Wahrscheinlichkeitstheorie

Wintersemester 2003/04

Peter Eichelsbacher

In der Wahrscheinlichkeitstheorie werden wir versuchen, Abläufe/Experimente, die vom Zufall gesteuert werden, zu untersuchen. Gibt es Gesetzmäßigkeiten solcher Abläufe? Was sind geeignete mathematische Modelle für das Studium von Zufallsexperimenten? Formal existiert die Theorie seit 1933; damals hat A. N. Kolmogorov sie mittels der bereits entwickelten Maßtheorie und Integrationstheorie fest in der Analysis verankert. Es gibt Verbindungen zur Zahlentheorie, zur Ergodentheorie sowie zur Theorie partieller Differentialgleichungen und zur Differentialgeometrie.

Wir setzen Grundkenntnisse aus der Maß- und Integrationstheorie voraus, etwa im Umfang meines Analysis III–Skripts. Auf diese Grundkenntnisse gehen wir jeweils durch kurze Wiederholung ein.

Inhaltsverzeichnis

KAPITEL 1

Wahrscheinlichkeitsräume

Beim Studium von Zufallsexperimenten interessieren wir uns für die Beobachtung spezieller „Ereignisse“ und „Zufallsgrößen“. Wir wollen „Wahrscheinlichkeiten“ berechnen, mit denen Ereignisse eintreten, bzw. „Erwartungswerte“ von Zufallsgrößen. In diesem und im folgenden Kapitel wollen wir diese Begriffe definieren und Beispiele betrachten.

Definition 1.1 Es seien Ω eine Menge, $\mathcal{A}$ eine σ-Algebra in Ω und P ein Maß auf $\mathcal{A}$ mit $P(\Omega) = 1$. P heißt *Wahrscheinlichkeitsmaß* (kurz W-Maß), der Maßraum $(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum* (kurz W-Raum). Elemente in $\mathcal{A}$ heißen *Ereignisse*, zu $A \in \mathcal{A}$ heißt $P(A)$ *Wahrscheinlichkeit* von A (oder für das Eintreten des Ereignisses A). Elemente ω von Ω mit $\{\omega\} \in \mathcal{A}$ heißen *Elementarereignisse*.

Wir wollen in diesem Kapitel intensiv $\mathcal{A}$ und P studieren. Zuvor führen wir noch etwas Sprache bzw. Notationen ein: $\emptyset$ bzw. Ω heißen das *unmögliche* bzw. *sichere* Ereignis. Ereignisse E mit $P(E) = 0$ bzw. $P(E) = 1$ heißen *fast unmöglich* bzw. *fast sicher*. Statt P-fast überall (siehe Definition 32.1, Analysis III) sagen wir auch P-*fast sicher* oder *mit Wahrscheinlichkeit Eins*, kurz P-f.s. oder m. W. 1.

Falls $E \subset F$, $E, F \in \mathcal{A}$, sagt man, ein Ereignis E *impliziert* F oder *zieht nach sich*. Gilt $E \cap F = \emptyset$, so nennt man E und F *disjunkt*, *fremd* oder *unvereinbar*. Man nennt $E \cup F$ bzw. $E \cap F$ bzw. $E \backslash F$ „mindestens eines der Ereignisse E und F tritt ein“ bzw. „E und F treten ein“ bzw. „es tritt E, nicht aber F ein“. Für eine Folge $(E_n)_{n \in \mathbb{N}}$ in $\mathcal{A}$ ist $\bigcup_{n=1}^\infty E_n$ bzw. $\bigcap_{n=1}^\infty E_n$ das Ereignis „E_n tritt für gewisse n ein“ bzw. „E_n tritt ein für alle n“. Schließlich setzen wir

$$\liminf_{n\to\infty} E_n =: \{E_n \text{ für schließlich alle } n\} ,$$
$$\limsup_{n\to\infty} E_n =: \{E_n \text{ für unendlich viele } n\}$$

mit

$$\liminf_{n\to\infty} E_n := \bigcup_{n\geq 1} \bigcap_{m=n}^\infty E_m ,$$

$$\limsup_{n\to\infty} E_n := \bigcap_{n\geq 1} \bigcup_{m=n}^\infty E_m .$$

Man schreibt auch $\{E_n \text{ u.o.}\} := \{E_n \text{ für unendlich viele } n\}$, wobei u.o. „unendlich oft“ bedeutet. Und wir lassen häufig $\{\dots\}$ weg:

$$P\{E_n \text{ u.o.}\} = P(\limsup_{n\to\infty} E_n)\ .$$

Nun wollen wir ausführlich zu $\mathcal{A}$ und zu P diskutieren.
Wir erinnern an die Definition einer σ-Algebra:

Definition 1.2 (siehe Analysis III, Kapitel 27) Ein System $\mathcal{A}$ von Teilmengen einer Menge Ω heißt *σ-Algebra*, wenn gilt:

(i) $\Omega \in \mathcal{A}$.
(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.
(iii) Für jede Folge $(A_n)_n$ von Mengen aus $\mathcal{A}$ liegt $\bigcup_{n\ge 1} A_n$ in $\mathcal{A}$.

Beispiele 1.3 **(a)** $\mathcal{P}(\Omega)$ ist eine σ-Algebra.
(b) Sei $\mathcal{E} \subset \mathcal{P}(\Omega)$ ein nicht-leeres Mengensystem. Dann ist

$$\sigma(\mathcal{E}) \quad := \bigcap_{\substack{\mathcal{E}\subset\mathcal{A}\\ \mathcal{A}\ \sigma\text{-Algebra}}} \mathcal{A}$$

eine σ-Algebra, die man die von $\mathcal{E}$ *erzeugte σ-Algebra* nennt, $\mathcal{E}$ heißt Erzeuger.
(c) In $\mathbb{R}^d$ kennen wir $\mathcal{B}^d := \sigma(\mathcal{F}^d)$, die σ-Algebra der Borelschen Mengen. Der Erzeuger $\mathcal{F}^d$ ist das System der d-dimensionalen Figuren, wobei eine d-dimensionale Figur eine endliche Vereinigung von nach rechts halboffenen Intervallen der Form $[a, b[$ mit $a, b \in \mathbb{R}^d$ ist. Wir kennen weitere Erzeuger von $\mathcal{B}^d$: das System aller offenen bzw. abgeschlossenen bzw. kompakten Teilmengen von $\mathbb{R}^d$ (siehe Satz 28.14, Analysis III).

Wir erinnern an die Definition eines Maßes:

Definition 1.4 Eine Abbildung $\mu : \mathcal{A} \to [0, \infty]$ mit $\mu(\emptyset) = 0$ und $\mu(\bigcup_{n\ge 1} A_n) = \sum_{n\ge 1} \mu(A_n)$ (σ-Additivität) für jede disjunkte Folge $(A_n)_n$ in $\mathcal{A}$ heißt *Maß* (auf $\mathcal{A}$). Gilt $\mu(\Omega) = 1$, so heißt μ W-Maß.

Beispiele 1.5 **(a)** Sei $\omega \in \Omega$ und

$$\delta_\omega(A) := \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A, \end{cases}$$

für $A \subset \Omega$. Dann ist $\delta_\omega : \mathcal{P}(\Omega) \to [0, \infty]$ ein W-Maß, das Dirac-*Maß* auf Ω.
(b) Für $A \subset \mathcal{P}(\Omega)$ sei $|A|$ die Anzahl ihrer Elemente, falls A eine endliche Menge ist, $+\infty$ sonst. Dies liefert das *Zählmaß*.
(c) λ^d bezeichne das d-dimensionale Lebesgue-*Maß* auf $\mathbb{R}^d$. Auf $([0, 1], \mathcal{B}_{[0,1]})$ ist dann das induzierte Maß ein W-Maß (Definition 28.12, Kapitel 28).

Im Folgenden führen wir noch den Begriff eines „Dynkin-Systems“ ein:

Definition 1.6 (siehe auch Lemma 35.6, Analysis III und dessen Beweis) Ein *Dynkin-System* $\mathcal{D}$ (über einer Menge Ω) ist ein System von Teilmengen von Ω, welches die folgenden Eigenschaften erfüllt:

(i) $\Omega \in \mathcal{D}$.
(ii) $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$.
(iii) Für jede Folge $(A_n)_n$ *paarweise disjunkter* Mengen aus $\mathcal{D}$ ist $\bigcup_{n\geq 1} A_n$ in $\mathcal{D}$.

Der Grund für die Einführung dieses Begriffs ist, dass (iii) in Definition 1.6 häufig leichter nachweisbar ist als (iii) in Definition 1.2. Es stellt sich die Frage: Wann ist ein Dynkin-System eine σ-Algebra?

Wir hatten im Beweis von Lemma 35.6 in Analysis III den folgenden Satz bereits bewiesen:

Satz 1.7 *Ist ein Dynkin-System durchschnittstabil, so ist es eine σ-Algebra.*

Beweis: $\mathcal{D}$ sei ein durchschnittstabiles Dynkin-System. Wir müssen zeigen, dass $\mathcal{D}$ abgeschlossen gegenüber abzählbaren Vereinigungen ist. Sei $(A_i)_i$ eine Folge in $\mathcal{D}$.

$$\begin{aligned} B_1 &:= A_1 \, , \\ B_n &:= A_n \backslash (A_1 \cup \cdots \cup A_{n-1}) \, ; \quad n \geq 2 \, . \end{aligned}$$

Wir zeigen mit Induktion nach n, dass B_n und $A_1 \cup \cdots \cup A_n$ zu $\mathcal{D}$ gehören. Für $n = 1$ ist nichts zu zeigen. Sei $n \geq 2$. B_n hat die Darstellung $B_n = A_n \cap ((A_1 \cup \cdots \cup A_{n-1})^c)$. Per Induktionsvoraussetzung ist $A_1 \cup \cdots \cup A_{n-1}$ in $\mathcal{D}$, also auch das Komplement. Da $\mathcal{D}$ durchschnittstabil ist, folgt $B_n \in \mathcal{D}$. $A_1 \cup \cdots \cup A_{n-1}$ und B_n sind disjunkt und $A_1 \cup \cdots \cup A_n = (A_1 \cup \cdots \cup A_{n-1}) \cup B_n$. Es gilt $A_1 \cup \cdots \cup A_n \in \mathcal{D}$. Die B_n sind paarweise disjunkt und $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$, also $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$. $\square$

Satz 1.8 *Ist $\mathcal{C}$ ein durchschnittstabiles Mengensystem in Ω, so gilt $d(\mathcal{C}) = \sigma(\mathcal{C})$, wobei $d(\mathcal{C})$ das kleinste von $\mathcal{C}$ erzeugte Dynkin-System bezeichnet.*

Beweis: Es folgt sofort $d(\mathcal{C}) \subset \sigma(\mathcal{C})$, denn jede σ-Algebra ist auch ein Dykin-System. Es bleibt zu zeigen, dass $d(\mathcal{C})$ eine σ-Algebra ist. Dazu zeigen wir mit Satz 1.7, dass $d(\mathcal{C})$ durchschnittstabil ist. Definiere

$$\mathcal{A} := \{A \subset \Omega : A \cap C \in d(\mathcal{C}) \; \forall C \in \mathcal{C}\} \, .$$

Da $\mathcal{C}$ durchschnittstabil ist, folgt $\mathcal{C} \subset \mathcal{A}$. Wir zeigen, dass $\mathcal{A}$ die Dynkin-Eigenschaften hat: (i) ist klar. (ii):

$$\begin{aligned} A \in \mathcal{A} &\Rightarrow A \cap C \in d(\mathcal{C}) \; \forall C \in \mathcal{C} \\ &\Rightarrow A^c \cap C = (C^c \cup (A \cap C))^c \in d(\mathcal{C}) \; \forall C \in \mathcal{C} \\ &\Rightarrow A^c \in \mathcal{A} \quad \text{(beachte: } C^c \text{ und } A \cap C \text{ sind disjunkt).} \end{aligned}$$

(iii): $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, seien paarweise disjunkt. Wegen $A_n \cap C \in d(\mathcal{C}) \; \forall C \in \mathcal{C}$ folgt $\left(\bigcup_{n\geq 1} A_n\right) \cap C \in d(\mathcal{C}) \; \forall C \in \mathcal{C}$, d. h. $\bigcup A_n \in \mathcal{A}$. Also gilt $d(\mathcal{C}) \subset \mathcal{A}$. Wir definieren

$$\bar{\mathcal{A}} := \{A \subset \Omega : A \cap A' \in d(\mathcal{C}) \text{ für alle } A' \in d(\mathcal{C})\} .$$

Nun ist nach dem vorangegangenen Schritt $\mathcal{C} \subset \bar{\mathcal{A}}$. Man zeigt nun analog zu eben, dass $\bar{\mathcal{A}}$ ein Dynkin-System ist. Damit folgt $d(\mathcal{C}) \subset \bar{\mathcal{A}}$, also ist $d(\mathcal{C})$ durchschnittstabil, was zu zeigen war. □

Wir leiten aus dem letzten Satz ein praktisches Verfahren ab, welches man *Dynkin-System-Argument* nennt:

Gegeben sei $(\Omega, \mathcal{A})$, ein Messraum, und eine Aussage $(*)$, deren Gültigkeit für alle $A \in \mathcal{A}$ behauptet wird. Es gebe einen durchschnittstabilen Erzeuger $\mathcal{E}$ von $\mathcal{A}$ derart, dass $(*)$ für alle $A \in \mathcal{E}$ nachweisbar ist. Betrachte dann

$$\mathcal{D} := \{A \in \mathcal{A} : A \text{ genügt der Behauptung } (*)\} .$$

Zeige, dass $\mathcal{D}$ ein Dynkin-System bildet. Dann folgt aus $\mathcal{E} \subset \mathcal{D} \subset \mathcal{A}$ und $d(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}$ nach Satz 1.8 die Inklusionskette

$$\mathcal{A} = d(\mathcal{E}) \subset \mathcal{D} \subset \mathcal{A} ,$$

also $\mathcal{A} = \mathcal{D}$, also ist die Behauptung $(*)$ für alle $A \in \mathcal{A}$ bewiesen! Dieses Argument wird häufig verwendet. Hier eine Anwendung

Satz 1.9 *Stimmen zwei Maße μ und ν, die auf einer σ-Algebra $\mathcal{A}$ definiert sind, auf einem durchschnittsstabilen Erzeuger $\mathcal{C}$ von $\mathcal{A}$ überein, und existiert eine Folge $\Omega_n \in \mathcal{C}$, $n \in \mathbb{N}$, mit $\Omega_n \nearrow \Omega$ und $\mu(\Omega_n) = \nu(\Omega_n) < \infty$, so gilt $\mu = \nu$ auf $\mathcal{A}$.*

Beweis: Sei zunächst $\mu(\Omega) = \nu(\Omega) < \infty$. Wir zeigen:

$$\mathcal{D} := \{A \in \mathcal{A} : \mu(A) = \nu(A)\} \text{ist ein Dynkin-System,}$$

denn dann folgt $\mathcal{A} = \mathcal{D}$. Es gilt $\Omega = \mathcal{D}$. Ist $D \in \mathcal{D}$, so ist $\mu(D^c) = \mu(\Omega) - \mu(D) = \nu(\Omega) - \nu(D) = \nu(D^c)$, also $D^c \in \mathcal{D}$. Für jede Folge $(D_n)_n$ von paarweise disjunkten Mengen aus $\mathcal{D}$ gilt

$$\mu\Big(\bigcup_{n=1}^{\infty} D_n\Big) = \sum_{n=1}^{\infty} \mu(D_n) = \sum_{n=1}^{\infty} \nu(D_n) = \nu\Big(\bigcup_{n=1}^{\infty} D_n\Big) ,$$

also $\bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$.

Für den allgemeinen Fall sei μ_n, ν_n definiert durch

$$\mu_n(A) := \mu(A \cap \Omega_n),\ \nu_n(A) := \nu(A \cap \Omega_n),\ A \in \mathcal{A} .$$

Es gilt $\mu_n = \nu_n$ für alle $n \in \mathbb{N}$. Es folgt für alle $A \in \mathcal{A}$

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

also $\mu = \nu$. □

Nun sammeln wir Rechenregeln: Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum. Für paarweise disjunkte Mengen $A_1, \ldots, A_n \in \mathcal{A}$ gilt

$$P\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n P(A_i)\ ; \qquad \text{setze } A_m = \emptyset \text{ für } m > n\ .$$

Weiter gilt für $A \subset B$, $A, B \in \mathcal{A}$: $P(B \backslash A) = P(B) - P(A)$, insbesondere $P(A^c) = P(\Omega \backslash A) = 1 - P(A)$. Für $A, B \in \mathcal{A}$ gilt

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &\leq P(A) + P(B)\ . \end{aligned} \tag{$*$}$$

Per Induktion folgt: Ist I eine endliche Indexmenge, so gilt

$$\begin{aligned} P\Big(\bigcup_{i\in I} A_i\Big) &= \sum_{\emptyset \neq J \subset I} (-1)^{|J|-1} P\Big(\bigcap_{j\in J} A_j\Big) \\ &\underset{I=\{1,\ldots,n\}}{=} \sum_{k=1}^n (-1)^{k-1} \sum_{1\leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}). \end{aligned}$$

(Siebformel von Poincaré–Sylvester)

Satz 1.10 *Es sei $\mathcal{A}$ eine σ-Algebra und $P : \mathcal{A} \to [0, +\infty]$ mit $P(\Omega) = 1$. Dann sind äquivalent:*

(i) *P ist ein W-Maß.*
(ii) *P ist* additiv*, d.h. $A, B \in \mathcal{A}$, $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$, und* isoton stetig*, d.h. $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, $A_i \subset A_{i+1}\ \forall i \in \mathbb{N} \Rightarrow P(\bigcup_{i\in\mathbb{N}} A_i) = \lim_{i\to\infty} P(A_i)$.*
(iii) *P ist* additiv *und* antiton stetig*, d.h. $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, $A_i \supset A_{i+1}\ \forall i \in \mathbb{N} \Rightarrow P(\bigcap_{i\in\mathbb{N}} A_i) = \lim_{i\to\infty} P(A_i)$.*

Beweis: siehe Satz 27.15, Analysis III.

Korollar 1.11 *$A_i \in \mathcal{A}$, $i \in \mathbb{N}$:*

$$P\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) \leq \sum_{i=1}^\infty P(A_i) \quad (\sigma\text{-}Subadditivität)\ .$$

Beweis: Es gilt

$$P\Big(\bigcup_{i=1}^\infty A_i\Big) = \lim_{n\to\infty} P\Big(\bigcup_{i=1}^n A_i\Big) \leq \lim_{n\to\infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^\infty P(A_i)$$

wobei die erste Gleichheit einfach aus 1.10 folgt und die Ungleichung aus $(*)$.

Lemma 1.12 (von Borel-Cantelli) *Es seien $A_i \in \mathcal{A}$, $i \in \mathbb{N}$. Dann gilt:*

$$\sum_{i=1}^\infty P(A_i) < \infty \;\Rightarrow\; P(\limsup_{n\to\infty} A_n) = P\Big(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m\Big) = 0\ .$$

Beweis: Da $\bigcup_{m\geq n} A_m \downarrow \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m$, folgt

$$P(\limsup_{n\to\infty} A_n) \overset{1.10}{=} \lim_{n\to\infty} P\Big(\bigcup_{m\geq n} A_m\Big) \overset{1.11}{\leq} \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) = 0$$

nach Vorraussetzung. □

Eine ganz wesentliche Aufgabe der Maßtheorie ist die *Konstruktion von Maßen* auf geeigneten σ-Algebren, siehe Kapitel 28, Analysis III. Eine der Probleme dabei ist, dass die Mengen in einer σ-Algebra häufig nicht direkt beschrieben werden können. Doch besitzen σ-Algebren in vielen Fällen handhabbare Erzeugendensysteme, die Ringe oder Algebren sind (zumindest in den für uns interessanten Fällen). Daher versucht man, gewünschte Maße auf einem Erzeuger zu konstruieren. Der Satz von CARATHÉODORY, Satz 28.7, Analysis III, sagt dann aus, dass jedes („σ-endliche“) Maß auf einem Ring/einer Algebra zu genau einem Maß auf der erzeugten σ-Algebra erweitert werden kann. Dabei heißt ein Maß μ auf einer σ-Algebra $\mathcal{A}$ σ-endlich, falls es $(A_n)_n \in \mathcal{A}^{\mathbb{N}}$ mit $\Omega = \bigcup_{n\geq 1} A_n$ und $\mu(A_n) < \infty$, $n \in \mathbb{N}$, gibt.
Natürlich interessiert uns im Rahmen der Wahrscheinlicheitstheorie dieser Begriff nicht so sehr, denn jedes endliche Maß ($\mu(\Omega) < \infty$) ist σ-endlich. Wir wiederholen die Konstruktion von CARATHÉODORY hier nicht, man sollte sie aber einmal gesehen haben. Sie führte uns zum LEBESGUE-*Maß* auf $(\mathbb{R}^d, \mathcal{B}^d)$ und in den Übungen zum HAUSDORFF-*Maß* und zum LEBESGUE-STIELTJES-*Maß*.
Nach dieser theoretisch orientierten Einführung der Bausteine eines W-Raumes wollen wir uns nun vielen Beispielen zuwenden.

Beispiel 1.13 (*Diskrete Maße*) Ω sei eine beliebige Menge und $\{\omega_i\}_{i\in I}$ eine höchstens abzählbare Menge von verschiedenen Punkten in Ω und $a_i \in [0,\infty)$ für alle $i \in I$. Für jede σ-Algebra auf Ω sei

$$\mu = \sum_{i\in I} a_i\, \delta_{\omega_i}$$

definiert durch

$$\mu(A) = \sum_{i\in I} a_i 1_A(\omega_i)\ , \quad A \in \mathcal{A}\ .$$

Dies definiert ein Maß. Ein Maß dieser Gestalt heißt *diskret*. Ein *diskretes W-Maß* auf $(\Omega, \mathcal{P}(\Omega))$ liegt vor, falls $\sum_{i\in I} a_i = 1$ gilt. Der W-Raum ist dann

$$\Big(\Omega, \mathcal{P}(\Omega), \sum_{i\in I} a_i\, \delta_{\omega_i}\Big)\ .$$

Wir werden in den meisten Beispielen jedem $\omega \in \Omega = \{\omega_i, i \in I\}$ ein Gewicht $p(\omega_i)$ mit $\sum_{i\in I} p(\omega_i) = 1$, $p : \Omega \to [0,1]$, zuordnen, und schreiben dann

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

Ist Ω endlich und jedes Ereignis gleichwahrscheinlich, $p(\omega) = \frac{1}{|\Omega|}$, $\omega \in \Omega$, so liegt ein LAPLACE-*Experiment* vor. Hier gilt

$$P(A) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \delta_\omega(A) = \frac{|A|}{|\Omega|}.$$

Dies liefert die LAPLACE-*Verteilung* auf Ω. Der n-malige Wurf eines Würfels wird beschrieben durch

$$\left(\{1, \dots, 6\}^n \, , \, \mathcal{P}(\{1, \dots, 6\}^n) \, , \, \frac{1}{6^n} \sum_{\omega \in \{1, \dots, 6\}^n} \delta_w \right) .$$

Ein Zufallsexperiment mit nur zwei möglichen Ausgängen heißt BERNOULLI-*Experiment*. Das zugehörige W-Maß ist von der Form

$$P = \theta \delta_1 + (1 - \theta) \delta_0$$

für ein $\theta \in [0, 1]$. P heißt BERNOULLI-*Verteilung* mit Parameter θ (Münzwurf). Der Fall $\theta = 1/2$ liefert die Laplace-Verteilung auf $\{0, 1\}$.

Beispiel 1.14 (*Münzwürfe*) Für n Münzwürfe nimmt man

$$\Omega = \big\{ (x_1, \dots, x_n) \,|\, x_i \in \{0, 1\} \big\} \, ,$$

für den ∞-fachen

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

Im einfachen Münzwurf ist $A = \{1\}$ das Ereignis „1 tritt ein“, im n-fachen

$$A = \left\{ (x_1, \dots, x_n) \,\middle|\, \sum_{i=1}^n x_i = k \right\} \quad \text{„genau } k \text{ Einsen“}$$

und beim ∞-fachen

$$A = \left\{ (x_i)_{i \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}} \,\middle|\, \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n x_i = p \right\} \quad \text{„die relative Häufigkeit der 1 ist } p\text{“} \, .$$

Setzt man $\mathcal{A} = \sigma(\mathcal{A}_0)$ mit

$$\begin{aligned} \mathcal{A}_0 := \big\{ B \subset \{0, 1\}^{\mathbb{N}} \,\big|\, & \exists n \in \mathbb{N} \, , \, \exists B_0 \in \mathcal{P}(\{0, 1\}^n) \text{ mit} \\ & B = B_0 \times \{0, 1\} \times \{0, 1\} \times \dots \big\} \end{aligned}$$

und

$$P\Big(\big\{ (x_1, x_2, \dots) \in \{0, 1\}^{\mathbb{N}} \,\big|\, x_1 = \bar{x}_1, \dots, x_n = \bar{x}_n \big\} \Big) := 2^{-n}$$

für $\bar{x}_1, \dots, \bar{x}_n \in \{0, 1\}$ fest, so gilt:

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

und obiges P ist fortsetzbar zu einem W-Maß auf $\mathcal{A} = \sigma(\mathcal{A}_0)$. Das beweisen wir etwas später.

Definition 1.15 Ein W-Maß auf $(\mathbb{R}^d, \mathcal{B})$ für $d \geq 1$ und eine beliebige σ-Algebra $\mathcal{B}$ wird als d-dimensionale *Wahrscheinlichkeitsverteilung* (W-Verteilung) bezeichnet.

Eine wichtige Klasse von W-Maßen auf $(\mathbb{R}, \mathcal{B}) := (\mathbb{R}^1, \mathcal{B}^1)$ sind diejenigen, die über eine „Dichtefunktion" definiert sind.
$f : \mathbb{R} \to \mathbb{R}^+$ heißt *Dichtefunktion* oder auch *W-Dichte*, wenn

$$\int_{\mathbb{R}} f \, d\lambda = 1$$

gilt. $\lambda := \lambda^1$ bezeichnet hierbei das LEBESGUE-Maß auf $(\mathbb{R}, \mathcal{B})$. Dann liefert

$$\mathcal{B} \ni A \mapsto P(A) := \int_A f \, d\lambda$$

ein W-Maß, denn $\int_{\mathbb{R}} f \, d\lambda = 1$ und es gilt für das LEBESGUE-Integral:

$$\int_{\bigcup_{i \in \mathbb{N}} A_i} f \, d\lambda = \sum_{i \in \mathbb{N}} \int_{A_i} f \, d\lambda$$

für $A_i \in \mathcal{B}$, $n \in \mathbb{N}$, paarweise disjunkt (mittels des Satzes von der monotonen Konvergenz, 31.10, Analysis III). Mit $\int f \, d\lambda$ meinen wir das in Kapitel 31 in Analysis III konstruierte Integral in Bezug auf das LEBESGUE-Maß. Die Konvergenzsätze und Rechenregeln für das LEBESGUE-Integral verwenden wir in diesem Kapitel ohne sie im Detail aufzulisten. Wir werden im nächsten Kapitel kurz an das Integral bezüglich eines beliebigen Maßes erinnern.
Wir wollen Dichten allgemeiner einführen und erinnern zunächst an

Definition 1.16 $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ seien zwei Messräume und $T : \Omega \to \Omega'$ eine Abbildung. T heißt $\mathcal{A}/\mathcal{A}'$-*messbar*, wenn $T^{-1}(A') \in \mathcal{A}$ für alle $A' \in \mathcal{A}'$ gilt. Ist $\Omega' = \mathbb{R}^d$, $\mathcal{A}' = \mathcal{B}^d$, so sagt man kurz BOREL-*messbar*. Eine Funktion $f : \Omega \to \bar{\mathbb{R}}$ heißt *numerisch*. f heißt $\mathcal{A}$-messbare numerische Funktion, falls $f^{-1}(-\infty)$, $f^{-1}(\infty)$ und $f^{-1}(O)$ für jede offene Teilmenge O in $\mathbb{R}$ zu $\mathcal{A}$ gehört. Die Menge aller $\mathcal{A}$-messbaren numerischen Funktionen auf Ω bezeichnen wir mit

$$\mathcal{L}_0(\Omega, \mathcal{A}, \bar{\mathbb{R}}) =: \mathcal{L}_0 \, .$$

Wir geben nun eine allgemeine Definition einer Dichte:

Definition 1.17 Es seien $(\Omega, \mathcal{A})$ ein Messraum und μ, ν zwei Maße auf $\mathcal{A}$. Eine $\mathcal{A}/\mathcal{B}$-messbare Funktion $f : \Omega \to \mathbb{R}^+$ heißt *Dichte* von ν bezüglich μ, wenn $\nu(A) = \int_A f \, d\mu$ für alle $A \in \mathcal{A}$ gilt (siehe Satz 36.3 (iii) in Analysis III)[1]. Wir schreiben $\nu = f\mu$ oder auch $f = d\nu / d\mu$.

Lemma 1.18 *Seien μ, ν zwei Maße auf $(\Omega, \mathcal{A})$. Falls eine Dichte von ν bezüglich μ existiert, so ist sie eindeutig bis auf μ-f.ü. Gleichheit.*

[1] Satz 36.3 hat unter anderem zum Inhalt, dass $\int_A f \, d\mu$ ein Maß auf $\mathcal{A}$ definiert

Beweis: f und g seien zwei Dichten. Es sei $A := \{x \in \Omega \,|\, f(x) > g(x)\}$. Dann ist

$$\int 1_A(f-g)\,d\mu = \nu(A) - \nu(A) = 0.$$

Da außerdem $1_A(f-g) \geq 0$, folgt $1_A(f-g) = 0$ μ-f.ü.[2], also $f \leq g$ μ-f.ü. $f \geq g$ μ-f.ü. folgt analog. □

Lemma 1.19 *Es seien μ, ν zwei Maße auf $(\Omega, \mathcal{A})$ und f eine Dichte von ν bezüglich μ. Eine $\mathcal{A}$-messbare numerische Funktion $\varphi : \Omega \to \bar{\mathbb{R}}$ ist genau dann ν-integrierbar, wenn φf μ-integrierbar ist und es gilt dann*

$$\int_\Omega \varphi\,d\nu = \int_\Omega \varphi f\,d\mu\ .$$

Der Beweis dieses Lemmas ist eine Übung (siehe Satz 36.3 (iii), Analysis III).

Diese Übung verwendet das sogenannte *Funktionserweiterungsargument*:

Gegeben sei ein Messraum $(\Omega, \mathcal{A})$ und eine Aussage $(*)$, die für alle $\mathcal{A}$-messbaren numerischen Funktionen behauptet wird. Man betrachte

$$\mathcal{M} := \{f : (\Omega, \mathcal{A}) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}}) \,|\, f \text{ erfüllt } (*)\}$$

und zeige

(i) $\mathcal{M} \supset \mathcal{EF}(\Omega, \mathbb{R}^+)$ (Menge der $\mathcal{A}$-einfachen Funktionen)
(ii) Für jede aufsteigende Folge $(f_n)_{n\geq 1}$ nicht-negativer Funktionen aus $\mathcal{M}$ folgt $\sup_{n\geq 1} f_n \in \mathcal{M}$.

Dann gilt $(*)$ für alle nicht-negativen $\mathcal{A}$-messbaren numerischen Funktionen, denn diese können (siehe Satz 30.12, Analysis III) punktweise durch eine wachsende Folge $\mathcal{A}$-einfacher Funktionen approximiert werden. Kann man außerdem

(iii) $f - g \in \mathcal{M}$ für $f, g \in \mathcal{M}$

zeigen, so gilt wegen $f = f^+ - f^-$ (Definitionen siehe 30.4, Analysis III) die Aussage $(*)$ für alle $\mathcal{A}$-messbaren numerischen Funktionen.

Zu zwei Maßen μ, ν auf $(\Omega, \mathcal{A})$ stellt sich die natürliche Frage, wie man entscheiden kann, ob ν eine Dichte bzgl. μ besitzt. Eine *notwendige* Bedingung ist offenbar, dass jede μ-Nullmenge auch eine ν-Nullmenge ist (denn $\int_N f\,d\mu = 0$ für jede μ-Nullmenge N, siehe Satz 32.3, Analysis III). Wir teilen an dieser Stelle inoffiziell mit:

Satz 1.20 (Satz von Radon-Nikodym) *Es seien μ, ν zwei Maße auf $(\Omega, \mathcal{A})$ und μ sei σ-endlich. Dann sind äquivalent:*

(i) *ν besitzt eine Dichte bezüglich μ,*
(ii) *jede μ-Nullmenge ist eine ν-Nullmenge.*

[2] Siehe Satz 32.3, Analysis III.

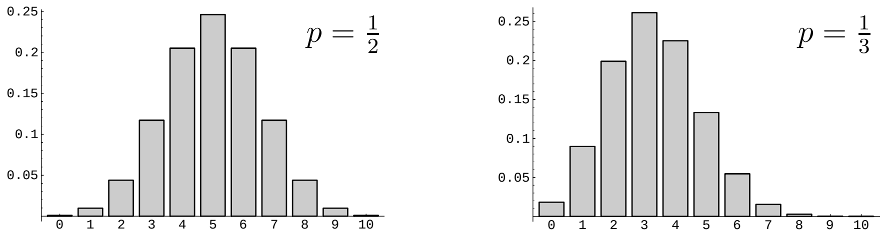


Abbildung 1.1. Histogramme der Binomial-Verteilung für $n = 10$.

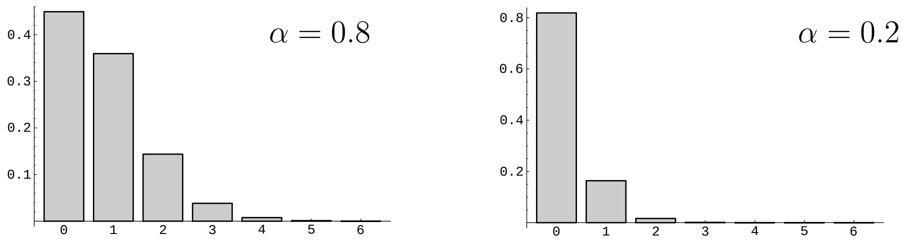


Abbildung 1.2. Histogramme der Poisson-Verteilung.

Der Beweis ist recht lang. Wir liefern ihn an anderer Stelle.

Ein W-Maß P auf $(\mathbb{R}^n, \mathcal{B}^n)$ hat die Dichte $f : \mathbb{R}^n \to \mathbb{R}^+$ bezüglich des n-dimensionalen Lebesgue-Maßes, wenn für $A \in \mathcal{B}^n$ gilt

$$P(A) = \int_A f \, d\lambda^n \, .$$

Diskrete Verteilungen haben offenbar keine Dichten bezüglich des Lebesgue-Maßes, denn $\lambda(\{x\}) = 0$, $x \in \mathbb{R}^n$.

Beispiele 1.21 (Verteilungen auf $\mathbb{R}$) **(a)** Die *Binomialverteilung* zu den Parametern n und p,

$$b(n,p) := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k \; ; \quad 0 \le p \le 1 \; , \; n \in \mathbb{N} \; ,$$

ist eine diskrete Verteilung auf $\mathcal{B}$ (siehe Abb. 1.1), denn

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1 \; .$$

(b) Die Poisson-*Verteilung* zum Parameter $\alpha > 0$,

$$\pi_\alpha := \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} \delta_k \; ,$$

ist ein diskretes W-Maß auf $\mathcal{B}$, $\alpha \in \mathbb{R}^+$ (siehe Abb. 1.2), denn $e^\alpha = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}$.

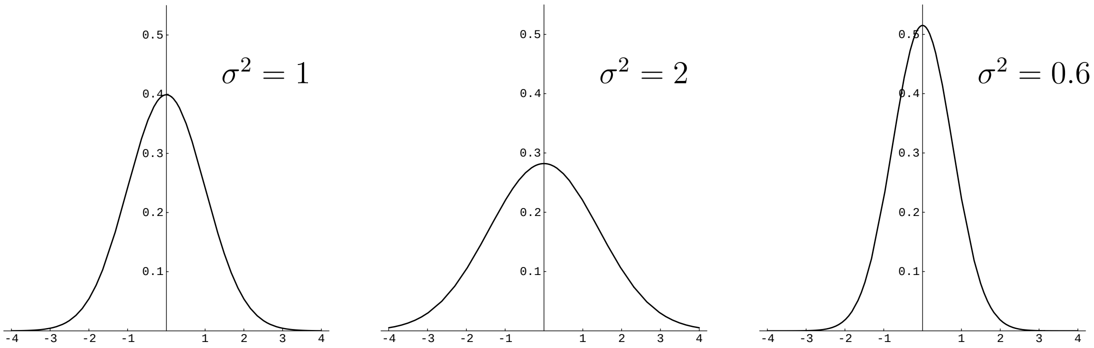


ABBILDUNG 1.3. Verschiedene Gaußdichten mit $a = 0$.

(c) Da

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2}\, dx = 1 \quad \text{(siehe Analysis III)}\ ,$$

folgt durch Substituition, dass

$$g_{a,\sigma^2}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

für jede Wahl von $a \in \mathbb{R}$ und $\sigma > 0$ eine W-Dichte auf $\mathbb{R}$ bezüglich λ^1 ist:

$$N(a, \sigma^2) := g_{a,\sigma^2} \cdot \lambda^1$$

ist ein W-Maß auf $\mathcal{B}$. Man nennt dies die *Normal-* oder *Gauß-Verteilung* auf $\mathbb{R}$ zu den Parametern a und σ^2. $N(0,1)$ heißt *standardisierte Normalverteilung* (siehe Abb. 1.3).

(d) Die Funktion

$$x \mapsto \frac{\alpha}{\pi}(\alpha^2 + x^2)^{-1} =: c_\alpha$$

ist für jedes $\alpha > 0$ eine W-Dichte auf $\mathbb{R}$ (bezüglich λ^1), denn

$$\int_{\mathbb{R}} (1+x^2)^{-1}\, dx = \lim_{n\to\infty} \big[\arctan x\big]_{-n}^{+n} = \pi\ .$$

$\gamma_\alpha := c_\alpha\, \lambda^1$ heißt CAUCHY-*Verteilung* zum Parameter $\alpha > 0$.

Wir lernen weitere diskrete Verteilungen und Dichten kennen.

Vor der Einführung einer wichtigen mehrdimensionalen Verteilung erinnern wir an Bildmaße und die Integration bezüglich eines Bildmaßes.

Definition 1.22 Sei $T : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ eine messbare Abbildung. Dann wird für jedes Maß μ auf $(\Omega, \mathcal{A})$ durch $A' \mapsto \mu(T^{-1}(A'))$ ein Maß μ' auf $\mathcal{A}'$ definiert. Es heißt *Bildmaß* von μ unter T und wird mit $T(\mu)$ bezeichnet. Wir schreiben auch μT^{-1} oder μ^T.

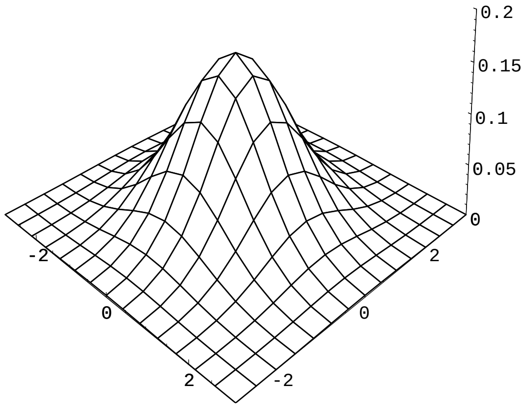


Abbildung 1.4. Zweidimensionale Standardnormalverteilung.

Satz 1.23 *Für $f' \in \mathcal{L}_0(\Omega', \mathcal{A}', \bar{\mathbb{R}}^+)$ gilt*

$$\int_{\Omega'} f'\, dT(\mu) = \int_{\Omega} f' \circ T\, d\mu\ . \tag{$*$}$$

Ist $f' \in \mathcal{L}_0(\Omega', \mathcal{A}', \bar{\mathbb{R}})$, dann ist f' genau dann $T(\mu)$-integrierbar, wenn $f' \circ T$ μ-integrierbar ist und es gilt $(*)$ (Beweis: 29.6, 36.3 (i), (ii), Analysis III).

Definition 1.24

(i) Das W-Maß auf $(\mathbb{R}^n, \mathcal{B}^n)$, das durch die Dichte

$$\varphi(x_1, \dots, x_n) := (2\pi)^{-n/2} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\Big)\ , \quad (x_1, \dots, x_n) \in \mathbb{R}^n\ ,$$

definiert wird, heißt *Standardnormalverteilung* auf $\mathbb{R}^n$ (siehe Abb. 1.4).

(ii) Ein W-Maß P auf $(\mathbb{R}^n, \mathcal{B}^n)$ heißt *Normalverteilung*, wenn eine $n \times n$-Matrix A und $b \in \mathbb{R}^n$ existieren, so dass $P = P_{\mathrm{st}}\phi^{-1}$ ist, wobei $\phi : \mathbb{R}^n \to \mathbb{R}^n$ die affine Abbildung $x \mapsto \phi(x) := Ax + b$ und P_{st} die Standardnormalverteilung sind.

Satz 1.25 *Das W-Maß P der obigen Definition besitzt genau dann eine Dichte, wenn A eine invertierbare Matrix ist. In diesem Fall ist die Dichte gegeben durch*

$$\varphi(x, b, \Sigma) := \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\Big(-\frac{1}{2}(x-b)^t \Sigma^{-1}(x-b)\Big)\ , \quad x \in \mathbb{R}^n\ ,$$

mit $\Sigma = AA^t$. [3]

Beweis: A sei invertierbar, dann ist ϕ invertierbar. Es gilt für $B \in \mathcal{B}^n$

$$\begin{aligned} P(B) &= P_{\mathrm{st}}(\phi^{-1}(B)) \\ &= \int_{\phi^{-1}(B)} \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2}\, \lambda^n(dx) \\ &= \int_{\mathbb{R}^n} 1_B(\phi(x)) \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}|\phi^{-1}\phi(x)|^2\right) \lambda^n(dx) \\ &= \int_{\mathbb{R}^n} 1_B(y) \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}|\phi^{-1}(y)|^2\right) (\lambda^n\phi^{-1})(dy). \end{aligned}$$

Die letzte Gleichheit folgt mittels Satz 1.23 Für bijektive, affine Abbildungen wissen wir[4]:

$$\lambda^n(\phi^{-1}(A)) = |\det \phi^{-1}|\lambda^n(A)\ , \quad A \in \mathcal{B}^n\ .$$

Also ist

$$\frac{d\lambda^n\phi^{-1}}{d\lambda^n}(x) = \det \phi^{-1} = (\det \Sigma)^{-1/2}\ .$$

Mit Lemma 1.19 folgt jetzt

$$\begin{aligned} P(B) &= \int_{\mathbb{R}^n} 1_B(y)(2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}|\phi^{-1}(y)|^2\right) \lambda^n(dy) \\ &= \int_B \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(y-b)^t\Sigma^{-1}(y-b)\right) \lambda^n(dy)\ , \end{aligned}$$

was zu zeigen war.

Ist nun ϕ nicht invertierbar, so ist $\lambda^n\{\phi(x), x \in \mathbb{R}^n\} = 0$, aber $P(\{\phi(x), x \in \mathbb{R}^n\}) = 1$. Also kann P keine Dichte bzgl. λ^n besitzen. □

Ausgangspunkt der Konstruktion des Lebesgue-Maßes λ^1 auf $(\mathbb{R}, \mathcal{B})$ bildete die Setzung $\lambda([a,b)) := b - a$ für nach rechts halboffene Intervalle (siehe 27.6, 27.7, 27.10 (iv) sowie 27.11). Wir wollen nun $b - a$ durch $F(b) - F(a)$ für ein monotones $F : \mathbb{R} \to \mathbb{R}$ ersetzen. Unter welchen Zusatzeigenschaften an F liefert dies ein Maß auf $(\mathbb{R}, \mathcal{B})$?

Satz 1.26 *Eine Funktion $F : \mathbb{R} \to \mathbb{R}$ definiert genau dann durch $\mu_F([a,b)) := F(b) - F(a)$ für alle $a, b \in \mathbb{R}$, $a \leq b$, und eindeutige Fortsetzung ein σ-endliches Maß μ_F auf $(\mathbb{R}, \mathcal{B})$, wenn sie monoton wachsend und linksseitig stetig ist.*

Beweisskizze: μ_F sei ein Maß auf $(\mathbb{R}, \mathcal{B})$, welches $\mu_F([a,b)) = F(b) - F(a)$ erfüllt. Da $\mu_F([a,b)) \geq 0$ für alle $a \leq b$, folgt die Monotonie von F. Weiter

[3] A^t bezeichnet hierbei die Transponierte von A.

[4] Siehe Satz 29.12, Korollar 29.13, Analysis III

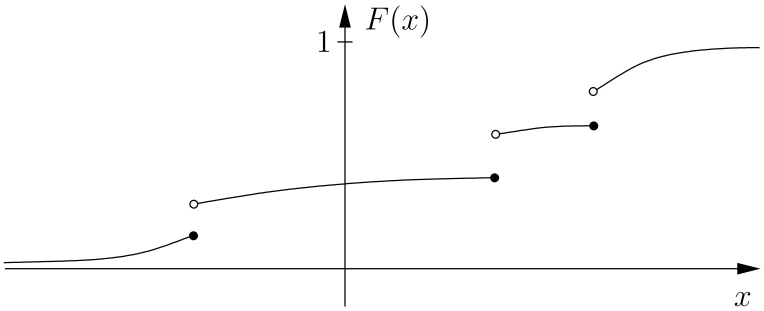


ABBILDUNG 1.5. Eine Verteilungsfunktion.

gilt für alle $a \in \mathbb{R}$ und jede Folge $(a_n)_n$ in $\mathbb{R}$ mit $a_n \uparrow a$, dass $[a_1, a_n) \uparrow [a_1, a)$. Dann folgt mit 1.10

$$\begin{aligned}\lim_{n\to\infty} F(a_n) - F(a_1) &= \lim_{n\to\infty} \mu_F([a_1, a_n)) \\ &= \mu_F([a_1, a)) = F(a) - F(a_1) \ ,\end{aligned}$$

also die linsseitige Stetigkeit in a.

Die Rückrichtung verläuft analog zu der Konstuktion von λ: Zu F existiert genau ein Inhalt μ auf dem Ring $\mathcal{F}$ der 1-dimensionalen Figuren (analog zu 27.10 (iv), Analysis III). Man benötigt die Monotonie von F. Da F linksseitig stetig ist, gibt es zu jedem $[a, b) \in \mathcal{R}$ und zu jedem $\varepsilon > 0$ ein $[a, c) \in \mathcal{R}$ mit $\overline{[a, c)} = [a, c] \subset [a, b]$ und

$$\mu([a, b)) - \mu([a, c)) = \mu([c, b)) = F(b) - F(c) \leq \varepsilon \ .$$

Dann aber folgt wie in Satz 27.11, dass μ ein σ-endliches Prämaß auf $\mathcal{F}$ ist. Dies kann nach dem Satz von CARATHÉODORY zu einem Maß $\tilde{\mu}$ auf $\mathcal{B}$ fortgesetzt werden. □

Bemerkung 1.27 Wir setzen $F(\pm\infty) = \lim_{x\to\pm\infty} F(x)$. Es folgt mit Satz 1.10

$$\mu_F(\mathbb{R}) = F(\infty) - F(-\infty) \ .$$

μ_F bildet also ein endliches Maß, wenn F beschränkt ist. Wir setzen $F(-\infty) = 0$, denn dann gilt

$$\mu_F((-\infty, x)) = F(x) - F(-\infty) = F(x) \ .$$

Definition 1.28 Sei μ eine W-Verteilung auf $(\mathbb{R}, \mathcal{B})$. Die monoton wachsende, linksseitig stetige Funktion

$$F_\mu(x) := \mu((-\infty, x)) \ , \quad x \in \mathbb{R} \ ,$$

mit $F_\mu(-\infty) = 0$ und $F_\mu(\infty) = 1$ heißt *Verteilungsfunktion* von μ.

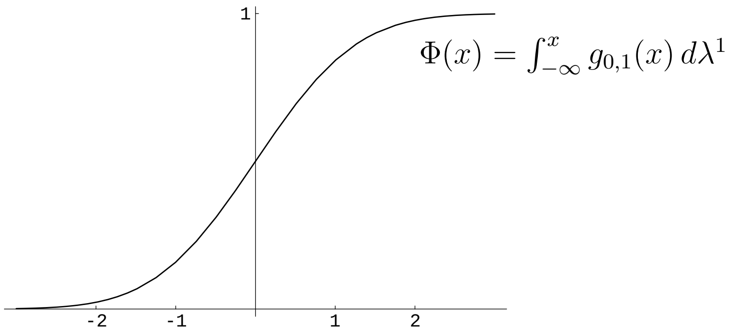


ABBILDUNG 1.6. Verteilungsfunktion Φ der Standardnormalverteilung.

Beispiele 1.29 **(a)** (zu 1.21 (b)):

$$F_{\pi_\alpha} = \begin{cases} e^{-\alpha} \sum_{k=0}^{n} \frac{\alpha^k}{k!} & \text{, falls } n < x \leq n+1 \\ 0 & \text{, falls } x < 0 \,. \end{cases}$$

(b) Sei $f : \mathbb{R} \to \mathbb{R}^+$ eine W-Dichte, so ist

$$P([a,b)) = \int_a^b f \, d\lambda^1$$

eine Verteilung und

$$F(t) := \int_{-\infty}^{t} f \, d\lambda^1$$

ihre Verteilungsfunktion (Stammfunktion der Dichte: $F' = f$, siehe Abb. 1.6). Natürlich ist

$$P(\{a\}) = \lim_{h\to 0} P([a-h,a)) = \lim_{h\to 0} \int_{a-h}^{a} f \, d\lambda^1 = 0 \,.$$

Abschließend liefern wir ein exaktes Modell für den unendlich häufigen Münzwurf einer fairen Münze:

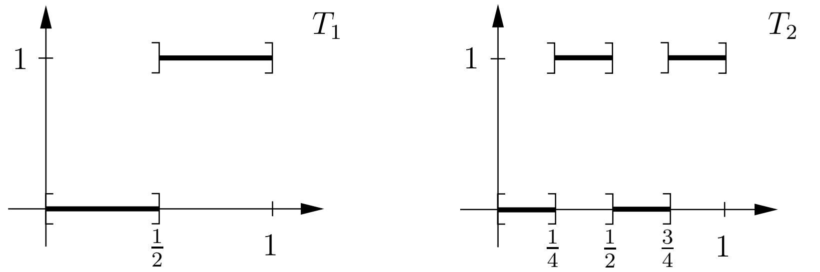
Beispiel 1.30 Es seien $\Omega = [0,1]$, $\mathcal{A} =$ Borel σ-Algebra auf $[0,1]$, und $P =$ Lebesgue-Maß restringiert auf $[0,1]$. Weiter sei

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

Wir setzen $X_i : \tilde{\Omega} \to \{0,1\}$ durch $X_i((x_i)_{i\in\mathbb{N}}) := x_i$, $i \in \mathbb{N}$. Es sei $\tilde{\mathcal{A}} := \sigma(\{X_i = 1\},\ i \in \mathbb{N})$. Die binäre Darstellung von $\omega \in [0,1]$ definiert eine Abbildung

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

mit



... usw.

Es gilt $T_i := X_i \circ T$, $i \in \mathbb{N}$. T ist $\mathcal{A}/\tilde{\mathcal{A}}$-messbar, denn $T^{-1}(\{X_i = 1\}) = \{T_i = 1\}$, und dies ist eine endliche Vereinigung von Intervallen aus $\mathcal{A}$. Sei nun

$$\tilde{P} := P \circ T^{-1} \,,$$

dann ist

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

Dies ist ein Modell für n faire Münzwürfe. Es existiert also ein W-Maß $\tilde{P}$ auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ mit $(*)$, vgl. Beispiel 1.14. Die Eindeutigkeit dieses W-Maßes diskutieren wir hier nicht.

KAPITEL 2

Zufallsvariable und Kenngrößen

Abbildungen können geeignet Information des Urbildraums fokussieren. Beim tausendfachen Münzwurf interessiert zum Beispiel die Anzahl der Einsen.

$$\begin{aligned} X : \{0,1\}^{1000} &\to \{0,\ldots,1000\} \\ (\omega_1,\ldots,\omega_{1000}) &\mapsto \sum_{i=1}^{1000} \omega_i \end{aligned}$$

ist dann beispielsweise eine geeignete Abbildung und uns interessiert $P(X = k)$, $k = 0,\ldots,1000$ (P ist ein W-Maß auf $\Big(\{0,1\}^{1000}, \mathcal{P}(\{0,1\}^{1000})\Big)$).

Definition 2.1 Es sei $(\Omega,\mathcal{A},P)$ ein W-Raum und $(\Omega',\mathcal{A}')$ ein Messraum. Dann heißt eine messbare Abbildung $X : (\Omega,\mathcal{A},P) \to (\Omega',\mathcal{A}')$ *Zufallsvariable*, im Fall $(\Omega',\mathcal{A}') \subset (\bar{\mathbb{R}},\bar{\mathcal{B}})$ *Zufallsgröße* und im Fall $(\Omega',\mathcal{A}') \subset (\bar{\mathbb{R}}^d,\bar{\mathcal{B}}^d)$ für $d \geq 2$ *Zufallsvektor* $(X = (X_1,\ldots,X_d))$. Das Bildmaß von P unter X heißt *Verteilung von X (unter P)* und wird mit $P^X := X(P) = PX^{-1}$ bezeichnet. Besitzt P^X eine Dichte bezüglich eines Maßes μ auf $(\Omega',\mathcal{A}')$, so wird f als *μ-Dichte von X bezeichnet*. Ein Maß μ auf $(\bar{\mathbb{R}},\bar{\mathcal{B}})$ heißt *stetig*, wenn $\mu(\{x\}) = 0$ für alle $x \in \mathbb{R}$. Eine Zufallsgröße $X : (\Omega,\mathcal{A},P) \to (\bar{\mathbb{R}},\bar{\mathcal{B}})$ heißt *diskret* bzw. *stetig*, wenn P^X diese Eigenschaft hat. $F_X(t) := P^X((-\infty,t)), t \in \mathbb{R}$, heißt *Verteilungsfunktion* von X. Für einen d–dimensionalen Zufallsvektor $X = (X_1,\ldots,X_d)$ heißen $P^{(X_i)_{i\in I}}, I \subset \{1,\ldots,d\}, |I| = k$, die zugehörigen k-dimensionalen *Rand-* oder *Marginalverteilungen* und $P^{X_1},\ldots,P^{X_d}$ die eindimensionalen Rand- oder Marginalverteilungen.

Bemerkungen 2.2 (i) Jedes Zufallsexperiment läßt sich mittels einer Zufallsvariablen beschreiben:

$$(\Omega,\mathcal{A},P),\ X \text{ identische Abbildung},\ P^X = P\ .$$

Die genaue Angabe von $(\Omega,\mathcal{A},P)$ tritt in den Hintergrund. Ein Würfelwurf ist zum Beispiel durch irgendeine Zufallsgröße

$$X : (\Omega,\mathcal{A},P) \to (\{1,\ldots,6\},\mathcal{P}(\{1,\ldots,6\}),P^X)$$

mit $P^X = \frac{1}{6}\sum_{i=1}^{6}\delta_i$ beschrieben (fair!).

(ii) In Analysis III hatten wir uns bereits an die abkürzenden Schreibweisen $\{X \leq t\}, \{X = Y\}, \ldots$ für $\{\omega : X(\omega) \leq t\}, \{\omega : X(\omega) = Y(\omega)\}$ gewöhnt (wir wählten allerdings eckige Klammern). Weiter schreiben wir $\{X \in A\}$ für $X^{-1}(A)$ und $P(X \leq t), P(X = Y), P(X \in A)$, wir lassen also hier die Mengenklammern weg.

Wir erinnern an Beobachtungen der Analysis III:

Bemerkungen 2.3 (i) Ist X eine Zufallsgröße auf $(\Omega, \mathcal{A}, P)$ und $h : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ Borel–messbar, so ist auch $h(X)$ eine Zufallsgröße, etwa $|X|, |X|^p, p \in \mathbb{N}, e^X$ usw. Die Klasse der Zufallsgrößen auf $(\Omega, \mathcal{A})$ ist abgeschlossen unter diversen Operationen. Sind zum Beispiel $(X_i)_{i\in\mathbb{N}}$ Zufallsgrößen, so auch $\sum \alpha_i X_i$, $\alpha_i \in \mathbb{R}$, $\sup_i X_i$, $\limsup_i X_i$, $\inf_i X_i$, $\liminf_i X_i$, usw., vergleiche Kapitel 30, Analysis III.

(ii) *Elementare Zufallsgrößen* sind $\mathcal{A}$-einfache Abbildungen (Definition 30.10, Analysis III), also von der Form

$$X = \sum_{j=1}^{m} c_j 1_{A_j} \quad \text{mit } (c_j, A_j) \in \mathbb{R} \times \mathcal{A} \text{ für } j = 1, \ldots, m\ .$$

(iii) $X : \Omega \to \bar{\mathbb{R}}$ sei eine Zufallsgröße. Dann ist $X^+ := X \vee 0$ bzw. $X^- := (-X) \vee 0$ der *Positiv-* bzw. der *Negativteil* von X. $X = X^+ - X^-, |X| = X^+ + X^-$. X^+, X^- sind Zufallsgrößen.

(iv) Für eine Zufallsgröße $X : \Omega \to \mathbb{R}^+$ gibt es eine wachsende Folge $(X_j)_{j\in\mathbb{N}}$ von elementaren Zufallsgrößen mit $\lim_j X_j = X$ (siehe Satz 30.12, Analysis III), nämlich z.B.

$$X_j = \sum_{k=0}^{j2^j-1} k2^{-j} 1_{A_{j,k}}\ , \ j \in \mathbb{N},$$

mit $A_{j,k} := \{k2^{-j} \leq X < (k+1)2^{-j}\}$ (vgl. Abb. 2.1).

Wir betrachten nun die *Integration* von Zufallsgrößen:

Definition 2.4 Es sei X eine Zufallsgröße auf einem W-Raum $(\Omega, \mathcal{A}, P)$. Ist dann $X \geq 0$ oder X P-integrierbar, so heißt

$$\mathbb{E}(X) := \mathbb{E}_P(X) := \int X\, dP \,(= \int_\Omega X\, dP)$$

der *Erwartungswert* von X (bzgl. P).

Wir erinnern uns kurz an die Integrationstheorie aus der Analysis III: Sei X eine Zufallsgröße . Ist $X = 1_A, A \in \mathcal{A}$, so definierte man $\int X\, dP := P(A)$. Ist X eine elementare Zufallsgröße, also $X = \sum_{j=1}^m c_j 1_{A_j}$ mit $(c_j, A_j) \in \mathbb{R} \times \mathcal{A}$, $j = 1, \ldots, m$, so ist

$$\int X\, dP := \sum_{j=1}^{m} c_j P(A_j)$$

(unabhängig von der speziellen Darstellung von X).

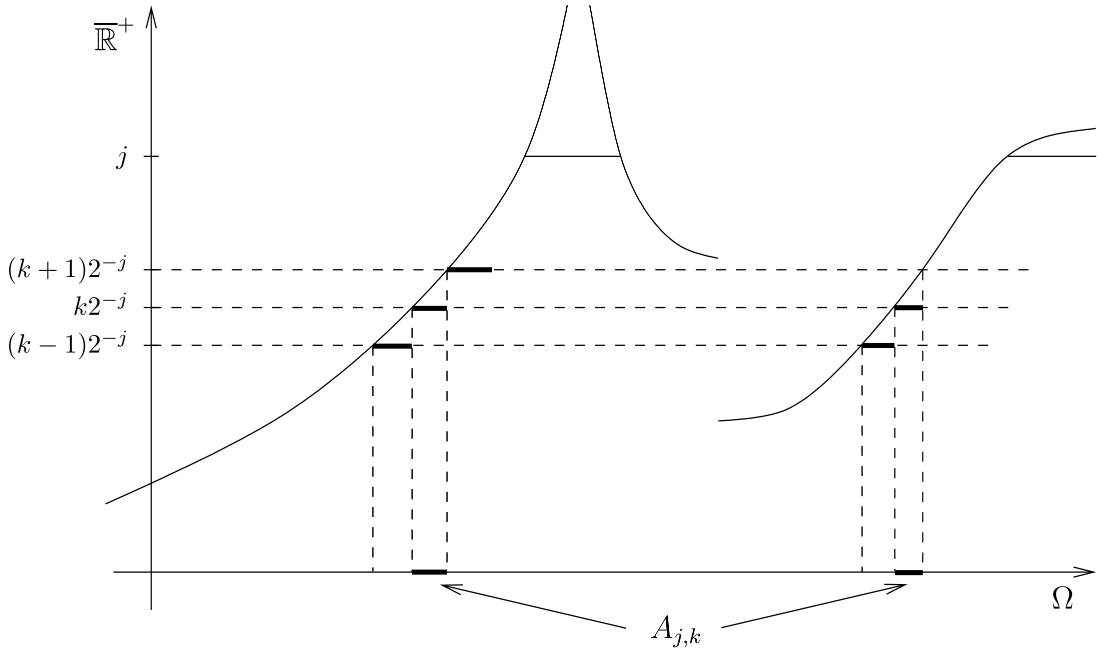


ABBILDUNG 2.1. Monotone Approximation durch elementare Zufallsgrößen.

Ist $X \geq 0$, so existiert eine Folge $(X_n)_n$ elementarer Zufallsgrößen mit $\lim X_n = X$ und

$$\int X\,dP := \lim_{n\to\infty} \int X_n\,dP \ \in [0,\infty]$$

(unabhängig von der speziellen Wahl der X_n).
Für allgemeines X zerlegten wir $X = X^+ - X^-$ und definierten

$$\int X\,dP := \int X^+\,dP - \int X^-\,dP, \quad \text{falls sinnvoll.}$$

Minimalbedingung: $\min(\int X^+\,dP, \int X^-\,dP) < \infty$ (quasi-integrierbar, Analysis III, 31.16)
$\mathcal{L} = \mathcal{L}(\Omega, \mathcal{A}, P)$ bezeichne die Menge der P-integrierbaren Zufallsgrößen ($\mathbb{E}(|X|) < \infty$). Die Verteilung P^X einer reellen Zufallsgröße ist ein W-Maß auf $\mathcal{B}^1$. Es gilt

$$\mathbb{E}(f \circ X) = \int f\,dP^X,$$

anders geschrieben $\mathbb{E}_P(f \circ X) = \mathbb{E}_{P^X}(f)$. Hier ist f als Borel-messbar und nicht-negativ oder als P^X-integrierbar angenommen. Ist also $X \geq 0$ oder X P-integrierbar und wählt man für f die Funktion $x \mapsto x$, so folgt

$$\mathbb{E}(X) = \int x\,dP^X(x).$$

Der Erwartungswert ist nur von der Verteilung von X abhängig! Die Integrierbarkeit von X ist äquivalent zur P^X-Integrierbarkeit von $x \mapsto x$ auf $\mathbb{R}$. Gilt

$X = Y$ P-fast sicher für zwei Zufallsgrößen X, Y definiert auf $(\Omega, \mathcal{A}, P)$, so folgt $\mathbb{E}(X) = \mathbb{E}(Y)$.

Satz 2.5 (i) $X \mapsto \mathbb{E}(X)$ *ist ein positives lineares Funktional auf* $\mathcal{L}^1$.

(ii) *Seien* $(X_n)_n$ *Zufallsgrößen mit* $0 \leq X_1 \leq X_2 \leq \ldots$, *so gilt*

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

(Satz von der monotonen Konvergenz)

(iii) $X_n \geq 0$, $n \in \mathbb{N}$ *:*

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

(iv) $X_n \geq 0$, $n \in \mathbb{N}$ *:*

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n) \ .$$

(Lemma von FATOU*)*

(v) $(X_n)_n$ *seien Zufallsgrößen,* $Y \in \mathcal{L}^1$ *und* $|X_n| \leq Y$ *P-f.s.,* $n \in \mathbb{N}$. *Existiert* $\lim_{n\to\infty} X_n$ *P-f.s., so gilt*

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

(Konvergenzsatz von LEBESGUE*; Satz von der majorisierten Konvergenz)*

(siehe 31.20, 31.10, 31.11, 31.14, 32.12)

Definition 2.6 Sei X eine P-integrierbare Zufallsgröße auf $(\Omega, \mathcal{A}, P)$ und $\mu := \mathbb{E}(X)$, so heißt

$$\operatorname{Var} X := \mathbb{E}(X-\mu)^2 = \int (X-\mu)^2 \, dP = \int (x-\mu)^2 \, dP^X(x)$$

Varianz von X.

$$\sigma(X) := (\operatorname{Var} X)^{1/2} = (\mathbb{E}(X-\mu)^2)^{1/2}$$

heißt *Standardabweichung* von X.
Für $k \in \mathbb{N}$ nennt man $\mathbb{E}X^k$ und $\mathbb{E}(X-\mu)^k$, wenn diese Größen existieren, das *k-te Moment* bzw. das *zentrale k-te Moment*, sowie für $p > 0$ nennt man $\mathbb{E}|X|^p$ und $\mathbb{E}|X-\mu|^p$ das p-te absolute bzw. das zentrale p-te absolute Moment.

Die Existenz des p-ten absoluten Moments von X bedeutet in der Sprache der Integrationstheorie die p-fache Integrierbarkeit von X, $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$. Die Halbnorm $||\cdot||_p$ war definiert durch

$$\begin{aligned} ||X||_p &:= (\mathbb{E}|X|^p)^{1/p}, \quad p \in [1, \infty) \\ ||X||_\infty &:= \inf\{\alpha \geq 0 : P(|X| > \alpha) = 0\} \end{aligned}$$

Satz 2.7 *Eine Zufallsgröße X auf einem W-Raum $(\Omega, \mathcal{A}, P)$ ist genau dann quadratisch integrierbar, wenn X integrierbar und* $\operatorname{Var} X < \infty$ *ist. Es gilt dann:*

$$\begin{aligned} \operatorname{Var} X &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \int x^2 \, dP^X(x) - \Big(\int x \, dP^X(x) \Big)^2 . \end{aligned}$$

Für integrierbares X gilt stets

$$\begin{aligned} \mathbb{E}(X)^2 &\leq \mathbb{E}(X^2) \quad \textit{sowie} \\ \operatorname{Var}(\alpha X + \beta) &= \alpha^2 \operatorname{Var} X, \quad \alpha, \beta \in \mathbb{R}. \end{aligned}$$

X hat genau dann Varianz 0, wenn X P-f.s. konstant ist.

Beweis: Es gilt $\mathcal{L}^2(P) \subset \mathcal{L}^1(P)$ mittels Hölder (siehe unten). Alle konstanten reellen Funktionen sind in $\mathcal{L}^2(P)$! Daher folgt aus $X \in \mathcal{L}^2(P)$ die Integrierbarkeit sowie $\operatorname{Var} X < \infty$. Sei umgekehrt $X \in \mathcal{L}^1(P)$ und $\operatorname{Var} X < \infty$, so liegt $X - \mathbb{E}(X) \in \mathcal{L}^2(P)$, also auch $X = X - \mathbb{E}(X) + \mathbb{E}(X)$. Da $\mathbb{E}$ linear ist, folgt $VarX = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. $\mathbb{E}(X^2) \leq \mathbb{E}(X)^2$ ist dann klar. Es gilt

$$\begin{aligned} \operatorname{Var}(\alpha X + \beta) &= \mathbb{E}((\alpha X + \beta) - (\alpha \mathbb{E} X + \beta))^2 \\ &= \mathbb{E}(\alpha X - \alpha \mathbb{E} X)^2 = \alpha^2 \operatorname{Var} X \ . \end{aligned}$$

Hat X Varianz 0, also $\mathbb{E}(X - \mathbb{E}X)^2 = 0$, so ist dies äquivalent zu $(X - \mathbb{E}X)^2 = 0$ P-f.s., das heißt $X = \mathbb{E}X$ P-f.s. □

Wir kommen zu einer Reihe von wichtigen Ungleichungen, denen wir teilweise bereits begegnet sind:

Satz 2.8 *X, Y seien Zufallsgrößen auf einem W-Raum $(\Omega, \mathcal{A}, P)$. Dann gilt:*

(i) (Markov-Ungleichung) *Für jedes $t > 0$ und jede monoton wachsende Funktion $g : [0, \infty) \to [0, \infty)$ ist*

$$P(|X| \geq t) \leq g(t)^{-1} \int_{\{|X| \geq t\}} g(|X|) \, dP \leq g(t)^{-1} \mathbb{E} g(|X|)$$

(ii) (Tschebyschev-Ungleichung) *Speziell im Fall $g(t) = t^2$ und $\mathbb{E}X = 0$ folgt*

$$P(|X| \geq t) \leq t^{-2} \int_{\{|X| \geq t\}} X^2 \, dP \leq \frac{\operatorname{Var} X}{t^2}$$

(iii) (Hölder-Ungleichung) *Aus $\mathbb{E}|X|^p < \infty$ und $\mathbb{E}|Y|^q < \infty$ für $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$ folgt $\mathbb{E}|XY| < \infty$ und*

$$|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq ||X||_p ||Y||_q$$

(iv) (Cauchy-Schwarz-Ungleichung) *Speziell im Fall $p = q = 2$:*

$$|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq ||X||_2 ||Y||_2$$

(v) (Minkowski-Ungleichung) *Aus $\mathbb{E}|X|^p < \infty$ und $\mathbb{E}|Y|^p < \infty$ für $p \in [1, \infty]$ folgt $\mathbb{E}|X + Y|^p < \infty$ und $||X + Y||_p \leq ||X||_p + ||Y||_p$*

(vi) (JENSENsche Ungleichung) *Es sei $I \subset \mathbb{R}$ ein Intervall, $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit $P(X \in I) = 1$, P-integrierbar und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion, dann ist $\int X\, dP = \mathbb{E}X \in I$, $\varphi \circ X$ ist quasi-integrierbar und es gilt*

$$\varphi(\mathbb{E}X) = \varphi(\int_\Omega X\, dP) \leq \int_\Omega \varphi \circ X\, dP = \mathbb{E}(\varphi(X))$$

Beweis: (i) Benutze $g(t)1_{\{|X|\geq t\}} \leq g(|X|)1_{\{|X|\geq t\}} \leq g(|X|)$ und die Monotonie des Integrals.

(ii) klar

(iii) Satz 33.4, Analysis III

(iv) klar

(v) Satz 33.6, Analysis III

(vi) Die JENSENsche Ungleichung wollen wir ausführlich beweisen. φ heißt konvex, wenn für alle $x, y \in I$ und $\lambda \in [0,1]$ gilt:

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y)\ .$$

In Analysis I hatten wir gesehen, dass dazu äquivalent ist

$$\frac{\varphi(t)-\varphi(x)}{t-x} \leq \frac{\varphi(y)-\varphi(x)}{y-x} \quad \forall x, y, t \in I : x < t < y$$

bzw.

$$\frac{\varphi(t)-\varphi(x)}{t-x} \leq \frac{\varphi(y)-\varphi(t)}{y-t} \quad \forall x, y, t \in I : x < t < y.$$

Daraus folgt, dass φ auf $\mathring{I}$ stetig ist, denn sei $x_0 \in \mathring{I}$, $s, t \in \mathring{I}$ mit $s < x_0 < t$. Ist nun $x_0 < x < t$, so folgt aus obigen Ungleichungen

$$\frac{\varphi(x_0)-\varphi(s)}{x_0-s} \leq \frac{\varphi(x)-\varphi(x_0)}{x-x_0} \leq \frac{\varphi(t)-\varphi(x_0)}{t-x_0}\ .$$

Daraus folgt sofort die rechtsseitige Stetigkeit. Die linksseitige Stetigkeit beweist man analog. Also ist φ ist höchstens in den Randpunkten von I unstetig und somit ist jede konvexe Funktion BOREL-messbar!
Zum Beweis der Ungleichung: wir zeigen zunächst $m := \mathbb{E}(X) \in I$. $a, b \in \bar{\mathbb{R}}$ seien linker bzw. rechter Randpunkt von I. Mit $a \leq X \leq b$ folgt $a \leq \mathbb{E}(X) \leq b$. Ist nun $a \in \mathbb{R}$ und $a \notin I$, so ist $0 < X(\omega) - a$ für alle $\omega \in \Omega$, also $a < m$. Analoges folgt für b, also $m \in I$.
Ist m kein innerer Punkt von I, so ist $m \in \mathbb{R}$ rechter oder linker Randpunkt von I. Also ist $X(\omega) = m$ für P-fast alle $\omega \in \Omega$, also $\varphi(X(\omega)) = \varphi(\mathbb{E}X) = \varphi(m)$ für P-fast alle $\omega \in \Omega$, also $\mathbb{E}(\varphi X) = \varphi(\mathbb{E}X)$.
Es sei nun $m \in \mathring{I}$. Nun konstruieren wir eine Stützgerade an den Graphen von φ im Punkt $(m, \varphi(m))$: Für $s, t \in I$ mit $s < m < t$ ist $\frac{\varphi(m)-\varphi(s)}{m-s} \leq \frac{\varphi(t)-\varphi(m)}{t-m}$, also ist $\alpha := \sup\{\frac{\varphi(m)-\varphi(s)}{m-s},\ s < m,\ s \in I\} < \infty$

und für alle $t \in I$ mit $t > m$ gilt:

$$\varphi(t) \geq \varphi(m) + \alpha(t - m) \qquad (\dagger)$$

Für $t = m$ ist $(\dagger)$ auch richtig und sie gilt nach Definition von α auch für alle $t \in I$ mit $t < m$. Somit gilt $(\dagger)$ für alle $t \in I$. Der Graph von φ auf I verläuft also stets oberhalb der durch $t \mapsto \varphi(m) + \alpha(t - m)$ definierten Stützgeraden. Es folgt

$$\varphi(X(\omega)) \geq \varphi(\mathbb{E}X) + \alpha(X(\omega) - \mathbb{E}X)$$

Integration dieser Ungleichung nach P liefert die JENSENsche Ungleichung.

Korollar 2.9 *Es sei $X \in \mathcal{L}^s(\Omega, \mathcal{A}, P)$. Dann ist $X \in \mathcal{L}^r$ für $1 \leq r \leq s$. (Wir hatten dies für $s = 2$ und $r = 1$ verwendet.) Ist X P-f.s. beschränkt, $X \in \mathcal{L}^\infty(\Omega, \mathcal{A}, P)$, so gilt*

$$||X||_r \uparrow ||X||_\infty \quad (r \to \infty).$$

Beweis: Es sei $\varphi(t) = t^{s/r}$, $t \geq 0$. Für $r \in [1, s]$ folgt aus der Jensenschen Ungleichung:

$$(\mathbb{E}|X|^r)^{1/r} = \Big(\varphi(\mathbb{E}|X|^r)\Big)^{1/s} \leq \Big(\mathbb{E}\varphi(|X|^r)\Big)^{1/s} = \mathbb{E}(|X|^s)^{1/s} \, . \qquad (\ddagger)$$

Ist X P-f.s. beschränkt, so ist X r-fach integrierbar für jedes $r \geq 1$. Mit $(\ddagger)$ folgt die Konvergenz von $||X||_r$ gegen einen Limes a. Aus $|X| \leq ||X||_\infty$ P-f.s. (siehe Analysis III) folgt $\|X\|_r \leq ||X||_\infty$ für jedes $r \geq 1$, also $a \leq ||X||_\infty$. Nun zeigen wir noch $a \geq ||X||_\infty$: Für $0 < c < ||X||_\infty$ ist $P(|X| > c) > 0$. Weiter gilt

$$\begin{aligned} (\mathbb{E}|X|^r)^{1/r} &\geq \Big(\int_{\{|X|>c\}} |X|^r \, dP\Big)^{1/r} \\ &\geq cP(|X| > c)^{1/r}. \end{aligned}$$

Also ist $a = \lim_{r\to\infty}(\mathbb{E}|X|^r)^{1/r} \geq c$ für alle $c < ||X||_\infty$, also ist $a \geq ||X||_\infty$. □

Definition 2.10 Ist $X = (X_1, \ldots, X_n)$ ein Zufallsvektor, so definiert man den Erwartungswert komponentenweise durch $\mathbb{E}(X) = (\mathbb{E}X_1, \ldots, \mathbb{E}X_n) \in \mathbb{R}^n$.

Definition 2.11 Sind X und Y aus $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ mit $X \cdot Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, so ist ihre *Kovarianz* $\mathrm{Cov}(X, Y)$ definiert durch

$$\begin{aligned} \mathrm{Cov}(X, Y) &:= \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \, . \end{aligned}$$

Ist $X = (X_1, \ldots, X_n)$ ein Zufallsvektor mit $X_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, $i = 1, \ldots, n$, und $X_iX_j \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ für alle $i, j \in \{1, \ldots, n\}$, so ist die *Kovarianzmatrix* $\Sigma(X) = (\sigma_{ij}(X))$ definiert durch $\sigma_{ij}(X) = \mathrm{Cov}(X_i, X_j)$.

Offenbar ist $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ für eine eindimensionale Zufallsgröße X. Ist X ein Zufallsvektor, als Spaltenvektor geschrieben, so ist

$$\Sigma(X) = \mathbb{E}((X - \mathbb{E}X)(X - \mathbb{E}X)^T).$$

Satz 2.12 (i) *Sind $X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, so existiert* $\mathrm{Cov}(X, Y)$ *und*

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

(ii) *Für $X_1, \ldots, X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ gilt*

$$\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j).$$

(iii) *Die Kovarianzmatrix ist symmetrisch und positiv semidefinit.*

Beweis: (i) Folgt aus Cauchy-Schwarz und einfachem Nachrechnen,

(ii) analog

(iii) Für $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ gilt

$$\begin{aligned} 0 &\le \mathbb{E}\Big(\sum_{i=1}^n \lambda_i (X_i - \mathbb{E}X_i)\Big)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \ \mathrm{Cov}(X_i, X_j)\ . \end{aligned}$$

$\square$

Definition 2.13 Zwei quadratisch integrierbare Zufallsgrößen X und Y heißen *unkorreliert*, wenn ihre Kovarianz verschwindet, d.h.

$$\mathrm{Cov}(X, Y) = 0.$$

Für eine Menge $X_1, \ldots, X_n$ von Zufallsgrößen mit endlicher Varianz gilt

$$\mathrm{Var}(\sum_{j=1}^n X_j) = \sum_{j=1}^n \mathrm{Var}(X_j),$$

wenn $X_1, \ldots, X_n$ *paarweise unkorreliert* sind.

Satz 2.14 $X, Y, Z \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$*;* $\alpha, \beta \in \mathbb{R}$. *Es gilt*

(i) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$

(ii) $\mathrm{Cov}(\alpha X + \beta Y, Z) = \alpha\,\mathrm{Cov}(X, Z) + \beta\,\mathrm{Cov}(Y, Z)$

(iii) $\mathrm{Cov}(X, \alpha Y + \beta Z) = \alpha\,\mathrm{Cov}(X, Y) + \beta\,\mathrm{Cov}(X, Z)$

(iv) $|\,\mathrm{Cov}(X, Y)| \le (\mathrm{Var}(X))^{1/2} (\mathrm{Var}(Y))^{1/2}$

Insgesamt ist also $\mathrm{Cov}(\cdot, \cdot)$ *eine symmetrische Bilinearform auf* $\mathcal{L}^2$.

Beweis: Nachrechnen und bei (iv) CAUCHY-SCHWARZ. $\square$

Satz 2.15 *Sei $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ eine Zufallsgröße. Dann gilt*

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \min_{a\in\mathbb{R}} \mathbb{E}(X-a)^2$$

Beweis: Für $a \in \mathbb{R}$ ist $\mathbb{E}(X-a)^2 = \mathbb{E}(X-\mathbb{E}X)^2 + 2\mathbb{E}(X-a)\mathbb{E}(X-\mathbb{E}X) + (\mathbb{E}X-a)^2$. Da $\mathbb{E}(X-\mathbb{E}X) = 0$, wird $\mathbb{E}(X-a)^2$ für $a = \mathbb{E}(X)$ minimiert.

Wir wollen uns speziellen Situationen und Beispielen zu den neuen Begriffen in diesem Kapitel widmen:

Beispiel 2.16 Sei $X \geq 0$ eine Zufallsgröße auf $(\Omega, \mathcal{A}, P)$ und $X(\Omega)$ abzählbar. Dann ist

$$\mathbb{E}(X) = \mathbb{E}\Big(\sum_{x\in X(\Omega)} x 1_{\{X=x\}}\Big) = \sum_{x\in X(\Omega)} xP(X=x)\ .$$

Falls nicht notwendig $X \geq 0$, jedoch X quasi-integrierbar ist, so gilt

$$\mathbb{E}(X) = \sum_{\substack{x\in X(\Omega)\\ x\geq 0}} xP(X=x) - \sum_{\substack{x\in X(\Omega)\\ x<0}} (-x)P(X=x).$$

Ist insbesondere Ω abzählbar und $X \geq 0$, so gilt $X = \sum_{\omega\in\Omega} X(\omega)1_{\{\omega\}}$, also

$$\begin{aligned}\mathbb{E}(X) &= \sum_{\omega\in\Omega} X(\omega)\mathbb{E}(1_{\{\omega\}}) = \sum_{\omega\in\Omega} X(\omega)P(\{\omega\})\\ &= \sum_{\omega\in\Omega} X(\omega)p(\omega), \quad \text{mit } \ p(\omega) = P(\{\omega\}).\end{aligned}$$

Beispiel 2.17 (*Fairer Münzwurf*) $\tilde{\Omega} = \{0,1\}^{\mathbb{N}}$, $\tilde{\mathcal{A}}$, $\tilde{P}$, siehe Beispiel 1.29.

(a) Sei $X_j\Big((x_i)_i\Big) := x_j$ für $(x_i)_i \in \tilde{\Omega}$ und $j \in \mathbb{N}$. Dann ist

$$\mathbb{E}(X_j) = 1 \cdot \tilde{P}(X_j = 1) + 0 \cdot \tilde{P}(X_j = 0) = 1/2\ .$$

(b) Sei $S_n := X_1 + \cdots + X_n$ (= Anzahl der Erfolge in n Würfen) mit $(X_j)_{j=1,\ldots,n}$ wie in (i), dann ist für $k \in \{0, \ldots, n\}$

$$P(S_n = k) = \sum_{\substack{(x_1,\ldots,x_n)\in\{0,1\}^n\\ \sum_{j=1}^n x_j = k}} P(X_1 = x_1, \ldots, X_n = x_n) = \binom{n}{k} 2^{-n}$$

und $\mathbb{E}(S_n) = \sum_{j=1}^n \mathbb{E}(X_j) = n \cdot \frac{1}{2}$.

(c) Sei $T : \tilde{\Omega} \to \mathbb{N} \cup \{\infty\}$ die Wartezeit auf die erste Eins, d.h. $T(\omega) := \min\{n \in \mathbb{N} | X_n(\omega) = 1\}$, dann ist

$$P(T = k) = P(X_1 = X_2 = \cdots = X_{k-1} = 0, X_k = 1) = 2^{-k}\ ,\ k \in \mathbb{N},$$

und $\mathbb{E}(T) = \sum_{k=1}^\infty kP(T=k) = \sum_{k=1}^\infty k2^{-k} = 2$.

(d) Wir starten mit einem Einsatz von einer Einheit und verdoppeln den Einsatz bis 1 auftritt. Der Einsatz in der n-ten Runde ist dann $X_n = 2^{n-1}1_{\{T>n-1\}}$ mit T wie in (iii). Daher ist

$$\mathbb{E}(X_n) = 2^{n-1}P(\{T > n-1\}) = 2^{n-1}(\frac{1}{2})^{n-1} = 1 \ .$$

Aber $\lim_{n\to\infty} X_n = 0$ P-fast sicher (für alle $\omega \neq (0,0,0,\dots)$).

Beispiel 2.18 (*geometrische Wahrscheinlichkeit*) Konstruktion einer Gleichverteilung auf S^1 (Sphäre)
Problem: $S^1 = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 = 1\}$ als Teilmenge von $\mathbb{R}^2$ hat Lebesgue-Maß 0! Die Abbildung $\Gamma : [0,2\pi) \to S^1$, $\phi \mapsto (\cos\phi, \sin\phi)$ ist bijektiv und $\mathcal{B}^1_{[0,2\pi)}/\mathcal{B}^1_{S^1}$-messbar. Γ überführt jedes Teilintervall I von $[0,2\pi)$ in ein Kreisrandsegment $K_I := \{(\cos\phi, \sin\phi), \phi \in I\}$ derselben Länge (zeigen wir gleich). Sei nun Φ eine auf $[0,2\pi)$ gleichverteilte Zufallsgröße. (Allgemein: Ist Ω eine Borel-messbare Teilmenge des $\mathbb{R}^d$ mit $0 < \lambda^d(\Omega) < \infty$, dann heißt das W-Maß $P : \mathcal{B}^d_\Omega \to [0,1]$, definiert durch $P(A) = \frac{\lambda^d(A)}{\lambda^d(\Omega)}$ *Gleichverteilung auf* Ω.) Dann besitzt $\Gamma \circ \Phi$ eine Verteilung Q mit

$$Q(K_I) = P(\Gamma \circ \Phi \in K_I) = \frac{\lambda(I)}{2\pi} = \frac{\text{Länge von } K_I}{2\pi}$$

für alle Intervalle $I \subset [0,2\pi)$. Q heißt Gleichverteilung auf S^1.
Dies sieht man so: Die Länge $L(K)$ einer differenzierbaren Kurve $K : I \to \mathbb{R}^2$ berechnet sich zu $L(K) = \int_a^b ||K'(t)||\,dt$, wobei $||\cdot||$ die euklidische Norm ist und a, b die Randpunkte von I (siehe Beispiel 42.14 (i), Analysis III). Für $K_I = \Gamma(I)$ liefert das speziell:

$$L(K_I) = \int_a^b ||\Gamma'(\phi)||\,d\phi = b - a = \lambda(I),$$

denn $\Gamma'(\phi) = (-\sin(\phi), \cos(\phi))$ und somit $||\Gamma'(\phi)|| = 1$ für alle $\phi \in [0,2\pi)$. Da $P(\Gamma \circ \Phi \in K_I) = P(\Phi \in I) = \frac{\lambda(I)}{2\pi}$, folgt die Behauptung.

Beispiele 2.19 **(a)** *Binomialverteilung* (siehe 1.20,a)
Es sei X nach $b(n,p)$ verteilt (also $X = X_1 + \cdots + X_n$ mit X_i, so dass $P(X_i = 1) = p$ und $P(X_i = 0) = 1-p$, $0 \le p \le 1$, wissen wir aber noch gar nicht!)

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = np(p+(1-p))^{n-1} = np \\
\mathbb{E}(X^2) &= \sum_{k=1}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} = np\sum_{k=1}^{n} k\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} \\
&= np[(n-1)p+1] = np(np+(1-p)),
\end{aligned}$$

also $\mathrm{Var}(X) = np(1-p)$.
Aber: Bitte, diese Herleitung nicht merken.

Sei $X' = 2X - n$. Die Verteilung von X' ist das Bild der Verteilung von X unter der Abbildung $x \mapsto 2x - n$. Also besitzt X' die Verteilung

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_{2k-n} =: b_s(n,p) \ .$$

X' hat die Werte $-n, -n+2, \ldots, n-2, n$ und heißt daher *symmetrische* oder *symmetrisierte Binomialverteilung*. Speziell ist $b_s(1,p) = (1-p)\delta_{-1} + p\delta_1$.

(b) POISSON-*Verteilung* (siehe 1.20,b) X sei π_α-verteilt, dann ist

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} k = \alpha \quad \text{und} \\ \mathbb{E}(X^2) &= \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} k^2 = \sum_{k=1}^{n} e^{-\alpha} \frac{\alpha^k}{(k-1)!}(k-1+1) = \alpha^2 + \alpha, \end{aligned}$$

also $\operatorname{Var}(X) = \alpha$.

(c) *Normalverteilung* (siehe 1.20,c) : Sei $\nu_{\alpha,\sigma^2} := g_{\alpha,\sigma^2} \cdot \lambda^1$, so gilt $T(\nu_{0,1}) = \nu_{\alpha,\sigma^2}$ für $T(x) = \sigma x + \alpha$, $\alpha \in \mathbb{R}$, $\sigma > 0$. Jede $N(0,1)$–verteilte Zufallsvariable X ist p–fach integrierbar für jedes $p \geq 0$, denn aus $\frac{t^k}{k!} < e^t$ für $t > 0$, $k = 0, 1, 2, \ldots$ folgt

$$|x|^p e^{-x^2/2} < 2^k k! |x|^{p-2k} \quad (x \neq 0) \ .$$

Nun ist $x \mapsto |x|^\alpha$ für $\alpha < -1$ über $\mathbb{R} \setminus [-1,1]$ λ^1–integrierbar und somit ist $x \mapsto |x|^p \, g_{0,1}(x)$ offenbar λ^1–integrierbar über $\mathbb{R}$ für jedes $p \geq 0$, indem man k hinreichend groß wählt. Also existieren

$$M_n := \mathbb{E}(X^n) = \int_{-\infty}^{\infty} x^n g_{0,1}(x) \, dx$$

für ganzzahlige $n \geq 0$. Natürlich ist $M_{2k-1} = 0$, $k \in \mathbb{N}$, da hier der Integrand ungerade ist. Also $\mathbb{E}(X) = 0$ für $N(0,1)$–verteiltes X und $\mathbb{E}(X) = \alpha$ für $N(\alpha, \sigma^2)$–verteiltes X. Für gerades $n \geq 2$ gilt

$$\int_{-\infty}^{\infty} x^n g_{0,1}(x) \, dx = -x^{n-1} g_{0,1}(x) \,\Big|_{-\infty}^{+\infty} + (n-1) \int_{-\infty}^{\infty} x^{n-2} g_{0,1}(x) \, dx \ ,$$

denn $g_{0,1}$ ist Stammfunktion von $x \mapsto -x g_{0,1}(x)$ und alle Integranden sind stetig, so dass diese Integrale auch als absolut konvergente Riemann–Integrale existieren. Wir können also partielle Integration anwenden. Es folgt

$$M_{2k} = (2k-1) M_{2k-2} \ , \ \ k \in \mathbb{N} \ .$$

Mit $M_0 = 1$ folgt $M_{2k} = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2k-1)$, $k \in \mathbb{N}$. Insbesondere ist $\mathbb{E}(X^2) = 1$ für $N(0,1)$–verteiltes X und $\mathbb{E}(X^2) = \alpha^2 + \sigma^2$ für $N(\alpha, \sigma^2)$–verteiltes X. Somit ist $\operatorname{Var}(X) = \sigma^2$. Weiter folgt für $N(\alpha, \sigma^2)$–verteiltes X:

$$\mathbb{E}(X^n) = \sum_{k=0}^{n} \binom{n}{k} \sigma^k \alpha^{n-k} M_k \ .$$

Dazu verwende $T(x) = \sigma x + \alpha$ und

$$\begin{aligned}\mathbb{E}(X^n) &= \int_{\mathbb{R}} x^n \, d\nu_{\alpha,\sigma^2}(x) = \int_{\mathbb{R}} x^n \, dT(\nu_{0,1})(x) \\ &= \int_{\mathbb{R}} (\sigma x + \alpha)^n \, d\nu_{0,1}(x) \ .\end{aligned}$$

(d) CAUCHY-*Verteilung* γ_α (siehe 1.20, d): X sei Cauchy–verteilt mit Parameter $\alpha > 0$. Der Erwartungswert existiert nicht, denn aus $\int_0^n \frac{x}{1+x^2} \, dx = \frac{1}{2} \log(1+n^2)$ folgt

$$\mathbb{E}(X^+) = \mathbb{E}(X^-) = \int_0^\infty \frac{x}{1+x^2} \, dx = +\infty \ .$$

X ist nicht einmal quasi–integrierbar.

(e) *Mehrdimensionale Normalverteilung* auf $\mathbb{R}^n$: X sei standardnormalverteilt auf $\mathbb{R}^n$ (siehe 1.23), dann ist

$$\mathbb{E}(X_i) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i \exp(-\frac{1}{2} \sum_{k=1}^n x_k^2) \, d\lambda^n(x) = 0 \ .$$

Für $i, j \in \{1, \ldots, n\}$ mit $i \neq j$ gilt

$$\mathbb{E}(X_i X_j) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i x_j \exp(-\frac{1}{2} \sum_{k=1}^n x_k^2) \, d\lambda^n(x) = 0$$

und weiter gilt

$$\mathbb{E}(X_i^2) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} x_i^2 \exp(-\frac{1}{2} \sum_{k=1}^n x_k^2) \, d\lambda^n(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x_i^2 e^{-x_i^2/2} \, d\lambda(x_i) = 1 \ .$$

Also ist $\Sigma(X)$ die Einheitsmatrix.

Sei X ein n–dimensionaler Zufallsvektor mit Kovarianzmatrix $\Sigma(X)$ und Erwartungswert $a \in \mathbb{R}^n$. Weiter sei A eine $m \times n$–Matrix, $b \in \mathbb{R}^m$ und $Y := AX + b$, dann ist $\mathbb{E}(Y) = Aa + b$ und

$$\begin{aligned}\Sigma(Y) &= \mathbb{E}\Big((Y - \mathbb{E}(Y))(Y - \mathbb{E}(Y))^t\Big) \\ &= \mathbb{E}\Big(A(X-a)(X-a)^t A^t\Big) = A\Sigma(X)A^t \ .\end{aligned}$$

Also ist für die allgemeine Normalverteilung in 1.23,ii) die Kovarianzmatrix gleich AA^t und der Vektor der Erwartungswerte gleich b.

Wir schließen dieses Kapitel mit einer nützlichen Abschätzung.

Satz 2.20 *Sei X eine $N(0, \sigma^2)$–verteilte Zufallsgröße, dann gilt für alle $\eta > 0$:*

$$(2\pi)^{-\frac{1}{2}} \frac{\sigma\eta}{\sigma^2 + \eta^2} e^{-\frac{\eta^2}{2\sigma^2}} < P(X \geq \eta) < (2\pi)^{-\frac{1}{2}} \frac{\sigma}{\eta} e^{-\frac{\eta^2}{2\sigma^2}} \ .$$

Beweis: Wir zeigen den Fall $\sigma = 1$. Der allgemeine Fall folgt, indem man X durch σX ersetzt, und somit η durch η/σ. Es gilt

$$P(X \geq \eta) = P^X([\eta, \infty)) = \nu_{0,1}([\eta, \infty)) = (2\pi)^{-1/2} \int_\eta^\infty e^{-x^2/2}\, dx$$

und

$$\int_\eta^\infty e^{-x^2/2}\, dx < \int_\eta^\infty \frac{x}{\eta} e^{-x^2/2}\, dx = \frac{1}{\eta} \int_\eta^\infty x e^{-x^2/2}\, dx = \frac{1}{\eta} e^{-\eta^2/2} \ .$$

Mittels partieller Integration folgt andererseits

$$\int_\eta^\infty x^{-2} e^{-x^2/2}\, dx = \frac{1}{\eta} e^{-\eta^2/2} - \int_\eta^\infty e^{-x^2/2}\, dx \ ,$$

also

$$\frac{1}{\eta} e^{-\eta^2/2} = \int_\eta^\infty (1 + x^{-2}) e^{-x^2/2}\, dx < \int_\eta^\infty (1 + \eta^{-2}) e^{-x^2/2}\, dx \ ,$$

und dazu äquivalent

$$\frac{\eta}{1 + \eta^2} e^{-\eta^2/2} < \int_\eta^\infty e^{-x^2/2}\, dx \ .$$

□

Den nächsten Satz werden wir in den folgenden Kapiteln noch deutlich verschärfen.

Satz 2.21 *[Schwaches Gesetz der großen Zahlen] Für jedes $n \in N$ seien paarweise unkorrelierte Zufallsgrößen $X_1, \ldots, X_n$ gegeben, die alle den gleichen Erwartungswert $E \in \mathbb{R}$ und die gleiche Varianz $V < \infty$ besitzen. Sei $\frac{1}{n} S_n := \frac{1}{n}(X_1 + \cdots + X_n)$, dann gilt für jedes $\varepsilon > 0$*

$$\lim_{n \to \infty} P\big(\big|\frac{1}{n} S_n - E\big| > \varepsilon\big) = 0.$$

Beweis: Die Linearität des Erwartungswerts liefert $\mathbb{E}(\frac{1}{n} S_n) = E$ und auf Grund der paarweisen Unkorreliertheit ist

$$\operatorname{Var}\left(\frac{1}{n} S_n\right) = \frac{1}{n^2} \operatorname{Var}(X_1 + \cdots + X_n) = \frac{1}{n^2} n \operatorname{Var}(X_1) = \frac{1}{n} V \ .$$

Also folgt mit der Tschebyschev-Ungleichung

$$P\left(\left|\frac{1}{n} S_n - E\right| > \varepsilon\right) \leq \frac{\operatorname{Var}(\frac{1}{n} S_n)}{\varepsilon^2} \ ,$$

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

Man spricht in Satz 2.21 auch von *Konvergenz in Wahrscheinlichkeit* oder *stochastischer Konvergenz*. Dies betrachten wir später genauer.

KAPITEL 3

Produkträume

In der Wahrscheinlichkeitstheorie möchte man endlich viele oder unendlich viele zufällige Experimente behandeln. Allgemein spricht man von *gekoppelten Zufallsexperimenten*. Wir stellen in diesem Kapitel die maßtheoretische Grundlage zur mathematischen Beschreibung gekoppelter Experimente bereit. Nach der Erinnerung an endliche Produkträume und Produktmaße konstruieren wir Produkt-Wahrscheinlichkeitsmaße auf unendlichen Produkträumen.

Gegeben seien fortan endlich viele Maßräume $(\Omega_j,\, \mathcal{A}_j,\, \mu_j)$, $j = 1, \dots, d$. Es seien

$$\Omega := \Omega_1 \times \cdots \times \Omega_d = \prod_{j=1}^{d} \Omega_j$$

und $p_j : \Omega \to \Omega_j$, $\omega = (\omega_1, \dots, \omega_d) \mapsto \omega_j$, $1 \le j \le d$, die *Projektions-Abbildungen*.

Für zwei Maßräume hatten wir die Produkt-σ-Algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ definiert als die von

$$\{A_1 \times A_2 :\, A_1 \in \mathcal{A}_1,\, A_2 \in \mathcal{A}_2\}$$

erzeugte σ-Algebra. Es ergab sich z.B. $\mathcal{B}^p \otimes \mathcal{B}^q = \mathcal{B}^{p+q}$, $p, q \in \mathbb{N}$ (siehe Definition 35.1 und Beispiel 35.2, Analysis III). Damit ist die Definition einer Produkt-σ-Algebra auf Ω zwar klar, nur geben wir hier aber eine zweite Definition:

Definition 3.1 Es sei Ω eine Menge. Ist $(\Omega_i,\, \mathcal{A}_i)_{i \in I}$ eine Familie messbarer Räume und $T_i : \Omega \to \Omega_i$ für jedes $i \in I$ eine Abbildung von Ω nach Ω_i, so heißt

$$\sigma\big((T_i)_{i \in I}\big) := \sigma\Big(\bigcup_{i \in I} T_i^{-1}(\mathcal{A}_i)\Big)$$

die von $(T_i)_{i \in I}$ (und $(\Omega_i,\, \mathcal{A}_i)_{i \in I}$) *erzeugte σ-Algebra*. Dabei ist

$$T_i^{-1}(\mathcal{A}_i) := \{T_i^{-1}(A_i),\, A_i \in \mathcal{A}_i\}\,.$$

Dies ist die kleinste σ-Algebra $\mathcal{A}$, so dass jedes T_i $\mathcal{A}/\mathcal{A}_i$-messbar ist. Für $I = \{1, \dots, n\}$ bzw. $I = \mathbb{N}$ schreiben wir $\sigma(T_1, \dots, T_n)$ bzw. $\sigma(T_1, T_2, \dots)$.

Definition 3.2 Die von den Projektionen $p_1, \dots, p_d$ auf $\Omega := \Omega_1 \times \cdots \times \Omega_d$ erzeugte σ-Algebra

$$\bigotimes_{j=1}^{d} \mathcal{A}_j := \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d := \sigma(p_1, \dots, p_d)$$

heißt das *Produkt* der σ-Algebra $\mathcal{A}_1, \dots, \mathcal{A}_d$ oder auch die *Produkt-σ-Algebra* von $\mathcal{A}_1, \dots, \mathcal{A}_d$.

Der folgende Satz schlägt die Brücke zu unserer alten Definition aus Analysis III:

Satz 3.3 *Für jedes $1 \le j \le d$ sei $\mathcal{E}_j$ ein Erzeuger der σ-Algebra $\mathcal{A}_j$ über Ω_j und $\mathcal{E}_j$ enthalte eine Folge $(E_{jk})_k$ mit $E_{jk} \uparrow \Omega_j$. Dann wird $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$ von dem System $\mathcal{E}_1 \times \cdots \times \mathcal{E}_d$ aller Mengen $E_1 \times \cdots \times E_d$ mit $E_j \in \mathcal{E}_j$ erzeugt:*

$$\bigotimes_{j=1}^{d} \mathcal{A}_j = \sigma(\mathcal{E}_1 \times \cdots \times \mathcal{E}_d)\ .$$

Insbesondere gilt $\bigotimes_{j=1}^{d} \mathcal{A}_j = \sigma(\mathcal{A}_1 \times \cdots \times \mathcal{A}_d)$.

Beweis: Es sei $\mathcal{A}$ eine σ-Algebra über Ω. Zu zeigen ist: Jede Projektion p_j ist genau dann $\mathcal{A}/\mathcal{A}_j$-messbar, wenn $E_1 \times \cdots \times E_d \in \mathcal{A}$ ist für alle $E_j \in \mathcal{E}_j$. Nach Satz 29.3 aus Analysis III ist p_j genau dann $\mathcal{A}/\mathcal{A}_j$-messbar, wenn aus $E_j \in \mathcal{E}_j$ stets $p_j^{-1}(E_j) \in \mathcal{A}$ folgt. Dann liegt auch $E_1 \times \cdots \times E_d = p_1^{-1}(E_1) \cap \cdots \cap p_d^{-1}(E_d)$ in $\mathcal{A}$.
Gilt umgekehrt $E_1 \times \cdots \times E_d \in \mathcal{A}$ für alle $E_j \in \mathcal{E}_j$, so liegen für jedes gegebene $j = 1, \dots, d$ und $E_j \in \mathcal{E}_j$ die Mengen

$$F_k := E_{1k} \times \cdots \times E_{j-1,k} \times E_j \times E_{j+1,k} \times \cdots \times E_{dk}\ , \quad k \ge 1,$$

in $\mathcal{A}$, und die Folge $(F_k)_k$ steigt gegen $\Omega_1 \times \cdots \times \Omega_{j-1} \times E_j \times \Omega_{j+1} \times \cdots \times \Omega_d = p_j^{-1}(E_j)$ auf, also ist $p_j^{-1}(E_j) \in \mathcal{A}$. $\square$

Satz 3.4 *Bezeichnen $(\Omega_j, \mathcal{A}_j)$, $0 \le j \le d$, messbare Räume, so gilt: Eine Abbildung $f : \Omega_0 \to \Omega_1 \times \cdots \times \Omega_d$ ist genau dann $\mathcal{A}_0/\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$-messbar, wenn jede der Abbildungen $p_j \circ f$, $1 \le j \le d$, $\mathcal{A}_0/\mathcal{A}_j$-messbar ist.*

Beweis: Da die Komposition messbarer Abbildungen messbar ist (Satz 29.5, Analysis III), folgt die eine Richtung sofort. Für die Rückrichtung nutzen wir, dass

$$\mathcal{E} = \bigcup_{j=1}^{d} p_j^{-1}(\mathcal{A}_j)$$

ein Erzeuger von $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$ ist. Jede Menge $E \in \mathcal{E}$ ist von der Form $E = p_j^{-1}(A_j)$ mit $A_j \in \mathcal{A}_j$ und $j \in \{1, \dots, d\}$. Somit gilt $f^{-1}(E) = (p_j \circ f)^{-1}(A_j) \in \mathcal{A}_0$. Damit ist nach Satz 29.3 aus Analysis III f $\mathcal{A}_0/\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$-messbar. $\square$

Unter welchen Voraussetzungen läßt sich die Existenz eines Maßes μ auf $(\Omega, \bigotimes_{j=1}^d \mathcal{A}_j)$ zeigen, das der Bedingung

$$\mu(A_1 \times \cdots \times A_d) = \mu_1(A_1) \ldots \mu_d(A_d)$$

für alle $A_j \in \mathcal{A}_j$ genügt? Für $d = 2$ tragen wir die Resultate aus Kapitel 35, Analysis III, zusammen:

Satz 3.5 (i) *Sind μ_1 und μ_2 σ-endlich, so gibt es genau ein Maß $\mu_1 \otimes \mu_2 : \mathcal{A}_1 \otimes \mathcal{A}_2 \to [0, \infty]$ mit $\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$, $A_i \in \mathcal{A}_i$, $i = 1, 2$. Das Maß $\mu_1 \otimes \mu_2$ ist σ-endlich (Satz 35.7) und heißt das* Produktmaß *von μ_1 und μ_2 (Definition 35.8). Wir hatten insbesondere gesehen:*

$$\lambda^p \otimes \lambda^q = \lambda^{p+q}, \quad p, q \in \mathbb{N}\,.$$

(ii) *(Satz von* FUBINI*) Wieder seien μ_1 und μ_2 σ-endlich. Für jede nicht-negative $\mathcal{A}_1 \otimes \mathcal{A}_2$-messbare numerische Funktion f sind die durch*

$$\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2) \quad \textit{bzw.} \quad \omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1) \tag{3.1}$$

auf Ω_1 bzw. Ω_2 definierten nicht-negativen Funktionen $\mathcal{A}_1$-messbar bzw. $\mathcal{A}_2$-messbar und es gilt

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2)\, d\mu_1 \otimes \mu_2(\omega_1, \omega_2) &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2)\,.\end{aligned} \tag{3.2}$$

Ist $f : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ $\mu_1 \otimes \mu_2$-integrierbar, so ist $f(\omega_1, \cdot\,)$ μ_2-integrierbar für μ_1-fast alle $\omega_1 \in \Omega_1$ und $f(\,\cdot\,, \omega_2)$ ist μ_1-integrierbar für μ_2-fast alle $\omega_2 \in \Omega_2$. Somit sind die Funktionen (3.1) *μ_1- bzw. μ_2-f.ü. definiert und μ_1- bzw. μ_2-integrierbar und es gilt* (3.2)*.*

Wir können $(\Omega_1 \times \cdots \times \Omega_{d-1}) \times \Omega_d$ und $\Omega_1 \times \cdots \times \Omega_d$ mittels $((\omega_1, \ldots \omega_{d-1}), \omega_d) \mapsto (\omega_1, \ldots, \omega_d)$ identifizieren und erhalten damit auch

$$(\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_{d-1}) \otimes \mathcal{A}_d = \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d\,. \tag{3.3}$$

Es gilt sogar

$$\Big(\bigotimes_{j=1}^k \mathcal{A}_j \Big) \otimes \Big(\bigotimes_{j=k+1}^d \mathcal{A}_j \Big) = \bigotimes_{j=1}^d \mathcal{A}_j \quad \text{für alle } 1 \le k \le d\,.$$

Mittels (3.3) kann dann die Existenz des Produktmaßes für beliebige $d \ge 2$ per Induktion bewiesen werden.

Es gilt also:

Satz 3.6 *Zu gegebenen σ-endlichen Maßen $\mu_1, \ldots, \mu_d$ auf $(\Omega_1, \mathcal{A}_1)$,..., $(\Omega_d, \mathcal{A}_d)$ gibt es genau ein Maß μ auf $(\prod_{j=1}^d \Omega_j, \bigotimes_{j=1}^d \mathcal{A}_j)$ mit*

$$\mu(A_1 \times \cdots \times A_d) = \mu_1(A_1) \ldots \mu_d(A_d)$$

für alle $A_j \in \mathcal{A}_j$, $j = 1, \ldots, d$. μ ist σ-endlich und heißt Produkt *der Maße $\mu_1, \ldots, \mu_d$ bzw.* Produktmaß *und wird mit*

$$\bigotimes_{j=1}^{d} \mu_j = \mu_1 \otimes \cdots \otimes \mu_d \,.$$

bezeichnet. Es gilt

$$\Big(\bigotimes_{j=1}^{k} \mu_j\Big) \otimes \Big(\bigotimes_{j=k+1}^{d} \mu_j\Big) = \bigotimes_{j=1}^{d} \mu_j \,, \quad 1 \leq k \leq d \,. \tag{3.4}$$

(Assoziativität)

Mit (3.4) und einer Induktion über d kann man den Satz von FUBINI übertragen (Übung!).

Wir betrachten nun den Fall, in dem jedes der Maße μ_j mit einer reellen Dichte $f_j \geq 0$ versehen wird. Dann ist mit μ_j auch $\nu_j := f_j \mu_j$ σ-endlich (dies beweisen wir hier nicht, da unsere Maße immer W-Maße sind) und es gilt:

Satz 3.7 *Für jedes $j = 1, \ldots, d$ seien $(\Omega_j, \mathcal{A}_j, \mu_j)$ σ-endliche Maßräume, $f_j \geq 0$ reelle $\mathcal{A}_j$-messbare numerische Funktionen auf Ω_j und $\nu_j := f_j \mu_j$. Dann ist das Produkt $\nu_1 \otimes \cdots \otimes \nu_d$ definiert und es gilt*

$$\nu_1 \otimes \cdots \otimes \nu_d = (f_1 \otimes \cdots \otimes f_d) \cdot (\mu_1 \otimes \cdots \otimes \mu_d)$$

mit

$$f_1 \otimes \cdots \otimes f_d(\omega_1, \ldots, \omega_d) := f_1(\omega_1) \cdots f_d(\omega_d)$$

*(*Tensorprodukt *von $f_1, \ldots, f_d$).*

Beweis: Wir beweisen hier nur den Fall $d = 2$, der Rest folgt per Induktion. Es seien $A_i \in \mathcal{A}_i$, $i = 1, 2$, dann ist

$$\begin{aligned}
\nu_1(A_1)\nu_2(A_2) &= \Big(\int_{A_1} f_1 \, d\mu_1\Big)\Big(\int_{A_2} f_2 \, d\mu_2\Big) \\
&= \iint 1_{A_1}(\omega_1) f_1(\omega_1) \, 1_{A_2}(\omega_2) f_2(\omega_2) \, \mu_1(d\omega_1)\mu_2(d\omega_2) \\
&= \iint 1_{A_1 \times A_2}(\omega_1, \, \omega_2) \, f_1 \otimes f_2(\omega_1, \, \omega_2) \, \mu_1(d\omega_1) \, \mu_2(d\omega_2) \\
&\overset{\text{FUBINI}}{=} \int_{A_1 \times A_2} f_1 \otimes f_2 \, d(\mu_1 \otimes \mu_2).
\end{aligned}$$

Es folgt mit Satz 3.5(i)

$$\nu_1 \otimes \nu_2 = (f_1 \otimes f_2)\,(\mu_1 \otimes \mu_2)\,. \qquad \square$$

Die folgende Anwendung des Satzes von FUBINI ist sehr nützlich. Sie erlaubt in bestimmten Situationen μ-Integrale durch LEBESGUE-Integrale auszudrücken:

Satz 3.8 *Es sei* $(\Omega, \mathcal{A}, \mu)$ *ein* σ*-endlicher Maßraum und* f *eine nicht-negative* $\mathcal{A}$*-messbare reelle Funktion. Ferner sei* $\varphi : [0, \infty) \to [0, \infty)$ *eine monoton wachsende stetige Funktion mit* $\varphi(0) = 0$*, die auf* $(0, \infty)$ *stetig differenzierbar ist. Dann gilt für alle* $A \in \mathcal{A}$

$$\begin{aligned}\int_A \varphi \circ f \, d\mu &= \int_{(0,\,\infty)} \varphi'(t)\, \mu(\{f \geq t\} \cap A)\, \lambda(dt)\\ &= \int_0^\infty \varphi'(t)\, \mu(\{f \geq t\} \cap A) dt\,.\end{aligned}$$

Beweis: Sei μ zunächst endlich. Dann kann ohne Einschränkung $A = \Omega$ gesetzt werden, denn mit

$$\tilde{\mu} := \mu(\,\cdot\, \cap A)$$

gilt

$$\int_A \varphi \circ f \, d\mu = \int \varphi \circ f \, d\tilde{\mu}$$

und

$$\mu\big(\{f \geq t\} \cap A\big) = \tilde{\mu}(f \geq t)$$

für alle $A \in \mathcal{A}$. Sei λ^* das auf $(0, \infty) \cap \mathcal{B}^1$ definierte Lebesgue-Maß. Da φ monoton wachsend ist, gilt $\varphi'(t) \geq 0$ für alle $t > 0$. Die stetige Funktion φ' ist wegen $[\frac{1}{n}, a] \uparrow (0, a]$ über jedem Intervall $(0, a]$, $a > 0$, λ^*-integrierbar und

$$\int_{(0,\,a]} \varphi'(t)\, \lambda^*(dt) = \lim_{n\to\infty} \int_{\frac{1}{n}}^a \varphi'(t)\, dt = \varphi(a) - \lim_{n\to\infty} \varphi(1/n) = \varphi(a)\,,$$

da $\varphi(0) = 0$ und φ in $t = 0$ stetig. Aus $f \geq 0$ folgt

$$\int_{(0,\,f(\omega)]} \varphi'(t)\, \lambda^*(dt) = \varphi(f(\omega))\,, \quad \omega \in \Omega\,.$$

Es folgt

$$\begin{aligned}\int \varphi \circ f \, d\mu &= \int \Big(\int_{(0,\,f(\omega)]} \varphi'(t)\, \lambda^*(dt) \Big)\, \mu(d\omega)\\ &= \iint \varphi'(t)\, 1_{(0,\,f(\omega)]}(t)\, \lambda^*(dt)\, \mu(d\omega)\\ &= \iint \varphi'(t)\, 1_E(\omega,\, t)\, \lambda^*(dt)\, \mu(d\omega)\end{aligned}$$

mit $E := \{(\omega, t) \in \Omega \times (0, \infty) \ : \ f(\omega) \geq t\}$. Die letzte Gleichheit folgt, da $E \in \mathcal{A} \otimes \big(\mathcal{B}^1 \cap (0, +\infty)\big)$. Wenn dies geklärt ist, liefert der Satz von Fubini

$$\begin{aligned}\int \varphi \circ f \, d\mu &= \iint \varphi'(t) \, 1_E(\omega, t) \, \mu(d\omega) \, \lambda^*(dt) \\ &= \int \varphi'(t) \, \mu(E_t) \, \lambda^*(dt) = \int \varphi'(t) \, \mu\big(\{f \geq t\}\big) \, \lambda^*(dt) \,,\end{aligned}$$

denn der t-Schnitt von E ist die Menge aller ω mit $f(\omega) \geq t$. Da $t \mapsto \mu\big(\{f \geq t\}\big)$ linksseitig stetig und monoton fallend ist, hat die Funktion höchstens abzählbar unendlich viele Unstetigkeitsstellen auf $(0, \infty)$. Zusammen mit der Stetigkeit von φ' liefert dies die uneigentliche Riemann-Integrierbarkeit von $t \mapsto \varphi'(t) \, \mu\big(\{f \geq t\}\big)$ auf $(0, \infty)$, wobei der Wert ∞ zugelassen ist.

Zu zeigen bleibt: $E \in \mathcal{A} \otimes \big(\mathcal{B}^1 \cap (0, \infty)\big)$:

Es sei $F(\omega, t) := \big(f(\omega), t\big)$, dann ist F $\mathcal{A} \otimes \big(\mathcal{B}^1 \cap (0, \infty)\big)/\mathcal{B}^2$-messbar, denn jede Komponente ist messbar, also verwenden wir Satz 3.4. Dann liegt E in $\mathcal{A} \otimes \big(\mathcal{B}^1 \cap (0, \infty)\big)$, denn E ist das Urbild unter F des abgeschlossenen Halbraums aller $(x, y) \in \mathbb{R}^2$ mit $x \geq y$.

Ist μ σ-endlich, so folgt die Aussage einfach mit dem Satz von der monotonen Konvergenz. Dies ist eine Übung. $\square$

Bemerkung 3.9 Die Aussage gilt analog, wenn man $\mu\big(\{f \geq t\}\big)$ durch $\mu\big(\{f > t\}\big)$ ersetzt, denn $t \mapsto \mu\big(\{f > t\}\big)$ hat höchstens abzählbar viele Unstetigkeitsstellen, also ist

$$\mu\big(\{f > t\}\big) = \mu\big(\{f \geq t\}\big) \quad \lambda\text{-f.ü.}$$

Beispiel 3.10 Es sei $\varphi(t) = t^p$, $p > 0$. Für jede $\mathcal{A}$-messbare reelle Funktion $f \geq 0$ folgt

$$\int f^p \, d\mu = p \int_0^\infty t^{p-1} \, \mu\big(\{f \geq t\}\big) \, dt$$

und für $p = 1$

$$\int f \, d\mu = \int_{\mathbb{R}^+} \mu\big(\{f \geq t\}\big) \, \lambda^1(dt) = \int_0^\infty \mu\big(\{f \geq t\}\big) \, dt \,.$$

Das Integral $\int f \, d\mu$ wird "vertikal", das Integral auf der rechten Seite jedoch "horizontal" gebildet.

Ist X eine nicht-negative Zufallsgröße auf einem W-Raum $(\Omega, \mathcal{A}, P)$, so folgt insbesondere

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty P(X > t) \, dt \\ \mathbb{E}(X^p) &= \int_0^\infty p \, t^{p-1} \, P(X > t) \, dt \,, \quad p > 0 \,.\end{aligned}$$

Aus der letzten Gleichung folgt

$$\sum_{n\geq 1} p(n-1)^{p-1}\, P(X>n) \leq \mathbb{E}(X^p)$$
$$\leq \sum_{n\geq 0} p(n+1)^{p-1}\, P(X>n)\,, \quad p\geq 1\,.$$

Ist X eine Zufallsgröße und existiert $\mathbb{E}(X)$, so folgt durch Zerlegung von X in Positiv- und Negativteil

$$\mathbb{E}(X) = \int_0^\infty \big(P(X>t) - P(X<-t)\big)\, dt\,.$$

Existiert auch $\mathrm{Var}(X)$, so gilt

$$\mathrm{Var}(X) = \int_0^\infty \big((2t-\mu)P(X>t) + (2t+\mu)P(X<-t)\big)\, dt$$

mit $\mu := \mathbb{E}(X)$.

Satz 3.6 liefert uns zu d Experimenten $(\Omega_j,\, \mathcal{A}_j,\, \mu_j)$, $j=1,\ldots,d$, einen gemeinsamen W-Raum $(\prod \Omega_j,\, \bigotimes \mathcal{A}_j,\, \bigotimes \mu_j)$, der die Ausführung der d Einzelexperimente beschreibt. Die Wahl des Produktmaßes bekommt in Kapitel 5 die Interpretation stochastisch unabhängig ausgeführter Teilexperimente. Zunächst wenden wir uns aber der Frage nach der Konstruktion eines W-Raumes für unendlich viele Experimente zu.

Es sei I eine nichtleere Indexmenge und $(\Omega_i,\, \mathcal{A}_i,\, P_i)_{i\in I}$ eine Familie von W-Räumen. Für $K \subset I$ setzen wir

$$\Omega_K := \prod_{i\in K} \Omega_i\,, \quad \Omega := \prod_{i\in I} \Omega_i\,.$$

Ω_K ist die Menge aller Abbildungen $\omega : K \to \bigcup_{i\in K} \Omega_i$ mit $\omega(i) \in \Omega_i$ für alle $i \in K$. Wir restringieren diese Abbildung auf eine nichtleere Teilmenge $J \subset K$ und erhalten die Projektionsabbildung

$$p_J^K : \Omega_K \to \Omega_J\,.$$

Für $K = I$ setzen wir $p_J := p_J^I$ und $p_i^K := p_{\{i\}}^K$ für $J = \{i\}$, speziell $p_i := p_i^I$. Es gilt

$$p_J^L = p_J^K \circ p_K^L\,, \quad J \subset K \subset L\,,$$

insbesondere

$$p_J = p_J^K \circ p_K\,, \quad J \subset K\,.$$

Es sei $\mathcal{H} = \mathcal{H}(I)$ das System aller *nichtleeren, endlichen* Teilmengen von I. Für $J \in \mathcal{H}$ sind

$$\mathcal{A}_J := \bigotimes_{i\in J} \mathcal{A}_i, \quad P_J := \bigotimes_{i\in J} P_i$$

definiert (Definition 3.2, Satz 3.6). Als Produkt-σ-Algebra für unendlich viele σ-Algebren wählen wir mit Definition 3.1

$$\mathcal{A} := \bigotimes_{i \in I} \mathcal{A}_i := \sigma\big((p_i)_{i \in I}\big)\,.$$

Für jedes $J \in \mathcal{H}$ ist dann p_J $\mathcal{A}/\mathcal{A}_J$-messbar, denn $\mathcal{A}_J = \sigma(p_i^J,\, i \in J)$ und $p_i = p_i^J \circ p_J$ für jedes $i \in J$ (wir verwenden also Satz 3.4). Es gilt also

$$\sigma\big((p_i)_{i \in I}\big) = \sigma(p_J,\, J \in \mathcal{H})\,.$$

Welche Eigenschaft erwarten wir von einem Maß P auf $(\Omega, \mathcal{A})$? Für $A_1 \in \mathcal{A}_1$, $\dots$, $A_n \in \mathcal{A}_n$ (bei einer Folge von W-Räumen $(\Omega_n, \mathcal{A}_n, P_n)$) soll $A = A_1 \times \dots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \dots$ in $\mathcal{A}$ liegen und $P(A) = P_1(A_1) \cdots P_n(A_n)$ gelten. Allgemein wünschen wir also

$$P\Big(p_J^{-1}\big(\prod_{i \in J} A_i\big)\Big) = \prod_{i \in J} P_i(A_i)\,, \quad J \in \mathcal{H},\, A_i \in \mathcal{A}_i,\, i \in J\,.$$

Also soll $p_J(P)$ gleich P_J sein für $J \in \mathcal{H}$, denn P_J ist das einzige derartige Maß (siehe Satz 3.6).

Gibt es ein W-Maß P auf $\mathcal{A}$ derart, dass dessen Bild unter jeder Projektion p_J mit $J \in \mathcal{H}$ gleich P_J ist?

Satz 3.11 (von Andersen und Jessen) *Auf der σ-Algebra $\mathcal{A} := \bigotimes_{i \in I} \mathcal{A}_i$ existiert genau ein Maß P derart, dass für jede Menge $J \in \mathcal{H}(I)$ gilt*

$$p_J(P) = P_J\,.$$

P ist ein W-Maß.

Bevor wir den Beweis führen, machen wir ein paar Vorbetrachtungen:

Für $J, K \in \mathcal{H}$, $J \subset K$, ist p_J^K $\mathcal{A}_K/\mathcal{A}_J$-messbar, denn die Mengen $\prod_{i \in J} A_i$, $A_i \in \mathcal{A}_i$, $i \in J$, erzeugen $\mathcal{A}_J$ und $(p_J^K)^{-1}\big(\prod_{i \in J} A_i\big) = \prod_{i \in K} A_i'$ mit $A_i' = A_i$ für $i \in J$ und $A_i' = \Omega$ für $i \in K \backslash J$.

Es gilt

$$\prod_{i \in K} P_i(A_i') = \prod_{i \in J} P_i(A_i),$$

also mit Satz 3.6

$$p_J^K(P_K) = P_J\,, \quad J \subset K,\, J, K \in \mathcal{H}\,.$$

Wir setzen $\mathcal{Z}_J := p_J^{-1}(\mathcal{A}_J)$, $J \in \mathcal{H}$. Diese σ-Algebra heißt die σ-Algebra der *J-Zylindermengen*. Es ist $(p_J^K)^{-1}(\mathcal{A}_J) \subset \mathcal{A}_K$. Mit $p_J = p_J^K \circ p_K$ folgt

$$\mathcal{Z}_J \subset \mathcal{Z}_K\,, \quad J \subset K,\, J, K \in \mathcal{H}\,. \tag{3.5}$$

Sei

$$\mathcal{Z} := \bigcup_{J \in \mathcal{H}} \mathcal{Z}_J.$$

Wir nennen $\mathcal{Z}$ das *System der Zylindermengen.* Je zwei Zylindermengen $Z_1, Z_2 \in \mathcal{Z}$ liegen in ein und derselben σ-Algebra $\mathcal{Z}_J$ für ein $J \in \mathcal{H}$. Für $Z_i \in \mathcal{Z}_{J_i}$, $i = 1, 2$, ist $J_1 \cup J_2$ nach (3.5) geeignet. $\mathcal{Z}$ ist also eine Algebra, i.a. jedoch keine σ-Algebra. Es gilt aber

$$\mathcal{A} = \sigma(\mathcal{Z})$$

nach Definition und $\mathcal{A} = \sigma(p_J,\ J \in \mathcal{H})$.
Nun kommen wir zum Beweis von Satz 3.11 in vier Schritten.

Beweis von Satz 3.11:
1. Schritt: Um $p_J(P) = P_J$ zu erreichen, muss das gesuchte P auf $Z = p_J^{-1}(A)$ den Wert $P_J(A)$ bekommen, $J \in \mathcal{H}$, $A \in \mathcal{A}_J$. Dieser Wert darf nur von Z, nicht von der speziellen Darstellung $Z = p_J^{-1}(A)$ abhängen!
Es sei $Z = p_J^{-1}(A) = p_K^{-1}(B)$, $J, K \in \mathcal{H}$, $A \in \mathcal{A}_J$, $B \in \mathcal{A}_K$. Wenn $J \subset K$ ist, so gilt

$$p_J^{-1}(A) = p_K^{-1}\big((p_J^K)^{-1}(A)\big)\,,$$

also

$$p_K^{-1}(B) = p_K^{-1}(B')$$

mit $B' := (p_J^K)^{-1}(A)$. Es gilt $B = B' = (p_J^K)^{-1}(A)$, also

$$P_K(B) = P_J(A)\,.$$

Für J und K beliebig setze $L := J \cup K$, also $J \subset L$, $K \subset L$. Also gibt es ein $C \in \mathcal{A}_L$ mit

$$p_L^{-1}(C) = p_J^{-1}(A) = p_K^{-1}(B)\,,$$

also $P_L(C) = P_J(A)$ und $P_L(C) = P_K(B)$, und somit $P_J(A) = P_K(B)$.
Somit ist durch $P_0\big(p_J^{-1}(A)\big) := P_J(A)$, $J \in \mathcal{H}$, $A \in \mathcal{A}_J$, eine Funktion P_0 auf $\mathcal{Z}$ definiert.
2. Schritt: $P_0 : \mathcal{Z} \to \mathbb{R}$ ist ein Inhalt, also $P_0 \geq 0$, $P_0(\emptyset) = 0$ und P_0 ist endlich additiv:
Seien $Y, Z \in \mathcal{Z}$, disjunkt, so existiert ein $J \in \mathcal{H}$ mit $Y = p_J^{-1}(A)$ und $Z = p_J^{-1}(B)$ für geeignete $A, B \in \mathcal{A}_J$. Mit Y und Z sind auch A und B disjunkt, $Y \cup Z = p_J^{-1}(A \cup B)$, also

$$P_0(Y \cup Z) = P_J(A \cup B) = P_J(A) + P_J(B) = P_0(Y) + P_0(Z)\,.$$

Wir zeigen noch die σ-Additivität von P_0, denn dann sind wir schon fertig: P ist dann die einzige Fortsetzung von P_0 zu einem Maß auf $\sigma(\mathcal{Z}) = \mathcal{A}$ (siehe Satz 28.8 und Satz 28.11, Analysis III). P ist ein W-Maß, denn $\Omega = p_J^{-1}(\Omega_J)$, $J \in \mathcal{H}$, also ist Ω eine J-Zylindermenge und es gilt

$$P(\Omega) = P_0(\Omega) = P_J(\Omega_J) = 1\,.$$

3. Schritt: Zu $Z \in \mathcal{Z}$ und $J \in \mathcal{H}$ betrachte

$$Z^{\omega_J} := \Big\{\omega \in \Omega \,:\, \big(\omega_J, p_{I\setminus J}(\omega)\big) \in Z\Big\}\,.$$

Dies ist für jedes $\omega_J \in \Omega_J$ eine Zylindermenge:

Aus $Z = p_K^{-1}(A)$ für ein $K \in \mathcal{H}$, $A \in \mathcal{A}_K$, und ohne Einschränkung $J \subset K$ folgt

$$\begin{aligned} Z^{\omega_J} &= \big\{\omega' : (\omega_J, \omega'_{K\setminus J}, \omega'_{I\setminus K}) \in Z\big\} \\ &= \big\{\omega' : (\omega_J, \omega_{K\setminus J}) \in A\big\} \\ &= \{\omega' : \omega'_{K\setminus J} \in A_{\omega_J}\} \end{aligned}$$

mit $A_{\omega_J} := \big\{\omega'_{K\setminus J} : (\omega_J, \omega_{K\setminus J}) \in A\big\}$, der ω_J-Schnitt von A in Ω_K. Also ist $Z^{\omega_J} = p_{K\setminus J}^{-1}(A_{\omega_J})$.

Es gilt

$$P_0(Z) = \int P_0(Z^{\omega_J})\, P_J(d\omega_J)\,.$$

Natürlich ist $A_{\omega_J} \in \mathcal{A}_{K\setminus J}$ und der Satz von Fubini liefert

$$P_0(Z) = P_K(A) = \int P_{K\setminus J}(A_{\omega_J})\, P_J(d\omega_J)\,.$$

Da $P_0(Z^{\omega_J}) = P_{K\setminus J}(A_{\omega_J})$, folgt also die Behauptung.

4. Schritt: Wir zeigen, dass P_0 $\emptyset$-stetig ist, d.h. dass für jede Folge $(B_n)_n$ von Ereignissen mit $B_n \downarrow \emptyset$ gilt: $\lim_{n\to\infty} P_0(B_n) = 0$. Daraus folgt die σ-Additivität nach Satz 27.15, Analysis III.

Sei $(Z_n)_n$ eine antitone Folge von Zylindermengen mit $\alpha := \inf_n P_0(Z_n) > 0$. Wir zeigen, dass $\bigcap_{n\geq 1} Z_n$ nicht leer sein kann.

Es gilt $Z_n = p_{J_n}^{-1}(A_n)$, $J_n \in \mathcal{H}$, $A_n \in \mathcal{A}_{J_n}$. Ohne Einschränkung kann $J_1 \subset J_2 \subset \dots$ angenommen werden. Die Funktion

$$\omega_{J_1} \mapsto P_0(Z_n^{\omega_{J_1}})$$

ist $\mathcal{A}_{J_1}$-messbar (nach Satz 3.5(ii)), also liegt

$$Q_n := \big\{\omega_{J_1} \in \Omega_{J_1} : P_0(Z_n^{\omega_{J_1}}) \geq \alpha/2\big\}$$

in $\mathcal{A}_{J_1}$, also folgt aus Schritt 3

$$\begin{aligned} \alpha \leq P_0(Z_n) &= \Big(\int_{Q_n} + \int_{Q_n^c}\Big) P_0(Z_n^{\omega_{J_1}})\, P_{J_1}(d\omega_{J_1}) \\ &\leq P_{J_1}(Q_n) + \frac{\alpha}{2}\,, \end{aligned}$$

also $P_{J_1}(Q_n) \geq \frac{\alpha}{2} > 0$. Mit $(Z_n)_n$ ist auch $(Q_n)_n$ antiton und P_{J_1} ist $\emptyset$-stetig, also kann $\bigcap_{n\geq 1} Q_n$ nicht leer sein, es existiert also ein $\omega_{J_1} \in \bigcap_{n\geq 1} Q_n$ mit

$$P_0(Z_n^{\omega_{J_1}}) \geq \frac{\alpha}{2} > 0 \quad \text{für alle } n \geq 1\,.$$

Ist $J_2 \neq J_1$, so liefert Schritt 3 die Existenz eines $\omega_{J_2\setminus J_1}$ mit

$$P_0\big((Z_n^{\omega_{J_1}})^{\omega_{J_2\setminus J_1}}\big) \geq \frac{\alpha}{4} \quad \text{für alle } n \geq 1\,.$$

Es gilt $\omega_{J_2} := (\omega_{J_1}, \omega_{J_2\setminus J_1}) \in \Omega_{J_2}$ und

$$(Z_n^{\omega_{J_1}})^{\omega_{J_2\setminus J_1}} = Z_n^{\omega_{J_2}}\,,$$

also

$$P_0(Z_n^{\omega_{J_2}}) \geq \frac{\alpha}{4} > 0 \quad \text{für alle } n \geq 1$$

und

$$\omega_{J_1} = p_{J_1}^{J_2}(\omega_{J_2})\,.$$

Ist $J_1 = J_2$, wählen wir $\omega_{J_2} = \omega_{J_1}$. Vollständige Induktion liefert: Zu jedem $k \in \mathbb{N}$ existiert ein $\omega_{J_k} \in \Omega_{J_k}$ mit

$$P_0(Z_n^{\omega_{J_k}}) \geq 2^{-k}\alpha > 0\,, \quad p_{J_k}^{J_{k+1}}(\omega_{J_{k+1}}) = \omega_{J_k}\,, \quad n, k \in \mathbb{N}\,.$$

Also existiert ein $\omega_0 \in \Omega$ mit $p_{J_k}(\omega_0) = \omega_{J_k}$ für alle $k \geq 1$ und $Z_n^{\omega_{J_n}} \neq \emptyset$, also existiert ein $\tilde{\omega}_n \in \Omega$ mit $\big(\omega_{J_n},\, p_{I \setminus J_n}(\tilde{\omega}_n)\big) \in Z_n$.

In der J_n-Zylindermenge Z_n liegt dann auch der Punkt $\big(\omega_{J_n},\, p_{I\setminus J_n}(\omega_0)\big) = \omega_0$. Also ist $\omega_0 \in Z_n$ für alle $n \geq 1$, d.h. $\bigcap_{n\geq 1} Z_n \neq \emptyset$. □

Definition 3.12 Das nach Satz 3.11 eindeutig bestimmte W-Maß P heißt das *Produktmaß der W-Maße* $(P_i)_{i\in I}$ und wird mit

$$\bigotimes_{i\in I} P_i$$

bezeichnet. Der W-Raum

$$\Big(\prod_{i\in I} \Omega_i\,, \bigotimes_{i\in I} \mathcal{A}_i\,, \bigotimes_{i\in I} P_i\Big)$$

heißt *Produkt der Familie* $\big((\Omega_i,\, \mathcal{A}_i,\, P_i)\big)_{i\in I}$ von W-Räumen und wird mit

$$\bigotimes_{i\in I} (\Omega_i,\, \mathcal{A}_i,\, P_i)$$

bezeichnet.

KAPITEL 4

Konvergenz von Zufallsvariablen und Verteilungen

Wir stellen vier verschiedene Konvergenzbegriffe für Folgen von Zufallsvariablen vor. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $(X_n)_n$ eine Folge von Zufallsgrößen und X eine weitere Zufallsgröße auf $(\Omega, \mathcal{A}, P)$.

Definition 4.1 Die Folge $(X_n)_n$ konvergiert P-fast sicher (P-f.s.) gegen die Zufallsgröße X, wenn gilt

$$P\big(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \text{ existiert und stimmt mit } X(\omega) \text{ überein}\}\big) = 1$$

oder kurz $P(X_n \to X) = 1$. Man schreibt dann $X_n \to X$ $P - f.s.$ oder $X_n \to X$ $f.s.$, wenn über P kein Zweifel besteht.

$A = \{X_n \to X\}$ ist eine messbare Menge, denn

$$\begin{aligned} A &= \bigcap_{k\geq 1} \bigcup_{m\geq 1} \bigcap_{n\geq m} \big(\{|X_n - X| \leq 1/k, |X| < \infty\} \\ &\quad \cup \{X_n > k, X = \infty\} \cup \{X_n < -k, X = -\infty\}\big) \,. \end{aligned}$$

Lemma 4.2 *Aus $X_n \to X$ f.s. und $X_n \to X'$ f.s. folgt $X = X'$ f.s.*

Beweis: In der Menge $\{X_n \to X\} \cap \{X_n \to X'\}$ gilt $X = X'$, also $P(X \neq X') \leq P(X_n \not\to X) + P(X_n \not\to X') = 0$. □

Wir betrachten nun ein notwendiges und hinreichendes Kriterium für die fast sichere Konvergenz:

Satz 4.3 *$(X_n)_n$ konvergiert genau dann f.s. gegen einen reellen Limes X, wenn*

$$\lim_{m\to\infty} P\Big(\sup_{n\geq m} |X_n - X| > \varepsilon\Big) = 0 \ \textit{für alle } \varepsilon > 0.$$

Beweis: Da X reellwertig ist, konvergiert $(X_n)_n$ genau dann f.s. gegen X, wenn

$$\begin{aligned} A^c &= \bigcup_{k\geq 1} \bigcap_{m\geq 1} \bigcup_{n\geq m} \{|X_n - X| > 1/k\} \\ &= \bigcup_{k\geq 1} \bigcap_{m\geq 1} \Big\{ \sup_{n\geq m} |X_n - X| > 1/k \Big\} \end{aligned}$$

Wahrscheinlichkeit 0 hat, also

$$P\Big(\bigcap_{m\geq 1}\Big\{\sup_{n\geq m}|X_n-X|>1/k\Big\}\Big)=\lim_{m\to\infty}P\Big(\sup_{n\geq m}|X_n-X|>1/k\Big)=0$$

für alle $k\geq 1$ gilt. □

Satz 4.4 (Cauchy-Kriterium) *$X_n\to X$ f.s. für einen reellen Limes genau dann, wenn*

$$\lim_{m\to\infty}P\Big(\sup_{n\geq m}|X_n-X_m|>\varepsilon\Big)=0$$

für alle $\epsilon>0$, wobei wir $|\infty-\infty|=|-\infty+\infty|=\infty$ setzen.

Beweis: „$\Rightarrow$“:

$$\begin{aligned}P\Big(\sup_{n\geq m}|X_n-X_m|>\varepsilon\Big) &\leq P\Big(\sup_{n\geq m}|X_n-X|>\frac{\varepsilon}{2}\Big)+P\Big(|X_m-X|>\frac{\varepsilon}{2}\Big)\\ &\leq 2P\Big(\sup_{n\geq m}|X_n-X|>\frac{\varepsilon}{2}\Big).\end{aligned}$$

Nach Satz 4.3 konvergiert die rechte Seite der Ungleichungskette gegen Null für $m\to\infty$.

„$\Leftarrow$“: Es sei $Y_n:=\sup_{j,k\geq n}|X_k-X_j|$, $n\geq 1$. $(Y_n)_n$ ist eine monoton fallende Folge nichtnegativer Zufallsgrößen und es gilt

$$P\Big(\sup_{n\geq m}Y_n>\varepsilon\Big)=P(Y_m>\varepsilon)\leq 2P\Big(\sup_{n\geq m}|X_n-X_m|>\frac{\varepsilon}{2}\Big)\ .$$

Die rechte Seite konvergiert nach Voraussetzung gegen Null für $m\to\infty$. Nach Satz 4.3 folgt somit $Y_n\to 0$ f.s. Es sei $A:=\{Y_n\to 0\}$, dann bildet $(X_n(\omega))_n$ für jedes $\omega\in A$ eine Cauchy-Folge in $\mathbb{R}$, hat also einen Limes $X(\omega)$. Für $\omega\in A^c$ setzen wir $X(\omega)=0$, so dass $X_n\to X$ auf A, also fast sicher. □

Es sei an die $\mathcal{L}_p$-Konvergenz erinnert, siehe Definition 33.11, Analysis III.

Definition 4.5 $(X_n)_n\subset\mathcal{L}_p(\Omega,\mathcal{A},P)$, $p>0$, *konvergiert im p-ten Mittel* gegen eine Zufallsgröße X, falls $X\in\mathcal{L}_p(\Omega,\mathcal{A},P)$ und

$$\mathbb{E}(|X_n-X|^p)\to 0 \text{ für } n\to\infty\ .$$

Man spricht auch von $\mathcal{L}_p$-Konvergenz.

Definition 4.6 $(X_n)_n$ *konvergiert in W.keit* (unter P) oder *P-stochastisch* gegen eine reelle Zufallsvariable X, wenn

$$\lim_{n\to\infty}P(|X_n-X|>\varepsilon)=0$$

für alle $\varepsilon>0$ gilt.

Wir schreiben $X_n\xrightarrow{P}X$.

Lemma 4.7 *Aus $X_n\xrightarrow{P}X$ und $X_n\xrightarrow{P}X'$ folgt $X=X'$ fast sicher.*

Beweis: $P(|X - X'| > \varepsilon) \leq P(|X_n - X| > \varepsilon/2) + P(|X_n - X'| > \varepsilon/2)$ für alle $n \geq 1$ und $\varepsilon > 0$. Da $\{X \neq X'\} = \bigcup_{k=1}^{\infty} \{|X - X'| \geq 1/k\}$, folgt die Behauptung. $\square$

Definition 4.8 $(X_n)_n$ konvergiert *schnell P-stochastisch* gegen X, wenn

$$\sum_{n \geq 1} P(|X_n - X| > \varepsilon) < \infty$$

für alle $\varepsilon > 0$ gilt.

Satz 4.9 *Konvergiert $(X_n)_n$ schnell P-stochastisch gegen X, so auch fast sicher.*

Beweis: Es gilt

$$P(|X_n - X| > \varepsilon \text{ unendlich oft}) = P(\limsup\{|X_n - X| > \varepsilon\}) = 0$$

nach dem Lemma von Borel-Cantelli (1.11). Dies ist zu $\lim_{m\to\infty} P(\sup_{n\geq m} |X_n - X| > \varepsilon) = 0$ äquivalent, womit die Behauptung aus Satz 4.3 folgt. Die Äquivalenz sieht man so:
Ist $B_m := \bigcup_{n=m}^{\infty} \{|X_n - X| > \varepsilon\}$ und $B := \limsup_{n\to\infty}\{|X_n - X| > \varepsilon\}$, so gilt $B_m \searrow B$ und $\lim_{m\to\infty} P(B_m) = P(B)$ und $B_m = \{\sup_{n\geq m} |X_n - X| > \varepsilon\}$. $\square$

Satz 4.10 *Fast sichere Konvergenz impliziert Konvergenz in W.keit. Konvergenz im p-ten Mittel, $p \geq 1$, impliziert Konvergenz in W.keit.*

Beweis: Die Ungleichung $P(|X_n - X| > \varepsilon) \leq P(\sup_{k\geq n} |X_k - X| > \varepsilon)$ liefert mit Satz 4.3 die erste Aussage, die zweite folgt aus der Markov-Ungleichung 2.8(i). $\square$

Die anderen denkbaren Implikationen sind nicht richtig, wie die folgenden Beispiele belegen:

Beispiele 4.11 Es sei $(\Omega, \mathcal{L}, P) = ([0,1], \mathcal{B} \cap [0,1], \lambda|_{[0,1]})$.

(a) Sei $X_n = n^{1/p} 1_{[0,1/n]}$, $p > 0$. Dann gilt $X_n \to 0$ f.s. und in W.keit, aber $\mathbb{E}(|X_n|^p) = 1$ für alle $n \in \mathbb{N}$, also konvergiert $(X_n)_n$ nicht im p-ten Mittel gegen Null.

(b) Für jede natürliche Zahl n gibt es genau ein Paar ganzer Zahlen $m \geq 0$, $k \geq 0$ mit $n = 2^m + k$ und für $k < 2^m$. Wir setzen $X_n = 1_{[k2^{-m}, (k+1)2^{-m}[}$. $(X_n(\omega))_n$ konvergiert für kein $\omega \in [0,1]$. Zu $\omega \in \Omega$ und $m = 0, 1, \ldots$ existiert genau ein $k \in \{0, \ldots, 2^m - 1\}$ mit $\omega \in [k2^{-m}, (k+1)2^{-m}[$. Im Fall $k < 2^m - 1$ ist $\omega \notin [(k+1)2^{-m}, (k+2)2^{-m}[$ und im Fall $k = 2^m - 1$ und $k \geq 1$ ist $\omega \notin [0, 2^{-(m+1)}[$. Aber es ist $P(|X_n| > \varepsilon) \leq 2^{-m}$ für alle $\varepsilon > 0$ und $\mathbb{E}(|X_n|^p) = 2^{-m}$ für $p > 0$.

Unter Zusatzbedingungen impliziert die fast sichere Konvergenz die Konvergenz im p-ten Mittel:

Satz 4.12 *Es sei $(X_n)_n$ eine Folge von Zufallsgrößen, die f.s. gegen X konvergiert. Gilt $|X_n| \leq Y$ fast sicher für ein $Y \in \mathcal{L}_p$, $p > 0$, so konvergiert X_n gegen X im p-ten Mittel.*

Beweis: Es gilt

$$|X_n - X|^p < (|X_n| + |X|)^p \leq (2Y)^p = 2^p Y^p \in \mathcal{L}_1.$$

Weiter ist $|X_n - X|^p \longrightarrow 0$ fast sicher. Nach dem Satz von der dominierten Konvergenz (Satz 2.5 bzw. Satz 32.12, Analysis III) folgt $\mathbb{E}(|X_n - X|^p) \to 0$. $\square$

Aus der Konvergenz in W.keit folgt nicht fast sichere Konvergenz (siehe Beispiel 4.11), aber es gilt:

Satz 4.13 *Es sei $(X_n)_n$ eine Folge, die in W.keit gegen X konvergiert, so existiert eine Teilfolge $(X_{n_k})_k$ mit $\lim_{k\to\infty} X_{n_k} = X$ fast sicher.*

Beweis: Zu jedem $k \in \mathbb{N}$ existiert ein $n_k \in \mathbb{N}$ mit $P(|X_{n_k} - X| \geq 1/k) \leq 1/k^2$. Wir können $n_{k+1} > n_k$ für alle $k \in \mathbb{N}$ annehmen. Da $\sum_{k\geq 1} 1/k^2 < \infty$ folgt die Behauptung aus Satz 4.9. $\square$

Alle drei bisherigen Konvergenztypen (fast sicher, im p-ten Mittel und in W.keit) sind vollständig, das heißt, dass jede CAUCHY-Folge konvergiert. Für fast sichere Konvergenz folgt dies unmittelbar aus der Vollständigkeit von $\mathbb{R}$, für Konvergenz im p-ten Mittel hatten wir es in Analysis III, Definition 33.11 und Satz 33.12 (Satz von Riesz-Fischer) gesehen, für Konvergenz in W.keit ist es der folgende Satz:

Satz 4.14 *Es sei $(X_n)_n$ eine Folge von Zufallsgrößen mit*

$$\lim_{n,m\to\infty} P(|X_n - X_m| \geq \varepsilon) = 0$$

für alle $\varepsilon > 0$. Dann existiert eine Zufallsgröße X mit $X_n \to X$ in W.keit.

Beweis: Wähle wie im Beweis von Satz 4.13 eine Teilfolge $(n_k)_k$ mit

$$P(|X_{n_k} - X_{n_{(k+1)}}| \geq 1/k) \leq 1/k^2 \ .$$

Nach dem Lemma von Borel-Cantelli (1.11) folgt

$$P\big(\limsup_{k\to\infty}\{|X_{n_k} - X_{n_{(k+1)}}| \geq 1/k\}\big) = 0 \ .$$

Für $\omega \notin \limsup_{k\to\infty}\{|X_{n_k} - X_{n_{(k+1)}}| \geq 1/k\}$ ist $(X_{n_k}(\omega))_k$ also eine CAUCHY-Folge in $\mathbb{R}$, also $(X_{n_k})_k$ konvergiert f.s. gegen ein X, also in W.keit. Für $\varepsilon > 0$ gilt

$$P(|X_m - X| \geq \varepsilon) \leq P(|X_m - X_{n_k}| > \varepsilon/2) + P(|X_{n_k} - X| > \varepsilon/2)$$

für alle m und k. Wähle k als die kleinste Zahl mit $n_k \geq m$, so folgt

$$\lim_{m\to\infty} P(|X_m - X| \geq \varepsilon) = 0$$

für jedes $\varepsilon > 0$. □

Die Verteilung einer Zufallsvariablen spielt eine zentrale Rolle. Implizieren die diskutierten Arten der Konvergenz von $(X_n)_n$ gegen X eine Konvergenz der Folge der Verteilungen P^{X_n} gegen die Verteilung P^X von X? Was soll dabei eine Konvergenz von Verteilungen genau bedeuten?

Definition 4.15 Eine Folge $(\mu_n)_n$ von W-Maßen auf $\mathcal{B}^d$ heißt *schwach konvergent* gegen ein W-Maß μ auf $\mathcal{B}^d$, wenn

$$\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu$$

für alle Funktionen $f \in C_b(\mathbb{R}^d)$ gilt (wobei $C_b(\mathbb{R}^d)$ den Vektorraum aller beschränkten, stetigen, reellen Funktionen auf $\mathbb{R}^d$ bezeichnet). Man schreibt $\lim_{n\to\infty} \mu_n = \mu$.

Sind X und $(X_n)_n$ $\mathbb{R}^d$-wertige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ und konvergiert $(P^{X_n})_n$ schwach gegen P^X bzw. allgemeiner gegen ein W-Maß ν auf $\mathcal{B}^d$, so nennt man die Folge $(X_n)_n$ *konvergent in Verteilung* gegen X bzw. gegen ν.

Bemerkungen 4.16 (i) Warum ist nicht $\mu_n(A) \to \mu(A)$ für alle $A \in \mathcal{B}^d$ der geeignete Konvergenzbegriff? Es sei μ_n die Verteilung einer binomialverteilten Zufallsgröße X_n zu den Parametern n und p. Dann existiert zu jedem n eine endliche Menge A_n mit $P\Big(\frac{X_n - np}{\sqrt{np(1-p)}} \in A_n\Big) = 1$. Für $A := \bigcup_{n=1}^\infty A_n$ gilt dann $P\Big(\frac{X_n - np}{\sqrt{np(1-p)}} \in A_n\Big) = 1$ für alle $n \in \mathbb{N}$, aber $\nu_{0,1}(A) = 0$ und nach dem Satz von Moivre und Laplace erwarten wir $\nu_{0,1}$ als „Limesverteilung“!

(ii) Sei $(x_n)_n$ eine Folge reeller Zahlen. Sie konvergiert genau dann gegen $x_0 \in \mathbb{R}$, wenn die Folge $(\delta_{x_n})_n$ schwach gegen δ_{x_0} konvergiert. Aus $\lim_{n\to\infty} x_n = x_0$ folgt sofort $\lim_{n\to\infty} \delta_{x_n} = \delta_{x_0}$. Für die Rückrichtung sei zu $\varepsilon > 0$

$$f(x) := \max(0, 1 - 1/\varepsilon |x - x_0|) \,,$$

also ist $f \in C_b(\mathbb{R})$. Es gilt $\{f > 0\} = (x_0 - \varepsilon, x_0 + \varepsilon)$ und $\lim_n f(x_n) = f(x_0) = 1$. Also gilt $|x_n - x_0| < \varepsilon$ für schließlich alle $n \in \mathbb{N}$.

(iii) $(\Omega, \mathcal{A}, P)$ sei gewählt wie in Beispiel 4.11, $X_n := 1_{[1/2,1]}$, $X := 1_{[0,1/2]}$. Dann ist $P^{X_n} = P^X = (1/2)(\delta_0 + \delta_1)$. $(X_n)_n$ konvergiert also in Verteilung sowohl gegen X als auch gegen X_1. Aber $|X(\omega) - X_1(\omega)| = 1$ für $\omega \in \Omega$, also $X(\omega) \neq X_1(\omega)$ für alle $\omega \in \Omega$ und $P(|X - X_1| \geq 1) = 1$ für jedes $n \in \mathbb{N}$. Somit ist bei Konvergenz in Verteilung der Limes nicht fast sicher eindeutig bestimmt und aus der Konvergenz in Verteilung folgt im Allgemeinen nicht die stochastische Konvergenz.

Es gilt aber

Satz 4.17 *Eine Folge $(X_n)_n$ reeller Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ konvergiere stochastisch gegen eine reelle Zufallsgröße X auf Ω. Dann konvergiert $(X_n)_n$ in Verteilung gegen X.*

Ist X fast sicher konstant, also P^X ein Dirac-Maß, so gilt hiervon auch die Umkehrung.

Wir fassen vor dem Beweis von Satz 4.17 die Zusammenhänge zwischen den Konvergenzbegriffen in einem Schema zusammen:

$$
\begin{array}{ccc}
 & & \mathcal{L}^p\text{-Konvergenz} \\
 & & \Downarrow \\
\text{fast sichere Konvergenz} & & \mathcal{L}^1\text{-Konvergenz} \\
 \searrow & & \swarrow \\
 & \text{stochastische Konvergenz} & \\
 & \Downarrow & \\
 & \text{schwache Konvergenz der Verteilungen} & \\
 & \text{(Konvergenz in Verteilung)} &
\end{array}
$$

Beweis: Sei $f \in C_b(\mathbb{R})$ zunächst gleichmäßig stetig auf $\mathbb{R}$. Zu $\varepsilon > 0$ existiert also ein $\delta > 0$, so dass für $|x-y| < \delta$ folgt $|f(x) - f(y)| < \varepsilon$, $x, y \in \mathbb{R}$. Es sei $A_n := \{|X_n - X| \geq \delta\}$, $n \in \mathbb{N}$. Es gilt $|f \circ X_n - f \circ X| \leq |f \circ X_n| + |f \circ X| \leq 2\|f\|_\infty$ (Supremums-Norm). Es folgt

$$
\begin{aligned}
\Big| \int f \, dP^{X_n} - \int f \, dP^X \Big| &= |\mathbb{E}(f \circ X_n - f \circ X)| \\
&\leq \mathbb{E}|(f \circ X_n - f \circ X)| \\
&= \int_{A_n} |f \circ X_n - f \circ X| \, dP + \int_{{A_n}^c} |f \circ X_n - f \circ X| \, dP \\
&\leq 2\|f\|_\infty P(A_n) + \varepsilon P({A_n}^c) \\
&\leq 2\|f\|_\infty P(A_n) + \varepsilon \; .
\end{aligned}
$$

Wegen der stochastischen Konvergenz ist $\lim_{n\to\infty} P(A_n) = 0$, also folgt $\lim_{n\to\infty} \int f \, dP^{X_n} = \int f \, dP^X$ für diese Klasse von Abbildungen.

Es sei f nun beliebig (in $C_b(\mathbb{R})$). Wähle $I_n := [-n, n] \nearrow \mathbb{R}$, also $P^X(I_n) \nearrow 1$. Zu jedem $\varepsilon > 0$ existiert daher ein $n_0 \in \mathbb{N}$ mit $1 - P^X(I_{n_0}) = P^X(\mathbb{R} \setminus I_{n_0}) < \varepsilon$. Eine Funktion $u_\varepsilon \in C_b(\mathbb{R})$ sei wie folgt definiert: auf I_{n_0} sei sie gleich 1, auf $[n_0, n_0 + 1]$ und $[-n_0 - 1, -n_0]$ affin-linear, $u_\varepsilon(n_0 + 1) = u_\varepsilon(-n_0 - 1) = 0$ und auf $I^c_{n_0+1}$ sei sie Null (siehe Abbildung 4.1).

Wir betrachten nun $f' := u_\varepsilon f$. Die Funktionen u_ε und f' sind auf $I^c_{n_0+1}$ Null und daher auf $\mathbb{R}$ gleichmäßig stetig. Damit folgt

$$
\lim_{n\to\infty} \int f' \, dP^{X_n} = \int f' \, dP^X \qquad \lim_{n\to\infty} \int u_\varepsilon \, dP^{X_n} = \int u_\varepsilon \, dP^X
$$

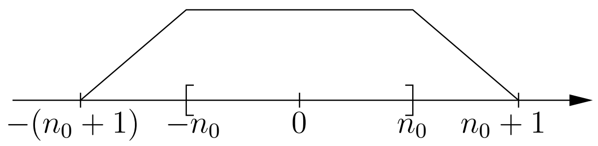


ABBILDUNG 4.1.

nach bereits Gezeigtem. Also folgt auch

$$\lim_{n\to\infty}\int (1-u_\varepsilon)\, dP^{X_n} = \int (1-u_\varepsilon)\, dP^X \ . \tag{4.1}$$

Nun gilt

$$\begin{aligned} &|\int f\, dP^{X_n} - \int f\, dP^X| \\ &\leq \int |f-f'|\, dP^{X_n} + |\int f'\, dP^{X_n} - \int f'\, dP^X| + \int |f'-f|\, dP^X \end{aligned} \tag{4.2}$$

und

$$\int |f-f'|\, dP^X = \int |f|(1-u_\varepsilon)\, dP^X \leq \|f\|_\infty \varepsilon \ ,$$

denn $\int (1-u_\varepsilon)\, dP^X \leq P^X(\mathbb{R}\setminus I_{n_0}) < \varepsilon$. Weiter ist dann $\int (1-u_\varepsilon)\, dP^{X_n} < \varepsilon$ nach (4.1) für schließlich alle n, etwa $n \geq n_1$. Also folgt auch

$$\int |f-f'|\, dP^{X_n} \leq \|f\|_\infty \varepsilon$$

für alle $n \geq n_1$. Also ist die rechte Seite von (4.2) für n hinreichend groß kleiner $2\|f\|_\infty\varepsilon + \varepsilon$, was zu zeigen war.

Ist umgekehrt $X = \eta$ P-fast sicher, also $P^X = \delta_\eta$, so wähle zu $(\eta-\varepsilon, \eta+\varepsilon)$ mit $\varepsilon > 0$ eine stückweise affin-lineare Funktion $f \in C_b(\mathbb{R})$ mit $f \leq 1_{(\eta-\varepsilon,\eta+\varepsilon)}$ und $f(\eta) = 1$. Dann ist

$$\int f\, dP^{X_n} \leq P^{X_n}\big((\eta-\varepsilon,\eta+\varepsilon)\big) = P\big(X_n \in (\eta-\varepsilon,\eta+\varepsilon)\big) \leq 1$$

und nach Voraussetzung $\lim_{n\to\infty}\int f\, dP^{X_n} = f(\eta) = 1$, also folgt

$$\lim_{n\to\infty} P\big(X_n \in (\eta-\varepsilon,\eta+\varepsilon)\big) = 1 \ .$$

Nun ist $\{X_n \in (\eta-\varepsilon,\eta+\varepsilon)\} = \{|X_n-\eta| < \varepsilon\}$, also $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$ für alle $\varepsilon > 0$, womit die stochastische Konvergenz gezeigt ist. $\square$

Der Beweis des obigen Satzes zeigt insbesondere, dass $(P^{X_n})_n$ schwach gegen P^X konvergiert genau dann, wenn

$$\lim_{n\to\infty}\int f\, dP^{X_n} = \int f\, dP^X$$

für alle *gleichmäßig stetigen* und beschränkten $f : \mathbb{R} \to \mathbb{R}$ gilt.

KAPITEL 5

Unabhängigkeit

In Kapitel 3 haben wir das Produktwahrscheinlichkeitsmaß auf unendlichen Produkträumen konstruiert, um ein Zufallsexperiment mit unendlich vielen Einzelexperimenten zu beschreiben. Wenn sich die Ausgänge der Einzelexperimente nicht gegenseitig beeinflussen, spricht man von „stochastisch unabhängigen“ Experimenten. Wir wollen diesen Begriff präzisieren. Er basiert letztendlich auf dem Begriff des Produktmaßes.

Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum. Wir setzen im Folgenden voraus, dass Familien von Teilmengen von Ω stets Ω enthalten.

Definition 5.1 (i) Teilmengen $\mathcal{E}_1, \ldots, \mathcal{E}_n$ von $\mathcal{A}$ mit $\Omega \in \mathcal{E}_i$ heißen *unabhängig*, wenn für $A_i \in \mathcal{E}_i$, $1 \leq i \leq n$, gilt:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)\,.$$

(ii) Es sei I eine Indexmenge und $\mathcal{E}_i$ für $i \in I$ seien Teilmengen von $\mathcal{A}$. Sie heißen *unabhängig*, wenn je endlich viele unabhängig sind.

(iii) Ereignisse A_i für $i \in I$ heißen *unabhängig*, wenn die Mengensysteme $\{A_i, \Omega\}$, $i \in I$, unabhängig sind.

Bemerkung 5.2 Die Voraussetzung, dass die Mengensysteme stets Ω enthalten, dient der bequemen Notation. Es hat nämlich zur Folge, dass für unabhängige Mengensysteme $\mathcal{E}_1, \ldots, \mathcal{E}_n$ auch stets

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}) \tag{5.1}$$

für $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ und $A_{i_j} \in E_{i_j}$ ist. Setzt man $\Omega \in \mathcal{E}_i$ nicht voraus, so muss man (5.1) als Definition verwenden.

Lemma 5.3 *Sind die $\mathcal{E}_i$ für $i \in I$ unabhängig und gilt $\mathcal{D}_i \subset \mathcal{E}_i$ für $i \in I$, so sind die $\mathcal{D}_i$ für $i \in I$ unabhängig. Ist $\mathcal{D}$ unabhängig von $\mathcal{E}_i$ für $i \in I$, so ist $\mathcal{D}$ unabhängig von $\bigcup_{i \in I} \mathcal{E}_i$.*

Beweis: Der erste Teil ist klar. Für $A \in \mathcal{D}$ und $B \in \bigcup_{i \in I} \mathcal{E}_i$ existiert ein $i \in I$ mit $B \in \mathcal{E}_i$, also ist

$$P(A \cap B) = P(A)\,P(B)\,. \qquad \square$$

Beispiele 5.4 **(a)** Es seien $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, \ldots, n$, endlich viele W-Räume, $(\Omega, \mathcal{A}, P)$ der Produktraum und $\tilde{A}_i \in \mathcal{A}_i$, $i = 1, \ldots, n$. Dann sind

$$\begin{aligned} A_1 &:= \tilde{A}_1 \times \Omega_2 \times \cdots \times \Omega_n\,, \\ A_2 &:= \Omega_1 \times \tilde{A}_2 \times \Omega_3 \times \cdots \times \Omega_n\,, \\ \ldots & \quad \ldots \\ A_n &:= \Omega_1 \times \cdots \times \Omega_{n-1} \times \tilde{A}_n \end{aligned}$$

unabhängig.

(b) Es seien $\Omega := [0, 1)$, $\mathcal{A} := \Omega \cap \mathcal{B}^1$, $P := \lambda^1|_\Omega$. Für $n \in \mathbb{N}$ sei

$$A_n := \left[0, \frac{1}{2^n}\right) \cup \left[\frac{2}{2^n}, \frac{3}{2^n}\right) \cup \ldots \cup \left[\frac{2^n - 2}{2^n}, \frac{2^n - 1}{2^n}\right).$$

Die $(A_n)_n$ sind unabhängig, denn $P(A_n) = \frac{1}{2}$, $n \in \mathbb{N}$, und

$$\begin{aligned} P(A_{i_1} \cap \cdots \cap A_{i_n}) &= \frac{1}{2} P(A_{i_1} \cap \cdots \cap A_{i_{n-1}}) \\ &= P(A_{i_1}) \cdots P(A_{i_n}) \end{aligned}$$

für je endlich viele paarweise verschiedene Zahlen $i_1, \ldots, i_n \in \mathbb{N}$. Die Menge A_n ist die Menge aller $x \in [0, 1)$ mit $\varepsilon_n = 0$ in der eindeutigen dyadischen Entwicklung

$$x = \sum_{k=1}^{\infty} \varepsilon_k 2^{-k}$$

mit $\varepsilon_k = 0$ oder 1 und nicht $\varepsilon_k = 1$ für schließlich alle k (siehe auch Beispiel 1.29).

Wir diskutieren nun Möglichkeiten, Unabhängigkeitsaussagen von Mengensystemen auf größere Mengensysteme hochzuziehen:

Satz 5.5 *Es seien $\mathcal{D}_i$ für $i \in I$ unabhängige Teilmengen von $\mathcal{A}$ mit $\Omega \in \mathcal{D}_i$. Sind die $\mathcal{D}_i$ durchschnittstabil, so sind die $\sigma(\mathcal{D}_i)$ für $i \in I$ unabhängig.*

Beweis: Ohne Einschränkung sei I endlich, etwa $I = \{1, \ldots, n\}$. Wir zeigen

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n) \tag{5.2}$$

für $A_i \in \sigma(\mathcal{D}_i)$. Für $0 \leq k \leq n$ sei L_k die folgende Aussage:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)\,, \qquad \begin{aligned} &\forall A_i \in \sigma(\mathcal{D}_i) \text{ für } i \leq k\,, \\ &\forall A_i \in \mathcal{D}_i \text{ für } i > k\,. \end{aligned}$$

L_0 gilt, da die $\mathcal{D}_i$ unabhängig sind. Wir zeigen

$$L_k \Rightarrow L_{k+1} \text{ für } 0 \leq k \leq n-1\,.$$

Betrachte das Mengensystem $\mathcal{A}_{k+1}$ bestehend aus den Mengen $A_{k+1} \in \sigma(\mathcal{D}_{k+1})$, die die Eigenschaft haben, dass die Gleichung (5.2) $\forall A_1 \in \sigma(\mathcal{D}_1), \ldots, \forall A_k \in \sigma(\mathcal{D}_k)$, $\forall A_{k+2} \in \mathcal{D}_{k+2}, \ldots, \forall A_n \in \mathcal{D}_n$ gilt.

Aus L_k folgt $\mathcal{A}_{k+1} \supset \mathcal{D}_{k+1}$. Wir zeigen, dass $\mathcal{A}_{k+1}$ ein Dynkin-System ist.
(i) $\Omega \in \mathcal{A}_{k+1}$ gilt, denn $\Omega \in \mathcal{D}_{k+1}$.
(ii) Für $D \in \mathcal{A}_{k+1}$ gilt

$$\begin{aligned} P\Big(\bigcap_{j=1}^{k} A_j \cap D^c \cap \bigcap_{j=k+2}^{n} A_j\Big) &\\ = P\big(\bigcap_{j=1}^{k} A_j \cap \bigcap_{j=k+2}^{n} A_j\Big) - P\big(\bigcap_{j=1}^{k} A_j \cap D \cap \bigcap_{j=k+2}^{n} A_j\Big) &\\ = \prod_{j,\, j\neq k+1} P(A_j) - P(D) \prod_{j,\, j\neq k+1} P(A_j) &\\ = \prod_{j,\, j\neq k+1} P(A_j)\, P(D^c) & \end{aligned}$$

für alle A_i gemäß den obigen Bedingungen, also $D^c \in \mathcal{A}_{k+1}$.
(iii) Für paarweise disjunkte $D_i \in \mathcal{A}_{k+1}$, $i \in \mathbb{N}$, folgt mittels der σ-Additivität von P

$$\bigcup_{i=1}^{\infty} D_i \in \mathcal{A}_{k+1}\,.$$

Nun folgt aus Satz 1.8

$$\mathcal{A}_{k+1} = \sigma(\mathcal{D}_{k+1})\,,$$

was aber heißt, dass L_{k+1} gilt. □

Bemerkung 5.6 Da das Mengensystem $\{A, \Omega\}$, $A \in \mathcal{A}$, durchschnittstabil ist, folgt: Sind A_i für $i \in I$ unabhängige Ereignisse, so sind die σ-Algebren $\{\emptyset, A_i, A_i^c, \Omega\}$ unabhängig, insbesondere auch die Komplemente A_i^c.

Korollar 5.7 (Blockbildung) *Es seien $\mathcal{D}_i \subset \mathcal{A}$ für $i \in I$ unabhängig und durchschnittstabil. Es sei $(I_k)_{k\in K}$ eine Famile von paarweise disjunkten Teilmengen von I. Dann sind die*

$$\sigma\Big(\bigcup_{j\in I_k} \mathcal{D}_j\Big)$$

für $k \in K$ unabhängig.

Beweis: Für $k \in K$ sei $\hat{\mathcal{D}}_k$ die Familie der endlichen Durchschnitte von Elementen aus $\mathcal{D}_j$ für $j \in I_k$. $\hat{\mathcal{D}}_k$ ist durchschnittstabil, und da die $\mathcal{D}_j$ durchschnittstabil sind, hat jedes Element aus $\hat{\mathcal{D}}_k$ die Gestalt $A_{j_1} \cap \cdots \cap A_{j_n}$ mit $n \in \mathbb{N}$, $A_j \in \mathcal{D}_j$ und $j_1, \ldots, j_n \in I_k$ verschieden. Daraus folgt, dass die $\hat{\mathcal{D}}_k$ für $k \in K$ unabhängig sind. Da $\hat{\mathcal{D}}_k \supset \mathcal{D}_j$ für alle $j \in I_k$, gilt

$$\sigma\Big(\bigcup_{j\in I_k} \mathcal{D}_j\Big) \subset \sigma\Big(\hat{\mathcal{D}}_k\Big)\,.$$

Also folgt die Behauptung aus Satz 5.5. □

Definition 5.8 Es sei $(\mathcal{A}_n)_n$ eine Folge von σ-Algebren von Ereignissen aus $\mathcal{A}$ und

$$\mathcal{T}_n := \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{A}_m\Big).$$

Dann heißt

$$\mathcal{T}_\infty := \bigcap_{n=1}^{\infty} \mathcal{T}_n$$

die σ-Algebra der *terminalen Ereignisse* der Folge $(\mathcal{A}_n)_n$.

Satz 5.9 (Null-Eins-Gesetz von Kolmogorov) *Es sei $(\mathcal{A}_n)_n$ eine unabhängige Folge von σ-Algebren $\mathcal{A}_n \subset \mathcal{A}$. Dann gilt*

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

für $A \in \mathcal{T}_\infty$.

Beweis: Nach Korollar 5.7 ist $\mathcal{T}_{n+1}$ unabhängig von $\sigma\big(\bigcup_{m=1}^{n} \mathcal{A}_m\big)$ und somit ist $\mathcal{T}_\infty$ unabhängig von $\sigma\big(\bigcup_{m=1}^{n} \mathcal{A}_m\big)$ (Lemma 5.3) für alle $n \in \mathbb{N}$. Dann ist $\mathcal{T}_\infty$ unabhängig von

$$\bigcup_{n=1}^{\infty} \sigma\Big(\bigcup_{m=1}^{n} \mathcal{A}_m\Big)$$

nach 5.3. Eine Vereinigung von aufsteigenden Mengen ist durchschnittstabil, also folgt mit Satz 5.5, dass $\mathcal{T}_\infty$ unabhängig ist von

$$\sigma\Big(\bigcup_{n=1}^{\infty} \sigma(\bigcup_{m=1}^{n} \mathcal{A}_m)\Big) = \sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{A}_n\Big).$$

Natürlich ist $\mathcal{T}_n \subset \sigma\big(\bigcup_{n=1}^{\infty} \mathcal{A}_n\big)$ für alle $n \in \mathbb{N}$, also auch

$$\mathcal{T}_\infty \subset \sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{A}_n\Big),$$

also ist nach Lemma 5.3 $\mathcal{T}_\infty$ unabhängig zu sich selbst! Das heißt für $A \in \mathcal{T}_\infty$ gilt

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

also $P(A) \in \{0,\, 1\}$. □

Korollar 5.10 (Null-Eins-Gesetz von Borel) *Für jede unabhängige Folge $(A_n)_n$ von Ereignissen aus $\mathcal{A}$ gilt*

$$P(\limsup_{n\to\infty} A_n) = 0 \ \textit{oder}\ = 1\,.$$

Beweis: Nach Satz 5.5 ist $\mathcal{A}_n := \sigma\big(\{A_n\}\big)$ eine unabhängige Folge. Es gilt $Q_n := \bigcup_{m=n}^{\infty} A_m \in \mathcal{T}_n$, sogar $Q_m \in \mathcal{T}_n$ für jedes $m \geq n$, $m \in \mathbb{N}$.

$$\limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} Q_k = \bigcap_{k=j}^{\infty} Q_k \in \mathcal{T}_j$$

für alle $j \in \mathbb{N}$, da $(Q_n)_n$ antiton ist, also ist $\limsup A_n \in \mathcal{T}_\infty$. □

Aus dem Lemma von Borel-Cantelli, 1.12, wissen wir

$$\sum_{n\geq 1} P(A_n) < \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 0\,.$$

Die Divergenz von $\sum P(A_n)$ führt im Allgemeinen nicht zum Schluss $P(\limsup A_n) = 1$. Wählt man nämlich ein $A_0 \in \mathcal{A}$ mit $0 < P(A_0) < 1$ und $(A_n)_n$ als konstante Folge $A_0, A_0, \ldots$, dann divergiert $\sum P(A_n)$, aber $P(\limsup A_n) = P(A_0) < 1$.

Nimmt man unabhängige $(A_n)_n$, so gilt die Umkehrung von 1.12, was auch von Borel und Cantelli bewiesen wurde. Es genügt, paarweise Unabhängigkeit zu fordern, was auf Erdős und Rényi zurückgeht:

Satz 5.11 (von Borel-Cantelli, Erdős-Rényi) *Sei $(A_n)_n$ eine Folge von Ereignissen in einem W-Raum $(\Omega, \mathcal{A}, P)$. Dann gilt:*

$$\sum_{n\geq 1} P(A_n) < \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 0\,.$$

Sind die Ereignisse wenigstens paarweise unabhängig, so gilt

$$\sum_{n\geq 1} P(A_n) = \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 1\,.$$

Beweis: Es sei $A := \limsup A_n$. Der erste Teil ist Lemma 1.12. Den zweiten Teil beweisen wir zunächst für unabhängige Ereignisse, weil der Beweis klassisch und kurz ist.

Mit der Stetigkeit von W-Maßen folgt

$$P(A^c) = P(\liminf A_n^c) = \lim_{n\to\infty} P\Big(\bigcap_{k\geq n} A_k^c\Big)\,.$$

Die $(A_n^c)_n$ sind unabhängig und $\sum_{k\geq n} P(A_k) = \infty$ für alle $n \in \mathbb{N}$, also

$$P(A^c) = \lim_{n\to\infty} \prod_{k\geq n} P(A_k^c) = \lim_{n\to\infty} \exp\Big(\sum_{k\geq n} \log\big(1 - P(A_k)\big)\Big)\,.$$

Für $x \in [0,\,1]$ gilt $\log(1-x) \leq -x$, also

$$P(A^c) \leq \lim_{n\to\infty} \exp\Big(-\sum_{k\geq n} P(A_k)\Big) = 0\,.$$

Im Fall paarweise unabhängiger Ereignisse $(A_n)_n$ setzen wir

$$I_n := 1_{A_n}\,, \quad S_n = \sum_{j=1}^n I_j \quad \text{und} \quad S := \sum_{n=1}^\infty I_n\,.$$

Die I_n sind nach Voraussetzung paarweise unkorreliert. Weiter ist $I_n^2 = I_n$. Also ist

$$\begin{aligned}\mathrm{Var}(S_n) &= \sum_{j=1}^n \mathrm{Var}(I_j) = \sum_{j=1}^n \big(\mathbb{E}(I_j^2) - \mathbb{E}(I_j)^2\big)\\ &= \mathbb{E}(S_n) - \sum_{j=1}^n \mathbb{E}(I_j)^2 \le \mathbb{E}(S_n)\,.\end{aligned}$$

Die Voraussetzung besagt $\sum_{n=1}^\infty \mathbb{E}(I_n) = +\infty$ und daher folgt wegen $S_n \uparrow S$

$$\lim_{n\to\infty} \mathbb{E}(S_n) = \mathbb{E}(S) = +\infty\,. \tag{5.3}$$

Ein Element $\omega \in \Omega$ liegt genau dann in A, also in A_n für unendlich viele n, wenn $S(\omega) = \infty$ ist. Zu zeigen ist also $P(S = +\infty) = 1$.
Nach Tschebyschev ist

$$P\big(|S_n - \mathbb{E}(S_n)| \le \eta\big) \ge 1 - \frac{\mathrm{Var}(S_n)}{\eta^2}$$

für $\eta > 0$. Mit (5.3) kann $\mathbb{E}(S_n) > 0$ für alle $n \in \mathbb{N}$ angenommen werden. Es folgt

$$\begin{aligned}P\big(S_n \ge (1/2)\mathbb{E}(S_n)\big) &\ge P\big(|S_n - \mathbb{E}(S_n)| \le (1/2)\mathbb{E}(S_n)\big)\\ &\ge 1 - 4\frac{\mathrm{Var}(S_n)}{\mathbb{E}(S_n)^2}\,.\end{aligned}$$

Es ist $\lim_{n\to\infty} \mathrm{Var}(S_n)/\mathbb{E}(S_n)^2 = 0$, und somit

$$P\big(S_n \ge (1/2)\mathbb{E}(S_n)\big) \ge 1 - \varepsilon$$

für jedes $\varepsilon > 0$ und schließlich alle n.
Da $S_n \le S$, folgt

$$P\big(S \ge (1/2)\mathbb{E}(S_n)\big) \ge P\big(S_n \ge (1/2)\mathbb{E}(S_n)\big) \ge 1 - \varepsilon$$

für schließlich alle n. Nach (5.3) gilt $\mathbb{E}(S_n) \uparrow \mathbb{E}(S) = +\infty$, also

$$P(S = +\infty) \ge 1 - \varepsilon$$

für alle $\varepsilon > 0$, also $P(S = \infty) = 1$. □

Die Ereignisse $(A_i)_{i\in I}$ sind genau dann unabhängig, wenn die $\mathcal{A}_i = \{\Omega, \emptyset, A_i, A_i^c\}$, $i \in I$, unabhängig sind (Satz 5.5). Es gilt weiter $\sigma(1_{A_i}) = \mathcal{A}_i$. Dies legt die folgende Definition nahe:

Definition 5.12 Eine Familie $(X_i)_{i\in I}$ von Zufallsvariablen auf einem W-Raum $(\Omega, \mathcal{A}, P)$ mit Werten in $(\Omega_i, \mathcal{A}_i)$ heißt *unabhängig*, wenn die Familie

$$\big(\sigma(X_i)\big)_{i\in I} = \big(X_i^{-1}(\mathcal{A}_i)\big)_{i\in I}$$

von σ-Algebren unabhängig ist.

Satz 5.13 *Für jedes $i = 1, \dots, n$ sei*

$$X_i : (\Omega, \mathcal{A}, P) \to (\Omega_i, \mathcal{A}_i)$$

eine Zufallsvariable und $\mathcal{E}_i$ ein durchschnittstabiler Erzeuger von $\mathcal{A}_i$ mit $\Omega_i \in \mathcal{E}_i$. Die $X_1, \dots, X_n$ sind genau dann unabhängig, wenn

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

für jede Auswahl $A_i \in \mathcal{E}_i$ gilt ($i = 1, \dots, n$).

Beweis: $X_i^{-1}(\mathcal{E}_i)$ ist ein durchschnittstabiler Erzeuger von $\sigma(X_i)$, der $\Omega = X_i^{-1}(\Omega_i)$ enthält. Die Behauptung folgt dann aus Satz 5.5. □

Korollar 5.14 *Eine Familie von Zufallsgrößen $(X_i)_{i\in I}$ ist genau dann unabhängig, wenn für alle $n \in \mathbb{N}$, $i_1, \dots, i_n \in I$ und $t_1, \dots, t_n \in \mathbb{R}$*

$$P(X_{i_1} \le t_1, \dots, X_{i_n} \le t_n) = \prod_{j=1}^{n} P(X_{i_j} \le t_j)$$

gilt.

Beweis: Dies folgt aus Definition 5.1, Satz 5.13 und der Tatsache, dass

$$\Big\{X_i^{-1}\big((-\infty, t]\big),\, t \in \mathbb{R}\Big\} \cup \Omega$$

ein durchschnittstabiles Erzeugendensystem von $X_i^{-1}(\mathcal{B})$ ist. □

Beispiel 5.15 (siehe Beispiel 5.4 (b)) Die Folge $X_n := 1_{A_n}$ der Rademacher-Funktionen X_n ist nach $b(1, \frac{1}{2})$ verteilt. Also ist die Folge $(X_n)_n$ konvergent in Verteilung. Weiter gilt:

$$\begin{aligned} P(|X_m - X_n| \ge \delta) &= P(X_m = 1,\, X_n = 0) + P(X_m = 0,\, X_n = 1) \\ &= P(X_m = 1)P(X_n = 0) + P(X_m = 0)P(X_n = 1) \\ &= \frac{1}{2} \end{aligned}$$

für alle $n \ne m$ und δ mit $0 < \delta < 1$.

Mit Satz 4.14 und 4.13 folgt: Die Folge $(X_n)_n$ kann keine stochastisch konvergente Teilfolge enthalten!

Satz 5.16 *Sei $(X_i)_{i\in I}$ eine unabhängige Familie $(\Omega_i, \mathcal{A}_i)$-wertiger Zufallsvariablen und*

$$f_i : (\Omega_i, \mathcal{A}_i) \to (\Omega_i', \mathcal{A}_i')$$

für jedes $i \in I$ eine messbare Abbildung. Dann ist auch die Familie $(f_i \circ X_i)_{i\in I}$ unabhängig.

Beweis: Für $A' \in \mathcal{A}_i'$ ist $(f_i \circ X_i)^{-1}(A') = X_i^{-1}\big(f_i^{-1}(A')\big)$ und somit $\sigma(f_i \circ X_i) \subset \sigma(X_i)$. Mit $\big(\sigma(X_i)\big)_{i\in I}$ ist daher auch $\big(\sigma(f_i \circ X_i)\big)_{i\in I}$ unabhängig. □

Die Unabhängigkeit von Zufallsvariablen ist eine wahrscheinlichkeitstheoretische Eigenschaft, also eine Eigenschaft ihrer Verteilungen:

Satz 5.17 *Eine Familie von Zufallsvariablen $(X_i)_{i\in I}$ auf einem W-Raum $(\Omega, \mathcal{A}, P)$ mit Werten in $(\Omega_i, \mathcal{A}_i)$ ist genau dann unabhängig, wenn ihre Verteilung die Produktverteilung ihrer Komponenten P^{X_i} ist:*

$$P^{(X_i)_{i\in I}} = \bigotimes_{i\in I} P^{X_i} .$$

($X_I = (X_i)_{i\in I}$ ist eine messbare Abbildung von $(\Omega, \mathcal{A})$ nach $(\prod \Omega_i, \bigotimes \mathcal{A}_i)$).

Beweis: Für jedes $J \subset I$ sei $X_J = (X_i)_{i\in J}$ und p_J die Projektion auf die Komponenten mit Index in J: $p_J \circ X_I = X_J$. Nach Satz 3.11 ist P^{X_I} genau dann das Produktmaß der P^{X_i}, $i \in I$, wenn für jedes $J = \{j_1, \ldots, j_n\} \in \mathcal{H}(I)$

$$P^{X_J} = p_J(P^{X_I}) = \bigotimes_{j\in J} P^{X_j}$$

gilt, also

$$P(X_{j_1} \in A_1, \ldots, X_{j_n} \in A_n) = \prod_{k=1}^{n} P^{X_{j_k}}(A_k) = \prod_{k=1}^{n} P(X_{j_k} \in A_k)$$

für messbare $A_1, \ldots, A_n$ gilt. Dies ist nach Satz 5.13 zur Unabhängigkeit der $(X_j)_{j\in J}$ und damit zur Unabhängigkeit der ganzen Familie $(X_i)_{i\in I}$ äquivalent. □

Korollar 5.18 *Zu jeder Familie $\big((\Omega_i, \mathcal{A}_i, P_i)\big)_{i\in I}$ von W-Räumen existiert eine unabhängige Familie $(X_i)_{i\in I}$ von $(\Omega_i, \mathcal{A}_i)$-wertigen Zufallsvariablen auf einem geeigneten W-Raum $(\Omega, \mathcal{A}, P)$, so dass für jedes $i \in I$ gilt $P_i = P^{X_i}$.*

Beweis: Sei $(\Omega, \mathcal{A}, P) := \bigotimes_{i\in I}(\Omega_i, \mathcal{A}_i, P_i)$ und X_i die i-te Projektionsabbildung. Dann ist $(X_i)_{i\in I}$ die identische Abbildung auf Ω und hat $P = \bigotimes_{i\in I} P_i$ als Verteilung, was die Unabhängigkeit unter P beweist. □

Beispiel 5.19 $\Omega_0 = \{0, 1\}$, $\mathcal{A}_0 = \mathcal{P}(\Omega_0)$, $A = \{1\}$, $P_0(A) = p$, $P_0(A^c) = q := 1 - p$, $0 \le p \le 1$.

Dann ist der BERNOULLI-Versuch gegeben durch den W-Raum

$$\Omega = \Omega_0^{\mathbb{N}}, \quad \mathcal{A} = \mathcal{A}_0^{\mathbb{N}}, \quad P := P_0^{\mathbb{N}},$$

und besteht aus abzählbar oft unabhängigen Wiederholungen. Es sei

$$X_n(\omega) := \omega_n \quad \text{für } \omega = (\omega_n)_n \in \Omega.$$

Wir berechnen die Wahrscheinlichkeit dafür, dass unendlich oft zweimal hintereinander Kopf geworfen wird, bei einer fairen Münze $p = q = \frac{1}{2}$. A_n sei das Ergebnis, dass beim n-ten und beim $(n+1)$-ten Wurf Kopf fällt. Dann ist $P(A_n) = \frac{1}{4}$ und

$$\sum_{n=1}^{\infty} P(A_{2n}) = +\infty.$$

$A := \limsup A_n$ interessiert uns. Es gilt $P(A) = 1$, denn $(A_{2n})_n$ ist eine Folge paarweise unabhängiger Ereignisse (sogar unabhängig) und

$$\limsup_{n\to\infty} A_{2n} \subset \limsup_{n\to\infty} A_n,$$

also wenden wir Satz 5.11 an.

Sei $(X_n)_n$ nun der Bernoulli-Versuch wie in Beispiel 5.19. Dann ist $S_n := X_1 + \cdots + X_n$ $b(n,p)$-verteilt, denn mit der Unabhängigkeit ist $P((X_1,\dots,X_n) = \omega) = p^k(1-p)^{n-k}$, wenn $S_n = k$ ist. Dann ist

$$\mathbb{E}(S_n) = n\mathbb{E}(X_1) = n\big(p \cdot 1 + (1-p) \cdot 0\big) = np$$

und dies ist eine deutlich schönere Herleitung als die bisher gegebene in Beispiel 2.19(a).

Satz 5.9 für von Zufallsvariablen erzeugte σ-Algebren besagt: Ist $(X_n)_n$ eine unabhängige Folge von Zufallsvariablen. Dann gilt für jedes terminale Ereignis

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_m\,;\, m \geq n) =: \mathcal{T}_\infty$$

entweder $P(A) = 0$ oder $P(A) = 1$. Wir betrachten ein Korollar dazu:

Korollar 5.20 *Es sei $(X_n)_{n\geq 1}$ eine unabhängige Folge reeller Zufallsvariablen. Dann ist jede $\mathcal{T}_\infty$-messbare numerische Zufallsvariable T fast sicher konstant, d.h. es existiert ein $\alpha \in \bar{\mathbb{R}}$ mit*

$$P(T = \alpha) = 1.$$

(T heißt manchmal terminale Funktion*)*

Beweis: Sei $\gamma \in \bar{\mathbb{R}}$, dann ist $\{T \leq \gamma\} \in \mathcal{T}_\infty$ und somit $P(T \leq \gamma) = 0$ oder 1. Für $\gamma = +\infty$ ist $P(T \leq \gamma) = P(\Omega) = 1$.

Es sei α das Infimum in $\bar{\mathbb{R}}$ der somit nichtleeren Menge C aller $\gamma \in \bar{\mathbb{R}}$ mit $P(T \leq \gamma) = 1$. Dann gilt $\gamma_n \downarrow \alpha$ für eine geeignete antitone Folge $(\gamma_n)_n$ in C und mit $\{T \leq \gamma_n\} \downarrow \{T \leq \alpha\}$ ist $\alpha \in C$. α ist also das kleinste Element von C. Hieraus folgt $P(T < \alpha) = 0$ und $P(T = \alpha) = 1$. $\square$

Wir sammeln noch ein paar Rechenregeln:

Satz 5.21 (Multiplikationssatz) *Seien $X_1, \ldots, X_n$ unabhängige reelle Zufallsvariablen. Dann gilt*

$$\mathbb{E}\Big(\prod_{i=1}^n X_i\Big) = \prod_{i=1}^n \mathbb{E}(X_i)$$

wenn alle $X_i \geq 0$ oder alle X_i integrierbar sind. Im zweiten Fall ist auch $\prod X_i$ integrierbar. Ist umgekehrt $\prod X_i$ integrierbar und verschwindet kein X_i fast sicher, so ist auch jedes X_i integrierbar

Beweis: Nach Satz 5.17 ist $Q = \bigotimes_{i=1}^n P^{X_i}$ die gemeinsame Verteilung der $X_1, \ldots, X_n$. Es gilt

$$\begin{aligned}
\mathbb{E}\Big(\big|\prod_{i=1}^n X_i\big|\Big) &= \int |x_1 \cdots x_n| \, Q(dx) \\
&= \int \cdots \int |x_1| \cdots |x_n| \, P^{X_1}(dx_1) \cdots P^{X_n}(dx_n) \\
&= \prod_{i=1}^n \int |x_i| \, P^{X_i}(dx_i) = \prod_{i=1}^n \mathbb{E}(|X_i|)
\end{aligned}$$

nach Satz 1.23 und Satz 3.5 (Fubini).
Also folgt die Behauptung für $X_i \geq 0$ und die Integrierbarkeit von $\prod X_i$ für den Fall, dass alle X_i integrierbar sind. Nach Fubini bleibt dann die obige Rechnung richtig, wenn die Absolut-Striche fehlen.
Gilt $\mathbb{E}\Big(\big|\prod_{i=1}^n X_i\big|\Big) < \infty$ und $\mathbb{E}(|X_i|) > 0$, $i = 1, \ldots, n$, so ist

$$\prod_{i=1}^n \mathbb{E}(|X_i|) = \mathbb{E}\Big(\big|\prod_{i=1}^n X_i\big|\Big) < \infty$$

und kein Faktor Null, also $\mathbb{E}(|X_i|) < \infty$, d.h. jedes X_i ist integrierbar. □

Korollar 5.22 *Zwei unabhängige Zufallsgrößen X und Y mit endlichem Erwartungswert sind unkorreliert.*

Beweis: Da X und Y unabhängig sind, sind auch $X - \mathbb{E}(X)$ und $Y - \mathbb{E}(Y)$ unabhängig (Satz 5.16) und ihr Erwartungswert ist Null. Somit folgt mit 5.21 die Behauptung. □

Im Fall des Bernoulli-Versuchs sind die $(X_n)_n$ also paarweise unkorreliert, also $\mathrm{Var}(S_n) = n \,\mathrm{Var}(X_1)$ (siehe Definition 2.13 und Folgerung). Nun gilt $\mathrm{Var}(X_1) = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2$ nach Satz 2.7 und hier ist $\mathbb{E}X_1^2 = \mathbb{E}X_1 = p$, also $\mathrm{Var}(X_1) = p - p^2 = p(1-p)$, also $\mathrm{Var}(S_n) = np(1-p)$, wieder eine deutlich schönere Herleitung als die bisher gegebene in Beispiel 2.19(a).

Wir untersuchen die Verteilung der Summe zweier unabhängiger Zufallsgrößen X, Y.

Definition 5.23 Die *Faltung* zweier W-Maße P_1 und P_2 auf $(\mathbb{R}, \mathcal{B})$ ist das Bildmaß

$$P_1 * P_2 := (P_1 \otimes P_2) \circ S^{-1},$$

wobei $S : \mathbb{R}^2 \to \mathbb{R}$ definiert ist als $S(x_1, x_2) = x_1 + x_2$ (kann analog für allgemeine Bildräume definiert werden).

X und Y seien Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$, die unabhängig sind. Seien $\mu = P^X$ und $\nu = P^Y$, dann ist

$$P(X + Y < t) = \int_{\mathbb{R}^2} f_t(x, y)\,(\mu \otimes \nu)\,d(x, y)$$

mit

$$f_t(x, y) = 1_{\{x+y<t\}} = 1_{(-\infty, t-y)}(x),$$

da $\mu \otimes \nu$ die Verteilung von (X, Y) ist. Nach FUBINI ist

$$\begin{aligned} P(X + Y < t) &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} 1_{(-\infty, t-y)}(x)\,\mu(dx) \Big) \nu(dy) \\ &= \int_{\mathbb{R}} F_X(t - y)\,\nu(dy) \end{aligned}$$

wobei F_X die Verteilungsfunktion von X ist.

Satz 5.24 *X und Y seien unabhängige Zufallsgrößen mit Lebesgue-Dichten f und g. Dann hat $X + Y$ die Lebesgue-Dichte*

$$h_{X+Y}(z) = \int f(y)\,g(z - y)\,\lambda(dy) = \int f(z - y)\,g(y)\,\lambda(dy),$$

XY die Lebesgue-Dichte

$$h_{XY}(z) = \int \frac{1}{|y|} f\Big(\frac{z}{y}\Big)\,g(y)\,\lambda(dy) = \int \frac{1}{|y|} f(y)\,g\Big(\frac{z}{y}\Big)\,\lambda(dy)$$

und X/Y die Lebesgue-Dichte

$$h_{X/Y}(z) = \int |y|\,f(zy)\,g(y)\,\lambda(dy) = \frac{1}{z^2} \int |y|\,f(y)\,g\Big(\frac{y}{z}\Big)\,\lambda(dy),$$

wobei $P(Y \neq 0) = 1$ durch die λ-Stetigkeit von Y garantiert wird.

Beweis: Wir beweisen nur die Formel für h_{X+Y}, der Rest ist eine Übung.

$$\begin{aligned} P(X+Y<t) &= \int_{\mathbb{R}} \Big(\int_{(-\infty,\, t-y)} f(x)\,\lambda(dx) \Big) g(y)\,\lambda(dy) \\ &= \int_{\mathbb{R}} \Big(\int_{(-\infty,\, t)} f(x-y)\,\lambda(dx) \Big) g(y)\,\lambda(dy) \\ &= \int_{(-\infty,\, t)} \Big(\int_{\mathbb{R}} f(x-y)\, g(y)\,\lambda(dy) \Big) \lambda(dx)\,. \end{aligned}$$

□

KAPITEL 6

Starkes Gesetz der großen Zahlen

Im Bernoulli Experiment aus Beispiel 5.19 nimmt $S_n := \sum_{i=1}^n X_i$ nur die Werte $0, 1, \dots, n$ an und gibt die Zahl der Erfolge bei den ersten n Ausführungen an. Die relative Häufigkeit $n^{-1}S_n$ sollte mit großer Wahrscheinlichkeit gegen p streben. Dieses wage Gefühl soll nun präzisiert werden.

Für $\omega = (0, 0, \dots)$ bzw. $\omega = (1, 1, \dots)$ ist $n^{-1}S_n(\omega) = 0$ bzw. $n^{-1}S_n(\omega) = 1$ für alle $n \in \mathbb{N}$, also konvergiert $n^{-1}S_n(\omega)$ für $n \to \infty$ offenbar nicht für jedes $\omega \in \Omega$. Konvergiert diese Größe stochastisch oder gar fast sicher? Im Bernoulli Experiment hat X_n die Verteilung $b(1, p)$ und den Erwartungswert $\mathbb{E}(X_n) = p$. Wir fragen also:

Gilt

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \big(X_i - \mathbb{E}(X_i)\big) = 0 \tag{6.1}$$

im Sinne der stochastischen bzw. der fast sicheren Konvergenz bzgl. P? Man sagt, dass eine Folge $(X_n)_n$ integrierbarer reeller Zufallsvariablen dem *schwachen* bzw. dem *starken Gesetz der großen Zahlen* genügt, wenn (6.1) im Sinne der stochastischen bzw. der P-fast sicheren Konvergenz gilt. Bei einer beliebigen Folge identisch verteilter Zufallsgrößen $(X_n)_n$ ist $\mathbb{E}(X_1) = \mathbb{E}(X_n)$ für alle $n \in \mathbb{N}$, wenn $\mathbb{E}(X_1)$ existiert, und (6.1) wird zu

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}(X_1) \ . \tag{6.2}$$

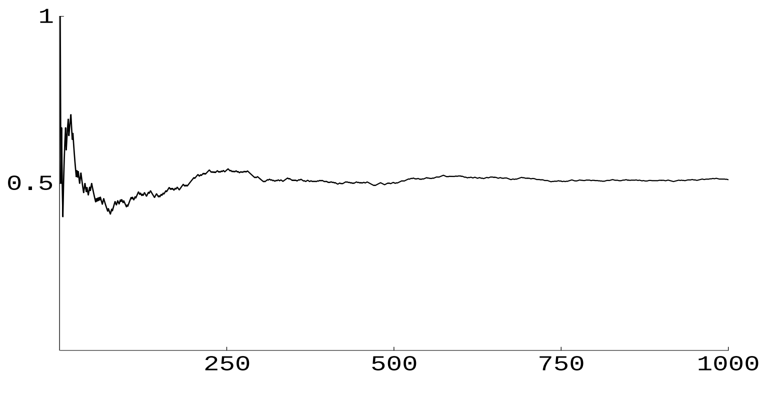


ABBILDUNG 6.1. Eine Simulation von $n^{-1}S_n$ bei der fairen Münze.

Man sagt daher allgemeiner, dass eine Folge reeller Zufallsvariablen $(X_n)_n$ dem schwachen bzw. dem starken Gesetz der großen Zahlen genügt, wenn (6.2) gilt für $\mathbb{E}(X_1)$ ersetzt durch eine reelle Zahl $\mu \in \mathbb{R}$ (wieder im Sinne der stochastischen bzw. der P-fast sicheren Konvergenz). Natürlich folgt aus der Gültigkeit eines starken Gesetzes die des korrespondierenden schwachen Gesetzes, siehe Satz 4.10.

Dem schwachen Gesetz hatten wir uns in Satz 2.21 bereits gewidmet. Der dort gegebene Beweis führt unmittelbar zu

Satz 6.1 (von KHINTCHINE) *Gilt für eine Folge $(X_n)_n$ integrierbarer und paarweise unkorrelierter reeller Zufallsvariablen*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = 0 \ ,$$

so genügt die Folge dem schwachen Gesetz der großen Zahlen.

Beweis: Aus der Voraussetzung folgt, dass alle X_i quadratisch integrierbar sind. Es gilt

$$\mathrm{Var}\Big(\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i),$$

also

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))\Big) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i), \ n \in \mathbb{N}.$$

Die Behauptung folgt mit der Tschebyschev-Ungleichung (Satz 2.8(ii)). □

Kommt man auch ohne quadratische Integrierbarkeit aus? Wir notieren hier das folgende Resultat ohne Beweis:

Satz 6.2 *Sind die $(X_n)_n$ unabhängig und identisch verteilt, so genügt $(X_n)_n$ (ohne der Annahme der Integrierbarkeit !) genau dann dem schwachen Gesetz der großen Zahlen mit Limes* 0*, wenn*

$$\lim_{n\to\infty} n\, P\big(|X_1| > n\big) = 0$$

und

$$\lim_{n\to\infty} \int_{\{|X_1|\le n\}} X_1 \, dP = 0$$

gilt.

Für einen Beweis siehe zum Beispiel das Buch von Gänssler und Stute, Wahrscheinlichkeitstheorie, Satz 2.1.11.

Das 0-1-Gesetz, Satz 5.9, lässt die Frage nach der Gültigkeit des starken Gesetzes der großen Zahlen in der folgenden Sicht erscheinen: Es sei $(X_n)_n$ eine Folge unabhängiger Zufallsgrößen, $\mathcal{T}_\infty$ definiert wie in 5.20 und $S_n := \sum_{j=1}^{n} X_j$. Dann gilt:

Lemma 6.3 *Sei $(\tau_n)_n$ eine Nullfolge reeller Zahlen, dann sind $\liminf_{n\to\infty} \tau_n S_n$ und $\limsup_{n\to\infty} \tau_n S_n$ $\mathcal{T}_\infty$-messbare Zufallsgrößen, also fast sicher konstant (Konstante in $[-\infty, +\infty]$).*

Beweis: Für $m \in \mathbb{N}$ gilt

$$\limsup_{n\to\infty} \tau_n S_n = \limsup_{n\to\infty} \tau_n \Big(\sum_{j=1}^m X_j + \sum_{j=m+1}^n X_j \Big) = \limsup_{n\to\infty} \tau_n \Big(\sum_{j=m+1}^n X_j \Big) .$$

Die Zufallsgröße auf der rechten Seite ist $\sigma(X_n, n \geq m+1)$-messbar für jedes $m \in \mathbb{N}$, also $\mathcal{T}_\infty$-messbar. Für $\liminf_{n\to\infty}$ folgt die Aussage analog. Mit Korollar 5.20 folgt die Behauptung. □

Es sei nun $A = \{\omega \in \Omega : \lim_{n\to\infty} \tau_n S_n(\omega) = 0\}$. Dann ist

$$A = \{\liminf_{n\to\infty} \tau_n S_n(\omega) = 0\} \cap \{\liminf_{n\to\infty} \tau_n S_n = 0\}$$

ein Ereignis (siehe Bemerkung 2.3) und nach Lemma 6.3 ist $A \in \mathcal{T}_\infty$. Also ist $P(A)$ gleich 0 oder 1. Ist jedes X_n integrierbar, so ist $(X_n - \mathbb{E}(X_n))_n$ auch eine unabhängige Folge und für $\tau_n = \frac{1}{n}$ gilt somit, dass

$$P\Big(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) = 0 \Big)$$

entweder 0 oder 1 ist. Das starke Gesetz der großen Zahlen gilt dann, wenn diese Wahrscheinlichkeit 1 ist. Tatsächlich kann das starke Gesetz der großen Zahlen für eine große Klasse von Folgen von Zufallsgrößen bewiesen werden:

Satz 6.4 (von Etemadi, 1981) *Jede Folge $(X_n)_n$ reeller, identisch verteilter und paarweise unabhängiger Zufallsvariablen genügt genau dann dem starken Gesetz der großen Zahlen mit Limes μ, wenn X_1 integrierbar ist mit $\mu = \mathbb{E}(X_1)$.*

Kolmogorov publizierte 1930 das starke Gesetz für unabhängige Zufallsvariablen:

Korollar 6.5 (von Kolmogorov, 1930) *Jede unabhängige Folge identisch verteilter, integrierbarer reeller Zufallsvariablen genügt dem starken Gesetz der großen Zahlen.*

Beweis: (des Satzes von Etemadi) „$\Rightarrow$“: Sei $S_n := \sum_{j=1}^n X_j$ und $\frac{S_n}{n} \to \mu$ f.s. für ein $\mu \in \mathbb{R}$. Es folgt

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \to 0 \text{ f.s.}$$

und somit

$$P\big(|X_n| > n \text{ unendlich oft}\big) = P\big(\limsup_{n\to\infty} \{|X_n| > n\}\big) = 0 .$$

Also konvergiert $\sum_{n\geq 1} P(|X_n| > n)$ nach Borel-Cantelli (Satz 5.11), denn die X_n sind paarweise unabhängig. Da die X_n auch identisch verteilt sind, folgt nach Beispiel 3.10

$$\mathbb{E}|X_1| \leq 1 + \sum_{n\geq 1} P(|X_1| > n) = 1 + \sum_{n\geq 1} P(|X_n| > n) < \infty .$$

„$\Leftarrow$": Natürlich ist dies das Herzstück des Satzes. Es wird behauptet: $\lim_{n\to\infty} \frac{1}{n} S_n = \mu = \mathbb{E}(X_1)$ P-fast sicher. Mit $(X_n)_n$ genügen auch $(X_n^+)_n$ und $(X_n^-)_n$ den Voraussetzungen des Satzes. Es kann daher ohne Einschränkung $X_n \geq 0$ für alle $n \in \mathbb{N}$ angenommen werden. Wir betrachten die gestutzte Folge

$$Y_n := X_n 1_{\{X_n \leq n\}} , \ n \in \mathbb{N} .$$

Es geht die ursprünglich gemeinsame Verteilung verloren. Wir gewinnen aber die quadratische Integrierbarkeit, denn

$$\mathbb{E}(Y_n^2) = \int_{[0,n)} x^2 \, P^{X_1}(x) < \infty .$$

Es genügt zu zeigen, dass $\frac{1}{n}\sum_{i=1}^n Y_i$ fast sicher gegen μ konvergiert. Dazu betrachte

$$\begin{aligned}\sum_{n\geq 1} P(X_n \neq Y_n) &= \sum_{n\geq 1} P(X_n > n) = \sum_{n\geq 1} P(X_1 > n) \\ &\leq \mathbb{E}X_1 < \infty,\end{aligned}$$

wobei wir wieder Beispiel 3.10 verwendet haben. Also folgt nach Borel-Cantelli $P(X_n \neq Y_n \text{ unendlich oft}) = 0$. Somit hat das Ereignis A aller $\omega \in \Omega$ mit $X_n(\omega) \neq Y_n(\omega)$ für höchstens endlich viele $n \in \mathbb{N}$ Wahrscheinlichkeit 1. Also kann aus der fast sicheren Konvergenz von $\frac{1}{n}\sum Y_i$ gegen μ auf die fast sichere Konvergenz von $\frac{1}{n}\sum X_i$ gegen μ geschlossen werden.

Es sei nun $T_n := \frac{1}{n}\sum_{i=1}^n Y_i$, $n \in \mathbb{N}$. Zu $\alpha > 1$ setzen wir $k_n := [\alpha^n] := \sup\{m \in \mathbb{N}_0 : \ m \leq \alpha^n\}$. Zu $\varepsilon > 0$ liefert die Tschebyschev-Ungleichung

$$\sum_{n\geq 1} P(|T_{k_n} - \mathbb{E}(T_{k_n})| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{n\geq 1} \frac{1}{k_n^2} \sum_{i=1}^{k_n} \mathrm{Var}(Y_i) ,$$

denn die $(Y_i)_i$ sind paarweise unabhängig, also paarweise unkorreliert. Wir nutzen $\mathrm{Var}(Y_i) \leq \mathbb{E}(Y_i^2)$ und vertauschen mit Fubini die Summationsreihenfolge (alle Summanden sind nicht negativ) und erhalten:

$$\sum_{n\geq 1} P(|T_{k_n} - \mathbb{E}(T_{k_n})| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \mathbb{E}(Y_i^2) \sum_{n : k_n \geq i} \frac{1}{k_n^2}.$$

Es sei n_i die kleinste natürliche Zahl n mit $k_n \geq i$. Wir schätzen $\sum_{n=n_i}^{\infty} \frac{1}{k_n^2}$ ab: Aus $[\alpha^n] \geq \alpha^n/2$ für alle $n \geq 1$ folgt

$$\sum_{n=n_i}^{\infty} \frac{1}{k_n^2} \leq 4 \sum_{n=n_i}^{\infty} \alpha^{-2n} = 4\alpha^{-2n_i} \sum_{n=n_i}^{\infty} \alpha^{-2(n-n_i)} \leq \frac{4}{i^2} \frac{1}{(1-\alpha^{-2})} ,$$

und somit

$$\sum_{n\geq 1} P(|T_{k_n} - \mathbb{E}(T_{k_n})| > \varepsilon) \leq \frac{4}{\varepsilon^2(1-\alpha^{-2})} \sum_{i=1}^{\infty} \frac{\mathbb{E}(Y_i^2)}{i^2} \ .$$

Nun wollen wir die Reihe auf der rechten Seite untersuchen. Dazu verwenden wir Beispiel 3.10 wie folgt. Für $a > 0$ und jede nicht negative Zufallsgröße X ist

$$\mathbb{E}\big(X^2 1_{\{X\leq a\}}\big) = \int_0^a 2tP(X > t, X \leq a)\,dt \leq \int_0^a 2tP(X > t)\,dt.$$

Also erhalten wir

$$\begin{aligned}
\sum_{i=1}^{\infty} \frac{\mathbb{E}(Y_i^2)}{i^2} &\leq \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^i 2tP(X_1 > t)\,dt \\
&= \sum_{i=1}^{\infty} \sum_{k=1}^{i} \frac{1}{i^2} \int_{k-1}^{k} 2tP(X_1 > t)\,dt \\
&= \sum_{k\geq 1} \Big(\int_{k-1}^{k} 2tP(X_1 > t)\,dt \sum_{i\geq k} \frac{1}{i^2} \Big) \\
&\leq 2 + \sum_{k\geq 2} \frac{2k-1}{k-1} P(X_1 > k-1) \leq 2 + 3\mathbb{E}(X_1) < \infty.
\end{aligned}$$

Wir müssen nur noch die letzte Abschätzung begründen. Wir verwenden

$$\sum_{i\geq 1} \frac{1}{i^2} \leq 1 + \sum_{i\geq 1} \frac{1}{i(i+1)} = 2$$

und

$$\sum_{i\geq k} \frac{1}{i^2} \leq \sum_{i\geq k-1} \frac{1}{i(i+1)} = \frac{1}{k-1}$$

und für $k \geq 2$

$$\int_{k-1}^{k} 2tP(X_1 > t)\,dt \leq P(X > k-1) \int_{k-1}^{k} 2t\,dt = (2k-1)P(X > k-1).$$

Also konvergiert $T_{k_n} - \mathbb{E}(T_{k_n})$ f.s. gegen Null gemäß Satz 4.9.

Nach dem Satz von der monotonen Konvergenz folgt $\mathbb{E}(Y_i) = \mathbb{E}\big(X_1\, 1_{\{X_1\leq i\}}\big) \to \mu = \mathbb{E}(X_1)$, und somit $\mathbb{E}(T_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(Y_i) \to \mu$ (siehe Analysis). Also impliziert $T_{k_n} - \mathbb{E}(T_{k_n}) \to 0$ f.s. $T_{k_n} \to \mu$ fast sicher.

Zu untersuchen bleiben die Werte in $\mathbb{N}\setminus\{k_n, n \geq 1\}$. k_n geht isoton gegen $+\infty$, also gibt es zu jedem $m \in \mathbb{N}$ mit $m > k_1$ genau ein $n \in \mathbb{N}$ mit $k_n < m \leq k_{n+1}$. Da alle $X_i \geq 0$, folgt

$$\frac{k_n}{k_{n+1}} T_{k_n} \leq T_m \leq T_{k_{n+1}} \frac{k_{n+1}}{k_n} \ .$$

Beachte

$$\frac{1}{\alpha} - \frac{1}{\alpha^{n+1}} = \frac{\alpha^n - 1}{\alpha^{n+1}} \leq \frac{[\alpha^n]}{[\alpha^{n+1}]} \leq \frac{\alpha^n}{\alpha^{n+1} - 1} = \frac{1}{\alpha - \frac{1}{\alpha^n}},$$

also $\lim_{n\to\infty} \frac{k_{n+1}}{k_n} = \alpha$ für $\alpha > 1$. Somit folgt

$$\frac{\mu}{\alpha} \leq \liminf_{m\to\infty} T_m \leq \limsup_{m\to\infty} T_m \leq \alpha\mu$$

fast sicher. Da $\alpha > 1$ beliebig vorgegeben war, folgt $T_m \to \mu$ fast sicher, was zu zeigen war. $\square$

Korollar 6.6 *Es sei $(X_n)_n$ eine Folge identisch verteilter und paarweise unabhängiger Zufallsgrößen. Aus $\mathbb{E}(X_1) = \infty$ bzw. $-\infty$ folgt $n^{-1}S_n \to +\infty$ bzw. $-\infty$ fast sicher.*

Beweis: Wir zeigen $\mathbb{E}(X_1) = \infty \Rightarrow n^{-1}S_n \to \infty$ fast sicher:
Es sei $X_n^c := X_n 1_{\{X_n \leq c\}}$, $c > 0$. $(X_n^c)_n$ ist dann paarweise unabhängig, identisch verteilt und integrierbar. Es gilt weiter $S_n \geq S_n^c := \sum_{j=1}^n X_j^c$, $n \in \mathbb{N}$, $c > 0$, also mit Satz 6.4

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \lim_{n\to\infty} \frac{S_n^c}{n} = \mathbb{E}(X_1^c) \text{ f.s.}$$

für alle $c > 0$, und somit folgt mit $\mathbb{E}(X_1^c) \nearrow \mathbb{E}(X_1) = \infty$ das Gewünschte.

Das starke Gesetz kann auch für nicht notwendig identisch verteilte Zufallsvariablen hergeleitet werden. Der sogenannte „klassische“Weg zum starken Gesetz (Kolmogorovsches Kriterium, 1928) führt über eine stochastische Ungleichung:

Satz 6.7 (Kolmogorov-Ungleichung) *Es sei $(X_n)_n$ eine Folge unabhängiger Zufallsgrößen mit $\mathbb{E}(X_i) = 0$ für alle $i \in \mathbb{N}$. Dann gilt für jedes $\varepsilon > 0$*

$$P\big(\max_{1\leq k\leq n} |S_k| \geq \varepsilon\big) \leq \frac{1}{\varepsilon^2} \sum_{k=1}^n \operatorname{Var}(X_k) \ .$$

Beweis: Es sei $A_1 := \{|S_1| \geq \varepsilon\}$ und

$$A_{k+1} := \{|S_{k+1}| \geq \varepsilon \text{ und } \max_{1\leq l\leq k} |S_l| < \varepsilon\}.$$

Dann sind die A_k's disjunkt und $B_n := \{\max_{1\leq k\leq n} |S_k| \geq \varepsilon\} = \cup_{k=1}^n A_k$. Nun gilt für $1 \leq k < n$

$$S_n^2 - S_k^2 = (S_n - S_k)^2 + 2(S_n - S_k)S_k \geq 2(S_n - S_k)S_k \ .$$

Da $\mathbb{E}(S_n - S_k) = 0$ und $S_n - S_k$ unabhängig von S_k, folgt $\mathbb{E}(S_n^2 - S_k^2) \geq 0$, also $\mathbb{E}(S_n^2) \geq \mathbb{E}(S_k^2)$. Diese Ungleichung verwenden wir nun wie folgt:

$$\begin{aligned}\mathbb{E}(S_n^2 1_{B_n}) &= \sum_{k=1}^n \mathbb{E}(S_n^2 1_{A_k}) \\ &\geq \sum_{k=1}^n \mathbb{E}(S_k^2 1_{A_k}) \\ &\geq \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(B_n)\ .\end{aligned}$$

Beachte nun noch $\mathbb{E}(S_n^2) = \sum_{k=1}^n \mathbb{E}(X_k^2) = \sum_{k=1}^n \operatorname{Var}(X_k)$. Damit ist die Ungleichung bewiesen. □

Satz 6.8 *Genügt eine unabhängige Folge $(X_n)_n$ von Zufallsgrößen mit $\mathbb{E}(X_i) = 0$ für alle $i \in \mathbb{N}$ der Bedingung $\sum_{n\geq 1} \operatorname{Var}(X_n) < \infty$, so ist die Reihe $\sum_{n\geq 1} X_n$ f.s. endlich.*

Beweis: Wir betrachten

$$\begin{aligned}\lim_{m\to\infty} P\Big(\sup_{n\geq m} |S_n - S_m| > \varepsilon\Big) &= \lim_{m\to\infty} \lim_{M\to\infty} P\Big(\max_{m\leq n\leq M} |S_n - S_m| > \varepsilon\Big) \\ &\leq \lim_{m\to\infty} \lim_{M\to\infty} \frac{1}{\varepsilon^2} \sum_{k=m}^M \operatorname{Var} X_k = 0.\end{aligned}$$

Hierbei verwenden wir die Voraussetzung für die Varianzen sowie die Ungleichung von Kolmogorov, Satz 6.7. Somit folgt die Behauptung aus Satz 4.4. □

Das anschließende Lemma wird die Verbindung zwischen der Konvergenz zufälliger Reihen und dem starken Gesetz der großen Zahlen herstellen:

Lemma 6.9 (von Kronecker) *Sei $(c_n)_n$ eine Zahlenfolge und $(a_n)_n$ eine aufsteigende Folge mit $0 < a_n$ für alle $n \in \mathbb{N}$. Dann gilt:*

$$\sum_{n\geq 1} \frac{c_n}{a_n} < \infty \Rightarrow \lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^n c_k = 0\ .$$

Beweis: Wurde in den Übungen besprochen.

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

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

Beweis: Die Folge $(Y_n)_n$ mit $Y_n := a_n^{-1}(X_n - \mathbb{E}(X_n))$, $n \geq 1$, genügt den Voraussetzungen in Satz 6.8. Somit ist $\sum_{n\geq 1} Y_n$ fast sicher endlich. Daraus folgt die Behauptung mit Hilfe des Lemmas von Kronecker, 6.9. □

Für $a_n = n$ ergibt sich die Aussage von Korollar 6.5, allerdings nur unter der stärkeren Voraussetzung $X_n \in \mathcal{L}_2$. Satz 6.10 kann im Fall identisch verteilter Zufallsgrößen auch für $a_n = n^{\delta+1/2}$ oder $a_n = n^{1/2}(\log n)^{\delta+1/2}$ für $\delta > 0$, nicht jedoch für $a_n = n^{1/2}(\log n)^{1/2}$ angewandt werden. Bei der Wahl $\delta = 1/2$ erhalten wir zum Beispiel

$$\lim_{n\to\infty} \frac{S_n}{\sqrt{n}\log n} = 0 \quad P\text{-f.s.}$$

(im Fall $\mathbb{E}(X_1) = 0$).

Wir fassen diese Beobachtungen zusammen: Es sei $S_n := \sum_{j=1}^n X_j$, wobei $(X_n)_n$ eine Folge identisch verteilter, reeller, quadratisch integrierbarer Zufallsgrößen mit $\mathbb{E}(X_1) = 0$ ist. Dann ist $\lim_{n\to\infty} \frac{S_n}{n} = 0$ und $\lim_{n\to\infty} \frac{S_n}{\sqrt{n}\log n} = 0$ P-f.s. Für jede isotone Folge $(a_n)_n \nearrow \infty$ wissen wir nach Lemma 6.3, dass es ein $\tau \in \bar{\mathbb{R}}$ gibt mit

$$P\Big(\limsup_{n\to\infty} \frac{1}{a_n} S_n = \tau\Big) = 1 \ .$$

Ebenfalls ist $P\Big(\liminf_{n\to\infty} \frac{1}{a_n} S_n = -\tau\Big) = 1$ aus Symmetriegründen. Für die schnell wachsenden Folgen $a_n = n$ und $a_n = \sqrt{n}\log n$ folgt $\tau = 0$. Für $a_n = \sqrt{n}$ werden wir $\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = +\infty$ P-f.s. sehen (als Folge aus dem zentralen Grenzwertsatz).

Kann man $(a_n)_n$ so bestimmen, dass τ reell und > 0 ist? Dann tritt das Ereignis $\{S_n \geq \eta a_n\}$ für $\eta < \tau$ unendlich oft und für $\eta > \tau$ nur endlich oft ein, jeweils mit Wahrscheinlichkeit 1! Tatsächlich gilt:

Satz 6.11 (Gesetz vom iterierten Logarithmus) *Es sei $(X_n)_n$ eine Folge unabhängiger, identisch verteilter, quadratisch integrierbarer Zufallsgrößen mit $\mathbb{E}(X_1) = 0$.*

Dann gelten

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = \sigma \textit{ fast sicher}$$

und

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = -\sigma \textit{ fast sicher}$$

mit $\sigma^2 := \mathrm{Var}(X_n)$.

Wir führen den Beweis hier nicht.

Das folgende numerische Gedankenexperiment gibt Auskunft über $\log(\log n)$: Jemand werfe in jeder Sekunde eine Münze einmal, n bezeichne die Anzahl

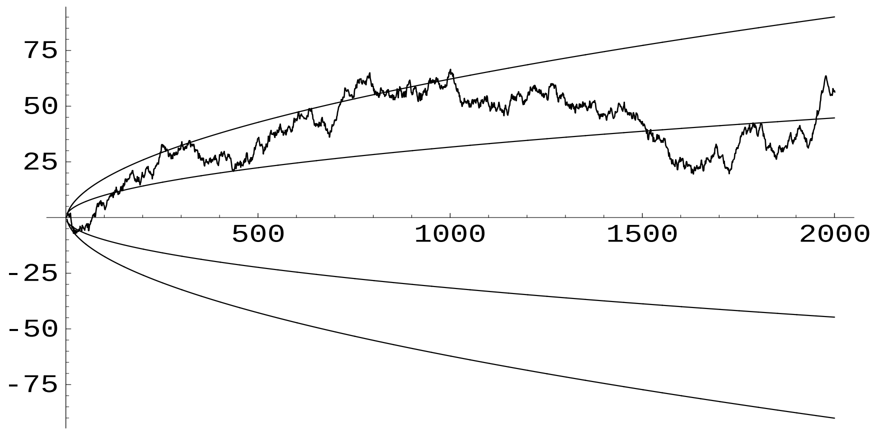


Abbildung 6.2. Realisierung $\big((S_n(\omega)\big)_{1\le n\le 2000}$ bei unabhängigen, $N(0,1)$-verteilten Zuwächsen zusammen mit den Abbildungen $n \mapsto \pm\sqrt{n}$ und $n \mapsto \pm\sqrt{2n\log\log n}$.

der Würfe. Ist dann x der Zeitpunkt, an dem das Experiment begonnen hat, welches er nun abbricht, so gilt für $\log(\log n)$:

x	$\log(\log n)$
vor einer Stunde	$2,103$
beim Tode Caesars	$3,214$
Geburt des Universums	$3,706$

Wir betrachten abschließend eine Anwendung des starken Gesetzes der großen Zahlen. Wie in Beispiel 5.4(b) bzw. 5.15 sei $(X_n)_n$ die Folge der unabhängigen Rademacher-Funktionen $X_n = 1_{A_n}$ mit

$$A_n = \bigcup_{k=0}^{2^{n-1}-1} \left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right) .$$

X_n ist $b(1,1/2)$ verteilt und nach dem starken Gesetz ist $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = 1/2$ P-f.s.

Jedes $\omega \in [0,1)$ besitzt bezüglich $g \ge 2$, $g \in \mathbb{N}$, genau eine g-adische Entwicklung

$$\omega = \sum_{n=1}^{\infty} \xi_n g^{-n} \ , \ \xi_n(\omega) \in \{0,\dots,g-1\} \tag{6.3}$$

(Nicht schließlich alle ξ_n sind gleich $g-1$).

$S_n^{\varepsilon,g}(\omega)$ sei die Anzahl aller $i = 1,\dots,n$ mit $\xi_i(\omega) = \varepsilon$ in der g-adischen Entwicklung von ω, $\varepsilon \in \{0,\dots,g-1\}$.

ω heißt *g-normal*, wenn
$$\lim_{n\to\infty} \frac{1}{n} S_n^{\varepsilon,g}(\omega) = \frac{1}{g}$$
für $\varepsilon = 0, \ldots, g-1$ gilt.
ω heißt *absolut normal*, wenn sie für *alle* $g = 2, 3, \ldots$ g-normal ist.
Wir betrachten den Fall $g = 2$. Es ist dann
$$S_n^{0,2}(\omega) = \sum_{i=1}^n X_i, \quad S_n^{1,2} = n - S_n^{0,2}, \ n \in \mathbb{N}.$$
Also sind P-fast alle Zahlen $\omega \in \Omega$ 2-normal. Tatsächlich sind P-fast alle $\omega \in \Omega$ auch g-normal für $g \geq 2$. Betrachte dazu
$$\xi_n : \Omega \to \{0, 1, \ldots, g-1\},$$
definiert durch (6.3). Dann gilt
$$\{\xi_n = \varepsilon\} = \bigcup_{k=0}^{g^{n-1}-1} \Big[\frac{kg+\varepsilon}{g^n}, \frac{kg+\varepsilon+1}{g^n}\Big)$$
für $\varepsilon \in \{0, \ldots, g-1\}$. Also sind alle ξ_n Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ mit $P(\xi_n = \varepsilon) = \frac{1}{g}$. Analog zu den Überlegungen in Beispiel 5.4(b) folgt, dass $(\xi_n)_n$ eine unabhängige Folge ist. Nun definieren wir für jedes $\varepsilon \in \{0, \ldots, g-1\}$
$$X_n^\varepsilon := 1_{\{\xi_n = \varepsilon\}}.$$
Dann sind diese alle $b(1, \frac{1}{g})$-verteilt und es gilt
$$S_n^{\varepsilon,g} = \sum_{i=1}^n X_i^\varepsilon.$$
Das starke Gesetz der großen Zahlen liefert nun
$$\lim_{n\to\infty} \frac{1}{n} S_n^{\varepsilon,g} = \frac{1}{g} \quad P-\text{f.s.}.$$
Wir haben also den folgenden Satz bewiesen:

Satz 6.12 *In Bezug auf das Lebesgue-Maß auf* $[0, 1)$ *sind fast alle Zahlen* $\omega \in [0, 1)$ *absolut normal.*

Bemerkung 6.13 Champernowne hat 1933 gezeigt, dass
$$0,12345678910111213141516171819 2021\ldots$$
10-normal ist. Die Frage, ob zum Beispiel $\sqrt{2}$, $\log 2$, e oder π normal sind, ist unbeantwortet. Man kennt kein konkretes Beispiel einer absolut normalen Zahl. Eine g-normale Zahl ist im Allgemeinen nicht absolut normal (dies haben Cassels 1959 und Schmidt 1962 beobachtet). Eine g-normale Zahl ist stets g^p-normal für jedes $p \in \mathbb{N}$ (bewiesen von Hlawka, 1979).

KAPITEL 7

Große Abweichungen

Wir betrachten in diesem Kapitel eine Folge $(X_n)_n$ von unabhängigen und identisch verteilten Zufallsgrößen auf einem W-Raum $(\Omega, \mathcal{A}, P)$. Wenn wir annehmen, dass $\mathbb{E}(X_1) = \mu \in \mathbb{R}$ ist, so besagt das starke Gesetz der großen Zahlen, dass $\frac{1}{n}S_n$ mit $S_n = X_1 + \cdots + X_n$ fast sicher gegen μ konvergiert für $n \to \infty$. Also konvergiert die Folge der Wahrscheinlichkeiten

$$P\big(S_n \geq (\mu + \alpha)n\big)\,, \quad \alpha > 0\,,$$

gegen Null für $n \to \infty$. Wir wollen in diesem Kapitel die *Rate* quantifizieren, mit der dies gegen Null konvergiert.

Wir betrachten zunächst den n-fachen unabhängigen Münzwurf. S_n sei $b(n, p)$-verteilt. Dann gilt mit der Tschebyschev-Ungleichung für $\alpha > 0$

$$P\left(\Big|\frac{S_n}{n} - p\Big| \geq \alpha\right) \leq \frac{p(1-p)}{n\alpha^2}\,,$$

was zum Beispiel für $p = \frac{1}{2}$, $\alpha = \frac{1}{10}$ und $n = 1000$ die Schranke $\frac{1}{40}$ liefert. Wie gut ist diese Abschätzung? Wir wählen alternativ die Markov-Ungleichung (Satz 2.8) mit $g(x) = e^{\lambda x}$, $\lambda > 0$. Dann gilt

$$P(S_n \geq n\alpha) \leq e^{-\alpha n \lambda}\,\mathbb{E}(e^{\lambda S_n})\,.$$

Mit den X_i sind die $e^{\lambda X_i}$ unabhängig, siehe Satz 5.16, und somit folgt mit Satz 5.21 für alle $\lambda > 0$

$$\begin{aligned} P(S_n \geq n\alpha) &\leq e^{-\alpha n \lambda}\big(pe^{\lambda} + (1-p)\big)^n \\ &= \exp\big(-n\{-\log M(\lambda) + \alpha\lambda\}\big) \end{aligned}$$

mit $M(\lambda) := pe^{\lambda} + (1-p)$. Jetzt optimieren wir in $\lambda > 0$, bestimmen also das Minimum von

$$f(\lambda) := \log M(\lambda) - \alpha\lambda\,.$$

Es gilt

$$f''(\lambda) = \frac{M''(\lambda)}{M(\lambda)} - \frac{M'(\lambda)^2}{M(\lambda)} = \frac{p(1-p)e^{\lambda}}{M(\lambda)^2} > 0$$

für $\lambda > 0$ und $0 < p < 1$. Also ist $f'(\lambda)$ streng monoton, es existiert also höchstens eine Nullstelle λ_0 von f'.

Ist $\alpha \in (p, 1)$, so ist $f'(\lambda_0) = 0$ für

$$\lambda_0 = \log \frac{\alpha(1-p)}{p(1-\alpha)} > 0\,.$$

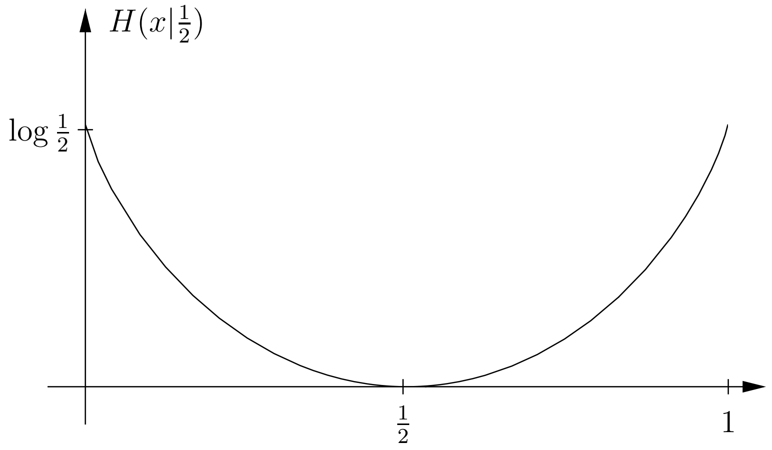


ABBILDUNG 7.1. Die KULLBACK-LEIBLER-Information für $p = 1/2$.

Einsetzen in f liefert

$$P(S_n \geq n\alpha) \leq \exp\bigl(-nH(\alpha|p)\bigr) \tag{7.1}$$

mit

$$H(\alpha|p) := \alpha \log \frac{\alpha}{p} + (1-\alpha)\log\frac{1-\alpha}{1-p}\,, \quad 0 < p < 1\,, \quad \alpha \in (p,\,1)\,.$$

Im obigen Zahlenbeispiel folgt das Minimum mit

$$\begin{aligned} P\left(\Bigl|\frac{S_n}{n} - \frac{1}{2}\Bigr| \geq \alpha\right) &= 2\,P\Bigl(\frac{S_n}{n} \geq \alpha + \frac{1}{2}\Bigr) \\ &\leq 2\exp\Bigl(-nH\bigl(\alpha + \frac{1}{2}\,|\,\frac{1}{2}\bigr)\Bigr) \end{aligned}$$

für $\alpha \in (0,\,\frac{1}{2})$:

$$P\left(\Bigl|\frac{S_{1000}}{1000} - \frac{1}{2}\Bigr| \geq \frac{1}{10}\right) \leq 2\left(\frac{5}{6}\right)^{600}\left(\frac{5}{4}\right)^{400} \leq 3.6\cdot 10^{-9}\,,$$

was phantastisch viel besser als $\frac{1}{40}$ ist.

$H(\cdot|p)$ ist eine berühmte Funktion, die *relative Entropie* oder KULLBACK-LEIBLER-*Information* genannt wird.

Das Ziel ist nun, die Güte von (7.1) zu untersuchen. Gibt es noch eine weitere Verbesserung? Kann eine Abschätzung wie in (7.1) für die allgemeine Situation einer Folge $(X_n)_n$ von unabhängigen und identisch verteilten Zufallsgrößen gefunden werden? Die Antwort liefert der

Satz 7.1 (CRAMÉR, 1938) *Es sei $(X_n)_n$ eine Folge unabhängiger und identisch verteilter Zufallsgrößen. Es sei*

$$\Lambda(\lambda) := \log M(\lambda) := \log \mathbb{E}(e^{\lambda X_1}) < \infty$$

für alle $\lambda \in \mathbb{R}$. Dann gilt:

(i) *Für jede abgeschlossene Menge $F \subset \mathbb{R}$ ist*

$$\limsup_{n\to\infty} \frac{1}{n} \log P\Big(\frac{S_n}{n} \in F\Big) \leq - \inf_{x\in F} I(x)\,.$$

(ii) *Für jede offene Menge $G \subset \mathbb{R}$ ist*

$$\liminf_{n\to\infty} \frac{1}{n} \log P\Big(\frac{S_n}{n} \in G\Big) \geq - \inf_{x\in G} I(x)\,.$$

Dabei ist I durch $I(x) := \sup_{\lambda\in\mathbb{R}}\big\{\lambda x - \Lambda(\lambda)\big\}$ gegeben.

Bemerkungen 7.2 (i) $M(\lambda)$ heißt *Momente-erzeugende Funktion*, $\Lambda(\lambda)$ daher *logarithmische Momente-erzeugende Funktion* oder *Cumulanten-erzeugende Funktion*. Zu gegebenem Anlass gehen auf diese Begriffe ein.
(ii) Ist X_1 $b(1,\, p)$-verteilt, so folgt

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

für $x \in [0,\, 1]$ und $I(x) = \infty$ sonst. Natürlich ist hier $\Lambda(\lambda) < \infty$ für alle $\lambda \in \mathbb{R}$, $I(\cdot)$ ist aber unstetig auf $\mathbb{R}$. Es ist

$$I(x) = H(x|p),\ x \in [0,1]\,.$$

(iii) Die Funktion I im Satz von Cramér heißt *Ratenfunktion*.

Wir bereiten den Beweis des Satzes vor, indem wir Eigenschaften der Funktionen Λ und I zusammenstellen.

Lemma 7.3 (i) *Λ und I sind konvexe Funktionen.*
(ii) *Aus $\Lambda(\lambda) < \infty$ für ein $\lambda > 0$ folgt $\mathbb{E}(X_1) < \infty$ und für $x \geq \mathbb{E}(X_1)$ ist*

$$I(x) = \sup_{\lambda\geq 0}\big\{\lambda x - \Lambda(\lambda)\big\}\,. \tag{7.2}$$

Für $x > \mathbb{E}(X_1)$ steigt I. Analog gilt: Aus $\Lambda(\lambda) < \infty$ für ein $\lambda < 0$ folgt $\mathbb{E}(X_1) > -\infty$ und für $x \leq \mathbb{E}(X_1)$ ist

$$I(x) = \sup_{\lambda\leq 0}\big\{\lambda x - \Lambda(\lambda)\big\}\,.$$

Für $x < \mathbb{E}(X_1)$ fällt I. Ist $\mathbb{E}(X_1) < \infty$, so ist $I\big(\mathbb{E}(X_1)\big) = 0$ und es gilt immer $\inf_{x\in\mathbb{R}} I(x) = 0$.
(iii) *Λ ist differenzierbar auf ganz $\mathbb{R}$, wenn $\Lambda(\lambda) < \infty$ für alle $\lambda \in \mathbb{R}$ gilt. Es gilt*

$$\Lambda'(\eta) = \frac{1}{M(\eta)} \mathbb{E}\left[X_1 e^{\eta X_1}\right]$$

und

$$\Lambda'(\eta) = y \quad \Rightarrow \quad I(y) = \eta y - \Lambda(\eta)\,.$$

Beweis: (i) Für jedes $\theta \in [0, 1]$ folgt aus der Hölder-Ungleichung

$$\begin{aligned}\Lambda\big(\theta\lambda_1 + (1-\theta)\lambda_2\big) &= \log \mathbb{E}\Big(\big(e^{\lambda_1 X_1}\big)^{\theta}\big(e^{\lambda_2 X_1}\big)^{1-\theta}\Big)\\ &\le \log\Big(\mathbb{E}\big(e^{\lambda_1 X_1}\big)^{\theta}\mathbb{E}\big(e^{\lambda_2 X_1}\big)^{1-\theta}\Big)\\ &= \theta\Lambda(\lambda_1) + (1-\theta)\Lambda(\lambda_2)\,.\end{aligned}$$

Aus der Definition folgt weiter

$$\begin{aligned}&\theta I(x_1) + (1-\theta) I(x_2)\\ &\quad= \sup_{\lambda\in\mathbb{R}}\big\{\theta\lambda x_1 - \theta\Lambda(\lambda)\big\} + \sup_{\lambda\in\mathbb{R}}\big\{(1-\theta)\lambda x_2 - (1-\theta)\Lambda(\lambda)\big\}\\ &\quad\ge \sup_{\lambda\in\mathbb{R}}\big\{\big(\theta x_1 + (1-\theta)x_2\big)\lambda - \Lambda(\lambda)\big\}\\ &\quad= I\big(\theta x_1 + (1-\theta)x_2\big)\,.\end{aligned}$$

(ii) Die Jensensche Ungleichung liefert

$$\lambda\mathbb{E}(X_1) = \mathbb{E}(\lambda X_1) = \mathbb{E}(\log e^{\lambda X_1}) \le \log\mathbb{E}(e^{\lambda X_1})\,, \tag{7.3}$$

also

$$\mathbb{E}(X_1) \le \frac{\Lambda(\lambda)}{\lambda}\,.$$

Ist die rechte Seite endlich für ein $\lambda > 0$, so ist $\mathbb{E}(X_1) < \infty$. Gilt $\mathbb{E}(X_1) = -\infty$, so ist mit (7.3) $\Lambda(\lambda) = \infty$ für negative λ, also gilt (7.2). Ist $\mathbb{E}(X_1)$ endlich, so folgt aus (7.3) $I\big(\mathbb{E}(X_1)\big) = 0$. Für $x \ge \mathbb{E}(X_1)$ und jedes $\lambda < 0$ gilt dann

$$\lambda x - \Lambda(\lambda) \le \lambda\mathbb{E}(X_1) - \Lambda(\lambda) \le I\big(\mathbb{E}(X_1)\big) = 0\,,$$

also gilt (7.2). Da für jedes $\lambda \ge 0$ die Abbildung $x \mapsto \lambda x - \Lambda(\lambda)$ steigt, ist I auf $\big(\mathbb{E}(X_1),\, +\infty\big)$ steigend. Man betrachte die logarithmische Momente-erzeugende Funktion für $-X$ für den anderen Fall. Es bleibt zu zeigen

$$\inf_{x\in\mathbb{R}} I(x) = 0\,.$$

Für $\mathbb{E}(X_1) < \infty$ ist $I\big(\mathbb{E}(X_1)\big) = 0$. Es sei $\mathbb{E}(X_1) = -\infty$ bei $\Lambda(\lambda) < \infty$ für ein $\lambda > 0$. Dann ist

$$\begin{aligned}\log P^{X_1}\big([x,\, \infty)\big) &\le \inf_{\lambda\ge 0}\log\mathbb{E}\big(e^{\lambda(X_1 - x)}\big)\\ &= -\sup_{\lambda\ge 0}\big\{\lambda x - \Lambda(\lambda)\big\} = -I(x)\end{aligned}$$

nach der Tschebyschev-Ungleichung. Also ist

$$\lim_{x\to-\infty} I(x) \le \lim_{x\to-\infty}\big\{-\log P^{X_1}\big([x,\, \infty)\big)\big\} = 0\,.$$

Der Fall $\mathbb{E}(X_1) = \infty$ für $\Lambda(\lambda) < \infty$ für ein $\lambda < 0$ geht analog.

(iii) Dies folgt, wenn wir Differentation und Integration vertauschen dürfen. Dies folgt mit dem Satz von der majorisierten Konvergenz, denn

$$f_\varepsilon(x) = \big(e^{(\eta+\varepsilon)x} - e^{\eta x}\big)\,/\,\varepsilon$$

konvergiert punktweise gegen $xe^{\eta x}$ für $\varepsilon \to 0$ und

$$|f_\varepsilon(x)| \le e^{\eta x}\left(e^{\delta|x|} - 1\right) / \delta =: h(x)$$

für $\varepsilon \in (-\delta, \delta)$ und $\mathbb{E}\big(|h(X_1)|\big) < \infty$ für $\delta > 0$ klein genug (Übung). Ist weiter $\Lambda'(\eta) = y$, so betrachten wir die Funktion $g(\lambda) := \lambda y - \Lambda(\lambda)$. Dann ist $g(\cdot)$ konkav und $g'(\eta) = 0$, also gilt

$$g(\eta) = \sup_{\lambda \in \mathbb{R}} g(\lambda)\,,$$

was zu zeigen war. □

Beweis des Satzes von Cramér*:* (i) Sei F eine nicht-leere abgeschlossene Teilmenge. Ist $\inf_{x\in F} I(x) =: I_F = 0$, so gilt die Aussage. Sei $I_F > 0$. Nach Voraussetzung ist $\mathbb{E}(X_1) < \infty$. Es gilt

$$\begin{aligned} P\Big(\frac{S_n}{n} \ge x\Big) &\le e^{-n\lambda x}\,\mathbb{E}\Big(\exp\Big(\lambda\sum_{i=1}^n X_i\Big)\Big) \\ &= e^{-n\lambda x}\prod_{i=1}^n \mathbb{E}\left(e^{\lambda X_i}\right) = e^{-n\big(\lambda x - \Lambda(\lambda)\big)} \end{aligned}$$

mit Hilfe der Markov-Ungleichung und Satz 5.21 für $\lambda \ge 0$. Es folgt nach Lemma 7.3 (ii)

$$P\Big(\frac{S_n}{n} \ge x\Big) \le \exp\big(-nI(x)\big) \tag{7.4}$$

für $x > \mathbb{E}(X_1)$. Für $\mathbb{E}(X_1) > -\infty$ und $x < \mathbb{E}(X_1)$ folgt analog

$$P\Big(\frac{S_n}{n} \le x\Big) \le \exp\big(-nI(x)\big)\,.$$

Nun wissen wir, dass $I\big(\mathbb{E}(X_1)\big) = 0$, und da $I_F > 0$, muss $\mathbb{E}(X_1)$ in der offenen Menge F^c sein. Sei (x_-, x_+) die Vereinigung aller offenen Intervalle (a, b) in F^c, die $\mathbb{E}(X_1)$ enthalten. Dann ist $x_- < x_+$ und einer der Werte muss endlich sein, da F nicht leer ist. Ist x_- endlich, so ist $x_- \in F$, also $I(x_-) \ge I_F$. Analog ist $I(x_+) \ge I_F$, wenn x_+ endlich ist. Somit ist

$$\begin{aligned} P\Big(\frac{S_n}{n} \in F\Big) &\le P\Big(\frac{S_n}{n} \le x_-\Big) + P\Big(\frac{S_n}{n} \ge x_+\Big) \\ &\le 2\exp(-n\,I_F)\,. \end{aligned}$$

Hieraus folgt die obere Abschätzung.

(ii) Nun zeigen wir

$$\liminf_{n\to\infty}\frac{1}{n}\log P\Big(\frac{S_n}{n} \in (-\delta, \delta)\Big) \ge \inf_{\lambda\in\mathbb{R}}\Lambda(\lambda) = -I(0) \tag{7.5}$$

für jedes $\delta > 0$. Dies genügt, denn für $Y = X - x$ ist $\log\mathbb{E}(e^{\lambda Y}) = \Lambda(\lambda) - \lambda x$ und $I_Y(\cdot) = I(\cdot + x)$, also folgt aus obiger Ungleichung für jedes x und $\delta > 0$

$$\liminf_{n\to\infty}\frac{1}{n}\log P\Big(\frac{S_n}{n} \in (x-\delta, x+\delta)\Big) \ge -I(x)\,.$$

Da für jede offene Menge G und jedes $x \in G$ und $\delta > 0$ klein genug $(x-\delta, x+\delta) \subset G$ ist, folgt damit die untere Schranke.

Angenommen $P^{X_1}\big((-\infty, 0)\big) > 0$ und $P^{X_1}\big((0, \infty)\big) > 0$. Dann ist $\lim_{|\lambda|\to\infty} \Lambda(\lambda) = \infty$. Weiter ist Λ stetig und differenzierbar nach Lemma 7.3 (iii), also existiert ein $\eta < \infty$ mit $\Lambda'(\eta) = 0$ und $\Lambda(\eta) = \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda)$. Wir definieren ein neues W-Maß durch

$$\tilde{\mu}(A) := \int_A e^{\eta x - \Lambda(\eta)}\, dP^{X_1}(x)\,,$$

denn

$$\int_{\mathbb{R}} e^{\eta x}\, dP^{X_1}(x) \cdot \frac{1}{M(\eta)} = 1\,.$$

Es sei nun $\hat{S}_n := \sum_{i=1}^n \hat{X}_i$ mit $P^{\hat{X}_i} = \tilde{\mu}$, $i = 1, \ldots, n$, und $(\hat{X}_i)_i$ unabhängig. Es gilt

$$\begin{aligned} P\Big(\frac{S_n}{n} \in (-\varepsilon, \varepsilon)\Big) &= \int_{\{|\sum_{i=1}^n x_i| < n\varepsilon\}} P^{X_1}(dx_1)\cdots P^{X_n}(dx_n) \\ &\geq e^{-n\varepsilon|\eta|} \int_{\{|\sum_{i=1}^n x_i| < n\varepsilon\}} \exp\Big(\eta \sum_{i=1}^n x_i\Big)\, P^{X_1}(dx_1)\cdots P^{X_n}(dx_n) \\ &= e^{-n\varepsilon|\eta|}\, e^{n\Lambda(\eta)}\, \hat{P}\Big(\frac{\hat{S}_n}{n} \in (-\varepsilon, \varepsilon)\Big)\,. \end{aligned}$$

Dabei bezeichnet $\hat{P} = \bigotimes_{i\geq 1} P^{\hat{X}_i}$. Nun liefert Lemma 7.3 (iii) und die Wahl von η

$$\mathbb{E}(\hat{X}_1) = \frac{1}{M(\eta)} \int x e^{\eta x}\, dP^{X_1}(x) = \Lambda'(\eta) = 0\,.$$

Also besagt das schwache Gesetz der großen Zahlen

$$\lim_{n\to\infty} \hat{P}\Big(\frac{\hat{S}_n}{n} \in (-\varepsilon, \varepsilon)\Big) = 1.$$

Somit folgt aus obiger Rechnung für $0 < \varepsilon < \delta$

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log P\Big(\frac{S_n}{n} \in (-\delta, \delta)\Big) & \\ &\geq \liminf_{n\to\infty} \frac{1}{n} \log P\Big(\frac{S_n}{n} \in (-\varepsilon, \varepsilon)\Big) \\ &\geq \Lambda(\eta) - \varepsilon|\eta|\,, \end{aligned}$$

also folgt (7.5) für $\varepsilon \to 0$.

Ist abschließend $P^{X_1}\big((-\infty, 0)\big) = 0$ oder $P^{X_1}\big((0, \infty)\big) = 0$, so ist Λ eine monotone Funktion mit

$$\inf_{x\in\mathbb{R}} \Lambda(\lambda) = \log P^{X_1}\big(\{0\}\big)\,.$$

Da

$$P\Big(\frac{S_n}{n} \in (-\delta, \delta)\Big) \geq P\Big(\frac{S_n}{n} = 0\Big) = \Big(P^{X_1}\big(\{0\}\big)\Big)^n,$$

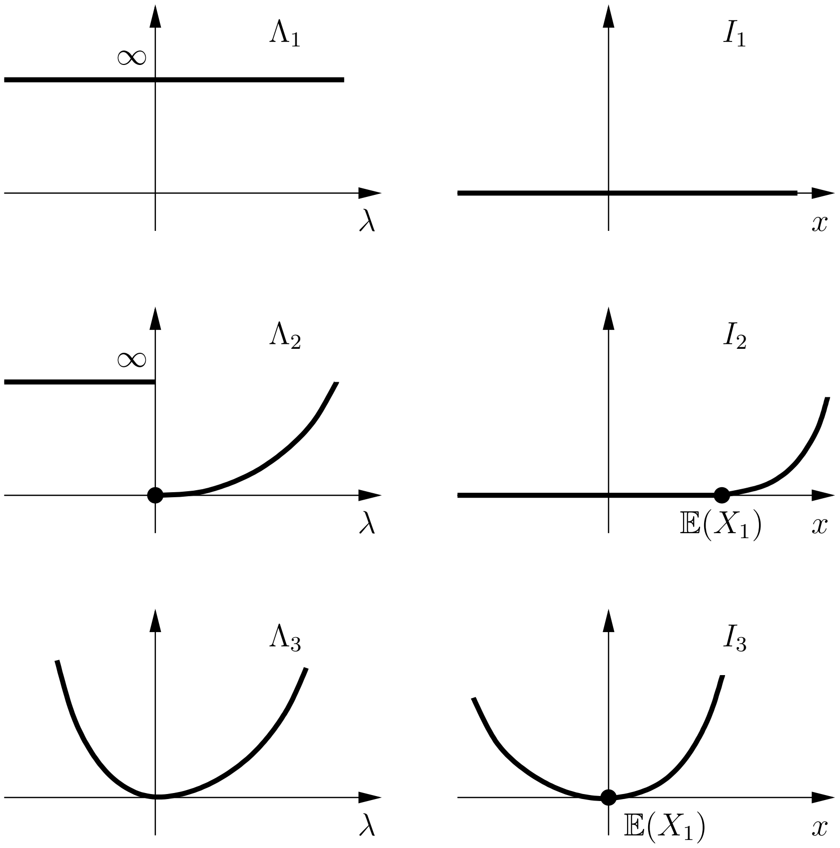


ABBILDUNG 7.2. Paare (Λ, I).

folgt erneut die Behauptung. □

Bemerkungen 7.4 (i) Für die obere Abschätzung bleibt bei allen technischen Details die Idee, die MARKOV-Ungleichung für $e^{\lambda x}$, $\lambda \geq 0$, anzuwenden. Bei der unteren Abschätzung ist die exponentielle Maßtransformation zu $\tilde{\mu}$ die Idee, um anschließend ein schwaches Gesetz für $\hat{S}_n/n$ nutzen zu können.

(ii) LUDWIG BOLTZMANN betrachtete 1877 ein Modell des idealen Gases. Dabei teilte er einen Gasbehälter und dachte sich die Gaspartikel rein zufällig und unabhängig auf die Teilbehälter verteilt. Er untersuchte untypische Besetzungszahlen und kam (im binomial- oder multinomial-verteilten Fall) zu einer Approximation

$$P\Big(\frac{S_n}{n} \approx p'\Big) \approx \exp\big(-nH(p'|p)\big)\,.$$

Dabei erkannte er, dass $H(\cdot|p)$ der Entropie der Thermodynamik entspricht und führte sie somit als eine zentrale Größe der sogenannten statistischen

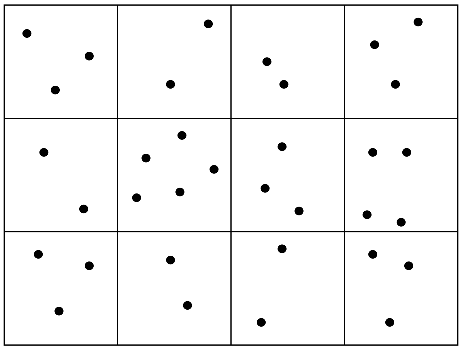


Abbildung 7.3. Gasbehälter, Boltzmann-Modell.

Mechanik ein. Formal entspricht das große Abweichungsresultat dem Boltzmann-Gesetz

$$S = k \log W$$

(S Entropie, W Wahrscheinlichkeit, k Boltzmann-Konstante).

(iii) In einer Übung bestimme man die Funktion I für diverse Verteilungen:

- X_1 sei Poisson-verteilt zu $\alpha > 0$, dann ist
$$I(x) = \alpha - x + x \log \frac{x}{\alpha} \qquad \text{für } x \geq 0$$
und $+\infty$ sonst.
- X_1 sei Normal-verteilt zu 0 und σ^2, dann ist
$$I(x) = \frac{x^2}{2\sigma^2} .$$

Unter den Bedingungen des Satzes von Cramér, die deutlich abgeschwächt werden können – es reicht $\Lambda(\lambda) < \infty$ für ein $\lambda > 0$ anzunehmen – gilt sogar, dass $\Lambda \in C^\infty(\mathbb{R})$ und I strikt konvex ist.

(iv) Für viele Teilmengen stimmen die obere und die untere Abschätzung überein, so dass Konvergenz folgt. Es gilt z.B. für $y \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n} \log P\Big(\frac{S_n}{n} \geq y\Big) = -\inf_{x \geq y} I(x) .$$

Dies wird eine einfache Folgerung des zentralen Grenzwertsatzes sein und im nächsten Kapitel betrachtet. Symbolisch schreibt man manchmal für die beiden Ungleichungen im Satz von Cramér:

$$P\Big(\frac{S_n}{n} \in A\Big) \approx \exp\big(-n \inf_{x \in A} I(x)\big) .$$

Da $I\big(\mathbb{E}(X_1)\big) = 0$, folgt aus dem Satz von Cramér auch das starke Gesetz. Dies wird in den Übungen behandelt.

Wir wollen eine Anwendung der exponentiellen Abschätzung aus Satz 7.1 kennenlernen:

Im BERNOULLI-Experiment nennen wir jede maximale Teilfolge von einander benachbarten gleichen Symbolen einen *Run*. Man teile die Gruppe der Hörerinnen und Hörer (dieser Vorlesung) in zwei gleich große Gruppen. In der einen Gruppe wird jeder gebeten, eine Münze 200 Mal zu werfen und die 0-1-Sequenz zu notieren. In der zweiten Gruppe wird jeder gebeten, eine zufällige 0-1-Folge der Länge 200 zu notieren, ohne eine Münze zu werfen. Die Notizen werden eingesammelt. Nun zieht ein Spielleiter aus dem Topf aller Zettel zufällig einen aus und kann, so die Behauptung, in der Regel den Zettel der Gruppe zuordnen, aus der er auch tatsächlich stammt.

Dazu ein heuristisches Argument. Wir groß ist die Länge des längsten Runs von Einsen in einer BERNOULLI-Kette der Länge n? Ein 1-Run der Länge m tritt mit Wahrscheinlichkeit p^m auf. Für den 1-Run gibt es ca. n mögliche Positionen, also ist die erwartete Anzahl der 1-Runs der Länge m ca. np^m. Ist der längste 1-Run eindeutig, so erfüllt seine Länge R_n die Gleichung $1 = np^{R_n}$, also

$$R_n = \frac{\log n}{\log(1/p)}\,.$$

Im Fall einer fairen Münze folgt für $n = 200$ $R_n \approx 7.64$. Die Mitglieder der zweiten Gruppe werden es in der Regel scheuen, mehr als vier oder fünf Mal hintereinander das Gleiche Symbol zu notieren! 1-Runs der Größe 5 führen mit großer Sicherheit zur Einordnung in die erste Gruppe! Dieses Lehrexperiment geht auf RÉNYI zurück und soll nun mathematisiert werden:

Es sei $(X_n)_n$ eine Folge unabhängiger und identisch verteilter Zufallsgrößen auf einem W-Raum $(\Omega,\, \mathcal{A},\, P)$, $S_0 := 0$, $S_k := \sum_{i=1}^k X_i$, $k = 1, 2, \dots$ R_m bezeichne Segmente maximaler Länge der Folge $(S_k)_{k\geq 0}$ bis zum Zeitpunkt m, deren empirischer Mittelwert zu einer Menge $A \in \mathcal{B}$ gehört:

$$R_m := \max\Big\{ l - k : 0 \leq k < l \leq m,\ \frac{S_l - S_k}{l - k} \in A \Big\}\,.$$

Bei Wahl der BERNOULLI-Kette und $A = \{1\}$ ist R_m die Länge des längsten Runs bis zum Zeitpunkt m. Als Hilfsgöße betrachten wir

$$T_r := \inf\Big\{ l : \frac{S_l - S_k}{l - k} \in A \text{ für ein } 0 \leq k \leq l - r \Big\}\,.$$

Es gilt $\{R_m \geq r\}$ tritt ein genau dann, wenn $\{T_r \leq m\}$.
Nun gilt der folgende Satz:

Satz 7.5 (ERDŐS und RÉNYI, 1970) *Angenommen, wir betrachten ein $A \in \mathcal{B}$ mit*

$$I_A := - \lim_{n\to\infty} \frac{1}{n} \log P\Big(\frac{S_n}{n} \in A\Big)$$

existiert, dann gilt

$$P\Big(\lim_{m\to\infty} \frac{R_m}{\log m} = \frac{1}{I_A}\Big) = P\Big(\lim_{r\to\infty} \frac{r}{\log T_r} = \frac{1}{I_A}\Big) = 1\,.$$

Korollar 7.6 *Im Fall des* BERNOULLI*-Experiments und $A = \{1\}$ gilt somit*

$$P\Big(\lim_{m\to\infty} \frac{R_m}{\log m} = \frac{1}{\log(1/p)}\Big) = 1\,.$$

Beweis: Der Satz von CRAMÉR liefert

$$I_{\{1\}} = H(1|p) = \log\frac{1}{p}\,.$$

Siehe Bemerkung 7.4 (iv) und 7.2 (ii). Alternativ sieht man aber auch sofort

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{S_n}{n} = 1\right) = \log p = -\log\frac{1}{p}.$$

□

Beweis von Satz 7.5: Wir setzen

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

Dann gilt

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

Mit $P(C_{k,l}) = P\big(\frac{1}{l-k}S_{l-k} \in A\big)$ folgt

$$P(T_r \le m) \le m\sum_{n=r}^{\infty} P\left(\frac{S_n}{n} \in A\right).$$

Wir nehmen zunächst an, dass $0 < I_A < \infty$. Wähle $m = \lfloor e^{r(I_A-\varepsilon)}\rfloor$, $\varepsilon > 0$. Dann folgt

$$\begin{aligned}\sum_{r=1}^{\infty} P\big(T_r \le e^{r(I_A-\varepsilon)}\big) &\le \sum_{r=1}^{\infty} e^{r(I_A-\varepsilon)}\sum_{n=r}^{\infty} c\, e^{-n(I_A-\varepsilon/2)}\\ &\le c'\sum_{r=1}^{\infty} e^{-r\varepsilon/2} < +\infty\end{aligned}$$

für positive Konstanten c, c', die von ε abhängen können. Für $I_A = \infty$ sei $m = \lfloor e^{r/\varepsilon}\rfloor$, dann gilt

$$\sum_{r=1}^{\infty} P(T_r \le e^{r/\varepsilon}) < \infty \quad \forall\varepsilon > 0\,.$$

Nach BOREL-CANTELLI ist

$$P\Big(\limsup_{r\to\infty}\big\{T_r \le e^{r(I_A-\varepsilon)}\big\}\Big) = 0\,.$$

Somit hat das Ereignis B aller $\omega \in \Omega$ mit $T_r \le e^{r(I_A-\varepsilon)}$ für höchstens endlich viele $r \in \mathbb{N}$ Wahrscheinlichkeit 1, also

$$\liminf_{r\to\infty} \frac{1}{r} \log T_r \ge I_A \quad \text{fast sicher}$$

(bzw. $T_r \le e^{r/\varepsilon}$).
Mit $\{R_m \ge r\} = \{T_r \le m\}$ folgt

$$\limsup_{m\to\infty} \frac{R_m}{\log m} \le \frac{1}{I_A} \quad \text{fast sicher}\,.$$

Sei nun $I_A < \infty$. Setze

$$B_l := \Big\{ \frac{S_{lr} - S_{(l-1)r}}{r} \in A \Big\}\,.$$

Dann sind $(B_l)_l$ unabhängige Ereignisse mit $P(B_l) = P\big(\frac{S_r}{r} \in A\big)$. Es gilt daher

$$\bigcup_{l=1}^{\lfloor m/r \rfloor} B_l \subset \{T_r \le m\}$$

und somit

$$\begin{aligned} P(T_r > m) &\le 1 - P\Big(\bigcup_{l=1}^{\lfloor m/r \rfloor} B_l\Big) = \big(1 - P(B_1)\big)^{\lfloor m/r \rfloor} \\ &\le e^{-\lfloor m/r \rfloor P(B_1)} = \exp\left(-\lfloor m/r \rfloor P\Big(\frac{S_r}{r} \in A\Big)\right)\,. \end{aligned}$$

Setze $m = \lceil e^{r(I_A+\varepsilon)} \rceil$, so folgt für alle $\varepsilon > 0$

$$\begin{aligned} \sum_{r=1}^{\infty} P\big(T_r > e^{r(I_A+\varepsilon)}\big) &\le \sum_{r=1}^{\infty} \exp\Big(-\frac{c_1}{r} e^{r(I_A+\varepsilon)} e^{-r(I_A+\varepsilon/2)}\Big) \\ &\le \sum_{r=1}^{\infty} \exp(-c_2 e^{c_3 r}) < \infty \end{aligned}$$

für c_1, c_2, c_3 positive Konstanten, die von ε abhängen können. Nun liefert Borel-Cantelli wie oben

$$\liminf_{m\to\infty} \frac{R_m}{\log m} = \liminf_{r\to\infty} \frac{r}{\log T_r} \ge \frac{1}{I_A} \quad \text{fast sicher,}$$

womit der Satz gezeigt ist. □

Bemerkungen 7.7 (i) Die Bedingung in Satz 7.5 ist nach dem Satz von Cramér erfüllt, wenn die logarithmischen Momente-erzeugende Funktion $\Lambda(\cdot)$ von X_1 überall endlich ist und

$$\inf_{x\in\bar{A}} I(x) = \inf_{x\in\mathring{A}} I(x) = I_A$$

gilt.

(ii) Satz 7.5 kann beim Vergleich von DNA-Sequenzen interpretiert werden. Wenn $X_i \in \{\text{Adenin, Guanin, Thymin, Cytosin}\}$ und Y_i ebenfalls und $(X_i)_i$ und $(Y_i)_i$ seien unabhängig und $p := P(X_1 = Y_1)$, $0 < p < 1$, so gilt für

$$R_m := \max\{l : A_{i+k} = B_{i+k} \text{ für } k = 1, \dots, l,\ 0 \leq i \leq m - l\}$$

die Aussage

$$P\Big(\lim_{m\to\infty} \frac{R_m}{\log_{1/p} m} = 1\Big) = 1\,.$$

Interessant ist die folgende Variante, bei der Verschiebungen erlaubt sind:

$$H_m := \max\{l : A_{i+k} = B_{j+k} \text{ für } k = 1, \dots, l,\ 0 \leq i, j \leq m - l\}\,.$$

Dann gilt

$$P\Big(\lim_{m\to\infty} \frac{H_m}{\log_{1/p} m} = 2\Big) = 1\,.$$

Der Beweis wird in den Übungen diskutiert und ist eine Anwendung eines allgemeinen Borel-Cantelli-Resultats.

KAPITEL 8

Der zentrale Grenzwertsatz

Für jedes $n \in \mathbb{N}$ seien $X_{n1}, \dots, X_{nk_n}$ unabhängige Zufallsgrößen, definiert auf einem W-Raum $(\Omega_n, \mathcal{A}_n, P_n)$, mit endlichen Varianzen $\sigma_{nk}^2 := \operatorname{Var}(X_{nk})$, $k = 1, \dots, k_n$. Betrachte

$$S_n := X_{n1} + \cdots + X_{nk_n} \ .$$

Bei der Familie $(X_{nk})_{n\geq 1}^{k=1,\dots,k_n}$ spricht man von einem *Dreiecksschema* mit gegen unendlich strebender Zeilenlänge k_n. Die S_n können hier für jedes n aus neuen Summanden bestehen.

Wir wollen die Folge der Verteilungen $(P_n^{S_n})_n$ untersuchen. Im Fall des Bernoulli-Experiments besagt der Grenzwertsatz von de Moivre und Laplace:

$$\lim_{n\to\infty} P\Big(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\operatorname{Var}(S_n)}} < t\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-x^2/2)\, dx =: \Phi(t)$$

für $t \in \mathbb{R}$, d.h. die Verteilungsfunktionen $F_n(\cdot) := P\Big(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\operatorname{Var}(S_n)}} < \cdot\Big)$ konvergieren punktweise gegen $\Phi(\cdot)$.

Zunächst klären wir den Zusammenhang zwischen Konvergenz in Verteilung bzw. schwacher Konvergenz und Konvergenz der zugehörigen Verteilungsfunktionen:

Satz 8.1 *Es seien P und $(P_n)_n$ W-Maße auf $(\mathbb{R}, \mathcal{B})$ und F, $(F_n)_n$ die zugehörigen Verteilungsfunktionen, d.h. $F(x) := P((-\infty, x))$ und $F_n(x) := P_n((-\infty, x))$, $x \in \mathbb{R}$. Weiter sei W eine Teilmenge von $C_b(\mathbb{R})$ mit der folgenden Eigenschaft:*

Für alle $x < y$ existiert ein $f \in W$ mit $0 \leq f \leq 1$, $f(z) = 1$ für alle $z \leq x$, $f(z) = 0$ für alle $z \geq y$. Dann sind die folgenden Aussagen äquivalent:

(i)
$$\lim_{n\to\infty} \int f\, dP_n = \int f\, dP \quad \text{für alle} \quad f \in C_b(\mathbb{R}).$$
(schwache Konvergenz)

(ii)
$$\lim_{n\to\infty} \int f\, dP_n = \int f\, dP \quad \text{für alle} \quad f \in W.$$

(iii)
$$\lim_{n\to\infty} F_n(x) = F(x) \ \text{für alle } x, \ \text{an denen } F \ \text{stetig ist.}$$

Bemerkungen 8.2 (i) Der Satz von DE MOIVRE und LAPLACE ist also eine Aussage über schwache Konvergenz bzw. Konvergenz in Verteilung, denn Φ ist überall stetig.

(ii) Die Menge W nennt man *Konvergenz-determinierend*. Wir hatten in Satz 4.17 bereits gesehen, dass die Menge der gleichmäßig stetigen und beschränkten Funktionen eine Konvergenz-determinierende Menge ist.

(iii) Gibt es die Funktionenmenge W? Es seien

$$W_0 := \{f \in C_b(\mathbb{R}) : f^{(k)} \in C_b(\mathbb{R}) \text{ für alle } k \in \mathbb{N}\}$$

und

$$f_0(t) := \begin{cases} 1, & \text{falls } t \leq 0, \\ 0, & \text{falls } t \geq 1, \\ \frac{\int_t^1 \exp\left(-\frac{1}{s(1-s)}\right) ds}{\int_0^1 \exp\left(-\frac{1}{s(1-s)}\right) ds}, & \text{falls } 0 \leq t \leq 1. \end{cases}$$

Dann ist f_0 wohldefiniert mit $0 \leq f_0 \leq 1$ und die Ableitungen $f_0^{(k)}$ existieren und sind in $C_b(\mathbb{R})$ für alle $k \in \mathbb{N}$. Also ist $f_0 \in W_0$. Für $x < y$ setzen wir nun $f(z) := f_0(\frac{z-x}{y-x})$. Dann hat f die gewünschten Eigenschaften.

Beweis: (ii) $\Rightarrow$ (iii): Zu $x \in \mathbb{R}$ und $\varepsilon > 0$ sei $f \in W$ so, dass

$$1_{(-\infty,x)} \leq f \leq 1_{(-\infty,x+\varepsilon)}.$$

Dann ist

$$\begin{aligned} \limsup_{n\to\infty} F_n(x) &= \limsup_{n\to\infty} P_n((-\infty,x)) \leq \limsup_{n\to\infty} \int f\, dP_n = \int f\, dP \\ &\leq P((-\infty,x+\varepsilon)) = F(x+\varepsilon) \end{aligned}$$

für alle $\varepsilon > 0$. Ganz analog folgt $\liminf_{n\to\infty} F_n(x) \geq F(x-\varepsilon)$. Daraus folgt $\lim_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitsstellen von F.

(iii) $\Rightarrow$ (i): (heißt in der Literatur der Satz von HELLY und BRAY) Sei D die Menge der Stetigkeitsstellen von F. Da F monoton wächst, ist $\mathbb{R} \setminus D$ abzählbar, D ist also dicht in $\mathbb{R}$. Sei $f \in C_b(\mathbb{R})$ und $\varrho := \sup_{x\in\mathbb{R}} |f(x)|$. Es existieren $\mu, \nu \in \mathbb{R}$ mit $F(\mu) < \varepsilon$ und $1 - F(\nu) < \varepsilon$, und da D dicht ist, existieren $\mu, \nu \in D$. Nun ist $f|_{[\mu,\nu]}$ gleichmäßig stetig, also existieren

$$\mu = \lambda_0 < \lambda_1 < \cdots < \lambda_{k-1} < \lambda_k = \nu \quad \text{mit } \lambda_0, \ldots, \lambda_k \in D\ , k \in \mathbb{N}_0\ ,$$

so dass $|f(x) - f(\lambda_{i-1})| < \varepsilon$ für $\lambda_{i-1} \leq x \leq \lambda_i$. Setze nun

$$g := \sum_{i=1}^k f(\lambda_{i-1}) 1_{[\lambda_{i-1},\lambda_i)}.$$

Dann ist

$$\int g\, dP_n = \sum_{i=1}^k f(\lambda_{i-1})(F_n(\lambda_i) - F_n(\lambda_{i-1}))\ .$$

Nach Voraussetzung gilt dann $\int g\, dP_n \to \int g\, dP$.

Sei nun $M := [\mu, \nu)$, $L = (-\infty, \mu)$ und $R = [\nu, \infty)$. Für $x \in L \cup R$ ist $|f(x) - g(x)| = |f(x)| \le \varrho$, für $x \in M$ ist $|f(x) - g(x)| < \varepsilon$, also

$$\begin{aligned} \Big| \int f\,dP - \int g\,dP \Big| &\le \int |f-g|\,dP \\ &\le \varrho P(L) + \varrho P(R) + \varepsilon P(M) \\ &= \varrho F(\mu) + \varrho(1 - F(\nu)) + \varepsilon P(M) \\ &\le (2\varrho + 1)\varepsilon \ . \end{aligned}$$

Analog folgt

$$\Big| \int f\,dP_n - \int g\,dP_n \Big| \le \varrho F_n(\mu) + \varrho(1 - F_n(\nu)) + \varepsilon \ .$$

Da μ, ν Stetigkeitstellen von F sind, folgt

$$\Big| \int f\,dP_n - \int g\,dP_n \Big| < (2\varrho + 1)\varepsilon$$

für n hinreichend groß und die Behauptung des Satzes folgt mit Hilfe der Dreiecks-Ungleichung. $\square$

Wir wollen nun den folgenden Satz beweisen:

Satz 8.3 (zentraler Grenzwertsatz) *Für jedes $n \in \mathbb{N}$ seien $X_{n1}, \dots, X_{nk_n}$ unabhängige Zufallsgrößen auf $(\Omega_n, \mathcal{A}_n, P_n)$ mit $\mathbb{E}(X_{nj}) = 0$ für alle $n \in \mathbb{N}$, $j = 1, \dots, k_n$ und endlichen Varianzen σ_{nk}^2, $n \ge 1$, $k = 1, \dots, k_n$. Es sei $s_n^2 := \sum_{k=1}^{k_n} \sigma_{nk}^2 > 0$ und*

$$\frac{1}{s_n^2} \sum_{k=1}^{k_n} \int_{\{|X_{nk}| \ge \varepsilon s_n\}} X_{nk}^2\,dP_n \to 0 \quad \text{für} \quad n \to \infty \text{ für alle } \varepsilon > 0 \ , \tag{8.1}$$

dann gilt:

$$\lim_{n \to \infty} P_n\Big(\frac{S_n}{s_n} < y \Big) = \Phi(y)$$

für alle $y \in \mathbb{R}$.

Zunächst diskutieren wir die Bedingung (8.1), die einen eigenen Namen bekommt:

Definition 8.4 Man sagt, dass die Folge von Zufallsgrößen $X_{n1}, \dots, X_{nk_n}$, $n \in \mathbb{N}$, der Lindeberg-*Bedingung* genügt, falls (8.1) gilt.

Die Bedingung (8.1) besagt, dass jeder Summand $\frac{X_{nk}}{s_n}$ von $S_n^* := \frac{S_n}{s_n}$ für große n nur einen kleinen Beitrag zur Gesamtsumme S_n^* liefert. Dies wird in dem folgenden Lemma präzisiert.

Lemma 8.5 *Genügt $X_{n1}, \dots, X_{nk_n}$, $n \in \mathbb{N}$, der* Lindeberg-*Bedingung, so folgt:*

(i) $\max_{1\le k\le k_n} \frac{|X_{nk}|}{s_n} \to 0$ stochastisch.
(ii) $\lim_{n\to\infty} \frac{1}{s_n} \max_{1\le k\le k_n} \sigma_{nk} = 0$ *(*FELLER-Bedingung*)*.

Beweis: Zu (i): Es gilt mittels der Markov-Ungleichung

$$\begin{aligned} P_n\Big(\max_{1\le k\le k_n} \frac{|X_{nk}|}{s_n} \ge \varepsilon\Big) &= P_n\Big(\bigcup_{k=1}^{k_n}\Big\{\frac{|X_{nk}|}{s_n} \ge \varepsilon\Big\}\Big) \le \sum_{k=1}^{k_n} P_n(|X_{nk}| \ge \varepsilon s_n) \\ &\le \frac{1}{s_n^2\varepsilon^2}\sum_{k=1}^{k_n}\int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\,dP_n\ . \end{aligned}$$

Damit folgt mit Hilfe der LINDEBERG-Bedingung die Behauptung.
Zu (ii): Wir bezeichnen zur Vereinfachung der Notation zu festem n mit $\mathbb{E}$ den Erwartungswert bezüglich P_n. Es gilt

$$\begin{aligned} \sigma_{nk}^2 &= \mathbb{E}(X_{nk}^2) = \int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\,dP_n + \int_{\{|X_{nk}|<\varepsilon s_n\}} X_{nk}^2\,dP_n \\ &\le \int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\,dP_n + \varepsilon^2 s_n^2, \end{aligned}$$

und somit

$$\limsup_{n\to\infty} \max_{1\le k\le k_n} \frac{\sigma_{nk}^2}{s_n^2} \le \varepsilon^2 \text{ für alle } \varepsilon > 0\ .$$

□

Lemma 8.6 *Sind die Zufallsgrößen* $\big(X_{nk}/\sigma_{nk}\big)_{n\ge1}^{k=1,\dots,k_n}$ *zusätzlich identisch verteilt, so folgt aus der* FELLER*schen Bedingung die* LINDEBERG-*Bedingung.*

Beweis: Setze $\varrho_n := \max_{1\le k\le k_n} \frac{\sigma_{nk}}{s_n}$, dann gilt

$$\begin{aligned} \int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\,dP_n &= \sigma_{nk}^2 \int_{\big\{\frac{|X_{nk}|}{\sigma_{nk}}\ge\frac{\varepsilon}{\sigma_{nk}/s_n}\big\}} \Big(\frac{X_{nk}}{\sigma_{nk}}\Big)^2 dP_n \\ &\le \sigma_{nk}^2 \int_{\big\{\frac{|X_{nk}|}{\sigma_{nk}}\ge\frac{\varepsilon}{\varrho_n}\big\}} \Big(\frac{X_{nk}}{\sigma_{nk}}\Big)^2 dP_n \\ &= \sigma_{nk}^2 \int 1_{[\varepsilon/\varrho_n,\infty)}(x)\ x^2\ P_n^{\frac{X_{nk}}{\sigma_{nk}}}(dx). \end{aligned}$$

Die rechte Seite konvergiert für $n\to\infty$ gegen Null, denn die Verteilung $P_n^{\frac{X_{nk}}{\sigma_{nk}}}$ ist unabhängig von n. □

Korollar 8.7 *Für jedes* $n \ge 1$ *seien* $X_{n1},\dots,X_{nk_n}$ *unabhängig und die* $(X_{nk})_{n\ge1}^{k=1,\dots,k_n}$ *seien identisch verteilt mit* $\mathbb{E}(X_{11}) = 0$, $\sigma_{11}^2 = \mathbb{E}(X_{11}^2) < \infty$ *und* $s_n^2 = \sum_{k=1}^{k_n}\sigma_{nk}^2 > 0$. *Dann gilt*

$$\lim_{n\to\infty} P_n\Big(\frac{\sum_{k=1}^{k_n} X_{nk}}{s_n} \le y\Big) = \Phi(y) \ \ \textit{für alle}\ \ y\in\mathbb{R}\ .$$

Erneut bezeichnet hier $\mathbb{E}$ den Erwatungswert bezüglich P_n zu festem n.

Beweis: Nach Lemma 8.6 genügt es, $\varrho_n = \frac{1}{s_n}\max_{1\le k\le k_n}\sigma_{nk} \to 0$ für $n\to\infty$ zu zeigen. Hier ist nun $\varrho_n = \frac{\sigma_{nk}}{\sqrt{k_n\sigma_{nk}^2}} \to 0$ für $n\to\infty$. □

Korollar 8.8 *Sei in der Situation von Korollar 8.7 $X_{nk} = X_k$, $1\le k\le k_n = n$, $n\ge 1$, also eine Folge von Zufallsgrößen, definiert auf einem W-Raum $(\Omega,\mathcal{A},P)$. Dann gilt*

$$\liminf_{n\to\infty}\frac{S_n}{\sqrt{n}} = -\infty \ \textit{und}\ \limsup_{n\to\infty}\frac{S_n}{\sqrt{n}} = +\infty$$

P-fast sicher.
(Nachtrag zur Diskussion vor Satz 6.11)

Beweis: Ohne Einschränkung sei $\mathrm{Var}(X_{n1}) = 1$. Wir wissen, dass $\liminf_{n\to\infty} S_n/\sqrt{n}$ und $\limsup_{n\to\infty} S_n/\sqrt{n}$ terminale Funktionen sind, also konstant α (bzw. $-\alpha$) mit $\alpha\in\bar{\mathbb{R}}$, siehe Korollar 5.20. Angenommen $-\alpha > -\infty$. Dann gilt für $t\in(-\infty,-\alpha)$

$$\begin{aligned} 0 &= P\Big(\liminf_{n\to\infty} S_n/\sqrt{n} < -\alpha\Big) = P\Big(\bigcap_{n\ge 1}\Big\{\inf_{m\ge n} S_m/\sqrt{m} < -\alpha\Big\}\Big) \\ &\ge \lim_{n\to\infty} P\Big(\inf_{m\ge n} S_m/\sqrt{m} < t\Big) \ge \lim_{n\to\infty} P\Big(S_n/\sqrt{n} < t\Big) = \Phi(t) > 0\ , \end{aligned}$$

also ein Widerspruch. □

Wir kommen nun zum Beweis von Satz 8.3. Nach Satz 8.1 wollen wir zeigen:

$$\lim_{n\to\infty}\int f\left(\frac{S_n}{s_n}\right) dP_n = \mathbb{E}\big(f(N)\big)\ \text{ für alle } f\in W\ ,$$

wobei N eine standardnormalverteilte Zufallsgröße bezeichne.
Die Idee, die auf P. Lévy zurückgeht, wird sein, auf der linken Seite die X_{nk} durch $N(0,\sigma_{nk}^2)$-verteilte Zufallsgrößen zu ersetzen, und zwar sukzessive. Man erhält dann eine Folge der Form

$$\begin{gathered} \int f\Big(\frac{X_{n1}+\cdots+X_{nk_n}}{s_n}\Big)\, dP_n\ , \\ \int f\Big(\frac{X_{n1}+\cdots+X_{nk_n-1}+Y_{nk_n}}{s_n}\Big)\, dP_n\ ,\ldots, \\ \int f\Big(\frac{X_{n1}+Y_{n2}+\cdots+Y_{nk_n}}{s_n}\Big)\, dP_n\ , \\ \int f\Big(\frac{Y_{n1}+\cdots+Y_{nk_n}}{s_n}\Big)\, dP_n\ . \end{gathered}$$

Wir zeigen dann, dass jedes Glied dieser Folge für große n so nahe beim Nächsten liegt, dass sogar das erste und letzte Glied nahe zusammen liegen! Wenn die Y_{nk}, $k=1,\ldots,k_n$, unabhängig gewählt werden, ist das letzte Glied gleich

$\mathbb{E}\big(f(N)\big)$. Diesen Sachverhalt kennt man aus der Vorlesung des dritten Semesters:

Satz 8.9 *Seien X_1 und X_2 unabhängige Zufallsgrößen und $P^{X_1} = N(\mu, \sigma^2)$, $P^{X_2} = N(\nu, \tau^2)$, dann ist $P^{X_1} * P^{X_2} = N(\mu + \nu, \sigma^2 + \tau^2)$.*

Beweis: Übung; wir werden in Beispiel 9.6 einen Beweis sehen.

Die Aussage und die Voraussetzung des Satzes 8.3 sind nur abhängig von $P_n^{(X_{n1},\dots,X_{nk_n})}$. Wir werden nun die gegebene Folge $X_{n1}, \dots, X_{nk_n}$ von Zufallsgrößen auf $(\Omega_n, \mathcal{A}_n, P_n)$ geeignet durch eine unabhängige Folge $\tilde{X}_{n1}, \dots, \tilde{X}_{nk_n}, Y_{n1}, \dots, Y_{nk_n}$ von Zufallsgrößen auf einem W-Raum $(\tilde{\Omega}_n, \tilde{\mathcal{A}}_n, \tilde{P}_n)$ ersetzen: Dazu sei für $n \geq 1$ $(\tilde{\Omega}_n, \tilde{\mathcal{A}}_n, \tilde{P}_n)$ ein W-Raum und

$$\mathcal{N} := \Big\{ \tilde{X} = \big((\tilde{X}_{n1}, \dots, \tilde{X}_{nk_n})\big)_{n\geq 1} \Big| \forall n \geq 1 : \tilde{X}_{n1}, \dots, \tilde{X}_{nk_n} \in \mathcal{L}^2(\tilde{P}_n),$$
$$\text{unabhängig, zentriert und } s_n^2 > 0,\ \tilde{X} \text{ genüge der Lindeberg-Bedingung} \Big\},$$

sowie

$$\mathcal{N}_0 := \Big\{ Y \in \mathcal{N} : \lim_{n\to\infty} \tilde{P}_n\Big(\frac{\sum_{k=1}^{k_n} Y_{nk}}{s_n} \leq y \Big) = \Phi(y) \text{ für alle } y \in \mathbb{R} \Big\} .$$

Wir zeigen dass W-Räume $(\tilde{\Omega}_n, \tilde{\mathcal{A}}_n, \tilde{P}_n)$, $n \geq 1$, und Zufallsgrößen $\tilde{X}_{n1}, \dots, \tilde{X}_{nk_n}, Y_{n1}, \dots, Y_{nk_n}$ darauf existieren mit:

(i) $\tilde{P}_n^{\tilde{X}_{nk}} = P_n^{X_{nk}}$, $n \geq 1, k = 1, \dots, k_n$
(ii) $\mathrm{Var}(\tilde{X}_{nk}) = \mathrm{Var}(Y_{nk})$, $n \geq 1, k = 1, \dots, k_n$
(iii) $\tilde{X}_{n1}, \dots, \tilde{X}_{nk_n}, Y_{n1}, \dots, Y_{nk_n}$ sind unabhängig
(iv) $Y = \big((Y_{n1}, \dots, Y_{nk_n})\big)_{n\geq 1} \in \mathcal{N}_0$.

In (ii) sind die Varianz bezüglich $\tilde{P}_n$ gemeint. Präziser müssen wir $\mathrm{Var}_{\tilde{P}_n}(\tilde{X}_{nk})$ schreiben. Im Fall der obigen Konstruktion gilt $\tilde{X} = \big((\tilde{X}_{n1}, \dots, \tilde{X}_{nk_n})\big)_{n\geq 1} \in \mathcal{N}$, denn

$$\mathrm{Var}(\tilde{X}_{nk}) = \mathrm{Var}(X_{nk}) = \sigma_{nk}^2 \text{ , also } s_n^2 > 0,$$

und

$$\begin{aligned}
\int_{\{|X_{nk}|\geq \varepsilon s_n\}} X_{nk}^2 \, dP_n &= \int 1_{[\varepsilon s_n, +\infty)}(|X_{nk}|)|X_{nk}|^2 \, dP_n \\
&= \int 1_{[\varepsilon s_n, +\infty)}(|x|)|x|^2 \, dP_n^{X_{nk}}(x) \\
&= \int 1_{[\varepsilon s_n, +\infty)}(|x|)|x|^2 \, d\tilde{P}_n^{\tilde{X}_{nk}}(x) ,
\end{aligned}$$

also erfüllt $\tilde{X}$ die Lindeberg-Bedingung.

Das sogenannte Invarianzprinzip führt uns sogar zu $\tilde{X} \in \mathcal{N}_0$, d.h. $\lim_{n\to\infty} \tilde{P}_n\Big(\sum_{k=1}^{k_n} \tilde{X}_{nk}/s_n \leq y\Big) = \Phi(y)$, aber $\tilde{P}_n\Big(\sum_{k=1}^{k_n} \tilde{X}_{nk}/s_n \leq y\Big) = P_n\Big(\sum_{k=1}^{k_n} X_{nk}/s_n \leq y\Big)$, denn

$$\begin{aligned} P_n^{(X_{n1},\dots,X_{nk_n})} &= P_n^{X_{n1}} \otimes \cdots \otimes P_n^{X_{nk_n}} \\ &= \tilde{P}_n^{\tilde{X}_{n1}} \otimes \cdots \otimes \tilde{P}_n^{\tilde{X}_{nk_n}} = \tilde{P}_n^{(\tilde{X}_{n1},\dots,\tilde{X}_{nk_n})}, \end{aligned}$$

womit alles gezeigt ist.

Satz 8.10 (Invarianzprinzip) *Es seien $\tilde{X} \in \mathcal{N}$ und $Y \in \mathcal{N}_0$ mit*

(i) $\operatorname{Var}(\tilde{X}_{nk}) = \operatorname{Var}(Y_{nk})$, $n \geq 1$, $k = 1, \dots, k_n$

(ii) $\tilde{X}_{n1}, \dots, \tilde{X}_{nk_n}, Y_{n1}, \dots, Y_{nk_n}$ *sind unabhängig für jedes* $n \geq 1$.

Dann gilt $\tilde{X} \in \mathcal{N}_0$.

Wir müssen dieses Prinzip beweisen, zeigen aber zuvor, dass obige Konstruktion mit (i), ..., (iv) möglich ist. Dazu seien

$$\tilde{\Omega}_n := \mathbb{R}^{2k_n}\ ,\ \tilde{\mathcal{A}}_n := \mathcal{B}^{2k_n} \text{ und } \tilde{P}_n := P_n^{X_{n1}} \otimes \cdots \otimes P_n^{X_{nk_n}} \otimes Q^{n1} \otimes \cdots \otimes Q^{nk_n}$$

mit $Q^{nk} = N(0, \sigma_{nk}^2)$, $k = 1, \dots, k_n$. Weiter seien

$$\tilde{X}_{nk} : \tilde{\Omega}_n \to \mathbb{R} \text{ mit } \tilde{X}_{nk}(x_{n1}, \dots, x_{nk_n}, y_{n1}, \dots, y_{nk_n}) = x_{nk}$$

und

$$\tilde{Y}_{nk} : \tilde{\Omega}_n \to \mathbb{R} \text{ mit } \tilde{Y}_{nk}(x_{n1}, \dots, x_{nk_n}, y_{n1}, \dots, y_{nk_n}) = y_{nk}$$

für $k = 1, \dots, k_n$. Die Abbildungen sind Projektionen. Dann gilt (i)-(iii) nach Konstruktion. Für (iv) müssen wir $Y \in \mathcal{N}$ zeigen, was nach Definition von $\mathcal{N}$ sich darauf beschränkt, die Lindeberg-Bedingung nachzurechnen. Nun ist $\frac{Y_{nk}}{\sigma_{nk}}$ $N(0,1)$-verteilt für alle $k = 1, \dots, k_n$, also zeigen wir gemäß Lemma 8.6

$$\lim_{n\to\infty} \frac{1}{s_n} \max_{1\leq k\leq k_n} \sigma_{nk} = 0.$$

Da $\sigma_{nk}^2 = \operatorname{Var}(X_{nk})$ und X nach Voraussetzung die Lindeberg-Bedingung erfüllt, folgt dies aber unmittelbar aus Lemma 8.5(ii). Also ist $Y \in \mathcal{N}$. Mit Satz 8.9 ist weiter

$$\tilde{P}_n\Big(\frac{\sum_{k=1}^{k_n} Y_{nk}}{s_n}\Big) = \Phi(y) \qquad \text{für alle } n \in \mathbb{N}$$

und für alle $y \in \mathbb{R}$, also ist $Y \in \mathcal{N}_0$.
Es bleibt also, das Invarianzprinzip zu beweisen:

Beweis: (von Satz 8.10) 1. Schritt: Es sei $f \in W_0$ und für $x \in \mathbb{R}$

$$g(x,h) := \begin{cases} \frac{2}{h^2}\big(f(x+h) - f(x) - h\,f'(x)\big) - f''(x), & \text{falls } h \neq 0, \\ 0, & \text{falls } h = 0, \end{cases}$$

also

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}(f''(x) + g(x,h)) \ , x, h \in \mathbb{R} \ . \tag{8.2}$$

Nach Taylor ist

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x + \theta h)$$

für ein $0 < \theta < 1$, welches von x und h abhängt. Damit folgt $|g(x,h)| \leq 2 \sup_{x\in\mathbb{R}} |f''(x)| =: A$ für alle $x, h \in \mathbb{R}$ (A existiert, da $f \in W_0$). Nach Taylor gilt auch

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + \frac{h^3}{6} f'''(x + \theta' h)$$

mit $0 < \theta' < 1$, also $|g(x,h)| \leq |h| \frac{1}{3} \sup_{x\in\mathbb{R}} |f'''(x)| =: |h|B$ (B existiert, da $f \in W_0$). Insgesamt haben wir (8.2) mit

$$|g(x,h)| \leq \min\{A, |h|B\} =: d(h), \ \ x, h \in \mathbb{R} \ .$$

2. Schritt: Wir setzen nun $\tilde{T}_n := \frac{1}{s_n} \sum_{k=1}^{k_n} \tilde{X}_{nk}$, $\hat{T}_n := \frac{1}{s_n} \sum_{k=1}^{k_n} Y_{nk}$, sowie $\tilde{P}_n^{\tilde{T}_n} =: \tilde{Q}_n$ und $\tilde{P}_n^{\hat{T}_n} =: \hat{Q}_n$. Zu zeigen:

$$\lim_{n\to\infty} \int f \, d\tilde{Q}_n = \mathbb{E}\big(f(N)\big) \quad \text{für alle } f \in W_0 \ .$$

Also ist zu zeigen:

$$\lim_{n\to\infty} \Big| \int f \, d\tilde{Q}_n - \int f \, d\hat{Q}_n \Big| = 0 \ \text{ für alle } f \in W_0 \ .$$

Es gilt $\int f \, d\tilde{Q}_n - \int f \, d\hat{Q}_n = \int f \circ \tilde{T}_n \, d\tilde{P}_n - \int f \circ \hat{T}_n \, d\tilde{P}_n$. Sei nun

$$U_{nk} := \frac{1}{s_n}(\tilde{X}_{n1} + \cdots + \tilde{X}_{n,k-1} + Y_{n,k+1} + \cdots + Y_{nk_n})$$

für jedes $k = 1, \ldots, k_n$, also

$$f \circ \tilde{T}_n - f \circ \hat{T}_n = \sum_{k=1}^{k_n} \Big(f\big(U_{nk} + \frac{\tilde{X}_{nk}}{s_n}\big) - f\big(U_{nk} + \frac{Y_{nk}}{s_n}\big)\Big) \ .$$

Es gilt

$$\begin{aligned}
&\mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk} + \frac{\tilde{X}_{nk}}{s_n}\big)\Big) \\
&= \mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk}\big)\Big) + \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}}{s_n} f'(U_{nk})\Big) + \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}^2}{2s_n^2} f''\big(U_{nk}\big)\Big) \\
&+ \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}^2}{2s_n^2} g\big(U_{nk}, \frac{\tilde{X}_{nk}}{s_n}\big)\Big) \\
&= \mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk}\big)\Big) + \frac{1}{2s_n^2} \operatorname{Var}_{\tilde{P}_n}(\tilde{X}_{nk}) \mathbb{E}_{\tilde{P}_n}\big(f''(U_{nk})\big) + \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}^2}{2s_n^2} g\big(U_{nk}, \frac{\tilde{X}_{nk}}{s_n}\big)\Big) \ ,
\end{aligned}$$

denn die $\tilde{X}_{nk}$'s sind zentriert und wir haben Satz 5.21 verwendet. Für $\mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk}+\frac{Y_{nk}}{s_n}\big)\Big)$ folgt die analoge Identität, wenn wir $\tilde{X}_{nk}$ durch Y_{nk} ersetzen. Mit $\mathrm{Var}_{\tilde{P}_n}(\tilde{X}_{nk}) = \mathrm{Var}_{\tilde{P}_n}(Y_{nk})$ folgt

$$\int f\,d\tilde{Q}_n - \int f\,d\hat{Q}_n = \frac{1}{2s_n^2}\sum_{k=1}^{k_n}\Big(\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 g\big(U_{nk},\frac{\tilde{X}_{nk}}{s_n}\big)\Big) - \mathbb{E}_{\tilde{P}_n}\Big(Y_{nk}^2 g\big(U_{nk},\frac{\tilde{X}_{nk}}{s_n}\big)\Big)\Big),$$

also

$$\Big|\int f\,d\tilde{Q}_n - \int f\,d\hat{Q}_n\Big| \le \frac{1}{2s_n^2}\sum_{k=1}^{k_n}\Big(\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big) + \mathbb{E}_{\tilde{P}_n}\Big(Y_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big)\Big)\ .$$

Wenn wir nun zeigen, dass die rechte Seite gegen Null konvergiert für $n \to \infty$, sind wir fertig. Es gilt

$$\begin{aligned}\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big) &= \int_{\{|\tilde{X}_{nk}|\ge\delta s_n\}} \tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\,d\tilde{P}_n + \int_{\{|\tilde{X}_{nk}|<\delta s_n\}} \tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\,d\tilde{P}_n \\ &\le A\int_{\{|\tilde{X}_{nk}|\ge\delta s_n\}} \tilde{X}_{nk}^2\,d\tilde{P}_n + B\delta\,\mathrm{Var}_{\tilde{P}_n}(\tilde{X}_{nk})\ ,\end{aligned}$$

also für alle $n \in \mathbb{N}$ und $\delta > 0$.

$$\frac{1}{2s_n^2}\sum_{k=1}^{k_n}\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big) \le \frac{A}{2s_n^2}\sum_{k=1}^{k_n}\int_{\{|\tilde{X}_{nk}|\ge\delta s_n\}} \tilde{X}_{nk}^2\,d\tilde{P}_n + \frac{B\delta}{2s_n^2}s_n^2\ .$$

Die rechte Seite wird gemäß der LINDEBERG-Bedingung klein in $n \in \mathbb{N}$. Analog folgt

$$\lim_{n\to\infty}\frac{1}{2s_n^2}\sum_{k=1}^{k_n}\mathbb{E}_{\tilde{P}_n}\Big(Y_{nk}^2 d\big(\frac{Y_{nk}}{s_n}\big)\Big) = 0\ .$$

□

Wir tragen ein Resultat zum Kapitel über große Abweichungen nach:

Satz 8.11 *Für jedes $y \in \mathbb{R}$ gilt*

$$\lim_{n\to\infty}\frac{1}{n}\log P\Big(\frac{S_n}{n} \ge y\Big) = -\inf_{x\ge y} I(x)\ .$$

Beweis: Da $[x, x+\delta) \subset [y,\infty)$ gilt für alle $x \ge y$ und für alle $\delta > 0$

$$\liminf_{n\to\infty}\frac{1}{n}\log P\Big(\frac{S_n}{n}\ge y\Big) \ge \sup_{x\ge y}\liminf_{n\to\infty}\frac{1}{n}\log P\Big(\frac{S_n}{n}\in[x,x+\delta)\Big)\ .$$

Wir müssen daher zeigen:

$$\liminf_{n\to\infty}\frac{1}{n}\log P\Big(\frac{S_n}{n}\in[x,x+\delta)\Big) \ge -I(x)\ .$$

Dazu gehen wir in den Beweis von Satz 7.1(ii). Erneut können wir ohne Einschränkung $x = 0$ wählen und $[0, \delta)$ und $[0, \varepsilon)$ ersetzt $(-\delta, \delta)$ und $(-\varepsilon, \varepsilon)$. Nun liefert der zentrale Grenzwertsatz

$$\lim_{n\to\infty} \hat{P}\Big(\frac{\hat{S}_n}{n} \in [0, \varepsilon)\Big) = \frac{1}{2} \ ,$$

womit alles gezeigt ist. $\square$

KAPITEL 9

Charakteristische Funktionen und Verteilungskonvergenz

Die Theorie charakteristischer Funktionen soll hier nur ganz kurz angerissen werden.

Definition 9.1 Es sei μ ein W-Maß auf $(\mathbb{R}^n, \mathcal{B}^n)$. Die *charakteristische Funktion* $\hat{\mu}$ von μ ist die Abbildung von $\mathbb{R}^n$ nach $\mathbb{C}$, die durch

$$\hat{\mu}(t) = \int e^{i\langle t,\, x\rangle}\, \mu(dx) = \int \cos\bigl(\langle t,\, x\rangle\bigr)\, \mu(dx) + i \int \sin\bigl(\langle t,\, x\rangle\bigr)\, \mu(dx), \ \ t \in \mathbb{R}^n,$$

definiert ist. Hier ist $\langle t,\, x\rangle = \sum_{j=1}^n t_j x_j$. Die charakteristische Funktion eines Zufallsvektors X ist die charakteristische Funktion der Verteilung von X, sie kann nach Satz 1.23 als $\mathbb{E}\bigl(\exp(i\,\langle t,\, X\rangle)\bigr)$ geschrieben werden.

$\hat{\mu}$ ist für alle $t \in \mathbb{R}^n$ definiert, denn Sinus und Cosinus sind beschränkt. Weiter ist $\hat{\mu}$ in t stetig nach dem Satz von der dominierten Konvergenz, Satz 2.5 (v).

Satz 9.2 (Eindeutigkeitssatz) *Es seien μ, ν zwei W-Maße auf $(\mathbb{R}^n, \mathcal{B}^n)$. Gilt $\hat{\mu}(t) = \hat{\nu}(t)$ für alle $t \in \mathbb{R}^n$, so gilt $\mu = \nu$.*

Beweis: Kompakte Mengen in $\mathbb{R}^n$ sind ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}^n$, also genügt es $\mu(K) = \nu(K)$ für alle kompakten Mengen K zu zeigen (siehe Satz 1.9). Für eine kompakte Menge K und $m \in \mathbb{N}$ sei

$$f_m(x) = \begin{cases} 1 & ,\ x \in K \\ 0 & ,\ \text{für } d(x,\, K) := \inf\bigl\{|x-y|,\ y \in K\bigr\} \geq 1/m \\ 1 - m\, d(x,\, K) & ,\ \text{sonst.} \end{cases}$$

Dann ist

(a) $0 \leq f_m(x) \leq 1$ für alle $x \in \mathbb{R}^n$,

(b) f_m ist stetig und

(c) $f_m(x) \downarrow 1_K(x)$ für $m \to \infty$.

Falls $\int f_m\, d\mu = \int f_m\, d\nu$ für alle $m \in \mathbb{N}$, so folgt $\mu(K) = \nu(K)$ mit dem Satz von der dominierten Konvergenz aus (c). Es genügt also nachzuweisen, dass $\int f\, d\mu = \int f\, d\nu$ für alle f gilt, die (a) und (b) erfüllen und einen kompakten Träger haben[1].

[1]der *Träger* von f ist die Menge $\overline{\{x : f(x) \neq 0\}}$

Sei $\varepsilon > 0$ und $N > 0$ so groß, dass

$$B_N := [-N,\, N]^n \supset \{x : f(x) \neq 0\}$$

und $\max\bigl\{\mu(B_N^c),\, \nu(B_N^c)\bigr\} \le \varepsilon$ gelten. Nach dem WEIERSTRASSschen Approximationssatz gibt es eine Funktion $g : \mathbb{R}^n \to \mathbb{C}$ der Form

$$g(x) = \sum_{j=1}^{m} c_j \exp\Bigl(i\bigl\langle \frac{\pi}{N} t_j,\, x\bigr\rangle\Bigr)$$

mit $c_j \in \mathbb{C}$ und $t_j \in \mathbb{Z}^n$, die periodisch in jeder Komponente ist und f in B_N bis auf ε approximiert:

$$\sup\bigl\{|f(x) - g(x)|;\, x \in B_N\bigr\} \le \varepsilon\,.$$

Es folgen $\sup_{x\in\mathbb{R}^n} |g(x)| \le 1 + \varepsilon$ und

$$\begin{aligned}\Bigl|\int f\,d\mu - \int f\,d\nu\Bigr| \le \Bigl|\int f\,d\mu - \int g\,d\mu\Bigr| + \Bigl|\int g\,d\mu - \int g\,d\nu\Bigr| \\ + \Bigl|\int g\,d\nu - \int f\,d\nu\Bigr|\,.\end{aligned}$$

Nach Voraussetzung ist $\hat{\mu} = \hat{\nu}$, also ist der zweite Summand gleich Null. Es gilt

$$\begin{aligned}\Bigl|\int f\,d\mu - \int g\,d\mu\Bigr| &\le \Bigl|\int_{B_N} f\,d\mu - \int_{B_N} g\,d\mu\Bigr| + \Bigl|\int_{B_N^c} f\,d\mu\Bigr| + \Bigl|\int_{B_N^c} g\,d\mu\Bigr| \\ &\le \int_{B_N} |f - g|\,d\mu + 0 + (1+\varepsilon)\mu(B_N^c) \\ &\le \varepsilon\mu(B_N) + (1+\varepsilon)\mu(B_N^c) \\ &\le \varepsilon(2+\varepsilon)\,.\end{aligned}$$

Der dritte Summand wird analog abgeschätzt. Da $\varepsilon > 0$ beliebig war, folgt $\int f\,d\mu = \int g\,d\nu$. □

Beispiele 9.3 **(a)** Für $a \in \mathbb{R}^n$ betrachte das DIRAC-Maß δ_a, dann ist $\hat{\delta}_a = e^{i\langle t,\, a\rangle}$, $t \in \mathbb{R}^n$, also gilt für $\mu = \sum_{i=1}^{\infty} a_i \delta_{x_i}$ mit $\sum_{i=1}^{\infty} a_i = 1$ und $x_i \in \mathbb{R}^n$:

$$\hat{\mu}(t) = \sum_{i=1}^{\infty} a_i e^{i\langle t, x_i\rangle}\,, \quad t \in \mathbb{R}^n\,.$$

Insbesondere

$$\begin{aligned}\widehat{b(n,\, p)}(t) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{ikt} \\ &= \bigl((1-p) + pe^{it}\bigr)^n\,, \quad t \in \mathbb{R}\end{aligned}$$

und

$$\hat{\pi}_\alpha(t) = e^{-\alpha} \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} e^{itn} = e^{\alpha(e^{it}-1)}\,, \quad t \in \mathbb{R}\,.$$

(b) μ sei standard-normalverteilt. Dann gilt

$$\hat{\mu}(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_{\mathbb{R}} e^{-(x-it)^2/2}\, dx.$$

Es gilt $\int_{\mathbb{R}} e^{-(x-it)^2/2}\, dx = \sqrt{2\pi}$ (Übung), also $\hat{\mu}(t) = e^{-t^2/2}$.

(c) μ sei die CAUCHY-Verteilung zum Parameter $\alpha > 0$. Dann gilt

$$\hat{\mu}(t) = \frac{\alpha}{\pi} \int_{\mathbb{R}} e^{itx} \frac{dx}{\alpha^2 + x^2}\,, \quad t \in \mathbb{R}\,.$$

Die Funktion $z \mapsto \frac{1}{\alpha^2+z^2}$ hat Pole in $\pm i\alpha$. Mit Hilfe des Residuensatzes ergibt sich $\hat{\mu}(t) = e^{-\alpha|t|}$.

(d) μ sei die Standardnormalverteilung in $(\mathbb{R}^n, \mathcal{B}^n)$. Mit (b) folgt

$$\hat{\mu}(t) = \exp\Big(-\frac{1}{2}\sum_{j=1}^{n} t_j^2\Big) = e^{-\langle t,\, t\rangle/2}$$

für $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$.

Gemäß Definition 1.23 ist die allgemeine Normalverteilung das Bildmaß $\nu = \mu\phi^{-1}$ der Standardnormalverteilung μ unter einer affinen Transformation $\mathbb{R}^n \ni x \mapsto \phi(x) = Ax + b$. Es gilt

$$\begin{aligned}
\hat{\nu}(t) = \int e^{i\langle t,\, x\rangle}\, \nu(dx) &= \int e^{i\langle t,\, \phi(x)\rangle}\, \mu(dx) \\
&= e^{i\langle t,\, b\rangle} \int e^{i\left\langle A^t t,\, x\right\rangle}\, \mu(dx) \\
&= e^{i\langle t,\, b\rangle} \hat{\mu}(A^t t) = e^{i\langle t,\, b\rangle} e^{-\left\langle A^t t,\, A^t t\right\rangle/2} \\
&= \exp\Big(i\,\langle t,\, b\rangle - \frac{1}{2}\,\langle t,\, \Sigma t\rangle\Big)
\end{aligned}$$

mit $\Sigma = AA^t$ als der Kovarianzmatrix von ν (siehe Satz 1.25, Beispiel 2.19 (e)).

Satz 9.4 *Für jedes $b \in \mathbb{R}^n$ und jede positiv semi-definite, symmetrische Matrix Σ gibt es genau eine Normalverteilung mit b als Erwartungswert und Σ als Kovarianzmatrix.*

Beweis: Eindeutigkeit: Satz 9.2 und Beispiel 9.3 (d). Existenz: Es existiert mindestens eine $n \times n$-Matrix A mit $AA^t = \Sigma$. $\square$

Satz 9.5 *Es seien X, Y zwei unabhängige Zufallsgrößen mit charakteristischen Funktionen χ_X bzw. χ_Y. Dann ist $\chi_X \cdot \chi_Y$ die charakteristische Funktion von $X + Y$.*

Beweis: Für alle $t \in \mathbb{R}$ gilt

$$\mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX} e^{itY}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY})\,,$$

da e^{itX} unabhängig von e^{itY} ist. Aber Satz 5.21 wurde nur für reellwertige Zufallsvariablen bewiesen. Eine Zerlegung in Real- und Imaginärteil liefert jedoch die entsprechende Aussage für komplexwertige Zufallsvariablen. □

Beispiele 9.6 **(a)** Sind X, Y unabhängig und Cauchy-verteilt, so ist für $\lambda \in (0,\, 1)$

$$\lambda X + (1-\lambda)Y$$

auch Cauchy-verteilt, denn für $t \in \mathbb{R}$ gilt

$$\begin{aligned}\chi_{\lambda X+(1-\lambda)Y}(t) &= \mathbb{E}\Big(\exp\big(it(\lambda X + (1-\lambda)Y)\big)\Big)\\ &= \chi_X(\lambda t)\cdot \chi_Y\big((1-\lambda)t\big)\\ &= \exp(-|\lambda t|)\exp\big(-|(1-\lambda)t|\big) = e^{-|t|}\,.\end{aligned}$$

(b) Beweis von Satz 8.9 (elegant!):

$$\begin{aligned}\chi_{X_1+X_2}(t) &= \chi_{X_1}(t)\chi_{X_2}(t)\\ &= \exp\Big(i\mu t - \frac{\sigma^2}{2}t^2\Big)\cdot \exp\Big(i\nu t - \frac{t^2}{2}t^2\Big)\\ &= \exp\Big(i(\mu+\nu)t - \frac{1}{2}(\sigma^2+\tau^2)t^2\Big)\,.\end{aligned}$$

□

Sind $(X_n)_n$ unabhängig und Cauchy-verteilt, so folgt mit Beispiel 9.6 (a), dass S_n/n auch Cauchy-verteilt ist. Für $c \in \mathbb{R}$ gilt

$$0 < \int_c^\infty \frac{1}{\pi}\frac{1}{1+x^2}\,dx = P\Big(\frac{S_n}{n} \geq c\Big) \leq P\Big(\sup_{k\geq n}\frac{S_k}{k} \geq c\Big) \overset{n\to\infty}{\longrightarrow} P\Big(\limsup_{n\to\infty}\frac{S_n}{n} \geq c\Big)\,.$$

Dann muss nach Lemma 6.3 und Kolmogorovs 0-1-Gesetz

$$P\Big(\limsup_{n\to\infty}\frac{S_n}{n} \geq c\Big) = 1$$

für alle $c \in \mathbb{R}$ gelten, also

$$P\Big(\limsup_{n\to\infty}\frac{S_n}{n} = \infty) = 1\,.$$

Analog zeigt man $P(\liminf_{n\to\infty} S_n/n = -\infty) = 1$.

S_n/n konvergiert nicht fast sicher, was dem Gesetz der großen Zahlen aber nicht widerspricht.

Sind $(X_n)_n$ unabhängig und standard-normalverteilt, so folgt mit Beispiel 9.6 (b), dass $S_n/\sqrt{n}$ $N(0,1)$-verteilt ist. Wie oben folgt dann, dass

$$\limsup_{n\to\infty}\frac{S_n}{\sqrt{n}} = \infty \quad \text{und} \quad \liminf_{n\to\infty}\frac{S_n}{\sqrt{n}} = -\infty$$

fast sicher, was wir mit Korollar 8.8 auch allgemein kennen.

Wir untersuchen den Begriff der Konvergenz in Verteilung, siehe Definition 4.15, für allgemeinere Wertebereiche. Sei $(S,\, d)$ ein metrischer Raum.

Definition 9.7 (i) Die Borel-σ-Algebra $\mathcal{B}_S$ sei die kleinste σ-Algebra auf S, die die offenen Mengen enthält. ($\mathcal{B}_S$ wird auch von der Familie der abgeschlossenen Mengen erzeugt!)
(ii) Mit $\mathcal{M}_1(S)$ sei die Menge der W-Maße auf $(S,\, \mathcal{B}_S)$ bezeichnet.
(iii) $C(S)$ bezeichne die Menge der beschränkten stetigen Funktionen von S nach $\mathbb{R}$.

Bemerkung 9.8 Es seien $\mu, \nu \in \mathcal{M}_1(S)$. Gilt $\int f\, d\mu = \int f\, d\nu$ für alle $f \in C(S)$, so gilt $\mu = \nu$.

Beweis: Sei $F \subset C$ abgeschlossen und für $n \in \mathbb{N}$ sei $f_n \in C(S)$ definiert durch

$$f_n(x) = \max\big\{\big(1 - n\, d(x,\, F)\big), 0\big\}\,.$$

Dann gilt $f_n \downarrow 1_F$ für $n \to \infty$ und aus dem Satz von der dominierten Konvergenz folgt

$$\mu(F) = \lim_{n\to\infty} \int f_n\, d\mu = \lim_{n\to\infty} \int f_n\, d\nu = \nu(F)\,.$$

Da die abgeschlossenen Mengen ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_S$ bilden, folgt die Behauptung $\mu = \nu$ wie gewohnt, also mit Satz 1.9. $\square$

Nun können wir Definition 4.15 übernehmen:

Definition 9.9 (i) Es seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. $(\mu_n)_n$ *konvergiert schwach gegen* μ (in Zeichen $\mu_n \xrightarrow{w} \mu$), wenn

$$\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu \quad \forall f \in C(S)$$

gilt.
(ii) Es seien X_n und X $(S,\, \mathcal{B}_S)$-wertige Zufallsvariablen für $n \in \mathbb{N}$, die auf einem W-Raum $(\Omega,\, \mathcal{A},\, P)$ definiert sind. Falls $P^{X_n} \xrightarrow{w} P^X$ gilt, so sagt man, die Folge $(X_n)_n$ *konvergiert in Verteilung* gegen X (und schreibt oft $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$).

Die Wahl dieser Definition wurde in Bemerkung 4.16 diskutiert. Formal ist sie bequem. Oft möchte man jedoch lieber wissen, für welche $A \in \mathcal{B}_S$ gilt: $\mu(A) = \lim_{n\to\infty} \mu_n(A)$.
Es gilt:

Satz 9.10 (Portmanteau) *Es seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. Die folgenden Aussagen sind äquivalent:*

(i) $\mu_n \xrightarrow{w} \mu$.
(ii) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *für jede abgeschlossene Menge* $F \subset S$.
(iii) $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$ *für jede offene Menge* $U \subset S$.
(iv) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *für alle* $A \in \mathcal{B}_S$ *mit* $\mu(\partial A) = 0$.

Beweis: (i) $\Rightarrow$ (ii) Sei F abgeschlossen, $\varepsilon > 0$, und $f_\varepsilon(x) := \max\{0, 1 - d(x, F)/\varepsilon\}$. f_ε ist beschränkt und stetig mit $1_F \le f_\varepsilon$, also

$$\limsup_{n\to\infty} \mu_n(F) \le \lim_{n\to\infty} \int f_\varepsilon \, d\mu_n = \int f_\varepsilon \, d\mu \, .$$

Mit $\varepsilon \downarrow 0$ folgt $f_\varepsilon \downarrow 1_F$. Der Satz von der dominierten Konvergenz liefert

$$\int f_\varepsilon \, d\mu \downarrow \mu(F)$$

für $\varepsilon \downarrow 0$, also $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$.
(ii) $\Leftrightarrow$ (iii) folgt aus der Tatsache, dass die offenen Mengen die Komplemente der abgeschlossenen sind.
((ii) und (iii)) $\Rightarrow$ (iv): Sei $A \in \mathcal{B}_S$ mit $\mu(\partial A) = 0$. Es gilt

$$\limsup_{n\to\infty} \mu_n(A) \le \limsup \mu_n(\bar{A}) \le \mu(\bar{A})$$

und

$$\liminf_{n\to\infty} \mu_n(A) \ge \liminf_{n\to\infty} \mu_n(\mathring{A}) \ge \mu(\mathring{A}) \, .$$

Aus $\mu(\partial A) = 0$ folgt $\mu(\bar{A}) = \mu(\mathring{A}) = \mu(A)$, also folgt (iv).
(iv) $\Rightarrow$ (ii): $F \subset S$ sei abgeschlossen. Für $\delta \ge 0$ sei

$$F^\delta := \big\{x : d(x, F) \le \delta\big\} \, .$$

Dann ist $\partial(F^\delta) \subset \{x : d(x, F) = \delta\}$. Die Mengen $\partial(F^\delta)$ mit $\delta > 0$ sind also paarweise disjunkt. Die Menge

$$\big\{\delta > 0 : \mu\big(\partial(F^\delta)\big) > 0\big\} = \bigcup_{m=1}^{\infty} \Big\{\delta > 0 : \mu\big(\partial(F^\delta)\big) \ge \frac{1}{m}\Big\}$$

ist höchstens abzählbar. Also existiert eine fallende Nullfolge $(\delta_k)_k$ mit $\mu\big(\partial(F^{\delta_k})\big) = 0$ für alle $k \in \mathbb{N}$. Dann gilt

$$\limsup_{n\to\infty} \mu_n(F) \le \limsup_{n\to\infty} \mu_n(F^{\delta_k}) = \mu(F^{\delta_k})$$

für alle k. Wegen $F^{\delta_k} \downarrow F$ haben wir $\mu(F^{\delta_k}) \downarrow \mu(F)$ für $k \to \infty$, also folgt (ii).
(iii) $\Rightarrow$ (i): Sei $f \ge 0$ und stetig. Dann folgt aus Satz 3.8 und dem Lemma von Fatou (2.5 (iv)):

$$\begin{aligned} \int f \, d\mu = \int_0^\infty \mu(f > t) \, dt &\le \int_0^\infty \liminf_{n\to\infty} \mu_n(f > t) \, dt \\ &\le \liminf_{n\to\infty} \int_0^\infty \mu_n(f > t) \, dt = \liminf_{n\to\infty} \int f \, d\mu_n \, . \end{aligned} \tag{9.1}$$

Sei nun f stetig mit $|f| \le c < \infty$. Wende (9.1) auf $c \pm f$ an. Dies liefert $\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu$. $\square$

„Portmanteau“ ist nicht der Name eines Mathematikers, sonder eine englische Vokabel für „Handkoffer“. Es soll zum Ausdruck bringen, dass man den Satz wie einen Koffer mit sich tragen sollte, wenn man in der Welt der schwachen Konvergenz spaziert. Das folgende Lemma gibt eine hinreichende (aber nicht notwendige) Bedingung für schwache Konvergenz:

Lemma 9.11 *Seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. Es sei $\mathcal{U}$ eine durchschnittstabile Teilfamilie von $\mathcal{B}_S$, die die Eigenschaft hat, dass jede offene Teilmenge von S als endliche oder abzählbare Vereinigung von Mengen aus $\mathcal{U}$ dargestellt werden kann. Gilt $\lim_{n\to\infty} \mu_n(U) = \mu(U)$ für alle $U \in \mathcal{U}$, so gilt $\mu_n \xrightarrow{w} \mu$.*

Beweis: Für $m \in \mathbb{N}$ und $A_1, \dots, A_m \in \mathcal{U}$ gilt

$$\begin{aligned}\mu_n\Big(\bigcup_{j=1}^m A_j\Big) &= \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<\dots<j_k} \mu_n(A_{j_1} \cap \dots \cap A_{j_k}) \\ &\underset{n\to\infty}{\longrightarrow} \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<\dots<j_k} \mu(A_{j_1} \cap \dots \cap A_{j_k}) = \mu\Big(\bigcup_{j=1}^m A_j\Big).\end{aligned}$$

Sei $G \subset S$ offen. Dann gilt $G = \bigcup_i A_i$ für eine Folge in $\mathcal{U}$. Dann existieren für jedes $\varepsilon > 0$ ein $m \in \mathbb{N}$ und $A_1, \dots, A_m \in \mathcal{U}$ mit $\mu(G) \leq \mu\big(\bigcup_{j=1}^m A_j\big) + \varepsilon$. Also gilt

$$\liminf_{n\to\infty} \mu_n(G) \geq \liminf_{n\to\infty} \mu_n\Big(\bigcup_{j=1}^m A_j\Big) = \mu\Big(\bigcup_{j=1}^m A_j\Big) \geq \mu(G) - \varepsilon\,.$$

Da $\varepsilon > 0$ beliebig war, folgt die Aussage (iii) von Satz 9.10. □

Wir wollen nun das Verhalten induzierter W-Maße untersuchen, wenn die Ursprungsmaße schwach konvergieren. Ist h eine messbare Abbildung auf $(S, \mathcal{B}_S)$ in einen zweiten metrischen Raum, so braucht aus $\mu_n \xrightarrow{w} \mu$ nicht $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$ zu folgen:

Beispiel 9.12 Sei $(x_n)_n$ eine Folge in $S\backslash\{x\}$, die gegen $x \in S$ konvergiert. Dann gilt $\delta_{x_n} \xrightarrow{w} \delta_x$, siehe 4.16 (ii). Ist $h : S \to \mathbb{R}$ durch $h(y) = 1_{\{x\}}(y)$ definiert, so gelten

$$\delta_{x_n} h^{-1} = \delta_0 \quad \text{und} \quad \delta_x h^{-1} = \delta_1\,,$$

also konvergiert $\delta_{x_n} h^{-1}$ nicht schwach gegen $\delta_x h^{-1}$.

Lemma 9.13 *Seien (S, d) und (S', d') zwei metrische Räume, und $h : S \to S'$ sei stetig. Es seien μ_n und μ W-Maße auf $(S, \mathcal{B}_S)$, $n \in \mathbb{N}$, mit $\mu_n \xrightarrow{w} \mu$. Dann gilt $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$ auf $(S', \mathcal{B}_{S'})$.*

Beweis: Ist $f \in C(S')$, so ist $f \circ h \in C(S)$, also

$$\begin{aligned}\lim_{n\to\infty} \int f\, d(\mu_n\, h^{-1}) &= \lim_{n\to\infty} \int (f \circ h)\, d\mu_n \\ &= \int (f \circ h)\, d\mu = \int f\, d(\mu\, h^{-1})\,. \end{aligned}$$ □

Die Forderung nach Stetigkeit ist stark einschränkend. Wir wollen sie abschwächen und bereiten dies vor:

Lemma 9.14 *Es seien (S, d) und (S', d') zwei metrische Räume, und $h : S \to S'$ sei $\mathcal{B}_S/\mathcal{B}_{S'}$-messbar. Dann ist*

$$D_h := \{x \in S : h \textit{ ist nicht stetig in } x\} \in \mathcal{B}_S\,.$$

Beweis: Für $m, n \in \mathbb{N}$ sei

$$\begin{aligned} A_{m,n} = \Big\{x \in S : &\text{ es gibt } y, z \in S \text{ mit } d(x,\, y) < \frac{1}{m} \\ &\text{und } d(x,\, z) < \frac{1}{m} \text{ sowie } d'\big(h(y),\, h(z)\big) \geq \frac{1}{n}\Big\} \end{aligned}$$

Die Menge $A_{m,n}$ ist offen. Daraus folgt

$$D_h = \bigcup_n \bigcap_m A_{m,n} \in \mathcal{B}_S\,.$$ □

Satz 9.15 *Es liege dieselbe Situation vor wie in Lemma 9.13, h sei jedoch nur als $\mathcal{B}_S/\mathcal{B}_{S'}$-messbar vorausgesetzt. Gilt $\mu(D_h) = 0$, so folgt $\mu_n h^{-1} \xrightarrow{w} \mu_n h^{-1}$.*

Beweis: Sei $F \subset S'$ abgeschlossen. Dann gilt

$$\limsup_{n\to\infty} \mu_n\big(h^{-1}(F)\big) \leq \limsup_{n\to\infty} \mu_n\left(\overline{h^{-1}(F)}\right) \leq \mu\left(\overline{h^{-1}(F)}\right)\,.$$

Es ist $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$. Wegen $\mu(D_h) = 0$ folgt

$$\mu\left(\overline{h^{-1}(F)}\right) = \mu\big(h^{-1}(F)\big)$$

und aus Kriterium (ii) von Satz 9.10 folgt die Behauptung. □

Häufig gibt es in Bezug auf schwache Konvergenz verhältnismäßig große kompakte bzw. relativ kompakte Mengen in $\mathcal{M}_1(S)$.

Definition 9.16 (i) Eine Teilmenge $\Gamma \subset \mathcal{M}_1(S)$ heißt *relativ kompakt*, wenn jede Folge $(\mu_n)_n$ in Γ eine schwach konvergente Teilfolge hat (Der Grenzwert muss nicht in Γ liegen).
(ii) Eine Teilmenge $\Gamma \subset \mathcal{M}_1(S)$ heißt *straff*, wenn für jedes $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subset S$ existiert, so dass $\mu(K_\varepsilon) \geq 1 - \varepsilon$ für jedes $\mu \in \Gamma$.

Bemerkungen 9.17 (i) Ist S kompakt, so ist offenbar $\mathcal{M}_1(S)$ straff.

(ii) $\mathcal{M}_1(\mathbb{R})$ ist nicht straff, weil schon $\{\delta_x : x \in \mathbb{R}\}$ nicht straff ist.

(iii) Ein einzelnes W-Maß $\mu \in \mathcal{M}_1(S)$ heißt straff, wenn $\{\mu\}$ straff ist, das heißt, wenn für jedes $\varepsilon > 0$ eine kompakte Menge K_ε existiert mit $\mu(K_\varepsilon) \geq 1 - \varepsilon$. Ist S σ-kompakt, das heißt, existiert eine Folge $(K_n)_n$ von kompakten Mengen in S mit $K_n \uparrow S$, so ist jedes W-Maß straff, denn $\mu(K_n) \uparrow \mu(S) = 1$. Dies ist für $S = \mathbb{R}$ oder $S = \mathbb{R}^d$ der Fall.

Es gibt eine erstaunlich große Klasse von metrischen Räumen, die nicht unbedingt σ-kompakt sind und in denen jedes W-Maß straff ist, nämlich vollständig separabel[2]. Diese Klasse umfasst separable Hilbert- und Banachräume, wie etwa den Folgenraum ℓ_2 oder den Raum $C[0, 1]$ der stetigen Funktionen $[0, 1] \to \mathbb{R}$, versehen mit der Supremumsmetrik. Unendlich-dimensionale Banachräume sind nie σ-kompakt.

Die Aussage, dass jedes W-Maß auf einem vollständigen, separablen metrischen Raum straff ist, ist ein Spezialfall des Satzes von Prohorov, der im Anhang A bewiesen wird:

Satz 9.18 (Satz von Prohorov, 1956) *Es sei S vollständig und separabel. Dann ist jede Teilmenge von $\mathcal{M}_1(S)$ genau dann relativ kompakt, wenn sie straff ist.*

Wie wollen diesen Satz anwenden. Dazu eine Vorbereitung:

Lemma 9.19 *Seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. Dann gilt $\mu_n \xrightarrow{w} \mu$ genau dann, wenn jede Teilfolge $(\mu'_n)_n$ von $(\mu_n)_n$ ihrerseits eine Teilfolge $(\mu''_n)_n$ hat mit $\mu''_n \xrightarrow{w} \mu$.*

Beweis: Folgt aus Definition 9.9 und der Tatsache, dass reelle Zahlenfolgen sich so verhalten. □

Wir leiten ein sehr nützliches Kriterium für schwache Konvergenz auf $\mathbb{R}^d$ her. Für $x \in \mathbb{R}^d$ sei $\pi_x : \mathbb{R}^d \to \mathbb{R}$ definiert durch $\pi_x(y) := \langle x, y\rangle$, Euklids Skalarprodukt.

Satz 9.20 (Satz von Cramér-Wold, 1936) *Es seien μ_n und μ W-Maße auf $(\mathbb{R}^d, \mathcal{B}^d)$ für $n \in \mathbb{N}$. Dann gilt $\mu_n \xrightarrow{w} \mu$ genau dann, wenn*

$$\mu_n \pi_x^{-1} \xrightarrow{w} \mu \pi_x^{-1} \text{ in } (\mathbb{R}, \mathcal{B})$$

für alle $x \in \mathbb{R}^d$ gilt.

Beweis: Da π_x stetig ist, folgt aus $\mu_n \xrightarrow{w} \mu$ und Lemma 9.13 die Behauptung $\mu_n \pi_x^{-1} \xrightarrow{w} \mu \pi_x^{-1}$. Zum Beweis der Umkehrung betrachten wir zunächst die Projektionen $\pi_i := \pi_{e_i}$, $1 \leq i \leq d$, auf die d Einheitsvektoren $e_i \in \mathbb{R}^d$. $\mu_n \pi_i^{-1}$

[2] *Separabel*: S enthält eine abzählbare, dichte Teilmenge

konvergiert schwach, also ist $\{\mu_n \pi_i^{-1},\, n \in \mathbb{N}\}$ relativ kompakt. Somit existiert für jedes $\varepsilon > 0$ eine kompakte Menge $K_i \subset \mathbb{R}$ mit

$$\mu_n\big(\pi_i^{-1}(K_i)\big) \geq 1 - \frac{\varepsilon}{d}$$

für alle $n \in \mathbb{N}$ und alle $i \in \{1, \ldots, d\}$. Die Menge

$$K := \bigcap_{i=1}^{d} \pi_i^{-1}(K_i) \subset \mathbb{R}^d$$

ist abgeschlossen und beschränkt in $\mathbb{R}^d$, also kompakt. Für alle $n \in \mathbb{N}$ gilt:

$$\mu_n(K^c) = \mu_n\Big(\bigcup_{i=1}^{d} \pi_i^{-1}(K_i)^c\Big) \leq \sum_{i=1}^{d} \mu_n\big(\pi_i^{-1}(K_i^c)\big) \leq \varepsilon\,.$$

Aus Satz 9.18 folgt, dass $\{\mu_n : n \in \mathbb{N}\}$ relativ kompakt ist. Sei $(\mu_n')_n$ eine beliebige Teilfolge von $(\mu_n)_n$. Diese hat eine konvergente Teilfolge $(\mu_n'')_n$ mit $\mu_n'' \xrightarrow{w} \mu''$ für ein $\mu'' \in \mathcal{M}_1(\mathbb{R}^d)$. Für alle $x \in \mathbb{R}^d$ folgt dann

$$\mu_n'' \pi_x^{-1} \xrightarrow{w} \mu'' \pi_x^{-1}\,.$$

Wegen $\mu_n \pi_x^{-1} \xrightarrow{w} \mu \pi_x^{-1}$ folgt $\mu \pi_x^{-1} = \mu'' \pi_x^{-1}$ für alle $x \in \mathbb{R}^d$. Damit stimmen auch die charakteristischen Funktionen von $\mu \pi_x^{-1}$ und $\mu'' \pi_x^{-1}$ überein, insbesondere im Punkt 1. Somit gilt

$$\hat{\mu}(x) = \int e^{i\langle x, y\rangle}\, \mu(dy) = \int e^{it}\, (\mu \pi_x^{-1})(dt) = \int e^{it} (\mu'' \pi_x^{-1})(dt) = \widehat{\mu''}(x)\,.$$

Aus Satz 9.2 folgt $\mu = \mu''$ und Lemma 9.19 führt zur Behauptung. □

Mit Satz 6.19 erhalten wir recht leicht eine mehrdimensionale Version des zentralen Grenzwertsatzes, $|\cdot|$ bezeichne die Euklidische Norm im $\mathbb{R}^d$.

Satz 9.21 *Es sei $(X_n)_n$ eine Folge unabhängiger, identisch verteilter d-dimensionaler Zufallsvektoren. Es gelte $\mathbb{E}|X_i|^2 < \infty$. Seien $a := \mathbb{E}X_1$ und Σ die Kovarianzmatrix der X_i. Dann gilt*

$$P^{\sum_{i=1}^n (X_i - a)/\sqrt{n}} \xrightarrow{w} \mu\,,$$

wobei μ die d-dimensionale Normalverteilung mit Mittelwert 0 *und Kovarianzmatrix Σ ist.*

Beweis: Sei $T_n := \sum_{i=1}^n (X_i - a)/\sqrt{n}$. Nach Satz 9.20 genügt es zu zeigen, dass für jedes $x \in \mathbb{R}^d$ gilt

$$P^{\langle x, T_n\rangle} \xrightarrow{w} \mu \pi_x^{-1}\,.$$

Es ist

$$\langle x, T_n\rangle = \sum_{i=1}^{n} \big(\langle x, X_i\rangle - \langle x, a\rangle\big)/\sqrt{n}\,.$$

Die $\langle x, X_i\rangle$, $i \in \mathbb{N}$, sind unabhängige, identisch verteilte, eindimensionale Zufallsgrößen mit Erwartungswert $\langle x, a\rangle$ und Varianz $\sigma_x^2 = \mathbb{E}(\langle x, X_i - a\rangle^2) = x^t\Sigma x$, wenn x als Spaltenvektor geschrieben wird.

Ist $\sigma_x^2 > 0$, so konvergiert $P^{\langle x, T_n/\sigma_x\rangle}$ nach Korollar 8.7 gegen die Standardnormalverteilung, also konvergiert $P^{\langle x, T_n\rangle}$ nach Lemma 9.13 mit $h : \mathbb{R} \to \mathbb{R}$, $h(y) = \sigma_x y$, gegen die Normalverteilung mit Mittelwert 0 und Varianz σ_x^2.

Gilt $\sigma_x^2 = 0$, so ist $\langle x, T_n\rangle = 0$ fast sicher nach Satz 2.7, und somit gilt $P^{\langle x, T_n\rangle} \xrightarrow{w} \delta_0$ (Normalverteilung mit Erwartungswert 0 und Varianz 0).

Nun ist aber $\mu\pi_x^{-1}$ die Normalverteilung mit Erwartungswert 0 und Varianz σ_x^2. Damit ist der Satz bewiesen. □

Abschließend soll definiert werden, was es heißt, dass eine Folge $(X_n)_n$ von $(S, \mathcal{B}_S)$-wertigen Zufallsvariablen „in Wahrscheinlichkeit" gegen eine Zufallsvariable X konvergiert. Es liegt nahe, $P\big(d(X_n, X) \geq \varepsilon\big)$ zu verwenden. Aber $d(X_n, X)$ ist nicht immer eine Zufallsgröße! Man muss voraussetzen, dass S separabel ist. Ist (S, d) ein metrischer Raum, so betrachten wir $(S \times S, d')$ mit

$$d'\big((x_1, x_2), (y_1, y_2)\big) := \big(d(x_1, y_1)^2 + d(x_2, y_2)^2\big)^{1/2}.$$

Lemma 9.22 *Ist S separabel, so ist $\mathcal{B}_{S\times S} = \mathcal{B}_S \otimes \mathcal{B}_S$.*

Beweis: Jede Produktmenge $A \times B$ mit $A \subset S$ und $B \subset S$ offen, ist offen in $S \times S$, liegt also in $\mathcal{B}_{S\times S}$. Da diese Mengen $\mathcal{B}_S \otimes \mathcal{B}_S$ erzeugen, folgt $\mathcal{B}_S \otimes \mathcal{B}_S \subset \mathcal{B}_{S\times S}$. Hier wurde die Separabilität nicht benutzt.

Ist S separabel, so existiert eine abzählbare Basis $\{U_i,\, i \in \mathbb{N}\}$ der Topologie von S und $\{U_i \times U_j,\, i, j \in \mathbb{N}\}$ ist dann eine abzählbare Basis der Topologie von $S \times S$. Somit ist jede offene Teilmenge von $S \times S$ in $\mathcal{B}_S \otimes \mathcal{B}_S$ enthalten, also ist $\mathcal{B}_{S\times S} \subset \mathcal{B}_S \otimes \mathcal{B}_S$. □

Sind nun X und Y zwei $(S, \mathcal{B}_S)$-wertige Zufallsvariablen, definiert auf einem gemeinsamen W-Raum $(\Omega, \mathcal{A}, P)$, so ist (X, Y) eine $(S \times S, \mathcal{B}_S \otimes \mathcal{B}_S)$-wertige Zufallsvariable. $d : S \times S \to \mathbb{R}$ ist eine d'-stetige Abbildung, also $\mathcal{B}_{S\times S}/\mathcal{B}$-messbar, also ist $d(X, Y)$ mit Hilfe von Lemma 9.22 eine Zufallsgröße:

Definition 9.23 Es sei S separabel und X sowie X_n für $n \in \mathbb{N}$ Zufallsvariablen mit Werten in $(S, \mathcal{B}_S)$. Die Folge $(X_n)_n$ *konvergiert in Wahrscheinlichkeit* gegen X, falls $d(X_n, X)$ in Wahrscheinlichkeit gegen 0 konvergiert, also für alle $\varepsilon > 0$

$$\lim_{n\to\infty} P\big(d(X_n, X) \geq \varepsilon\big) = 0$$

gilt.

Satz 9.24 *Konvergiert $(X_n)_n$ in W-keit gegen X, so gilt $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$.* [3]

[3] Vgl. Satz 4.17

Beweis: Sei $A \in \mathcal{B}_S$ mit $P(X \in \partial A) = 0$. Für $\varepsilon > 0$ gilt

$$P(X_n \in A,\, X \notin A) \leq P\big(d(X_n,\, X) \geq \varepsilon\big) + P\big(d(X,\, A) < \varepsilon,\, X \notin A\big)\,.$$

Dies und dasselbe mit A^c ergibt auf Grund $\lim_{n\to\infty} P\big(d(X_n,\, X) \geq \varepsilon\big) = 0$ die Abschätzung

$$\begin{aligned}\limsup_{n\to\infty} P(\{X_n \in A\}\Delta\{X \in A\}) &\leq P\big(d(X,\, A) < \varepsilon,\, X \notin A\big) \\ &\quad + P\big(d(X,\, A^c) < \varepsilon,\, X \in A\big)\,.\end{aligned}$$

Dabei sei an die Definition $A\Delta B := (A \setminus B) \cup (B \setminus A)$ erinnert. Für $\varepsilon \downarrow 0$ gilt

$$\{d(X,\, A) < \varepsilon,\, X \notin A\} \downarrow \{X \in \partial A \cap A^c\}$$

und

$$\{d(X,\, A^c) < \varepsilon,\, X \in A\} \downarrow \{X \in A \cap \partial A\}\,.$$

Wegen $P(X \in \partial A) = 0$ folgt

$$\lim_{n\to\infty} P(\{X_n \in A\}\Delta\{X \in A\}) = 0\,,$$

also

$$\lim_{n\to\infty} P(X_n \in A) = P(X \in A)\,.$$

Nach Satz 9.10 (iv) folgt $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$. □

Lemma 9.25 *Es sei S separabel, und $(X_n)_n$ und $(X'_n)_n$ seien zwei Folgen von $(S,\, \mathcal{B}_S)$-wertigen Zufallsvariablen. Gelten $\mathcal{L}(X_n) \xrightarrow{w} \mu$ und $d(X_n,\, X'_n) \to 0$ in Wahrscheinlichkeit, so gilt $\mathcal{L}(X'_n) \xrightarrow{w} \mu$.*

Beweis: Seien $F \subset S$ abgeschlossen, $\varepsilon > 0$ und $F^\varepsilon = \{x : d(x,\, F) \leq \varepsilon\}$. Dann gilt

$$\begin{aligned}\limsup_{n\to\infty} P(X'_n \in F) &\leq \limsup_{n\to\infty} P\big(d(X_n,\, X'_n) \geq \varepsilon\big) \\ &\quad + \limsup_{n\to\infty} P(X_n \in F^\varepsilon) \leq \mu(F^\varepsilon)\,.\end{aligned}$$

Für $\varepsilon \to 0$ gilt $\mu(F^\varepsilon) \to \mu(F)$. □

KAPITEL 10

Der Satz von Donsker

Fortan betrachten wir $S = C[0,1]$ die Menge der stetigen Funktionen $[0,1] \to \mathbb{R}$. Für $f, g \in [0,1]$ sei $d(f,g) := \sup_{t\in[0,1]} |f(t) - g(t)|$.

Satz 10.1 *$(C[0,1], d)$ ist ein vollständiger und separabler metrischer Raum.*

Beweis: Die Banachraum-Eigenschaft kennen wir aus der Analysis-Vorlesung. Die Polynome mit rationalen Koeffizienten bilden eine abzählbare, dichte Teilmenge in $C[0,1]$ nach dem Approximationssatz von Weierstraß (der dies für jede kompakte Teilmenge in $\mathbb{R}^k$ liefern würde). $\square$

Für $m \in \mathbb{N}$ und $0 \le t_1 < t_2 < \cdots < t_m \le 1$ sei $\pi_{t_1,\ldots,t_m} : C[0,1] \to \mathbb{R}^m$ die *endlich dimensionale Projektion* $f \mapsto (f(t_1), \ldots, f(t_m))$.

Lemma 10.2 *Es gilt*

$$\mathcal{B}_C = \sigma(\pi_t^{-1}(\mathcal{B}), t \in [0,1])$$

mit der Notation $C := C[0,1]$.

Beweis: Mit $\mathcal{B}' := \sigma(\pi_t^{-1}(\mathcal{B}), t \in [0,1])$ wollen wir $\mathcal{B}_C = \mathcal{B}'$ zeigen. Da π_t stetig ist, ist für $U \subset \mathbb{R}$ offen auch $\pi_t^{-1}(U)$ offen, liegt also in $\mathcal{B}_C$. Daraus folgt $\mathcal{B}' \subset \mathcal{B}_C$. Für $f \in C[0,1]$ und $\varepsilon > 0$ sei

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

Dann ist, da f stetig,

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

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

Der Satz von Donsker ist eine Verallgemeinerung des zentralen Grenzwertsatzes, indem nicht nur die Asymptotik der Verteilung von $S_n/\sqrt{n}$ ($S_n = \sum_{i=1}^n X_i$, unabhängig, identisch verteilte X_i) untersucht wird, sondern die Verteilung des „gesamten Pfades".

Es sei $(X_n)_n$ eine Folge unabhängiger, identisch verteilter, $\mathbb{R}$-wertiger Zufallsvariablen, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Es sei $\mathbb{E}X_i = 0$ (falls nicht, ersetzen wir X_i durch $X_i - \mathbb{E}X_i$) und $\sigma^2 := \mathbb{E}X_i^2 \in (0, \infty)$. Wir

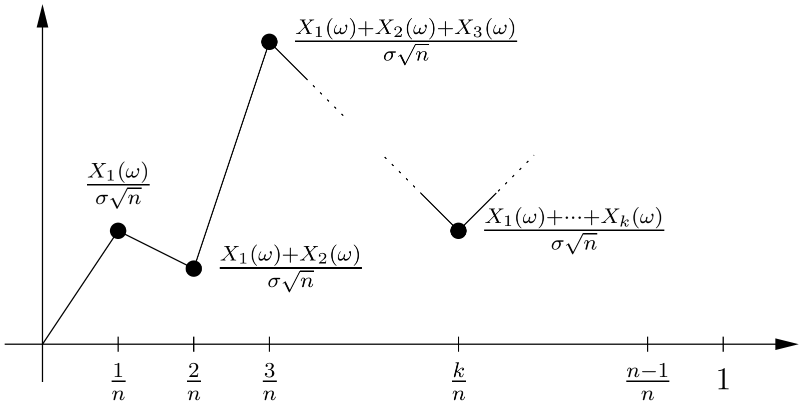


ABBILDUNG 10.1. Broken-Line Prozess.

setzen $S_0 = 0$ und $S_n = \sum_{i=1}^{n} X_i$, $n \in \mathbb{N}$. Für $n \in \mathbb{N}$ und $\omega \in \Omega$ definieren wir die Abbildung $Y_n(\omega, \cdot) : [0,1] \to \mathbb{R}$ durch

$$Y_n\big(\omega, \frac{k}{n}\big) := \frac{S_k(\omega)}{\sigma\sqrt{n}}\ ,\ k = 0, 1, \ldots, n$$

und lineare Interpolation („Broken-Line-Prozeß“).
(Bild einer Irrfahrt: Teilchen schneller springen lassen und die Sprünge immer kleiner werden lassen, „Zeit- und Ortsskala simultan ändern“.)
Eine äquivalente Defintion ist

$$Y_n(\omega, t) = \frac{S_{[nt]}(\omega)}{\sigma\sqrt{n}} + \frac{nt - [nt]}{\sigma\sqrt{n}} X_{[nt]+1}(\omega)\ , 0 \le t \le 1\ . \tag{10.1}$$

($[x]$ sei der ganzzahlige Anteil der reellen Zahl x).
Y_n kann als Abbildung von Ω nach $C[0,1]$ aufgefasst werden. Für ein festes $t \in [0,1]$ ist $Y_n(\cdot, t)$ offenbar $\mathcal{A}$-$\mathcal{B}$-messbar. Nach Lemma 10.2 folgt, dass $Y_n : \Omega \to C[0,1]$ eine $(C, \mathcal{B}_C)$-wertige Zufallsvariable ist. Ohne Einschränkung ist $\sigma^2 = 1$, sonst ersetze X_i durch X_i/σ. Der Satz von DONSKER wird liefern, dass $\mathcal{L}(Y_n)$ schwach gegen ein W-Maß auf $(C, \mathcal{B}_C)$ konvergiert. Da $\pi_{t_1,\ldots,t_m}$ stetig ist, ist nach Lemma 9.13 für die Konvergenz von

$$\mu_n := P^{Y_n}$$

notwendig, dass $\mu_n \pi_{t_1,\ldots,t_m}^{-1}$ für $n \to \infty$ auf $(\mathbb{R}^m, \mathcal{B}^m)$ konvergiert.

Satz 10.3 *Für jedes $m \in \mathbb{N}$ und $0 \le t_1 < t_2 < \cdots < t_m \le 1$ konvergiert $\mu_n \pi_{t_1,\ldots,t_m}^{-1}$ schwach auf $(\mathbb{R}^m, \mathcal{B}^m)$ gegen die m-dimensionale Normalverteilung mit Erwartungswert 0 und Kovarianzmatrix $(\min\{t_i, t_j\})_{i,j}$.*

$m = 1$, $t_1 = 1$: $\mu_n \pi_1^{-1} = \mathcal{L}(Y_n(1)) = \mathcal{L}(\frac{S_n}{\sqrt{n}})$ konvergiert gegen die Standard-Normalverteilung. Für $m = 1$ und $t_1 = 0$ ist $\mu_n \pi_0^{-1} = \mathcal{L}(Y_n(0)) = \delta_0$.
Wir müssen noch etwas vorbereiten:

Lemma 10.4 *Sei $d \in \mathbb{N}$ und für $j = 1, \dots, d$ sei $(\mu_n^{(j)})_n$ eine Folge von W-Maßen auf $(\mathbb{R}, \mathcal{B})$ mit $\mu_n^{(j)} \xrightarrow{w} \mu^{(j)} \in \mathcal{M}_1(\mathbb{R})$. Dann gilt*

$$\mu_n^{(1)} \otimes \cdots \otimes \mu_n^{(d)} \xrightarrow{w} \mu^{(1)} \otimes \cdots \otimes \mu^{(d)}$$

auf $(\mathbb{R}^d, \mathcal{B}^d)$.

Beweis: Es sei $A_j := \{x \in \mathbb{R} : \mu^{(j)}(\{x\}) = 0\}$. A_j^c ist abzählbar und somit ist A_j dicht. Sei $B_j \subset A_j$ eine abzählbare dichte Teilmenge von A_j. Dann ist $\{(a_j, b_j) : a_j, b_j \in B_j\}$ eine abzählbare Basis der Topologie von $\mathbb{R}$, also ist

$$\mathcal{U} := \{(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d) : a_j, b_j \in B_j \text{ für } j = 1, \dots, d\}$$

eine Basis der Topologie von $\mathbb{R}^d$. $\mathcal{U}$ ist durchschnittstabil und für $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d) \in \mathcal{U}$ gilt wegen Satz 9.10 (iv).

$$\mu_n^{(1)} \otimes \cdots \otimes \mu_n^{(d)}((a_1, b_1) \times \cdots \times (a_d, b_d)) = \prod_{j=1}^d \mu_n^{(j)}((a_j, b_j))$$
$$\xrightarrow{n\to\infty} \prod_{j=1}^d \mu^{(j)}((a_j, b_j)) = \mu^{(1)} \otimes \cdots \otimes \mu^{(d)}((a_1, b_1) \times \cdots \times (a_d, b_d)) \ .$$

Das Lemma folgt nun aus Lemma 9.11. □

Beweis: (von Satz 10.3) Wir können annehmen, dass $t_1 > 0$ gilt. Setze $\sum_{i=1}^0 := 0$ und

$$Z_1^{(n)} := \sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}}, Z_2^{(n)} := \sum_{i=[nt_1]+1}^{[nt_2]} \frac{X_i}{\sqrt{n}}, \dots, Z_m^{(n)} := \sum_{i=[nt_{m-1}]+1}^{[nt_m]} \frac{X_i}{\sqrt{n}} \ .$$

$Z_1^{(n)}, \dots, Z_m^{(n)}$ sind für jedes $n \in \mathbb{N}$ unabhängig. Mit Lemma 10.4 untersuchen wir das Konvergenzverhalten von $(Z_j^{(n)})_n$ für festes j:

$\mathcal{L}(Z_j^{(n)}) = \mathcal{L}(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}})$, wo wir $t_0 := 0$ und $k(n) := [nt_j] - [nt_{j-1}]$ setzen. Der zentrale Grenzwertsatz liefert:

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le s\Big) = \Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^s e^{-x^2/2}\, dx \ .$$

Nun gilt $\lim_{n\to\infty}\frac{k(n)}{n} = t_j - t_{j-1}$. Für $\varepsilon > 0$ und $s \in \mathbb{R}$ folgt

$$\begin{aligned}
\limsup_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &\le \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le \frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big) \\
&= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big) \quad \text{und} \\
\liminf_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &\ge \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{k(n)}} \le \frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big) \\
&= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big)\ , \\
\text{also } \lim_{n\to\infty} P\Big(\sum_{i=1}^{k(n)} \frac{X_i}{\sqrt{n}} \le s\Big) &= \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}}\Big)
\end{aligned}$$

Dies ist die Verteilungsfunktion der eindimensionalen Normalverteilung mit Erwartungswert 0 und Varianz $t_j - t_{j-1}$. Nach Lemma 10.4 folgt, dass $\mathcal{L}(Z_1^{(n)}, \ldots, Z_m^{(n)})$ für $n \to \infty$ gegen die Produktverteilung konvergiert, und dies ist die m-dimensionale Normalverteilung ν mit Erwartungswert 0 und Kovarianzmatrix $(\delta_{ij}(t_j - t_{j-1}))_{i,j}$.

Sei nun $f : \mathbb{R}^m \to \mathbb{R}^m$ durch $f(x_1, \ldots, x_m) := (x_1, x_1+x_2, x_1+x_2+x_3, \ldots, x_1 + \cdots + x_m)$ definiert. Nach Lemma 9.13 konvergiert die Verteilung von

$$f(Z_1^{(n)}, \ldots, Z_m^{(n)}) = (\sum_{i=1}^{[nt_1]} \frac{X_i}{\sqrt{n}}, \sum_{i=1}^{[nt_2]} \frac{X_i}{\sqrt{n}}, \ldots, \sum_{i=1}^{[nt_m]} \frac{X_i}{\sqrt{n}})$$

gegen νf^{-1}. Sei $(U_1, \ldots, U_m)$ eine Zufallsgröße mit Verteilung ν, dann besitzt die Normalverteilung νf^{-1} den Erwartungswert 0 und die Kovarianzmatrix mit Komponenten

$$\begin{aligned}
\mathbb{E}(\sum_{k=1}^{i} U_k \sum_{s=1}^{j} U_s) &= \sum_{k=1}^{\min\{i,j\}} \mathbb{E}(U_k^2) + \sum_{k=1, k\neq s}^{i} \sum_{s=1}^{j} \mathbb{E}(U_k U_s) = \sum_{k=1}^{\min\{i,j\}} (t_k - t_{k-1}) \\
&= \min\{t_i, t_j\}\ .
\end{aligned}$$

Sei nun $W_j^{(n)} := \sum_{i=1}^{[nt_j]} \frac{X_i}{\sqrt{n}} - Y_n(t_j)$. Dann gilt $|W_j^{(n)}| \le \frac{|X_{[nt_j]+1}|}{\sqrt{n}}$, falls $t_j < 1$ und $W_j^{(n)} = 0$ sonst. Damit ist für $\varepsilon > 0$

$$\begin{aligned}
&P(|(W_1^{(n)}, \ldots, W_m^{(n)})| \ge \varepsilon) \le P\Big(\bigcup_{j=1}^{m} \{|W_j^{(n)}| \ge \varepsilon/m\}\Big) \\
&\le \sum_{j=1}^{m} P(|X_{[nt_j]+1}| \ge \sqrt{n}\varepsilon/m\Big) = mP(|X_1| \ge \sqrt{n}\varepsilon/m) \longrightarrow 0
\end{aligned}$$

für $n \to \infty$, also konvergiert $(W_1^{(n)}, \dots, W_m^{(n)})$ in Wahrscheinlichkeit gegen 0. Nach Lemma 9.25 konvergiert dann auch $\mathcal{L}(Y_n(t_1), \dots, Y_n(t_m))$ gegen νf^{-1}. $\square$

Konvergiert, wie behauptet wird, $\mu_n = \mathcal{L}(Y_n)$ gegen ein Wahrscheinlichkeitsmaß μ auf $(C, \mathcal{B}_C)$, so konvergiert für alle $m \in \mathbb{N}$ und $0 \leq t_1 < t_2 < \cdots < t_m \leq 1$ auch $\mu_n \pi_{t_1,\dots,t_m}^{-1}$ gegen $\mu \pi_{t_1,\dots,t_m}^{-1}$. Dieses Wahrscheinlichkeitsmaß muß dann nach Satz 10.3 das dort angegebene Wahrscheinlichkeitsmaß auf $(\mathbb{R}^m, \mathcal{B}^m)$ sein. Gibt es ein solches Maß μ?

Satz 10.5 *Es gibt genau ein Wahrscheinlichkeitsmaß μ auf $(C[0,1], \mathcal{B}_C)$ derart, dass für alle $m \in \mathbb{N}$ und $0 \leq t_1 < t_2 < \cdots < t_m \leq 1$ das Maß $\mu \pi_{t_1,\dots,t_m}^{-1}$ die m-dimensionale Normalverteilung mit Erwartungswert 0 und Kovarianzmatrix $(\min\{t_i, t_j\})_{i,j}$ ist.*

Definition 10.6 Das Maß μ aus Satz 10.5 ist das sogenannte Wiener-*Maß* oder die Verteilung der eindimensionalen Brown*schen Bewegung.*

Die Eindeutigkeit in Satz 10.5 folgt aus der Tatsache, dass die Mengen

$$\{\pi_{t_1,\dots,t_m}^{-1}(A), m \in \mathbb{N}, 0 \leq t_1 < t_2 < \cdots < t_m \leq 1, A \in \mathcal{B}^m\}$$

ein durchschnittsstabiles Erzeugendensystem von $\mathcal{B}_C$ bilden. (vgl. Beweis von Satz 9.2). Die Existenz des Wiener-Maßes μ wird nun simultan mit dem folgenden Satz bewiesen:

Satz 10.7 (Satz von Donsker, 1951) *Es gilt $\mu_n \xrightarrow{w} \mu$ auf $(C, \mathcal{B}_C)$.*

Die Existenz in Satz 10.7 und die in Satz 10.5 folgen aus der folgenden Aussage:

Satz 10.8 *Die Folge $\{\mu_n, n \in \mathbb{N}\}$ ist straff.*

Aus Satz 10.8 folgt mit Satz 9.18 (Satz von Prohorov), dass $(\mu_n)_n$ konvergente Teilfolgen enthält. Jedes Grenzelement μ einer derartigen Teilfolge hat aber nach Satz 10.3 die richtigen endlichdimensionalen Randverteilungen $\mu \pi_{t_1,\dots,t_m}^{-1}$. Damit ist die Existenz in Satz 10.5 gezeigt.

Aus Satz 10.8 folgt weiter, dass jede Teilfolge von $(\mu_n)_n$ wieder eine konvergente Teilfolge hat. Deren Grenzelement stimmt mit Satz 10.3 mit dem Wiener-Maß überein. Aus Lemma 9.19 folgt dann $\mu_n \xrightarrow{w} \mu$. Damit ist der Satz von Donsker bewiesen.

Der Beweis von Satz 10.8 wird noch etwas verschoben. Zunächst folgt aus 10.7

Satz 10.9 *Ist $h : C[0,1] \to \mathbb{R}$ eine Borel-messbare Abbildung mit $\mu(D_h) = 0$* [1] *und ist $(X_i)_i$ eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit*

[1] siehe Notation Lemma 9.14

$\mathbb{E}X_i = 0$ *und* $\mathbb{E}X_i^2 = 1$, *so gilt* $\mathcal{L}(h(Y_n)) \xrightarrow{w} \mu h^{-1}$, *wobei* Y_n *die durch* (10.1) *definierte* $(C, \mathcal{B}_C)$*-wertige Zufallsvariable sei.*

Beweis: Der Satz folgt sofort aus Satz 9.15.

Der Satz liefert die asymptotische Verteilung von $h(Y_n)$, wenn man μh^{-1} kennt. Der Grenzwert hängt aber gar nicht von der speziellen Gestalt der Verteilung der X_i ab. Daher kann der Satz auch zur Berechnung von μh^{-1} dienen, wenn man die Verteilung der $h(Y_n)$ kennt. Man kann dazu die Verteilung der X_i beliebig wählen, solange $\mathbb{E}X_i = 0$ und $\mathbb{E}X_i^2 = 1$ erfüllt sind. Meist ist die Berechnung von $\mathcal{L}(h(Y_n))$ am einfachsten, wenn $P(X_i = \pm 1) = 1/2$ ist. Die für diesen Spezialfall gewonnene Grenzverteilung gilt dann für jede Verteilung der X_i. Man nennt dies das *Invarianzprinzip* von DONSKER. Wir betrachten im folgenden Kapitel Anwendungen dazu!

Sind $(X_i)_i$ und $(X_i')_i$ zwei Folgen unabhängig, identisch verteilter Zufallsgrößen mit $\mathbb{E}X_i = \mathbb{E}X_i' = 0$ und $\mathbb{E}X_i^2 = \mathbb{E}(X_i')^2 = 1$, und sind Y_n und Y_n' die dazugehörigen interpolierten Irrfahrten, so gilt für jede messbare Funktion $h : C[0,1] \to \mathbb{R}$ mit $\mu(D_h) = 0$

$$\lim_{n\to\infty} \mathcal{L}(h(Y_n)) = \lim_{n\to\infty} \mathcal{L}(h(Y_n')) \, . \tag{10.2}$$

Aus Satz 10.3 wissen wir, dass für $0 \le t_1 < \cdots < t_m \le 1$ gilt

$$\lim_{n\to\infty} \mathcal{L}(Y_n(t_1), \ldots, Y_n(t_m)) = \lim_{n\to\infty} \mathcal{L}(Y_n'(t_1), \ldots, Y_n'(t_m)),$$

und somit für jede stetige Funktion $h : \mathbb{R}^m \to \mathbb{R}$

$$\lim_{n\to\infty} \mathcal{L}\big(h(Y_n(t_1), \ldots, Y_n(t_m))\big) = \lim_{n\to\infty} \mathcal{L}\big(h(Y_n'(t_1), \ldots, Y_n'(t_m))\big). \tag{10.3}$$

Wir betrachten die spezielle Abbildung $h : C[0,1] \to \mathbb{R}$ mit $h(f) := \max_{0\le t\le 1} f(t)$, bzw. $h : \mathbb{R}^m \to \mathbb{R}$ mit $h(x_1, \ldots, x_m) := \max_{i=1,\ldots,m} x_i$. Wir wollen aus (10.3) nun (10.2) herleiten. Dies ist eine Limesvertauschung von $m \to \infty$ und $n \to \infty$. Genauer: Sei $(t_0^{(m)}, \ldots, t_m^{(m)})_m$ eine Folge von Einteilungen $0 = t_0^{(m)} \le t_1^{(m)} < \cdots < t_m^{(m)} \le 1$ des Einheitsintervalls, wobei wir einfach annehmen, dass die $m+1$-te Einteilung durch Hinzunahme eines Punktes aus der m-ten entsteht. Es gelte $\max_{1\le i\le m}(t_i^{(m)} - t_{i-1}^{(m)}) \to 0$ für $m \to \infty$. Es gilt für jedes $n \in \mathbb{N}$

$$\lim_{m\to\infty} \max_{1\le i\le m} Y_n(t_i^{(m)}) = \max_{0\le t\le 1} Y_n(t),$$

also

$$\mathcal{L}\Big(\max_{0\le t\le 1} Y_n(t)\Big) = \lim_{m\to\infty} \mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big) \, .$$

Somit folgt (10.2) aus (10.3), sofern man

$$\lim_{n\to\infty} \lim_{m\to\infty} \mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big) = \lim_{m\to\infty} \lim_{n\to\infty} \mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big) \tag{10.4}$$

zeigen kann (falls Limites existieren).

Erinnerung an Doppelfolgen $(a_{nm})_{n,m\in\mathbb{N}}$: Falls $b_n := \lim_{m\to\infty} a_{nm}$ und $c_m := \lim_{n\to\infty} a_{nm}$ sowie $b := \lim_{n\to\infty} b_n$ existieren, wann gilt $b = \lim_{m\to\infty} c_m$? Hinreichend dafür ist die Konvergenz von $a_{nm} \overset{m\to\infty}{\longrightarrow} b_n$ gleichmäßig in n, bzw. $(a_{nm})_m$ ist eine in n gleichmäßige Cauchyfolge: $\sup_n \lim_{m,m'\to\infty} |a_{nm} - a_{nm'}| = 0$.

Zeige also: $\mathcal{L}\Big(\max_{1\le i\le m} Y_n(t_i^{(m)})\Big)$ liegt für große m, m' nahe bei $\mathcal{L}\Big(\max_{1\le i\le m'} Y_n(t_i^{(m')})\Big)$, gleichmäßig in n.

Für $f \in C[0,1]$ und $\delta > 0$ sei

$$\omega_\delta(f) := \sup\{|f(s) - f(t)| : s,t \in [0,1] \text{ mit } |s-t| \le \delta\}$$

Es gilt

$$|\max_{1\le i\le m} Y_n(t_i^{(m)}) - \max_{1\le i\le m'} Y_n(t_i^{(m')})| \le \omega_\delta(Y_n)$$

für $m' \ge m$, falls m so groß ist, dass $\max_i(t_i^{(m)} - t_{i-1}^{(m)}) \le \delta$. Nun gilt (Übung): Wenn

$$\sup_n P\big(\omega_\delta(Y_n) \ge \varepsilon\big) \overset{\delta\to 0}{\longrightarrow} 0 \text{ für alle } \varepsilon > 0\ , \tag{10.5}$$

so folgt (10.4).

Es ist erstaunlich, dass der sehr viel allgemeinere Satz von Donsker sich ebenfalls aus (10.5) ergibt, diese Aussage also die wirklich entscheidende für den Beweis von Satz 10.8 sein wird: Es geht um eine Charakterisierung relativ kompakter Teilmengen in $C[0,1]$.

Das Stetigkeitsmodul $\omega_\delta(f)$ sei wie oben definiert. Es gilt $|\omega_\delta(f) - \omega_\delta(g)| \le 2d(f,g)$, also ist für $\delta > 0$ fest ω_δ stetig. Da ein $f \in C[0,1]$ gleichmäßig stetig ist, gilt $\lim_{\delta\to 0} \omega_\delta(f) = 0$.

Satz 10.10 (Satz von Arzelà-Ascoli) *Eine Teilmenge $A \subset C[0,1]$ hat genau dann kompakten Abschluss, wenn*

(i) $\sup\{|f(0)|, f \in A\} < \infty$ *ist und*
(ii) $\lim_{\delta\to 0} \sup_{f\in A} \omega_\delta(f) = 0$.

gelten.
(Wir geben im Anhang einen Beweis.)

Dieser Satz kann in ein Kriterium für die Straffheit einer Folge von W-Maßen auf C übersetzt werden:

Satz 10.11 *Eine Folge $(\nu_n)_n$ von W-Maßen auf $(C, \mathcal{B}_C)$ ist genau dann straff, wenn*

$$\lim_{a\nearrow\infty} \sup_n \nu_n(\{f : |f(0)| > a\}) = 0 \text{ und} \tag{10.6}$$

$$\lim_{\delta\searrow 0} \limsup_{n\to\infty} \nu_n(\{f : \omega_\delta(f) \ge \varepsilon\}) = 0 \text{ für alle } \varepsilon > 0 \tag{10.7}$$

gelten.

Nach obiger Bemerkung ist $\{f : \omega_\delta(f) \geq \varepsilon\} \in \mathcal{B}_C$. Die Bedingungen (10.6) und (10.7) in Satz 10.11 können wie folgt übersetzt werden:

$$\forall \eta > 0\ \exists a > 0\ \forall n \in \mathbb{N} : \quad \nu_n(\{f : |f(0)| > a\}) \leq \eta\ , \tag{10.8}$$
$$\forall \varepsilon > 0, \eta > 0\ \exists \delta > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0 :$$
$$\nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta. \tag{10.9}$$

Bemerkung 10.12 $C[0,1]$ 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 (10.9) äquivalent zu

$$\forall \varepsilon > 0, \eta > 0\ \exists \delta > 0\ \forall n \in \mathbb{N} : \quad \nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) \leq \eta\ . \tag{10.10}$$

Beweis: (von Satz 10.11) Sei $\{\nu_n, n \in \mathbb{N}\}$ straff. Für $\eta > 0$ sei K eine kompakte Menge mit $\nu_n(K) \geq 1-\eta$ für alle n. Daraus folgen mit dem Satz von Arzelà-Ascoli die Aussagen (10.8) und (10.10), 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 (10.8) und (10.10) erfüllt. Sei $\eta > 0$ vorgegeben. Nach (10.8) existiert ein $a \in \mathbb{R}$, so dass $A := \{f : |f(0)| \leq a\}$ erfüllt: $\nu_n(A) \geq 1-\eta/2$ für alle $n \in \mathbb{N}$. Für $k \in \mathbb{N}$ sei 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 Arzelà-Ascoli hat

$$K := A \cap \bigcap_{k=1}^{\infty} \{f : \omega_{\delta_k}(f) < 1/k\}$$

kompakten Abschluss und es gilt

$$\nu_n(\bar{K}^c) \leq \nu(K^c) \leq \eta/2 + \sum_{k=1}^{\infty} \eta/2^{k+1} = \eta$$

für alle $n \in \mathbb{N}$, was zu zeigen war. □

Bemerkung 10.13 Hinreichend für (10.8) ist $\nu_n(\{f : f(0) = 0\}) = 1$, was für die μ_n im Satz von Donsker erfüllt ist.

Lemma 10.14 *Hinreichend für* (10.9) *ist:*

$$\forall \varepsilon, \eta > 0\ \exists \delta \in (0,1),\ \exists n_0 \in \mathbb{N} : \quad \forall n \geq n_0,\ \forall t \in [0, 1-\delta] :$$
$$\frac{1}{\delta} \nu_n\Big(\Big\{f : \sup_{t \leq s \leq t+\delta} |f(s) - f(t)| \geq \varepsilon\Big\}\Big) \leq \eta\ . \tag{10.11}$$

Beweis: Seien $\varepsilon, \eta > 0$. Zu $\varepsilon/2$ und $\eta/3$ wählen wir $\delta_0 \in (0,1)$ und $n_0 \in \mathbb{N}$ wie in (10.11). $m \in \mathbb{N}$ sei die kleinste natürliche Zahl mit $1/m < \delta_0$. Setze $\delta := \frac{1}{2m}$. Ist $f \in C[0,1]$ mit $\omega_\delta(f) \geq \varepsilon$, so existieren $t < s$ mit $|f(t) - f(s)| \geq \varepsilon$ und $|t-s| \leq \delta$. Zu t, s existiert ein $k \in \mathbb{N}_0$ mit $k \leq 2m-2$ und $\frac{k}{2m} \leq t < s \leq \frac{k}{2m} + \frac{1}{m}$.

Dann ist $\left|f(t) - f\left(\frac{k}{2m}\right)\right| \geq \varepsilon/2$ oder $\left|f(s) - f\left(\frac{k}{2m}\right)\right| \geq \varepsilon/2$. Also ist

$$\{f : \omega_\delta(f) \geq \varepsilon\} \subset \bigcup_{k=0}^{2m-2} \Big\{f : \sup_{\frac{k}{2m} \leq s \leq \frac{k}{2m}+\delta_0} \big|f(s) - f\big(\frac{k}{2m}\big)\big| \geq \varepsilon/2\Big\},$$

und somit gilt für alle $n \geq n_0$:

$$\begin{aligned}
\nu_n(\{f : \omega_\delta(f) \geq \varepsilon\}) &\leq \sum_{k=0}^{2m-2} \nu_n\Big(\Big\{f : \sup_{\frac{k}{2m} \leq s \leq \frac{k}{2m}+\delta_0} \big|f(s) - f\big(\frac{k}{2m}\big)\big| \geq \varepsilon/2\Big\}\Big) \\
&\leq (2m-1)\delta_0 \frac{\eta}{3} \leq (2+\delta_0)\frac{\eta}{3} \leq \eta \ .
\end{aligned}$$

Damit ist (10.9) gezeigt. □

Bemerkung 10.15 Die Bedingung in Lemma 10.14 folgt aus der folgenden Aussage: Für alle $\varepsilon > 0$ gilt:

$$\lim_{\delta \searrow 0} \limsup_{n\to\infty} \sup_{t\in[0,1-\delta]} \frac{1}{\delta}\nu_n\Big(\Big\{f : \sup_{t\leq s\leq t+\delta} |f(s) - f(t)| \geq \varepsilon\Big\}\Big) = 0 \ .$$

Die Bedingung aus Bemerkung 10.15 soll nun für $\mu_n = P^{Y_n}$ untersucht werden: Für $\delta \in (0,1)$ und $t \in [0, 1-\delta]$ ist

$$\mu_n\Big(\Big\{f : \sup_{t\leq s\leq t+\delta} |f(s) - f(t)| \geq \varepsilon\Big\}\Big) = P\Big(\sup_{t\leq s\leq t+\delta} |Y_n(s) - Y_n(t)| \geq \varepsilon\Big).$$

Für $t = k/n$ und $t + \delta = j/n$ $(k < j)$ ist

$$\sup_{t\leq s\leq t+\delta} |Y_n(s) - Y_n(t)| = \max_{1\leq i\leq n\delta} \frac{|S_{k+i} - S_k|}{\sqrt{n}}.$$

Für allgemeine $t \in [0,1]$ und $\delta \in (0,1)$ mit $t+\delta \leq 1$ kann man so abschätzen: Es existieren $j, k \in \{0, 1, \ldots, n\}$ mit $k < j$ und $\frac{k}{n} \leq t < \frac{k+1}{n}$ sowie $\frac{j-1}{n} < t+\delta \leq \frac{j}{n}$. Dann gilt für jedes $s \in [t, t+\delta]$:

$$\begin{aligned}
|Y_n(s) - Y_n(t)| &\leq \Big|Y_n(t) - Y_n\Big(\frac{k}{n}\Big)\Big| + \max_{1\leq i\leq j-k} \Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big| \\
&\leq 2 \max_{1\leq i\leq j-k} \Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big|,
\end{aligned}$$

also

$$\begin{aligned}
\sup_{t\leq s\leq t+\delta} |Y_n(s) - Y_n(t)| &\leq 2 \max_{1\leq i\leq j-k} \Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big| \\
&= 2 \max_{1\leq i\leq j-k} \Big|\sum_{r=k+1}^{k+i} X_r\Big| / \sqrt{n} \ .
\end{aligned}$$

Es ist $\frac{j-k-2}{n} \leq \delta$. Für $n \geq \frac{1}{\delta}$ folgt $j - k \leq 3n\delta$. Somit ist die rechte Seite der letzten Ungleichung nicht größer als $2\max_{1\leq i\leq 3n\delta} \Big|\sum_{r=k+1}^{k+i} X_r\Big|/\sqrt{n}$. Die

Verteilung dieser Zufallsvariablen hängt nicht von k ab. Für $n \geq \frac{1}{\delta}$ gilt somit

$$\sup_{t\in[0,1-\delta]} P\Big(\sup_{t\leq s\leq t+\delta} |Y_n(s) - Y_n(t)| \geq \varepsilon\Big) \leq P\Big(\max_{1\leq i\leq 3n\delta} \frac{|S_i|}{\sqrt{n}} \geq \frac{\varepsilon}{2}\Big) .$$

Sei $m := [3\,n\,\delta]$, so ist $\sqrt{n} \geq \sqrt{m/3\delta}$ und somit

$$P\Big(\max_{1\leq i\leq 3n\delta} \frac{|S_i|}{\sqrt{n}} \geq \frac{\varepsilon}{2}\Big) \leq P\Big(\max_{1\leq i\leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{12\delta}}\Big) .$$

Für jedes feste $\delta > 0$ geht $m \to \infty$ für $n \to \infty$. Nach Bemerkung 10.15 müssen wir für jedes $\varepsilon > 0$ zeigen, dass:

$$\lim_{\delta\searrow 0} \limsup_{m\to\infty} \frac{1}{\delta} P\Big(\max_{1\leq i\leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big) = 0 \tag{10.12}$$

gilt. Leider hilft die Abschätzung

$$P\Big(\max_{1\leq i\leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big) \leq \sum_{i=1}^{m} P\Big(\frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big)$$

analog zum Beweis von Lemma 10.14 nicht. Wir müssen diese Wahrscheinlichkeit wesentlich genauer abschätzen:

Lemma 10.16 *Für alle $\lambda > 0$ und $m \in \mathbb{N}$ gilt*

$$P\Big(\max_{1\leq i\leq m} |S_i| \geq \lambda\sqrt{m}\Big) \leq 2P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) .[2]$$

Beweis: Für $\lambda \leq \sqrt{2}$ ist nichts zu zeigen. Sei $\lambda > \sqrt{2}$.

$$A_i := \bigcap_{j=1}^{i-1} \{|S_j| < \lambda\sqrt{m}\} \cap \{|S_i| \geq \lambda\sqrt{m}\} \ , \ i = 1, \ldots, m.$$

Die A_i sind disjunkt und $A = \{\max_{1\leq i\leq m} |S_i| \geq \lambda\sqrt{m}\} = \bigcup_{i=1}^m A_i$. Also

$$\begin{aligned} P(A) &= P\Big(A \cap \{|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\}\Big) + P\Big(A \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}\Big) \\ &\leq P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) \\ &\quad + \sum_{j=1}^{m-1} P\Big(A_j \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\}\Big) \ , \end{aligned}$$

denn $A_m \cap \{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\} = \emptyset$. Weiter gilt

$$A_j \cap \big\{|S_m| < (\lambda - \sqrt{2})\sqrt{m}\big\} \subset A_j \cap \big\{|S_m - S_j| \geq \sqrt{2m}\big\} \ , \ j = 1, \ldots, m-1 \ .$$

Die Ereignisse A_j und $\{|S_m - S_j| \geq \sqrt{2}\sqrt{m}\}$ sind unabhängig, also haben wir

$$P(A) \leq P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) + \sum_{j=1}^{m-1} P(A_j) P\Big(|S_m - S_j| \geq \sqrt{2m}\Big).$$

[2]vgl. Kolmogorov-Ungleichung, Satz 6.7

Wegen

$$P\Big(|S_m - S_j| \geq \sqrt{2m}\Big) \leq \frac{1}{2m}\mathbb{E}\Big(\big(\sum_{k=j+1}^{m} X_k\big)^2\Big) = \frac{1}{2m}\sum_{k=j+1}^{m}\mathbb{E}(X_k^2) \leq \frac{1}{2}$$

folgt

$$\begin{aligned} P(A) &\leq P\big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\big) + \frac{1}{2}\sum_{j=1}^{m} P(A_j) \\ &= P\Big(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}\Big) + \frac{1}{2}P(A)\ , \end{aligned}$$

also folgt die Behauptung. □

Wir schließen mit dem Beweis von (10.12) ab:
Mit Lemma 10.16 und dem zentralen Grenzwertsatz folgt:

$$\begin{aligned} \limsup_{m\to\infty} \frac{1}{\delta} P\Big(\max_{1\leq i\leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big) &\leq \limsup_{m\to\infty} \frac{2}{\delta} P\Big(\frac{|S_m|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big) \\ &= \frac{2}{\delta} P\Big(|N| \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big), \end{aligned}$$

wenn N eine $N(0,1)$-verteilte Zufallsgröße bezeichnet. Die Markov-Ungleichung liefert

$$P\Big(|N| \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big) \leq \frac{\mathbb{E}(|N|^3)}{\big(\frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\big)^3}.$$

Dies führt zu (10.12). Somit ist die Straffheit der Folge $(\mu_n)_n$ bewiesen und somit Satz 10.5. □

Wir sammeln noch ein paar Eigenschaften des Wiener-Maß μ. Natürlich gilt

(i) $\mu(C[0,1]) = 1$
(ii) Die Familie der Projektionen $(\pi_t)_{t\in[0,1]}$ erfüllt:

$$\mu(\pi_t \leq \alpha) = \lim_{n\to\infty} \mu_n(\pi_t \leq \alpha) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \exp\big(-\frac{s^2}{2t}\big)\, ds$$

für $t > 0$ und $\mu(\pi_0 = 0) = \lim_{n\to\infty} \mu_n(\pi_0 = 0) = 1$.
(iii) Die Familie $(\pi_t)_{t\in[0,1]}$ hat *unabhängige Zuwächse*, d.h. für $0 = t_0 \leq \cdots \leq t_m \leq 1$, $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$, $m \in \mathbb{N}$, gilt

$$\mu(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i, i = 1, \ldots, m) = \prod_{i=1}^{m} \mu(\pi_{t_i} - \pi_{t_{i-1}} \leq \alpha_i).$$

Den Punkt (iii) sieht man mit Hilfe von Satz 10.3 so:

$$\mu(\pi_{t_i} - \pi_{t_{i-1}} \le \alpha_i, i = 1, \dots, m) = \lim_{n\to\infty} \mu_n(\pi_{t_i} - \pi_{t_{i-1}} \le \alpha_i, i = 1, \dots, m)$$
$$= \mu_{t_1} \otimes \mu_{t_2-t_1} \otimes \cdots \otimes \mu_{t_m-t_{m-1}} \Big(\times_{i=1}^m (-\infty, \alpha_i] \Big)$$
$$= \prod_{i=1}^m \lim_{n\to\infty} \mu_n(\pi_{t_i} - \pi_{t_{i-1}} \le \alpha_i)$$
$$= \prod_{i=1}^m \mu(\pi_{t_i} - \pi_{t_{i-1}} \le \alpha_i).$$

Definition 10.17 Die Familie $(\pi_t)_{t\in[0,1]}$ heißt BROWN*sche Bewegung*. Präziser nennt man das Tupel $(C, \mathcal{B}_C, \mu, (\pi_t)_{t\in[0,1]})$ so.

Damit ist uns schon der wohl wichtigste stochastische Prozess, gemeint ist die Familie $(\pi_t)_{t\in[0,1]}$, begegnet, die BROWNsche Bewegung eines Teilchens (z.B. Pollen- oder Staubkorns, eines markierten Moleküls) in einer Flüssigkeit oder einem Gas. Der Ort $x_t \in \mathbb{R}^3$ (bei uns zunächst in $\mathbb{R}$) wird durch eine Zufallsvariable π_t beschrieben. BROWN entdeckte 1828 das Phänomen dieser Bewegung. EINSTEIN entwickelte 1905 die physikalische Theorie, unabhängig davon 1906 SMOLUCKOWSKI. EINSTEIN beschreibt die Bewegung eines Teilchens unter Berücksichtigung von Kollisionen mit vielen Teilchen und nimmt unabhängige Zuwächse und zeitlich stationäre Zuwächse an. Er bestimmt die Verteilung des Zuwachses in $[0, t]$ als Normalverteilung $N(0, \sigma^2)$ mit $\sigma^2 = 2t$. BACHELIER untersuchte 1900 in seiner bei POINCARÉ geschriebenen Dissertation ökonomische Agenten zur Beschreibung von Kursschwankungen an der Pariser Börse. Dabei nahm er für Fluktuationen in $[0, t]$ eine Normalverteilung $N(0, 2t)$ an! Der mathematische Begriff der BROWNschen Bewegung wurde 1920 von N.WIENER geprägt.

ANHANG: Beweis des Satzes von ARZELÀ-ASCOLI:

Wir bereiten den Beweis durch ein Kriterium für Kompaktheit von Mengen in metrischen Räumen vor.

Satz 10.18 *Eine Teilmenge eines metrischen Raumes (X, d) ist genau dann kompakt, wenn sie vollständig und totalbeschränkt ist. Dabei heißt $K \subset X$* totalbeschränkt, *wenn es zu jedem $r > 0$ ein $m \in \mathbb{N}$ und $x_0, \dots, x_m \in K$ gibt mit $K \subset \bigcup_{k=0}^m B(x_k, r)$ (womit jede totalbeschränkte Menge beschränkt ist).*

Beweis: Es sei $K \subset X$ kompakt, $(x_j)_j$ sei eine Cauchyfolge in K. K ist folgenkompakt (denn eine Teilmenge eines metrischen Raums ist genau dann kompakt, wenn sie folgenkompakt ist, Analysis I), also besitzt $(x_j)_j$ eine in K konvergente Teilfolge. Damit konvergiert die Folge (denn besitzt eine Cauchyfolge eine konvergente Teilfolge, so ist sie selbst konvergent, Analysis I) in K, also ist K vollständig. Für jedes $r > 0$ ist $\{B(x,r), x \in K\}$ eine offene Überdeckung von K. Da K kompakt, gibt es eine endliche Teilüberdeckung, also ist K auch totalbeschränkt.

Sei nun K vollständig und totalbeschränkt. $(x_j)_j$ sei eine Folge in K. Zu jedem $n \in \mathbb{N}$ existieren endlich viele Bälle mit Mittelpunkten in K und Radius $1/n$, die K überdecken. Es existiert also eine Teilfolge $(x_{1,j})_j$ von $(x_j)_j$, die ganz in einem Ball mit Radius 1 enthalten ist. Dann gibt es eine Teilfolge $(x_{2,j})_j$ von $(x_{1,j})_j$, die ganz in einem Ball mit Radius $1/2$ enthalten ist, etc. Also gibt es zu jedem $n \in \mathbb{N}$ eine Teilfolge $(x_{n+1,j})_j$ von $(x_{n,j})_j$, die ganz in einem Ball mit Radius $1/(n+1)$ enthalten ist. Sei $y_n := x_{n,n}$, $n \in \mathbb{N}$ (Diagonalfolge). Dann ist $(y_n)_n$ offensichtlich eine Cauchyfolge in K, also konvergiert $(y_n)_n$ in K, da K vollständig. $(x_j)_j$ hat also eine in K konvergente Teilfolge: $(y_n)_n$, also ist K folgenkompakt, also kompakt. $\square$

Im zweiten Teil des Beweises haben wir das *Diagonalfolgenprinzip* verwendet. Wir wählen aus eine Folge gemäß einer Vorschrift sukzessive Teilfolgen aus und bilden dann die Diagonalfolge, indem man von der n-ten Teilfolge das n-te Glied auswählt. Hier ist $(x_{n+1,j})_j$ Für jedes $n \in \mathbb{N}$ eine Teilfolge von $(x_{n,j})_j$. Die Diagonalfolge $(y_n)_n$ hat dann die Eigenschaft, dass $(y_n)_{n \geq N}$ für jedes $N \in \mathbb{N}$ eine Teilfolge von $(x_{N,j})_j$ ist, also dieselben infinitären Eigenschaften wie jede der Teilfolgen $(x_{n,j})_j$ besitzt.

Da $A \subset X$ 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 ARZELÀ-ASCOLI) Sei $\bar{A} \subset C[0,1]$ kompakt. Dann ist A totalbeschränkt: zu $\varepsilon > 0$ existieren $f_1, \ldots, f_n \in A$ mit $d(f, f_j) < \varepsilon/3$ für ein $j \in \{1, \ldots, n\}$ für alle $f \in A$. Jedes f_j in $C[0,1]$ ist gleichmäßig stetig, also gilt für die endliche Menge $\{f_1, \ldots, f_n\}$: Wähle $\delta > 0$, so dass $|x-y| < \delta$ $|f_j(x) - f_j(y)| < \varepsilon/3$ für alle $j = 1, \ldots, n$ und $x, y \in [0,1]$ zur Folge hat. Also ist $|f(x) - f(y)| < \varepsilon$ für alle $f \in A$, somit gilt $\lim_{\delta \to 0} \sup_{f \in A} \omega_\delta(f) = 0$. A ist auch beschränkt bezüglich d, was (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

$$|f(t)| \leq |f(0)| + \sum_{i=1}^{k} \left| f\Big(\frac{i}{k}t\Big) - f\Big(\frac{i-1}{k}t\Big) \right| ,$$

folgt mit (i)

$$\sup_{t \in [0,1]} \sup_{f \in A} |f(t)| < \infty \,. \tag{10.13}$$

Wir zeigen nun, dass aus (ii) und (10.13) folgt, dass A totalbeschränkt ist, also auch $\bar{A}$. Nun ist $C[0,1]$ vollständig, also auch $\bar{A}$, damit ist $\bar{A}$ dann kompakt.

Sei $\varepsilon > 0$ und

$$\alpha := \sup_{t\in[0,1]} \sup_{f\in A} |f(t)| \ .$$

Ferner sei $H := \{\frac{u}{v}\alpha, u = 0, \pm 1, \pm 2, \ldots, \pm v, v \in \mathbb{N}\}$ mit $v \in \mathbb{N}$ so, dass $\frac{\alpha}{v} < \varepsilon$. H hat dann die Eigenschaft, dass zu jedem $t \in [-\alpha, \alpha]$ ein $t_k \in H$ existiert mit $|t - t_k| < \varepsilon$ Nun wähle k groß genug, so dass $\omega_{1/k}(f) < \varepsilon$ für alle $f \in A$. B sei die Teilmenge in $C[0,1]$ derjenigen Funktionen, die in jedem Intervall $[\frac{i-1}{k}, \frac{i}{k}]$, $i = 1, \ldots, k$, linear sind und Werte aus H an den Endpunkten $\frac{i}{k}$, $i = 0, \ldots, k$, annehmen. B ist endlich (besteht aus $(2\nu + 1)^{k+1}$ Punkten). Wir zeigen nun, dass jedes $f \in A$ in einem 2ε-Ball um ein Element aus B liegt: Sei $f \in A$, also $|f(\frac{i}{k})| \le \alpha$. Dann existiert ein $g \in B$ mit

$$|f(\frac{i}{k}) - g(\frac{i}{k})| < \varepsilon \ , i = 0, \ldots, k \ . \tag{10.14}$$

Da $\omega_{1/k}(f) < \varepsilon$ und g linear in jedem Teilintervall $[\frac{i-1}{k}, \frac{i}{k}]$ ist, folgt aus (10.14) $d(f, g) < 2\varepsilon$. Dies war zu zeigen. □

KAPITEL 11

Anwendungen des Invarianzprinzips, die eindimensionale Irrfahrt

Im ersten Beispiel soll Satz 10.9, das Invarianzprinzip von DONSKER, auf $h : C[0,1] \to \mathbb{R}$, $h(f) := \sup_{0 \le t \le 1} f(t)$ angewendet werden. Es ist leicht einzusehen, dass h stetig ist.

Nun seien die Zufallsgrößen $(X_i)_i$ unabhängig mit

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

also $\mathbb{E}X_i = 0$ und $\operatorname{Var} X_i = \mathbb{E}X_i^2 = 1$. Wieder sei $S_0 = 0$ und $S_n = \sum_{i=1}^n X_i$, sowie $M_n := \max_{1 \le i \le n} S_i$ für $n \in \mathbb{N}$. Wir interessieren uns für die Verteilung von M_n. Man beachte:

$$\sup_{0 \le t \le 1} Y_n(t) = \max_{0 \le i \le n} \frac{S_i}{\sqrt{n}} = \frac{M_n}{\sqrt{n}}\,.$$

$(M_n)_n$ heißt auch die Folge der *Maximalgewinne* beim Münzwurfspiel.

Satz 11.1 *Für die Folge $(M_n)_n$ der Maximalgewinne beim Münzwurfspiel gilt für alle $t \ge 0$*

$$\lim_{n \to \infty} P\Big(\frac{M_n}{\sqrt{n}} \le t\Big) = 2\Phi(t) - 1\,.$$

Für $t < 0$ gilt

$$P\Big(\frac{M_n}{\sqrt{n}} \le t\Big) = 0\,.$$

Satz 11.2 *Erfüllen die $(X_i)_i$ die Voraussetzungen des Satzes 10.9, so gilt für alle $t \in \mathbb{R}$*

$$\lim_{n \to \infty} P\Big(\max_{0 \le i \le n} \frac{S_i}{\sqrt{n}} \le t\Big) = \max\{2\Phi(t) - 1,\, 0\}\,.$$

Für den Beweis von Satz 11.1 bereiten wir das sogenannte *Spiegelungsprinzip/Reflexionsprinzip* vor.

Für $i, j \in \mathbb{Z}$, $i < j$, nennen wir eine Folge $(i, s_i), \ldots, (j, s_j)$ mit $s_k \in \mathbb{Z}$, $i \le k \le j$, und $|s_{k+1} - s_k| = 1$ für $i \le k \le j-1$ einen *Pfad* von (i, s_i) nach (j, s_j). Oft schreibt man einfach $(s_i, s_{i+1}, \ldots, s_j)$. $j - i$ ist die *Länge* des Pfades. Wir sagen, dass ein Pfad $(s_i, s_{i+1}, \ldots, s_j)$ die x-Achse berührt, falls ein k mit $i \le k \le j$ existiert, für das $s_k = 0$ ist.

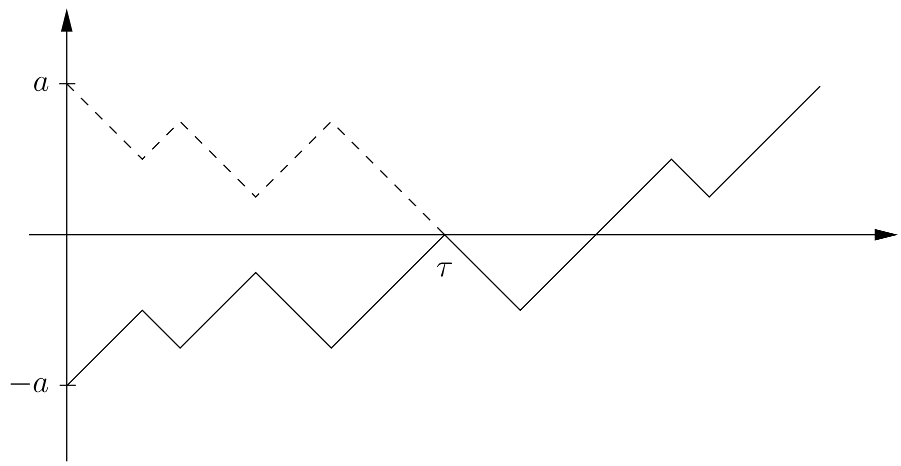
Lemma 11.3 (Reflexionsprinzip) (i) *Es seien $a, b \in \mathbb{N}$ und $i, j \in \mathbb{Z}$ mit $i < j$. Die Anzahl der Pfade von $(i,\, a)$ nach $(j,\, b)$, welche die x-Achse berühren, ist gleich der Anzahl der Pfade von $(i,\, -a)$ nach $(j,\, b)$.*

(ii) *Sei $a \in \mathbb{N}$, $b \in \mathbb{Z}$ und $n \in \mathbb{N}$. Die Anzahl der Pfade von $(0,0)$ nach (n,b), die $s_j = a$ für ein $j \in \{1,\ldots,n\}$ erfüllen, ist gleich der Anzahl der Pfade von $(0,0)$ nach $(n,\, 2a-b)$, die $s_j = a$ für ein $j \in \{1,\ldots,n\}$ erfüllen.*

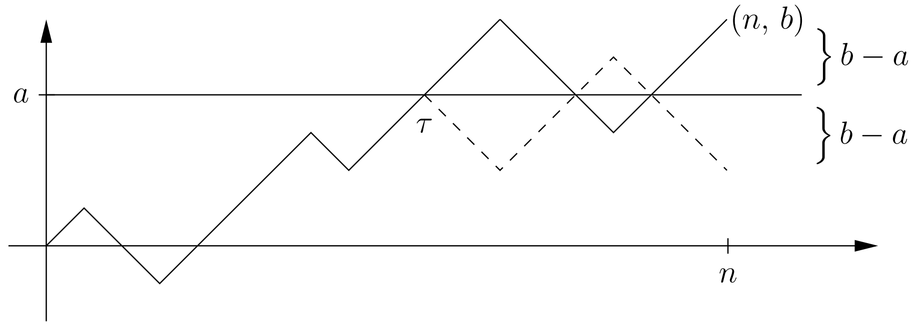
Beweis: (i) Sei $(s_i = -a,\, s_{i+1},\ldots,s_{j-1}, s_j = b)$. Dieser Pfad muss die x-Achse berühren. τ sei die kleinste Zahl größer als i, für welche $s_\tau = 0$ gilt. Dann ist

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

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



Das Bild für den Beweis von (ii) ist



τ ist das erstmalige Erreichen des Wertes a. □

Beweis von Satz 11.1: Für $l, k \in \mathbb{N}_0$ gilt

$$P(S_n = l+k) = P(S_n = l+k,\, M_n \geq k)\,.$$

Nun ist nach Teil (ii) von Lemma 11.3

$$P(M_n \geq a,\, S_n = b) = P(M_n \geq a,\, S_n = 2a-b)$$

für jedes $b \in \mathbb{Z}$. Also ist

$$P(S_n = l + k) = P(M_n \geq k,\, S_n = k - l)\,.$$

Damit ist

$$\begin{aligned} P(M_n \geq k) &= \sum_{l=-\infty}^{\infty} P(M_n \geq k,\, S_n = l + k) \\ &= \sum_{l=-\infty}^{-1} P(M_n \geq k,\, S_n = l + k) + \sum_{l=1}^{\infty} P(S_n = l + k) + P(S_n = k) \\ &= 2P(S_n > k) + P(S_n = k) \\ &= 2P(S_n \geq k) - P(S_n = k)\,. \end{aligned}$$

Sei $t \in \mathbb{R}_+$. Für $n \in \mathbb{N}$ bezeichne k_n die kleinste ganze Zahl größer-gleich $t\sqrt{n}$. Es gilt

$$P^{S_n/\sqrt{n}} \xrightarrow{w} N(0,1)\,.$$

Da $\{S_n/\sqrt{n} \geq t\} = \{S_n \geq k_n\}$, folgt

$$\lim_{n\to\infty} P(S_n \geq k_n) = \nu_{0,1}\big([t,\,\infty)\big)\,.$$

Wegen $t\sqrt{n} \leq k_n < t\sqrt{n} + 1$ gilt weiter für jedes $\varepsilon > 0$ und alle $n \in \mathbb{N}$ mit $1/\sqrt{n} \leq \varepsilon$

$$\{S_n = k_n\} = \Big\{\frac{S_n}{\sqrt{n}} = \frac{k_n}{\sqrt{n}}\Big\} \subset \Big\{t \leq \frac{S_n}{\sqrt{n}} < t + \varepsilon\Big\}\,,$$

und daraus folgt

$$\limsup_{n\to\infty} P(S_n = k_n) \leq \int_t^{t+\varepsilon} g_{0,1}(x)\,dx \quad \forall \varepsilon \geq 0\,,$$

also

$$\lim_{n\to\infty} P(S_n = k_n) = 0\,.$$

Zusammen erhalten wir

$$\begin{aligned} \lim_{n\to\infty} P\Big(\frac{M_n}{\sqrt{n}} \geq t\Big) &= 2\nu_{0,1}\big([t,\,\infty)\big) \\ &= \Big(\frac{2}{\pi}\Big)^{1/2} \int_t^{\infty} e^{-x^2/2}\,dx = 2\big(1 - \Phi(t)\big)\,, \end{aligned}$$

womit die Behauptung des Satzes folgt. □

Im zweiten Beispiel betrachten wir die Abbildung $g : C[0,1] \to \mathbb{R}$ mit

$$g(f) := \lambda\big(\{t \in [0,1] : f(t) \geq 0\}\big)\,,$$

wobei λ das Lebesgue-Maß sei. Die Funktion g ist nicht überall stetig, z.B. ist sie unstetig in $f \equiv 0$. Es gilt jedoch

Lemma 11.4 *g ist $\mathcal{B}_C/\mathcal{B}$-messbar und $\mu(D_g) = 0$, wobei μ das Wiener-Maß bezeichnet.*

Beweis: Es sei $\psi : C[0,1] \times [0,1] \to \mathbb{R}$ definiert durch $\psi(f, t) = f(t)$. ψ ist stetig (Übung!), also $\mathcal{B}_{C\times[0,1]}/\mathcal{B}$-messbar, wobei wir wieder kurz $C := C[0,1]$ schreiben. Da C und $[0,1]$ separabel sind, folgt analog zu Lemma 9.22

$$\mathcal{B}_{C\times[0,1]} = \mathcal{B}_C \otimes \mathcal{B}_{[0,1]} .$$

Also ist ψ $\mathcal{B}_C \otimes \mathcal{B}_{[0,1]}/\mathcal{B}$-messbar.

Sei nun $A = \{(f, t) : f(t) \geq 0\} = \psi^{-1}\big([0,\infty)\big) \in \mathcal{B}_C \otimes \mathcal{B}_{[0,1]}$. Für $f \in C$ ist $g(f) = \lambda\big(\{t : (f, t) \in A\}\big)$. Also ist $f \mapsto g(f)$ $\mathcal{B}_C/\mathcal{B}$-messbar, siehe Satz 35.13 (Fubini), Analysis III.

Es gilt

$$g(f) = \int_0^1 1_{[0,\infty]}\big(f(t)\big)\, dt .$$

Ist $f \in C$ mit $\lambda\big(\{t : f(t) = 0\}\big) = 0$, und ist $(f_n)_n$ eine Folge in C mit $d(f_n, f) \to 0$, so gilt $1_{[0,\infty]}\big(f_n(t)\big) \to 1_{[0,\infty]}\big(f(t)\big)$ für λ-fast alle $t \in [0,1]$. Nach dem Satz von der dominierten Konvergenz folgt

$$g(f_n) \to g(f) .$$

Also ist $D_g \subset \Big\{f : \lambda\big(\{t : f(t) = 0\}\big) > 0\Big\}$ gezeigt.

Wir zeigen

$$\mu\Big(\Big\{f : \lambda\big(\{t : f(t) = 0\}\big) > 0\Big\}\Big) = 0.$$

Dazu müssen wir zeigen, dass $f \mapsto \lambda\big(\{t : f(t) = 0\}\big)$ messbar ist. Dies geht analog zur Messbarkeit von g. Es ist zu zeigen:

$$0 = \int_C \lambda\big(\{t : f(t) = 0\}\big)\, \mu(df) = \int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f, t)\, dt\, \mu(df) .$$

Nach dem Satz von Fubini gilt

$$\begin{aligned}\int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f, t)\, dt\, \mu(df) &= \int_{[0,1]} \int_C (1_{\{0\}} \circ \psi)(f, t)\, \mu(df)\, dt \\ &= \int_{[0,1]} \mu\big(\{f : f(t) = 0\}\big)\, dt \\ &= \int_{[0,1]} \mu\pi_t^{-1}\big(\{0\}\big)\, dt .\end{aligned}$$

Das letzte Integral ist tatsächlich gleich Null, denn $\mu\pi_t^{-1}$ ist für $t > 0$ die Normalverteilung mit Erwartungswert 0 und Varianz t. Damit ist das Lemma bewiesen. □

Die Abbildung g erfüllt also die Voraussetzung des Invarianzprinzips, Satz 10.9. Es folgt nun die Berechnung von $\mathcal{L}\big(g(Y_n)\big)$ im Spezialfall, wo $P(X_i = \pm 1) = 1/2$ ist. Dies ist eine elementare und schöne Auseinandersetzung mit der eindimensionalen, symmetrischen Irrfahrt und hebt die Bedeutung des Reflexionsprinzips eindrücklich hervor. Es gilt:

Satz 11.5 *Sind die $(X_i)_i$ unabhängig und $P(X_i = \pm 1) = 1/2$, so gilt für $t \in [0,1]$*

$$\lim_{n\to\infty} P\big(g(Y_n) \le t\big) = \frac{2}{\pi} \arcsin \sqrt{t}\,.$$

Dies liefert somit die Verteilungsfunktion von $\mu\, g^{-1}$, wenn μ das Wiener-Maß ist. Es folgt mit Satz 10.9:

Satz 11.6 (Arcussinus-Gesetz) *Die auf $(C, \mathcal{B}_C, \mu)$ definierte Zufallsgröße $f \mapsto \lambda\big(\{t : f(t) \ge 0\}\big)$ hat die Verteilungsfunktion*

$$t \mapsto \frac{2}{\pi} \arcsin \sqrt{t}\,, \quad t \in [0,1]\,.$$

Erfüllen die $(X_i)_i$ die Voraussetzungen von Satz 10.9, so gilt für $t \in [0,1]$:

$$\lim_{n\to\infty} P\big(g(Y_n) \le t\big) = \frac{2}{\pi} \arcsin \sqrt{t}\,.$$

Bemerkung 11.7 Es ist nicht sehr schwer zu zeigen, dass

$$g(Y_n) - \frac{1}{n}\big|\{m \le n : S_m > 0\}\big|$$

in Wahrscheinlichkeit gegen 0 konvergiert. Nach Lemma 9.25 folgt dann, dass auch

$$\mathcal{L}\Big(\frac{1}{n}\big|\{m \le n : S_m > 0\}\big|\Big)$$

asymptotisch nach der Arcussinus-Verteilung verteilt ist. Wir zeigen dies hier nicht.

Zunächst betrachten wir einige kombinatorische Resultate zu Pfaden, so wie sie von unserem Münzwurfspiel der $(X_i)_i$ erzeugt werden. Wir betrachten zwei verschiedene Zufallsexperimente:

(I) *Der Endpunkt liegt fest:* Ist $n \in \mathbb{N}$ und hat s, $s \in \mathbb{Z}$, dieselbe Parität wie n, so bezeichne $\Omega_{(n,s)}$ die Menge der Pfade von $(0,0)$ nach (n,s). Auf dieser Menge betrachten wir die Gleichverteilung. Wir müssen zunächst die Anzahl der Pfade zählen: Hat ein Pfad $\omega \in \Omega_{(n,s)}$ p ansteigende Verbindungen und q absteigende (d.h. $p := |\{i \in \{0,\dots,n-1\} : s_{i+1} = s_i + 1\}|$), so gelten $p+q = n$, $p - q = s$, das heißt $p = (n+s)/2$, $q = (n-s)/2$. p und q sind also durch n und s vollständig festgelegt.

$|\Omega_{(n,s)}|$ ist die Anzahl der Möglichkeiten, die p aufsteigenden Verbindungen in der Gesamtzahl von n Schritten zu plazieren, das heißt, es gilt

$$|\Omega_{(n,s)}| = \binom{n}{(n+s)/2} = \binom{p+q}{p}. \tag{11.1}$$

(II) *Freier Endpunkt:* Ω_n bezeichne die Menge aller Pfade der Länge n mit Startpunkt $(0,0)$. $|\Omega_n|$ ist hier offenbar 2^n.

Wir betrachten zunächst den Fall (I), das heißt das Zufallsexperiment, das durch die Gleichverteilung auf $\Omega_{(n,s)} = \Omega_{(p+q,p-q)}$ beschrieben wird.

Wir können uns etwa vorstellen, dass eine Wahl zwischen zwei Kandidaten K_1, K_2 stattgefunden hat, wobei nun p Stimmen für K_1 und q Stimmen für K_2 in einer Wahlurne liegen. Diese Stimmen werden nun eine um die andere ausgezählt. Wir wollen zunächst das folgende Ereignis betrachten. Sei $p > q$ (d.h. K_1 hat gewonnen). Mit welcher Wahrscheinlichkeit liegt K_1 stets vorn bei der Auszählung? Diese Wahrscheinlichkeit ist gleich $|A|/|\Omega_{(p+q,p-q)}| = |A|/\binom{p+q}{p}$, wobei

$$A = \{\, \omega = (0, s_1, \ldots, s_{p+q}) \in \Omega_{(p+q,p-q)} : s_k > 0 \text{ für } 1 \leq k \leq p+q \,\}$$

ist. Zum Abzählen der Pfade in A verwenden wir Lemma 11.3. Für $\omega = (0, s_1, \ldots, s_n) \in A$ gilt notwendigerweise $s_1 = 1$. $|A|$ ist somit die Anzahl der Pfade von $(1,1)$ nach $(p+q, p-q)$, die die x-Achse nicht berühren. Dies ist gleich der Anzahl aller Pfade von $(1,1)$ nach $(p+q, p-q)$, minus der Anzahl derjenigen, die die x-Achse berühren. Letztere ist nach Lemma 11.3 gleich der Anzahl aller Pfade von $(1,-1)$ nach $(p+q, p-q)$. Wenden wir (11.1) an, so ergibt sich also

$$|A| = \binom{p+q-1}{p-1} - \binom{p+q-1}{p} = \frac{p-q}{p+q}\binom{p+q}{p}. \tag{11.2}$$

(Wir haben hier natürlich $p > q$ vorausgesetzt.) Die Anzahl aller Elemente in $\Omega_{(p+q,p-q)}$ ist nach (11.1) $\binom{p+q}{p}$. Somit ergibt sich das folgende Resultat:

Satz 11.8 (Ballot-Theorem, von ballot (engl.) = geheime Abstimmung) *Die Wahrscheinlichkeit dafür, dass der Kandidat mit der größeren Anzahl p der Stimmen während des gesamten Verlaufs der Auszählung führt, ist $(p-q)/(p+q)$, wobei q die Anzahl der Stimmen des Unterlegenen bezeichnet.*

Eine kleine Modifikation des obigen Arguments gestattet auch die Diskussion des Falles $p = q$. Natürlich kann dann keiner der Kandidaten dauernd führen, da nach der Auszählung Gleichstand herrscht. Wir können aber die beiden folgenden Ereignisse betrachten:

(i) Kandidat K_1 führt während der gesamten Auszählung, erst am Schluß tritt Gleichstand ein.
(ii) Kandidat K_2 führt nie.

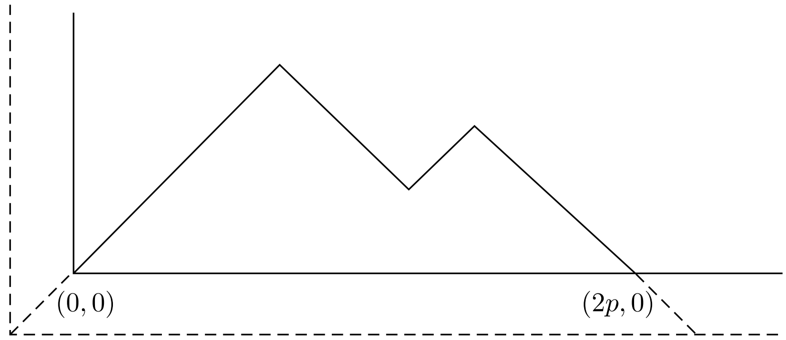
Da der zugrunde liegende W.-Raum $\binom{2p}{p}$ Elementarereignisse hat, die alle die gleiche Wahrscheinlichkeit haben, ergeben sich aus dem folgenden Satz die Wahrscheinlichkeiten für diese beiden Ereignisse:

Satz 11.9 (i) *Es gibt $\frac{1}{p}\binom{2p-2}{p-1}$ Pfade von $(0,0)$ nach $(2p,0)$ mit $s_1 > 0$, $s_2 > 0, \ldots, s_{2p-1} > 0$.*
(ii) *Es gibt $\frac{1}{p+1}\binom{2p}{p}$ Pfade von $(0,0)$ nach $(2p,0)$ mit $s_1 \geq 0$, $s_2 \geq 0, \ldots, s_{2p-1} \geq 0$.*

Beweis: (i) Natürlich ist notwendigerweise $s_{2p-1} = 1$. Wir suchen somit nach der Anzahl der Pfade von $(0,0)$ nach $(2p-1,1)$ mit $s_1 > 0$, $s_2 > 0, \ldots, s_{2p-1} = 1$. Nach der Formel (11.2) mit $q = p-1$ ist dies gleich

$$\frac{1}{2p-1}\binom{2p-1}{p} = \frac{1}{p}\binom{2p-2}{p-1}.$$

(ii) Wir verlängern jeden Pfad, der die Bedingung erfüllt, indem wir noch die beiden Punkte $(-1,-1)$ und $(2p+1,-1)$ anfügen und mit $(0,0)$ bzw. $(2p,0)$ verbinden.

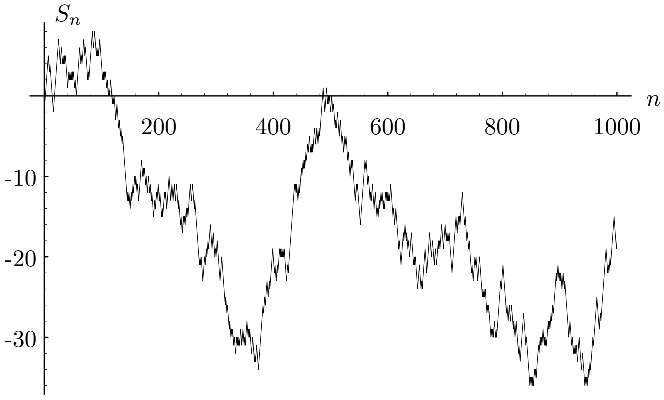


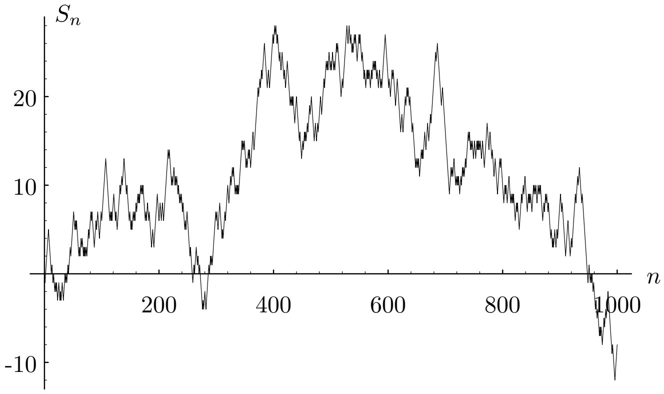
Auf diese Weise wird eine bijektive Abbildung von der gesuchten Menge von Pfaden auf die Menge der Pfade von $(-1,-1)$ nach $(2p+1,-1)$, welche die Bedingung $s_0 > -1$, $s_1 > -1, \ldots, s_{2p} > -1$ erfüllen, hergestellt. Die Anzahl der Pfade in dieser Menge ist gleich der Anzahl der Pfade von $(0,0)$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

Aus (ii) des obigen Satzes folgt, dass bei Gleichstand der Stimmen mit Wahrscheinlichkeit $1/(p+1)$ der Kandidat K_2 zu keinem Zeitpunkt der Auszählung führt. Das Gleiche gilt auch für den Kandidaten K_1. Mit Wahrscheinlichkeit $2/(p+1)$ wechselt somit die Führung nie.

Wir wenden uns nun der Situation (II) zu, das heißt dem Zufallsexperiment, das durch die Gleichverteilung auf Ω_n beschrieben wird. Dies ist nichts anderes als eine Umformulierung unseres Münzwurfexperiments mit Werten -1, 1. Einem Element $(a_1, \ldots, a_n) \in \{-1,1\}^n$ können wir einen Pfad $(s_0 = 0, s_1, \ldots, s_n) \in \Omega_n$ durch $s_k = \sum_{j=1}^k a_j$, $1 \le k \le n$, zuordnen. Dies definiert eine bijektive Abbildung $\{-1,1\}^n \to \Omega_n$. Der Gleichverteilung auf $\{-1,1\}^n$ entspricht dann via dieser bijektiven Abbildung die Gleichverteilung auf Ω_n.

Nachfolgend sind zwei Simulationen einer derartigen Irrfahrt mit $n = 1000$ abgebildet. Aus dem Gesetz der großen Zahlen folgt, dass zum Beispiel $S_{1000}/1000$ mit großer Wahrscheinlichkeit nahe bei 0 liegt. Um etwas zu „sehen“ müssen wir die y-Achse gegenüber der x-Achse strecken.





Zunächst betrachten wir für $k \leq n$ das Ereignis $A_k = \{S_k = 0\}$. A_k ist das unmögliche Ereignis, falls k ungerade ist. Wir betrachten also A_{2k}, $2k \leq n$. Um die Anzahl der Pfade der Länge n zu bestimmen, die zu A_{2k} gehören, multiplizieren wir die Anzahl der Pfade der Länge $2k$ von $(0,0)$ nach $(2k,0)$ mit der Anzahl der Pfade der Länge $n-2k$, die in $(2k,0)$ starten (bei freiem Ende). Somit ist

$$|A_{2k}| = \binom{2k}{k} 2^{n-2k}.$$

Ω_n enthält 2^n Elemente. Also gilt

$$P(A_{2k}) = \binom{2k}{k} 2^{-2k}.$$

Wir kürzen diese Größe auch mit u_{2k} ab ($u_0 = 1$). Man sieht zunächst nicht, von welcher Größenordnung $u_{2k} = P(A_{2k})$ für große k ist. Da

$$u_{2k} = \frac{(2k)!}{(k!)^2} 2^{-2k}$$

ist, benötigen wir eine genauere Kenntnis des Verhaltens der Fakultätsfunktion für große Argumente. Diese erhält man über die STIRLING-Approximation.

$$\lim_{n\to\infty} n!/(\sqrt{2\pi} n^{n+1/2} e^{-n}) = 1. \tag{11.3}$$

Für zwei reelle Zahlenfolgen $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, mit $a_n, b_n > 0$ schreiben wir $a_n \sim b_n$, sofern

$$\lim_{n\to\infty} a_n/b_n = 1$$

gilt. Dies bedeutet keineswegs, dass $|a_n - b_n|$ gegen 0 konvergiert. So gilt etwa

$$\lim_{n\to\infty} |n! - \sqrt{2\pi}n^{n+1/2}e^{-n}| = \infty.$$

Setzen wir die Stirling Approximation ein, so erhalten wir

Satz 11.10 *Es gilt*

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

Satz 11.10 ist eine recht gute Näherung für u_{2k}. Um dies genauer zu diskutieren, brauchte man gute Abschätzungen für die Differenz $n! - \sqrt{2\pi}n^{n+1/2}e^{-n}$. Wir wollen diesen Punkt jedoch nicht weiter verfolgen.

Interessanterweise lassen sich die Wahrscheinlichkeiten einer Reihe anderer Ereignisse in Beziehung zu u_{2k} setzen. Es sei zunächst für $k \in \mathbb{N}$ f_{2k} die Wahrscheinlichkeit, dass die erste Nullstelle der Irrfahrt nach dem Zeitpunkt 0 die Zeitkoordinate $2k$ hat, das heißt

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

Lemma 11.11 (i) $f_{2k} = \frac{1}{2k}u_{2k-2} = P(S_1 \geq 0,\, S_2 \geq 0, \ldots, S_{2k-2} \geq 0,\, S_{2k-1} < 0) = u_{2k-2} - u_{2k}$.

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

(iii) $u_{2k} = \sum_{j=1}^{k} f_{2j}u_{2k-2j}$.

Beweis: (i) Nach Satz 11.9, (i) gibt es $\frac{1}{k}\binom{2k-2}{k-1}$ Pfade von $(0,0)$ nach $(2k,0)$ mit $s_1 > 0, \ldots, s_{2k-1} > 0$ und natürlich genauso viele mit $s_1 < 0, \ldots, s_{2k-1} < 0$. Es folgt

$$f_{2k} = \frac{2}{k}\binom{2k-2}{k-1}2^{-2k} = \frac{1}{2k}\binom{2k-2}{k-1}2^{-2(k-1)} = \frac{1}{2k}u_{2k-2}.$$

Wir beweisen die nächste Gleichung: Falls $s_{2k-2} \geq 0$ und $s_{2k-1} < 0$ sind, so gelten $s_{2k-2} = 0$ und $s_{2k-1} = -1$. Die Anzahl der Pfade von $(0,0)$ nach $(2k-1,-1)$ mit $s_1 \geq 0, \ldots, s_{2k-3} \geq 0$, $s_{2k-2} = 0$ ist gleich der Anzahl der Pfade von $(0,0)$ nach $(2k-2,0)$ mit allen y-Koordinaten ≥ 0. Die zweite Gleichung in (i) folgt dann mit Hilfe von Satz 11.9, (ii). Die dritte ergibt sich aus

$$u_{2k} = \binom{2k}{k}2^{-2k} = \frac{2k(2k-1)}{k\cdot k}\binom{2k-2}{k-1}\cdot\frac{1}{4}\cdot 2^{-2k+2} = \left(1 - \frac{1}{2k}\right)u_{2k-2}.$$

(ii) C_{2j} sei das Ereignis $\{S_1 \neq 0, S_2 \neq 0, \ldots, S_{2j-1} \neq 0, S_{2j} = 0\}$. Diese Ereignisse schließen sich gegenseitig aus und haben Wahrscheinlichkeiten $f_{2j} = u_{2j-2} - u_{2j}$. Somit ist mit $u_0 = 1$

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

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

(iii) Für $1 \leq j \leq k$ sei $B_j = \{S_1 \neq 0, S_2 \neq 0, \ldots, S_{2j-1} \neq 0, S_{2j} = 0, S_{2k} = 0\}$. Diese Ereignisse sind paarweise disjunkt, und ihre Vereinigung ist $\{S_{2k} = 0\}$. $|B_j|$ ist offenbar gleich der Anzahl der Pfade von $(0,0)$ nach $(2j,0)$, die die x-Achse dazwischen nicht berühren, multipliziert mit der Anzahl aller Pfade von $(2j,0)$ nach $(2k,0)$, das heißt $|B_j| = 2^{2j} f_{2j} 2^{2k-2j} u_{2k-2j}$. Somit gilt $P(B_j) = f_{2j} u_{2k-2j}$, das heißt

$$u_{2k} = \sum_{j=1}^{k} P(B_j) = \sum_{j=1}^{k} f_{2j} u_{2k-2j}.$$

$\square$

Eine interessante Folgerung ergibt sich aus der ersten Gleichung in (ii). Da nach Satz 11.10 $\lim_{k\to\infty} u_{2k} = 0$ gilt, folgt, dass die Wahrscheinlichkeit für keine Rückkehr der Irrfahrt bis zum Zeitpunkt $2k$ mit $k \to \infty$ gegen 0 konvergiert. Man kann das folgendermaßen ausdrücken: „Mit Wahrscheinlichkeit 1 findet irgendwann eine Rückkehr statt.“ Man sagt auch, die Irrfahrt sei rekurrent. Wir wollen das noch etwas genauer anschauen und bezeichnen mit T den Zeitpunkt der ersten Nullstelle nach dem Zeitpunkt 0. T muß gerade sein, und es gilt $P(T = 2k) = f_{2k}$. Aus (i) und $u_{2k} \to 0$ folgt

$$\sum_{k=1}^{\infty} f_{2k} = \lim_{N\to\infty} \sum_{k=1}^{N} f_{2k} = \lim_{N\to\infty} \sum_{k=1}^{N}(u_{2k-2} - u_{2k}) = \lim_{N\to\infty}(u_0 - u_{2N}) = 1.$$

Wir sehen also, dass $(f_{2k})_{k\in\mathbb{N}}$ eine Wahrscheinlichkeitsverteilung auf den geraden natürlichen Zahlen definiert, die Verteilung von T. Daraus läßt sich der Erwartungswert von T berechnen:

$$\mathbb{E}T = \sum_{k=1}^{\infty} 2k f_{2k} = \sum_{k=1}^{\infty} u_{2k-2},$$

wobei wir die Gleichung (i) in Lemma 11.11 anwenden. Nach Satz 11.10 divergiert jedoch diese Reihe! Man kann auch sagen, dass $\mathbb{E}T$ gleich ∞ ist. Mit Wahrscheinlichkeit 1 findet also ein Ausgleich statt; man muß jedoch im Schnitt unendlich lange darauf warten.

Obgleich $P(S_1 \neq 0, \ldots, S_{2k} \neq 0) = P(S_1 \geq 0, \ldots, S_{2k} \geq 0) \sim 1/\sqrt{\pi k}$ gegen 0 konvergiert, ist diese Wahrscheinlichkeit erstaunlich groß. Wieso erstaunlich? Wir betrachten das Ereignis $F_j^{(k)}$, dass die Irrfahrt während genau $2j$ Zeiteinheiten bis $2k$ positiv ist. Aus formalen Gründen präzisieren

wir ,,positiv sein“ wie folgt: Die Irrfahrt ist positiv im Zeitintervall von l bis $l+1$, falls S_l *oder* $S_{l+1} > 0$ ist. Es kann also auch $S_l = 0$, $S_{l+1} > 0$ oder $S_l > 0$, $S_{l+1} = 0$ sein. Man überzeugt sich leicht davon, dass die Anzahl der Intervalle, wo dieses der Fall ist, gerade ist. $F_k^{(k)}$ ist natürlich gerade das Ereignis $\{S_1 \geq 0, S_2 \geq 0, \ldots, S_{2k} \geq 0\}$. Aus Gründen der Symmetrie ist $P(F_0^{(k)}) = P(F_k^{(k)})$, was nach Lemma 11.11 (b) gleich $u_{2k} \sim 1/\sqrt{\pi k}$ ist.
Die $F_j^{(k)}$ sind für $0 \leq j \leq k$ paarweise disjunkt, und es gilt

$$\sum_{j=0}^{k} P(F_j^{(k)}) = 1.$$

Mithin können nicht allzuviele der $P(F_j^{(k)})$ von derselben Größenordnung wie $P(F_k^{(k)})$ sein, denn sonst müßte die obige Summe > 1 werden. Anderseits ist wenig plausibel, dass unter diesen Wahrscheinlichkeiten gerade $P(F_k^{(k)})$ und $P(F_0^{(k)})$ besonders groß sind. Genau dies ist jedoch der Fall, wie aus dem folgenden bemerkenswerten Resultat hervorgehen wird.

Satz 11.12 (Satz von Chung und Feller) *Für* $0 \leq j \leq k$ *gilt*

$$P(F_j^{(k)}) = u_{2j} u_{2k-2j}.$$

Beweis: Wir führen einen Induktionsschluß nach k. Für $k = 1$ gilt

$$P(F_0^{(1)}) = P(F_1^{(1)}) = \frac{1}{2} = u_2.$$

Wir nehmen nun an, die Aussage des Satzes sei bewiesen für alle $k \leq n-1$, und beweisen sie für $k = n$.
Wir hatten in Lemma 11.11 (ii), schon gesehen, dass $P(F_0^{(n)}) = P(F_n^{(n)}) = u_{2n}$ ist (u_0 ist $= 1$). Wir brauchen deshalb nur noch $1 \leq j \leq n-1$ zu betrachten. Zunächst führen wir einige spezielle Mengen von Pfaden ein.
Für $1 \leq l \leq n$, $0 \leq m \leq n-l$ sei $G_{l,m}^{+}$ die Menge der Pfade der Länge $2n$ mit: $s_0 = 0$, $s_1 > 0$, $s_2 > 0, \ldots, s_{2l-1} > 0$, $s_{2l} = 0$ und $2m$ Strecken des Pfades zwischen den x-Koordinaten $2l$ und $2n$ sind positiv.
Analog bezeichne $G_{l,m}^{-}$ für $1 \leq l \leq n$, $0 \leq m \leq n-l$, die Menge der Pfade mit: $s_0 = 0$, $s_1 < 0$, $s_2 < 0, \ldots, s_{2l-1} < 0$, $s_{2l} = 0$ und $2m$ Strecken des Pfades zwischen den x-Koordinaten $2l$ und $2n$ sind positiv. Die $G_{l,m}^{+}, G_{l,m}^{-}$ sind offensichtlich alle paarweise disjunkt. Ferner gilt

$$G_{l,m}^{+} \subset F_{l+m}^{(n)},\ G_{l,m}^{-} \subset F_m^{(n)}.$$

Man beachte, dass für $1 \leq j \leq n-1$ jeder Pfad aus $F_j^{(n)}$ zu genau einer der Mengen $G_{l,m}^{+}$, $G_{l,m}^{-}$ gehört. Dies folgt daraus, dass ein solcher Pfad mindestens einmal das Vorzeichen wechseln, also auch die 0 passieren muß. Ist $2l$ die x-Koordinate der kleinsten Nullstelle > 0, so gehört der Pfad zu $G_{l,j-l}^{+}$, falls der

Pfad vor $2l$ positiv, und zu $G^-_{l,j}$, falls er vor $2l$ negativ ist. Demzufolge ist

$$P(F_j^{(n)}) = \sum_{l=1}^{j} P(G^+_{l,j-l}) + \sum_{l=1}^{n-j} P(G^-_{l,j}).$$

Es bleibt noch die Aufgabe, die Summanden auf der rechten Seite dieser Gleichung zu berechnen.

Offensichtlich enthalten $G^+_{l,m}$ und $G^-_{l,m}$ gleich viele Pfade. $|G^+_{l,m}|$ ist gleich der Anzahl der Pfade von $(0,0)$ nach $(2l,0)$ mit $s_1 > 0, s_2 > 0, \ldots, s_{2l-1} > 0$ multipliziert mit der Anzahl der Pfade der Länge $2n-2l$ mit Start in $(2l,0)$ und $2m$ positiven Strecken, das heißt

$$|G^+_{l,m}| = |G^-_{l,m}| = \frac{1}{2} f_{2l} 2^{2l} P(F_m^{(n-l)}) 2^{2n-2l},$$

und

$$P(G^+_{l,m}) = P(G^-_{l,m}) = \frac{1}{2} f_{2l} P(F_m^{(n-l)}).$$

Nach der weiter oben stehenden Gleichung ist also

$$P(F_j^{(n)}) = \frac{1}{2}\sum_{l=1}^{j} f_{2l} P(F_{j-l}^{(n-l)}) + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l} P(F_j^{(n-l)}).$$

Nach Induktionsvoraussetzung ist das

$$= \frac{1}{2}\sum_{l=1}^{j} f_{2l}\, u_{2j-2l}\, u_{2n-2j} + \frac{1}{2}\sum_{l=1}^{n-j} f_{2l}\, u_{2n-2j-2l}\, u_{2j} = u_{2j}\, u_{2n-2j}$$

nach Lemma 11.11,(iii). □

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

$$\frac{P(F_j^{(k)})}{P(F_{j+1}^{(k)})} = \frac{\binom{2j}{j}\binom{2k-2j}{k-j}}{\binom{2j+2}{j+1}\binom{2k-2j-2}{k-j-1}} = \frac{(2k-2j-1)(j+1)}{(2j+1)(k-j)}.$$

Dieser Quotient ist > 1, $= 1$ oder < 1, je nachdem, ob $j < \frac{k-1}{2}$, $j = \frac{k-1}{2}$ oder $j > \frac{k-1}{2}$ ist.

Als Funktion von j fällt also $P(F_j^{(k)})$ für $j < \frac{k-1}{2}$ und steigt an für $j > \frac{k-1}{2}$.

$P(F_0^{(k)}) = P(F_k^{(k)})$ ist also der größte vorkommende Wert und $P(F_{\lceil \frac{k-1}{2} \rceil})$ der kleinste. Es ist bedeutend wahrscheinlicher, dass die Irrfahrt über das ganze betrachtete Zeitintervall positiv ist, als dass sich positive und negative Zahlen ausgleichen. Dies scheint im Widerspruch zum Gesetz der großen Zahlen zu stehen. Ohne dies hier genauer zu diskutieren, sei aber daran erinnert, dass die Rückkehrzeit T nach 0 keinen endlichen Erwartungswert hat, wie wir oben gezeigt haben.

Mit Hilfe von Satz 11.10 läßt sich eine einfach Approximation für $P(F_j^{(k)})$ für große j und $k-j$ gewinnen:

Satz 11.13 *Für $j \to \infty$, $k-j \to \infty$ gilt $P(F_j^{(k)}) \sim \frac{1}{\pi} \frac{1}{\sqrt{j(k-j)}}$, das heißt*

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

Betrachten wir speziell $x \in (0,1)$ so gilt für $j, k \to \infty$ mit $j/k \sim x$

$$P(F_j^{(k)}) \sim \frac{1}{\pi k} \frac{1}{\sqrt{x(1-x)}}.$$

Diese Wahrscheinlichkeiten sind also von der Größenordnung $1/k$, das heißt asymptotisch viel kleiner als

$$P(F_0^{(k)}) = P(F_k^{(k)}) \sim \frac{1}{\sqrt{\pi k}}.$$

Die Funktion $(x(1-x))^{-1/2}$ hat für $x = 0$ und 1 Pole. Das steht in Übereinstimmung damit, dass für $j/k \sim 0$ und $j/k \sim 1$ die Wahrscheinlichkeiten $P(F_j^{(k)})$ von einer anderen Größenordnung als $1/k$ sind.

Eine Aussage wie die in Satz 11.13 nennt man einen *lokalen Grenzwertsatz*, da wir damit Informationen über die Wahrscheinlichkeit, dass der Zeitraum der Führung exakt $= 2j$ ist, erhalten. Da diese Wahrscheinlichkeiten jedoch alle für große k klein werden, interessiert man sich eher zum Beispiel für die Wahrscheinlichkeit, dass der relative Anteil der Zeit, wo die Irrfahrt positiv ist, $\geq \alpha$ ist.

Es seien $0 < \alpha < \beta < 1$. $\gamma_k(\alpha, \beta)$ sei die Wahrscheinlichkeit, dass dieser relative Anteil der Zeit zwischen α und β liegt. Genauer: T_k sei (die auf Ω_{2k} definierte) Zufallsgröße, die die Dauer der Führung zählt:

$$T_k := \sum_{j=1}^{2k} 1_{\{S_{j-1}\geq 0,\, S_j \geq 0\}}.$$

Dann ist

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

Wir sind übrigens nun bei der in Satz 11.5 diskutierten Abbilgung $g(Y_n)$ angekommen, denn $T_k = 2k\, g(Y_{2k})$. Wir wollen nun aus Satz 11.13 für $k \to \infty$ folgern:

$$\gamma_k(\alpha,\beta) \sim \frac{1}{\pi} \sum_{j:\alpha\leq\frac{j}{k}\leq\beta} \frac{1}{k} \frac{1}{\sqrt{\frac{j}{k}\big(1-\frac{j}{k}\big)}}. \tag{11.4}$$

Die rechte Seite ist nichts anderes als die Riemann-Approximation für

$$\int_\alpha^\beta \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi}(\arcsin\sqrt{\beta} - \arcsin\sqrt{\alpha}).$$

Es folgt nun (und damit Satz 11.5):

Satz 11.14 (Arcussinus-Gesetz)

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

Beweis: Wir müssen (11.4) zeigen. Wir schreiben die Stirling-Approximation als $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n F(n)$ mit $\lim_{n\to\infty} F(n) = 1$. Es folgt

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

Wir wählen nun ein $\delta > 0$ mit $0 < \delta < 1/2$ und betrachten für jedes k nur die Werte j für die gilt

$$\delta \le \frac{j}{k} \le 1-\delta,$$

womit $k\delta \le j$ und $k\delta \le k-j$ folgt. Für $k \to \infty$ konvergiert nun jedes $F(j), F(k-j), F(2j), F(2(k-j))$ gleichmäßig für alle obigen Werte von j. Somit existiert für $\delta \le \alpha < \beta \le 1-\delta$ ein $G_{\alpha,\beta}(k)$ für jedes $k = 1, 2, \ldots$, so dass für jedes obige $\delta > 0$ gilt:

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

und

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

Nun folgt die Behauptung gleichmäßig für $\delta \le \alpha < \beta \le 1-\delta$, wie auch immer $0 < \delta < 1/2$ gewählt war. Damit folgt die Behauptung. □

Bemerkung 11.15 Die Aussage von in Satz 11.14 ist auch richtig für $\alpha = 0$ oder $\beta = 1$. Das heißt etwa, dass $\gamma_k(0,\beta)$ — die Wahrscheinlichkeit dafür, dass der relative Anteil der Zeit, in der K_1 führt, $\le \beta$ ist — gegen $\frac{2}{\pi}\arcsin\sqrt{\beta}$ konvergiert.

Beweis: Offensichtlich gilt $\lim_{k\to\infty}\gamma_k\left(0,\frac{1}{2}\right) = 1/2$. Ist $\beta \in (0, 1/2)$, so folgt

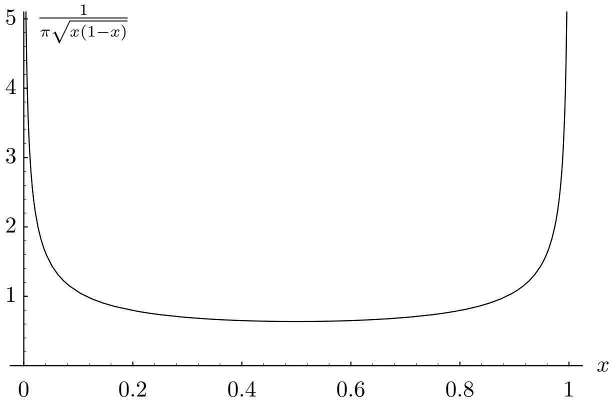
$$\lim_{k\to\infty}\gamma_k(0,\beta) = \lim_{k\to\infty}(\gamma_k(0,1/2) - \gamma_k(\beta,1/2)) = \frac{2}{\pi}\arcsin\sqrt{\beta},$$

für $\beta > 1/2$

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

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

Der Beweis des Arcus-Sinus-Gesetzes wurde zuerst von P. Lévy im Jahre 1939 gegeben. Die Funktion $\frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}}$ hat das folgende Aussehen:



Zur Illustration des Arcus-Sinus-Gesetzes diene die folgende Tabelle der sogenannten Arcus-Sinus-Verteilungsfunktion $A(x) = \frac{2}{\pi}\arcsin\sqrt{x}$. Für $x \in (\frac{1}{2}, 1]$ kann $A(x)$ mit der Formel $A(x) = 1 - A(1-x)$ berechnet werden.

$$A(x) = \frac{2}{\pi}\arcsin\sqrt{x}$$

x	$A(x)$	x	$A(x)$	x	$A(x)$
0.00	0.000	0.20	0.295	0.40	0.436
0.01	0.064	0.21	0.303	0.41	0.442
0.02	0.090	0.22	0.311	0.42	0.449
0.03	0.111	0.23	0.318	0.43	0.455
0.04	0.128	0.24	0.326	0.44	0.462
0.05	0.144	0.25	0.333	0.45	0.468
0.06	0.158	0.26	0.341	0.46	0.474
0.07	0.171	0.27	0.348	0.47	0.481
0.08	0.183	0.28	0.355	0.48	0.487
0.09	0.194	0.29	0.362	0.49	0.494
0.10	0.205	0.30	0.369	0.50	0.500
0.11	0.215	0.31	0.376		
0.12	0.225	0.32	0.383		
0.13	0.235	0.33	0.390		
0.14	0.244	0.34	0.396		
0.15	0.253	0.35	0.403		
0.16	0.262	0.36	0.410		
0.17	0.271	0.37	0.416		
0.18	0.279	0.38	0.423		
0.19	0.287	0.39	0.429		

ANHANG A

Beweis des Satzes von Prohorov

Inspiziert man den Beweis von Satz 9.20 (Cramér-Wold) und den von Satz 10.5 und 10.7 (Satz von Donsker), so haben wir nur die folgende einfachere Version des Satzes von Prohorov verwendet:

Ist die Folge $(\mu_n)_n$ von W-Maßen auf einem vollständigen, separablen metrischen Raum straff, so hat $(\mu_n)_n$ eine schwach konvergente Teilfolge, ist also insbesondere relativ kompakt. Ist die Folge $(\mu_n)_n$ schwach konvergent, so ist sie auch straff (Satz 9.20). Wir beweisen daher auch nur diese Variante. Für einen Beweis von Satz 9.18 in seiner vollen Allgemeinheit verweisen wir auf Billingsley, „Convergence of Probablitiy measures“, oder Parthasarathy, „Probabilty measures on metric spaces“.

Satz A.1 *Es sei S ein separabler metrischer Raum und $(\mu_n)_n$ eine Folge von W-Maßen auf $(S, \mathcal{B}_S)$, die straff ist. Dann hat $(\mu_n)_n$ eine schwach konvergente Teilfolge. Ist S ein vollständiger, separabler metrischer Raum und konvergiert $(\mu_n)_n$ schwach, dann ist $\{\mu_n,\, n \in \mathbb{N}\}$ straff.*

Alle Beweise in der Literatur verwenden tiefere Resultate der Analysis, so auch der hier vorgestellte Beweis.

Lemma A.2 *Es sei S ein kompakter metrischer Raum und $(\mu_n)_n$ eine Folge von W-Maßen auf $(S, \mathcal{B}_S)$. Dann hat $(\mu_n)_n$ eine schwach konvergente Teilfolge.*

Der Beweis des Lemmas verwendet die folgende Variante des Darstellungssatzes von Riesz:

Eine Abbildung $\Lambda : C(S) \to \mathbb{R}$ heißt ein *normiertes, nicht-negatives lineares Funktional*, wenn $\Lambda(1) = 1$, $\Lambda(f) \geq 0$ für $f \geq 0$ und $\Lambda(af+bg) = a\Lambda(f)+b\Lambda(g)$ für alle $a, b \in \mathbb{R}$ und $f, g \in C(S)$ gilt.

Satz A.3 (Darstellungssatz von Riesz) *Es sei S ein kompakter metrischer Raum. Dann existiert zu jedem normierten, nicht-negativem linearem Funktional $\Lambda : C(S) \to \mathbb{R}$ ein eindeutig bestimmtes W-Maß μ auf $(S, \mathcal{B}_S)$ mit*

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

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

Beweis: Z.B. in PARTHASARATHY, „Introduction to Probability and Measure“.

Beweis von Lemma A.2: Für $f \in C(S)$ sei

$$||f|| := \sup_{x \in S} |f(x)| \, .$$

Da S kompakt ist, ist $C(S)$ ein separabler metrischer Raum; dies folgt aus dem Satz von WEIERSTRASS. Sei $(f_n)_n$ eine dichte Folge in $C(S)$. Mit Hilfe des Diagonalfolgenverfahrens, beschrieben im Anhang zu Kapitel 10, finden wir eine Teilfolge $(\mu_{n_k})_k$ von $(\mu_n)_n$, so dass

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

für alle $j \in \mathbb{N}$ existiert. Zu einem $f \in C(S)$ und $\varepsilon > 0$ sei f_j so gewählt, dass $||f - f_j|| < \varepsilon$. Dann ist

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

Der erste Summand konvergiert gegen Null für $k, m \to \infty$, also

$$\lim_{k,m \to \infty} \Big| \int f \, d\mu_{n_k} - \int f \, d\mu_{n_m} \Big| = 0 \, ,$$

und somit konvergiert $\int f \, d\mu_{n_k}$ für $k \to \infty$ für jedes $f \in C(S)$. Setzen wir

$$\Lambda(f) := \lim_{k \to \infty} \int f \, d\mu_{n_k} \, , \quad f \in C(S) \, ,$$

so ist Λ ein nicht-negatives lineares Funktional auf $C(S)$ mit $\Lambda(1) = 1$, also existiert nach Satz A.3 ein $\mu \in \mathcal{M}_1(S)$ mit

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

womit die schwache Konvergenz von $(\mu_{n_k})_k$ gegen μ folgt. □

Für den Beweis von Satz A.1 bereiten wir vor:

Lemma A.4 (URYSOHN) *Ist S ein separabler metrischer Raum, so ist er homöomorph zu einer Teilmenge in* $[0, 1]^{\mathbb{N}}$.

Beweis: d bezeichne die Metrik auf S und $(s_n)_n$ eine dichte, abzählbare Teilmenge von S. $h : S \to [0, 1]^{\mathbb{N}}$ sei definiert durch die n-ten Koordinatenfunktionen

$$h_n(x) = \frac{d(x, \, s_n)}{1 + d(x, \, s_n)} \, , \quad x \in S \, , \, n \in \mathbb{N} \, .$$

Es ist eine schöne Übung zu sehen, dass dies ein Homöomorphismus ist. □

Nun ist $[0,1]$ kompakt. Tatsächlich ist auch $[0,1]^{\mathbb{N}}$ kompakt. Ist (K, d) ein kompakter metrischer Raum, so ist die Metrik d offenbar beschränkt:

$$\sup_{x,y\in K} d(x, y) < \infty\,.$$

Auf $K^{\mathbb{N}}$ definieren wir

$$\bar{d}(x, y) := \sum_{i=1}^{\infty} \frac{d(x_i, y_i)}{2^i}$$

für $x = (x_i)_i$ und $y = (y_i)_i$. Dann ist $\bar{d}$ eine Metrik und eine Folge in $(K^{\mathbb{N}}, \bar{d})$ konvergiert genau dann, wenn alle ihre Komponenten konvergieren. Es gilt

Satz A.5 (Tychonov) *$(K^{\mathbb{N}}, \bar{d})$ ist kompakt.*

Beweisskizze: (Diagonalfolgen-Verfahren) Sei $(x^n)_n = \big((x_i^n)_i\big)_n$ eine Folge in $K^{\mathbb{N}}$. Wir wählen eine Teilfolge $(x^{n_{1,m}})_m$, so dass $(x_1^{n_{1,m}})_m$ konvergiert, dann eine Teilfolge $(x^{n_{2,m}})_m$ dieser Folge, so dass $(x_2^{n_{2,m}})_m$ konvergiert, etc. Dann konvergiert $(x^{n_{m,m}})_m$. □

Also ist nach Lemma A.4 und Satz A.5 ein separabler metrischer Raum homöomorph zu einer Teilmenge eines kompakten metrischen Raumes. Wir gehen zu dieser Teilmenge über und wählen die Relativtopologie (eine Menge O ist offen, wenn $O = [0,1]^{\mathbb{N}} \cap O'$ mit O' offen in $[0,1]^{\mathbb{N}}$ gilt).

Beweis von Satz A.1: Wir fassen den separablen metrischen Raum entsprechend der Vorbetrachtung als Teilmenge eines kompakten metrischen Raumes $\tilde{S}$ auf. Für $\mu \in \mathcal{M}_1(S)$ definieren wir $\tilde{\mu} \in \mathcal{M}_1(\tilde{S})$ durch

$$\tilde{\mu}(A) := \mu(A \cap S)\,, \quad A \in \mathcal{B}_{\tilde{S}}\,.$$

Mit Lemma A.2 hat $(\tilde{\mu}_n)_n$ eine konvergente Teilfolge $(\tilde{\mu}_{n_k})_k$, die schwach gegen ein W-Maß 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 Satz von Portmanteau gilt

$$\limsup_{n\to\infty} \tilde{\mu}_{n_k}(K_r) \leq \nu(K_r)\,, \quad r \in \mathbb{N}\,.$$

Dann folgt auch $\nu(K_r) \geq 1 - 1/r$ für $r \in \mathbb{N}$.

Sei $E_0 := \bigcup_r K_r$, dann ist $E_0 \subset S$, $E_0 \in \mathcal{B}_{\tilde{S}}$ und $\nu(E_0) = 1$. Wir behaupten nun, dass es ein $\mu \in \mathcal{M}_1(S)$ gibt mit $\tilde{\mu} = \nu$.

Es ist $\mathcal{B}_S = \mathcal{B}_{\tilde{S}} \cap S$. Für jedes $A \in \mathcal{B}_S$ existiert ein $B_1 \in \mathcal{B}_{\tilde{S}}$ mit $A = B_1 \cap S$. Sei $\mu(A) := \nu(B_1)$. Wenn $B_2 \in \mathcal{B}_{\tilde{S}}$ und $A = B_2 \cap S$, dann ist $B_1 \Delta B_2 \subset S^c \subset E_0^c$ und $\nu(B_1 \Delta B_2) = 0$, also $\nu(B_1) = \nu(B_2)$, also ist $\mu(A)$ wohldefiniert.

Es sei nun $(A_i)_i$ mit $A_i = B_i \cap S$, $i \in \mathbb{N}$, eine Folge von disjunkten Mengen mit $B_i \in \mathcal{B}_{\tilde{S}}$, $i \in \mathbb{N}$. Da $B_i \cap E_0 \subset B_i \cap S$ für alle i, sind die $B_i \cap E_0$ auch disjunkt. Also

$$\begin{aligned}\mu\Big(\bigcup_i A_i\Big) &= \nu\Big(\bigcup_i B_i\Big) = \nu\Big(\bigcup_i (B_i \cap E_0)\Big) \\ &= \sum_i \nu(B_i \cap E_0) = \sum_i \nu(B_i) = \sum_i \mu(A_i)\,.\end{aligned}$$

Also ist μ ein W-Maß mit $\tilde{\mu} = \nu$.

Sei C eine abgeschlossene Menge in S. Dann existiert ein D abgeschlossen in $\tilde{S}$ mit $C = D \cap S$. Da $\tilde{\mu}_{n_k} \xrightarrow{w} \tilde{\mu}$, folgt

$$\limsup_{k\to\infty} \mu_{n_k}(C) = \limsup_{k\to\infty} \tilde{\mu}_{n_k}(D) \le \tilde{\mu}(D) = \mu(C)\,.$$

Der Satz von Portmanteau liefert $\mu_{n_k} \xrightarrow{w} \mu$. Damit ist der erste Teil des Satzes bewiesen.

Sei nun S vollständig und separabel und $\mu_n \xrightarrow{w} \mu$. Da S separabel ist, existiert eine Folge offener Bälle $B_{n1}, B_{n2}, \dots$ mit Radius $1/n$, so dass

$$S = \bigcup_{j=1}^{\infty} B_{nj}\,, \quad n \in \mathbb{N}\,.$$

Wir zeigen nun, dass für jedes $\delta > 0$ ein $k_n \in \mathbb{N}$ existiert mit

$$\mu_i\Big(\bigcup_{j=1}^{k_n} B_{nj}\Big) > 1 - \delta\,, \quad i \in \mathbb{N}\,.$$

Angenommen, dies stimmt nicht. Also existiert ein $\delta_0 > 0$ und Folgen $i_1 < i_2 < \cdots$ und $k_1 < k_2 < \cdots$ mit

$$\mu_{i_m}\Big(\bigcup_{j=1}^{k_m} B_{nj}\Big) \le 1 - \delta_0 \quad \text{für } m = 1, 2, \dots$$

Es gilt $\bigcup_{j=1}^{k_r} B_{nj} \subset \bigcup_{j=1}^{k_m} B_{nj}$ für $m \ge r$, also

$$\mu_{i_m}\Big(\bigcup_{j=1}^{k_r} B_{nj}\Big) \le \mu_{i_m}\Big(\bigcup_{j=1}^{k_m} B_{nj}\Big) \le 1 - \delta_0$$

für $m \ge r$.

Da $\mu_{i_m} \xrightarrow{w} \mu$ und $\bigcup_{j=1}^{k_r} B_{nj}$ offen, sagt der Satz von Portmanteau

$$\mu\Big(\bigcup_{j=1}^{k_r} B_{nj}\Big) \le \liminf_{m\to\infty} \mu_{i_m}\Big(\bigcup_{j=1}^{k_r} B_{nj}\Big) \le 1 - \delta_0\,.$$

Für $r \to \infty$ folgt $\mu(S) \le 1 - \delta_0$. Ein Widerspruch!

Sei nun $n \in \mathbb{N}$ fest und $\delta = \varepsilon/2^n$ und k_n so gewählt, dass

$$\mu_i\Big(\bigcup_{j=1}^{k_n} B_{nj}\Big) > 1 - \frac{\varepsilon}{2^n}, \quad i \in \mathbb{N}.$$

Sei $C_n := \bigcup_{j=1}^{k_n} \bar{B}_{nj}$ und $K := \bigcap_{n=1}^{\infty} C_n$. Dann folgt $\mu_i(K) > 1 - \varepsilon$ für alle $i \in \mathbb{N}$. Tatsächlich ist K kompakt:

Da die C_n abgeschlossen sind, ist auch K abgeschlossen. $(x_n)_n$ sei eine Folge in K. Da $K \subset C_1$, existiert ein $n_1 \le k_1$, so dass $K \cap \bar{B}_{1n_1} =: K_1$ unendlich viele der x_i enthält. Da $K_1 \subset C_2$, existiert ein $n_2 \le k_2$, so dass $K_1 \cap \bar{B}_{2n_2} =: K_2$ unendlich viele der x_i enthält. Wir gelangen so zu einer Kette $K_1 \supset K_2 \supset \dots$, und jedes K_j enthält unendlich viele der x_i. Nun ist $K_j \subset \bar{B}_{jn_j}$, also ist der Durchmesser von K_j kleiner-gleich $2/j$, $j \in \mathbb{N}$. Nun liefert die Vollständigkeit von S

$$\bigcap_{j=1}^{\infty} K_j = \{x_0\}, \quad x_0 \in S.$$

Nun enthält ein Ball um x_0 ein K_j für j hinreichend groß, also enthält der Ball unendlich viele der x_i. x_0 ist also Limespunkt der Folge $(x_n)_n$, also ist K kompakt und er Satz ist bewiesen. $\square$

Literaturverzeichnis

[1] Heinz Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, fifth edition, 2002. ISBN 3-11-017236-4.

[2] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. ISBN 0-471-00710-2. A Wiley-Interscience Publication.

[3] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

[4] Leo Breiman. *Probability*. Addison-Wesley Publishing Company, Reading, Mass., 1968.

[5] Kai Lai Chung. *A course in probability theory*. Academic Press Inc., San Diego, CA, third edition, 2001. ISBN 0-12-174151-6.

[6] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 1998.

[7] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. ISBN 0-521-00754-2. Revised reprint of the 1989 original.

[8] Richard Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996. ISBN 0-534-24318-5.

[9] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[10] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[11] H.-O. Georgii. *Stochastik*. Walter de Gruyter, Berlin, 2002.

[12] Peter Gänssler and Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin, 1977. ISBN 3-540-08418-5.

[13] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[14] K. R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press Inc., New York, 1967.

[15] Kalyanapuram Rangachari Parthasarathy. *Introduction to probability and measure*. Springer-Verlag New York Inc., New York, 1978. ISBN 0-387-91135-9.

[16] S. R. S. Varadhan. *Probability theory*, volume 7 of *Courant Lecture Notes in Mathematics*. New York University Courant Institute of Mathematical Sciences, New York, 2001. ISBN 0-8218-2852-2

Index