann folgt aus der gleichmäßigen Stetigkeit von $t \mapsto B_t$, dass $\lim_{k \rightarrow \infty} \varepsilon_{n_k} = 0$. Damit ergibt sich, dass

$$\sum_{j=0}^{n_k-1} |\Delta B_j| \geq \sum_{j=0}^{n_k-1} \frac{1}{\varepsilon_{n_k}} |\Delta B_j|^2 \sim \frac{1}{\varepsilon_{n_k}} T \rightarrow \infty \quad \text{as } k \rightarrow \infty.$$

□

Um auch die Nicht-Differenzierbarkeit der Brownschen Bewegung herzuleiten, erinnern wir uns an Satz 4.46. Dies impliziert natürlich (mit $c = 1/a$), dass für $c > 0$ mit $(B_t)_{0 \leq t \leq T}$ auch $(c \cdot B_{t/c^2})_{0 \leq t \leq T}$ eine Brownsche Bewegung (auf $[0, T/c^2]$) ist.

Nun beweisen wir

Theorem 4.54 *Die Pfade der Brownschen Bewegung*

$$t \mapsto B_t$$

sind fast sicher nirgends differenzierbar.

Beweis: Sei $X_{n,k} := \max_{j=k, k+1, k+2} |B_{\frac{j}{2^n}} - B_{\frac{j-1}{2^n}}|$. Für $\varepsilon > 0$ gilt

$$\mathbb{P}(X_{n,k} \leq \varepsilon) = \mathbb{P}(|B_{\frac{1}{2^n}}| \leq \varepsilon)^3 = P(|B_1| \leq 2^{n/2}\varepsilon)^3 \leq (2^{n/2+1}\varepsilon)^3$$

wobei der zweite Schritt aus der obigen Skaleninvarianz der Brownschen Bewegung folgt. Also erhalten wir für $Y_n := \min_{k \leq T \cdot 2^n} X_{n,k}$

$$\mathbb{P}(Y_n \leq \varepsilon) \leq T 2^n (2^{n/2+1}\varepsilon)^3. \quad (54)$$

Bezeichne mit

$$A := \{\omega \in \Omega : t \mapsto B_t(\omega) \text{ is differentiable somewhere}\}.$$

Sei $\omega \in A$ und $t \mapsto B_t(\omega)$ differenzierbar in $t_0 := t_0(\omega)$, and bezeichne D die diese Ableitung. Dann gibt es ein $\delta := \delta(\omega, t_0)$ sodass

$$|t - t_0| \leq \delta \Rightarrow |B_t - B_{t_0}| \leq (|D| + 1)|t - s|.$$

Nun wählen wir $n_0 := n_0(\omega, t_0)$ hinreichend groß, dass

$$\frac{1}{2^{n_0}} < \frac{\delta}{2}, \quad n_0 > (|D| + 1) \text{ and } n_0 > t_0.$$

Für $n \geq n_0$ wähle k sodass $\frac{k}{2^n} \leq t_0 < \frac{k+1}{2^n}$. Dann gilt

$$|t_0 - \frac{j}{2^n}| < \delta \quad \text{for } j = k, k+1, k+2.$$

Also

$$X_{n,k}(\omega) \leq (|D| + 1) \frac{1}{2^n} \leq \frac{n}{2^n},$$

und, da $n > t_0 > k/2^n$, auch $Y_n(\omega) \leq \frac{n}{2^n}$. Damit folgt

$$A \subset A_n := \{Y_n(\omega) \leq \frac{n}{2^n}\}$$

für alle hinreichend großen n . Somit folgt auch

$$A \subseteq \liminf A_n.$$

Aber aus (54) folgt

$$\sum_n \mathbb{P}(A_n) \leq \sum_n n 2^n (2^{n/2+1} n 2^{-n})^3 < \infty$$

wenn $n \rightarrow \infty$, sodass $\mathbb{P}(\liminf A_n) = 0$. Somit ist $t \rightarrow B_t(\omega)$ fast sicher nirgends differenzierbar. \square

Die Brownsche Bewegung ist also nicht differenzierbar. Eine natürliche Frage ist, ob es irgendetwas zwischen Differenzierbarkeit und Stetigkeit gibt, also, ob man vielleicht sagen kann "wie sehr" die Brownsche Bewegung nicht differenzierbar ist. Wir versuchen das im folgenden zu quantifizieren.

Zunächst einmal sei bemerkt, dass die Realisierungen der Brownschen Bewegung als stetige Funktionen auf dem kompakten Intervall $[0, 1]$ (oder jedem anderen kompakten Intervall) gleichmäßig stetig ist. Also gibt es eine *zufällige* Funktion φ mit $\lim_{h \downarrow 0} \varphi(h) = 0$, die **Stetigkeitsmodul** der Funktion $B : [0, 1] \rightarrow \mathbb{R}$ heißt (omega-weise) und für die gilt

$$\limsup_{h \downarrow 0} \sup_{0 \leq t \leq 1-h} \frac{|B(t+h) - B(t)|}{\varphi(h)} \leq 1 \quad (55)$$

Natürlich wird (55) für jedes ω gleich von einer ganzen Funktionenschar erfüllt. Trotzdem bleibt die Frage, ob wir die gleiche Schranke auch für eine nicht-zufällige Funktion erhalten können. Dies würde uns erlauben, zu sagen, „wie stetig“ fast alle Brownschen Bewegungen sind. Der folgende Satz zeigt, dass das möglich ist.

Theorem 4.55 *Fast sicher gilt*

$$\limsup_{h \downarrow 0} \sup_{0 \leq t \leq 1-h} \frac{|B(t+h) - B(t)|}{\sqrt{2h \log(1/h)}} = 1 \quad (56)$$

5 Extremwertverteilungen von iid-Beobachtungen

Rekorde sind ein Phänomen, das uns im täglichen Leben interessiert. Wir sind auf der Suche nach dem höchsten Berg, dem längsten Fluss, dem heißesten Tag. Bei Sportveranstaltungen küren wir die schnellste Frau oder den Mann, der am höchsten springt, einen Speer am weitesten wirft.

Rekorde sind aber auch von großer Wichtigkeit für praktische Belange: In vorangegangenen Kapiteln und Vorlesungen über Wahrscheinlichkeitstheorie haben wir gelernt, wie sich Summen von Zufallsvariablen verhalten. Dies hilft uns, etliche praktische Problemstellungen lösen zu können. Will man aber zum Beispiel einen Deich bauen, so interessiert einen die durchschnittliche Höhe einer Flut wenig. Ebenso ist man am Extremalverhalten interessiert, wenn man die notwendige Sicherheitsreserve einer Bank oder eines Versicherungsunternehmens festlegen möchte. Auch hier hilft es einem wenig zu wissen, wie hoch die durchschnittlichen Forderungen an eine Versicherung sind – wir wollen wissen, wie groß eine solche Forderung maximal sein kann (und dies mit großer Wahrscheinlichkeit). Nun sind Wasserstände ebenso einem stochastischen Einfluss unterworfen wie die Forderungen an ein Versicherungsunternehmen. Es bleibt uns wenig anderes übrig als sie als stochastischen Prozess zu modellieren. Wir wollen daher in der Folge das Extremwertverhalten von stochastischen Prozessen untersuchen. Wir beginnen mit Folgen von iid Zufallsvariablen. Es sei also (X_1, X_2, \dots) eine Folge von iid Zufallsvariablen mit der Verteilungsfunktion F . Da wir uns für die Extrema der $(X_i)_i$ interessieren, bezeichnen wir mit

$$M_n := \max_{1 \leq i \leq n} X_i$$

das Maximum der ersten n der X_i und mit

$$m_n := \min_{1 \leq i \leq n} X_i$$

ihr Minimum. Es liegt nahe, dass wir – fragen wir nach einer Grenzwertverteilung von M_n – das M_n skalieren müssen (auch im Zentralen Grenzwertsatz benötigen wir ja z. B. eine Skalierung). Tatsächlich ist das Grenzwertverhalten ohne eine solche Skala leicht festzustellen: Mit Hilfe des Borel-Cantelli-Lemmas sieht man leicht, dass M_n \mathbb{P} -f.s. gegen den rechten Randpunkt der Verteilung, also gegen

$$x_R = \sup_x \{x : F(x) = 1\}$$

konvergiert. Wir fragen also, ob es Folgen (A_n) und (B_n) bzw. (a_n) und (b_n) gibt, sowie Grenzverteilungen G bzw. H gibt, so dass

$$A_n(M_n - B_n) \xrightarrow{\mathcal{D}} G$$

bzw.

$$a_n(m_n - b_n) \xrightarrow{\mathcal{D}} H$$

gilt.

Die Frage ist natürlich, wie wir überhaupt an die Verteilung der Extrema kommen. Dazu dient das folgende Lemma.

Lemma 5.1 Für alle $n \in \mathbb{N}$ und $x \in \mathbb{R}$ gilt

$$\begin{aligned} F^{M_n}(x) &= F^n(x) \quad \text{und} \\ F^{m_n}(x) &= 1 - (1 - F(x))^n. \end{aligned}$$

Hierbei sind F^{M_n} und F^{m_n} die Verteilungsfunktionen von M_n bzw. m_n .

Beweis: Es gilt

$$\begin{aligned} F^{M_n}(x) &= \mathbb{P}(\max_{1 \leq i \leq n} X_i \leq x) \\ &= \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \leq x\}\right) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq x) \\ &= F^n(x). \end{aligned}$$

Hierbei haben wir die Unabhängigkeit und die identische Verteilung der X_i benutzt. Analog folgt

$$\begin{aligned} F^{m_n}(x) &= \mathbb{P}(\min X_i \leq x) \\ &= 1 - \mathbb{P}\left(\bigcap_{i=1}^n \{X_i > x\}\right) \\ &= 1 - \prod_{i=1}^n \mathbb{P}(X_i > x) \\ &= 1 - (1 - F(x))^n. \end{aligned}$$

□

Die eingangs gestellte Frage lässt sich also so umformulieren, dass wir Konstanten A_n, B_n bzw. a_n, b_n und Verteilungsfunktionen G bzw. H , so dass

$$F^n\left(\frac{x}{A_n} + B_n\right) \rightarrow G(x)$$

bzw.

$$\left(1 - F\left(\frac{x}{a_n} + b_n\right)\right)^n \rightarrow 1 - H(x).$$

Dies soll jeweils für alle Stetigkeitspunkte x der Funktionen G bzw. H gelten. Man bezeichnet dies folgendermaßen:

Definition 5.2 Es seien F und G Verteilungsfunktionen.

- a) F liegt im Max-Anziehungsbereich von G , falls es Konstanten $A_n > 0, B_n$ gibt, so dass

$$F^n \left(\frac{x}{A_n} + B_n \right) \xrightarrow{n \rightarrow \infty} G(x)$$

für alle x , in denen G stetig ist (Bezeichnung für diese Menge: $C(G)$). Wir schreiben:

$$F \in \mathcal{D}(G).$$

- b) F liegt im Min-Anziehungsbereich von G , falls es Konstanten $a_n > 0$ und b_n gibt mit

$$\left(1 - F \left(\frac{x}{a_n} + b_n \right) \right)^n \rightarrow 1 - G(x)$$

für alle $x \in C(G)$.

Beispiel 5.3 a) (X_n) seien iid $\text{Exp}(\lambda)$ -verteilt. Diese haben die Verteilungsfunktion

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

Man kann sich z. B. vorstellen, dass die X_i die Zerfallsdauer von radioaktiven Isotopen modellieren. Wir wählen $A_n = \lambda$ und $B_n = \frac{\log n}{\lambda}$. Dann folgt für alle $x \in \mathbb{R}$:

$$\begin{aligned} F^{\lambda M_n - \log n}(x) &= F^{M_n} \left(\frac{x}{\lambda} + \frac{\log n}{\lambda} \right) \\ &= F^n \left(\frac{x}{\lambda} + \frac{\log n}{\lambda} \right) \\ &= (1 - (e^{-x - \log n}))^n \mathbb{1}_{(0, \infty)} \left(\frac{x}{\lambda} + \frac{\log n}{\lambda} \right) \\ &= (1 - (e^{-x - \log n}))^n \mathbb{1}_{[-\log n, \infty)}(x) \\ &= \left(1 - \frac{e^{-x}}{n} \right)^n \mathbb{1}_{[-\log n, \infty)}(x) \\ &\xrightarrow{n \rightarrow \infty} e^{-e^{-x}} =: \Lambda(x), \end{aligned}$$

da $-\log n$ für $n \rightarrow \infty$ gegen $-\infty$ konvergiert. Man prüft leicht nach, dass $\Lambda(\cdot)$ monoton ist und dass

$$\lim_{x \rightarrow -\infty} e^{-e^{-x}} = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} e^{-e^{-x}} = 1$$

gilt. Also ist $\Lambda(\cdot)$ eine Verteilungsfunktion mit $C(\Lambda) = \mathbb{R}$.

- b) Die Variablen X_1, X_2, \dots seien nun iid normalverteilt, genauer $X_1 \sim \mathcal{N}(0, 1)$. Wieder suchen wir $A_n > 0$ und B_n , so dass $F^n(\frac{x}{A_n} + B_n)$ einen nicht-entarteten Limes hat. Nun ist

$$\mathbb{P}[A_n(M_n - B_n) \leq x] = \mathbb{P} \left[M_n \leq \frac{x}{A_n} + B_n \right] = \left(\Phi \left(\frac{x}{A_n} + B_n \right) \right)^n,$$

wobei Φ die Verteilungsfunktion der $\mathcal{N}(0, 1)$ -Verteilung ist. Wenn wir

$$x_n := \frac{x}{A_n} + B_n$$

setzen, schreiben wir daher

$$\left(\Phi \left(\frac{x}{A_n} + B_n \right) \right)^n = (1 - (1 - \Phi(x_n))^n).$$

Dies konvergiert, wenn

$$(1 - \Phi(x_n)) = \frac{g(x)}{n} + \mathcal{O}(1/n)$$

und in diesem Falle ist der Limes

$$\lim_{n \rightarrow \infty} (1 - (1 - \Phi(x_n))^n) = e^{-g(x)}.$$

Somit müssen wir die Folgen A_n und B_n so finden, dass

$$\frac{1}{\sqrt{2\pi}} \int_{x_n}^{\infty} e^{-y^2/2} dy = \frac{g(x)}{n}.$$

Nun lässt sich das Tail-Verhalten dieses Integrals recht gut abschätzen: Es ist

$$\frac{1}{\sqrt{2\pi}} \int_{x_n}^{\infty} e^{-y^2/2} dy \leq \int_{x_n}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{y}{x_n} e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi} x_n} e^{-x_n^2/2}.$$

Ähnlich zeigt man

$$\frac{1}{\sqrt{2\pi}} \int_{x_n}^{\infty} e^{-y^2/2} dy \geq \frac{1}{x_n \sqrt{2\pi}} e^{-x_n^2/2} \left(1 - \frac{2}{x_n^2}\right).$$

Somit genügt es, x_n so zu wählen, dass

$$\frac{1}{x_n \sqrt{2\pi}} e^{-x_n^2/2} = \frac{g(x)}{n}.$$

Setzen wir x_n ein, so erhalten wir

$$\frac{g(x)}{n} = \frac{e^{-\frac{1}{2}(\frac{x}{A_n} + B_n)^2}}{\sqrt{2\pi}(\frac{x}{A_n} + B_n)} = \frac{e^{-B_n^2/2 - \frac{x^2}{2A_n^2} - \frac{B_n}{A_n}x}}{\sqrt{2\pi}(\frac{x}{A_n} + B_n)}. \quad (57)$$

Setzen wir $x = 0$, sehen wir, dass

$$e^{-B_n^2/2} / \sqrt{2\pi} B_n = g(0)/n$$

gilt. Wir setzen

$$B_n = \sqrt{2 \log n} + C_n$$

und bekommen dann für C_n

$$e^{-\sqrt{2 \log n} C_n - C_n^2/2} = g(0) \sqrt{2\pi} (\sqrt{2 \log n} + C_n).$$

Wir normieren g , so dass $g(0) = 1$ gilt. Es lässt sich dann zeigen, dass die führenden Terme für C_n von der Gestalt

$$C_n = -\frac{\log \log n + \log(4\pi)}{2\sqrt{2 \log n}}$$

sind. Die höheren Ordnungsterme tragen nichts mehr bei. Damit uns der Faktor von x in (57) nicht abhaut, muss A_n approximativ so groß wie B_n sein. Wir setzen daher

$$A_n = \sqrt{2 \log n}.$$

Setzt man dies in (57), so erhält man

$$g(x) = e^{-x} + o\left(\frac{1}{n}\right).$$

Mit unserer Eingangsüberlegung haben wir somit gezeigt, dass mit der Wahl

$$B_n := \sqrt{2 \log n} - \frac{\log \log n + \log(4\pi)}{2\sqrt{2 \log n}}$$

und

$$A_n := \sqrt{2 \log n}$$

für alle $x \in \mathbb{R}$ gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n(M_n - B_n) \leq x) = e^{-e^{-x}}.$$

Bemerkung 5.4 Wir werden den Sachverhalt aus Beispiel 5.3 b) später noch in leicht modifizierter Form verwenden. Mit den gleichen Konstantenfolgen A_n und B_n sei

$$u_n(x) := B_n + \frac{x}{A_n}.$$

Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}[M_n \leq u_n(x)] = e^{-e^{-x}}.$$

Wir sehen hier, dass die Extreme einer Folge von Zufallsvariablen zwar wachsen, aber relativ langsam, nämlich wie $\sqrt{\log n}$. Für Gaußsche Zufallsvariablen lässt sich mit relativ guter Präzision behaupten, dass

$$M_n \sim \sqrt{2 \log n}$$

gilt. Darüber hinaus haben wir gesehen, dass für zwei verschiedene Verteilungen, die Exponential- und die Gaußverteilung, die Extreme gegen die gleiche Limesverteilung konvergieren. Die Verteilung mit der Verteilungsfunktion $\Lambda(x) = e^{-e^{-x}}$ heißt auch die Gumbelverteilung. Es erhebt sich die Frage, ob dies die einzige Limesverteilung unter geeigneten Voraussetzungen ist (ähnlich wie die Normalverteilung im Zentralen Grenzwertsatz) oder ob noch andere Verteilungen auftreten können. Hierbei muss man die gleichen Unterscheidungen wie beim Zentralen Grenzwertsatz treffen. Sind

die X_i iid mit $\mathbb{E}(X_1) = \mu$ und $0 \neq \mathbb{E}((X_1 - \mu)^2) < +\infty$, so besagt der Zentrale Grenzwertsatz, dass

$$\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n \mathbb{V} X_1}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

aber auch

$$\frac{\sum_{i=1}^n (X_i - \mu) + 2\mu}{\sqrt{2n \mathbb{V} X_1}} \xrightarrow{\mathcal{D}} \mathcal{N}(2\mu, \frac{1}{2}).$$

Natürlich würde niemand sagen, dass dies verschiedene Limiten für die Summe der X_i sind. Eine entsprechende Untersuchung trifft die folgende Definition.

Definition 5.5 *Zwei Verteilungsfunktionen G und \tilde{G} heißen vom selben Typ, falls Konstanten $c > 0$ und $d \in \mathbb{R}$ existieren mit*

$$\tilde{G}(x) = G(cx + d) \quad \forall x \in \mathbb{R}.$$

Ist also $F \in \mathcal{D}(G)$, dann gilt auch $F \in \mathcal{D}(\tilde{G})$.

Darüber hinaus bemerken wir noch, dass wegen

$$M_n^- := \max_{1 \leq i \leq n} (-X_i) = - \min_{1 \leq i \leq n} X_i$$

und

$$m_n^- := \min_{1 \leq i \leq n} (-X_i) = - \max_{1 \leq i \leq n} X_i$$

folgendes gilt:

- Gilt für $A_n > 0$, $B_n \in \mathbb{R}$ und stetiges G , dass

$$\mathbb{P}^{A_n(M_n - B_n)} \xrightarrow{\mathcal{D}} G,$$

so folgt

$$\mathbb{P}^{A_n(m_n^- + B_n)} \xrightarrow{\mathcal{D}} H$$

mit $H(x) := 1 - G(-x) \quad \forall x \in \mathbb{R}$.

- Gilt für $a_n > 0$, $b_n \in \mathbb{R}$ und stetiges H , dass

$$\mathbb{P}^{a_n(m_n - b_n)} \xrightarrow{\mathcal{D}} H,$$

so folgt

$$\mathbb{P}^{a_n(M_n^- + b_n)} \xrightarrow{\mathcal{D}} G,$$

wobei $G(x) := 1 - H(-x) \quad \forall x \in \mathbb{R}$. Will man dies begründen, so berechnet man

$$\begin{aligned} \mathbb{P}(A_n(m_n^- + B_n) \leq x) &= \mathbb{P}(-A_n(M_n - B_n) \leq x) \\ &= 1 - \mathbb{P}(A_n(M_n - B_n) < -x) \xrightarrow{n \rightarrow \infty} 1 - G(-x). \end{aligned}$$

Die zweite Tatsache beweist man analog. Wir können uns also auf das Studium der Maxima eines Prozesses konzentrieren (oder, wenn dies praktischer sein sollte, auf die Minima).

Wir haben in Beispiel 5.3 schon die Gumbel-Verteilung $\Lambda(x) = e^{-e^{-x}}$ als mögliche Limesverteilung kennengelernt. Wir sollen in der Folge sehen, welche Verteilungen noch mögliche Limes-Verteilungen sind. Denkt man an den CLT zurück, so ist ein wesentlicher Grund für das Auftreten der Normalverteilung dort, dass die Normalverteilung stabil ist unter Faltung, d. h. sind die $X_i \sim \mathcal{N}(0, 1)$ und iid, so ist

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Wir wollen sehen, ob Ähnliches für die Gumbelverteilung gilt:

Beispiel 5.6 *Es gelte $F^{X_i} = \Lambda(x) = e^{-e^{-x}}$. Dann gilt für $M_n - \log n$, d. h. für $A_n = 1$ und $B_n = \log n$*

$$\begin{aligned} F^{M_n - \log n}(x) &= \Lambda^n(x + \log n) \\ &= (e^{-e^{-x - \log n}})^n = (e^{-\frac{e^{-x}}{n}})^n \\ &= e^{-e^{-x}} = \Lambda(x) \quad \forall x \in \mathbb{R}, \forall n \in \mathbb{N}. \end{aligned}$$

Λ erhält man also auch als Verteilung des Maximums von richtig skalierten Λ -verteilten iid Zufallsvariablen.

Wir nehmen dieses Beispiel zum Anlass für folgende Definition:

Definition 5.7 *Eine nicht-entartete Verteilungsfunktion G heißt genau dann max-stabil, wenn für geeignete Konstanten $a_n > 0$ und $b_n \in \mathbb{R}$ gilt*

$$G^n(a_n^{-1}x + b_n) = G(x) \quad \forall x \in \mathbb{R} \quad \forall n \in \mathbb{N}.$$

Eine nicht-entartete Verteilung heißt max-stabil, wenn die zugehörige Verteilungsfunktion max-stabil ist.

Definition 5.8 • *Die Verteilung mit der Verteilungsfunktion*

$$\Phi_\alpha(x) = e^{-x^{-\alpha}} \mathbb{1}_{(0, \infty)}(x), \quad \alpha > 0,$$

heißt Fréchet-Verteilung.

• *Die Verteilung mit der Verteilungsfunktion*

$$\psi_\alpha(x) = e^{-(-x)^\alpha} \mathbb{1}_{(-\infty, 0)}(x) + \mathbb{1}_{[0, \infty)}(x)$$

(für $\alpha > 0$) heißt Weibull-Verteilung.

Beispiel 5.9 (i) Es sei $F^{X_n} = \Phi_\alpha$. Für $A_n = n^{-1/\alpha}$ und $B_n = 0$ gilt dann

$$\begin{aligned} F^{n^{-1/\alpha} M_n}(x) &= F^{M_n}(x n^{1/\alpha}) = \Phi_\alpha^n(x n^{1/\alpha}) \\ &= \left(e^{-(x n^{1/\alpha})^{-\alpha}} \right)^n \mathbb{1}_{(0,\infty)}(x n^{1/\alpha}) \\ &= \left(e^{\frac{-x^{-\alpha}}{n}} \right)^n \mathbb{1}_{(0,\infty)}(x) \\ &= e^{-x^{-\alpha}} \mathbb{1}_{(0,\infty)}(x) \\ &= \Phi_\alpha(x) \quad \forall x \in \mathbb{R}, \forall n \in \mathbb{N}. \end{aligned}$$

Also ist die Fréchet-Verteilung max-stabil.

(ii) Es sei $F^{X_n} = \psi_\alpha$, $\alpha > 0$. Sei $A_n = n^{1/\alpha}$ und $B_n = 0$. Dann gilt:

$$\begin{aligned} F^{M_n n^{1/\alpha}}(x) &= F^{M_n}(x n^{-1/\alpha}) = \psi_\alpha^n(x n^{-1/\alpha}) \\ &= \left(e^{-(-x n^{-1/\alpha})^\alpha} \right)^n \mathbb{1}_{(-\infty,0)}(x n^{-1/\alpha}) + \mathbb{1}_{[0,\infty)}(x n^{-1/\alpha}) \\ &= \left(e^{-((-x)^\alpha n^{-1})} \right)^n \mathbb{1}_{(-\infty,0)}(x) + \mathbb{1}_{[0,\infty)}(x) \\ &= e^{-(-x)^\alpha} \mathbb{1}_{(-\infty,0)}(x) + \mathbb{1}_{[0,\infty)}(x) \\ &= \psi_\alpha(x) \quad \forall x \in \mathbb{R}, n \in \mathbb{N}. \end{aligned}$$

Somit ist auch die Weibull-Verteilung max-stabil.

Wir wollen nun die möglichen Limesverteilungen für Maxima von iid Zufallsvariablen charakterisieren. Ein erster Schritt in diese Richtung ist die Beschreibung der max-stabilen Verteilungen. Wir beginnen mit dem folgenden Satz.

Satz 5.10 Eine nicht-entartete Verteilungsfunktion G ist genau dann max-stabil, wenn es Folgen (F_n) von Verteilungsfunktionen und $(A_n), (B_n)$ von Konstanten mit $A_n > 0$ gibt, so dass

$$F_n(A_{nk}^{-1}x + B_{nk}) \xrightarrow{n \rightarrow \infty} G^{1/k}(x)$$

für alle $x \in C(G)$ und alle $k \in \mathbb{N}$ gilt.

Für den Beweis dieses Satzes benötigen wir den Satz von Khintchin, der wiederum einiger Vorbereitung bedarf.

Definition 5.11 Sei φ eine reellwertige, monoton nicht-fallende, rechtsseitig stetige Funktion. Ferner sei

$$\begin{aligned} \inf(\varphi) &:= \inf\{\varphi(x), x \in \text{dom}(\varphi)\} \quad \text{und} \\ \sup(\varphi) &:= \sup\{\varphi(x), x \in \text{dom}(\varphi)\}. \end{aligned}$$

Dann kann die Pseudo-Inverse φ^- von φ durch

$$\begin{aligned} \varphi^- &: (\inf(\varphi), \sup(\varphi)) \rightarrow \overline{\text{dom}(\varphi)} \\ \varphi^-(y) &:= \inf\{x \in \text{dom}(\varphi) : \varphi(x) \geq y\} \end{aligned}$$

definiert werden. Dabei ist $\text{dom}(\varphi)$ der Definitionsbereich von φ und $\overline{\text{dom}(\varphi)}$ sein Abschluss.

Wir sammeln ein paar wichtige Eigenschaften der Pseudo-Inversen.

Lemma 5.12 φ^- sei die Pseudo-Inverse einer rechtsseitig stetigen, monoton nicht-fallenden Funktion. Dann gilt:

- (i) φ^- ist monoton nicht-fallend und linksseitig stetig.
- (ii) Ist φ in $\varphi^-(y)$ stetig, so gilt $\varphi(\varphi^-(y)) = y$.
- (iii) Ist φ^- in $\varphi(x) \in \text{dom}(\varphi^-)$ stetig, so gilt $\varphi^-(\varphi(x)) = x$.
- (iv) Es gilt
 - $\varphi(\varphi^-(y)) \geq y \quad \forall y \in \text{dom}(\varphi^-)$.
 - $\varphi^-(\varphi(x)) \leq x \quad \forall x : \varphi(x) \in \text{dom}(\varphi^-)$.

Beweis:

- (iv) Die zweite Aussage folgt direkt aus der Definition von φ^- . Die erste Aussage folgt mit Hilfe der rechtsseitigen Stetigkeit von φ aus der Definition von φ^- .
- (i) Die Monotonie von φ^- folgt direkt aus der Monotonie von φ . Es bleibt die linksseitige Stetigkeit von φ^- . Dazu sei $(y_n)_n$ monoton nicht-fallend aus $\text{dom}(\varphi^-)$ und es gelte

$$\lim_{n \rightarrow \infty} y_n = y \in \text{dom}(\varphi^-).$$

Dann konvergiert auch die Folge $(\varphi^-(y_n))_n$, da sie monoton nicht-fallend und durch $\varphi^-(y)$ beschränkt ist. Der Grenzwert heie z . Zu zeigen ist, dass

$$z := \lim_{n \rightarrow \infty} \varphi^-(y_n) = \varphi^-(y)$$

gilt. Zunchst ist

$$z \leq \varphi^-(y), \quad \text{da} \quad \varphi^-(y_n) \leq \varphi^-(y)$$

fr alle $n \in \mathbb{N}$ gilt. Angenommen, dass $z < \varphi^-(y)$ gilt. Dann glte nach Definition bzw. mit Hilfe von (iv)

$$\varphi(z) < y \quad \text{und} \quad \varphi(\varphi^-(y_n)) \geq y_n \quad \forall n \in \mathbb{N}.$$

Daraus wrde dann

$$\liminf_{n \rightarrow \infty} \varphi(\varphi^-(y_n)) \geq y > \varphi(z)$$

folgen, im Widerspruch zur Monotonie von φ , da $(\varphi^-(y_n))$ monoton nicht-fallend ist. Also gilt $z = \varphi^-(y)$ und φ^- ist linksseitig stetig.

- (ii) Sei $y \in \text{dom}(\varphi^-)$ und $\varphi^-(y)$ ein Stetigkeitspunkt von φ . $(X_n)_n$ sei monoton nicht-fallend mit

$$X_n < \varphi^-(y) \quad \forall n \in \mathbb{N} \quad \text{und} \quad \lim_{n \rightarrow \infty} x_n = \varphi^-(y).$$

Dann gilt nach Definition $\varphi(x_n) < y \quad \forall n \in \mathbb{N}$ und es folgt mit (iv)

$$\lim_{n \rightarrow \infty} \varphi(x_n) \leq y \leq \varphi(\varphi^-(y)) = \varphi(\lim_{n \rightarrow \infty} x_n) = \lim_{n \rightarrow \infty} \varphi(x_n),$$

wobei die letzte Gleichheit gilt, da $\varphi^-(y)$ ein Stetigkeitspunkt von φ ist. Damit folgt

$$\varphi(\varphi^-(y)) = y.$$

- (iii) Dies geht analog zu (ii). $\varphi(x)$ sei ein Stetigkeitspunkt von φ^- und (y_n) monoton nicht-wachsend mit

$$y_n > \varphi(x) \quad \forall n \in \mathbb{N} \quad \text{und} \quad \lim_{n \rightarrow \infty} y_n = \varphi(x).$$

Dann gilt nach Definition $\varphi^-(y_n) > x$ für alle $n \in \mathbb{N}$, und es folgt

$$\lim_{n \rightarrow \infty} \varphi^-(y_n) \geq x \geq \varphi^-(\varphi(x)) = \varphi^-(\lim_{n \rightarrow \infty} y_n) = \lim_{n \rightarrow \infty} \varphi^-(y_n),$$

wobei die letzte Gleichung wieder wegen der Stetigkeit von φ^- in $\varphi(x)$ gilt. Also folgt

$$\varphi^-(\varphi(x)) = x.$$

□

Seinen Zweck erfüllt dieses Lemma durch Anwendung auf die Verteilungsfunktion von Zufallsgrößen. Das folgende Lemma haben wir schon beim Kolmogorov-Smirnov-Test kennengelernt.

Lemma 5.13 (i) Sei X eine Zufallsvariable mit stetiger Verteilungsfunktion F . Dann gilt

$$F(X) \sim \mathcal{R}(0, 1)$$

(dabei ist $\mathcal{R}(a, b)$ die Rechteck-Verteilung auf dem Intervall $[a, b]$).

(ii) Sei F eine Verteilungsfunktion und $U \sim \mathcal{R}(0, 1)$. Dann gilt

$$F^-(U) \sim F.$$

Beweis:

- (i) Da F stetig ist, ist $F(X)$ messbar und daher eine Zufallsvariable. Außerdem ist F^- auf $(0,1)$ definiert, da F eine Verteilungsfunktion ist. Somit sind die Mengen

$$A_x := \{y \in \mathbb{R} : F(y) \leq x\} \quad \text{und} \\ B_x := \{y \in \mathbb{R} : y \leq F^-(x)\}$$

für $0 < x < 1$ wohldefiniert. Zunächst zeigen wir

$$A_x = B_x \quad \forall x \in C(F^-). \quad (58)$$

“ \subseteq ”: Es sei $x \in C(F^-)$ und $y \in A_x$.

1. Fall: $F(y) = x$. Dann gilt mit Lemma 5.12 (iii)

$$F^-(x) = F^-(F(y)) = y,$$

also $y \in B_x$.

2. Fall: $F(y) < x$. Angenommen, es gälte $y \geq F^-(x)$. Dann folgt mit der schwachen Monotonie von F und Lemma 5.12 (iv)

$$F(y) \geq F(F^-(x)) \geq x$$

im Widerspruch zu $F(y) < x$. Also gilt

$$y < F^-(x), \quad \text{also} \quad y \in B_x.$$

“ \supseteq ”: Es sei $0 < x < 1$ und $y \in B_x$, d. h. $y \leq F^-(x)$. Mit der Monotonie und der Stetigkeit von F folgt daraus

$$F(y) \leq F(F^-(x)) = x,$$

also $y \in A_x$.

Somit ist (58) gezeigt. Da F^- monoton nicht-fallend ist, ist die Menge der Unstetigkeitsstellen von F^- und somit auch die Menge

$$M_{\neq} := \{x \in (0,1) : A_x \neq B_x\}$$

höchstens abzählbar. Für alle $x \in (0,1) \setminus M_{\neq}$ gilt aber:

$$\begin{aligned} \{\omega \in \Omega : F(X(\omega)) \leq x\} &= \{\omega \in \Omega : X(\omega) \in A_x\} \\ &= \{\omega \in \Omega : X(\omega) \in B_x\} \\ &= \{\omega \in \Omega : X(\omega) \leq F^-(x)\}. \end{aligned}$$

D. h. für alle $x \in (0,1) \setminus M_{\neq}$ gilt

$$\mathbb{P}(F(X) \leq x) = \mathbb{P}(X \leq F^-(x)) = F(F^-(x)) = x,$$

wobei die letzte Gleichheit aus der Stetigkeit von F folgt. Mit der Stetigkeit von F folgt die Aussage aber schon für alle $x \in \mathbb{R}$, da die höchstens abzählbar vielen Unstetigkeitsstellen von F^- isoliert liegen müssen. Also gilt $F(X) \sim \mathcal{R}(0,1)$.

- (ii) Mit Lemma 5.12 (iv) und der Monotonie von F und F^- gilt für alle $0 < F(x) < 1$

$$\{\omega \in \Omega : F^-(U(\omega)) \leq x\} \subseteq \{\omega \in \Omega : U(\omega) \leq F(F^-(U(\omega))) \leq F(x)\}$$

und

$$\{\omega \in \Omega : U(\omega) \leq F(x)\} \subseteq \{\omega \in \Omega : F^-(U(\omega)) \leq F^-(F(x)) \leq x\}.$$

Daraus folgt

$$\mathbb{P}(F^-(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$$

für alle $0 < F(x) < 1$, wobei die letzte Gleichheit gilt, da U $\mathcal{R}(0,1)$ -verteilt ist. Also besitzt $F^-(U)$ die Verteilungsfunktion F .

□

Des weiteren wird es wichtig sein, den Zusammenhang zwischen Verteilungsfunktionen und Zufallsvariablen herzustellen. Dies leistet der Satz von Skohorod.

Satz 5.14 (*Übertragungssatz von Skohorod*)

Es sei $(F_n)_n$ eine Folge von Verteilungsfunktionen und G eine Verteilungsfunktion, so dass

$$F_n(x) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in C(G).$$

Dann existieren Zufallsvariablen $(X_n)_n$ und Y , so dass X_n die Verteilungsfunktion F_n hat für jedes n und Y die Verteilungsfunktion G besitzt und

$$X_n \xrightarrow{n \rightarrow \infty} Y \quad \mathbb{P}\text{-f.s.}$$

gilt.

Beweis: Es sei U $\mathcal{R}(0,1)$ -verteilt. Es wird nun gezeigt, dass die durch $X_n = F_n^-(U)$ und $Y = G^-(U)$ definierten Zufallsvariablen den gewünschten Bedingungen genügen. Aus Lemma 5.13 folgt, dass F_n für jedes n die Verteilungsfunktion von X_n ist und Y die Verteilungsfunktion G besitzt. Es bleibt die behauptete fast-sichere Konvergenz zu zeigen, also

$$\mathbb{P}(\lim X_n = Y) = 1.$$

Nun ist aber

$$\begin{aligned} \mathbb{P}(\lim_{n \rightarrow \infty} X_n = Y) &= \mathbb{P}(\lim_{n \rightarrow \infty} F_n^-(U) = G^-(U)) \\ &= \mathbb{P}^U(\lim_{n \rightarrow \infty} F_n^-(x) = G^-(x)). \end{aligned}$$

Da U auf $(0,1)$ gleichverteilt ist und somit vom Lebesguemaß dominiert wird und jede höchstens abzählbare Menge eine \mathbb{X} -Nullmenge ist, genügt es zu zeigen:

$$F_n^-(x) \xrightarrow{n \rightarrow \infty} G^-(x) \quad \text{für } \mathbb{X}\text{-fast alle } x \in (0,1).$$

Es sei daher $x_0 \in (0, 1) \cap C(G^-)$ und $\varepsilon > 0$. Da G höchstens abzählbar viele Unstetigkeitsstellen besitzt, liegt $C(G)$ dicht in $\text{dom}(G)$. Daher existieren $y, z \in C(G)$ mit

$$y < G^-(x_0) < z \quad \text{und} \quad z - y < \varepsilon. \quad (59)$$

Wir zeigen nun

$$G(y) < x_0 < G(z). \quad (60)$$

Angenommen, es gelte $G(y) \geq x_0$, dann folgt aus Lemma 5.12 (iv) und der Monotonie von G^-

$$y \geq G^-(G(y)) \geq G^-(x_0)$$

im Widerspruch zu $y < G^-(x_0)$. Umgekehrt folgt aus $G^-(x_0) < z$ mit Lemma 5.12 (iv) und der Monotonie von G

$$x_0 \leq G(G^-(x_0)) \leq G(z).$$

Angenommen, es gelte $x_0 = G(z)$. Dann wäre aber für alle $\delta > 0$ mit $x_0 + \delta < \sup(G)$

$$G^-(x_0 + \delta) = \inf\{u : G(u) \geq x_0 + \delta\} > z.$$

Da voraussetzungsgemäß $G^-(x_0) < z$ gilt, wäre das ein Widerspruch zur Stetigkeit von G^- in x_0 . Damit ist (60) gezeigt. Da

$$F_n(x) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in C(G)$$

gilt und wegen $G(y) < x_0 < G(z)$ und $y, z \in C(G)$ existiert ein $u(\varepsilon) \in \mathbb{N}$ mit

$$F_n(y) < x_0 < F_n(z) \quad \forall n > u(\varepsilon).$$

Daraus folgt

$$y \leq F_n^-(x_0) \leq z \quad \forall n > u(\varepsilon),$$

denn mit der Monotonie von F_n^- und Lemma 5.12 (iv) gilt

$$F_n^-(x_0) \leq F_n^-(F_n(z)) \leq z.$$

Für die andere Ungleichung führt man die Annahme $y > F_n^-(x_0)$ zum Widerspruch. Denn aus dieser Annahme würde mit der Monotonie von F_n und Lemma 5.12 (iv)

$$F_n(y) \geq F_n(F_n^-(x_0)) \geq x_0$$

folgen. Es gilt somit

$$y \leq F_n^-(x_0) \quad \text{und} \quad G^-(x_0) \leq z \quad \forall n \geq n(\varepsilon).$$

Mit (59) ergibt sich daher

$$|F_n^-(x_0) - G^-(x_0)| \leq z - y < \varepsilon \quad \forall n \geq n_\varepsilon.$$

Da G^- höchstens abzählbar viele Unstetigkeitsstellen besitzt, folgt die Behauptung. \square

Eine direkte Konsequenz ist

Lemma 5.15 *Es sei $(F_n)_n$ eine Folge von Verteilungsfunktionen und G eine Verteilungsfunktion, so dass*

$$F_n(x) \rightarrow G(x) \quad \forall x \in C(G).$$

$(\alpha_n)_n$ und (β_n) seien Folgen reeller Zahlen mit $\alpha_n > 0 \quad \forall n$,

$$\alpha_n \rightarrow \alpha > 0 \quad \text{und} \quad \beta_n \rightarrow \beta \in \mathbb{R}.$$

Dann gilt

$$F_n(\alpha_n x + \beta_n) \xrightarrow{n \rightarrow \infty} G(\alpha x + \beta)$$

für alle $\alpha x + \beta \in C(G)$.

Beweis: Es seien $U \sim \mathcal{R}(0, 1)$, $Y := G^-(U)$ und $X_n := F_n^-(U)$ für jedes $n \in \mathbb{N}$. Wieder haben die X_n die Verteilungsfunktion F_n für jedes n und Y hat die Verteilungsfunktion G . Mit Satz 5.14 folgt

$$X_n \rightarrow Y \quad \mathbb{P}\text{-f.s.},$$

daher auch

$$\frac{X_n - \beta_n}{\alpha_n} \rightarrow \frac{Y - \beta}{\alpha} \quad \mathbb{P}\text{-f.s.}$$

Da fast-sichere Konvergenz Verteilungskonvergenz impliziert, gilt auch

$$\mathbb{P} \left(\frac{X_n - \beta_n}{\alpha_n} \leq x \right) \rightarrow \mathbb{P} \left(\frac{Y - \beta}{\alpha} \leq x \right)$$

für alle $\alpha x + \beta \in C(G)$, d. h.

$$F_n(\alpha_n x + \beta_n) \xrightarrow{n \rightarrow \infty} G(\alpha x + \beta)$$

für alle $\alpha x + \beta \in C(G)$. □

Damit kann man ein wichtiges Hilfsmittel zeigen, das auch eigenständige Bedeutung besitzt.

Satz 5.16 (Khinchin)

Es sei G eine nicht-entartete Verteilungsfunktion und $(F_n)_n$ eine Folge von Verteilungsfunktionen. Es seien (α_n) und (β_n) Folgen reeller Zahlen mit $\alpha_n > 0$ und es gelte

$$F_n(\alpha_n x + \beta_n) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in C(G).$$

Dann gilt die folgende Äquivalenz:

Für Folgen reeller Zahlen $(\tilde{\alpha}_n)$ und $(\tilde{\beta}_n)$ mit $\tilde{\alpha}_n > 0$ für alle n und eine geeignete nicht-entartete Verteilungsfunktion \tilde{G} gilt genau dann

$$F_n(\tilde{\alpha}_n x + \tilde{\beta}_n) \xrightarrow{n \rightarrow \infty} \tilde{G}(x) \quad \forall x \in C(G),$$

wenn Konstanten $\alpha > 0$ und $\beta \in \mathbb{R}$ existieren mit

$$\frac{\tilde{\alpha}_n}{\alpha_n} \xrightarrow{n \rightarrow \infty} \alpha \quad \text{und} \quad \frac{\tilde{\beta}_n - \beta_n}{\alpha_n} \xrightarrow{n \rightarrow \infty} \beta.$$

In diesem Fall gilt

$$\tilde{G}(x) = G(\alpha x + \beta),$$

d. h. G und \tilde{G} sind vom selben Typ.

Neben seiner technischen Bedeutung beantwortet dieser Satz auch die Frage nach der Eindeutigkeit der Grenzverteilung normierter Maxima. Existiert für geeignet normierte Maxima $A_n(M_n - B_n)$ von iid Zufallsvariablen eine nicht-entartete Grenzverteilung, so ist diese bis auf Typ-Gleichheit auch eindeutig bestimmt. Darüber hinaus charakterisiert dieser Satz sämtliche Konstantenfolgen $A_n > 0$ und B_n , für die $F^n(A_n x + B_n)$ gegen eine nicht-entartete Verteilungsfunktion konvergiert in Abhängigkeit von einem schon gefundenen Paar (α_n) , $\alpha_n > 0$ und (β_n) .

Beweis: Wir setzen

$$\alpha_n^* := \frac{\tilde{\alpha}_n}{\alpha_n} \quad \text{und} \quad \beta_n^* := \frac{\tilde{\beta}_n - \beta_n}{\alpha_n} \quad \text{und} \quad F_n^*(x) = F_n(\alpha_n x + \beta_n) \quad \forall x \in \mathbb{R}.$$

“ \Leftarrow ”: Es gebe also Konstanten $\alpha > 0$ und β mit $\alpha_n^* \rightarrow \alpha$ und $\beta_n^* \rightarrow \beta$ für $n \rightarrow \infty$. Wegen $\alpha_n, \tilde{\alpha}_n > 0 \quad \forall n$, gilt auch $\alpha_n^* > 0 \quad \forall n \in \mathbb{N}$. Nach Voraussetzung gilt

$$F_n^*(x) \rightarrow G(x) \quad \forall x \in C(G).$$

Damit sind die Voraussetzungen von Lemma 5.15 erfüllt. Nach diesem ergibt sich:

$$\lim_{n \rightarrow \infty} F_n(\tilde{\alpha}_n x + \tilde{\beta}_n) = \lim_{n \rightarrow \infty} F_n^*(\alpha_n^* x + \beta_n^*) = G(\alpha x + \beta) =: \tilde{G}(x) \quad \forall \alpha x + \beta \in C(G).$$

Somit gilt die Konvergenz auch für alle $x \in C(\tilde{G})$.

“ \Rightarrow ”: Es gelte

$$F_n^*(x) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in C(G)$$

und

$$F_n^*(\alpha_n^* x + \beta_n^*) \xrightarrow{n \rightarrow \infty} \tilde{G}(x) \quad \forall x \in C(\tilde{G}) \quad (61)$$

für eine nicht-entartete Verteilungsfunktion \tilde{G} .

Wir zeigen zunächst: (α_n^*) und (β_n^*) sind beschränkt: Da \tilde{G} eine nicht entartete Verteilungsfunktion ist, gibt es $x, y \in C(\tilde{G})$ mit $x \neq y$ und $0 < \tilde{G}(x), \tilde{G}(y) < 1$. (Zunächst ist nur die Existenz eines solchen x klar, aufgrund der rechtsseitigen Stetigkeit gibt es aber ein weiteres y mit derselben Eigenschaft.) Dann folgt aber die Beschränktheit von $(\alpha_n^* x + \beta_n^*)_n$ und $(\alpha_n^* y + \beta_n^*)_n$, da sonst mit (61)

$$\begin{aligned} F_n^*(\alpha_n^* x + \beta_n^*) &\xrightarrow{n \rightarrow \infty} \tilde{G}(x) \in (0, 1) \quad \text{und} \\ F_n^*(\alpha_n^* y + \beta_n^*) &\xrightarrow{n \rightarrow \infty} \tilde{G}(y) \in (0, 1) \end{aligned}$$

nicht gelten könnte. Dann ist aber auch die Folge

$$(\alpha_n^* x + \beta_n^* - (\alpha_n^* y + \beta_n^*)) = (\alpha_n^* (x - y))$$

beschränkt. Dann ist auch die Folge (α_n^*) beschränkt. Dann aber auch (β_n^*) , da $(\alpha_n^* x + \beta_n^*)$ beschränkt ist.

Nach dem Satz von Bolzano-Weierstrass gibt es also Teilfolgen $(\alpha_{n_j})_j$ und $(\beta_{n_j})_j$ von $(\alpha_n)_n$ bzw. $(\beta_n)_n$ und Konstanten $\alpha \geq 0$ und $\beta \in \mathbb{R}$ mit

$$\alpha_{n_j}^* \xrightarrow{j \rightarrow \infty} \alpha \quad \text{und} \quad \beta_{n_j}^* \xrightarrow{j \rightarrow \infty} \beta.$$

Mit (61) folgt daher

$$F_{n_j}^*(\alpha_{n_j}^* x + \beta_{n_j}^*) \xrightarrow{j \rightarrow \infty} \tilde{G}(x) \quad \forall x \in C(\tilde{G}).$$

Daher kann nicht $\alpha = 0$ gelten, denn sonst wäre \tilde{G} entartet oder konstant. Somit gilt $\alpha > 0$.

Es bleibt zu zeigen: $\alpha_n^* \rightarrow \alpha$ und $\beta_n^* \rightarrow \beta$. Dazu seien $\alpha' > 0$ und $\beta' \in \mathbb{R}$ weitere Konstanten und $(\alpha_{n_k}^*)$ und $(\beta_{n_k}^*)$ weitere Teilfolgen von (α_n^*) und (β_n^*) mit

$$\alpha_{n_k}^* \xrightarrow{k \rightarrow \infty} \alpha' \quad \text{und} \quad \beta_{n_k}^* \xrightarrow{k \rightarrow \infty} \beta'.$$

Wir wollen zeigen: $\alpha = \alpha'$ und $\beta = \beta'$. Aus Lemma 5.15 folgt

$$G(\alpha x + \beta) = \tilde{G}(x) = G(\alpha' x + \beta') \quad \forall x \in C(\tilde{G}).$$

Nun gilt diese Gleichheit aber schon für alle $x \in \mathbb{R}$, denn da $C(\tilde{G})$ dicht in \mathbb{R} liegt, existiert für jedes x aus dem Unstetigkeitsbereich von G eine Folge $(x_n)_n$ aus $C(\tilde{G})$ mit $x_n \downarrow x$. Mit der rechtsseitigen Stetigkeit von G und \tilde{G} folgt die Behauptung.

Da G nicht entartet ist, existieren $x_1 < x_2$ mit

$$0 < y_1 := G(x_1) < y_2 := G(x_2) \leq 1.$$

Es werde zunächst $y_2 < 1$ angenommen. Dann gilt für $i \in \{1, 2\}$

$$\begin{aligned} \tilde{G}^-(y_i) &= \inf\{\omega : \tilde{G}(\omega) \geq y_i\} = \inf\{\omega : G(\alpha\omega + \beta) \geq y_i\} \\ &= \inf\left\{\frac{1}{\alpha}(\omega - \beta) : G(\omega) \geq y_i\right\} \\ &= \frac{1}{\alpha}(G^-(y_i) - \beta) \end{aligned}$$

und

$$\begin{aligned} \tilde{G}^-(y_i) &= \inf\{w : \tilde{G}(w) \geq y_i\} = \inf\{w : G(\alpha'w + \beta') \geq y_i\} \\ &= \inf\left\{\frac{1}{\alpha'}(w - \beta') : G(w) \geq y_i\right\} \\ &= \frac{1}{\alpha'}(G^-(y_i) - \beta'). \end{aligned}$$

Also folgt

$$\begin{aligned}\tilde{G}^-(y_1) &= \frac{1}{\alpha}(G^-(y_1) - \beta) = \frac{1}{\alpha'}(G^-(y_1) - \beta') \\ \tilde{G}^-(y_2) &= \frac{1}{\alpha}(G^-(y_2) - \beta) = \frac{1}{\alpha'}(G^-(y_2) - \beta').\end{aligned}\tag{62}$$

Durch Subtraktion folgt

$$\frac{1}{\alpha}(G^-(y_1) - G^-(y_2)) = \frac{1}{\alpha'}(G^-(y_1) - G^-(y_2)).$$

Ist der Klammerausdruck verschieden von null, so folgt $\alpha = \alpha'$ und dann wegen (62) auch $\beta = \beta'$. Dies ist aber der Fall, da $G^-(y_1) < G^-(y_2)$ gilt. Denn zunächst folgt mithilfe von Lemma 5.12 (iv) aus $y_1 = G(x_1)$

$$G^-(y_1) = G^-(G(x_1)) \leq x_1$$

und dann folgt mithilfe der Monotonie von G und der Definition von G^- aus $G(x_1) = y_1 < y_2$

$$G^-(y_2) > x_1 \geq G^-(y_1).$$

Somit ist in diesem Fall $\alpha' = \alpha$ und $\beta' = \beta$. Nun sei $y_2 = 1$. Wegen $G(x_2) = y_2$ ist

$$\{x : G(x) \geq 1\} \neq \emptyset.$$

Also kann die Pseudo-Inverse auf $(0, 1]$ definiert werden und somit gilt auch in diesem Fall analog $\alpha' = \alpha$ und $\beta' = \beta$. Damit galt aber schon

$$\alpha_n^* \rightarrow \alpha \quad \text{und} \quad \beta_n^* \rightarrow \beta.$$

Dies zeigt den Satz. □

Unter Ausnutzung des Satzes von Khinchin lässt sich nun auch Satz 5.10 zeigen.

Beweis von Satz 5.10: “ \Rightarrow ”: G sei max-stabil. Dann existieren nach Definition der max-Stabilität für $F_n := G^n$, $n \in \mathbb{N}$, Folgen reeller Zahlen $(a_n)_n$ und (b_n) mit $a_n > 0$ und

$$F_n(a_{nk}x + b_{nk}) = G^{nk}(a_{nk}x + b_{nk})^{1/k} = (G(x))^{1/k}$$

für alle $k \in \mathbb{N}$, für alle $x \in \mathbb{R}$ und für alle $n \in \mathbb{N}$. Somit gilt insbesondere

$$F_n(a_{nk}x + b_{nk}) \xrightarrow{n \rightarrow \infty} G^{1/k}(x)$$

für alle $x \in C(G)$, für alle $k \in \mathbb{N}$.

“ \Leftarrow ”: Mit G ist auch $G^{1/k}$ für alle k eine nicht-entartete Verteilungsfunktion. Nach Voraussetzung gilt

$$F_n(a_{nk}x + b_{nk}) \xrightarrow{n \rightarrow \infty} G^{1/k}(x)$$

für alle $k \in \mathbb{N}$, für $x \in C(G)$. Es sei nun $k \in \mathbb{N}$ und

$$\tilde{\alpha}_n := a_{nk}, \quad \tilde{\beta}_n := b_{nk}, \quad \tilde{G} = G^{1/k}.$$

Damit ergibt sich aus den Voraussetzungen

$$\begin{aligned} F_n(\tilde{\alpha}_n x + \tilde{\beta}_n) &\rightarrow \tilde{G}(x) \quad \forall x \in C(\tilde{G}) = C(G) \\ F_n(a_n x + b_n) &\rightarrow G(x) \quad \forall x \in C(G), \end{aligned}$$

wobei die zweite Zeile aus $k = 1$ folgt. Damit sind die Voraussetzungen des Satzes von Khinchin erfüllt. Daher folgt für geeignete Konstanten $\alpha_k > 0$ und $\beta_k \in \mathbb{R}$

$$\tilde{G}(x) = G^{1/k}(x) = G(\alpha_k x + \beta_k)$$

für alle $x \in \mathbb{R}$. D. h. für alle $x \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt

$$G(x) = G^n(\alpha_n x + \beta_n).$$

Also ist G max-stabil. □

Wir kümmern uns nun darum, die nicht-entarteten Extremwertverteilungen genauer zu beschreiben. Der folgende Satz zeigt, dass wir mit der Gumbel-, der Fréchet- und der Weibull-Verteilung schon alle möglichen Typen von Extremwertverteilungen kennengelernt haben.

Satz 5.17 (*Fisher-Tippett-Theorem*)

Eine nicht-entartete Verteilung G ist genau dann eine Extremwertverteilung, wenn G zum Typ der Weibull-, Gumbel- oder Fréchet-Verteilung gehört.

Dieser Satz wird in zwei großen Schritten gezeigt. Zunächst leitet man her, dass die nicht-entarteten Verteilungen gerade die max-stabilen Verteilungen sind, dann beweist man, dass diese gerade mit den im Satz erwähnten Verteilungen übereinstimmen.

Wir beginnen mit dem ersten Schritt.

Satz 5.18 *Für eine nicht-entartete Verteilungsfunktion G sind die folgenden Aussagen äquivalent:*

(i) $\mathcal{D}(G) \neq \emptyset$.

(ii) G ist max-stabil.

Beweis: (ii) \Rightarrow (i): ist offensichtlich, denn wir wissen definitionsgemäß, dass stets $G \in \mathcal{D}(G)$ gilt, wenn G eine max-stabile Verteilung ist.

(i) \Rightarrow (ii): Es sei $F \in \mathcal{D}(G)$, also gelte

$$F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in C(G)$$

für geeignete Folgen (a_n) , $a_n > 0$ für alle n und $(b_n), b \in \mathbb{R}$. Dann gilt aber schon

$$F^{nk}(a_{nk}x + b_{nk}) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in C(G) \quad \forall k \in \mathbb{N}$$

und somit

$$F^n(a_{nk}x + b_{nk}) \xrightarrow{n \rightarrow \infty} G^{1/k}(x) \quad \forall x \in C(G) \quad \forall k \in \mathbb{N}.$$

Mit Satz 5.10 folgt, dass G max-stabil ist. □

Dies ist der erste und einfachere Teil des Satzes. Alle uns bekannten Extremwertverteilungen sind also zwingend max-stabil. Wir zeigen nun, dass die bekannten Extremwertverteilungen auch die einzigen sind. Wir beginnen mit

Lemma 5.19 *Sei G eine max-stabile, nicht-entartete Verteilungsfunktion. Dann existieren messbare Funktionen*

$$\begin{aligned} a : \mathbb{R}^+ &\rightarrow \mathbb{R}^+ \\ b : \mathbb{R}^+ &\rightarrow \mathbb{R}^+ \\ \text{mit} \quad G^s(a(s)x + b(s)) &= G(x) \quad \forall x \in \mathbb{R}, \forall s > 0. \end{aligned}$$

Beweis: Sei G max-stabil. Bezeichne mit $\lceil \cdot \rceil$ die obere Gauß-Klammer. Dann gilt mit der Definition der Max-Stabilität

$$G^{\lceil ns \rceil}(a_{\lceil ns \rceil}x + b_{\lceil ns \rceil}) = G(x)$$

für alle $x \in \mathbb{R}$, für alle $s > 0$ und für alle $n \in \mathbb{N}$. Da

$$\frac{n}{\lceil ns \rceil} \xrightarrow{n \rightarrow \infty} \frac{1}{s} \quad \forall s > 0$$

konvergiert, ergibt sich

$$G^n(a_{\lceil ns \rceil}x + b_{\lceil ns \rceil}) = (G^{\lceil ns \rceil}(a_{\lceil ns \rceil}x + b_{\lceil ns \rceil}))^{\frac{n}{\lceil ns \rceil}} = G^{\frac{n}{\lceil ns \rceil}}(x) \rightarrow G^{1/s}(x)$$

für alle $x \in \mathbb{R}$ und alle $s > 0$. Wählen wir insbesondere $s = 1$, so ergibt sich

$$G^n(a_nx + b_n) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in \mathbb{R}.$$

Wählen wir in den letzten beiden Konvergenzen

$$\alpha_n := a_n, \beta_n := b_n, \tilde{\alpha}_n := a_{\lceil ns \rceil} \quad \text{und} \quad \tilde{\beta}_n := b_{\lceil ns \rceil},$$

so sind die Voraussetzungen des Satzes von Khinchin erfüllt. Somit existieren für jedes $s > 0$ geeignete $a(s) > 0$ und $b(s) \in \mathbb{R}$ mit

$$\frac{a_{\lceil ns \rceil}}{a_n} \xrightarrow{n \rightarrow \infty} a(s) \quad \text{und} \quad \frac{b_{\lceil ns \rceil} - b_n}{a_n} \xrightarrow{n \rightarrow \infty} b(s).$$

Dies ergibt Funktionen

$$\begin{aligned} a : \mathbb{R}^+ &\rightarrow \mathbb{R}^+ \\ s &\mapsto \lim_{n \rightarrow \infty} \frac{a_{[ns]}}{a_n} \\ \text{und} \quad b : \mathbb{R}^+ &\rightarrow \mathbb{R} \\ s &\mapsto \lim_{n \rightarrow \infty} \frac{b_{[ns]}}{a_n}. \end{aligned}$$

Diese sind als Grenzwerte messbarer Funktionen messbar und besitzen die Eigenschaft

$$G^{1/s}(x) = G(a(s)x + b(s)) \quad \forall x \in \mathbb{R} \quad \forall s > 0.$$

□

Nun wenden wir uns dem angekündigten letzten Schritt zu:

Satz 5.20 *Für eine nicht-entartete Verteilungsfunktion G gilt die folgende Äquivalenz:*

(i) *G ist max-stabil.*

(ii) *G gehört zum Typ der Gumbel-, Fréchet- oder Weibullverteilung.*

Beweis: Der Beweis ist aufwändig. Wir beginnen mit

(ii) \Rightarrow (i) (dem einfachen Teil): Wir haben bereits nachgerechnet, dass die Gumbel-, Weibull- und Fréchet-Verteilungen max-stabil sind (siehe Beispiel 5.6 und 5.9).

(i) \Rightarrow (ii): Die Idee in diesem Beweisschritt besteht darin, die Funktionalgleichung aus Lemma 5.19 zu verwenden und zu zeigen, dass deren einzige Lösung Verteilungsfunktionen der gewünschten Typen sind. Wir betrachten die Funktion

$$\begin{aligned} \psi : T &\rightarrow \mathbb{R} \\ x &\mapsto \psi(x) := -\log(-\log G(x)), \end{aligned}$$

wobei

$$T = \{x \in \mathbb{R} : 0 < G(x) < 1\}$$

ist.

Behauptung: ψ besitzt eine Pseudo-Inverse ψ^- auf \mathbb{R} .

Beweis: Da G als Verteilungsfunktion monoton nicht-fallend ist und $\log(\cdot)$ strikt monoton steigt, ist ψ monoton nicht-fallend. Außerdem ist ψ rechtsseitig stetig. Somit bleibt noch

$$\inf_{x \in T} \psi(x) = -\infty \tag{63}$$

und

$$\sup_{x \in T} \psi(x) = +\infty \tag{64}$$

nachzuweisen. (63) ist gezeigt, wenn wir zeigen können, dass T kein kleinstes Element enthält, das unter G Masse hat, denn dann gilt

$$\inf_{x \in T} G(x) = 0 \quad \text{und somit} \\ \inf_{x \in T} \psi(x) = -\infty.$$

Angenommen, es gibt ein $x_L \in \mathbb{R}$ mit

$$G(x_L) > 0 \quad \text{und} \quad \lim_{x \uparrow x_L} G(x) = 0. \quad (65)$$

Aus der Max-Stabilität von G folgt

$$G^2(ax + b) = G(x) \quad \forall x \in \mathbb{R}$$

für geeignetes $a > 0$ und $b \in \mathbb{R}$. Gibt es nun ein x_L , das (65) erfüllt, so gilt:

$$(i) \text{ für } ax_L + b < x_L : 0 = G^2(ax_L + b) = G(x_L) > 0,$$

$$(ii) \text{ für } ax_L + b > x_L : \text{Wegen } x_L > \frac{x_L - b}{a}$$

$$0 < G^2(x_L) = G^2(a \frac{x_L - b}{a} + b) = G(\frac{x_L - b}{a}) = 0,$$

$$(iii) \text{ für } ax_L + b = x_L :$$

$$G^2(x_L) = G^2(ax_L + b) = G(x_L) > 0.$$

In jedem dieser Fälle ergibt sich ein Widerspruch. Dieser ist in (i) und (ii) evident. In (iii) ergäbe sich sofort, dass $G(x_L) = 1$ ist und somit, dass G entartet ist. Also ist (65) zum Widerspruch geführt und es gilt (63).

Analog zeigt man, dass T kein größtes Element enthält, das unter G Masse trägt, denn dann gilt

$$\sup_{x \in T} G(x) = 1, \quad \text{also} \quad \sup_{x \in T} \psi(x) = +\infty.$$

Angenommen, es gäbe ein Element x_R mit

$$G(x_R) = 1 \quad \text{und} \quad \lim_{x \uparrow x_R} G(x) < 1.$$

Dann folgte wie oben

$$(i) \text{ für } ax_R + b < x_R : 1 > G^2(ax_R + b) = G(x_R) = 1,$$

$$(ii) \text{ für } ax_R + b > x_R : 1 = G^2(x_R) = G^2(a \frac{x_R - b}{a} + b) = G(\frac{x_R - b}{a}) < 1,$$

$$(iii) \text{ für } ax_R + b = x_R : \lim_{x \uparrow x_R} G^2(x) = \lim_{x \uparrow x_R} G^2(ax + b) = \lim_{x \uparrow x_R} G.$$

Wie oben erhalten wir einen Widerspruch, womit auch (64) gezeigt ist. Daher ist gezeigt, dass ψ eine Pseudo-Inverse ψ^- auf \mathbb{R} besitzt.

Da G eine nicht-entartete max-stabile Verteilung ist, können wir Lemma 5.19 anwenden. Dies liefert die Existenz messbarer Funktionen $a : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ und $b : \mathbb{R}^+ \rightarrow \mathbb{R}$ mit

$$G^s(a(s)x + b(s)) = G(x) \quad \forall x \in \mathbb{R}, \quad \forall s > 0.$$

Daraus ergibt sich

$$\begin{aligned} \psi(x) &= -\log(-\log G(x)) = -\log(-\log(G^s(a(s)x + b(s)))) \\ &= -\log(-\log(G(a(s)x + b(s))) \cdot s) \\ &= \psi(a(s)x + b(s)) - \log s \end{aligned} \tag{66}$$

für alle $x \in \mathbb{R}$ und $s > 0$. Damit ergibt sich für ψ^- :

$$\begin{aligned} \psi^-(y) &= \inf\{x : \psi(x) \geq y\} \\ &= \inf\{x : \psi(a(s)x + b(s)) - \log s \geq y\} \\ &= \inf\left\{\frac{x - b(s)}{a(s)} : \psi(x) \geq y + \log(s)\right\} \\ &= \frac{\psi^-(y + \log(s)) - b(s)}{a(s)} \end{aligned} \tag{67}$$

für alle $y \in \mathbb{R}$ und für alle $s > 0$, wobei die zweite Gleichung aus (66) folgt. Setzen wir $y = 0$, so ergibt sich

$$\psi^-(0) = \frac{\psi^-(\log(s)) - b(s)}{a(s)} \quad \forall s > 0. \tag{68}$$

Wir subtrahieren (68) von (67) und erhalten:

$$\psi^-(y + \log(s)) - \psi^-(\log s) = a(s)(\psi^-(y) - \psi^-(0))$$

für alle $y \in \mathbb{R}$ und $s > 0$. Setzen wir

$$z := \log s, \quad \tilde{a}(z) := a(e^z), \quad g(y) := \psi^-(y) - \psi^-(0).$$

Damit erhalten wir

$$g(y + z) - g(z) = g(y)\tilde{a}(z) \quad \forall y, z \in \mathbb{R}. \tag{69}$$

Wir bestimmen nun die Lösungen von (69). Hierfür nehmen wir zunächst an, $\tilde{a}(z)$ wäre identisch gleich 1. Dann ist

$$g(y + z) = g(y) + g(z) \quad \forall y, z \in \mathbb{R} \tag{70}$$

zu lösen. Die Lösung dieser Funktionalgleichung ist eindeutig (bis auf Konstanten), d. h., setzen wir

$$c = g(1),$$

so ist die einzige Lösung dieser Funktionalgleichung die lineare Funktion

$$g(y) = cy.$$

Aus der Monotonie von ψ^- folgt nun

$$g(1) = \psi^-(1) - \psi^-(0) \geq 0.$$

Dann muss aber schon $g(1) > 0$ gelten, denn sonst wäre $g \equiv 0$ und somit ψ^- konstant und G entartet im Widerspruch zu den Voraussetzungen. Daher können wir $\alpha := \frac{1}{c}$ definieren und es folgt

$$g(y) = \frac{y}{\alpha} \quad \text{mit} \quad \alpha > 0.$$

Mit g ist dann aber ψ^- stetig. Setzen wir

$$\beta := -\alpha\psi^-(0),$$

so erhalten wir

$$\begin{aligned} x &= \psi^-(\psi(x)) = g(\psi(x)) + \psi^-(0) \\ &= \frac{\psi(x)}{\alpha} + \psi^-(0) \\ &= \frac{\psi(x)}{\alpha} - \frac{\beta}{\alpha} \quad \forall x \in T. \end{aligned}$$

Hierbei haben wir für die erste Gleichheit die Stetigkeit von ψ^- ausgenutzt, während die zweite Gleichheit aus der Definition von g folgt. Damit ergibt sich

$$\psi(x) = \alpha x + \beta \quad \forall x \in T,$$

also insbesondere

$$G(x) = e^{-e^{-\alpha x + \beta}} \quad \forall x \in \mathbb{R}.$$

Also ist G vom Typ der Gumbel-Verteilung.

Nun betrachten wir den Fall, dass $\tilde{a}(z) \not\equiv 1$ gilt. Es gibt also mindestens ein $z_0 \in \mathbb{R}$ mit $\tilde{a}(z_0) \neq 1$. Aus (69) folgt durch Vertauschen von y und z :

$$g(y+z) - g(y) = g(z)\tilde{a}(y) \quad \forall y, z \in \mathbb{R}. \quad (71)$$

Subtrahieren wir (71) von (69), so ergibt sich

$$-g(z) + g(y) = g(y)\tilde{a}(z) - g(z)\tilde{a}(y) \quad \forall y, z \in \mathbb{R},$$

also

$$g(z)(-\tilde{a}(y) + 1) = g(y)(-\tilde{a}(z) + 1) \quad \forall y, z \in \mathbb{R}.$$

Insbesondere ergibt sich für $z = z_0$ und $d := \frac{g(z_0)}{1-\tilde{a}(z_0)}$

$$g(y) = \frac{g(z_0)}{1-\tilde{a}(z_0)}(1-\tilde{a}(y)) = d \cdot (1-\tilde{a}(y)) \quad \forall y \in \mathbb{R}. \quad (72)$$

Dabei ist $d \neq 0$, denn sonst wäre wieder $g \equiv 0$ und G entartet. Damit folgt aus (71)

$$d(1 - \tilde{a}(y+z)) - d(1 - \tilde{a}(z)) = d(1 - \tilde{a}(y))\tilde{a}(z) \quad \forall y, z \in \mathbb{R}.$$

Dividieren wir durch d , so erhalten wir

$$\tilde{a}(y+z) = \tilde{a}(y) \cdot \tilde{a}(z) \quad \forall y, z.$$

Diese Gleichung aber lässt sich auf bekannte Weise lösen: Da $\tilde{a}(y) = a(e^y) > 0$ ist, dürfen wir beide Seiten logarithmieren und erhalten

$$\log(\tilde{a}(y+z)) = \log(\tilde{a}(y)) + \log(\tilde{a}(z))$$

für alle $y, z \in \mathbb{R}$. Die einzige Lösung dieser Gleichung ist

$$\log(\tilde{a}(y)) = cy \quad \forall y \in \mathbb{R}$$

mit $c = \log \tilde{a}(1) = \log a(e)$. Also gilt

$$\tilde{a}(y) = e^{cy} \quad \forall y \in \mathbb{R}.$$

Nun ist aber $c \neq 0$, denn sonst wäre $\tilde{a} \equiv 1$ im Widerspruch zur Voraussetzung. Mit $\gamma = \frac{1}{c}$ erhalten wir also, dass die einzigen Lösungen von der Form

$$\tilde{a}(y) = e^{y/\gamma} \quad \forall y \in \mathbb{R}. \quad (73)$$

Dabei ist auch $\gamma \neq 0$. Setzen wir $\beta := \psi^-(0)$, so ergibt sich

$$\psi^-(y) = g(y) + \psi^-(0) = d(1 - \tilde{a}(y)) + \beta = d(1 - e^{y/\gamma}) + \beta \quad (74)$$

für alle $y \in \mathbb{R}$. Hierbei folgt die Gültigkeit der zweiten Gleichung aus (72) und die der dritten Gleichung aus (5.17). Wegen der Monotonie von ψ^- muss $\frac{d}{\gamma} < 0$ gelten, denn für $z > y$ ist wegen dieser Monotonie

$$\varphi^-(z) - \psi^-(y) \geq 0,$$

woraus wegen (74)

$$d(e^{y/\gamma} - e^{z/\gamma}) \geq 0$$

folgt. Für $d > 0$ muss also $\frac{y}{\gamma} \geq \frac{z}{\gamma}$, also $\gamma < 0$, gelten. Analog folgt aus $d < 0$, dass $\gamma > 0$ gilt. Nun gilt aber

$$\begin{aligned} x &= \psi^-(\psi(x)) = \beta + d(1 - e^{\psi(x)/\gamma}) \\ &= \beta + d \left(1 - \exp \left(\frac{-\log(-\log(G(x)))}{\gamma} \right) \right) \\ &= \beta + d(1 - (-\log G(x))^{-1/\gamma}) \quad \forall x \in T. \end{aligned}$$

Hierbei folgt die erste Identität aus (74), da dann ψ^- wieder stetig ist und die dritte Gleichheit ergibt sich direkt aus der Definition von ψ . Lösen wir obige Gleichung nach $G(x)$, so folgt

$$G(x) = \exp \left(- \left(1 - \frac{x - \beta}{d} \right)^{-\gamma} \right) \quad \forall x \in T. \quad (75)$$

Sei zunächst $\gamma > 0$. Mit

$$\hat{a} := -\frac{1}{d} > 0 \quad \text{und} \quad \hat{b} := 1 + \frac{\beta}{d}$$

ergibt sich

$$G(x) = \exp(-(\hat{a}x + \hat{b})^{-\gamma}) \quad \forall x \quad \text{mit} \quad 0 < G(x) < 1,$$

also für alle x mit $\hat{a}x + \hat{b} > 0$. Damit gilt

$$G(x) = \exp(-(\hat{a}x + \hat{b})^{-\gamma}) \mathbb{1}_{(0, \infty)}(\hat{a}x + \hat{b}) = \Phi_\gamma(\hat{a}x + \hat{b}).$$

Also ist G vom Fréchet-Typ.

Nun gelte $\gamma < 0$ und es sei $\alpha := -\gamma > 0$. Dann folgt mit

$$\hat{a} := \frac{1}{d} \quad \text{und} \quad \hat{b} := -1 - \frac{\beta}{d}$$

aus (75)

$$G(x) = \exp(-(-(\hat{a}x + \hat{b}))^\alpha) \quad \forall x \quad \text{mit} \quad 0 < G(x) < 1,$$

also alle x mit $\hat{a}x + \hat{b} < 0$. Damit gilt

$$G(x) = \exp(-(-(\hat{a}x + \hat{b}))^\alpha) \mathbb{1}_{(-\infty, 0)}(\hat{a}x + \hat{b}) + \mathbb{1}_{[0, \infty)}(\hat{a}x + \hat{b}) = \psi_\alpha(\hat{a}x + \hat{b}).$$

Also ist G vom Typ der Weibull-Verteilung mit Parameter $-\gamma$. Damit folgt (ii). \square

Damit sind wir auch in der Lage, das Fisher-Tippett-Theorem endgültig zu beweisen.

Beweis von Satz 5.17: G sei eine nicht-entartete Verteilungsfunktion. Dann ist der Anziehungsbereich von G nach Satz 5.18 genau dann nicht-leer, wenn G max-stabil ist. Das ist nach Satz 5.20 genau dann der Fall, wenn G vom Typ der Gumbel-, Fréchet- oder Weibull-Verteilung ist. \square

Wir haben somit die Klasse der nicht-entarteten Extremwertverteilungen vollständig beschrieben. Insbesondere sehen wir, dass alle in Frage kommenden Verteilungen stetig sind. Durch den bekannten Rechentrick

$$\max = -\min -$$

lassen sich so auch die Verteilungen der Minima von iid Folgen beschreiben.

Es bleibt noch die Frage zu untersuchen, ob jede Verteilungsfunktion im Anziehungsbereich einer der drei Extremwertverteilungen liegt. Diese ist zu verneinen, wie das folgende Beispiel zeigt.

Beispiel 5.21 *Es sei F eine Verteilungsfunktion mit*

$$x_R := \sup\{x \in \mathbb{R} : F(x) < 1\} < \infty$$

und es gebe ein $\rho < 1$ mit

$$\lim_{x \uparrow x_R} F(x) = \rho.$$

Dann liegt F nicht im Anziehungsbereich einer nicht-entarteten Extremwertverteilung, denn angenommen, G sei eine nicht-entartete Extremwertverteilung und $F \in \mathcal{D}(G)$. Dann gäbe es Folgen $(a_n)_n$ und $(b_n)_n$ mit $a_n > 0$ für alle $n \in \mathbb{N}$ und

$$F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x) \quad \forall x \in \mathbb{R},$$

denn nach dem Fisher-Tippett-Theorem gilt für jede nicht-entartete Extremwertverteilung G $C(G) = \mathbb{R}$. Da aus der rechtsseitigen Stetigkeit $F(x_R) = 1$ folgt, gälte für alle $n \in \mathbb{N}$

$$F^n(a_n x + b_n) \begin{cases} = 1 & \text{für } x \geq \frac{x_R - b_n}{a_n} \\ \leq \rho^n & \text{für } x < \frac{x_R - b_n}{a_n} \end{cases}.$$

Daraus würde aber wegen $F \in \mathcal{D}(G)$ für alle $x \in \mathbb{R}$

$$F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} \mathbb{1}_{[\lim_{n \rightarrow \infty} \frac{x_R - b_n}{a_n}, \infty)}(x) = G(x)$$

folgen. Damit wäre G aber entartet in Widerspruch zur Voraussetzung.

Aus dem letzten Beispiel sehen wir, dass jedes F mit endlichem rechten Endpunkt und positiver Masse in diesem nicht im Anziehungsbereich einer nicht-entarteten Extremwertfunktion liegt. Wir haben also gesehen, dass es drei verschiedene Typen von Extremwertverteilungen gibt und dass es Verteilungen gibt, die nicht im Anziehungsbereich von einer von diesen liegen. Umgekehrt stellt sich auch die Frage, ob wir im Fall der Konvergenz auch beschreiben können, welchen Typ von Limesverteilung wir erwarten sollen, d. h. ob wir die Anziehungsbereiche der Extremwertverteilungen charakterisieren können. Wie schon eingangs erwähnt, konvergiert das nicht-normierte M_n gegen x_R , d. h. den rechten Randpunkt der Verteilung F der X_1 , konvergiert. Es liegt daher nahe, dass die Frage, gegen welche Verteilung M_n konvergiert, vom Verhalten von F nahe x_R abhängt. Dies wird auch durch das letzte Beispiel gestützt. Es stellt sich dabei heraus, dass die wichtige Frage ist, "wie schnell" F in x_R hineingeht. Dies wird mit dem Begriff der Variation beschrieben. Genauer nennen wir eine Funktion regulär variierend (in $+\infty$), wenn sie sich asymptotisch wie eine Potenzfunktion verhält.

Definition 5.22 Eine messbare Funktion $f : (0, \infty) \rightarrow (0, \infty)$ heißt regulär variierend mit Index $\alpha \in \mathbb{R}$, falls

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha$$

für alle $x > 0$ gilt. Die Klasse der regulär variierenden Funktionen zum Index $\alpha \in \mathbb{R}$ bezeichnen wir in der Folge mit $\mathcal{R}_\alpha^\infty$.

Da für jede Verteilungsfunktion F gilt $F(x) \rightarrow 1$, wenn $x \rightarrow \infty$, ist diese immer in ∞ langsam variierend, d. h. regulär variierend mit Index 0. Allerdings ist dies nur eine sehr grobe Beschreibung des Verhaltens von F nahe x_R . Um dies genauer zu analysieren, definieren wir die Tailfunktion zu F und analysieren deren Verhalten.

Definition 5.23 a) F sei eine Verteilungsfunktion. Die durch

$$\bar{F}(x) := 1 - F(x) \quad \forall x \in \mathbb{R}$$

definierte Funktion heißt **Tailfunktion** von F .

b) \bar{F} heißt **regulär variierend mit Index α in ∞** , wenn

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^\alpha \quad \forall x > 0$$

gilt.

c) Wir nennen

$$x_R := \sup\{x : F(x) < 1\}$$

rechten Endpunkt der Verteilungsfunktion x .

Mit diesen Definitionen lassen sich die Anziehungsbereiche der Extremwertverteilungen folgendermaßen charakterisieren:

Satz 5.24 Es sei F eine Verteilungsfunktion und γ_n sei für $n \geq 2$ definiert durch

$$\gamma_n := F^{-1}\left(1 - \frac{1}{n}\right).$$

Dann gilt:

1. F liegt genau dann im Anziehungsbereich der Fréchet-Verteilung Φ_α , wenn \bar{F} in ∞ regulär variierend mit Index $-\alpha$ und $x_R = +\infty$ ist, d. h. wenn

$$x_R = +\infty \quad \text{und} \quad \lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha} \quad \forall x > 0 \quad (76)$$

gilt. In diesem Fall kann man die Folgen (A_n) und (B_n) für die schwache Konvergenz von $A_n(M_n - B_n)$ als $A_n = \gamma_n^{-1}$ und $B_n = 0$ wählen.

2. F liegt genau dann im Anziehungsbereich der Weibull-Verteilung ψ_α , wenn die zu

$$F^*(x) := F\left(x_R - \frac{1}{x}\right), \quad x > 0 \quad (77)$$

gehörende Tailfunktion $1 - F^*(x) = \bar{F}^*(x)$ in ∞ regulär variierend mit Index $-\alpha$ ist und $x_R < +\infty$ gilt, d. h. falls

$$x_R < +\infty \quad \text{und} \quad \lim_{t \rightarrow \infty} \frac{\bar{F}^*(tx)}{\bar{F}^*(t)} = x^{-\alpha} \quad \forall x \in \mathbb{R}^+$$

gilt. Die Konstantenfolgen für die Konvergenz von $A_n(M_n - B_n)$ lassen sich in diesem Fall als

$$A_n := \frac{1}{x_R - \gamma_n} \quad \text{und} \quad B_n := x_R$$

wählen.

3. F liegt genau dann im Anziehungsbereich der Gumbel-Verteilung, wenn eine positive, messbare Funktion g existiert, so dass

$$\lim_{t \uparrow x_R} \frac{\bar{F}(t + xg(t))}{\bar{F}(t)} = e^{-x} \quad \forall x \in \mathbb{R} \quad (78)$$

gilt. In diesem Fall können die Folgen (A_n) und (B_n) für die schwache Konvergenz von $A_n(M_n - B_n)$ als $A_n = \frac{1}{g(\gamma_n)}$ und $B_n = \gamma_n$ gewählt werden.

Es sei zunächst darauf hingewiesen, dass für hinreichend großes n stets $A_n > 0$ gilt. Dies ist im dritten Fall evident, da g positiv ist. Für die anderen Fälle betrachte $(\gamma_n)_{n \geq 2}$. Da

$$\gamma_n := \inf \left\{ t : F(t) \geq 1 - \frac{1}{n} \right\} \quad \forall n \geq 2$$

und

$$x_R := \sup \{ t : F(t) < 1 \}$$

gilt, folgt

$$\gamma_n \rightarrow x_R \quad \text{für } n \rightarrow \infty.$$

Da im Fréchet-Fall $x_R = +\infty$ ist, folgt $\gamma_n > 0$ für hinreichend großes n und somit ist auch A_n positiv. Im Weibullfall gilt natürlich stets $\gamma_n \leq x_R$. Es muss noch gezeigt werden, dass nicht $\gamma_n = x_R$ gelten kann. Wäre aber $\gamma_n = x_R$ für ein n , so wäre auch $F(\gamma_n) = F(x_R)$ und daher besäße x_R unter F Wahrscheinlichkeitsmasse. Beispiel 5.21 zeigt aber, dass dann F nicht im Anziehungsbereich einer Extremwertverteilung liegen kann.

Der Beweis ist relativ lang. Bevor wir ihn beginnen, wollen wir ein Beispiel betrachten:

Beispiel 5.25 1. Die $\mathcal{R}(0, 1)$ -Verteilung liegt im Anziehungsbereich der Weibullverteilung: Sei F die Verteilungsfunktion der $\mathcal{R}(0, 1)$ -Verteilung. Dann ist

$$x_R = \sup_x \{ x : F(x) < 1 \} = \sup \{ x : x < 1 \} = 1.$$

Für die durch

$$F^*(x) := F\left(1 - \frac{1}{x}\right)$$

definierte Funktion F^* gilt:

$$\begin{aligned} \lim_{t \uparrow \infty} \frac{\bar{F}^*(tx)}{\bar{F}^*(t)} &= \lim_{t \uparrow \infty} \frac{1 - F(1 - \frac{1}{tx})}{1 - F(1 - \frac{1}{t})} \\ &= \lim_{t \rightarrow \infty} \frac{\frac{1}{tx}}{\frac{1}{t}} = \frac{1}{x} \quad \forall x > 0. \end{aligned}$$

Also liegt F nach Satz 5.24 im Anziehungsbereich von ψ_1 . Die Konstanten für die schwache Konvergenz von $A_n(M_n - B_n)$ können als

$$A_n = \frac{1}{x_R - \gamma_n} = \frac{1}{1 - F^{-1}(1 - \frac{1}{n})} = n \quad \text{und} \quad B_n = 1$$

gewählt werden.

2. Die Pareto-Verteilung mit Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x \leq 1 \\ 1 - x^{-\alpha} & x \geq 1 \end{cases}$$

für festes $\alpha > 0$ liegt im Anziehungsbereich der Fréchet-Verteilung Φ_α : Offenbar ist

$$\frac{1}{x_R} = \sup\{x : F(x) < 1\} = +\infty.$$

Da mit $t \uparrow \infty$ auch $tx \uparrow \infty$ für alle $x > 0$ gilt, ergibt sich für alle $x > 0$

$$\lim_{t \uparrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = \lim_{t \uparrow \infty} \frac{1 - F(tx)}{1 - F(t)} = \lim_{t \uparrow \infty} \frac{(tx)^{-\alpha}}{t^{-\alpha}} = x^{-\alpha}.$$

Somit liegt F nach Satz 5.24 im Anziehungsbereich der Fréchet-Verteilung mit Index α , Φ_α . Für die Konstanten (A_n) und (B_n) kann man folgende Wahl treffen:

$$A_n = \frac{1}{F^-(1 - \frac{1}{n})} = \left(\inf \left\{ t : (1 - t^{-\alpha}) \mathbb{1}_{[1, \infty)}(t) \geq 1 - \frac{1}{n} \right\} \right)^{-1} = n^{-1/\alpha}$$

und $B_n = 0$.

3. Die Exponentialverteilung mit Parameter $\lambda > 0$ liegt im Anziehungsbereich der Gumbelverteilung Λ : Sei also F die Verteilungsfunktion der Exponentialverteilung

$$F(x) := 1 - e^{-\lambda x} \mathbb{1}_{[0, \infty)}(x).$$

Zunächst gilt offenbar

$$x_R = \sup\{x : F(x) < 1\} = +\infty.$$

Aus der Form der Verteilungsfunktion lässt sich schon ablesen, dass \bar{F} schneller als jede Potenzfunktion gegen 0 konvergiert. Dies lässt schon vermuten, dass $F \in \mathcal{D}(\Lambda)$ gilt. Tatsächlich gilt für die konstante Funktion $g \equiv \lambda^{-1}$ (die natürlich messbar ist)

$$\lim_{t \uparrow x_R} \frac{\bar{F}(t + g(t)x)}{\bar{F}(t)} = \lim_{t \uparrow \infty} \frac{\bar{F}(t + \frac{x}{\lambda})}{\bar{F}(t)} = \lim_{t \rightarrow \infty} \frac{e^{-\lambda t - x}}{e^{-\lambda t}} = e^{-x}.$$

Somit folgt aus Satz 5.24, dass $F \in \mathcal{D}(\lambda)$ gilt und dass sich für die schwache Konvergenz der Folge $A_n(M_n - B_n)$ die Konstanten

$$A_n = \frac{1}{g(\gamma_n)} = \lambda \quad \text{und} \quad B_n = \gamma_n = F^-\left(1 - \frac{1}{n}\right) = \frac{\log n}{n}$$

wählen lassen.

Den sehr aufwändigen Beweis von Satz 5.24 werden wir in mehreren Schritten führen. Wir beginnen mit häufiger verwendeten Lemmata:

Lemma 5.26 *Es sei F eine Verteilungsfunktion und es gelte (76), (77) oder (78). Ferner sei*

$$\gamma_n := F_n^-(1 - \frac{1}{n}) \quad \forall n \geq 2.$$

Dann folgt

$$\lim_{n \rightarrow \infty} n\bar{F}(\gamma_n) = 1.$$

Beweis: Zunächst folgern wir aus Lemma 5.12 (iv)

$$n\bar{F}(\gamma_n) = n \left(1 - F \left(F^- \left(1 - \frac{1}{n} \right) \right) \right) \leq n \left(1 - \left(1 - \frac{1}{n} \right) \right) = 1,$$

somit folgt auch

$$\limsup_{n \rightarrow \infty} n\bar{F}(\gamma_n) \leq 1.$$

Bleibt

$$\liminf_{n \rightarrow \infty} n\bar{F}(\gamma_n) \geq 1 \tag{79}$$

zu zeigen. Zunächst nehmen wir an, dass (76) gilt. Für $x \in (0, 1)$ und $n \geq 2$ sei $\delta_n(x) := \gamma_n x < \gamma_n$. γ_n ist für genügend großes n positiv, da $\gamma_n \rightarrow x_R = +\infty$ gilt. Somit gilt für diese n (und alle x) $F(\delta_n(x)) < 1 - \frac{1}{n}$. Somit auch

$$n\bar{F}(\delta_n(x)) > n \left(1 - \left(1 - \frac{1}{n} \right) \right) = 1.$$

Damit ergibt sich

$$n\bar{F}(\gamma_n) > \frac{n\bar{F}(\gamma_n)}{n\bar{F}(\delta_n(x))} = \frac{\bar{F}(\gamma_n)}{\bar{F}(\gamma_n x)} \xrightarrow{n \rightarrow \infty} x^\alpha$$

für alle $x \in (0, 1)$. Dabei folgt die Konvergenz aus der Tatsache, dass \bar{F} regulär variierend ist mit Index $-\alpha$ und $\gamma_n \rightarrow x_R = +\infty$ konvergiert. Lassen wir $x \uparrow 1$ gehen, so folgt (79).

Nun gelte (77). $(h_n)_n$ sei eine Folge positiver reeller Zahlen mit $x_R - \frac{1}{h_n} = \gamma_n$ für alle $n \geq 2$. Insbesondere folgt $h_n \rightarrow \infty$ für $n \rightarrow \infty$. Für $x \in (0, 1)$ und $n \geq 2$ gilt

$$\delta_n(x) := x_R - \frac{1}{h_n x} < x_R - \frac{1}{h_n} = \gamma_n,$$

daher folgt wie oben

$$n\bar{F}(\delta_n(x)) > 1$$

für alle $x \in (0, 1)$ und $n \geq 2$. Nutzt man zusätzlich (77), so folgt

$$n\bar{F}(\gamma_n) > \frac{n\bar{F}(\gamma_n)}{n\bar{F}(\delta_n(x))} = \frac{\bar{F}(x_R - \frac{1}{h_n})}{\bar{F}(x_R - \frac{1}{h_n x})} = \frac{F^*(h_n)}{F^*(h_n x)} \rightarrow x^\alpha \quad \forall x \in (0, 1).$$

Mit $x \uparrow 1$ folgt wieder (79).

Schließlich gelte (78). Für $x < 0$ und $n \geq 2$ sei

$$\delta_n(x) := \gamma_n + xg(\gamma_n).$$

Dann gilt $\delta_n(x) < \gamma_n$ für alle $x < 0$ und $n \geq 2$. Daher folgt wie oben mit $n\bar{F}(\delta_n(x)) > 1$
 $x_n \rightarrow x_R$ und somit

$$n\bar{F}(\gamma_n) > \frac{n\bar{F}(\gamma_n)}{n\bar{F}(\delta_n(x))} = \frac{\bar{F}(\gamma_n)}{\bar{F}(\gamma_n + xg(\gamma_n))} \xrightarrow{n \rightarrow \infty} e^x \quad \forall x < 0.$$

Schicken wir x gegen 0, so folgt (79). □

Lemma 5.27 Für $0 \leq \tau \leq +\infty$, eine Verteilungsfunktion F und eine Folge reeller Zahlen (u_n) sind äquivalent:

$$(i) \quad n(1 - F(u_n)) \xrightarrow{n \rightarrow \infty} \tau.$$

$$(ii) \quad \mathbb{P}(M_n \leq u_n) = F^n(u_n) \xrightarrow{n \rightarrow \infty} e^{-\tau}.$$

Beweis: Sei zunächst $\tau < +\infty$. Es gelte 5.27 (i). Dann folgt

$$F^n(u_n) = \left[1 - \frac{1}{n} (n(1 - F(u_n))) \right]^n \rightarrow e^{-\tau}.$$

Nun gelte 5.27 (ii). Als erstes zeigen wir, dass dann

$$1 - F(u_n) \xrightarrow{n \rightarrow \infty} 0 \tag{80}$$

konvergiert, denn sonst gäbe es eine Teilfolge (u_{n_k}) von (u_n) und ein $\beta \in (0, 1)$ mit $1 - F(u_{n_k}) \geq \beta$ für alle k . Dann gälte aber

$$F^{n_k}(u_{n_k}) \leq (1 - \beta)^{n_k} \xrightarrow{k \rightarrow \infty} 0,$$

was einen Widerspruch zu $\tau < +\infty$ bildet. Durch logarithmieren folgt aus 5.27 (ii)

$$n \log(1 - (1 - F(u_n))) = n \log F(u_n) \rightarrow -\tau,$$

was wegen $\tau < +\infty$ möglich ist. Entwickeln wir den Logarithmus um 1, so erhalten wir

$$\log(1 - h) = -h + O(h^2), \quad h \downarrow 0$$

und daher mit (80)

$$n(1 - F(u_n)) \xrightarrow{n \rightarrow \infty} \tau.$$

Nun sei $\tau = +\infty$.

(i) \Rightarrow (ii) zeigen wir mit Widerspruch. Angenommen, es gäbe eine Verteilungsfunktion F und eine Folge reeller Zahlen $(u_n)_n$, für die 5.27 (i) aber nicht 5.27 (ii) gilt. Nun

ist $(F_n(u_n))_n \subseteq [0, 1]$ beschränkt und wenn 5.27(ii) nicht gilt, so besitzt $(F_n(u_n))$ eine konvergente Teilfolge, deren Limes wir $e^{-\mu}$ nennen wollen ($\mu \in [0, \infty)$). Mit dem oben Gezeigten folgt dann

$$\lim_{k \rightarrow \infty} n_k(1 - F(u_{n_k})) = \mu < +\infty$$

im Widerspruch zu 5.27(i).

Schließlich betrachten wir (ii) \Rightarrow (i). Wieder nehmen wir an, dass eine Verteilungsfunktion F und eine Folge $(u_n)_n$ existieren, so dass

$$\lim_{n \rightarrow \infty} F^n(u_n) = 0 \quad \text{und} \quad \liminf_{n \rightarrow \infty} n(1 - F(u_n)) = \mu < +\infty$$

gilt. Dann gibt es eine Teilfolge $(u_{n_k})_k$ mit

$$\lim_{k \rightarrow \infty} n_k(1 - F(u_{n_k})) = \mu.$$

Da $\mu < +\infty$ ist, folgt (im Widerspruch zur Voraussetzung) aus dem oben Gezeigten

$$\lim_{h \rightarrow \infty} F^{n_k}(u_{n_k}) = e^{-\mu} > 0.$$

□

Wir widmen uns nun dem Anziehungsbereich der Fréchet-Verteilungen.

Satz 5.28 *Ist die Tailfunktion \bar{F} einer Verteilungsfunktion F in ∞ regulär variierend mit Index $-\alpha$ und gilt $x_R = +\infty$, so gilt $F \in \mathcal{D}(\Phi_\alpha)$. Man kann $A_n = \frac{1}{\gamma_n}$ und $B_n = 0 \quad \forall n \geq 2$ wählen.*

Beweis: Sei $A_n = \frac{1}{\gamma_n}$ und $a_n = \frac{1}{A_n} = \gamma_n$. Wir zeigen, dass für alle x

$$n(1 - F(a_n x)) \xrightarrow{n \rightarrow \infty} \begin{cases} x^{-\alpha} & x > 0 \\ \infty & x \leq 0 \end{cases}, \quad (81)$$

denn dann folgt mit Lemma 5.27

$$\lim_{n \rightarrow \infty} F^n(a_n x) = e^{-x^{-\alpha}} \mathbb{1}_{(0, \infty)}(x) = \Phi_\alpha(x)$$

für alle x , also die Behauptung. Sei zunächst $x < 0$. Dann folgt aus $x_R = +\infty$, $a_n = \gamma_n \rightarrow \infty$ ($n \rightarrow \infty$). Da $x < 0$ ist, ergibt sich $a_n x \rightarrow -\infty$, also $1 - F(a_n x) \rightarrow 1$, d. h. (81) für $x < 0$. Sei $x = 0$. Dann ist $(1 - F(a_n x))$ konstant und ungleich 0. Damit folgt wieder (81). Schließlich sei $x > 0$. Dann gilt

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x)) = \lim_{n \rightarrow \infty} n(1 - F(a_n)) \frac{(1 - F(a_n x))}{1 - F(a_n)} = x^{-\alpha} \lim_{n \rightarrow \infty} n(1 - F(a_n)),$$

wobei wir im letzten Schritt wieder die reguläre Variation von \bar{F} und $\gamma_n \rightarrow \infty$ ausgenutzt haben. Da mit Lemma 5.26

$$\lim_{n \rightarrow \infty} n(1 - F(a_n)) = 1$$

gilt, folgt auch für $x > 0$ (81). Also gilt die Behauptung. \square

Nun wollen wir die Notwendigkeit (75) überprüfen.

Satz 5.29 *Gilt $F \in \mathcal{D}(\Phi_\alpha)$ für ein α , so ist die Tailfunktion \bar{F} regulär variierend mit Index $-\alpha$ und es gilt $x_R = +\infty$.*

Wir schieben zwei Lemmata voran, die aber in dieser Vorlesung nicht bewiesen werden sollen. Ihre Beweise finden sich in dem Buch von L. de Haan “On regular variation and its applications”.

Lemma 5.30 *Die Tailfunktion \bar{F} einer Verteilungsfunktion F ist in ∞ regulär variierend, falls für jedes $\varepsilon > 0$ Folgen $(\lambda_n(\varepsilon))_{n \in \mathbb{N}}$ und $(a_n(\varepsilon))_{n \in \mathbb{N}}$ existieren, für die folgende Aussagen gelten:*

1. $\lim_{n \rightarrow \infty} \frac{\lambda_n(\varepsilon)}{\lambda_{n+1}(\varepsilon)} > 1 - \varepsilon$.
2. $\lim_{n \rightarrow \infty} a_n(\varepsilon) = +\infty$.
3. $\lim_{n \rightarrow \infty} \lambda_n(\varepsilon) \bar{F}(a_n(\varepsilon)x) =: \lambda(x)$ existiert und ist für alle $x > 0$ positiv und endlich.

Lemma 5.31 $(a_n), (b_n)$ seien Folgen reeller Zahlen mit

- $a_n > 0 \quad \forall n$;
- $\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = \gamma > 1$;
- $\lim_{n \rightarrow \infty} \frac{b_{n+1} - b_n}{a_n} = 0$.

Dann gilt:

$$\lim_{n \rightarrow \infty} \frac{b_n}{a_n} = 0.$$

Nun sind alle Hilfsmittel bereit gestellt, um Satz 5.29 zu beweisen:

Beweis von Satz 5.29: Nach Voraussetzung gilt für geeignete $a_n > 0, b_n \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \Phi_\alpha(x) \quad \forall x \in \mathbb{R}. \quad (82)$$

Wir zeigen zunächst:

$$\frac{b_{[ns]} - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{} 0 \quad \forall s > 1 \quad \text{und} \quad \frac{a_{[ns]}}{a_n} \xrightarrow[n \rightarrow \infty]{} s^{1/\alpha} \quad \forall s > 1. \quad (83)$$

Wegen (82) gilt für alle $s > 1$

$$\lim_{n \rightarrow \infty} F^{\lceil ns \rceil}(a_{\lceil ns \rceil}x + b_{\lceil ns \rceil}) = \Phi_\alpha(x) \quad \forall x \in \mathbb{R}.$$

Damit folgt wie im Beweis von Lemma 5.19

$$\lim_{n \rightarrow \infty} F^n(a_{\lceil ns \rceil}x + b_{\lceil ns \rceil}) = \Phi_\alpha^{1/s} \quad \forall x \in \mathbb{R}.$$

Der Satz von Khinchin liefert

$$\frac{a_{\lceil ns \rceil}}{a_n} \xrightarrow{n \rightarrow \infty} A_s > 0 \quad \text{und} \quad \frac{b_{\lceil ns \rceil} - b_n}{a_n} \rightarrow B_s \in \mathbb{R}$$

und $\Phi_\alpha^{1/s}(x) = \Phi_\alpha(A_s x + B_s)$ für alle $x \in \mathbb{R}$. Es folgt $A_s = s^{1/\alpha}$ und $B_s = 0$, also (83). Durch

$$n(1) := \left\lceil \frac{s}{s-1} \right\rceil \quad \text{und} \quad n(k+1) := \lceil n(k) \cdot s \rceil, \quad k \geq 1$$

ist eine Folge $(n(k))_k$ definiert. Für diese gilt

$$n(k) \xrightarrow{k \rightarrow \infty} \infty \quad \text{und} \quad \frac{n(k+1)}{n(k)} \xrightarrow{k \rightarrow \infty} s. \quad (84)$$

Mit dem ersten Grenzübergang und (83) folgt

$$\lim_{k \rightarrow \infty} \frac{a_{n(k+1)}}{a_{n(k)}} = \lim_{k \rightarrow \infty} \frac{a_{\lceil n(k) \cdot s \rceil}}{a_{n(k)}} = A_s = s^{1/\alpha} > 1$$

und

$$\lim_{k \rightarrow \infty} \frac{b_{n(k+1)} - b_{n(k)}}{a_{n(k)}} = \lim_{k \rightarrow \infty} \frac{b_{\lceil n(k) \cdot s \rceil} - b_{\lceil n(k) \rceil}}{a_{n(k)}} = B_s = 0.$$

Mit Lemma 5.31 folgt aus den letzten beiden Gleichungen

$$\frac{b_{n(k)}}{a_{n(k)}} \xrightarrow{k \rightarrow \infty} 0. \quad (85)$$

Aus dem Satz von Khinchin folgt nun aber

$$\lim_{k \rightarrow \infty} F^{n(k)}(a_{n(k)}x) = \Phi_\alpha(x) \quad \forall x \in \mathbb{R}, \quad (86)$$

da für $k \rightarrow \infty$ trivialerweise $\frac{a_{n(k)}}{a_{n(k)}} = 1$ mit (85) auch

$$\frac{b_{n(k)} - 0}{a_{n(k)}} \rightarrow 0$$

und mit (82) wegen $n(k) \rightarrow \infty$ auch

$$\lim_{k \rightarrow \infty} F^{n(k)}(a_{n(k)}x + b_{n(k)}) = \Phi_\alpha(x) \quad (87)$$

für alle $x \in \mathbb{R}$ gilt. Nun soll gezeigt werden, dass \bar{F} regulär variierend mit Index $-\alpha$ ist. Wegen (86) gilt mit Lemma 5.27

$$n(k)\bar{F}(a_{n(k)}x) \xrightarrow{k \rightarrow \infty} x^{-\alpha} \quad \forall x > 0.$$

Bleibt noch zu zeigen, dass die Voraussetzungen von Lemma 5.30 erfüllt sind:

- (i) Definitionsgemäß folgt wegen $\frac{a_{n(k)+1}}{a_{n(k)}} \rightarrow s^{1/\alpha}$, dass $a_{n(k)} \rightarrow \infty$ für $k \rightarrow \infty$ gilt.
- (ii) Man erinnere sich, dass die Folge $(n(k))$ von s abhängt. Für $s \downarrow 1$ gilt $\frac{1}{s} \uparrow 1$, so dass mit dem zweiten Grenzübergang in (84) für alle $\varepsilon > 0$ ein $s > 1$ und somit eine Folge $(n(k))$ existiert mit

$$\lim_{k \rightarrow \infty} \frac{n(k)}{n(k+1)} = \frac{1}{s} > 1 - \varepsilon.$$

Also folgt mit (87) aus Lemma 5.30, dass \bar{F} regulär variierend mit Index $-\alpha$ ist. Es bleibt zu zeigen, dass $x_R = +\infty$ gilt. Gälte nun $x_R < +\infty$, so gäbe es wegen $a_{n(k)} > 0$ für jedes k und $a_{n(k)} \rightarrow \infty$ für $k \rightarrow \infty$ ein $x > 0$, so dass $a_{n(k)}x > x_R$ gelten würden für alle k . Für dieses x wäre dann

$$1 = \lim_{k \rightarrow \infty} F^{n(k)}(a_{n(k)}x) = \Phi_\alpha(x),$$

was im Widerspruch zu $0 < \Phi_\alpha(x) < 1 \quad \forall x > 0$ stünde. Damit ist die Behauptung bewiesen. \square

Wir wenden uns nun dem Anziehungsbereich der Weibull-Verteilung zu. Wieder zerfällt die Behauptung in zwei Sätze.

Satz 5.32 *Eine Verteilungsfunktion F liegt im Anziehungsbereich der Weibullverteilung ψ_α , falls $x_R < +\infty$ und die Tailfunktion \bar{F}^* der durch $F^*(x) = F(x_R - \frac{1}{x})$ definierten Funktion F^* regulär variierend mit Index $-\alpha$ ist. Eine mögliche Wahl der Konstanten ist*

$$A_n^{-1} = a_n = x_R - \gamma_n \quad \text{und} \quad B_n = b_n = x_R.$$

Beweis: Mit F ist auch F^* eine Verteilungsfunktion, x_R^* bezeichne ihren rechten Randpunkt. Mit Satz 5.28 wissen wir, dass F^* im Anziehungsbereich der Fréchet-Verteilung Φ_α liegt und dass für

$$a_n^* := \gamma_n^* := F^{*-} \left(1 - \frac{1}{n} \right) \quad \text{und} \quad b_n^* := 0$$

$$\lim_{n \rightarrow \infty} F^{*n}(a_n^*x + b_n^*) = \lim_{n \rightarrow \infty} F^{*n}(a_n^*x) = \lim_{n \rightarrow \infty} F^n \left(x_R - \frac{1}{a_n^*x} \right) = \Phi_\alpha(x) \quad \forall x > 0$$

gilt. Mit $a_n := \frac{1}{a_n^*}$ und $b_n := x_R$ gilt daher

$$\lim_{n \rightarrow \infty} F^n \left(b_n - \frac{a_n}{x} \right) = \Phi_\alpha(x) \quad \forall x > 0$$

und somit

$$\lim_{n \rightarrow \infty} F^n(a_nx + b_n) = \Phi_\alpha \left(-\frac{1}{x} \right) = \psi_\alpha(x) \quad \forall x < 0.$$

Da $b_n = x_R$ und $a_n > 0$ gilt, ist $F(a_n x + b_n) \equiv 1$ für $x \geq 0$. Deswegen gilt natürlich

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = 1 \quad \forall x \geq 0.$$

Somit ist gezeigt, dass $F \in \mathcal{D}(\psi_\alpha)$ gilt. Bleibt noch zu zeigen: $a_n = x_R - \gamma_n$.

$$\begin{aligned} a_n^* = \gamma_n^* &= \inf \left\{ x > 0 : F^*(x) \geq 1 - \frac{1}{n} \right\} \\ &= \inf \left\{ x > 0 : F \left(x_R - \frac{1}{x} \right) \geq 1 - \frac{1}{n} \right\} \\ &= \inf \left\{ x > 0 : u = x_R - \frac{1}{x} \wedge F(u) \geq 1 - \frac{1}{n} \right\} \\ &= \inf \left\{ x > 0 : x = (x_R - u)^{-1} \wedge F(u) \geq 1 - \frac{1}{n} \right\} \\ &= \inf \left\{ (x_R - u)^{-1} > 0 : F(u) \geq 1 - \frac{1}{n} \right\} \\ &= \left(\sup \left\{ x_R - u > 0 : F(u) \geq 1 - \frac{1}{n} \right\} \right)^{-1} \\ &= \left(x_R - \inf \left\{ u < x_R : F(u) \geq 1 - \frac{1}{n} \right\} \right)^{-1} \\ &= (x_R - \gamma_n)^{-1}. \end{aligned}$$

Also gilt: $a_n = (a_n^*)^{-1} = x_R - \gamma_n$. □

Wieder bleibt noch zu zeigen, dass die Bedingungen auch notwendig sind, um im Anziehungsbereich der Weibull-Verteilung zu liegen.

Satz 5.33 *Liegt F in $\mathcal{D}(\psi_\alpha)$, so gilt $x_R < +\infty$ und die Teilfunktion \bar{F}^* der durch $F^*(x) := F(x_R - \frac{1}{x})$ für alle $x > 0$ definierten Funktion F^* ist in ∞ regulär variierend mit Index $-\alpha$.*

Wir beginnen auch diesen Beweis mit einem Lemma, für dessen Beweis wir auf das Buch von de Haan verweisen.

Lemma 5.34 *(a_n) und (b_n) seien zwei Folgen in \mathbb{R} mit*

- (i) $a_n > 0$ für alle n ;
- (ii) $\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = \gamma$ mit $0 < \gamma < 1$ und
- (iii) $\lim_{n \rightarrow \infty} \frac{b_{n+1} - b_n}{a_n} = 0$.

Dann existiert $b = \lim_{n \rightarrow \infty} b_n$ und ist endlich und es gilt

$$\lim_{n \rightarrow \infty} \frac{b - b_n}{a_n} = 0.$$

Beweis von Satz 5.33: Der Beweis ist dem Beweis des Fréchet-Falls sehr ähnlich. Es gebe also Folgen $(a_n)_n$ und $(b_n)_n$ mit $a_n > 0$ für alle $n \in \mathbb{N}$ und

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \psi_\alpha(x) \quad \forall x \in \mathbb{R}.$$

Wie im Beweis von Satz 5.29 kann nun gezeigt werden:

$$\lim_{n \rightarrow \infty} \frac{a_{[ns]}}{a_n} = A_s = s^{-1/\alpha} \quad \text{und} \quad \lim_{n \rightarrow \infty} \frac{b_{[ns]} - b_n}{a_n} = B_s = 0 \quad (88)$$

für alle $s > 1$. Wie im Beweis von Satz 5.29 wird dieselbe Folge $(n(k))_k$ in Abhängigkeit von $s > 1$ gewählt. Mit dieser Folge gilt wegen (88)

$$\lim_{k \rightarrow \infty} \frac{a_{n(k+1)}}{a_{n(k)}} = \lim_{k \rightarrow \infty} \frac{a_{[n(k) \cdot s]}}{a_{n(k)}} = s^{-1/\alpha} \quad (89)$$

und

$$\lim_{k \rightarrow \infty} \frac{b_{n(k+1)} - b_{n(k)}}{a_{n(k)}} = \lim_{k \rightarrow \infty} \frac{b_{[n(k) \cdot n]} - b_{n(k)}}{a_{n(k)}} = 0. \quad (90)$$

Mit Lemma 5.34 folgt daraus

$$\lim_{k \rightarrow \infty} b_{n(k)} = b \quad \text{mit} \quad -\infty < b < \infty \quad \text{und} \quad \lim_{k \rightarrow \infty} \frac{b - b_{n(k)}}{a_{n(k)}} = 0.$$

Der Satz von Khinchin liefert wie im Beweis von Satz 5.29

$$\lim_{k \rightarrow \infty} F^{n(k)}(a_{n(k)}x + b) = \psi_\alpha(x) \quad \forall x \in \mathbb{R}. \quad (91)$$

Es soll nun gezeigt werden, dass $x_R < +\infty$ gilt. Aus $\psi_\alpha(0) = 1$ folgt direkt $F(b) = 1$, also $x_R \leq b < +\infty$. Nun zeigen wir noch, dass \bar{F}^* in $+\infty$ regulär variierend mit Index $-\infty$ ist. Da $\psi_\alpha(-1) < 1$ gilt, ist $F(-a_{n(k)} + b) < 1$ für hinreichend große k . Da $(a_{n(k)})$ nach (89) eine Nullfolge ist, kann nicht $b > x_R$ gelten, woraus mit $x_R \leq b$ trivialerweise $x_R = b$ folgt. Damit folgt aus (91) mit $a_{n(k)}^* := \frac{1}{a_{n(k)}}$

$$\begin{aligned} \lim_{k \rightarrow \infty} F^{*n(k)}(a_{n(k)}^* x) &= \lim_{k \rightarrow \infty} F^{n(k)} \left(b - \frac{1}{a_{n(k)}^*} x \right) \\ &= \lim_{k \rightarrow \infty} F^{n(k)} \left(a_{n(k)} \left(-\frac{1}{x} \right) + b \right) \\ &= \psi_\alpha \left(-\frac{1}{x} \right) \quad \text{für alle } x > 0. \end{aligned}$$

Da $\psi_\alpha(-\frac{1}{x}) = \Phi_\alpha(x)$ für alle $x > 0$ gilt, folgt aus der letzten Gleichung

$$\lim_{k \rightarrow \infty} F^{*n(k)}(a_{n(k)}^* x) = \Phi_\alpha(x) \quad \forall x > 0.$$

Im Beweis von Satz 5.29 wurde bereits gezeigt, dass daraus folgt, dass \bar{F}^* regulär variierend mit Parameter $-\alpha$ ist. \square

Wir werden nun den Gumbel-Fall diskutieren.

Satz 5.35 Eine Verteilungsfunktion F liegt in $\mathcal{D}(\Lambda)$, falls eine positive, messbare Funktion g existiert, so dass

$$\lim_{t \uparrow x_R} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}$$

für alle $x \in \mathbb{R}$ gilt. Die Konstanten können dann als $a_n = g(\gamma_n)$ und $b_n = \gamma_n$ gewählt werden.

Beweis: Der Beweis beruht auf Lemma 5.27. Da $\gamma_n \uparrow x_R$ für $n \rightarrow \infty$ gilt, ergibt sich mit Lemma 5.26 und den Voraussetzungen des Satzes

$$\lim_{n \rightarrow \infty} n(1 - F(\gamma_n + xg(\gamma_n))) = \lim_{n \rightarrow \infty} \frac{n(1 - F(\gamma_n + xg(\gamma_n)))}{n(1 - F(\gamma_n))} = e^{-x} \quad \text{für alle } x \in \mathbb{R}.$$

Lemma 5.27 gibt dann die Behauptung des Satzes. \square

Dass die genannten Bedingungen auch notwendig sind, ist wesentlich aufwendiger zu zeigen und soll hier nur skizziert werden.

Satz 5.36 Liegt die Verteilungsfunktion F in $\mathcal{D}(\Lambda)$, so existiert eine positive, messbare Funktion g mit

$$\lim_{t \uparrow x_R} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}$$

für alle $x \in \mathbb{R}$.

Als ersten Beweisschritt benötigt man eine allgemeinere Version des Satzes von Khinchin, die im wesentlichen genauso bewiesen wird wie der Satz von Khinchin selbst.

Lemma 5.37 G sei eine nicht-entartete Verteilungsfunktion, $t_0 \in \mathbb{R} \cup \{\infty\}$ und $(F_t)_{t < t_0}$ eine Familie von monoton nicht-fallenden, rechtsseitig stetigen Funktionen mit

$$\lim_{x \rightarrow \infty} F_t(x) = 1 \quad \text{und} \quad \lim_{x \rightarrow -\infty} F_t(x) = c_t \geq 0$$

für alle $t < t_0$. Außerdem seien $a : \mathbb{R} \rightarrow \mathbb{R}^+$ und $b : \mathbb{R} \rightarrow \mathbb{R}$ Funktionen mit

$$F_t(a(t)x + b(t)) \rightarrow G(x) \quad (t \uparrow t_0)$$

für alle $x \in C(G)$. Dann gilt folgende Äquivalenz: Für Funktionen $\alpha : \mathbb{R} \rightarrow \mathbb{R}^+$ und $\beta : \mathbb{R} \rightarrow \mathbb{R}$ und eine geeignete nicht-entartete Verteilungsfunktion \tilde{G} gilt genau dann

$$F_t(\alpha(t)x + \beta(t)) \rightarrow \tilde{G}(x) \quad (t \uparrow t_0) \quad \forall x \in C(G),$$

wenn es Konstanten $A > 0$ und B mit

$$\frac{\alpha(t)}{a(t)} \rightarrow A \quad \text{und} \quad \frac{\beta(t) - b(t)}{a(t)} \rightarrow B$$

für $t \uparrow t_0$ gibt. Dann gilt

$$\tilde{G}(x) = G(Ax + B) \quad \forall x \in \mathbb{R}.$$

Außerdem benötigen wir die folgenden Lemmata, für deren Beweis wir wieder auf das Buch von de Haan verweisen:

Lemma 5.38 *Für eine Verteilungsfunktion F , die im Anziehungsbereich der Gumbel-Verteilung liegt, gilt*

$$\lim_{t \downarrow 0} \frac{F^-(1-tx) - F^-(1-t)}{F^-(1-ty) - F^-(1-t)} = \frac{\log x}{\log y} \quad \forall x \in \mathbb{R}^+ \quad \forall y \in \mathbb{R}^+ \setminus \{1\}.$$

Lemma 5.39 *Für Verteilungsfunktionen F , die im Anziehungsbereich der Gumbel-Verteilung liegen, gilt*

$$\lim_{s \rightarrow \infty} s(1 - F(a(s)x + b(s))) = e^{-x} \quad \forall x \in \mathbb{R}.$$

Hierbei sei $a : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ durch

$$a(s) := F^- \left(1 - \frac{1}{se} \right) - F^- \left(1 - \frac{1}{s} \right)$$

und $b : \mathbb{R}^+ \rightarrow \mathbb{R}$ durch $b(s) := F^-(1 - \frac{1}{s})$ definiert.

Nun sind wir in der Lage, Satz 5.36 zu beweisen.

Beweis von Satz 5.36: Mithilfe der vorhergehenden drei Lemmata folgern wir die Behauptung des Satzes. Wichtig ist vor allem die erweiterte Version des Satzes von Khinchin. Sei F eine Verteilungsfunktion mit $F \in \mathcal{D}(\Lambda)$. Mit Lemma 5.39 wissen wir, dass

$$\lim_{s \rightarrow \infty} s(1 - F(a(s)x + b(s))) = e^{-x}$$

mit den dort definierten Funktionen $a(\cdot)$ und $b(\cdot)$ gilt. Wir substituieren s durch die Funktion

$$\begin{aligned} s : (-\infty, x_R) &\rightarrow \mathbb{R}^+ \\ t &\mapsto s(t) := \frac{1}{1 - F(t)}. \end{aligned}$$

Da $s(t) \rightarrow \infty$ für $t \uparrow x_R$ gilt, ergibt sich

$$\lim_{t \uparrow x_R} s(t)(1 - F(a(s(t))x + b(s(t)))) = \lim_{t \uparrow x_R} \frac{1 - F(a(s(t))x + b(s(t)))}{1 - F(t)} = e^{-x}$$

für alle $x \in \mathbb{R}$. Daraus folgt direkt

$$\lim_{t \uparrow x_R} \exp \left(- \frac{1 - F(a(s(t))x + b(s(t)))}{1 - F(t)} \right) = e^{-e^{-x}} \quad (92)$$

für alle $x \in \mathbb{R}$. Wir definieren für alle $t < x_R$ die Funktion F_t vermöge

$$F_t(x) := \exp \left(- \frac{1 - F(x)}{1 - F(t)} \right), \quad x \in \mathbb{R}.$$

Für $t < x_R$ ist die Funktion rechtsseitig stetig, monoton nicht-fallend und es gilt

$$\lim_{x \rightarrow \infty} F_t(x) = 1 \quad \text{und} \quad \lim_{x \rightarrow -\infty} F_t(x) = \exp\left(-\frac{1}{1-F(t)}\right) > 0.$$

(92) impliziert für diese Funktion

$$\lim_{t \uparrow x_R} F_t(a(s(t))x + b(s(t))) = e^{-e^{-x}} \quad \forall x \in \mathbb{R}.$$

Es bleibt also zu zeigen, dass

$$\frac{t - b(s(t))}{a(s(t))} \rightarrow 0 \tag{93}$$

gilt, denn dann folgt mit der Erweiterung des Satzes von Khinchin

$$\lim_{t \uparrow x_R} F_t(a(s(t))x + t) = e^{-e^{-x}} \quad \forall x \in \mathbb{R}.$$

Dies ist aber äquivalent zu

$$\lim_{t \uparrow x_R} \frac{1 - F(a(s(t))x + t)}{1 - F(t)} = e^{-x} \quad \forall x \in \mathbb{R},$$

also zur Behauptung mit einer Funktion g , die auf $(-\infty, x_R)$ durch $g(t) := a(s(t))$ definiert ist und auf ganz \mathbb{R} messbar und positiv fortgesetzt werden kann. Also zeigen wir (93). Nach Definition von b und s gilt:

$$b(s(t)) = F^-(F(t)) \leq t < b(s(t)(1 + \varepsilon)),$$

da

$$F(t) = 1 - (1 - F(t)) < 1 - \frac{1 - F(t)}{1 + \varepsilon}$$

$\forall \varepsilon > 0$. Somit folgt

$$0 \leq \frac{t - b(s(t))}{a(s(t))} \leq \frac{b(s(t)(1 + \varepsilon)) - b(s(t))}{a(s(t))}.$$

Nach Lemma 5.38 konvergiert die rechte Seite dieser Ungleichung für $t \uparrow x_R$ gegen $\log(1 + \varepsilon)$, denn:

$$\begin{aligned} \frac{b(s(t)(1 + \varepsilon)) - b(s(t))}{a(s(t))} &= \frac{b((1 - F(t))^{-1}(1 + \varepsilon)) - b((1 - F(t))^{-1})}{a((1 - F(t))^{-1})} \\ &= \frac{F^-(1 - (1 - F(t))(1 + \varepsilon)^{-1}) - F^-(1 - (1 - F(t)))}{F^-(1 - (1 - F(t))e^{-1}) - F^-(1 - (1 - F(t)))} \\ &\xrightarrow[\substack{t \uparrow x_R \\ (1 - F(t)) \downarrow 0}]{\log((1 + \varepsilon)^{-1})} \frac{\log((1 + \varepsilon)^{-1})}{\log(e^{-1})} = \log(1 + \varepsilon). \end{aligned}$$

Also gilt (93) und die Behauptung des Satzes. \square

Literatur

[1] Feller, :

[2] Hardy / Wright: 1959