Skript zur Vorlesung Wahrscheinlichkeitstheorie

E. Bolthausen

Inhaltsverzeichnis

Notationen

Ist Ω eine Menge, so bezeichnet $\mathcal{P}(\Omega)$ die Potenzmenge von Ω, das heißt die Menge der Teilmengen von Ω. Teilmengen von $\mathcal{P}(\Omega)$ werden auch als Klassen oder Familien von Teilmengen bezeichnet beziehungsweise Mengensysteme genannt und meist mit $\mathcal{A}$, $\mathcal{B}$, $\mathcal{F}$, $\mathcal{R}$ etc. bezeichnet.

Sind $A, B \subset \Omega$, so bezeichnen $A^c = \{\, x \in \Omega \mid x \notin A \,\}$ das Komplement von A in Ω und $A \setminus B = \{\, x \in A \mid x \notin B \,\}$ die Mengendifferenz. Die symmetrische Differenz ist definiert durch $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

Sind $\{A_n\}_{n\in\mathbb{N}}$ eine Folge von Mengen und A eine weitere Menge, so bedeutet $A_n \uparrow A$, daß $A_{n+1} \supset A_n$ für alle $n \in \mathbb{N}$ und $A = \bigcup_{n=1}^{\infty} A_n$ gelten. Entsprechend bedeutet $A_n \downarrow A$, daß $A_{n+1} \subset A_n$ für alle $n \in \mathbb{N}$ und $A = \bigcap_{n=1}^{\infty} A_n$ gelten. Man schreibt statt dessen auch $A = \lim \uparrow_{n\to\infty} A_n$ beziehungsweise $A = \lim \downarrow_{n\to\infty} A_n$.

Es sei $\{A_i\}_{i\in I}$ eine Familie von Mengen. Die Mengen A_i heißen paarweise disjunkt, falls $A_i \cap A_j = \emptyset$ für alle $i, j \in I$ mit $i \neq j$ gilt.

Ist $\{A_n\}_{n\in\mathbb{N}}$ eine beliebige Folge von Mengen, so sind

$$\overline{\lim_{n\to\infty}} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \lim_{n\to\infty}\downarrow \bigcup_{m=n}^{\infty} A_m,$$

$$\underline{\lim_{n\to\infty}} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \lim_{n\to\infty}\uparrow \bigcap_{m=n}^{\infty} A_m.$$

Es gelten offenbar:

$$x \in \overline{\lim_{n\to\infty}} A_n \iff \text{Es gilt } x \in A_n \text{ für unendlich viele } n \in \mathbb{N}.$$

$$x \in \underline{\lim_{n\to\infty}} A_n \iff \text{Es gilt } x \notin A_n \text{ für höchstens endlich viele } n \in \mathbb{N}.$$

Ist $f\colon A \to B$ eine Abbildung, so seien $f(C) := \{\, f(x)\colon x \in C \,\}$ für $C \subset A$ und $f^{-1}(D) := \{\, x \in A\colon f(x) \in D \,\}$ für $D \subset B$. Man beachte, daß $f^{-1}(C^c) = f^{-1}(C)^c$ sowie $f^{-1}\left(\bigcup_{i\in I} C_i\right) = \bigcup_{i\in I} f^{-1}(C_i)$ und $f^{-1}\left(\bigcap_{i\in I} C_i\right) = \bigcap_{i\in I} f^{-1}(C_i)$ gelten.

Ist $\mathcal{A}$ eine Familie von Teilmengen von B, so bezeichnet $f^{-1}(\mathcal{A})$ (etwas mißbräuchlich) die Familie $\{\, f^{-1}(C)\colon C \in \mathcal{A} \,\}$ von Teilmengen von A.

Sind $f, g\colon A \to B$ zwei Abbildungen, so sind unter anderem folgende Schreibweisen gebräuchlich:

$$\begin{aligned}
\{f \in C\} &= f^{-1}(C),\\
\{f = g\} &= \{\, x \in A\colon f(x) = g(x) \,\},\\
\{f \neq g\} &= \{\, x \in A\colon f(x) \neq g(x) \,\}.
\end{aligned}$$

Wenn B eine Teilmenge der reellen Zahlen ist, kann

$$\{f \leq t\} = \{\, x \in A \colon f(x) \leq t \,\}$$

für $t \in \mathbb{R}$ geschrieben werden.

Es bezeichnen $\mathbb{R}$ die Menge der reellen Zahlen, $\mathbb{N}$ die Menge der natürlichen Zahlen, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ und $\mathbb{R}_+ := \{\, x \in \mathbb{R} \colon x \geq 0 \,\}$.

Zu $\mathbb{R}$ können zwei Punkte ∞ und $-\infty$ adjungiert werden: Dazu wird $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ definiert. Auf $\overline{\mathbb{R}}$ wird in „natürlicher Weise" addiert und multipliziert, bis auf die Konventionen $0 \cdot \infty = 0 \cdot (-\infty) = 0$. Ausdrücke der Art $\infty - \infty$ oder $\infty + (-\infty)$ sind *nicht* definiert. Für jede Folge $\{x_n\}_{n \in \mathbb{N}}$ in $\overline{\mathbb{R}}$ existieren $\overline{\lim}_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{m \geq n} x_m$ und $\underline{\lim}_{n\to\infty} x_n = \lim_{n\to\infty} \inf_{m \geq n} x_m$ in $\overline{\mathbb{R}}$.

In jedem Kapitel werden Formeln, Sätze und Definitionen fortlaufend numeriert. Das Ende eines Beweises wird mit $\square$ bezeichnet.

Kapitel 1

Maßtheoretische Grundlagen

Nachfolgend wird eine Zusammenstellung der für die Wahrscheinlichkeitstheorie wichtigsten Begriffe und Sätze aus der Maß- und Integrationstheorie (im allgemeinen ohne Beweise) gegeben. Für ausführliche Beweise und Darstellungen sind die folgenden Bücher empfehlenswert:

- H. Bauer: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, de Gruyter 1978 (3. neubearbeitete Auflage)
- H. Bauer: Maß- und Integrationstheorie, de Gruyter 1990
- D. L. Cohn: Measure Theory, Birkhäuser 1980
- P. R. Halmos: Measure Theory, Springer 1974 (Reprint)

Verweise in diesem Kapitel beziehen sich auf das erstgenannte Buch von Bauer.

1a. Maße und Erweiterungen

Definition 1.1. Sei Ω eine Menge. Eine nichtleere Familie $\mathcal{F}$ von Teilmengen von Ω heißt *Algebra,* falls $A, B \in \mathcal{F} \Rightarrow A^c,\ A \cap B,\ A \cup B \in \mathcal{F}$ gilt. Eine Algebra heißt *σ-Algebra,* wenn zusätzlich folgendes gilt:

$$A_n \in \mathcal{F} \text{ für alle } n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Jede Algebra enthält $\emptyset$ und Ω, weil $\emptyset = A \cap A^c$ für $A \in \mathcal{F}$ und $\Omega = \emptyset^c$ gelten.

Bemerkung 1.2. Zu jedem Mengensystem $\mathcal{C}$ in Ω gibt es eine kleinste Algebra $a(\mathcal{C})$ und eine kleinste σ-Algebra $\sigma(\mathcal{C})$, die $\mathcal{C}$ enthalten. Der Beweis ist einfach: Die kleinste Algebra, die $\mathcal{C}$ enthält, ist der Durchschnitt aller Algebren, die $\mathcal{C}$ enthalten (mindestens eine Algebra, nämlich $\mathcal{P}(\Omega)$, enthält $\mathcal{C}$). Analog wird $\sigma(\mathcal{C})$ „konstruiert". Das Mengensystem $\mathcal{C}$ heißt ein *Erzeugendensystem* der Algebra $a(\mathcal{C})$ beziehungsweise der σ-Algebra $\sigma(\mathcal{C})$. In der Regel hat eine σ-Algebra sehr viele Erzeugendensysteme. Von besonderer Bedeutung in der Wahrscheinlichkeitstheorie sind durchschnittstabile Erzeugendensysteme. (Eine Familie $\mathcal{A}$ von Teilmengen von Ω heißt *durchschnittstabil,* falls $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$.)

Definition 1.3. Eine Familie $\mathcal{D}$ von Teilmengen von Ω heißt *Dynkin-System,* falls die folgenden Bedingungen erfüllt sind:

(a) $\Omega \in \mathcal{D}$

(b) $D \in \mathcal{D} \Rightarrow D^c \in \mathcal{D}$
(c) Für jede Folge $\{D_n\}_{n\in\mathbb{N}}$ von *paarweise disjunkten* Mengen aus $\mathcal{D}$ ist $\bigcup_{n=1}^{\infty} D_n$ ebenfalls in $\mathcal{D}$.

Ist $\mathcal{C} \subset \mathcal{P}(\Omega)$, so gibt es analog wie bei den Algebren und σ-Algebren ein kleinstes Dynkin-System $d(\mathcal{C})$, das $\mathcal{C}$ enthält.

Satz 1.4. *Ist $\mathcal{C} \subset \mathcal{P}(\Omega)$ durchschnittstabil, so gilt $d(\mathcal{C}) = \sigma(\mathcal{C})$.*

Beweis. Siehe Bauer, Satz 2.4.

Ist $\{\mathcal{A}_i\}_{i\in I}$ eine beliebige Familie von σ-Algebren auf derselben Menge Ω, so ist $\bigcap_{i\in I} \mathcal{A}_i$ wieder eine σ-Algebra, $\bigcup_{i\in I} \mathcal{A}_i$ jedoch im allgemeinen nicht. Mit $\bigvee_{i\in I} \mathcal{A}_i$ wird die σ-Algebra $\sigma\left(\bigcup_{i\in I} \mathcal{A}_i\right)$ bezeichnet. (Im Fall $I = \{1, \ldots, n\}$ ist auch die Schreibweise $\mathcal{A}_1 \vee \cdots \vee \mathcal{A}_n$ gebräuchlich.)

Definition 1.5.

(a) Ein *Inhalt* μ auf einer Algebra $\mathcal{F}$ ist eine Abbildung $\mathcal{F} \to [0, \infty]$ mit den Eigenschaften $\mu(\emptyset) = 0$ und $\mu(A \cup B) = \mu(A) + \mu(B)$ für alle $A, B \in \mathcal{F}$ mit $A \cap B = \emptyset$.
(b) Ein Inhalt μ heißt *σ-endlich,* falls eine Folge $\{A_n\}_{n\in\mathbb{N}}$ von Mengen aus $\mathcal{F}$ existiert, für die $\Omega = \bigcup_{n=1}^{\infty} A_n$ und $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$ gelten.
(c) Ein Inhalt μ heißt *σ-additiv,* falls für jede Folge $\{A_n\}_{n\in\mathbb{N}}$ von paarweise disjunkten Mengen aus $\mathcal{F}$, für die $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ gilt, $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$ erfüllt ist. (Ein σ-additiver Inhalt heißt auch *Prämaß.*)
(d) Ein σ-additiver Inhalt, der auf einer σ-Algebra definiert ist, heißt *Maß.*

Jeder Inhalt ist *monoton,* das heißt $A \supset B \Rightarrow \mu(A) \geq \mu(B)$. Es werden im folgenden nur σ-endliche Inhalte und Maße betrachtet; dies wird stets stillschweigend vorausgesetzt.

Satz 1.6. *Es seien μ ein Maß auf einer σ-Algebra $\mathcal{F}$ und $\{A_n\}_{n\in\mathbb{N}}$ eine Folge von Mengen aus $\mathcal{F}$. Sei $A \in \mathcal{F}$ eine weitere Menge. Dann gelten:*

(a) *$A_n \downarrow A$ und $\exists n_0 \in \mathbb{N}$ mit $\mu(A_{n_0}) < \infty \Rightarrow \mu(A) = \lim_{n\to\infty} \mu(A_n)$.*
(b) *$A_n \uparrow A \Rightarrow \mu(A) = \lim_{n\to\infty} \mu(A_n)$.*
(c) *$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$.*

Bemerkung 1.7. Ein endlicher Inhalt (das heißt $\mu(\Omega) < \infty$) ist genau dann σ-additiv, wenn für jede Folge $\{A_n\}_{n\in\mathbb{N}}$ aus $\mathcal{F}$ die Implikation $A_n \downarrow \emptyset \Rightarrow \mu(A_n) \downarrow 0$ gilt. (Man sagt auch, daß μ in $\emptyset$ stetig ist.)

Satz 1.8 (ERWEITERUNGSSATZ VON CARATHÉODORY)**.** *Es sei μ_0 ein σ-additiver Inhalt (stets σ-endlich!) auf einer Algebra $\mathcal{A}$. Dann gibt es genau ein Maß μ auf $\sigma(\mathcal{A})$, das μ_0 erweitert, das heißt, das auf $\mathcal{A}$ mit μ_0 übereinstimmt.*

Ein Beweis findet sich in jedem Buch über Maßtheorie. Schwierig ist der Beweis der Existenz. Die Eindeutigkeit folgt aus Satz 1.4:

Bemerkung 1.9. Stimmen zwei Maße μ und ν, die auf einer σ-Algebra $\mathcal{F}$ definiert sind, auf einem durchschnittstabilen Erzeugendensystem $\mathcal{D}$ von $\mathcal{F}$ überein, und gilt $\mu(\Omega_n) = \nu(\Omega_n) < \infty$ für eine Folge $\Omega_n \in \mathcal{D}$ mit $\Omega_n \uparrow \Omega$, so gilt $\mu = \nu$ auf $\mathcal{F}$.

Beweis. (Für den Spezialfall $\mu(\Omega) = \nu(\Omega) < \infty$, das heißt insbesondere für Wahrscheinlichkeitsmaße.) Sei $\mathcal{F}' = \{A \in \mathcal{F} \colon \mu(A) = \nu(A)\}$. Dann ist $\mathcal{D} \subset \mathcal{F}'$, und $\mathcal{F}'$ ist ein Dynkin-System (siehe Definition 1.3), da

(a) $\Omega \in \mathcal{F}'$.
(b) $D \in \mathcal{F}' \Rightarrow \mu(D^c) = \mu(\Omega \setminus D) = \mu(\Omega) - \mu(D) = \nu(\Omega) - \nu(D) = \nu(D^c) \Rightarrow D^c \in \mathcal{F}'$.
(c) Für jede Folge $\{D_n\}_{n\in\mathbb{N}}$ von paarweise disjunkten Mengen aus $\mathcal{F}'$ gilt

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

woraus $\bigcup_{n=1}^{\infty} D_n \in \mathcal{F}'$ folgt.

Somit folgt $\mathcal{F}' \supset d(\mathcal{D}) = \sigma(\mathcal{D}) = \mathcal{F}$ aus Satz 1.4. □

Definition 1.10.

(a) Ist $\mathcal{F}$ eine σ-Algebra auf einer Menge Ω, so heißt $(\Omega, \mathcal{F})$ *meßbarer Raum* und die Mengen in $\mathcal{F}$ heißen *meßbar.*
(b) Ist $(\Omega, \mathcal{F})$ ein meßbarer Raum und μ ein Maß auf $\mathcal{F}$, so heißt $(\Omega, \mathcal{F}, \mu)$ ein *Maßraum.* Ist $\mu(\Omega) < \infty$, so heißt der Maßraum *endlich.* Gilt $\mu(\Omega) = 1$, so spricht man von einem *Wahrscheinlichkeitsraum.*

1b. Beispiele von meßbaren Räumen

Beispiel 1.11. Es seien $\Omega = \mathbb{R}$ und $\mathcal{B}$ die kleinste σ-Algebra, die alle Intervalle enthält, das heißt $\mathcal{B} = \sigma(\mathcal{J})$, wobei $\mathcal{J}$ die Familie der Intervalle in $\mathbb{R}$ ist. Die σ-Algebra $\mathcal{B}$ heißt die *Borel-σ-Algebra,* und die Mengen in $\mathcal{B}$ heißen *Borel-Mengen.*

Bemerkung 1.12. Es gilt auch $\mathcal{B} = \sigma(\mathcal{J}_0)$ mit $\mathcal{J}_0 = \{\, (-\infty, t] \colon t \in \mathbb{R} \,\}$.

Beweis. Da $\mathcal{J}_0 \subset \mathcal{J}$, folgt $\sigma(\mathcal{J}_0) \subset \sigma(\mathcal{J}) = \mathcal{B}$. Andererseits zeigt man, daß $\mathcal{J} \subset \sigma(\mathcal{J}_0)$ gilt: Für jedes $t \in \mathbb{R}$ ist $(-\infty, t) = \bigcup_{n=1}^{\infty} (-\infty, t - 1/n] \in \sigma(\mathcal{J}_0)$. Somit sind für $t \in \mathbb{R}$ auch $[t, \infty) = (-\infty, t)^c \in \sigma(\mathcal{J}_0)$ und $(t, \infty) = (-\infty, t]^c \in \sigma(\mathcal{J}_0)$. Jedes Intervall in $\mathbb{R}$ läßt sich als Durchschnitt solcher einseitig unbeschränkten Intervalle darstellen (bis auf $\mathbb{R}$, aber $\mathbb{R} \in \sigma(\mathcal{J}_0)$ gilt sowieso). Aus $\mathcal{J} \subset \sigma(\mathcal{J}_0)$ folgt $\sigma(\mathcal{J}) \subset \sigma(\sigma(\mathcal{J}_0)) = \sigma(\mathcal{J}_0)$. □

Beispiel 1.13. Es seien $\Omega = \mathbb{R}^n$ und $\mathcal{B}^n$ die kleinste σ-Algebra, die alle n-dimensionalen Rechtecke (Produktmengen von Intervallen) enthält.

Bemerkung 1.14.

(a) $\mathcal{B}^n = \sigma(\mathcal{O})$, wobei $\mathcal{O}$ die Familie der offenen Mengen in $\mathbb{R}^n$ ist.
(b) $\mathcal{B}^n = \sigma(\mathcal{J}_0)$ mit $\mathcal{J}_0 = \{\, (-\infty, t_1] \times \cdots \times (-\infty, t_n] \colon t_1, \ldots, t_n \in \mathbb{R} \,\}$.

Beweis. Als Übungsaufgabe.

Beispiel 1.15. Sei I eine beliebige (Index)-Menge. Dann ist $\Omega = \mathbb{R}^I$ die Menge der Abbildungen $I \to \mathbb{R}$. Sei $\mathcal{B}_I$ die kleinste σ-Algebra, die alle Mengen der Form $\{ x \in \mathbb{R}^I : x(t_1) \in \mathcal{J}_1, \dots, x(t_n) \in \mathcal{J}_n \}$ enthält, wobei $n \in \mathbb{N}$, $t_1, \dots, t_n \in I$ und $\mathcal{J}_1, \dots, \mathcal{J}_n$ Intervalle in $\mathbb{R}$ sind.

1c. Beispiele von Maßen

Beispiel 1.16 (Lebesgue-Mass auf $\mathbb{R}$)**.** Sei $\mathcal{B}_0$ die Familie der endlichen Vereinigungen von Intervallen in $\mathbb{R}$. Für eine derartige Menge $A \in \mathcal{B}_0$ sei $\mu_0(A)$ die „normale" Länge von A. Dann ist μ_0 ein Inhalt (klar!), der σ-additiv ist (weniger klar; Beweis in Bauer). Nach Satz 1.8 existiert genau ein Maß λ auf $\mathcal{B}$, das μ_0 erweitert, dieses heißt *Lebesgue-Maß* auf $(\mathbb{R}, \mathcal{B})$.

Beispiel 1.17 (Lebesgue-Mass auf $\mathbb{R}^n$)**.** Analog definiert man das n-dimensionale Lebesgue-Maß λ_n auf $(\mathbb{R}^n, \mathcal{B}^n)$.

Beispiel 1.18 (Diskrete Masse)**.** Es seien Ω beliebig, $\{x_n\}_{n\in\mathbb{N}}$ eine Folge von verschiedenen Punkten in Ω und $a_n \in \mathbb{R}_+$ für alle $n \in \mathbb{N}$. Für jede σ-Algebra $\mathcal{A}$ ist $\mu = \sum_{n=1}^{\infty} a_n \delta_{x_n}$ wie folgt definiert:

$$\mu(A) = \sum_{n=1}^{\infty} a_n 1_A(x_n).$$

Dies definiert ein Maß μ, jedoch nicht in jedem Fall ein σ-endliches, wie das Beispiel $a_n = 1$ für alle $n \in \mathbb{N}$ und $\mathcal{A} = \{\emptyset, \Omega\}$ zeigt. Falls jedoch $\{x_n\} \in \mathcal{A}$ für jedes $n \in \mathbb{N}$ gilt, so ist μ offenbar σ-endlich. Ein Maß dieser Gestalt soll *diskret* genannt werden. Die a_n heißen *Gewichte* auf den Punkten x_n. Im Fall eines Wahrscheinlichkeitsmaßes gilt $\sum_{n=1}^{\infty} a_n = 1$.

1d. Meßbare Abbildungen

Definition 1.19. Es seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ zwei meßbare Räume. Eine Abbildung $f : \Omega \to \Omega'$ heißt *$\mathcal{F}$-$\mathcal{F}'$-meßbar,* falls $f^{-1}(\mathcal{F}') = \{ f^{-1}(A) : A \in \mathcal{F}' \} \subset \mathcal{F}$ gilt. Ist aus dem Zusammenhang klar, welche σ-Algebren gemeint sind, so spricht man auch einfach von einer meßbaren Abbildung.
Ist $(\Omega, \mathcal{F})$ ein meßbarer Raum, so ist eine *$\mathcal{F}$-meßbare Funktion* eine Abbildung $f : \Omega \to \mathbb{R}$, die $\mathcal{F}$-$\mathcal{B}$-meßbar ist. Eine *$\mathcal{F}$-meßbare numerische Funktion* ist eine Abbildung $f : \Omega \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ mit $f^{-1}(A) \in \mathcal{F}$ für alle $A \in \mathcal{B}$ sowie $f^{-1}(\{\infty\}) \in \mathcal{F}$ und $f^{-1}(\{-\infty\}) \in \mathcal{F}$.

Bemerkung 1.20. Ist $\mathcal{C}$ ein Erzeugendensystem der σ-Algebra $\mathcal{F}'$ (das heißt $\mathcal{F}' = \sigma(\mathcal{C})$), so ist f genau dann $\mathcal{F}$-$\mathcal{F}'$-meßbar, wenn $f^{-1}(\mathcal{C}) \subset \mathcal{F}$ gilt. Zusammen mit

Bemerkung 1.12 ergibt sich: Eine Funktion $f : \Omega \to \mathbb{R}$ ist genau dann $\mathcal{F}$-meßbar, wenn $\{f \leq t\} \in \mathcal{F}$ für alle $t \in \mathbb{R}$ gilt.

Satz 1.21. *Es sei $(\Omega, \mathcal{F})$ ein meßbarer Raum. Die Menge der meßbaren Funktionen $f : \Omega \to \mathbb{R}$ ist abgeschlossen bezüglich punktweiser Konvergenz, das heißt: Ist $\{f_n\}_{n\in\mathbb{N}}$ eine Folge meßbarer Funktionen, für die $f(\omega) := \lim_{n\to\infty} f_n(\omega)$ für jedes $\omega \in \Omega$ existiert, so ist die Grenzfunktion f meßbar.*

Beweis. Sei $g_n(\omega) = \sup_{m\geq n} f_m(\omega)$ für alle $n \in \mathbb{N}$ und alle $\omega \in \Omega$. Dann gilt

$$\{g_n > t\} = \bigcup_{m\geq n} \{f_m > t\} = \bigcup_{m\geq n} \{f_m \leq t\}^c \in \mathcal{F},$$

also auch $\{g_n \leq t\} \in \mathcal{F}$ für alle $t \in \mathbb{R}$, woraus

$$\{g_n < t\} = \bigcup_k \Big\{g_n \leq t - \frac{1}{k}\Big\} \in \mathcal{F}$$

für alle $t \in \mathbb{R}$ folgt. Somit ist

$$\{f < t\} = \bigcup_{n=1}^{\infty} \{g_n < t\} \in \mathcal{F}$$

für alle $t \in \mathbb{R}$, also auch

$$\{f \leq t\} = \bigcap_{n=1}^{\infty} \Big\{f < t + \frac{1}{n}\Big\} \in \mathcal{F}$$

für alle $t \in \mathbb{R}$. $\square$

Analog zeigt man, daß für jede Folge $\{f_n\}_{n\in\mathbb{N}}$ von meßbaren numerischen Funktionen $\overline{\lim}\, f_n$ und $\underline{\lim}\, f_n$ meßbar sind.

Der folgende Satz zeigt, daß die Menge der meßbaren Funktionen auch abgeschlossen gegenüber den üblichen arithmetischen Operationen ist.

Satz 1.22.

(a) *Sind f, g meßbare Funktionen, so sind auch $f + g$ (definiert durch $(f + g)(\omega) = f(\omega) + g(\omega)$ für alle $\omega \in \Omega$) und $f \cdot g$ sowie af für $a \in \mathbb{R}$ meßbar.*

(b) *Sind f, g meßbare Funktionen und gilt $g(\omega) \neq 0$ für alle $\omega \in \Omega$, so ist f/g meßbar.*

(c) *Jede konstante Funktion ist meßbar.*

(d) *Ist $A \in \mathcal{F}$, so ist die Indikatorfunktion $1_A : \Omega \to \mathbb{R}$ meßbar.*

(e) *Zusammensetzungen von meßbaren Abbildungen sind meßbar: Es seien $(\Omega_i, \mathcal{F}_i)$ für $i = 1, 2, 3$ meßbare Räume, und $f_i : \Omega_i \to \Omega_{i+1}$ seien $\mathcal{F}_i$-$\mathcal{F}_{i+1}$-meßbar für $i = 1, 2$. Dann ist $f_2 \circ f_1$ eine $\mathcal{F}_1$-$\mathcal{F}_3$-meßbare Abbildung.*

Beweis von (e). $(f_2 \circ f_1)^{-1}(A) = f_1^{-1}(f_2^{-1}(A)) \in \mathcal{F}_1$ für alle $A \in \mathcal{F}_3$. $\square$

Satz 1.23. *Jede stetige Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ ist $\mathcal{B}^n$-$\mathcal{B}^m$-meßbar.*

Beweis. Ist $U \subset R^m$ offen, so ist $f^{-1}(U)$ offen, also nach Bemerkung 1.14 in $\mathcal{B}^n$. Da gemäß Bemerkung 1.14 die offenen Mengen $\mathcal{B}^m$ erzeugen, folgt aus Bemerkung 1.20 die Meßbarkeit von f. □

Aus Satz 1.22(a) und (e) und Satz 1.23 folgt:

Satz 1.24. *Ist f eine meßbare Funktion, so sind $f^+ = \max(f, 0)$, $f^- = \max(-f, 0)$ und $|f|$ meßbar.*

Definition 1.25. Sei $(\Omega, \mathcal{F})$ ein meßbarer Raum. Funktionen der Form $\sum_{i=1}^n a_i 1_{A_i}$ mit $n \in \mathbb{N}$, $a_i \in \mathbb{R}$ und $A_i \in \mathcal{F}$ für $i \in \{1, \ldots, n\}$ bezeichnet man als *einfach.*

Bemerkung 1.26. Jede nichtnegative, meßbare Funktion f ist punktweiser Limes einer Folge $\{f_n\}_{n\in\mathbb{N}}$ monoton ansteigender, nichtnegativer, einfacher Funktionen.

Beweis. Wähle

$$f_n := \sum_{k=1}^{n2^n} (k-1)2^{-n} 1_{\{(k-1)2^{-n} \le f < k2^{-n}\}} + n 1_{\{f \ge n\}}.$$

□

Notation: Ist $f : \Omega \to \Omega'$ eine Abbildung und $\mathcal{F}'$ eine σ-Algebra auf Ω', so bezeichnet $\sigma(f) := f^{-1}(\mathcal{F}')$. Sind $f_i : \Omega \to \Omega_i$ Abbildungen und $\mathcal{F}_i$ σ-Algebren auf Ω_i für i aus einer beliebigen Indexmenge I, so ist $\sigma(f_i : i \in I) := \bigvee_{i\in I} f_i^{-1}(\mathcal{F}_i)$.

Wichtig ist der folgende *Faktorisierungssatz:*

Satz 1.27. *Es seien Ω eine Menge, $(\Omega', \mathcal{F}')$ ein meßbarer Raum, $f : \Omega \to \overline{\mathbb{R}}$ eine numerische Funktion und $g : \Omega \to \Omega'$ eine Abbildung. Dann ist f genau dann $\sigma(g)$-meßbar, wenn eine $\mathcal{F}'$-meßbare numerische Funktion $h : \Omega' \to \overline{\mathbb{R}}$ existiert mit $f = h \circ g$.*

Beweis. Eine Richtung ist klar: Falls h existiert, so ist

$$f^{-1}(\mathcal{B}) = g^{-1}(h^{-1}(\mathcal{B})) \subset g^{-1}(\mathcal{F}') = \sigma(g).$$

Analog gelten $f^{-1}(\{\infty\}) \in \sigma(g)$ und $f^{-1}(\{-\infty\}) \in \sigma(g)$. Für die andere Richtung: siehe Bauer, Lemma 55.1. □

Aus Bemerkung 1.26 folgt sofort die folgende Charakterisierung der nichtnegativen, meßbaren Funktionen:

Satz 1.28. *Es sei $(\Omega, \mathcal{F})$ ein meßbarer Raum, und Γ sei eine Menge nichtnegativer, meßbarer Funktionen, für die folgende Bedingungen erfüllt sind:*

(a) $f, g \in \Gamma$ *und* $a, b \in \mathbb{R}_+ \Rightarrow af + bg \in \Gamma$
(b) $f_n \in \Gamma$ *für alle* $n \in \mathbb{N}$, $f_n(\omega) \uparrow f(\omega)$ *für alle* $\omega \in \Omega \Rightarrow f \in \Gamma$
(c) $1_A \in \Gamma$ *für alle* $A \in \mathcal{F}$.

Dann ist Γ die Menge aller nichtnegativen, meßbaren Funktionen.

1e. Integration

Für den ganzen Abschnitt sei ein fester Maßraum $(\Omega, \mathcal{F}, \mu)$ vorgegeben.

Definition 1.29.

(a) Sei $f = \sum_{i=1}^n a_i 1_{A_i}$ eine nichtnegative (das heißt $a_i \geq 0$), meßbare, einfache Funktion. Dann wird das Integral $\int f\, d\mu = \int f(\omega)\mu(d\omega)$ von f definiert durch $\int f\, d\mu = \sum_{i=1}^n a_i \mu(A_i)$. (Man weist nach, daß das Integral nicht von der speziellen Darstellung von f abhängt.)
(b) Sei $f\colon \Omega \to \mathbb{R}$ nichtnegativ und meßbar. Nach Bemerkung 1.26 existiert eine Folge von nichtnegativen, einfachen Funktionen $\{f_n\}_{n\in\mathbb{N}}$ mit $f_n \uparrow f$. Dann ist $\int f\, d\mu = \int f(\omega)\, \mu(d\omega) \in [0,\infty]$ definiert durch $\lim_{n\to\infty} \int f_n\, d\mu$. (Man weist wieder nach, daß das Integral nicht von der speziell gewählten Folge $\{f_n\}_{n\in\mathbb{N}}$ abhängt.)

Für nichtnegative, meßbare Funktionen ist also das Integral stets definiert; es kann aber unendlich sein. Das Integral ist insbesondere *monoton*, das heißt $f \leq g \Rightarrow \int f\, d\mu \leq \int g\, d\mu$, und *linear*, das heißt, für $a, b \geq 0$ gilt $\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu$.

Definition 1.30. Eine meßbare Funktion f heißt *μ-integrierbar*, falls $\int |f| d\mu < \infty$ ist.

Wegen $|f| = f^+ + f^-$ (mit $f^+ = \max(f, 0)$ und $f^- = \max(-f, 0)$) ist das gleichbedeutend damit, daß $\int f^+ d\mu < \infty$ und $\int f^- d\mu < \infty$ gelten.

Definition 1.31. Sei f eine μ-integrierbare Funktion.

(a) Das Integral $\int f\, d\mu = \int f(\omega)\mu(d\omega)$ ist definiert durch $\int f^+ d\mu - \int f^- d\mu$.
(b) Ist $A \in \mathcal{F}$, so ist $\int_A f\, d\mu := \int (1_A f)\, d\mu$ das Integral von f über A.

Notation: Wir schreiben $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ beziehungsweise kurz $f \in \mathcal{L}^1(\mu)$ oder $f \in \mathcal{L}^1$, wenn f μ-integrierbar ist.

Satz 1.32. *Seien $f, g \in \mathcal{L}^1$. Dann gelten:*

(a) $f \geq g \Rightarrow \int f\, d\mu \geq \int g\, d\mu$ *(Monotonie des Integrals).*
(b) $af + bg \in \mathcal{L}^1$ *und* $\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu$ *für alle* $a, b \in \mathbb{R}$ *(Linearität des Integrals).*
(c) $\int_{A\cup B} f\, d\mu = \int_A f\, d\mu + \int_B f\, d\mu$ *für alle* $A, B \in \mathcal{F}$ *mit* $A \cap B = \emptyset$ *(Additivität des Integrals).*

Für die Formulierung der nachfolgenden Konvergenzsätze benötigt man folgende Begriffsbildung:

Definition 1.33. Eine Eigenschaft, die die Elemente von Ω haben können, gilt *μ-fast überall,* falls die Menge der $\omega \in \Omega$, für die die Eigenschaft nicht gilt, in einer meßbaren Menge vom μ-Maß null enthalten ist.

Bemerkung 1.34. Ist die Menge der $\omega \in \Omega$, für die eine Eigenschaft nicht gilt, selbst meßbar, so ist dies natürlich gleichbedeutend damit, daß diese Menge μ-Maß null hat.

Beispiel 1.35. Es seien f, g, f_n für alle $n \in \mathbb{N}$ meßbare Funktionen. Dann gelten:

(a) $f = g$ fast überall $\Leftrightarrow \mu(\{\, \omega \colon f(\omega) \neq g(\omega) \,\}) = 0$.
(b) $\lim_{n\to\infty} f_n = g$ fast überall $\Leftrightarrow \mu(\{\, \omega \colon \lim_{n\to\infty} f_n(\omega)$ existiert nicht oder ist von $g(\omega)$ verschieden $\}) = 0$.

Satz 1.36. *Seien $f, g \in \mathcal{L}^1$. Dann gelten:*

(a) *$f = g$ μ-fast überall $\Rightarrow \int f\, d\mu = \int g\, d\mu$.*
(b) *$f \geq g$ und $\int f\, d\mu = \int g\, d\mu \Rightarrow f = g$ μ-fast überall.*

Satz 1.37 (MONOTONER KONVERGENZSATZ)**.** *Es sei $\{f_n\}_{n\in\mathbb{N}}$ eine Folge nichtnegativer, meßbarer Funktionen, und es gelte $f_n \uparrow f$ μ-fast überall. Dann ist $\int f\, d\mu = \lim_{n\to\infty} \int f_n\, d\mu$.*

Lemma 1.38 (LEMMA VON FATOU)**.** *Es sei $\{f_n\}_{n\in\mathbb{N}}$ eine Folge nichtnegativer, meßbarer Funktionen. Dann gilt:*

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

Bemerkung 1.39. Natürlich muß $g := \liminf_{n\to\infty} f_n$ keine reellwertige Funktion sein. Sei $U = \{g = \infty\}$. Man definiert $\int g\, d\mu = \infty$, falls $\mu(U) > 0$ ist, und $\int g\, d\mu = \int (1_{U^c} g)\, d\mu$, falls $\mu(U) = 0$ ist.

Satz 1.40 (BESCHRÄNKTER KONVERGENZSATZ)**.** *Es seien f_n, $n \in \mathbb{N}$, und f meßbare Funktionen, und es gebe ein $g \in \mathcal{L}^1$ mit $|f_n| \leq g$ μ-fast überall für alle $n \in \mathbb{N}$. Ist dann $f = \lim_{n\to\infty} f_n$ μ-fast überall, so gelten $f \in \mathcal{L}^1$ und*

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

Der folgende Satz ist eine Anwendung von Satz 1.37.

Satz 1.41. *Sei $f\colon \Omega \to [0,\infty)$ meßbar. Dann wird durch $\mathcal{F} \ni A \mapsto \nu(A) := \int_A f\,d\mu$ ein (stets σ-endliches!) Maß auf $\mathcal{F}$ definiert. (Man beachte, daß μ als σ-endlich vorausgesetzt ist.)*

Beweis. Offensichtlich gilt $\nu(\emptyset) = 0$. Ist $\{A_n\}_{n\in\mathbb{N}}$ eine Folge paarweise disjunkter Mengen aus $\mathcal{F}$, und ist $A = \bigcup_{n=1}^{\infty} A_n$, so gilt

$$\begin{aligned}\nu(A) &= \int_A f\,d\mu \\ &= \lim_{n\to\infty} \int_{\bigcup_{k=1}^{n} A_k} f\,d\mu \text{ (nach Definition 1.31(b) und Satz 1.37)} \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \int_{A_k} f\,d\mu \text{ (nach Satz 1.32(c))} \\ &= \sum_{k=1}^{\infty} \nu(A_k).\end{aligned}$$

Der Beweis, daß ν σ-endlich ist, bleibt als Übungsaufgabe. $\square$

Definition 1.42. Es seien $(\Omega, \mathcal{F})$ ein meßbarer Raum und μ, ν zwei Maße auf $\mathcal{F}$. Dann heißt ν *absolutstetig* bezüglich μ (Notation: $\nu \ll \mu$), falls folgende Bedingung erfüllt ist:

$$\forall A \in \mathcal{F}:\ \mu(A) = 0 \Rightarrow \nu(A) = 0.$$

Satz 1.43 (SATZ VON RADON-NIKODYM). *Es gilt $\nu \ll \mu$ genau dann, wenn eine nichtnegative, meßbare Funktion $f\colon \Omega \to \mathbb{R}$ existiert mit $\nu(A) = \int_A f\,d\mu$ für alle $A \in \mathcal{F}$.*

Definition 1.44. In der Situation von Satz 1.43 heißt f *Dichte von ν bezüglich μ,* und man schreibt $f = d\nu/d\mu$ oder $f(\omega) = \nu(d\omega)/\mu(d\omega)$.

Bemerkung 1.45. Die Dichte f ist nicht eindeutig durch ν und μ bestimmt. Sind f, g zwei Dichten von ν bezüglich μ, so stimmen sie jedoch μ-fast überall überein.

Satz 1.46. *Es seien $f = d\nu/d\mu$ und g eine meßbare Funktion.*

(a) *Ist $g \geq 0$, so gilt $\int g\,d\nu = \int gf\,d\mu$.*
(b) *Es gilt $g \in \mathcal{L}^1(\nu)$ genau dann, wenn $gf \in \mathcal{L}^1(\mu)$ ist, und in diesem Fall gilt ebenfalls die obige Gleichung.*

1f. Induzierte Maße (Bildmaße)

Es seien $(\Omega_1, \mathcal{F}_1, \mu)$ ein Maßraum und $(\Omega_2, \mathcal{F}_2)$ ein meßbarer Raum. Die Abbildung $\varphi\colon \Omega_1 \to \Omega_2$ sei $\mathcal{F}_1$-$\mathcal{F}_2$-meßbar, und $\mu\varphi^{-1}\colon \mathcal{F}_2 \to [0,\infty]$ sei definiert durch $(\mu\varphi^{-1})(A) = \mu(\varphi^{-1}(A))$. Eine einfache Nachprüfung ergibt:

Bemerkung 1.47. $\mu\varphi^{-1}$ ist ein Maß auf $\mathcal{F}_2$.

Definition 1.48. Das Maß $\mu\varphi^{-1}$ heißt das *durch* φ *auf* $\mathcal{F}_2$ *induzierte Maß* oder das *Bildmaß von* μ *unter* φ.

Satz 1.49 (INTEGRATION BEZÜGLICH BILDMASSEN)**.** *Sei* $f\colon \Omega_2 \to \mathbb{R}$ *eine* $\mathcal{F}_2$*-meßbare Funktion.*

(a) *Für* $f \geq 0$ *gilt* $\int_{\Omega_2} f \, d(\mu\varphi^{-1}) = \int_{\Omega_1} (f \circ \varphi) \, d\mu$.
(b) *Es gilt* $f \in \mathcal{L}^1(\mu\varphi^{-1})$ *genau dann, wenn* $f \circ \varphi \in \mathcal{L}^1(\mu)$ *ist, und die obige Gleichung gilt in diesem Fall ebenfalls.*

Beispiel 1.50. Sei λ_n das Lebesgue-Maß auf $(\mathbb{R}^n, \mathcal{B}^n)$ und sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ eine Kongruenz. Dann gilt $\lambda_n\varphi^{-1} = \lambda_n$.

Wichtig ist das Transformationsverhalten von Dichten:

Satz 1.51. *Seien* μ, ν *zwei Maße auf* $(\Omega, \mathcal{F})$ *mit* $\mu \ll \nu$ *und* $d\mu/d\nu = f$. *Es sei* $\varphi\colon \Omega \to \Omega$ *bijektiv, und* φ *und* φ^{-1} *seien* $\mathcal{F}$*-*$\mathcal{F}$*-meßbar. Gilt* $\nu\varphi^{-1} \ll \nu$, *so gelten auch* $\mu\varphi^{-1} \ll \nu$ *und* $d(\mu\varphi^{-1})/d\nu = (f \circ \varphi^{-1})(d(\nu\varphi^{-1})/d\nu)$.

1g. Produktmaße

Es seien $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ zwei meßbare Räume.

Definition 1.52.

(a) Sei $\Omega = \Omega_1 \times \Omega_2$ die Produktmenge. Die *Produkt-*σ*-Algebra* $\mathcal{F}_1 \otimes \mathcal{F}_2$ ist die σ-Algebra, die vom Mengensystem $\mathcal{C} = \{\, A_1 \times A_2 \colon A_i \in \mathcal{F}_i \,\}$ erzeugt wird. (Das Mengensystem $\mathcal{C}$ ist selbst *keine* σ-Algebra, es ist aber durchschnittstabil.)
(b) Sei $f\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ eine Abbildung. Für $\omega_1 \in \Omega_1$ sei $f^{(1)}_{\omega_1}\colon \Omega_2 \to \mathbb{R}$ definiert durch $f^{(1)}_{\omega_1}(\omega_2) = f(\omega_1, \omega_2)$ für alle $\omega_2 \in \Omega_2$, und analog für $\omega_2 \in \Omega_2$ sei $f^{(2)}_{\omega_2}(\omega_1) = f(\omega_1, \omega_2)$ für alle $\omega_1 \in \Omega_1$ definiert. Die Funktionen $f^{(1)}_{\omega_1}$, $f^{(2)}_{\omega_2}$ heißen *Schnitte* von f.

Satz 1.53. *Ist* $f\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ *eine* $\mathcal{F}_1 \otimes \mathcal{F}_2$*-meßbare Funktion, so sind alle Schnitte meßbar bezüglich der jeweiligen* σ*-Algebra, das heißt, für alle* $\omega_2 \in \Omega_2$ *ist* $f^{(2)}_{\omega_2}$ $\mathcal{F}_1$*-meßbar und für alle* $\omega_1 \in \Omega_1$ *ist* $f^{(1)}_{\omega_1}$ $\mathcal{F}_2$*-meßbar.*

Satz 1.54. *Seien* μ_1, μ_2 σ*-endliche Maße auf* $\mathcal{F}_1$ *beziehungsweise* $\mathcal{F}_2$. *Dann gibt es genau ein Maß* μ *auf* $\mathcal{F}_1 \otimes \mathcal{F}_2$ *mit der Eigenschaft, daß* $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ *für alle* $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$ *gilt.*

Definition 1.55. Dieses Maß μ heißt das *Produktmaß* von μ_1 und μ_2 und wird mit $\mu = \mu_1 \otimes \mu_2$ bezeichnet.

Satz 1.56 (SATZ VON FUBINI). *In der Situation von Satz 1.54 sei* $f: \Omega_1 \times \Omega_2 \to \mathbb{R}$ *in* $\mathcal{L}^1(\Omega_1 \times \Omega_2,\ \mathcal{F}_1 \otimes \mathcal{F}_2,\ \mu_1 \otimes \mu_2)$*. Dann ist für* μ_1*-fast alle* $\omega_1 \in \Omega_1$ *die Funktion* $f^{(1)}_{\omega_1} \in \mathcal{L}^1(\Omega_2, \mathcal{F}_2, \mu_2)$ *und für* μ_2*-fast alle* $\omega_2 \in \Omega_2$ *die Funktion* $f^{(2)}_{\omega_2} \in \mathcal{L}^1(\Omega_1, \mathcal{F}_1, \mu_1)$*. Setzt man*

$$f_1(\omega_1) = \begin{cases} \int f^{(1)}_{\omega_1}(\omega_2)\,\mu_2(d\omega_2), & \textit{falls } f^{(1)}_{\omega_1} \in \mathcal{L}^1, \\ 0, & \textit{sonst,} \end{cases}$$

und $f_2(\omega_2)$ *analog, so gilt:*

$$\int f\,d(\mu_1 \otimes \mu_2) = \int f_1(\omega_1)\,\mu_1(d\omega_1) = \int f_2(\omega_2)\,\mu_2(d\omega_2).$$

In einfacherer Sprechweise (die nicht ganz präzise ist) besagt der obige Satz: Ist $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, so darf man die Integration vertauschen:

$$\begin{aligned} \int f\,d(\mu_1 \otimes \mu_2) &= \int \Big(\int f(\omega_1, \omega_2)\,\mu_1(d\omega_1) \Big)\,\mu_2(d\omega_2) \\ &= \int \Big(\int f(\omega_1, \omega_2)\,\mu_2(d\omega_2) \Big)\,\mu_1(d\omega_1). \end{aligned} \tag{1.57}$$

Satz 1.58 (SATZ VON FUBINI-TONELLI). *Ist* $f: \Omega_1 \times \Omega_2 \to [0, \infty)$ *meßbar und nichtnegativ, so gilt die Formel (1.57) ohne weitere Voraussetzungen.*

Bemerkung 1.59. Daraus ergibt sich, daß eine meßbare Funktion $g: \Omega_1 \times \Omega_2 \to \mathbb{R}$ genau dann $\mu_1 \otimes \mu_2$-integrierbar ist, wenn

$$\int \Big(\int |g(\omega_1, \omega_2)|\,\mu_1(d\omega_1) \Big)\,\mu_2(d\omega_2) < \infty$$

gilt.

Bemerkung 1.60. Die obigen Betrachtungen können leicht auf endlich viele Faktoren ausgedehnt werden: Sind $(\Omega_i, \mathcal{F}_i, \mu_i)$ für $1 \le i \le n$ Maßräume, so definiert man die Produkt-σ-Algebra $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$ durch $(\dots((\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3) \otimes \cdots \otimes \mathcal{F}_{n-1}) \otimes \mathcal{F}_n$, wobei die Produktmenge $\Omega_1 \times \cdots \times \Omega_n$ mit $(\dots((\Omega_1 \times \Omega_2) \times \Omega_3) \times \cdots \times \Omega_{n-1}) \times \Omega_n$ identifiziert wird. Analog wird das Produktmaß $\mu_1 \otimes \cdots \otimes \mu_n$ durch $(\dots((\mu_1 \otimes \mu_2) \otimes \mu_3) \otimes \cdots \otimes \mu_{n-1}) \otimes \mu_n$ definiert.

Man weist leicht nach, daß man beliebig umklammern kann: Ist $1 \le k < n$, so gelten die Beziehungen $(\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k) \otimes (\mathcal{F}_{k+1} \otimes \cdots \otimes \mathcal{F}_n) = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$ und $(\mu_1 \otimes \cdots \otimes \mu_k) \otimes (\mu_{k+1} \otimes \cdots \otimes \mu_n) = \mu_1 \otimes \cdots \otimes \mu_n$ beim Identifizieren von $(\Omega_1 \times \cdots \times \Omega_k) \times (\Omega_{k+1} \times \cdots \times \Omega_n)$ mit $\Omega_1 \times \cdots \times \Omega_n$.

Beispiel 1.61. Es seien $(\Omega_i, \mathcal{F}_i, \mu_i) = (\mathbb{R}, \mathcal{B}, \lambda)$ mit dem Lebesgue-Maß λ für $1 \le i \le n$. Man zeigt, daß $\mathcal{B}^n = \mathcal{B} \otimes \cdots \otimes \mathcal{B}$ gilt, und $\lambda_n = \lambda \otimes \cdots \otimes \lambda$ heißt n-dimensionales Lebesgue-Maß (λ_n ist dasselbe Maß wie das in Beispiel 1.17 angedeutete).

Unendliche Produkte von Wahrscheinlichkeitsräumen werden im nächsten Kapitel betrachtet.

Kapitel 2

Zufallsgrößen

Es sei $(\Omega, \mathcal{F})$ ein meßbarer Raum. Wahrscheinlichkeitsmaße auf $\mathcal{F}$ werden oft (aber nicht immer) mit P, Q usw. statt mit μ, ν usw. bezeichnet (P für "probability"). Die Elemente von $\mathcal{F}$ bezeichnet man in der Wahrscheinlichkeitstheorie meist als *Ereignisse*. Die Mengenoperationen in $\mathcal{F}$ werden oft in die „Sprache der Ereignisse" übersetzt:

$$
\begin{array}{rlrl}
A^c : & \text{nicht } A & A \subset B & : A \text{ impliziert } B \\
A \cup B : & A \text{ oder } B & A \cap B : & A \text{ und } B
\end{array}
$$

Da für ein Wahrscheinlichkeitsmaß $P(\Omega) = 1$ gilt, folgt $0 \leq P(A) \leq 1$ für jedes $A \in \mathcal{F}$. Ferner gilt $P(A^c) = 1 - P(A)$. Statt P-fast überall sagt man meist P-*fast sicher* (Abkürzung: P-f. s.).

Im folgenden sei $(\Omega, \mathcal{F}, P)$ ein fester Wahrscheinlichkeitsraum.

2a. Zufallsgrößen und ihre Verteilungen

Definition 2.1. Es sei $(E, \mathcal{E})$ ein meßbarer Raum. Eine $(E, \mathcal{E})$-*wertige Zufallsgröße* (kurz E-wertige Zufallsgröße) ist eine $\mathcal{F}$-$\mathcal{E}$-meßbare Abbildung $X : \Omega \to E$. Im Spezialfall $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ spricht man auch einfach von einer Zufallsgröße, im Falle $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}^n)$ von einem n-dimensionalen *Zufallsvektor* und schreibt ihn oft als $X = (X_1, \ldots, X_n)$, wobei die Abbildungen $X_i : \Omega \to \mathbb{R}$ die Koordinaten sind.

Bemerkung 2.2. Für einen Zufallsvektor $X = (X_1, \ldots, X_n)$ gilt $X_i = \pi_i \circ X$, wobei die Abbildungen $\pi_i : \mathbb{R}^n \to \mathbb{R}$ die Projektionen auf die einzelnen Faktoren des kartesischen Produkts $\mathbb{R}^n$ sind. Nach der Definition von $\mathcal{B}^n$ und der Bemerkung 1.20 sind die Projektionen π_i $\mathcal{B}^n$-$\mathcal{B}$-meßbar, und daher sind nach Satz 1.22(e) die X_i Zufallsgrößen. Sind umgekehrt $X_1, \ldots, X_n$ Zufallsgrößen, so definiert $X = (X_1, \ldots, X_n)$ eine Abbildung $\Omega \to \mathbb{R}^n$. Mit Bemerkung 1.20 folgt, daß X ein n-dimensionaler Zufallsvektor ist.

Weil eine Abbildung $X : \Omega \to E$ genau dann eine $(E, \mathcal{E})$-wertige Zufallsgröße ist, wenn $X^{-1}(\mathcal{E}) \subset \mathcal{F}$ gilt, entspricht die folgende Notation der Notation in Abschnitt 1d.

Notation: Ist X_i für jedes $i \in I$ eine $(E_i, \mathcal{E}_i)$-wertige Zufallsgröße, so bezeichnet man mit $\sigma(X_i : i \in I)$ die σ-Algebra $\bigvee_{i \in I} X_i^{-1}(\mathcal{E}_i)$. (Dies entspricht der Notation aus Abschnitt 1d.)

Definition 2.3. Es sei X eine $(E,\mathcal{E})$-wertige Zufallsgröße. Das Wahrscheinlichkeitsmaß PX^{-1} auf $\mathcal{E}$ heißt die *Verteilung* von X.

Notation: Haben zwei $(E,\mathcal{E})$-wertige Zufallsgrößen X, X' dieselbe Verteilung, so schreiben wir $\mathcal{L}(X)=\mathcal{L}(X')$ (dabei steht $\mathcal{L}$ für "law", den englischen Begriff für Verteilung); X, X' müssen dabei nicht auf demselben Wahrscheinlichkeitsraum definiert sein.

Definition 2.4. Sei X eine ($\mathbb{R}$-wertige) Zufallsgröße. Die Abbildung $\mathbb{R}\ni t\mapsto F_X(t)=P(\{X\le t\})$ heißt *Verteilungsfunktion* von X.

Notation: Um geschweifte Klammern zu sparen, schreibt man meist $P(X\le t)$, $P(X\ne Y)$ usw. anstelle von $P(\{X\le t\})$, $P(\{X\ne Y\})$ (oder noch aufwendiger: $P(\{\,\omega\colon X(\omega)\le t\,\})$). Ein Komma in den Klammern hinter P bedeutet „und", zum Beispiel schreibt man $P(X\le t,\ X'\le t')$ anstelle von $P(\{X\le t\}\cap\{X'\le t'\})$.

Man weist leicht nach, daß die Verteilungsfunktion F_X einer Zufallsgröße X monoton wachsend und rechtsstetig ist mit $\lim_{t\to\infty}F(t)=1$ und $\lim_{t\to-\infty}F(t)=0$.

Bemerkung 2.5. Haben zwei Zufallsgrößen X, X' dieselbe Verteilungsfunktion, so gilt $\mathcal{L}(X)=\mathcal{L}(X')$.

Beweis. Weil $\{\,(-\infty,t]\colon t\in\mathbb{R}\,\}$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}$ ist, folgt die Bemerkung sofort aus Bemerkung 1.9. □

Ohne Beweis sei hier das folgende Ergebnis erwähnt, das wir nicht weiter benützen werden:

Satz 2.6. *Es sei $F\colon\mathbb{R}\to[0,1]$ eine Verteilungsfunktion, das heißt eine rechtsstetige, monoton wachsende Funktion mit $\lim_{t\to\infty}F(t)=1$ und $\lim_{t\to-\infty}F(t)=0$. Dann existiert genau ein Wahrscheinlichkeitsmaß μ auf $\mathbb{R}$ mit $F(t)=\mu((-\infty,t])$ für alle $t\in\mathbb{R}$.*

Zwei wichtige Spezialfälle von Verteilungen auf $(\mathbb{R},\mathcal{B})$ (beziehungsweise allgemeiner auf $(\mathbb{R}^n,\mathcal{B}^n)$) sind besonders wichtig:

(a) *Diskrete Verteilungen:* Nimmt eine Zufallsgröße beziehungsweise ein Zufallsvektor höchstens abzählbar viele Werte $\{x_i\}_{i\in I}\subset\mathbb{R}^n$ an, so ist die zugehörige Verteilung ein diskretes Wahrscheinlichkeitsmaß, also von der Form $\sum_i a_i\delta_{x_i}$ mit $a_i\ge 0$ für alle $i\in I$ und $\sum_i a_i=1$ (siehe Beispiel 1.18).

(b) *Verteilungen mit Dichte* bezüglich des Lebesgue-Maßes λ (beziehungsweise λ_n in $(\mathbb{R}^n,\mathcal{B}^n)$, siehe Beispiel 1.17).

Beispiel 2.7 (VERTEILUNGEN AUF $\mathbb{R}$).

(a) *Poisson-Verteilung zum Parameter* $\lambda > 0$*:*

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n.$$

(b) *Standardnormalverteilung:*

$$\frac{d\mu}{d\lambda}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

(c) *Normalverteilung mit Mittelwert* $a \in \mathbb{R}$ *und Varianz* $\sigma^2 > 0$*:*

$$\frac{d\mu}{d\lambda}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-a)^2}{2\sigma^2}\Big).$$

(d) *Cauchy-Verteilung zum Parameter* $c > 0$*:*

$$\frac{d\mu}{d\lambda}(x) = \frac{c}{\pi(c^2 + x^2)}.$$

Man spricht von Poisson-verteilten Zufallsgrößen, normalverteilten Zufallsgrößen, usw.

Von besonderer Bedeutung in $(\mathbb{R}^n, \mathcal{B}^n)$ ist die mehrdimensionale Normalverteilung:

Definition 2.8.

(a) Das Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}^n, \mathcal{B}^n)$, das durch die Dichte

$$\frac{d\mu}{d\lambda_n}(x_1, \dots, x_n) = (2\pi)^{-n/2} \exp\Big(-\frac{1}{2} \sum_{i=1}^{n} x_i^2\Big)$$

definiert wird, heißt *Standardnormalverteilung* in $\mathbb{R}^n$.

(b) Ist A eine reelle $n \times n$-Matrix und $b \in \mathbb{R}^n$, so definiert $x \mapsto T(x) = Ax + b$ eine affine Abbildung $\mathbb{R}^n \to \mathbb{R}^n$. Verteilungen der Form μT^{-1}, wobei μ die Standard-Normalverteilung ist, heißen *Normalverteilungen*.

Bemerkung 2.9.

(a) Ist A regulär, so ist T invertierbar, und eine Anwendung von Satz 1.51 ergibt

$$\frac{d\mu T^{-1}}{d\lambda_n}(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\Big(-\frac{1}{2}(x-b)^t\, \Sigma^{-1}(x-b)\Big) \text{ mit } \Sigma = AA^t.$$

Ist A nicht regulär, so ist $\lambda_n(\mathrm{Im}(T)) = 0$ für das Bild $\mathrm{Im}(T)$ von T, und somit gilt $\mu T^{-1} \not\ll \lambda_n$.

(b) Die Klasse der Normalverteilungen hat die Eigenschaft, daß sie invariant unter beliebigen affinen Abbildungen $\mathbb{R}^m \to \mathbb{R}^n$ ist: Ist μ eine Normalverteilung auf $\mathbb{R}^m$ und $T\colon \mathbb{R}^m \to \mathbb{R}^n$ eine Abbildung der Form $T(x) = Ax + b$, wobei A eine $m \times n$-Matrix und $b \in \mathbb{R}^n$ sind, so ist μT^{-1} eine Normalverteilung auf $\mathbb{R}^n$. Obgleich dies hier nicht allzu schwierig zu beweisen ist, verschieben wir den Beweis besser auf später, wenn charakteristische Funktionen als Hilfsmittel zur Verfügung stehen.

Definition 2.10. Es sei $(E, \mathcal{E})$ ein meßbarer Raum.

(a) Eine Folge $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ von $(E, \mathcal{E})$-wertigen Zufallsgrößen, die auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind, heißt (abzählbarer) $(E, \mathcal{E})$-wertiger *stochastischer Prozeß*. Im Falle $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ sprechen wir einfach von einem stochastischen Prozeß.

(b) Es bezeichne $E^{\mathbb{N}}$ die Menge der Folgen von Elementen aus E, das heißt die Menge aller Abbildungen $\mathbb{N} \to E$. Seien $\pi_i\colon E^{\mathbb{N}} \to E$ für $i \in \mathbb{N}$ die Projektionsabbildungen. Dann ist auf E die σ-Algebra $\mathcal{E}^{\mathbb{N}}$ wie folgt definiert:

$$\mathcal{E}^{\mathbb{N}} := \bigvee_{i\in\mathbb{N}} \pi_i^{-1}(\mathcal{E}).$$

Diese Definition verallgemeinert die Definition der Produkt-σ-Algebra (Definition 1.52(a)). Wir zeigen nun, daß ein $(E, \mathcal{E})$-wertiger stochastischer Prozeß als $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$-wertige Zufallsgröße betrachtet werden kann. Weil eine Folge $(X_n)_{n\in\mathbb{N}}$ von Abbildungen $X_n\colon \Omega \to E$ nichts anderes als eine Abbildung $\mathbb{X}\colon \Omega \to E^{\mathbb{N}}$ ist, haben wir uns nur um die Meßbarkeit zu kümmern:

Satz 2.11. *Es sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ eine Folge von Abbildungen $X_n\colon \Omega \to E$. Dann ist $\mathbb{X}$ genau dann $\mathcal{F}$-$\mathcal{E}^{\mathbb{N}}$-meßbar, wenn für alle $i \in \mathbb{N}$ die Abbildungen X_i $\mathcal{F}$-$\mathcal{E}$-meßbar sind.*

Beweis. Nach Definition von $\mathcal{E}^{\mathbb{N}}$ sind alle Projektionen $\pi_i\colon E^{\mathbb{N}} \to E$ $\mathcal{E}^{\mathbb{N}}$-$\mathcal{E}$-meßbar. Ist $\mathbb{X}$ $\mathcal{F}$-$\mathcal{E}^{\mathbb{N}}$-meßbar, so ist somit $X_i = \pi_i \circ \mathbb{X}$ für jedes $i \in \mathbb{N}$ eine $\mathcal{F}$-$\mathcal{E}$-meßbare Abbildung. Sind umgekehrt alle X_i $\mathcal{F}$-$\mathcal{E}$-meßbar, so gilt $\mathbb{X}^{-1}(\pi_i^{-1}(A)) = X_i^{-1}(A) \in \mathcal{F}$ für alle $i \in \mathbb{N}$ und alle $A \in \mathcal{E}$. Da per Definition $\{\, \pi_i^{-1}(A) : i \in \mathbb{N},\ A \in \mathcal{E} \,\}$ ein Erzeugendensystem von $\mathcal{E}^{\mathbb{N}}$ ist, folgt die Meßbarkeit von $\mathbb{X}$ aus der Bemerkung 1.20. $\square$

Die Verteilung eines stochastischen Prozesses $\mathbb{X}$ ist einfach seine Verteilung als $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$-wertige Zufallsgröße. Die meisten der uns interessierenden Fragen hängen nur von der Verteilung des stochastischen Prozesses ab. Ist $\mathbb{X}$ ein stochastischer Prozeß, so ist die Folge $(\pi_i)_{i\in\mathbb{N}}$ der Projektionen $\pi_i\colon E^{\mathbb{N}} \to E$ ein auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}, P\mathbb{X}^{-1})$ definierter stochastischer Prozeß, der dieselbe Verteilung wie $\mathbb{X}$ hat. Es ist daher meist keine Einschränkung anzunehmen, daß der zugrunde liegende Wahrscheinlichkeitsraum von der Form $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}, P)$ ist und die X_i die Projektionen von $E^{\mathbb{N}}$ auf E sind.

Die Beschreibung von Verteilungen auf einem unendlichen Produktraum ist nicht ganz einfach. Sie geschieht in der Regel über die sogenannten endlichdimensionalen Verteilungen.

Definition 2.12. Ist Q ein Wahrscheinlichkeitsmaß auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ und ist $\pi^{(n)}: E^{\mathbb{N}} \to E^n$ für $n \in \mathbb{N}$ durch $\pi^{(n)} = (\pi_1, \dots, \pi_n)$ definiert, so ist $Q^{(n)} := Q\pi^{(n)-1}$ ein Wahrscheinlichkeitsmaß auf $(E^n, \mathcal{E}^n)$. Die Maße $Q^{(n)}$ heißen die *endlichdimensionalen Verteilungen* von Q.

Wir wollen im folgenden zwei Fragen nachgehen:

(a) Legen die endlichdimensionalen Verteilungen $Q^{(n)}$, $n \in \mathbb{N}$, die Verteilung Q eindeutig fest?
(b) Gibt es zu einer vorgegebenen Folge $(Q^{(n)})_{n\in\mathbb{N}}$ von Wahrscheinlichkeitsmaßen auf $(E^n, \mathcal{E}^n)$ stets ein Wahrscheinlichkeitsmaß Q auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$, dessen endlichdimensionale Verteilungen die $Q^{(n)}$ sind?

Die Antwort auf (a) lautet uneingeschränkt „Ja".

Satz 2.13. *Die endlichdimensionalen Verteilungen $(Q^{(n)})_{n\in\mathbb{N}}$ bestimmen das Wahrscheinlichkeitsmaß Q eindeutig.*

Beweis. Durch die endlichdimensionalen Verteilungen ist das Maß Q auf der Algebra $\bigcup_{n\in\mathbb{N}} \pi^{(n)-1}(\mathcal{E}^n)$ festgelegt. Weil diese Algebra ein durchschnittstabiles Erzeugendensystem von $\mathcal{E}^{\mathbb{N}}$ ist, folgt die Eindeutigkeit aus Bemerkung 1.9. $\square$

Im Spezialfall $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ ist das folgende Lemma hilfreich.

Lemma 2.14. *Sind $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ und $\mathbb{X}' = (X'_n)_{n\in\mathbb{N}}$ zwei $(\mathbb{R}, \mathcal{B})$-wertige stochastische Prozesse auf den Wahrscheinlichkeitsräumen $(\Omega, \mathcal{F}, P)$ beziehungsweise $(\Omega', \mathcal{F}', P')$, so gilt $\mathcal{L}(\mathbb{X}) = \mathcal{L}(\mathbb{X}')$ genau dann, wenn*

$$P(X_1 \le t_1, \dots, X_n \le t_n) = P'(X'_1 \le t_1, \dots, X'_n \le t_n)$$

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

Beweis. Die Mengen der Form $\{x \in \mathbb{R}^{\mathbb{N}} : x_1 \le t_1, x_2 \le t_2, \dots, x_n \le t_n\}$ bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}^{\mathbb{N}}$. Nach Voraussetzung stimmen $P\mathbb{X}^{-1}$ und $P'\mathbb{X}'^{-1}$ auf diesen Mengen überein. Somit folgt $P\mathbb{X}^{-1} = P'\mathbb{X}'^{-1}$ auf $\mathcal{B}^{\mathbb{N}}$ aus Bemerkung 1.9. $\square$

Die Antwort auf (b) ist viel schwieriger und kann im Moment nicht vollständig diskutiert werden. Es ist jedoch offensichtlich, daß die $Q^{(n)}$ einer Verträglichkeitsbedingung genügen müssen, damit überhaupt eine Chance besteht, ein Wahrscheinlichkeitsmaß Q auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ zu finden mit

$$Q^{(n)} = Q\pi^{(n)-1}. \tag{2.15}$$

Ist nämlich $\varphi_n\colon E^n \to E^{n-1}$ die Projektion auf die ersten $n-1$ Koordinaten, so gilt $\pi^{(n-1)} = \varphi_n \circ \pi^{(n)}$ und aus (2.15) ergibt sich

$$Q^{(n-1)} = Q^{(n)}\varphi_n^{-1}, \quad n \geq 2. \tag{2.16}$$

Definition 2.17. Eine Folge von Wahrscheinlichkeitsmaßen $Q^{(n)}$ auf $(E^n, \mathcal{E}^n)$ für $n \geq 1$ heißt *verträglich,* wenn (2.16) gilt.

Die Frage (b) muß also dahin präzisiert werden, ob zu jeder verträglichen Folge ein Wahrscheinlichkeitsmaß Q auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ existiert, das (2.15) erfüllt. Die Antwort ist leider „Nein“. Es gibt jedoch wichtige Spezialfälle, in denen die Antwort „Ja“ lautet.

Satz 2.18 (SATZ VON KOLMOGOROFF)**.** *Ist $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$, so existiert zu jeder verträglichen Familie $Q^{(n)}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^n, \mathcal{B}^n)$ genau ein Wahrscheinlichkeitsmaß Q auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ mit (2.15).*

Bemerkung 2.19. Der Satz gilt allgemeiner, wenn E ein vollständiger, separabler metrischer Raum ist und $\mathcal{E}$ die von den offenen Mengen erzeugte σ-Algebra ist.

Der Satz von Kolmogoroff ist ein Spezialfall des Satzes von Ionescu-Tulcea, der im folgenden kurz diskutiert werden soll. Wir werden diesen in einem Anhang beweisen. Daß der Satz von Kolmogoroff eine Folgerung des Satzes von Ionescu-Tulcea ist, ist jedoch auch nicht offensichtlich und wird sich im Kapitel 5 ergeben. Um den Satz von Ionescu-Tulcea vorzustellen, benötigen wir eine wichtige Verallgemeinerung des Begriffs der stochastischen Matrix.

Definition 2.20. Es seien $(E_1, \mathcal{E}_1)$, $(E_2, \mathcal{E}_2)$ zwei meßbare Räume. Ein *Markoffkern* K von $(E_1, \mathcal{E}_1)$ nach $(E_2, \mathcal{E}_2)$ ist eine Abbildung $K\colon E_1 \times \mathcal{E}_2 \to [0,1]$ mit den folgenden zwei Eigenschaften:

(a) Für alle $x \in E_1$ ist $K(x, \cdot)$ ein Wahrscheinlichkeitsmaß auf $(E_2, \mathcal{E}_2)$.
(b) Für alle $A \in \mathcal{E}_2$ ist $K(\cdot, A)$ eine $\mathcal{E}_1$-meßbare Funktion auf E_1.

Beispiel 2.21. Zwei (extreme) Spezialfälle von Markoffkernen sind die folgenden:

(a) $K(x, A) := \mu(A)$ für ein Wahrscheinlichkeitsmaß μ auf $(E_2, \mathcal{E}_2)$.
(b) $K(x, A) := 1_A(f(x))$ für eine meßbare Abbildung f von $(E_1, \mathcal{E}_1)$ nach $(E_2, \mathcal{E}_2)$.

Man stellt sich einen Kern am besten als eine Art „fuzzy“ Abbildung vor. Wichtig ist, daß sich mit Kernen der Begriff des Produktmaßes verallgemeinern läßt:

Definition 2.22. Sei ν ein Wahrscheinlichkeitsmaß auf $(E_1, \mathcal{E}_1)$ und K ein Markoffkern wie oben. Dann ist das Wahrscheinlichkeitsmaß $\mu \otimes K$ auf $\mathcal{E}_1 \otimes \mathcal{E}_2$ wie folgt definiert:

$$(\mu \otimes K)(A) = \int \mu(d\omega_1) K(\omega_1, A_{\omega_1}) \quad \text{für alle } A \in \mathcal{E}_1 \otimes \mathcal{E}_2, \tag{2.23}$$

wobei $A_{\omega_1} = \{\, \omega_2 \in E_2 \colon (\omega_1, \omega_2) \in A \,\}$ der ω_1-Schnitt der Menge A ist.

Bemerkung 2.24.

(a) Nach Satz 1.53 ist $A_{\omega_1} \in \mathcal{E}_2$, und demzufolge ist $K(\omega_1, A_{\omega_1})$ für jedes $\omega_1 \in E_1$ wohldefiniert.

(b) Damit die obige Definition sinnvoll ist, muß nachgewiesen werden, daß $\omega_1 \mapsto K(\omega_1, A_{\omega_1})$ eine $\mathcal{E}_1$-meßbare Funktion ist. Dies sieht man wie folgt ein:

$$\mathcal{D} = \{\, A \in \mathcal{E}_1 \otimes \mathcal{E}_2 \colon \omega_1 \mapsto K(\omega_1, A_{\omega_1}) \text{ ist } \mathcal{E}_1\text{-meßbar} \,\}$$

ist ein durchschnittstabiles Dynkinsystem (einfache Übungsaufgabe) und enthält die Mengen der Form $A_1 \times A_2$ mit $A_i \in \mathcal{E}_i$ für $i = 1, 2$, denn es gilt:

$$K\big(\omega_1, (A_1 \times A_2)_{\omega_1}\big) = 1_{A_1}(\omega_1) K(\omega_1, A_2).$$

Nach Satz 1.4 folgt $\mathcal{D} = \mathcal{E}_1 \otimes \mathcal{E}_2$. Damit ist $(\mu \otimes K)(A)$ für jedes $A \in \mathcal{E}_1 \otimes \mathcal{E}_2$ definiert.

(c) Mit Hilfe des Satzes von der majorisierten Konvergenz zeigt man, daß $\mu \otimes K$ ein Wahrscheinlichkeitsmaß ist.

Bemerkung 2.25. Man beachte, daß für die Projektion $\pi_1 \colon E_1 \times E_2 \to E_1$

$$(\mu \otimes K)\pi_1^{-1} = \mu \tag{2.26}$$

gilt. Ist $\pi_2 \colon E_1 \times E_2 \to E_2$ die Projektion, so wird die zweite Randverteilung $(\mu \otimes K)\pi_2^{-1}$ meist mit μK bezeichnet. Es gilt

$$(\mu K)(A) = \int \mu(dx) K(x, A).$$

Nun zurück zu den endlichdimensionalen Verteilungen und dem Satz von Ionescu-Tulcea. Sei $(E, \mathcal{E})$ ein beliebiger meßbarer Raum, und seien μ ein Wahrscheinlichkeitsmaß auf $(E, \mathcal{E})$ und K_n für jedes $n \in \mathbb{N}$ ein Markoffkern von $(E^n, \mathcal{E}^n)$ nach $(E, \mathcal{E})$. Wir können damit rekursiv eine Folge von Wahrscheinlichkeitsmaßen Q_n auf $(E^n, \mathcal{E}^n)$ definieren:

$$Q_1 = \mu, \qquad Q_{n+1} = Q_n \otimes K_n \quad \text{für } n \in \mathbb{N}.$$

Nach (2.26) sind diese Wahrscheinlichkeitsmaße verträglich.

Satz 2.27 (SATZ VON IONESCU-TULCEA). *In der oben beschriebenen Situation existiert stets ein Wahrscheinlichkeitsmaß Q auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$, das (2.15) erfüllt.*

Beweis. Siehe Anhang A1 oder Theorem 9.1 im Buch von S. Ethier, T. Kurtz: Markov Processes, Wiley 1986, S. 504.

Bemerkung 2.28. Der Satz von Kolmogoroff (Satz 2.18) ist ein Spezialfall von Satz 2.27, weil sich für $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ *jede* verträgliche Folge in dieser Weise mit Markoffkernen beschreiben läßt (siehe Kapitel 5 und Anhang A2).

Beispiel 2.29.

(a) *Produktwahrscheinlichkeiten:* Sei $(E, \mathcal{E})$ ein beliebiger meßbarer Raum (zum Beispiel $(\mathbb{R}, \mathcal{B})$). Für jedes $n \in \mathbb{N}$ sei μ_n ein Wahrscheinlichkeitsmaß auf $(E, \mathcal{E})$. Die endlichen Produktmaße $Q_n = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$ sind offenbar von der im Satz von Ionescu-Tulcea geforderten Form, denn wir können den Kern K_n von $(E^{n-1}, \mathcal{E}^{n-1})$ nach $(E, \mathcal{E})$ trivial durch $K_n(x, A) = \mu_n(A)$ wählen. Somit sind die Q_n die endlichdimensionalen Verteilungen eines eindeutig bestimmten Wahrscheinlichkeitsmaßes Q auf $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$. Sind die $X_n : E^{\mathbb{N}} \to E$ für alle $n \in \mathbb{N}$ die Projektionen, so hat X_n die Verteilung μ_n. Man sagt, die X_n seien *unabhängig.* Sind alle μ_i gleich, so heißen die X_i *identisch verteilt.*

(b) Ist $(X_n)_{n\in\mathbb{N}}$ ein Prozeß mit den endlichdimensionalen Verteilungen von Beispiel 2.29(a) und sind alle μ_i gleich, so hat der Prozeß $(X'_n)_{n\in\mathbb{N}}$, definiert durch $X'_n = X_1$, offenbar dieselben eindimensionalen Verteilungen $QX_n^{-1} = QX'^{-1}_n = \mu_1$. Aber abgesehen von trivialen Fällen gilt $\mathcal{L}((X_n)) \neq \mathcal{L}((X'_n))$, denn wenn $0 < \mu_1(A) < 1$ ist, dann gilt $Q_2(A \times A) = \mu_1(A)^2 \neq Q'_2(A \times A) = \mu_1(A)$.

(c) *Markoffketten:* Es seien I eine abzählbare Menge und $(p_{ij})_{i,j\in I}$ eine stochastische Matrix, das heißt, es gelten $p_{ij} \geq 0$ und $\sum_{j\in I} p_{ij} = 1$ für alle $i \in I$. Sei $(q(i))_{i\in I}$ eine beliebige Wahrscheinlichkeitsverteilung auf I. Abzählbare Mengen versehen wir mit der Potenzmenge $\mathcal{P}(I)$ als σ-Algebra. Dann legen q und (p_{ij}) für jedes $n \in \mathbb{N}$ eine Wahrscheinlichkeitsverteilung Q_n auf $(I^n, \mathcal{P}(I)^n)$ fest mittels $Q_1 = q$ und $Q_n = Q_{n-1} \otimes K_n$ mit $K_n((i_1, \ldots, i_{n-1}), i_n) = p(i_{n-1}, i_n)$. Die Verteilung Q_n ist nichts anderes als die in der „Einführung in die Stochastik“ definierte Verteilung $Q_n((i_1, \ldots, i_n)) = q(i_1)p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$. Nach dem Satz von Ionescu-Tulcea existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß auf $(I^{\mathbb{N}}, \mathcal{P}(I)^{\mathbb{N}})$, dessen endlichdimensionale Verteilungen die Q_n sind. Ein stochastischer Prozeß $(X_n)_{n\in\mathbb{N}}$ mit dieser Verteilung heißt *Markoffkette* mit Übergangsmatrix (p_{ij}) und Startverteilung q. (*Bemerkung:* Für jedes $n \in \mathbb{N}$ gilt zwar $\mathcal{P}(I)^n = \mathcal{P}(I^n)$, aber $\mathcal{P}(I)^{\mathbb{N}} \neq \mathcal{P}(I^{\mathbb{N}})$.)

(d) *Stationäre Markoffketten:* Ein wichtiger Spezialfall von Beispiel 2.29(c) liegt vor, wenn $\sum_i q(i)p_{ij} = q(j)$ für alle $j \in I$ gilt. Nicht für alle stochastischen Matrizen existieren derartige stationäre Verteilungen. Sie existieren jedoch für den Fall, daß (p_{ij}) positiv rekurrent ist (siehe „Einführung in die Stochastik“). Sei $(X_n)_{n\in\mathbb{N}}$ eine Markoffkette mit einer stationären Startverteilung, und sei $X'_n = X_{n+1}$. Dann gilt:

$$\mathcal{L}((X_n)) = \mathcal{L}((X'_n)). \tag{2.30}$$

Beweis von (2.30). Es genügt nach Lemma 2.14 nachzuweisen, daß

$$P(X_1 = i_1, \ldots, X_n = i_n) = P(X'_1 = i_1, \ldots, X'_n = i_n)$$

für alle $n \in \mathbb{N}$ und alle $i_1, \ldots, i_n \in I$ gilt. Es gilt

$$\begin{aligned} P(X'_1 = i_1, \ldots, X'_n = i_n) &= P(X_2 = i_1, \ldots, X_{n+1} = i_n) \\ &= \sum_{j \in I} P(X_1 = j,\ X_2 = i_1, \ldots, X_{n+1} = i_n) \\ &= \sum_{j \in I} q(j) p_{ji_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \\ &= q(i_1) p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \\ &= P(X_1 = i_1, \ldots, X_n = i_n). \end{aligned}$$ □

Definition 2.31. Ein stochastischer Prozeß mit der Eigenschaft (2.30) heißt *stationär.*

Stationäre Prozesse sind Gegenstand von Kapitel 4.

2b. Erwartungswerte

Definition 2.32. Sei X eine reelle Zufallsgröße, die auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert ist. Ist $X \geq 0$ oder $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, so heißt $EX = \int X dP$ der *Erwartungswert von* X.

Der Erwartungswert ist also für positive Zufallsgrößen stets definiert, kann in diesem Fall jedoch gleich unendlich sein.

Lemma 2.33. *Es seien X eine $(E, \mathcal{E})$-wertige Zufallsgröße und $\varphi\colon E \to \mathbb{R}$ eine $\mathcal{E}$-$\mathcal{B}$-meßbare Abbildung. Dann ist $\varphi \circ X$ (meist $\varphi(X)$ geschrieben) eine reelle Zufallsgröße. Sei $\hat{P} = PX^{-1}$ die Verteilung von X auf $(E, \mathcal{E})$.*

(a) *Ist $\varphi \geq 0$, so gilt $E(\varphi(X)) = \hat{E}(\varphi)$, wobei $\hat{E}$ der Erwartungswert bezüglich $\hat{P}$ ist.*
(b) *Es gilt $\varphi(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ genau dann, wenn $\varphi \in \mathcal{L}^1(E, \mathcal{E}, \hat{P})$ ist, und in diesem Fall gilt ebenfalls die obige Gleichung.*

Beweis. Das Lemma ist eine Umformulierung von Satz 1.49. □

Bemerkung 2.34.

(a) Aus Lemma 2.33 ergibt sich, daß $E(\varphi(X))$ nur von der Verteilung von X (und natürlich φ) abhängt. Wir betrachten den Spezialfall $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ mit $\varphi =$ Identität. Sei $\mu = PX^{-1}$ die Verteilung der Zufallsgröße X. Dann folgt aus Lemma 2.33, daß $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ genau dann gilt, wenn $\int |x|\, \mu(dx) < \infty$ gilt. In diesem Fall ist $E(X) = \int x\, \mu(dx)$.
(b) Ist $\mu = \sum_i a_i \delta_{x_i}$ mit $x_i \in \mathbb{R}$, $a_i \geq 0$ und $\sum_i a_i = 1$, so gilt $\int |x|\, \mu(dx) = \sum_i a_i |x_i|$. Somit gilt $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ genau dann, wenn $\sum_i a_i |x_i| < \infty$, und in diesem Fall ist $E(X) = \sum_i x_i a_i = \sum_i x_i P(X = x_i)$.

(c) Besitzt μ eine Dichte f bezüglich des Lebesgue-Maßes λ, so gilt

$$X \in \mathcal{L}^1(\Omega, \mathcal{F}, P) \Leftrightarrow \int |x| \, \mu(dx) = \int |x| f(x) \, \lambda(dx) < \infty,$$

und in diesem Fall gilt $E(X) = \int x f(x) \, \lambda(dx)$.

(d) Allgemeiner gilt für eine Zufallsgröße X und eine meßbare Funktion $\varphi\colon \mathbb{R} \to \mathbb{R}$ die folgende Äquivalenz:

$$\varphi(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, P) \Leftrightarrow \begin{cases} \sum_i |\varphi(x_i)| P(X = x_i) < \infty, & \text{im diskreten Fall,} \\ \int |\varphi(x)| f(x) \, \lambda(dx) < \infty, & \text{falls } f = dPX^{-1}/d\lambda. \end{cases}$$

Beispiel 2.35.

(a) Ist X Cauchy-verteilt zum Parameter $c > 0$ (siehe Beispiel 2.7(d)), so gilt

$$\int |x| f(x) \, \lambda(dx) = \int \frac{c}{\pi} \frac{|x|}{c^2 + x^2} \, \lambda(dx) = \infty,$$

das heißt $X \notin \mathcal{L}^1$.

(b) Ist X standard-normalverteilt (siehe Beispiel 2.7(b), so gelten

$$\frac{1}{\sqrt{2\pi}} \int |x| e^{-x^2/2} \, \lambda(dx) < \infty \quad \text{und} \quad EX = \frac{1}{\sqrt{2\pi}} \int x e^{-x^2/2} \, \lambda(dx) = 0.$$

(c) Ist X normalverteilt mit Mittelwert a und Varianz σ^2 (siehe Beispiel 2.7(c)), so ergibt sich $X \in \mathcal{L}^1$ und $EX = a$.

Das folgende Lemma ergibt sich aus den Eigenschaften des Integrals aus Kapitel 1.

Lemma 2.36.

(a) *Sind $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ und $a, b \in \mathbb{R}$, so gelten $aX + bY \in \mathcal{L}^1$ und $E(aX + bY) = aEX + bEY$.*

(b) *Ist $X \geq 0$ mit $EX = 0$, so folgt $X = 0$ P-fast sicher.*

Definition 2.37. Ist X eine Zufallsgröße und $p > 0$, so ist $|X|^p$ eine nichtnegative Zufallsgröße. Es bezeichne $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ die Menge der auf $(\Omega, \mathcal{F}, P)$ definierten Zufallsgrößen mit $E|X|^p < \infty$.

Lemma 2.38. *Für $p \geq p' > 0$ gilt $\mathcal{L}^p \subset \mathcal{L}^{p'}$.*

Beweis. Sei $X \in \mathcal{L}^p$ und $A = \{|X| \leq 1\}$. Dann gilt

$$E|X|^{p'} = \int_A |X|^{p'} dP + \int_{A^c} |X|^{p'} dP \leq P(A) + \int_{A^c} |X|^p dP \leq 1 + E|X|^p < \infty.$$

□

Satz 2.39 (MARKOFF-UNGLEICHUNG). *Sei $X \in \mathcal{L}^p$ mit $p > 0$. Dann gilt für alle $a > 0$ die Abschätzung $P(|X| \geq a) \leq a^{-p} E|X|^p$.*

Beweis. Sei $A = \{|X| \geq a\}$. Dann gilt $1_A \leq a^{-p}|X|^p$ und somit $P(A) = E(1_A) \leq a^{-p} E|X|^p$. □

Für $p = 2$ nennt man diese Ungleichung auch *Tschebyscheff-Ungleichung.* Sie ist im allgemeinen sehr grob, wie schon in „Einführung in die Stochastik" gezeigt wurde.

Satz 2.40 (SCHWARZSCHE UNGLEICHUNG). *Sind $X, Y \in \mathcal{L}^2$, so gelten $X \cdot Y \in \mathcal{L}^1$ und $E|X \cdot Y| \leq (E(X^2)E(Y^2))^{1/2}$.*

Beweis. Siehe „Einführung in die Stochastik".

Ohne Beweis sei hier auch die allgemeinere Höldersche Ungleichung erwähnt.

Satz 2.41 (HÖLDERSCHE UNGLEICHUNG). *Seien $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Für $X \in \mathcal{L}^p$, $Y \in \mathcal{L}^q$ gelten $X \cdot Y \in \mathcal{L}^1$ und $E|X \cdot Y| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$.*

Definition 2.42. Sei $X \in \mathcal{L}^1$. Die *Varianz* von X ist definiert durch

$$\operatorname{var}(X) = E((X - EX)^2) \in [0, \infty].$$

Lemma 2.43. *Sei $X \in \mathcal{L}^1$.*

(a) $\operatorname{var}(X) = E(X^2 - 2XEX + (EX)^2) = E(X^2) - (EX)^2$.
(b) *Es gilt offenbar* $\operatorname{var}(X) < \infty \Leftrightarrow X \in \mathcal{L}^2$.
(c) $\operatorname{var}(X) = 0 \Leftrightarrow X = EX$ *fast sicher (nach 2.36 (b)).*
(d) *Die Tschebyscheff-Ungleichung angewandt auf $X - EX$ ergibt:*

$$P(|X - EX| \geq a) \leq \frac{1}{a^2} \operatorname{var}(X).$$

2c. Mehrdimensionale Zufallsgrößen

Definition 2.44. Ist $X = (X_1, \ldots, X_n)$ ein Zufallsvektor, so definiert man $EX \in \mathbb{R}^n$ komponentenweise durch $EX = (EX_1, \ldots, EX_n)$ (falls dies existiert).

An die Stelle der Varianz tritt der Begriff der Kovarianz:

Definition 2.45.

(a) Sind X und Y zwei Zufallsgrößen aus $\mathcal{L}^1$ mit $X \cdot Y \in \mathcal{L}^1$, so ist die *Kovarianz* $\operatorname{cov}(X,Y)$ definiert durch

$$\operatorname{cov}(X,Y) = E(X \cdot Y) - E(X)E(Y) = E((X-EX)(Y-EY)).$$

(b) Ist X ein Zufallsvektor mit $X_i \in \mathcal{L}^1$ und $X_iX_j \in \mathcal{L}^1$ für alle i und j, so ist die *Kovarianzmatrix* $\Sigma(X) = (\sigma_{ij}(X))$ definiert durch $\sigma_{ij}(X) = \operatorname{cov}(X_i, X_j)$.

Bemerkung 2.46.

(a) Sind $X, Y \in \mathcal{L}^2$, so ist die Kovarianz $\operatorname{cov}(X,Y)$ definiert (siehe Satz 2.40 und Lemma 2.38).
(b) $\operatorname{cov}(X,X) = \operatorname{var}(X)$.
(c) Die Kovarianzmatrix ist symmetrisch und positiv semidefinit.
(d) Wird X als Spaltenvektor geschrieben, so gilt $\Sigma(X) = E((X-EX)(X-EX)^t)$.

Beweis von Bemerkung 2.46(c). Für $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ gilt

$$0 \le E\Big(\Big(\sum_{i=1}^n \lambda_i(X_i - EX_i)\Big)^2\Big) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i\lambda_j \operatorname{cov}(X_i, X_j).$$

$\square$

Beispiel 2.47 (ERWARTUNGSWERT UND KOVARIANZ DER NORMALVERTEILUNG).

(a) Sei X standard-normalverteilt, das heißt, die Verteilung μ von X auf $\mathbb{R}^n$ hat die Dichte $(2\pi)^{-n/2} \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big)$. Dann gilt

$$EX_i = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i \exp\Big(-\frac{1}{2}\sum_{j=1}^n x_j^2\Big)\lambda_n(dx) = 0.$$

Für $i \neq j$ gelten

$$E(X_iX_j) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_ix_j \exp\Big(-\frac{1}{2}\sum_j x_j^2\Big)\, \lambda_n(dx) = 0,$$

$$EX_i^2 = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i^2 \exp\Big(-\frac{1}{2}\sum_{j=1}^n x_j^2\Big)\, \lambda_n(dx)$$

$$= (2\pi)^{-1/2} \int_{\mathbb{R}^n} x^2 e^{-x^2/2}\, \lambda(dx) = 1,$$

das heißt, $\Sigma(X)$ ist die Einheitsmatrix.

(b) Ist $Y = AX + b$, wobei A eine $m \times n$-Matrix und $b \in \mathbb{R}^m$ sind (Y, X, b werden als Spaltenvektoren betrachtet), so gelten $EY = b$ und

$$\Sigma(Y) = E((Y-b)(Y-b)^t) = E(AXX^tA^t) = AE(XX^t)A^t = AA^t.$$

2d. Charakteristische Funktionen

Definition 2.48. Sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathcal{B}^n)$. Die *charakteristische Funktion* $\hat{\mu}$ von μ ist die Abbildung $\mathbb{R}^n \to \mathbb{C}$, die durch

$$\hat{\mu}(t) = \int e^{i\langle t,x\rangle}\mu(dx) = \int \cos(\langle t,x\rangle)\mu(dx) + i\int \sin(\langle t,x\rangle)\mu(dx)$$

definiert wird. Dabei ist $\langle t,x\rangle = \sum_{i=1}^n t_i x_i$. Die charakteristische Funktion eines Zufallsvektors X ist die charakteristische Funktion der Verteilung von X; sie kann nach Satz 1.49 als $E(\exp(i\langle t,X\rangle))$ geschrieben werden.

Die charakteristische Funktion ist offenbar für alle $t \in \mathbb{R}^n$ definiert (da sin und cos beschränkt sind) und stetig in t (nach dem beschränkten Konvergenzsatz 1.40). Die Theorie charakteristischer Funktionen soll hier nur rudimentär behandelt werden. Als einziges Ergebnis werden wir das folgende benötigen:

Satz 2.49. *Es seien μ, ν zwei Wahrscheinlichkeitsmaße 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. Da die Familie der kompakten Mengen in $\mathbb{R}^n$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}^n$ ist, genügt es nach Bemerkung 1.9 nachzuweisen, daß $\mu(K) = \nu(K)$ für alle kompakten Mengen K gilt. Für eine derartige Menge K und $m \in \mathbb{N}$ sei

$$f_m(x) = \begin{cases} 1, & \text{falls } x \in K, \\ 0, & \text{falls } d(x,K) := \inf\{\,|x-y| : y \in K\,\} \geq 1/m, \\ 1 - md(x,K), & \text{sonst.} \end{cases}$$

Dann hat f_m die folgenden Eigenschaften:

(a) $0 \leq f_m(x) \leq 1$ für alle x,
(b) f_m ist stetig,
(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 $m \in \mathbb{N}$ gilt, so folgt $\mu(K) = \nu(K)$ mit dem beschränkten Konvergenzsatz aus (c). Es genügt also nachzuweisen, daß $\int f d\mu = \int f d\nu$ für alle f gilt, die (a), (b) erfüllen und kompakten Träger haben. (Der Träger von f ist die Menge $\overline{\{\,x : f(x) \neq 0\,\}}$.)
Sei $\varepsilon > 0$, und $N > 0$ sei so groß gewählt, daß

$$B_N := [-N,N] \times [-N,N] \times \cdots \times [-N,N] \supset \{\,x : f(x) \neq 0\,\}$$

und $\mu(B_N^c)$, $\nu(B_N^c) \leq \varepsilon$ gelten. Nach dem Weierstraßschen Approximationssatz gibt es eine Funktion $g : \mathbb{R}^n \to \mathbb{R}$ der Form $g(x) = \sum_{j=1}^m c_j \exp\big(i\langle \frac{\pi}{N} t_j, x\rangle\big)$ 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, das heißt, $\sup\{\,|f(x) - g(x)| : x \in B_N\,\} \leq \varepsilon$. Es folgen $\sup_{x\in\mathbb{R}^n} |g(x)| \leq 1 + \varepsilon$ und

$$\left|\int f\,d\mu - \int f\,d\nu\right| \leq \left|\int f\,d\mu - \int g\,d\mu\right| + \left|\int g\,d\mu - \int g\,d\nu\right| + \left|\int g\,d\nu - \int f\,d\nu\right|.$$

Der zweite Summand ist nach der Voraussetzung $\hat{\mu} = \hat{\nu}$ gleich null. Der erste Summand kann wegen $|g(x)| \leq 1 + \varepsilon$ für alle x folgendermaßen abgeschätzt werden:

$$\begin{aligned}\Big|\int f\,d\mu - \int g\,d\mu\Big| &\leq \Big|\int_{B_N} f\,d\mu - \int_{B_N} g\,d\mu\Big| + \Big|\int_{B_N^c} f\,d\mu\Big| + \Big|\int_{B_N^c} g\,d\mu\Big| \\ &\leq \int_{B_N} |f-g|\,d\mu + 0 + (1+\varepsilon)\mu(B_N^c) \\ &\leq \varepsilon\mu(B_N) + (1+\varepsilon)\mu(B_N^c) \\ &\leq \varepsilon(2+\varepsilon).\end{aligned}$$

Der dritte Summand wird analog abgeschätzt. Da $\varepsilon > 0$ beliebig war, folgt $\int f\,d\mu = \int f\,d\nu$. $\square$

Beispiel 2.50.

(a) Sei μ die Standardnormalverteilung. Dann gilt

$$\hat{\mu}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} dx.$$

Das Integral ergibt $\sqrt{2\pi}$ (einfache Übungsaufgabe aus der Funktionentheorie). Somit gilt $\hat{\mu}(t) = e^{-t^2/2}$.

(b) Sei μ die Cauchy-Verteilung zum Parameter $c > 0$. Dann gilt

$$\hat{\mu}(t) = \frac{c}{\pi} \int_{-\infty}^{\infty} e^{itx} \frac{dx}{c^2 + x^2}.$$

Die Funktion $\mathbb{C} \ni z \mapsto \frac{1}{c^2+z^2}$ hat Pole in $\pm ic$. Eine Anwendung des Residuensatzes ergibt somit $\hat{\mu}(t) = e^{-c|t|}$.

(c) Sei μ die Standardnormalverteilung in $(\mathbb{R}^n, \mathcal{B}^n)$. Dann folgt aus 2.50(a)

$$\hat{\mu}(t) = \exp\Big(-\frac{1}{2}\sum_{j=1}^{n} t_j^2\Big) \quad \text{für alle } t = (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

(d) Die allgemeine Normalverteilung ist das Bildmaß $\nu = \mu\varphi^{-1}$ der Standardnormalverteilung μ unter einer affinen Transformation $\mathbb{R}^n \ni x \mapsto \varphi(x) = Ax + b$. Bezeichnet A' die Transponierte von A, so gilt

$$\begin{aligned}\hat{\nu}(t) &= \int e^{i\langle t,x\rangle}\,\nu(dx) = \int e^{i\langle t,\varphi(x)\rangle}\,\mu(dx) = e^{i\langle t,b\rangle} \int e^{i\langle A't,x\rangle}\,\mu(dx) \\ &= e^{i\langle t,b\rangle}\hat{\mu}(A't) = e^{i\langle t,b\rangle} e^{-\langle A't,A't\rangle/2} = \exp\big(i\langle t,b\rangle - \tfrac{1}{2}\langle t,\Sigma t\rangle\big),\end{aligned}$$

mit $\Sigma = AA'$ als der Kovarianzmatrix von ν (siehe Beispiel 2.47(b)).

Es ergibt sich aus Beispiel 2.50(d):

Satz 2.51. *Für jedes $b \in \mathbb{R}^n$ und jede positiv semidefinite, symmetrische Matrix Σ gibt es genau eine Normalverteilung mit b als Erwartungswert und Σ als Kovarianzmatrix.*

Beweis. Die Eindeutigkeit folgt aus Satz 2.49 und der Rechnung in 2.50(d). Die Existenz folgt sofort daraus, daß mindestens eine $n \times n$-Matrix A existiert mit $AA' = \Sigma$. □

2e. Konvergenz von Folgen von Zufallsgrößen

Im folgenden sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen, das heißt ein stochastischer Prozeß. In der Wahrscheinlichkeitstheorie sind drei Konvergenzbegriffe besonders wichtig.

Definition 2.52.

(a) Die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert fast sicher* gegen eine Zufallsgröße X, falls

$$P(\{\,\omega\colon \lim_{n\to\infty} X_n(\omega) = X(\omega)\,\}) = 1$$

gilt (*Notation:* $X_n \to X$ P-fast sicher).

(b) Die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert im p-ten Mittel* ($p > 0$) gegen eine Zufallsgröße X, falls X_n, $X \in \mathcal{L}^p$ und

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

(c) Die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert in Wahrscheinlichkeit* gegen eine Zufallsgröße X, falls

$$P(|X_n - X| \geq \varepsilon) \to 0 \quad \text{für } n \to \infty$$

für alle $\varepsilon > 0$ gilt.

Satz 2.53.

(a) *Fast sichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit.*
(b) *Konvergenz im p-ten Mittel impliziert Konvergenz in Wahrscheinlichkeit.*

Beweis. Der Beweis von (b) folgt sofort aus der Markoff-Ungleichung.

(a) Sei $Y_n = 1_{\{|X_n - X| \geq \varepsilon\}}$ für $\varepsilon > 0$. Gilt $X_n \to X$ fast sicher, so gilt $Y_n \to 0$ fast sicher. Wegen $|Y_n| \leq 1$ folgt aus dem beschränkten Konvergenzsatz

$$P(|X_n - X| \geq \varepsilon) = E(Y_n) \to 0.$$

□

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

Beispiel 2.54. Sei $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}_{[0,1]}, \lambda)$.

(a) Wähle $X_n = n^{1/p} 1_{[0,1/n]}$ für $p > 0$. Dann gilt $X_n \to 0$ fast sicher und in Wahrscheinlichkeit, aber $E|X_n|^p = 1$ für alle n, das heißt, X_n konvergiert nicht im p-ten Mittel gegen null.

(b) Ist $n = 2^m + k$ für $m \in \mathbb{N}_0$ und $0 \leq k < 2^m$, so setzt man $X_n = 1_{[k2^{-m},(k+1)2^{-m}]}$. Offenbar konvergiert die Folge $(X_n(\omega))_{n \in \mathbb{N}}$ für kein $\omega \in [0,1]$. Andererseits gelten $P(|X_n| \geq \varepsilon) \leq 2^{-m}$ für alle $\varepsilon > 0$ und $E|X_n|^p = 2^{-m}$ für $p > 0$, das heißt $X_n \to 0$ in Wahrscheinlichkeit und im p-ten Mittel.

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

Satz 2.55. *Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsgrößen, die fast sicher gegen X konvergiert. Gilt $|X_n| \leq Y$ fast sicher für eine Zufallsgröße $Y \in \mathcal{L}^p$ (für $p > 0$), so gilt $X_n \to X$ im p-ten Mittel.*

Beweis. Es gelten $|X_n - X|^p \leq (|X_n| + |X|)^p \leq (2Y)^p \leq 2^p Y^p \in \mathcal{L}^1$ und $|X_n - X|^p \to 0$ fast sicher. Daher folgt aus dem beschränkten Konvergenzsatz $E|X_n - X|^p \to 0$. □

Wie aus Beispiel 2.54(b) hervorgeht, folgt aus der Konvergenz in Wahrscheinlichkeit nicht die fast sichere Konvergenz. Es gilt aber der folgende Satz:

Satz 2.56. *Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsgrößen, die in Wahrscheinlichkeit gegen X konvergiert. Dann existiert eine Teilfolge $(X_{n_k})_{k \in \mathbb{N}}$ mit $\lim_{k \to \infty} X_{n_k} = X$ fast sicher.*

Zum Beweis benötigt man das folgende sehr einfache, aber wichtige Lemma.

Lemma 2.57 (1. BOREL-CANTELLI-LEMMA). *Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge von Ereignissen mit $\sum_{n=1}^{\infty} P(A_n) < \infty$. Dann gilt $P(\limsup_{n \to \infty} A_n) = 0$.*

Beweis. Aus $B_k := \bigcup_{n=k}^{\infty} A_n \downarrow \limsup_{n \to \infty} A_n$ folgt

$$P(\limsup_{n \to \infty} A_n) = \lim_{k \to \infty} P(B_k) \leq \lim_{k \to \infty} \sum_{n=k}^{\infty} P(A_n) = 0.$$

□

Beweis von Satz 2.56. Zu jedem $k \in \mathbb{N}$ existiert nach Voraussetzung 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 $\sum_{k=1}^{\infty} 1/k^2 < \infty$ gilt, folgt $P(\limsup_{k \to \infty} \{ |X_{n_k} - X| \geq 1/k \}) = 0$ aus Lemma 2.57. Für $\omega \notin \limsup_{k \to \infty} \{ |X_{n_k} - X| \geq 1/k \}$ gilt $|X_{n_k}(\omega) - X(\omega)| < 1/k$ für genügend große k, das heißt $\lim_{k \to \infty} X_{n_k}(\omega) = X(\omega)$. □

Bemerkung 2.58. Alle drei Konvergenztypen sind *vollständig,* das heißt, daß jede Cauchy-Folge konvergiert. Für die fast sichere Konvergenz ist das klar, denn wenn $X_n - X_m \to 0$ fast sicher für $n, m \to \infty$ gilt, dann folgt aufgrund der Vollständigkeit von $\mathbb{R}$, daß $(X_n(\omega))_{n\in\mathbb{N}}$ für fast alle $\omega \in \Omega$ konvergiert. Aus Lemma 2.57 folgt das Entsprechende für die Konvergenz in Wahrscheinlichkeit:

Satz 2.59. *Sei $(X_n)_{n\in\mathbb{N}}$ 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 Wahrscheinlichkeit.

Beweis. Wähle wie im Beweis von Satz 2.56 eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit

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

Aus dem 1. Borel-Cantelli-Lemma folgt

$$P(\limsup_{k\to\infty}\{\, |X_{n_k} - X_{n_{k+1}}| \geq 1/k^2 \,\}) = 0.$$

Für $\omega \notin \limsup_{k\to\infty}\{\, |X_{n_k} - X_{n_{k+1}}| \geq 1/k^2 \,\}$ ist $(X_{n_k}(\omega))_{k\in\mathbb{N}}$ offenbar eine Cauchy-Folge in $\mathbb{R}$, das heißt, X_{n_k} konvergiert für $k \to \infty$ fast sicher gegen eine Zufallsgröße X, also nach Satz 2.53(a) auch in Wahrscheinlichkeit. Für $\varepsilon > 0$ gilt

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

für alle m und k. Wählt man k als die kleinste Zahl mit $n_k \geq m$, dann folgt

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

□

Für die Konvergenz im p-ten Mittel gilt die Vollständigkeit auch, soll aber hier nicht bewiesen werden.

Kapitel 3

Unabhängigkeit

Es sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum. Im folgenden wird von Familien von Teilmengen von Ω *stets* stillschweigend vorausgesetzt, daß sie Ω enthalten.

Definition 3.1.

(a) Teilmengen $\mathcal{E}_1, \dots, \mathcal{E}_n$ von $\mathcal{F}$ (mit $\Omega \in \mathcal{E}_i$!) heißen *unabhängig,* wenn für $A_i \in \mathcal{E}_i$, $1 \leq i \leq n$, gilt: $P(A_1 \cap \dots \cap A_n) = P(A_1) \cdots P(A_n)$.

(b) Seien eine Indexmenge und $\mathcal{E}_i$ für $i \in I$ Teilmengen von $\mathcal{F}$. Sie heißen *unabhängig,* wenn je endlich viele unabhängig sind.

Notation: Für zwei unabhängige Teilmengen $\mathcal{E}_1$, $\mathcal{E}_2$ von $\mathcal{F}$ schreiben wir $\mathcal{E}_1 \perp\!\!\!\perp \mathcal{E}_2$.

Bemerkung 3.2.

(a) 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.

(b) Gilt $\mathcal{D} \perp\!\!\!\perp \mathcal{E}_i$ für $i \in I$, so gilt $\mathcal{D} \perp\!\!\!\perp \bigcup_{i \in I} \mathcal{E}_i$.

Beweis. (a) ist klar.

(b) Für $A \in \mathcal{D}$ und $B \in \bigcup_{i \in I} \mathcal{E}_i$ existiert ein $i \in I$ mit $B \in \mathcal{E}_i$, das heißt, daß $P(A \cap B) = P(A)P(B)$ gilt. $\square$

Satz 3.3. *Es seien $\mathcal{D}_i$ für $i \in I$ unabhängige Teilmengen von $\mathcal{F}$ (stets $\Omega \in \mathcal{D}_i$). Sind die $\mathcal{D}_i$ durchschnittstabil, so sind die $\sigma(\mathcal{D}_i)$ für $i \in I$ unabhängig.*

Beweis. Es genügt, das zu zeigen, wenn I endlich ist. Sei etwa $I = \{1, \dots, n\}$. Es ist zu zeigen, daß

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^{n} P(A_i) \tag{3.4}$$

für $A_i \in \sigma(\mathcal{D}_i)$ gilt.

Für $0 \leq k \leq n$ sei L_k die folgende Aussage: Die Gleichung (3.4) gilt für alle $A_i \in \sigma(\mathcal{D}_i)$ für $i \leq k$ und für alle $A_i \in \mathcal{D}_i$ für $i > k$.

Die Aussage L_0 gilt wegen der Unabhängigkeit der $\mathcal{D}_i$. Wir zeigen $L_k \Rightarrow L_{k+1}$ für $0 \leq k \leq n-1$.

Sei

$$\begin{aligned}\mathcal{A} = \{\, B \in \sigma(\mathcal{D}_{k+1}) : {} & P(A_1 \cap \cdots \cap A_k \cap B \cap A_{k+2} \cap \cdots \cap A_n) \\ & = P(A_1) \cdots P(A_k) P(B) P(A_{k+2}) \cdots P(A_n) \\ & \text{für alle } A_i \in \sigma(\mathcal{D}_i),\ i \le k, \text{ und alle } A_i \in \mathcal{D}_i,\ \ i \ge k+2 \,\}.\end{aligned}$$

Aus L_k folgt $\mathcal{A} \supset \mathcal{D}_{k+1}$. Wir zeigen, daß $\mathcal{A}$ ein Dynkin-System ist (siehe Definition 1.3).

(a) $\Omega \in \mathcal{A}$ gilt wegen $\Omega \in \mathcal{D}_{k+1}$.

(b) Für $D \in \mathcal{A}$ gilt

$$\begin{aligned} & P(A_1 \cap \cdots \cap A_k \cap D^c \cap A_{k+2} \cap \cdots \cap A_n) \\ &\qquad = P(A_1 \cap \cdots \cap A_k \cap A_{k+2} \cap \cdots \cap A_n) \\ &\qquad\qquad - P(A_1 \cap \cdots \cap A_k \cap D \cap A_{k+2} \cap \cdots \cap A_n) \\ &\qquad = P(A_1)P(A_2) \cdots P(A_k) P(A_{k+2}) \cdots P(A_n) \\ &\qquad\qquad - P(A_1) \cdots P(A_k) P(D) P(A_{k+2}) \cdots P(A_n) \\ &\qquad = P(A_1) \cdots P(A_k) P(D^c) P(A_{k+2}) \cdots P(A_n) \end{aligned}$$

für alle A_i gemäß den obigen Bedingungen, das heißt $D^c \in \mathcal{A}$.

(c) Für paarweise disjunkte $D_i \in \mathcal{A}$, $i \in \mathbb{N}$, folgt analog $\bigcup_{i=1}^{\infty} D_i \in \mathcal{A}$.

Nach Satz 1.4 folgt $\mathcal{A} = \sigma(\mathcal{D}_{k+1})$, das heißt, L_{k+1} gilt. □

Definition 3.5. Die Ereignisse A_i für $i \in I$ heißen *unabhängig,* wenn die Mengensysteme $\{A_i, \Omega\}$, $i \in I$, unabhängig sind.

Bemerkung 3.6.

(a) Ereignisse A_i für $i \in I$ sind genau dann unabhängig, wenn $P(A_{i_1} \cap \cdots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_n})$ für jede Auswahl $i_1, \ldots, i_n \in I$.

(b) Da das Mengensystem $\{A, \Omega\}$ durchschnittstabil ist, folgt, wenn die Ereignisse A_i für $i \in I$ unabhängig sind, daß auch die σ-Algebren $\{\emptyset, A_i, A_i^c, \Omega\}$ unabhängig sind; insbesondere dann auch die Komplemente A_i^c.

Lemma 3.7. *Es seien $\mathcal{D}_i \subset \mathcal{F}$ für $i \in I$ unabhängig und durchschnittstabil. Es sei $(I_k)_{k \in K}$ eine Familie von paarweise disjunkten Teilmengen von I. Dann sind die $\sigma\left(\bigcup_{j \in I_k} \mathcal{D}_j\right)$ 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 $j \in I_k$. Das Mengensystem $\hat{\mathcal{D}}_k$ ist offenbar 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 paarweise disjunkten $j_1, \ldots, j_n \in I_k$. Daraus folgt sofort, daß die $\hat{\mathcal{D}}_k$ für $k \in K$ unabhängig sind. Da $\hat{\mathcal{D}}_k \supset \mathcal{D}_j$ für alle $j \in I_k$ ist, gilt $\sigma(\bigcup_{j \in I_k} \mathcal{D}_j) = \sigma(\hat{\mathcal{D}}_k)$. Das Lemma folgt nun aus Satz 3.3. □

Als Folgerung:

Satz 3.8 (KOLMOGOROFFSCHES 0-1-GESETZ). *Sei* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *eine Folge von unabhängigen Teil-*σ*-Algebren* $\mathcal{F}_n$ *von* $\mathcal{F}$. *Seien* $\check{\mathcal{F}}_n := \bigvee_{k=n}^{\infty} \mathcal{F}_k$ *und* $\mathcal{T}_\infty = \bigcap_{n=1}^{\infty} \check{\mathcal{F}}_n$. *Für* $A \in \mathcal{T}_\infty$ *gilt* $P(A) \in \{0,1\}$.

Beweis. Nach Lemma 3.7 gilt $\check{\mathcal{F}}_{n+1} \perp\!\!\!\perp \bigvee_{k=1}^{n} \mathcal{F}_k$, also nach Bemerkung 3.2(a): $\mathcal{T}_\infty \perp\!\!\!\perp \bigvee_{k=1}^{n} \mathcal{F}_k$ für alle $n \in \mathbb{N}$. Nach Bemerkung 3.2(b) folgt dann $\mathcal{T}_\infty \perp\!\!\!\perp \bigcup_{n=1}^{\infty} \bigvee_{k=1}^{n} \mathcal{F}_k$. Da die rechte Seite als Vereinigung einer aufsteigenden Folge von σ-Algebren durchschnittstabil ist, folgt nach Satz 3.3

$$\mathcal{T}_\infty \perp\!\!\!\perp \bigvee_{n=1}^{\infty} \mathcal{F}_n \Big(= \sigma\Big(\bigcup_{n=1}^{\infty} \bigvee_{k=1}^{n} \mathcal{F}_k \Big)\Big).$$

Nun ist aber $\check{\mathcal{F}}_n \subset \bigvee_{k=1}^{\infty} \mathcal{F}_k$ für alle $n \in \mathbb{N}$, also auch $\mathcal{T}_\infty \subset \bigvee_{n=1}^{\infty} \mathcal{F}_n$. Nach Bemerkung 3.2(a) folgt also $\mathcal{T}_\infty \perp\!\!\!\perp \mathcal{T}_\infty$, das heißt, für $A \in \mathcal{T}_\infty$ gilt $P(A) = P(A \cap A) = P(A)^2$, das heißt $P(A) \in \{0,1\}$. $\square$

Die Ereignisse in $\mathcal{T}_\infty$ werden *terminal* genannt. Der obige Satz ist vor allem für unabhängige Zufallsgrößen interessant.

Definition 3.9. Die X_i, $i \in I$, seien auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definierte $(E_i, \mathcal{E}_i)$-wertige Zufallsgrößen. (Die meßbaren Räume $(E_i, \mathcal{E}_i)$ können verschieden sein.) Die X_i heißen *unabhängig,* wenn die Teil-σ-Algebren $X_i^{-1}(\mathcal{E}_i)$ unabhängig sind.

Beachte: Die X_i müssen auf demselben Wahrscheinlichkeitsraum definiert sein, damit die Aussage einen Sinn hat.
Notation: Sind zwei Zufallsgrößen unabhängig, so schreiben wir $X \perp\!\!\!\perp Y$.

Bemerkung 3.10.

(a) Sind $\mathcal{D}_i \subset \mathcal{F}$ für $i \in I$ unabhängige Teilmengen und sind $(E_i, \mathcal{E}_i)$-wertige Zufallsgrößen X_i $\mathcal{D}_i$-$\mathcal{E}_i$-meßbar, so sind diese offenbar unabhängig.
(b) Sind X_i für $i \in I$ unabhängige $(E_i, \mathcal{E}_i)$-wertige Zufallsgrößen und sind $\varphi_i \colon E_i \to E_i'$ (mit meßbaren Räumen $(E_i', \mathcal{E}_i')$) $\mathcal{E}_i$-$\mathcal{E}_i'$-meßbare Abbildungen, so sind die $\varphi_i \circ X_i$ offenbar auch unabhängig (da $(\varphi_i \circ X_i)^{-1}(\mathcal{E}_i') = X_i^{-1}(\varphi_i^{-1}(\mathcal{E}_i')) \subset X_i^{-1}(\mathcal{E}_i)$ ist).
(c) Die X_i sind genau dann unabhängig, wenn für $n \in \mathbb{N}$ und $i_1, \ldots, i_n \in I$ sowie $A_1 \in \mathcal{E}_{i_1}, \ldots, A_n \in \mathcal{E}_{i_n}$ die Gleichung

$$P(X_{i_1} \in A_1, \ldots, X_{i_n} \in A_n) = \prod_{j=1}^{n} P(X_{i_j} \in A_j)$$

gilt. Man sieht daraus, daß für eine Folge $(X_n)_{n\in\mathbb{N}}$ von $(E, \mathcal{E})$-wertigen Zufallsgrößen diese genau dann unabhängig sind, wenn die Verteilung von $(X_n)_{n\in\mathbb{N}}$ die Produktwahrscheinlichkeit der Verteilungen der X_n ist.

Im folgenden betrachten wir $(\mathbb{R}, \mathcal{B})$-wertige Zufallsgrößen.

Lemma 3.11. *Zufallsgrößen X_i, $i \in I$, sind genau dann unabhängig, wenn für alle $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ und $t_1, \ldots, t_n \in \mathbb{R}$*

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

gilt.

Beweis. $\{ X_i^{-1}((-\infty, t]) : t \in \mathbb{R} \} \cup \{\Omega\}$ ist ein durchschnittstabiles Erzeugendensystem von $X_i^{-1}(\mathcal{B})$. Die Behauptung folgt aus Satz 3.3. $\square$

Eine andere Formulierung der Unabhängigkeit ist die folgende:

Satz 3.12. *Es seien X, Y zwei unabhängige reelle Zufallsgrößen.*

(a) *Falls X und Y nichtnegativ sind, so gilt $E(XY) = E(X)E(Y)$.*
(b) *Sind $X, Y \in \mathcal{L}^1$, so gilt $XY \in \mathcal{L}^1$ und $E(XY) = E(X)E(Y)$.*

Beweis. (a) Es seien $\mathcal{F}_1 := X^{-1}(\mathcal{B})$ und $\mathcal{F}_2 := Y^{-1}(\mathcal{B})$. Dann gilt $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2$. Für $A \in \mathcal{F}_1$ erfüllt die Menge der nichtnegativen $\mathcal{F}_2$-meßbaren Zufallsgrößen Y' mit $E(1_A Y') = P(A)E(Y')$ die Eigenschaften (a)–(c) von Satz 1.28. Demzufolge gilt diese Gleichung für alle nichtnegativen $\mathcal{F}_2$-meßbaren Y', also insbesondere für Y selbst. Die Menge der nichtnegativen $\mathcal{F}_1$-meßbaren Zufallsgrößen X' mit $E(X'Y) = E(X')E(Y)$ erfüllt ebenfalls die Bedingungen von Satz 1.28. Das gleiche Argument wie oben belegt, daß diese Gleichung für $X' = X$ gilt.

(b) Aus $X \perp\!\!\!\perp Y$ folgt $|X| \perp\!\!\!\perp |Y|$. Somit folgt aus (a), daß gilt: $E(|XY|) = E(|X|)E(|Y|) < \infty$, wenn $X, Y \in \mathcal{L}^1$ sind, das heißt $XY \in \mathcal{L}^1$. Die Gleichung $E(XY) = E(X)E(Y)$ folgt, indem X und Y in Positiv- und Negativteil zerlegt werden. $\square$

Besonders nützlich für die Untersuchung von unabhängigen Zufallsgrößen sind die charakteristischen Funktionen.

Satz 3.13. *Es seien X, Y zwei unabhängige Zufallsgrößen mit charakteristischen Funktionen $\mathcal{X}_X$ beziehungsweise $\mathcal{X}_Y$. Dann ist $\mathcal{X}_X \cdot \mathcal{X}_Y$ die charakteristische Funktion von $X + Y$.*

Beweis. Für alle $t \in \mathbb{R}$ gilt

$$E(e^{it(X+Y)}) = E(e^{itX} e^{itY}) = E(e^{itX})E(e^{itY}),$$

da $e^{itX} \perp\!\!\!\perp e^{itY}$ gilt. Der Beweis ist insofern unvollständig, als Satz 3.12 nur für reellwertige Zufallsgrößen bewiesen wurde. Eine Zerlegung in Real- und Imaginärteil liefert jedoch sofort die entsprechende Aussage für komplexwertige Zufallsgrößen. $\square$

Aus Satz 3.13 und Satz 2.49 folgt, daß die Verteilung von $X+Y$ für unabhängige Zufallsgrößen X, Y nur von den Verteilungen von X und Y abhängt. Man nennt diese Verteilung auch die *Faltung* der einzelnen Verteilungen.

Die Verteilungsfunktion von $X+Y$ kann auch wie folgt berechnet werden: Seien $\mu = PX^{-1}$, $\nu = PY^{-1}$. Dann gilt:

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

mit $f_t(x,y) = 1_{\{x+y \leq t\}} = 1_{(-\infty,t-y]}(x)$, da $\mu \otimes \nu$ die Verteilung von (X,Y) ist. Nach dem Satz von Fubini gibt das:

$$= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} 1_{(-\infty,t-y]}(x)\mu(dx)\Big)\nu(dy) = \int_{\mathbb{R}} F(t-y)\nu(dy),$$

wobei F die Verteilungsfunktion von X ist. Gilt ferner $\mu \ll \lambda$, $\nu \ll \lambda$, $f = d\mu/d\lambda$, $g = d\nu/d\lambda$, so gibt das weiter:

$$\begin{aligned} &= \int_{\mathbb{R}} \Big(\int_{(-\infty,t-y]} f(x)dx\Big) g(y)dy = \int_{\mathbb{R}} \Big(\int_{(-\infty,t]} f(x-y)dx\Big) g(y)dy \\ &= \int_{(-\infty,t]} \Big(\int_{\mathbb{R}} f(x-y)g(y)dy\Big) dx. \end{aligned}$$

Demzufolge hat dann auch die Verteilung von $X+Y$ eine Dichte bezüglich λ, nämlich die Abbildung

$$x \mapsto \int_{\mathbb{R}} f(x-y)g(y)dy = \int_{\mathbb{R}} f(y)g(x-y)dy.$$

Charakteristische Funktionen sind für die Berechnung jedoch oft einfacher.

Beispiel 3.14.

(a) *Cauchy-Verteilung:*
Behauptung: Sind X, Y unabhängig und Cauchy-verteilt, so ist für $\lambda \in (0,1)$ die Zufallsgröße $\lambda X + (1-\lambda)Y$ auch Cauchy-verteilt.
Beweis: Für $\lambda \in (0,1)$ und $t \in \mathbb{R}$ gilt:

$$\begin{aligned} \mathcal{X}_{\lambda X+(1-\lambda)Y}(t) &= E(\exp(it(\lambda X + (1-\lambda)Y))) = \mathcal{X}_X(\lambda t) \cdot \mathcal{X}_Y((1-\lambda)t) \\ &= \exp(-|\lambda t|)\exp(-|(1-\lambda)t|) = e^{-|t|}. \end{aligned}$$

□

(b) *Normalverteilung:*
Behauptung: Ist X normalverteilt mit Mittelwert a und Varianz σ^2, Y normalverteilt mit Parametern a', σ'^2, und gilt $X \perp\!\!\!\perp Y$, so ist $X+Y$ normalverteilt mit Parametern $a+a'$ und $\sigma^2+\sigma'^2$.
Beweis: Für $t \in \mathbb{R}$ gilt:

$$\begin{aligned} \mathcal{X}_{X+Y}(t) = \mathcal{X}_X(t)\mathcal{X}_Y(t) &= \exp\Big(iat - \frac{\sigma^2}{2}t^2\Big) \cdot \exp\Big(ia't - \frac{\sigma'^2}{2}t^2\Big) \\ &= \exp\Big(i(a+a')t - \frac{1}{2}(\sigma^2+\sigma'^2)t^2\Big). \end{aligned}$$

□

Das Kolmogoroffsche $0-1$-Gesetz (Satz 3.8) soll nun auf unabhängige Zufallsgrößen angewandt werden. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrößen, und $\mathcal{F}_n := X_n^{-1}(\mathcal{B})$. Es sei $\mathcal{T}_\infty$ die terminale σ-Algebra der $\mathcal{F}_n$ wie in Satz 3.8. Mit S_n sei $\sum_{j=1}^n X_j$ bezeichnet.

Lemma 3.15. *Sei $(a_n)_{n\in\mathbb{N}}$ eine Folge positiver Zahlen mit $a_n \to \infty$. Dann sind $\overline{Y} := \limsup_{n\to\infty} \frac{S_n}{a_n}$ und $\underline{Y} := \liminf_{n\to\infty} \frac{S_n}{a_n}$ $\mathcal{T}_\infty$-meßbare $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$-wertige Zufallsgrößen.*

Beweis. Für jedes $m \in \mathbb{N}$ gilt

$$\overline{Y} = \limsup_{n\to\infty} \frac{S_n}{a_n} = \limsup_{n\to\infty} \frac{1}{a_n}\Big(\sum_{j=1}^m X_j + \sum_{j=m+1}^n X_j\Big) = \limsup_{n\to\infty} \frac{1}{a_n}\Big(\sum_{j=m+1}^n X_j\Big),$$

was offenbar $\check{\mathcal{F}}_m$-meßbar ist. Daher ist $\overline{Y}$ $\mathcal{T}_\infty$-meßbar. Für $\underline{Y}$ geht der Beweis gleich. □

Zufallsgrößen, die meßbar bezüglich einer σ-Algebra sind, die die Eigenschaft aus Satz 3.8 hat, sind fast sicher konstant:

Lemma 3.16. *Es sei $\mathcal{T} \subset \mathcal{F}$ eine σ-Algebra mit $P(A) \in \{0,1\}$ für alle $A \in \mathcal{T}$. Ist Z eine $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$-wertige, $\mathcal{T}$-meßbare Zufallsgröße , so existiert ein $c \in \overline{\mathbb{R}}$ mit $P(Z = c) = 1$.*

Beweis. Sei $F(t) = P(Z \le t) \in \{0,1\}$ für $t \in \mathbb{R}$. Die Funktion F ist nichtfallend. Demzufolge sind drei Fälle möglich:

(a) $F(t) = 0$ für alle $t \in \mathbb{R} \Rightarrow P(Z > n) = 1$ für alle $n \Rightarrow P(Z = \infty) = 1$.
(b) $F(t) = 1$ für alle $t \in \mathbb{R} \Rightarrow P(Z \le n) = 1$ für alle $n \Rightarrow P(Z = -\infty) = 1$.
(c) F springt an einer Stelle $t_0 \in \mathbb{R}$ von 0 nach 1. Dann gilt

$$F\Big(t_0 + \frac{1}{n}\Big) - F\Big(t_0 - \frac{1}{n}\Big) = P\Big(Z \in \Big(t_0 - \frac{1}{n},\ t_0 + \frac{1}{n}\Big]\Big) = 1 \quad \text{für alle } n \in \mathbb{N}.$$

Somit ist dann $P(Z = t_0) = \lim_{n\to\infty} P\left(Z \in \left(t_0 - \frac{1}{n},\ t_0 + \frac{1}{n}\right]\right) = 1$. □

Aus Satz 3.8, Lemma 3.15 und Lemma 3.16 ergibt sich also:

Satz 3.17. *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsgrößen und $(a_n)_{n\in\mathbb{N}}$ eine positive Zahlenfolge mit $\lim_{n\to\infty} a_n = \infty$. Dann sind $\overline{Y} = \limsup_{n\to\infty} \frac{S_n}{a_n}$ und $\underline{Y} = \liminf_{n\to\infty} \frac{S_n}{a_n}$ fast sicher konstant in $[-\infty, \infty]$.*

Beispiel 3.18.

(a) Die X_n seien unabhängig und Cauchy-verteilt, $a_n = n$. Aus Beispiel 3.14(a) folgt sofort, daß S_n/n auch Cauchy-verteilt ist. Somit ist für $c \in \mathbb{R}$:

$$\begin{aligned}0 < \int_c^\infty \frac{1}{\pi}\frac{1}{1+x^2}dx = P\Big(\frac{S_n}{n} \ge c\Big) \\ \le P\Big(\sup_{k\ge n} \frac{S_k}{k} \ge c\Big) \overset{n\to\infty}{\to} P\Big(\limsup_{n\to\infty} \frac{S_n}{n} \ge c\Big) = P(\overline{Y} \ge c).\end{aligned}$$

Nach Lemma 3.15 und Satz 3.8 folgt $P(\overline{Y} \geq c) = 1$ für alle $c \in \mathbb{R}$, das heißt $P(\overline{Y} = \infty) = 1$. Analog zeigt man $P(\underline{Y} = -\infty) = 1$.

(b) Die Zufallsgrößen X_n seien unabhängig und standard-normalverteilt. Nach Beispiel 3.14(b) ist auch $S_n/\sqrt{n}$ standard-normalverteilt. Wie oben folgt dann, daß $\limsup_{n\to\infty} S_n/\sqrt{n} = \infty$ fast sicher und $\liminf_{n\to\infty} S_n/\sqrt{n} = -\infty$ fast sicher gilt.

Mit Hilfe des zentralen Grenzwertsatzes (siehe Kapitel 6 oder Einführung in die Stochastik) läßt sich leicht nachweisen, daß für alle Folgen von unabhängigen Zufallsgrößen X_n, die die gleiche Verteilung und positive und endliche Varianz haben, gelten: $\limsup_{n\to\infty} S_n/\sqrt{n} = +\infty$ und $\liminf_{n\to\infty} S_n/\sqrt{n} = -\infty$.

Aus dem Beispiel 3.18(a) ist ersichtlich, daß S_n/n im Cauchy-verteilten Fall nicht fast sicher konvergiert. Dies widerspricht nicht dem Gesetz der großen Zahlen, weil Cauchy-verteilte Zufallsgrößen keinen Erwartungswert besitzen.

Satz 3.19 (STARKES GESETZ DER GROSSEN ZAHLEN). *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsgrößen, die alle dieselbe Verteilung haben. Es gelte $X_i \in \mathcal{L}^1$ für alle $i \in \mathbb{N}$, und es sei $S_n = \sum_{i=1}^n X_i$. Dann gilt $\lim_{n\to\infty} S_n/n = EX_1$ fast sicher.*

Beweis. Ein vollständiger Beweis ergibt sich als Spezialfall des Ergodensatzes in Kapitel 4. Unter der Zusatzbedingung $X_i \in \mathcal{L}^4$ kann ein simpler Beweis wie folgt geführt werden: Sei $a = EX_n$ (unabhängig von n), $X_n' := X_n - a$. Es gilt $EX_n' = 0$, und $S_n/n \to a$ fast sicher gilt genau dann, wenn $\sum_{i=1}^n X_i'/n \to 0$ fast sicher gilt. Man kann also annehmen, daß $a = 0$ ist.

Sei $A_n = \left\{\left|\frac{S_n}{n}\right| \geq \frac{1}{n^{1/8}}\right\}$; $\overline{A} = \limsup_{n\to\infty} A_n$. Wir schätzen $P(A_n)$ mit der Markoff-Ungleichung (2.39 mit $p = 4$) ab:

$$P(A_n) \leq \frac{n^{1/2}}{n^4} E(S_n^4). \tag{3.20}$$

Nun ist

$$\begin{aligned}E(S_n^4) &= E\Big(\Big(\sum_{i=1}^n X_i\Big)^4\Big) = E\Big(\sum_{i_1,i_2,i_3,i_4=1}^n X_{i_1}X_{i_2}X_{i_3}X_{i_4}\Big)\\ &= \sum_{i_1,i_2,i_3,i_4=1}^n E(X_{i_1}X_{i_2}X_{i_3}X_{i_4}) = \sum_{i=1}^n E(X_i^4) + 3\sum_{i\neq j} E(X_i^2X_j^2) + \text{eine Summe}\end{aligned}$$

von Erwartungswerten, in denen mindestens eine der Zufallsgrößen $X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}$ als einzelner Faktor auftritt.

Eine Anwendung von Beispiel 3.12(b) ergibt, daß diese letzten Terme null sind. Somit folgt $E(S_n^4) = nE(X_1^4) + 3n(n-1)(E(X_1^2))^2$. Mit (3.20) zusammen folgt $\sum_{n=1}^\infty P(A_n) < \infty$, und aus dem 1. Borel-Cantelli-Lemma (Lemma 2.57) folgt $P(\overline{A}) = 0$. Für $\omega \notin \overline{A}$ konvergiert $S_n(\omega)/n$ offenbar gegen Null. □

Der Satz hat viele Verallgemeinerungen. Eine wichtige für nichtunabhängige Zufallsgrößen wird in Kapitel 4 diskutiert werden.

Ein Gesetz der großen Zahlen für nicht notwendig identisch verteilte Zufallsgrößen ist der folgende Satz, der hier nicht bewiesen werden soll.

Satz 3.21. *Seien X_n für $n \in \mathbb{N}$ unabhängige Zufallsgrößen in $\mathcal{L}^2$ mit $EX_n = 0$. Ist $(a_n)_n$ eine Folge positiver Zahlen mit $a_n \to \infty$ und gilt $\sum_{n=1}^{\infty} EX_n^2/a_n^2 < \infty$, so gilt $\lim_{n\to\infty} S_n/a_n = 0$ fast sicher.*

Angewandt auf identisch verteilte Zufallsgrößen und $a_n = n$ ergibt sich daraus die Aussage von Satz 3.19, allerdings unter der Voraussetzung $X_n \in \mathcal{L}^2$. Man kann jedoch in Satz 3.21 in diesem Fall auch $a_n = n^{\delta+\frac{1}{2}}$ oder $a_n = n^{1/2}(\log n)^{\delta+\frac{1}{2}}$ für $\delta > 0$ wählen, nicht jedoch $a_n = n^{1/2}(\log n)^{1/2}$ (da dann $\sum a_n^{-2}$ divergiert). Tatsächlich ist jedoch die Aussage von Satz 3.21 noch richtig für diese letzte Folge. Aus Beispiel 3.18(b) folgt, daß dies für $a_n = \sqrt{n}$ nicht mehr richtig ist. Wo liegt die Grenze? Das heißt, für welche Folgen $a_n \to \infty$ konvergiert S_n/a_n noch gegen Null fast sicher? Eine Antwort gibt das berühmte Gesetz vom iterierten Logarithmus.

Satz 3.22. *Es seien X_n für $n \in \mathbb{N}$ unabhängige Zufallsgrößen in $\mathcal{L}^2$, alle mit derselben Verteilung, und es gelte $EX_n = 0$. Dann gelten*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = \sigma \text{ fast sicher und}$$
$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = -\sigma \text{ fast sicher,}$$

wobei $\sigma^2 = \operatorname{var}(X_n)$ ist.

Der Beweis ist trickreich und soll hier nicht geführt werden. Zum Schluß noch eine partielle Umkehrung des 1. Borel-Cantelli-Lemmas aus Kapitel 2:

Lemma 3.23 (2. BOREL-CANTELLI-LEMMA). *Es seien A_n für $n \in \mathbb{N}$ unabhängige Ereignisse mit $\sum_{n=1}^{\infty} P(A_n) = \infty$. Dann gilt $P(\limsup_{n\to\infty} A_n) = 1$.*

Beweis. Für $n \in \mathbb{N}$ gilt

$$\begin{aligned} P\Big(\bigcup_{m=n}^{\infty} A_m\Big) &= 1 - P\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) = 1 - \lim_{k\to\infty} P\Big(\bigcap_{m=n}^{k} A_m^c\Big) \\ &= 1 - \prod_{m=n}^{\infty} P(A_m^c) = 1 - \Big(\prod_{m=n}^{\infty}(1 - P(A_m)\Big). \end{aligned}$$

Ein einfaches Lemma aus der Analysis besagt, daß wenn $\sum_{n=1}^{\infty} a_n$ divergiert für eine Folge von Zahlen $a_n \in [0,1]$, das unendliche Produkt $\prod_{n=1}^{\infty}(1-a_n) = 0$ ist. (Falls nicht bekannt: Übungsaufgabe.) Somit ist $P\left(\bigcup_{m=n}^{\infty} A_m\right) = 1$ für alle $n \in \mathbb{N}$, das heißt

$$P(\limsup_{n\to\infty} A_n) = P\Big(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\Big) = \lim_{n\to\infty} P\Big(\bigcup_{m=n}^{\infty} A_m\Big) = 1.$$

□

Bemerkung 3.24. Man kann auf die Voraussetzung der Unabhängigkeit in Lemma 3.23 nicht verzichten. Zum Beispiel gilt mit $A_n = A$ für alle $n \in \mathbb{N}$ und $0 < P(A) < 1$ natürlich $\Sigma_n P(A_n) = \infty$, aber $P(\limsup_{n\to\infty} A_n) = P(A) < 1$.

Das Lemma spielt bei vielen feineren Untersuchungen über Folgen von unabhängigen Zufallsgrößen eine große Rolle (zum Beispiel beim Beweis von Satz 3.22) Eine einfache Anwendung ist das folgende Ergebnis:

Satz 3.25. *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrößen, die alle die gleiche Verteilung haben. Sei $p > 0$. Dann gilt $X_n \in \mathcal{L}^p$ genau dann, wenn $P(\limsup_{n\to\infty}\{|X_n| > n^{1/p}\}) = 0$ ist.*

Beweis. Es gilt $X_n \in \mathcal{L}^p$ genau dann, wenn $|X_n|^p \in \mathcal{L}^1$ ist, und $\{|X_n| > n^{1/p}\}$ ist gleich $\{|X_n|^p > n\}$. Es genügt also, $p = 1$ zu betrachten.

Nach den beiden Borel-Cantelli-Lemmata (Lemma 3.23 und Lemma 2.57) gilt genau dann $P(\limsup_{n\to\infty}\{|X_n| > n\}) = 0$, wenn $\sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X_1| > n) < \infty$ gilt. Sei μ die Verteilung von $|X_n|$. Dann ist

$$\sum_{n=1}^{\infty} P(|X_1| > n) = \sum_{n=1}^{\infty} \mu((n,\infty)) = \sum_{n=1}^{\infty} n\,\mu((n,n+1]).$$

Dies ist einerseits $\leq \int_0^\infty x\,\mu(dx) = E|X_1|$ und andererseits $\geq -1 + \sum_{n=1}^{\infty}(n+1)\,\mu((n,n+1]) \geq -1 + \int_0^\infty x\,\mu(dx) = -1 + E|X_1|$. $\square$

Kapitel 4

Stationäre Prozesse

Im ganzen Kapitel 4 sei $(\Omega, \mathcal{F}, P)$ ein fester, aber beliebiger Wahrscheinlichkeitsraum.

Definition 4.1. Eine $\mathcal{F}$-$\mathcal{F}$-meßbare Abbildung $T\colon \Omega \to \Omega$ heißt *maßerhaltende Transformation,* wenn $PT^{-1} = P$ gilt. Wir sagen dafür auch, daß P T-*invariant* (oder invariant unter T) ist.

Mit Hilfe einer maßerhaltenden Transformation und einer Zufallsgröße kann man kanonisch einen stationären Prozeß definieren:

Lemma 4.2. *Es seien Y eine Zufallsgröße und T eine maßerhaltende Transformation auf $(\Omega, \mathcal{F}, P)$. Definiere $X_1 := Y$ und $X_{n+1} = X_n \circ T$ für $n \in \mathbb{N}$. Dann ist der Prozeß $(X_n)_n$ stationär.*

Beweis. Seien $n \in \mathbb{N}$ und $t_1, \dots, t_n \in \mathbb{R}$. Dann gilt:

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

Lemma 4.2 folgt nun aus Lemma 2.14. $\square$

Umgekehrt kann jeder stationäre Prozeß in gewisser Weise so beschrieben werden. Genauer: Zu jedem auf $(\Omega, \mathcal{F}, P)$ definierten, stationären Prozeß $\mathbb{X} = (X_n)_n$ existiert ein mit Hilfe einer maßerhaltenden Transformation beschriebener, der dieselbe Verteilung hat. Wie wir in Satz 2.11 gesehen haben, ist $\mathbb{X}$ nichts anderes als eine $\mathcal{F}$-$\mathcal{B}^{\mathbb{N}}$-meßbare Abbildung $\Omega \to \mathbb{R}^{\mathbb{N}}$. Sei $P' = P\mathbb{X}^{-1}$ die Verteilung von $\mathbb{X}$. Die sogenannte Verschiebungsabbildung $T\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, definiert durch $T((x_1, x_2, \dots)) = (x_2, x_3, \dots)$, ist maßerhaltend auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$: Für $A \in \mathcal{B}^{\mathbb{N}}$ gilt nämlich $P'(A) = P(\mathbb{X} \in A) = P(T \circ \mathbb{X} \in A) = P\mathbb{X}^{-1}T^{-1}(A) = P'T^{-1}(A)$ wegen der Stationarität von $\mathbb{X}$. Natürlich ist P' auch die Verteilung der Folge der Projektionen $\pi_1, \pi_2, \dots : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, denn $\Pi = (\pi_n)_{n \in \mathbb{N}}$ ist die identische Abbildung auf $\mathbb{R}^{\mathbb{N}}$. Offensichtlich gilt $\pi_n = \pi_{n-1} \circ T$ für $n \ge 2$. Der Prozeß $(\pi_n)_n$ auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ ist also von der in Lemma 4.2 definierten Form und besitzt dieselbe Verteilung wie $\mathbb{X}$.

Wir werden die in Definition 4.1 beschriebene Situation zugrundelegen. Alle gewonnenen Resultate, insbesondere der Ergodensatz (Satz 4.19), sind unmittelbar auf stationäre Folgen von Zufallsgrößen übertragbar.

Bemerkung 4.3. Eine meßbare Abbildung $T\colon \Omega \to \Omega$ ist genau dann maßerhaltend, wenn $PT^{-1}(A) = P(A)$ für alle A aus einem durchschnittstabilen Erzeugendensystem von $\mathcal{F}$ gilt. Dies folgt sofort aus Bemerkung 1.9.

Beispiel 4.4.

(a) Eine Folge von unabhängigen identisch verteilten Zufallsgrößen ist ein stationärer Prozeß.

(b) Sei $\Omega = D := \{\, z \in \mathbb{C}\colon |z| = 1 \,\}$. Wir können D als Kreis mit Radius 1 in $\mathbb{R}^2$ auffassen. Überträgt man von dem Intervall $[0, 2\pi)$ die Borel-σ-Algebra und die gleichförmige Verteilung (=Lebesgue-Maß/2π) auf D, so erhält man einen Wahrscheinlichkeitsraum $(D, \mathcal{B}_D, \lambda_D)$. Für $c \in D$ sei der Multiplikations- oder auch Drehoperator $T_c\colon D \to D$ definiert durch $T_c(\omega) := c\omega$. Ist $A \subset D$ ein Intervall, so gilt offensichtlich $\lambda_D T_c^{-1}(A) = \lambda_D(A)$. Aus Bemerkung 4.3 folgt, daß T_c maßerhaltend ist.

(c) Auf $([0,1), \mathcal{B}_{[0,1)}, \lambda)$ betrachten wir die Transformation $T\colon [0,1) \to [0,1)$, die gegeben ist durch

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

Für $0 \le a < b \le 1$ ist $T^{-1}([a,b)) = \left[\frac{a}{2}, \frac{b}{2}\right) \cup \left[\frac{a+1}{2}, \frac{b+1}{2}\right)$, also folgt $PT^{-1}([a,b)) = b - a = P([a,b))$. Die Mengen der Form $[a,b)$ mit $0 \le a < b \le 1$ bilden zusammen mit $\emptyset$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_{[0,1)}$. Aus Bemerkung 4.3 folgt $PT^{-1} = P$.

Definition 4.5.

(a) Für eine Abbildung $T\colon \Omega \to \Omega$ heißt eine Teilmenge A von Ω *T-invariant* (oder invariant unter T), wenn $T^{-1}(A) = A$ gilt.

(b) Eine maßerhaltende Abbildung T auf $(\Omega, \mathcal{F}, P)$ heißt *ergodisch,* wenn für jede T-invariante Menge $A \in \mathcal{F}$ gilt: $P(A) \in \{0, 1\}$.

Bemerkung 4.6. Die Familie der T-invarianten Mengen aus $\mathcal{F}$ ist offenbar eine Teil-σ-Algebra von $\mathcal{F}$. Diese wird meist mit $\mathcal{J}$ bezeichnet. Ist T ergodisch, so sagt man auch, daß $\mathcal{J}$ *P-trivial* ist. Nach Satz 3.17 ist in diesem Fall jede $\mathcal{J}$-meßbare Zufallsgröße fast sicher konstant.

Für das folgende Lemma erinnern wir an die Diskussion nach Lemma 4.2:

Lemma 4.7. *Es sei $\mathbb{X} = (X_n)_n$ eine Folge von unabhängigen identisch verteilten Zufallsgrößen auf $(\Omega, \mathcal{F}, P)$ mit Verteilung $P' := P\mathbb{X}^{-1}$. Dann ist die Verschiebungsabbildung $T\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, $(x_n)_n \mapsto (x_{n+1})_n$, ergodisch auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$.*

Beweis. Für eine T-invariante Menge $A \in \mathcal{B}^{\mathbb{N}}$ gilt für alle $n \in \mathbb{N}$ auch $A = T^{-n}(A) := (T^n)^{-1}(A) = \{\, (x_k)_k \in \mathbb{R}^{\mathbb{N}} \mid (x_{n+k})_k \in A \,\}$. Demzufolge liegt die Menge $\mathbb{X}^{-1}(A) =$

$\{\,\omega\colon (X_{n+1}(\omega),\, X_{n+2}(\omega), \dots) \in A\,\}$ in $\check{\mathcal{F}}_{n+1} := \bigvee_{k=n+1}^{\infty} X_k^{-1}(\mathcal{B})$. Da dies für alle $n \in \mathbb{N}$ gilt, ist $\mathbb{X}^{-1}(A) \in \mathcal{T}_\infty$ in der Notation von Satz 3.8, und aus diesem Satz folgt $P'(A) = P(\mathbb{X}^{-1}(A)) \in \{0, 1\}$. $\square$

Bemerkung 4.8. Der obige Beweisgang zeigt folgendes: Die σ-Algebra J der verschiebungsinvarianten meßbaren Mengen in $\mathbb{R}^{\mathbb{N}}$ ist eine Teil-σ-Algebra der terminalen σ-Algebra $\mathcal{T}_\infty = \bigcap_{n=1}^{\infty} \sigma(\pi_k\colon k \geq n)$. Es gilt jedoch keinesfalls $J = \mathcal{T}_\infty$, und ohne Beweis sei bemerkt, daß es ergodische Maße auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ gibt, bezüglich denen $\mathcal{T}_\infty$ nicht trivial ist.

Wir untersuchen nun den Multiplikations- beziehungsweise Drehoperator T_c aus Beispiel 4.4(b) auf Ergodizität:

Satz 4.9. *Es seien $D = \{\, z \in \mathbb{C}\colon |z| = 1\,\}$, $\mathcal{B}_D$ die Borel-σ-Algebra auf D und λ_D das normierte Lebesgue-Maß auf $(D, \mathcal{B}_D)$ sowie $T_c\colon D \to D$ definiert durch $T_c(\omega) := c\omega$ für $c, \omega \in D$. Dann ist T_c genau dann ergodisch auf $(D, \mathcal{B}_D, \lambda_D)$, wenn c keine Einheitswurzel ist.*

Beweis. Wir schreiben $T = T_c$. Ist c eine Einheitswurzel, das heißt existiert ein $n \in \mathbb{N}$ mit $c^n = 1$, so ist T^n die Identität. Somit ist für jedes $A \in \mathcal{F}$ die Menge $A \cup T^{-1}A \cup \dots \cup T^{-n+1}A$ invariant unter T. Wir wählen ein $A \in \mathcal{F}$ mit $0 < \lambda_D(A) < 1/n$ und erhalten

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

Somit ist T nicht ergodisch. Der Beweis der Umkehrung erfordert einige Vorbereitungen:

Lemma 4.10. *Sei $c \in D$ keine Einheitswurzel. Dann liegt $\{\, c^n : n \in \mathbb{N}_0\,\}$ dicht in D.*

Beweis. Die Folge $(c^n)_n$ besitzt mindestens einen Häufungspunkt $\omega_0 \in D$. Seien $\varepsilon > 0$ und $m > n$ zwei natürliche Zahlen mit $|c^n - \omega_0| < \varepsilon$ und $|c^m - \omega_0| < \varepsilon$. Dann gilt $|c^{m-n} - 1| \in (0, 2\varepsilon)$. Daraus folgt, daß für jedes $\omega \in D$ ein $k \in \mathbb{N}$ existiert mit $|\omega - c^{k(m-n)}| < 2\varepsilon$. Da $\varepsilon > 0$ beliebig ist, folgt die Behauptung. $\square$

Wir benötigen noch das folgende maßtheoretische Resultat (Beweis als Übungsaufgabe oder Satz 5.7 in dem Buch von H. Bauer):

Lemma 4.11. *Es sei $(\Omega, \mathcal{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $\mathcal{A}$ eine Algebra mit $\mathcal{F} = \sigma(\mathcal{A})$. Zu $\varepsilon > 0$ und $A \in \mathcal{F}$ existiert $B \in \mathcal{A}$ mit $P(A \triangle B) < \varepsilon$.*

Fortsetzung des Beweises von 4.9. Sei c keine Einheitswurzel. Sei $\mathcal{A}$ die Familie der endlichen Vereinigungen paarweise disjunkter Intervalle in D, dann ist $\mathcal{A}$ eine Algebra. Ist A invariant unter T_c mit $\lambda_D(A) > 0$ und $\varepsilon \in \left(0, \frac{1}{2}\right)$, so existiert nach

Lemma 4.11 eine endliche Vereinigung $B = \bigcup_{i=1}^n I_i$ disjunkter Intervalle $I_1, \dots, I_n \subset D$ mit $\lambda_D(A \triangle B) \le \varepsilon \lambda_D(A)$. Wir können annehmen, daß $\lambda_D(I_i) \le \varepsilon$ ist für $i = 1, \dots, n$.

Wir haben

$$\lambda_D(A \triangle B) \le \varepsilon\lambda_D(A) \le 2\varepsilon(1-\varepsilon)\lambda_D(A) \le 2\varepsilon(\lambda_D(A) - \lambda_D(A \triangle B)) \le 2\varepsilon\lambda_D(B),$$

und daraus folgt

$$\sum_{i=1}^n \lambda_D(A\cap I_i) = \lambda_D(A\cap B) \ge \lambda_D(B)-\lambda_D(A\triangle B) \ge (1-2\varepsilon)\lambda_D(B) = (1-2\varepsilon)\sum_{i=1}^n \lambda_D(I_i).$$

Mindestens eines der I_i — kurz mit I bezeichnet — erfüllt also die Ungleichung $\lambda_D(A\cap I) \ge (1-2\varepsilon)\lambda_D(I)$. Nach Lemma 4.10 existieren ein $k \in \mathbb{N}$ und $n_1, \dots, n_k \in \mathbb{N}$ derart, daß die Intervalle $T^{-n_1}I$, $T^{-n_2}I, \dots, T^{-n_k}I$ paarweise disjunkt sind und D bis auf eine Menge von kleinerem Maß als 2ε ausfüllen. Wegen der T-Invarianz von A und λ_D gilt für $j = 1, \dots, k$:

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

und dies führt zu

$$\lambda_D(A) \ge \sum_{j=1}^k \lambda_D(A\cap T^{-n_j}I) \ge (1-2\varepsilon)\sum_{j=1}^k \lambda_D(T^{-n_j}I) \ge (1-2\varepsilon)^2.$$

Daraus folgt $\lambda_D(A) = 1$. □

Die Transformation aus Beispiel 4.4(c) ist ebenfalls ergodisch. Um das einzusehen, führen wir den folgenden Begriff ein:

Definition 4.12. Eine maßerhaltende Transformation T auf $(\Omega, \mathcal{F}, P)$ heißt *mischend,* wenn für alle $A, B \in \mathcal{F}$ gilt:

$$\lim_{n\to\infty} P(A\cap T^{-n}B) = P(A)P(B). \tag{4.13}$$

Lemma 4.14. *Jede mischende Transformation ist ergodisch.*

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

$$P(A) = P(A\cap A) = P(A\cap T^{-n}A) \to P(A)P(A) \quad \text{für } n\to\infty.$$

Daraus folgt sofort $P(A) = P(A)^2$, das heißt $P(A) \in \{0,1\}$. □

Lemma 4.15. *Es sei $\mathcal{A}$ eine Algebra, die $\mathcal{F}$ erzeugt. Falls die Gleichung (4.13) für alle $A, B \in \mathcal{A}$ gilt, so ist T mischend.*

Beweis. Seien $A, B \in \mathcal{F}$ und $\varepsilon > 0$. Nach Lemma 4.11 existieren $A_0, B_0 \in \mathcal{A}$ mit $P(A \triangle A_0) < \varepsilon$ und $P(B \triangle B_0) < \varepsilon$. Daraus ergibt sich

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

Die Folge $(P(A_0 \cap T^{-n}B_0))_n$ konvergiert gegen $P(A_0)P(B_0)$, und es gilt

$$|P(A_0) - P(A)| \leq P(A \triangle A_0) < \varepsilon \quad \text{sowie} \quad |P(B_0) - P(B)| < \varepsilon.$$

Da $\varepsilon > 0$ beliebig ist, folgt $\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B)$. □

Nicht jede ergodische Transformation ist mischend. Man sieht leicht ein, daß der Drehoperator T_c auf D (siehe Beispiel 4.4(b)) für kein c mischend ist.

Satz 4.16. *Die in Beispiel 4.4 (c) definierte Transformation T auf $([0,1), \mathcal{B}_{[0,1)}, \lambda)$ ist mischend.*

Beweis. Für jede Menge $A \subset [0,1)$ sind $T^{-1}A \cap \left[0, \frac{1}{2}\right)$ und $T^{-1}A \cap \left[\frac{1}{2}, 1\right)$ nur um 1/2 gegeneinander verschobene Mengen. Daraus folgt:

$$P\big(T^{-1}A \cap \left[\tfrac{1}{2}, 1\right)\big) = P\big(T^{-1}A \cap \left[0, \tfrac{1}{2}\right)\big) = P(T^{-1}A)/2 = P(A)P\big(\left[0, \tfrac{1}{2}\right)\big).$$

Eine analoge Überlegung führt dazu, daß für $n \in \mathbb{N}$ und $k = 0, 1, \dots, 2^n - 1$ gilt:

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

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

$$\lim_{m\to\infty} P(T^{-m}A \cap I) = P(A)P(I) \quad \text{für alle } I \in \mathcal{A}_0.$$

Diese Gleichung bleibt auch richtig, wenn I eine endliche Vereinigung paarweise disjunkter Intervalle dieser Form ist. Diese Intervallfiguren bilden jedoch eine Algebra, die die Borel-σ-Algebra auf $[0,1)$ erzeugt. Satz 4.16 folgt somit aus Lemma 4.15. □

Wir wollen nun nachweisen, daß eine irreduzible, positiv rekurrente und aperiodische Markoffkette mischend ist. Sei I eine abzählbare Menge und $\mathbb{P} = (p_{ij})_{i,j\in I}$ eine irreduzible, positiv rekurrente und aperiodische stochastische Matrix. Wie wir aus der Vorlesung „Einführung in die Stochastik" wissen, existiert eine eindeutige stationäre Wahrscheinlichkeitsverteilung auf I, das heißt ein Wahrscheinlichkeitsmaß π mit $\sum_{i\in I} \pi(i)p_{ij} = \pi(j)$ für alle j. Das Maß π und die Matrix $\mathbb{P}$ definieren zusammen nach Beispiel 2.29(c) ein Wahrscheinlichkeitsmaß P auf $(I^{\mathbb{N}}, \mathcal{P}(I)^{\mathbb{N}})$, das invariant unter dem Verschiebungsoperator $T\colon I^{\mathbb{N}} \to I^{\mathbb{N}}$ ist.

Satz 4.17. *Es sei I eine abzählbare Menge und P die Verteilung einer stationären irreduziblen, positiv rekurrenten und aperiodischen Markoffkette auf $(I^{\mathbb{N}}, \mathcal{P}(I)^{\mathbb{N}})$. Dann ist $T\colon I^{\mathbb{N}} \to I^{\mathbb{N}}$, definiert durch $T((x_n)_n) := (x_{n+1})_n$, mischend.*

Beweis. Für $k \in \mathbb{N}$ sei $X_k\colon I^{\mathbb{N}} \to I$ die k-te Projektion, definiert durch $X_k((i_n)_n) := i_k$. Sei $\mathcal{A}$ die Algebra $\bigcup_{m=1}^{\infty} \sigma(X_1, \dots, X_m)$. Beachte, daß gilt: $\sigma(X_1, \dots, X_m) = X^{(m)-1}(\mathcal{P}(I^m)) = \{\, X^{(m)-1}(C)\colon C \subset I^m \,\}$, wobei $X^{(m)} = (X_1, \dots, X_m)\colon I^{\mathbb{N}} \to I^m$ die Projektion auf die ersten m Faktoren bezeichnet. Für jedes $A \in \mathcal{A}$ existiert also ein $m \in \mathbb{N}$ und ein $C \subset I^m$ mit $A = X^{(m)-1}(C) = \{\, i \in I^{\mathbb{N}}\colon X^{(m)}(i) \in C \,\}$.

Da $\sigma(\mathcal{A}) = \mathcal{P}(I)^{\mathbb{N}}$ ist, genügt es nach Lemma 4.15 nachzuweisen, daß für $A, B \in \mathcal{A}$ gilt:

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

Dies ist äquivalent zu

$$\lim_{n\to\infty} P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) = P(X^{(m)} \in C)P(X^{(k)} \in D) \tag{4.18}$$

für $k, m \in \mathbb{N}$ und $C \subset I^m$ sowie $D \subset I^k$. Wir überlegen uns zunächst, daß es genügt, die Gleichung (4.18) für endliche Mengen C und D nachzuweisen. Ist nämlich D eine beliebige Teilmenge von I^k, so existiert zu $\varepsilon > 0$ eine endliche Teilmenge D' von D mit

$$|P(X^{(k)} \in D) - P(X^{(k)} \in D')| \leq \varepsilon.$$

Wegen der T-Invarianz von P folgt

$$\begin{aligned}
&|P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) - P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D')| \\
&\quad \leq |P(T^n(X^{(k)}) \in D) - P(T^n(X^{(k)}) \in D')| \\
&\quad \leq |P(X^{(k)} \in D) - P(X^{(k)} \in D')| \leq \varepsilon.
\end{aligned}$$

Da $\varepsilon > 0$ beliebig ist, genügt es also, die Gleichung (4.18) für endliches D zu zeigen. Für C ist das Argument dasselbe. Seien also $k, m \in \mathbb{N}$ und $C \subset I^m$, $D \subset I^k$ endliche Mengen. Für $n > m$ gilt:

$$\begin{aligned}
&P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) \\
&\quad = P((X_1, \dots, X_m) \in C,\ (X_{n+1}, \dots, X_{n+k}) \in D) \\
&\quad = \sum_{(i_1,\dots,i_m)\in C} \sum_{(j_1,\dots,j_k)\in D} P(X_1 = i_1, \dots, X_m = i_m, X_{n+1} = j_1, \dots,\ X_{n+k} = j_k) \\
&\quad = \sum_{(i_1,\dots,i_m)\in C} \sum_{(j_1,\dots,j_k)\in D} \pi(i_1) p_{i_1 i_2} \cdots p_{i_{m-1} i_m} p_{i_m j_1}^{(n-m)} p_{j_1 j_2} \cdots p_{j_{k-1} j_k}.
\end{aligned}$$

Aus den Sätzen (9.26) und (9.29) aus dem Skript „Einführung in die Stochastik“ folgt $\lim_{n\to\infty} p_{i_m j_1}^{(n-m)} = \pi(j_1)$. Da die obigen Summen endlich sind, können wir natürlich die

Summation mit dem Limes $n \to \infty$ vertauschen und erhalten

$$\begin{aligned}
&\lim_{n\to\infty} P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) \\
&= \sum_{(i_1,\dots,i_m)\in C} \pi(i_1) p_{i_1 i_2} \cdots p_{i_{m-1} i_m} \sum_{(j_1,\dots,j_k)\in D} \pi(j_1) p_{j_1 j_2} \cdots p_{j_{k-1} j_k} \\
&= P(X^{(m)} \in C) P(X^{(k)} \in D).
\end{aligned}$$

□

Das wichtigste Ergebnis des 4. Kapitels ist der folgende Satz. Es sei daran erinnert, daß $\mathcal{J}$ die σ-Algebra der T-invarianten meßbaren Mengen bezeichnet (siehe Bemerkung 4.6).

Satz 4.19 (ERGODENSATZ). *Sei T eine maßerhaltende Transformation auf $(\Omega, \mathcal{F}, P)$, und es sei $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$. Dann konvergiert $\frac{1}{n}\sum_{j=0}^{n-1} X \circ T^j$ fast sicher für $n \to \infty$ gegen eine $\mathcal{J}$-meßbare Zufallsgröße Y, für die $\int X\,dP = \int Y\,dP$ gilt.*

Bemerkung 4.20.

(a) Ist T ergodisch, so ist nach Satz 3.17 jede $\mathcal{J}$-meßbare Zufallsgröße fast sicher konstant. Wegen $\int X\,dP = \int Y\,dP$ muß $Y = \int X\,dP = EX$ gelten. Insbesondere folgt aus dem Ergodensatz zusammen mit Lemma 4.7 das Gesetz der großen Zahlen.

(b) Es gilt auch Konvergenz im 1. Mittel. Dies soll hier nicht bewiesen werden. Man beachte, daß die $\mathcal{L}^1$-Konvergenz nicht aus der fast sicheren Konvergenz und Satz 2.55 folgt, da im allgemeinen keine integrierbare Dominante für die Folge $\left(\frac{1}{n}\sum_{j=0}^{n-1} X \circ T^j\right)_n$ existiert. Ein einfacher funktionalanalytischer Beweis für die Konvergenz im 1. Mittel findet sich etwa in dem Buch 'Ergodic Theory and Information' von P. Billingsley.

Beweis von Satz 4.19 *(Katznelson & Weiss, 1981)*. Wegen der Zerlegung $X = X^+ - X^-$ reicht es aus, den Satz für nichtnegative Zufallsgrößen zu beweisen. Sei also $X \geq 0$, und sei $S_n(\omega) = \frac{1}{n}\sum_{j=0}^{n-1} X(T^j\omega)$ (per Konvention: $T^0\omega = \omega$).

Man definiert $\overline{X}(\omega) := \limsup_{n\to\infty} S_n(\omega)$ sowie $\underline{X}(\omega) := \liminf_{n\to\infty} S_n(\omega)$. Offenbar gelten $\underline{X} \circ T = \underline{X}$ und $\overline{X} \circ T = \overline{X}$. Daraus folgt sofort, daß $\underline{X}$ und $\overline{X}$ $\mathcal{J}$-meßbar sind.

Man zeigt nun, daß

$$\int \overline{X}\,dP \leq \int X\,dP \leq \int \underline{X}\,dP \tag{4.21}$$

gilt. Wegen $\underline{X} \leq \overline{X}$ folgt daraus $\underline{X} = \overline{X}$ fast sicher, das heißt, $(S_n)_n$ konvergiert fast sicher gegen eine $\mathcal{J}$-meßbare Zufallsgröße mit den gewünschten Eigenschaften.

Wir beweisen nun die 1. Ungleichung in (4.21). Die Grundidee des Beweises ist sehr einfach: Man schaut auf die „Zeitpunkte“ $0 = n_0 < n_1 < n_2 < n_3 < \dots$, für die der durchschnittliche Zuwachs $\frac{1}{n_{j+1}-n_j}[X(T^{n_j}\omega) + \cdots + X(T^{n_{j+1}-1}\omega)]$ dem lim sup

„nahe“ kommt. Es ergeben sich zwei technische Probleme: Erstens kann man nicht von vorneherein ausschließen, daß $\overline{X} = \infty$ ist. Um diese Schwierigkeit zu umgehen, schneidet man $\overline{X}$ ab. Für $M \in (0, \infty)$ sei $\overline{X}_M = \min(\overline{X}, M)$. Sei $\varepsilon > 0$ beliebig. Für $\omega \in \Omega$ sei $n(\omega) \in \mathbb{N}$ die kleinste Zahl $k \in \mathbb{N}$ mit $S_k(\omega) \geq \overline{X}_M(\omega) - \varepsilon$. Die Abbildung $n \colon \Omega \to \mathbb{N}$ ist $\mathcal{F}$-$\mathcal{P}(\mathbb{N})$-meßbar, denn für $k \in \mathbb{N}$ gilt

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

Die zweite technische Schwierigkeit besteht darin, daß n als Funktion von ω nicht beschränkt zu sein braucht. Allerdings gilt natürlich $\lim_{N\to\infty} P(\{\, \omega \colon n(\omega) > N \,\}) = 0$. Es existiert also ein $N_\varepsilon \in \mathbb{N}$ mit

$$P(\{\, \omega \colon n(\omega) > N_\varepsilon \,\}) \leq \varepsilon. \tag{4.22}$$

Wir stutzen nun auch $n(\omega)$ und X zurecht: Mit $A := \{\, \omega \colon n(\omega) \leq N_\varepsilon \,\}$ seien

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

Gilt $X(\omega) > M$, so ist $n(\omega) = 1$ (wegen $S_1(\omega) = X(\omega)$), das heißt, es gilt $\omega \in A$. Es folgt also

$$X(\omega) \leq \widetilde{X}(\omega) \quad \text{für alle } \omega. \tag{4.23}$$

Für $\omega \notin A$ gilt wegen $\widetilde{n}(\omega) = 1$ und $\widetilde{X}(\omega) = M$ die Abschätzung

$$\frac{1}{\widetilde{n}(\omega)} \sum_{j=0}^{\widetilde{n}(\omega)-1} \widetilde{X}(T^j\omega) \geq \overline{X}_M(\omega) - \varepsilon. \tag{4.24}$$

Diese Ungleichung gilt auch für $\omega \in A$, denn dann ist $\widetilde{n}(\omega) = n(\omega)$, und wegen (4.23) ist die linke Seite von (4.24) nicht kleiner als $S_{n(\omega)}(\omega)$, also auch nicht kleiner als $\overline{X}_M(\omega) - \varepsilon$ nach der Definition von $n(\omega)$.

Wegen (4.22) folgt

$$\int \widetilde{X}\, dP = \int_A \widetilde{X}\, dP + \int_{A^c} \widetilde{X}\, dP \leq \int X\, dP + M\varepsilon. \tag{4.25}$$

Man definiert nun rekursiv:

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

Für $m \in \mathbb{N}$ sei $K(\omega)$ die größte Zahl $k \in \mathbb{N}$ mit $n_k(\omega) \leq m$. (Der Index $K(\omega)$ hängt natürlich von m ab.) Wegen $\widetilde{n}(\omega) \leq N_\varepsilon$ folgt $m - n_{K(\omega)}(\omega) \leq N_\varepsilon$.

Es folgt

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

Die Ungleichung (4.24), angewandt auf ω, $T^{n_1(\omega)}(\omega)$, $T^{n_2(\omega)}(\omega), \ldots,$ $T^{n_{K(\omega)-1}(\omega)}(\omega)$, ergibt für die rechte Seite die Abschätzung

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

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

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

Dividiert man durch m und beachtet $\int \widetilde{X}(T^j\omega)P(d\omega) = \int \widetilde{X}\, dP$ (denn T ist maßerhaltend), so folgt:

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

Somit gilt wegen (4.25):

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

Da dies für alle $m \in \mathbb{N}$ und $\varepsilon > 0$ gilt, folgt $\int X dP \geq \int \overline{X}_M dP$ für alle $M > 0$.

Nach dem monotonen Konvergenzsatz (Satz 1.37) folgt $\int X dP \geq \int \overline{X} dP$. Damit ist die 1. Ungleichung von (4.21) bewiesen; die zweite folgt völlig analog und ist sogar noch etwas einfacher, da $\underline{X}$ nicht nach unten abgeschnitten werden muß. □

Bemerkung 4.26. Sei $\mathbb{X} = (X_n)_n$ ein stationärer Prozeß. Die Verschiebungstransformation $T\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ ist maßerhaltend für $P' := P\mathbb{X}^{-1}$. Sei $\pi_1 : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ die Projektion auf den 1. Faktor. Die Zufallsgröße X_1 liegt genau dann in $\mathcal{L}^1(P)$, wenn π_1 in $\mathcal{L}^1(P')$ liegt. In diesem Fall konvergiert nach dem Ergodensatz die Folge $\left(\frac{1}{n}\sum_{j=0}^{n-1} \pi_1 \circ T^j\right)_n$ fast sicher gegen eine auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$ definierte T-invariante Zufallsgröße ψ mit $\int \psi\, dP' = \int \pi_1\, dP'$. Das bedeutet aber nichts anderes als die P-fast sichere Konvergenz von $\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)_n$ gegen $\psi \circ \mathbb{X}$. Ist T ergodisch, so konvergiert $\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)_n$ P-fast sicher gegen EX_1.

Wir können das auch auf irreduzible, aperiodische, positiv rekurrente Markoffketten anwenden. Ist $(X_n)_n$ eine derartige stationäre Kette auf einer abzählbaren Menge I

mit stationärer Startverteilung π und ist $f : I \to \mathbb{R}$ eine π-integrierbare Abbildung, so folgt aus dem Ergodensatz und Satz 4.17 sofort

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(X_j) = \int f d\pi = \sum_{j\in I} f(j)\pi(j)$$

P-fast sicher.

Kapitel 5

Bedingte Wahrscheinlichkeiten und Erwartungswerte

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

Definition 5.1. Sind $A, B \in \mathcal{F}$ mit $P(B) > 0$, so ist die *bedingte Wahrscheinlichkeit von A gegeben B* durch $P(A \,|\, B) := P(A \cap B)/P(B)$ definiert.

Sei X eine Zufallsgröße mit Werten in einer höchstens abzählbaren Menge E (versehen mit der Potenzmenge als σ-Algebra), wobei wir annehmen, daß $P(X = x) > 0$ für alle $x \in E$ gilt. Für Ereignisse $A \in \mathcal{F}$ ist $P(A \,|\, X = x)$ in diesem Fall durch $P(A \cap \{X = x\})/P(X = x)$ definiert. Diese Definition ist so nicht mehr möglich, wenn $P(X = x) = 0$ für eine Menge von $x \in E$ mit positiver PX^{-1}-Wahrscheinlichkeit gilt. Ist X zum Beispiel eine reelle Zufallsgröße, deren Verteilung eine Dichte bezüglich des Lebesgue-Maßes hat, so gilt $P(X = x) = 0$ für alle $x \in \mathbb{R}$. Man ist natürlich versucht, $P(A \,|\, X = x)$ als Limes von $P(A \,|\, X \in U_n)$ zu definieren, wobei $(U_n)_{n \in \mathbb{N}}$ eine Folge von „Umgebungen“ von x ist, für die $P(X \in U_n) > 0$ und $U_n \downarrow \{x\}$ gilt. Ein solcher Ansatz führt jedoch zu endlosen Schwierigkeiten. Zum Glück gibt es den Satz von Radon-Nikodym. Eine Definition von $P(A \,|\, X = x)$ ist damit nämlich stets möglich, jedoch nicht isoliert für einzelne x, sondern als Funktion von x.

Wir untersuchen zunächst Eigenschaften der Funktion $x \mapsto P(A \,|\, X = x)$ im Spezialfall $P(X = x) > 0$ für alle $x \in E$. Für jede Teilmenge $C \subset E$ gilt offenbar:

$$(5.2) \quad P(A \cap \{X \in C\}) = \sum_{x \in C} P(A \,|\, X = x) P(X = x) = \int_C P(A \,|\, X = x) PX^{-1}(dx).$$

Notation: Für eine allgemeine $(E, \mathcal{E})$-wertige Zufallsgröße X sei

$$\mathbb{P}_{A \,|\, X = .} := \Big\{ \psi \colon E \to [0, \infty) \colon \psi \text{ ist } \mathcal{E}\text{-}\mathcal{B}\text{-meßbar, und es gilt } \int_C \psi \, dPX^{-1} = P(A \cap \{X \in C\}) \text{ für alle } C \in \mathcal{E} \Big\}.$$

Definition 5.3. Ein Element $\psi \in \mathbb{P}_{A \,|\, X = .}$ heißt *Version der bedingten Wahrscheinlichkeit* von A gegeben X.

Das Maß ν auf $\mathcal{E}$ sei durch $\nu(C) = P(A \cap \{X \in C\})$ für alle $C \in \mathcal{E}$ definiert. Offenbar gilt $\nu(C) \leq PX^{-1}(C)$; insbesondere ist ν also absolutstetig bezüglich PX^{-1}, und die Bedingung $\nu(C) = \int_C \psi \, dPX^{-1}$ für alle $C \in \mathcal{E}$ besagt nichts anderes, als daß ψ eine Radon-Nikodym-Ableitung von ν bezüglich PX^{-1} ist. Aus dem Satz von Radon-Nikodym ergibt sich somit:

Satz 5.4. *Zu jedem $A \in \mathcal{F}$ und jeder $(E, \mathcal{E})$-wertigen Zufallsgröße X existiert eine Version der bedingten Wahrscheinlichkeit von A gegeben X und ist bis auf PX^{-1}-fast sichere Gleichheit eindeutig.*

Bemerkung 5.5. Es existiert eine Version $\psi \in \mathbb{P}_{A \mid X=.}$ mit $\psi(x) \leq 1$ für alle $x \in E$.

Beweis. Seien $\psi \in \mathbb{P}_{A \mid X=.}$ und $C := \{x \in E \colon \psi(x) > 1\} \in \mathcal{E}$. Die Annahme $PX^{-1}(C) > 0$ widerspricht $\int_C \psi(x) PX^{-1}(dx) \leq PX^{-1}(C)$. Somit gilt $PX^{-1}(C) = 0$. Dann ist $\psi' = \min(\psi, 1)$ offenbar auch in $\mathbb{P}_{A \mid X=.}$. □

Wir unterscheiden in Zukunft oft nicht zwischen $\mathbb{P}_{A \mid X=.}$ und den Elementen darin. Für $\psi \in \mathbb{P}_{A \mid X=.}$ schreibt man einfach $\psi = P(A \mid X = \,.\,)$ und $\psi(x) = P(A \mid X = x)$. Man muß sich jedoch darüber klar sein, daß $P(A \mid X = x)$ nur als Funktion von x definiert ist und nur bis auf PX^{-1}-fast sichere Gleichheit eindeutig ist. Dies scheint von marginaler Bedeutung zu sein, führt jedoch auf die folgende Schwierigkeit: Für jedes $A \in \mathcal{F}$ sei $P(A \mid X = \,.\,) \in \mathbb{P}_{A \mid X=.}$ gewählt. Ist dann für jedes $x \in E$ die Funktion $\mathcal{F} \ni A \mapsto P(A \mid X = x)$ ein Wahrscheinlichkeitsmaß? Im allgemeinen natürlich nicht, denn man kann ja $P(A \mid X = \,.\,)$ auf einer PX^{-1}-Nullmenge ganz nach Belieben festlegen. Man mag jedoch den Eindruck haben, daß dieses Nullmengenproblem mit einigen kosmetischen Operationen analog zu Bemerkung 5.5 behoben werden kann: Sind $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, paarweise disjunkte Mengen und $A = \bigcup_{n \in \mathbb{N}} A_n$, so sieht man leicht, daß $\sum_{n=1}^{\infty} P(A_n \mid X = \,.\,) \in \mathbb{P}_{A \mid X=.}$ gilt, das heißt $\sum_{n=1}^{\infty} P(A_n \mid X = \,.\,) = P(A \mid X = \,.\,)$ PX^{-1}-fast sicher. Durch eine Korrektur von $P(A \mid X = x)$ auf einer Nullmenge kann man daher erreichen, daß $P(A \mid X = x) = \sum_{n=1}^{\infty} P(A_n \mid X = x)$ gilt. Eine derartige Korrektur müßte jedoch für *jede* Zerlegung von A erfolgen, und da eine Menge A in der Regel überabzählbar viele Zerlegungen in abzählbar viele, paarweise disjunkte, meßbare Mengen besitzt, stößt dieses naive Vorgehen offenbar auf Schwierigkeiten. Zunächst eine Definition:

Definition 5.6. Es sei X eine $(E, \mathcal{E})$-wertige Zufallsgröße. Eine *reguläre bedingte Wahrscheinlichkeit gegeben X* ist ein Markoffkern Q von $(E, \mathcal{E})$ nach $(\Omega, \mathcal{F})$ mit der Eigenschaft $Q(\,.\,, A) \in \mathbb{P}_{A \mid X=.}$ für alle $A \in \mathcal{F}$.

Reguläre bedingte Wahrscheinlichkeiten existieren nicht immer. Es gilt jedoch:

Satz 5.7. *Ist Ω ein vollständiger, separabler metrischer Raum und $\mathcal{B}_\Omega$ die Borel-σ-Algebra, das heißt die von den offenen Mengen in Ω erzeugte σ-Algebra, so existiert für jede $(E, \mathcal{E})$-wertige auf $(\Omega, \mathcal{B}_\Omega, P)$ definierte Zufallsgröße eine reguläre bedingte Wahrscheinlichkeit.*

Beweis. Ein Beweis findet sich in Anhang A. 2.

Als nächstes sollen bedingte Erwartungswerte eingeführt werden. Sei Y eine zweite, $(\mathbb{R}, \mathcal{B})$-wertige Zufallsgröße. Existiert eine reguläre bedingte Wahrscheinlichkeit Q für P gegeben X, so wird man $E(Y \mid X = x)$ einfach durch $\int Y(\omega) Q(x, d\omega)$ definieren (falls $Y \in \mathcal{L}^1(Q(x, \,.\,)$ für PX^{-1}-fast alle $x \in E$). Reguläre bedingte Wahrscheinlichkeiten sind jedoch für die Definition bedingter Erwartungswerte nicht erforderlich.

Definition 5.8. Sei $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Ein *bedingter Erwartungswert* $E(Y \mid X = x)$ ist eine $\mathcal{E}$-meßbare, PX^{-1}-integrierbare Funktion in x mit

$$\int_C E(Y \mid X = x)\, PX^{-1}(dx) = E(Y 1_{X^{-1}(C)}) \tag{5.9}$$

für alle $C \in \mathcal{E}$.

Ist $Y \geq 0$, so besagt die Bedingung $\int_C E(Y \mid X = x)\, PX^{-1}(dx) = E(Y 1_{X^{-1}(C)})$ nichts anderes, als daß $E(Y \mid X = \,.\,)$ eine Radon-Nikodym-Ableitung des Maßes $\mathcal{E} \ni C \mapsto E(Y 1_{X^{-1}(C)})$ bezüglich PX^{-1} ist. Existenz und Eindeutigkeit folgen in diesem Fall aus dem Satz von Radon-Nikodym. Ist $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, so setzt man $E(Y \mid X = x) = E(Y^+ \mid X = x) - E(Y^- \mid X = x)$. Es folgt also:

Satz 5.10. *Die $\mathcal{E}$-meßbare Funktion $E \ni x \mapsto E(Y \mid X = x)$ existiert und ist bis auf PX^{-1}-fast sichere Gleichheit eindeutig bestimmt.*

Man beachte, daß bedingte Wahrscheinlichkeiten Spezialfälle von bedingten Erwartungswerten mit $Y = 1_A$ sind. Bedingte Wahrscheinlichkeiten und Erwartungswerte (bei bedingender Zufallsgröße $X : \Omega \to E$) sind auf E definierte, $\mathcal{E}$-meßbare, reellwertige Abbildungen. Man kann sie daher mit $X : \Omega \to E$ zusammensetzen und erhält so auf Ω definierte Abbildungen. Man schreibt dafür meist

$$P(A \mid X) := P(A \mid X = \,.\,) \circ X \quad \text{beziehungsweise} \quad E(Y \mid X) := E(Y \mid X = \,.\,) \circ X.$$

Also sind $P(A \mid X)$ und $E(Y \mid X)$ auf $(\Omega, \mathcal{F}, P)$ definierte reelle Zufallsgrößen. Welche Eigenschaften haben sie?

Satz 5.11. *Seien $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ und $A \in \mathcal{F}$.*

(a) *Die Zufallsgrößen $E(Y \mid X)$ und $P(A \mid X)$ sind $X^{-1}(\mathcal{E})$-meßbar.*
(b) *Für alle $C \in X^{-1}(\mathcal{E})$ gelten*

$$\int_C E(Y \mid X)\, dP = E(1_C Y) \quad \textit{und} \quad \int_C P(A \mid X)\, dP = P(A \cap C).$$

Beweis. (a) ist klar nach Definition, und (b) folgt sofort aus dem Transformationssatz 1.49. □

Es folgt genauso wie oben, daß die Eigenschaften (a) und (b) $E(Y \mid X)$ beziehungsweise $P(A \mid X)$ bis auf P-fast sichere Gleichheit eindeutig festlegen. Die Eigenschaften (a) und (b) legen eine allgemeinere Definition von bedingten Erwartungswerten nahe:

Definition 5.12. Es seien $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ und $\mathcal{D}$ eine Teil-σ-Algebra von $\mathcal{F}$. Ein *bedingter Erwartungswert von Y gegeben $\mathcal{D}$* ist eine auf Ω definierte Zufallsgröße $E(Y \,|\, \mathcal{D})$ aus $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ mit

(a) $E(Y \,|\, \mathcal{D})$ ist $\mathcal{D}$-meßbar.
(b) Für alle $C \in \mathcal{D}$ gilt $\int_C E(Y \,|\, \mathcal{D})\, dP = \int_C Y\, dP$.

Existenz und Eindeutigkeit folgen wieder wie oben: Ist $Y \geq 0$, so ist $\mathcal{D} \ni C \mapsto \int_C Y\, dP$ ein endliches Maß auf $\mathcal{D}$, das absolutstetig bezüglich P ist. Somit existiert die Radon-Nikodym-Ableitung auf $(\Omega, \mathcal{D})$, die der gewünschte bedingte Erwartungswert ist. Für $Y \in \mathcal{L}^1$ setzt man $E(Y \,|\, \mathcal{D}) = E(Y^+ \,|\, \mathcal{D}) - E(Y^- \,|\, \mathcal{D})$. Die Eindeutigkeit folgt ebenso einfach: Sind $\varphi_1, \varphi_2 \in \mathcal{L}_1(\Omega, \mathcal{D}, P)$ mit $\int_C \varphi_1\, dP = \int_C \varphi_2\, dP$ für alle $C \in \mathcal{D}$, so folgt sofort $\varphi_1 = \varphi_2$ P-fast sicher. Eindeutigkeit hat man also wieder nur bis auf P-fast sichere Gleichheit.

Beispiel 5.13. Seien $A_1, \ldots, A_n \in \mathcal{F}$ paarweise disjunkt mit $\Omega = A_1 \cup \cdots \cup A_n$. Sei $\mathcal{D} := \sigma(\{A_1, \ldots, A_n\})$. Dann besteht $\mathcal{D}$ aus $\emptyset$ und den Vereinigungen von Elementen aus $\{A_1, \ldots, A_n\}$. Sei $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Für $\omega' \in A_i$ mit $P(A_i) > 0$ setzen wir $\overline{X}(\omega') = \int_{A_i} X(\omega)\, P(d\omega)/P(A_i)$, und für $\omega' \in A_i$ mit $P(A_i) = 0$ setzen wir $\overline{X}(\omega') = 0$. Die Abbildung $\overline{X}$ ist offenbar $\mathcal{D}$-meßbar, da sie auf jeder der Mengen A_i konstant ist. Für $D \in \mathcal{D}$ mit Darstellung $D = A_{i_1} \cup \cdots \cup A_{i_m}$ gilt

$$\int_D \overline{X}(\omega) P(d\omega) = \sum_{j=1}^m \int_{A_{i_j}} \overline{X}(\omega) P(d\omega) = \sum_{j=1}^m \int_{A_{i_j}} X(\omega) P(d\omega) = \int_D X(\omega) P(d\omega).$$

Also erfüllt $\overline{X}$ die Bedingungen (a) und (b) in Definition 5.12. Somit gilt $\overline{X} = E(X \,|\, \mathcal{D})$ fast sicher.

Der folgende Satz ist eine Auflistung der wichtigsten Eigenschaften bedingter Erwartungswerte.

Satz 5.14. *Es seien $X, X_1, X_2 \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, und $\mathcal{D}$ sowie $\mathcal{D}'$ seien Teil-σ-Algebren von $\mathcal{F}$. Dann gelten:*

(a) *Ist $\mathcal{D}$ die triviale σ-Algebra, das heißt $\mathcal{D} = \{\emptyset, \Omega\}$, so gilt $E(X \,|\, \mathcal{D}) = E(X)$ fast sicher.*
(b) *Ist X $\mathcal{D}$-meßbar, so gilt $E(X \,|\, \mathcal{D}) = X$ fast sicher.*
(c) *Sind $a_1, a_2 \in \mathbb{R}$, so gilt $E(a_1 X_1 + a_2 X_2 \,|\, \mathcal{D}) = a_1 E(X_1 \,|\, \mathcal{D}) + a_2 E(X_2 \,|\, \mathcal{D})$ fast sicher.*
(d) *Ist $X \geq 0$, so ist $E(X \,|\, \mathcal{D}) \geq 0$ fast sicher.*
(e) *Es gilt $|E(X \,|\, \mathcal{D})| \leq E(|X| \,|\, \mathcal{D})$ fast sicher.*
(f) *Ist $\mathcal{D}' \subset \mathcal{D}$, so gilt $E(X \,|\, \mathcal{D}') = E(E(X \,|\, \mathcal{D}) \,|\, \mathcal{D}')$ fast sicher.*
(g) *Sind X und $\mathcal{D}$ unabhängig, so gilt $E(X \,|\, \mathcal{D}) = E(X)$ fast sicher.*

Beweis.

(a) Dies ist ein Spezialfall von Beispiel 5.13.
(b) Folgt unmittelbar aus der Definition des bedingten Erwartungswertes.
(c) Sei Y die rechte Seite der behaupteten Gleichung. Dann gilt für alle $D \in \mathcal{D}$:

$$\begin{aligned}\int_D Y\,dP &= a_1 \int_D E(X_1 \,|\, \mathcal{D})\,dP + a_2 \int_D E(X_2 \,|\, \mathcal{D})\,dP \\ &= a_1 \int_D X_1\,dP + a_2 \int_D X_2\,dP \quad \text{(nach Definition 5.12(b))} \\ &= \int_D (a_1X_1 + a_2X_2)\,dP.\end{aligned}$$

Außerdem ist Y natürlich $\mathcal{D}$-meßbar, erfüllt also die definierenden Eigenschaften für einen bedingten Erwartungswert von $a_1X_1 + a_2X_2$ gegeben $\mathcal{D}$.
(d) Für alle $D \in \mathcal{D}$ gilt $\int_D E(X \,|\, \mathcal{D})\,dP = \int_D X\,dP \geq 0$. Ist $D = \{\, E(X \,|\, \mathcal{D}) < 0 \,\}$, so folgt $P(D) = 0$.
(e) $E(X \,|\, \mathcal{D}) = E(X^+ \,|\, \mathcal{D}) - E(X^- \,|\, \mathcal{D}) \Rightarrow |E(X \,|\, \mathcal{D})| \leq E(X^+ \,|\, \mathcal{D}) + E(X^- \,|\, \mathcal{D}) = E(|X| \,|\, \mathcal{D})$.
(f) Sei wieder Y die rechte Seite der behaupteten Gleichung, und sei $D' \in \mathcal{D}'$. Dann gilt $\int_{D'} Y\,dP = \int_{D'} E(X \,|\, \mathcal{D})\,dP = \int_{D'} X\,dP$, da $D' \in \mathcal{D}$ ist. Da Y $\mathcal{D}'$-meßbar ist, folgt die Aussage.
(g) Die konstante Abbildung $\omega \mapsto E(X)$ ist natürlich $\mathcal{D}$-meßbar. Für $D \in \mathcal{D}$ gilt wegen der Unabhängigkeit $\int_D X\,dP = E(1_D X) = P(D)E(X) = \int_D E(X)\,dP$. $\square$

Für bedingte Erwartungswerte gelten die üblichen Konvergenzsätze entsprechend:

Satz 5.15. *Seien X_n für alle $n \in \mathbb{N}$ und X integrierbare Zufallsgrößen mit $X_n \to X$ fast sicher, und $\mathcal{D}$ sei eine Teil-σ-Algebra von $\mathcal{F}$.*

(a) *Falls $X_n \geq 0$ fast sicher für alle $n \in \mathbb{N}$ gilt, so gilt das Lemma von Fatou:*

$$E(X \,|\, \mathcal{D}) \leq \liminf_{n\to\infty} E(X_n \,|\, \mathcal{D}) \quad \textit{fast sicher.}$$

(b) *Existiert $Y \in \mathcal{L}^1$ mit $|X_n| \leq Y$ fast sicher für alle $n \in \mathbb{N}$, so gilt der beschränkte Konvergenzsatz:*

$$E(X \,|\, \mathcal{D}) = \lim_{n\to\infty} E(X_n \,|\, \mathcal{D}) \quad \textit{fast sicher.}$$

Beweis. Der Beweis von (b) ergibt sich durch Anwenden von (a) auf $X_n + Y$ und $-X_n + Y$. Der Beweis von (a) folgt aus einer monotonen Version von (b):

$$(5.16) \qquad X \in \mathcal{L}^1,\ 0 \leq X_n \uparrow X \text{ fast sicher} \quad \Rightarrow \quad E(X_n \,|\, \mathcal{D}) \uparrow E(X \,|\, \mathcal{D}) \text{ fast sicher.}$$

Gelten nämlich $X_n \geq 0$ fast sicher für alle $n \in \mathbb{N}$ und $X_n \to X$ fast sicher (wie in (a)), dann gilt $Y_n := \inf_{m \geq n} X_m \uparrow X$ fast sicher, und aus (5.16) folgt

$$E(X \mid \mathcal{D}) = \lim_{n\to\infty} E(Y_n \mid \mathcal{D}) \leq \liminf_{n\to\infty} E(X_n \mid \mathcal{D}).$$

Es bleibt noch (5.16) zu zeigen: Aus $X_n \leq X_{n+1} \leq X$ fast sicher für alle $n \in \mathbb{N}$ folgt $E(X_n \mid \mathcal{D}) \leq E(X_{n+1} \mid \mathcal{D}) \leq E(X \mid \mathcal{D})$ fast sicher für alle $n \in \mathbb{N}$. Die Folge $(E(X_n \mid \mathcal{D}))_{n\in\mathbb{N}}$ konvergiert also fast sicher gegen eine $\mathcal{D}$-meßbare Zufallsgröße Y mit $Y \leq E(X \mid \mathcal{D})$. Für alle $D \in \mathcal{D}$ gilt

$$\begin{aligned} \int_D (E(X \mid \mathcal{D}) - Y)\, dP &= \int_D X\, dP - \int_D Y\, dP \\ &= \int_D X\, dP - \int_D \lim_{n\to\infty} E(X_n \mid \mathcal{D})\, dP \\ &= \int_D X\, dP - \lim_{n\to\infty} \int_D E(X_n \mid \mathcal{D})\, dP \\ &= \int_D X\, dP - \lim_{n\to\infty} \int_D X_n\, dP = 0. \end{aligned}$$

Somit folgt $Y = E(X \mid \mathcal{D})$ fast sicher. □

Satz 5.17. *Seien $X, Y \in \mathcal{L}^1$ und $X \cdot Y \in \mathcal{L}^1$. Ist Y $\mathcal{D}$-meßbar, so gilt $E(YX \mid \mathcal{D}) = YE(X \mid \mathcal{D})$ fast sicher.*

Beweis. Sind $X \geq 0$ und $Y = 1_D$ mit $D \in \mathcal{D}$, so folgt für jedes $D' \in \mathcal{D}$

$$\int_{D'} YE(X \mid \mathcal{D})\, dP = \int_{D\cap D'} E(X \mid \mathcal{D})\, dP = \int_{D\cap D'} X\, dP = \int_{D'} 1_D X\, dP.$$

Somit erfüllt $YE(X \mid \mathcal{D})$ die definierenden Eigenschaften eines bedingten Erwartungswertes von YX gegeben $\mathcal{D}$, das heißt, es gilt $YE(X \mid \mathcal{D}) = E(XY \mid \mathcal{D})$ fast sicher.

Ist $Y = \sum_{i=1}^n a_i 1_{D_i}$ mit $a_i \geq 0$ und $D_i \in \mathcal{D}$, also eine einfache, nichtnegative Zufallsgröße, so folgt diese Gleichung aus der Linearität. Sind $Y, X \geq 0$, so approximiert man Y von unten punktweise durch einfache, $\mathcal{D}$-meßbare Funktionen und verwendet Satz 5.15(b). Der allgemeine Fall mit $X, Y,\ X \cdot Y \in \mathcal{L}^1$ folgt durch die Zerlegungen $X = X^+ - X^-$ und $Y = Y^+ - Y^-$. □

Bedingte Wahrscheinlichkeiten sind Spezialfälle von bedingten Erwartungswerten: Man definiert $P(A \mid \mathcal{D}) := E(1_A \mid \mathcal{D})$. Die bedingte Wahrscheinlichkeit $P(A \mid \mathcal{D})$ ist eine $\mathcal{D}$-meßbare Zufallsgröße, die $\int_D P(A \mid \mathcal{D})\, dP = P(A \cap D)$ für alle $D \in \mathcal{D}$ erfüllt.

Mit Hilfe von bedingten Erwartungswerten können wir nun den Limes im Ergodensatz auch im nicht ergodischen Fall identifizieren.

Satz 5.18 (ERGODENSATZ). *Unter den Voraussetzungen von Satz 4.19 gilt*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} X \circ T^j = E(X \mid \mathcal{J}),$$

wobei $\mathcal{J}$ die σ-Algebra der T-invarianten Ereignisse ist.

Beweis. Sei $A \in \mathcal{J}$. Dann gilt $1_A \circ T^n = 1_A$ für alle $n \in \mathbb{N}$. Somit ist $(X1_A) \circ T^n = (X \circ T^n)1_A$. Dann folgt sofort aus Satz 4.19

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (X1_A) \circ T^j = \lim_{n\to\infty} 1_A \frac{1}{n} \sum_{j=0}^{n-1} (X \circ T^j) = 1_A \overline{X} \qquad \text{fast sicher,}$$

wobei $\overline{X}$ der fast sichere Limes von $\frac{1}{n} \sum_{j=0}^{n-1} X \circ T^j$ ist. Wendet man Satz 4.19 auf $X1_A$ an, so folgt aber auch, daß $\frac{1}{n} \sum_{j=0}^{n-1} (X1_A) \circ T^j$ gegen eine $\mathcal{J}$-meßbare Zufallsgröße $\overline{X}_A$ konvergiert mit $\int \overline{X}_A \, dP = \int X1_A \, dP$. Es muß fast sicher $\overline{X}_A = 1_A \overline{X}$ gelten. Somit folgt $\int_A X \, dP = \int_A \overline{X} \, dP$ für alle $A \in \mathcal{J}$. Damit ist $\overline{X} = E(X \mid \mathcal{J})$ bewiesen. $\square$

Kapitel 6

Verteilungskonvergenz

6a. Einführung und der zentrale Grenzwertsatz

Bisher wurde nur das Konvergenzverhalten von Zufallsgrößen untersucht (zum Beispiel fast sichere Konvergenz). Wichtiger ist aber oft das Verhalten der Verteilungen.

Beispiel 6.1. Es sei $(X_n)_n$ eine Folge unabhängiger Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$ für alle i mit einem $p \in (0, 1)$. Dann ist $S_n := X_1 + \cdots + X_n$ binomialverteilt mit Parametern p, n, und es gilt für alle $t \in \mathbb{R}$ (siehe Satz (5.10) in „Einführung in die Stochastik“)

$$\lim_{n\to\infty} P\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \le t\Big) = \Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx. \tag{6.2}$$

Konvergiert die Folge dieser Zufallsgrößen? Aus dem Kolmogoroff'schen 0-1-Gesetz und (6.2) folgt leicht (siehe das Argument in Beispiel 3.18), daß $\limsup_{n\to\infty}(S_n - np)/\sqrt{np(1-p)} = \infty$ fast sicher und $\liminf_{n\to\infty}(S_n - np)/\sqrt{np(1-p)} = -\infty$ fast sicher ist. Die Zufallsgrößen konvergieren also nicht fast sicher. Man kann auch leicht zeigen, daß die Folge nicht in Wahrscheinlichkeit konvergiert. Tatsächlich aber konvergiert die Folge der Verteilungen von $(S_n - np)/\sqrt{np(1-p)}$. Sei μ_n diese Verteilung, und μ sei die Standard-Normalverteilung.

Was bedeutet die Aussage „μ_n konvergiert gegen μ“? Gilt etwa $\mu_n(A) \to \mu(A)$ für alle $A \in \mathcal{B}$? Offensichtlich nicht: Für jedes n existiert eine endliche Menge A_n mit $\mu_n(A_n) = 1$. Setzt man $A := \bigcup_{n=1}^{\infty} A_n$, so ist A abzählbar. Es gilt $\mu_n(A) = 1$ für alle n, aber $\mu(A) = 0$.

Die grundlegende Definition soll nun ganz allgemein für Wahrscheinlichkeitsmaße auf einem metrischen Raum (S, d) gegeben werden, der ab jetzt fest gewählt sei.

Definition 6.3.

(a) 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.)

(b) Mit $\mathbb{P}(S)$ sei die Menge der Wahrscheinlichkeitsmaße auf $(S, \mathcal{B}_S)$ bezeichnet.

(c) Das Symbol $C(S)$ bezeichne die Menge der beschränkten stetigen Funktionen $S \to \mathbb{R}$.

Bemerkung 6.4. Es seien $\nu, \mu \in \mathbb{P}(S)$. 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 $f_n(x) = \max\{(1 - n\,d(x,F)), 0\}$. Offenbar gilt $f_n \downarrow 1_F$ für $n \to \infty$.

Aus dem beschränkten Konvergenzsatz 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 aus Bemerkung 1.9 die Behauptung $\mu = \nu$. $\square$

Definition 6.5.

(a) Seien $\mu_n, \mu \in \mathbb{P}(S)$ für $n \in \mathbb{N}$. Die Folge $(\mu_n)_{n\in\mathbb{N}}$ *konvergiert schwach* gegen μ (Notation: $\mu_n \xrightarrow{w} \mu$), wenn $\lim_{n\to\infty} \int f\,d\mu_n = \int f\,d\mu$ für alle $f \in C(S)$ gilt.

(b) Es seien X_n und X $(S, \mathcal{B}_S)$-wertige Zufallsgrößen, die auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind. Falls gilt $PX_n^{-1} \xrightarrow{w} PX^{-1}$, so sagt man, die Folge $(X_n)_n$ *konvergiert in Verteilung* gegen X, und schreibt oft $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$.

Für $S = \mathbb{R}$ ist die schwache Konvergenz äquivalent zu einem gewissen Konvergenzverhalten der Verteilungsfunktionen. Das soll weiter unten diskutiert werden (siehe Satz 6.9).

Der klassische zentrale Grenzwertsatz kann nun aus dem Stand bewiesen werden:

Satz 6.6 (ZENTRALER GRENZWERTSATZ). *Es seien X_n, $n \in \mathbb{N}$, unabhängige, identisch verteilte Zufallsgrößen in $\mathcal{L}^2$. Sei $a = EX_n$, $\sigma^2 = \operatorname{var}(X_n) > 0$, $S_n = \sum_{k=1}^n X_k$. Dann konvergiert $\mathcal{L}((S_n - na)/(\sqrt{n}\sigma))$ schwach gegen die Standardnormalverteilung.*

Beweis. Der einfachste Beweis verwendet charakteristische Funktionen, benötigt dann aber einige Vorbereitungen. Der folgende mirakulöse Beweis stammt von Charles Stein (1972). Sei $X_k' = (X_k - a)/\sigma$. Dann ist $a' := EX_k' = 0$ und $\sigma'^2 := \operatorname{var} X_k' = 1$. Ferner ist $(S_n - na)/\sqrt{n}\sigma = \sum_{j=1}^n X_j'/\sqrt{n}$. Man kann also ohne Einschränkung der Allgemeinheit annehmen, daß $a = 0$ und $\sigma^2 = 1$ gelten. Sei $f \in C(\mathbb{R})$ und $\bar{f}(x) := f(x) - \int_{-\infty}^{\infty} f(y)\varphi(y)\,dy$, wobei $\varphi(y) := e^{-y^2/2}/\sqrt{2\pi}$ sei. Nach Definition 6.5 ist zu zeigen, daß

$$\int f\,d(P(S_n/\sqrt{n})^{-1}) = Ef\Big(\frac{S_n}{\sqrt{n}}\Big) \overset{n\to\infty}{\longrightarrow} \int_{-\infty}^{\infty} f(y)\varphi(y)\,dy$$

gilt, das heißt daß $\lim_{n\to\infty} E\bar{f}(S_n/\sqrt{n}) = 0$ ist. Sei $h(x) := \int_{-\infty}^{x} \bar{f}(y)\varphi(y)\,dy/\varphi(x)$. Wegen $\int_{-\infty}^{\infty} \bar{f}\,\varphi\,dy = 0$ ist das auch gleich $-\int_x^{\infty} \bar{f}(y)\varphi(y)\,dy/\varphi(x)$. Sei $M = \sup_{x\in\mathbb{R}} |\bar{f}(x)|$. Mit Hilfe der Tatsache $\varphi'(x) = -x\varphi(x)$ errechnet man für $x > 0$:

$$|h(x)| \le \frac{M}{\varphi(x)} \int_x^{\infty} \varphi(t)\,dt \le \frac{M}{\varphi(x)} \int_x^{\infty} \frac{t}{x}\varphi(t)\,dt = \frac{M}{x},$$

und analog für $x < 0$ die Abschätzung $|h(x)| \leq M/|x|$. Daher ist die Abbildung $x \mapsto xh(x)$ beschränkt. Man errechnet auch leicht $h'(x) - xh(x) = \bar{f}(x)$ für $x \in \mathbb{R}$. Somit ist h' beschränkt und stetig. Es gilt:

$$\begin{aligned} E\bar{f}\Big(\frac{S_n}{\sqrt{n}}\Big) &= Eh'\Big(\frac{S_n}{\sqrt{n}}\Big) - E\Big(\frac{S_n}{\sqrt{n}}\, h\Big(\frac{S_n}{\sqrt{n}}\Big)\Big) \\ &= Eh'\Big(\frac{S_n}{\sqrt{n}}\Big) - \sum_{j=1}^{n} \frac{1}{\sqrt{n}}\, E\Big(X_j\, h\Big(\frac{S_n}{\sqrt{n}}\Big)\Big) \\ &= Eh'\Big(\frac{S_n}{\sqrt{n}}\Big) - \sqrt{n} E\Big(X_1\, h\Big(\frac{S_n}{\sqrt{n}}\Big)\Big) \\ &= Eh'\Big(\frac{S_n}{\sqrt{n}}\Big) - \sqrt{n} E\Big(X_1\, h\Big(\frac{\widetilde{S}_n}{\sqrt{n}} + \frac{X_1}{\sqrt{n}}\Big)\Big), \end{aligned}$$

wobei $\widetilde{S}_n = \sum_{j=2}^{n} X_j$ sei.

Wir entwickeln den zweiten Summanden nach der Taylor-Formel:

$$E\bar{f}\Big(\frac{S_n}{\sqrt{n}}\Big) = E\Big(h'\Big(\frac{S_n}{\sqrt{n}}\Big)\Big) - \underbrace{\sqrt{n}\, E\Big(X_1\, h\Big(\frac{\widetilde{S}_n}{\sqrt{n}}\Big)\Big)}_{=0 \text{ wegen } X_1 \perp\!\!\!\perp \widetilde{S}_n \text{ und } EX_1 = 0} - E\Big(X_1^2 \int_0^1 h'\Big(\frac{\widetilde{S}_n + sX_1}{\sqrt{n}}\Big) ds\Big).$$

Wegen $EX_1^2 = 1$ ist dies

$$= \underbrace{E\left(h'\Big(\frac{S_n}{\sqrt{n}}\Big) - h'\Big(\frac{\widetilde{S}_n}{\sqrt{n}}\Big)\right)}_{=:U_n} + \underbrace{E\left(X_1^2 \int_0^1 \Big(h'\Big(\frac{\widetilde{S}_n}{\sqrt{n}}\Big) - h'\Big(\frac{\widetilde{S}_n + sX_1}{\sqrt{n}}\Big)\Big) ds\right)}_{=:V_n}.$$

Für $m \in \mathbb{N}$ sei $A_{n,m}$ das Ereignis $\Big\{\Big|\frac{\widetilde{S}_n}{\sqrt{n}}\Big| \leq m\Big\}$, und es sei $H := \sup\{\, |h'(x)| : x \in \mathbb{R}\}$. Dann ist der obige Ausdruck

$$= E\big(U_n 1_{A_{n,m}}\big) + E\big(V_n 1_{A_{n,m}}\big) + E\big(U_n 1_{A^c_{n,m}}\big) + E\big(V_n 1_{A^c_{n,m}}\big).$$

Da h' auf $[-m, m]$ gleichmäßig stetig ist und $|U_n| \leq 2H$ sowie $|V_n| \leq 2HX_1^2 \in \mathcal{L}^1$ gelten, folgt wegen $\frac{X_1}{\sqrt{n}} \to 0$ nach dem beschränkten Konvergenzsatz:

$$\lim_{n\to\infty} E\big(U_n 1_{A_{n,m}}\big) = 0 \quad \text{und} \quad \lim_{n\to\infty} E\big(V_n 1_{A_{n,m}}\big) = 0 \quad \text{für alle } m \in \mathbb{N}.$$

Ferner ist

$$\begin{aligned} \big|E\big(U_n 1_{A^c_{n,m}}\big)\big| &\leq 2H\, P(A^c_{n,m}) \leq \frac{2H}{m^2} E\Big(\frac{\widetilde{S}_n}{\sqrt{n}}\Big)^2 \leq 2H/m^2 \quad \text{und} \\ \big|E\big(V_n 1_{A^c_{n,m}}\big)\big| &\leq 2H\, E\big(X_1^2 1_{A^c_{n,m}}\big) = 2H\, E(X_1^2)\, P(A^c_{n,m}) \leq 2H/m^2. \end{aligned}$$

Kombiniert man das, so ergibt sich $\limsup_{n\to\infty} \Big|E\bar{f}\Big(\frac{S_n}{\sqrt{n}}\Big)\Big| \leq \frac{4H}{m^2}$. Da $m \in \mathbb{N}$ beliebig ist, folgt $\lim_{n\to\infty} E\bar{f}\Big(\frac{S_n}{\sqrt{n}}\Big) = 0$. $\square$

Die Definition 6.5 ist formal sehr bequem; oft möchte man jedoch lieber wissen, für welche $A \in \mathcal{B}_S$ gilt: $\mu(A) = \lim_{n\to\infty} \mu_n(A)$.

Satz 6.7. *Es seien $\mu_n, \mu \in \mathbb{P}(S)$ für $n \in \mathbb{N}$. Die folgenden Aussagen sind äquivalent:*

(a) $\mu_n \xrightarrow{w} \mu$.
(b) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *für jede abgeschlossene Menge* $F \subset S$.
(c) $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$ *für jede offene Menge* $U \subset S$.
(d) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *für alle* $A \in \mathcal{B}_S$ *mit* $\mu(\partial A) = 0$.

Beweis. (a) $\Rightarrow$ (b). Seien F abgeschlossen, $\varepsilon > 0$, und sei $f_\varepsilon(x) := \max\{0, 1 - d(x,F)/\varepsilon\}$. Die Funktion f_ε ist beschränkt und stetig mit $1_F \leq f_\varepsilon$. Somit gilt $\limsup_{n\to\infty} \mu_n(F) \leq \lim_{n\to\infty} \int f_\varepsilon \, d\mu_n = \int f_\varepsilon \, d\mu$. Es gilt $f_\varepsilon \downarrow 1_F$ für $\varepsilon \downarrow 0$. Aus dem beschränkten Konvergenzsatz folgt $\int f_\varepsilon \, d\mu \downarrow \mu(F)$ für $\varepsilon \downarrow 0$. Demzufolge gilt $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.

(b) $\Leftrightarrow$ (c) folgt sofort aus der Tatsache, daß die offenen Mengen genau die Komplemente der abgeschlossenen sind.

((b) und (c)) $\Rightarrow$ (d). Sei $A \in \mathcal{B}_S$ mit $\mu(\partial A) = 0$; es sei $\mathring{A}$ sei das Innere, $\bar{A}$ der Abschluß von A. Dann gelten

$$\limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A}),$$
$$\liminf_{n\to\infty} \mu_n(A) \geq \liminf_{n\to\infty} \mu_n(\mathring{A}) \geq \mu(\mathring{A}).$$

Aus $\mu(\partial A) = 0$ folgt $\mu(\bar{A}) = \mu(\mathring{A}) = \mu(A)$. Somit folgt (d).

(d) $\Rightarrow$ (b). Sei $F \subset S$ abgeschlossen. Eine einfache Überlegung zeigt, daß $\partial(F^\delta) \subset \{\, x : d(x,F) = \delta \,\}$ für alle $\delta \geq 0$ gilt, wobei $F^\delta := \{\, x : d(x,F) \leq \delta \,\}$ sei. Die Mengen $\partial(F^\delta)$ mit $\delta > 0$ sind also paarweise disjunkt. Die Menge $\{\, \delta > 0 : \mu(\partial(F^\delta)) > 0 \,\} = \bigcup_{m=1}^\infty \{\, \delta > 0 : \mu(\partial(F^\delta)) \geq 1/m \,\}$ ist als abzählbare Vereinigung endlicher Mengen höchstens abzählbar. Es gibt somit eine fallende Nullfolge $(\delta_k)_k$ mit $\mu(\partial(F^{\delta_k})) = 0$ für alle $k \in \mathbb{N}$. Somit gilt $\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F^{\delta_k}) = \mu(F^{\delta_k})$ für alle k. Wegen $F^{\delta_k} \downarrow F$ haben wir $\mu(F^{\delta_k}) \downarrow \mu(F)$ für $k \to \infty$, also folgt (b).

(b) $\Rightarrow$ (a). Sei $f \in C(S)$. Dann existieren $a > 0$ und $b \in \mathbb{R}$ mit $\bar{f}(x) := af(x) + b \in (0,1)$ für alle $x \in S$. Die Menge $F_i^{(k)} := \{\, x : \bar{f}(x) \geq i/k \,\}$ ist abgeschlossen für $k \in \mathbb{N}$ und $0 \leq i \leq k$. Für $\nu \in \mathbb{P}(S)$ und alle $k \in \mathbb{N}$ gilt

$$\sum_{i=1}^k \frac{i-1}{k} \big(\nu(F_{i-1}^{(k)}) - \nu(F_i^{(k)})\big) \leq \int \bar{f} \, d\nu \leq \sum_{i=1}^k \frac{i}{k} \big(\nu(F_{i-1}^{(k)}) - \nu(F_i^{(k)})\big),$$

das heißt $\frac{1}{k} \sum_{i=1}^{k-1} \nu(F_i^{(k)}) \leq \int \bar{f} \, d\nu \leq \frac{1}{k} \sum_{i=0}^{k-1} \nu(F_i^{(k)})$ (man beachte $F_k^{(k)} = \emptyset$). Angewandt auf μ_n und μ an Stelle von ν gibt das zusammen mit (b): $\limsup_{n\to\infty} \int \bar{f} \, d\mu_n \leq \int \bar{f} \, d\mu + 1/k$. Da k beliebig war, folgt $\limsup_{n\to\infty} \int \bar{f} \, d\mu_n \leq \int \bar{f} \, d\mu$, und wegen $a > 0$ folgt $\limsup_{n\to\infty} \int f \, d\mu_n \leq \int f \, d\mu$. Diese Ungleichung, angewandt auf $-f$, gibt $\liminf_{n\to\infty} \int f \, d\mu_n \geq \int f \, d\mu$. $\square$

In diesem Zusammenhang ist das folgende Lemma nützlich, das eine hinreichende (aber nicht notwendige) Bedingung für schwache Konvergenz gibt:

Lemma 6.8. *Seien $\mu_n, \mu \in \mathbb{P}(S)$ für $n \in \mathbb{N}$. Es sei $\mathcal{U}$ eine durchschnittsstabile 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 nach dem Ein- und Ausschlußprinzip

$$\mu_n\Big(\bigcup_{j=1}^m A_j\Big) = \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<j_2<\dots<j_k} \mu_n(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k})$$
$$\xrightarrow{n\to\infty} \sum_{k=1}^m (-1)^{k+1} \sum_{j_1<j_2<\dots<j_k} \mu_n(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = \mu\Big(\bigcup_{j=1}^m A_j\Big).$$

Sei $G \subset S$ offen. Dann existieren für jedes $\varepsilon > 0$ ein $m \in \mathbb{N}$ und $A_1, \dots, A_m \in \mathcal{U}$ mit $\mu(G) \le \mu\left(\bigcup_{j=1}^m A_j\right) + \varepsilon$ und $\bigcup_{j=1}^m A_j \subset G$. Somit ist

$$\liminf_{n\to\infty} \mu_n(G) \ge \liminf_{n\to\infty} \mu_n\Big(\bigcup_{j=1}^m A_j\Big) = \mu\Big(\bigcup_{j=1}^m A_j\Big) \ge \mu(G) - \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, folgt die Aussage (c) von Satz 6.7. □

Mit Hilfe dieses Lemmas leiten wir für $(S, \mathcal{B}_S) = (\mathbb{R}, \mathcal{B})$ ein einfaches notwendiges und hinreichendes Kriterium für schwache Konvergenz her.

Satz 6.9. *Es seien $\mu_n, \mu \in \mathbb{P}(\mathbb{R})$ für $n \in \mathbb{N}$ mit Verteilungsfunktionen F_n von μ_n beziehungsweise F von μ. Es gilt genau dann $\mu_n \xrightarrow{w} \mu$, wenn für jedes $t \in \mathbb{R}$, in dem F stetig ist, $\lim_{n\to\infty} F_n(t) = F(t)$ ist.*

Beweis. „$\Rightarrow$". Es gelte $\mu_n \xrightarrow{w} \mu$. Ist F stetig in $t \in \mathbb{R}$, so gilt $\mu(\partial(-\infty, t]) = \mu(\{t\}) = 0$. Somit folgt $\lim_{n\to\infty} F_n(t) = F(t)$ nach Kriterium (d) aus Satz 6.7.

„$\Leftarrow$". Es gelte $\lim_{n\to\infty} F_n(t) = F(t)$ für alle $t \in D := \{\, x \in \mathbb{R}\colon\ F \text{ ist stetig in } x \,\}$. Das Komplement $D^c = \{\, t\colon \mu(\{t\}) > 0 \,\} = \bigcup_{n=1}^\infty \{\, t\colon \mu(\{t\}) \ge 1/n \,\}$ ist abzählbar. Demzufolge ist D dicht in $\mathbb{R}$.

Das System $\mathcal{U} := \{(a, b]\colon a \le b;\ a, b \in D\}$ ist durchschnittsstabil. Jedes offene Intervall läßt sich als abzählbare Vereinigung von Elementen aus $\mathcal{U}$ schreiben, und da jede offene Menge sich als abzählbare Vereinigung von offenen Intervallen schreiben läßt, erfüllt $\mathcal{U}$ die Bedingungen von Lemma 6.8. Es gilt

$$\lim_{n\to\infty} \mu_n((a, b]) = \lim_{n\to\infty} (F_n(b) - F_n(a)) = F(b) - F(a) = \mu((a, b]) \quad \text{für } (a, b] \in \mathcal{U}.$$

Nach Lemma 6.8 gilt $\mu_n \xrightarrow{w} \mu$. □

Als Folgerung von Satz 6.6 und Satz 6.9 ergibt sich der *zentrale Grenzwertsatz* in der üblichen Formulierung: (Die Funktion Φ sei wie in (6.2) definiert.)

Satz 6.10 (ZENTRALER GRENZWERTSATZ). *Es sei $(X_n)_n$ eine Folge von Zufallsgrößen, die die Bedingungen von Satz 6.6 erfüllen. Dann gilt für alle $t \in \mathbb{R}$:*

$$\lim_{n\to\infty} P\Big(\frac{S_n - na}{\sqrt{n}\sigma} \le t\Big) = \Phi(t).$$

Beweis. Φ ist überall stetig. $\square$

6b. Weitere allgemeine Eigenschaften der Verteilungskonvergenz und Verteilungskonvergenz auf $\mathbb{R}^d$

Wir wollen zunächst das Verhalten induzierter Wahrscheinlichkeitsmaße untersuchen, wenn die Ursprungsmaße schwach konvergieren. Es sei also $(\mu_n)_n$ eine Folge von Wahrscheinlichkeitsmaßen auf $(S, \mathcal{B}_S)$ mit $\mu_n \xrightarrow{w} \mu \in \mathbb{P}(S)$. Ist h eine meßbare Abbildung von S in einen zweiten metrischen Raum, so braucht natürlich nicht $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$ zu gelten.

Beispiel 6.11. Sei $(x_n)_n$ eine Folge in $S \setminus \{x\}$, die gegen ein $x \in S$ konvergiert. Dann gilt $\delta_{x_n} \xrightarrow{w} \delta_x$. Ist $h\colon S \to \mathbb{R}$ durch $h(y) = 1_{\{x\}}(y)$ definiert, so gelten $\delta_{x_n} h^{-1} = \delta_0$ und $\delta_x h^{-1} = \delta_1$, also konvergiert $\delta_{x_n} h^{-1}$ nicht schwach gegen $\delta_x h^{-1}$.

Ist h jedoch stetig, so überträgt sich die schwache Konvergenz auf die induzierten Maße:

Lemma 6.12. *Seien (S, d) und (S', d') zwei metrische Räume, und $h\colon S \to S'$ sei stetig. Es seien μ_n und μ Wahrscheinlichkeitsmaße auf $(S, \mathcal{B}_S)$ 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 $f \circ h \in C(S)$. Somit gilt

$$\lim_{n\to\infty} \int f \, d(\mu_n h^{-1}) = \lim_{n\to\infty} \int (f \circ h)\, d\mu_n = \int (f \circ h)\, d\mu = \int f\, d(\mu h^{-1}).$$

$\square$

Für gewisse Anwendungen, auf die wir später noch stoßen, ist die Forderung nach Stetigkeit von h zu stark einschränkend. Wir benötigen zunächst den folgenden Hilfssatz:

Lemma 6.13. *Es seien (S, d) und (S', d') zwei metrische Räume, und $h\colon S \to S'$ sei $\mathcal{B}_S$-$\mathcal{B}_{S'}$-meßbar. Dann ist $D_h := \{\, x \in S :\ h \text{ ist nicht stetig in } x \,\} \in \mathcal{B}_S$.*

Beweis. Für $m, n \in \mathbb{N}$ sei $A_{m,n} = \{\, x \in S$: Es gibt $y, z \in S$ mit $d(x, y) < 1/m$ und $d(x, z) < 1/m$ sowie $d'(h(y), h(z)) \ge 1/n \,\}$. Offensichtlich ist $A_{m,n}$ offen. Somit ist

$$D_h = \bigcup_n \bigcap_m A_{m,n} \in \mathcal{B}_S.$$

$\square$

Satz 6.14. *Es liege dieselbe Situation wie in Lemma 6.12 vor, h sei jedoch nur als $\mathcal{B}_S$-$\mathcal{B}_{S'}$-meßbar vorausgesetzt. Gilt $\mu(D_h) = 0$, so folgt ebenfalls $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$.*

Beweis. Sei $F \subset S'$ abgeschlossen. Dann gilt

$$\limsup_{n\to\infty} \mu_n(h^{-1}(F)) \leq \limsup_{n\to\infty} \mu_n(\overline{h^{-1}(F)}) \leq \mu(\overline{h^{-1}(F)}).$$

Es ist $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$. Wegen $\mu(D_h) = 0$ folgt

$$\mu(\overline{h^{-1}(F)}) = \mu(h^{-1}(F)),$$

und aus Kriterium (b) von Satz 6.7 folgt die Behauptung. □

Eine der sehr nützlichen Eigenschaften der schwachen Konvergenz ist die, daß es „verhältnismäßig große“ kompakte beziehungsweise relativ kompakte Mengen gibt.

Definition 6.15.

(a) Eine Teilmenge $\Gamma \subset \mathbb{P}(S)$ heißt (sequentiell) *relativ kompakt,* wenn jede Folge $(\mu_n)_n$ in Γ eine schwach konvergente Teilfolge hat. (Der Grenzwert muß nicht in Γ liegen.)
(b) Eine Teilmenge $\Gamma \subset \mathbb{P}(S)$ 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 $\mu \in \Gamma$ ist.

Bemerkung 6.16.

(a) Ist S kompakt, so ist $\mathbb{P}(S)$ offenbar straff.
(b) $\mathbb{P}(\mathbb{R})$ ist nicht straff, weil nämlich schon die Familie $\{\, \delta_x : x \in \mathbb{R} \}$ nicht straff ist.
(c) Ein einzelnes Wahrscheinlichkeitsmaß $\mu \in \mathbb{P}(S)$ 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 Wahrscheinlichkeitsmaß straff. (Es gilt ja $\mu(K_n) \uparrow \mu(S) = 1$.) Dies ist für $S = \mathbb{R}$ oder $S = \mathbb{R}^d$ der Fall.

Erstaunlicherweise gibt es jedoch eine große Klasse von metrischen Räumen, die nicht unbedingt σ-kompakt sind und in denen jedes Wahrscheinlichkeitsmaß straff ist: nämlich vollständige separable Räume. 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 Supremummetrik. Unendlichdimensionale Banachräume sind nie σ-kompakt.

Die Aussage, daß jedes Wahrscheinlichkeitsmaß auf einem vollständigen, separablen metrischen Raum straff ist, ist ein Spezialfall des Satzes von Prohorov, der in Anhang A3 bewiesen wird:

Satz 6.17 (SATZ VON PROHOROV). *Es sei S vollständig und separabel. Dann ist jede Teilmenge von $\mathbb{P}(S)$ genau dann relativ kompakt, wenn sie straff ist.*

Um den Satz anzuwenden, benötigt man eine Eigenschaft der schwachen Konvergenz, die jeder „vernünftige" Konvergenzbegriff hat:

Lemma 6.18. *Seien $\mu_n, \mu \in \mathbb{P}(S)$ 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. Das Lemma folgt unmittelbar aus der Definition 6.5 und der Tatsache, daß reelle Zahlenfolgen sich so verhalten. $\square$

Als Anwendung davon kann ein sehr nützliches Kriterium für schwache Konvergenz auf $\mathbb{R}^d$ bewiesen werden. Für $x \in \mathbb{R}^d$ sei $\pi_x \colon \mathbb{R}^d \to \mathbb{R}$ definiert durch $\pi_x(y) = \langle x, y \rangle$ (Hier sei $\langle \cdot\,, \cdot \rangle$ das Euklid'sche Skalarprodukt).

Satz 6.19 (SATZ VON CRAMÉR-WOLD). *Es seien μ_n und μ Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d)$ für $n \in \mathbb{N}$. Dann gilt $\mu_n \xrightarrow{w} \mu$ genau dann, wenn $\mu_n \pi_x^{-1} \xrightarrow{w} \mu \pi_x^{-1}$ 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 6.12 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 $e_i \in \mathbb{R}^d$.

Da $\mu_n \pi_i^{-1}$ schwach konvergiert, also insbesondere $\{\, \mu_n \pi_i^{-1} \colon n \in \mathbb{N} \}$ relativ kompakt ist, existiert für jedes $\varepsilon > 0$ eine kompakte Menge $K_i \subset \mathbb{R}$ mit $\mu_n(\pi_i^{-1}(K_i)) \geq 1 - \frac{\varepsilon}{d}$ für alle $n \in \mathbb{N}$ und alle $i \in \{1, \dots, d\}$. Die Menge $K := \bigcap_{i=1}^d \pi_i^{-1}(K_i) \subset \mathbb{R}^d$ ist abgeschlossen und beschränkt in $\mathbb{R}^d$, also kompakt. Für alle $n \in \mathbb{N}$ gilt:

$$\mu_n(K^c) = \mu_n\Big(\bigcup_{i=1}^d (\pi_i^{-1}(K_i))^c\Big) \leq \sum_{i=1}^d \mu_n(\pi_i^{-1}(K_i^c)) \leq \varepsilon.$$

Aus Satz 6.17 folgt, daß $\{\, \mu_n \colon n \in \mathbb{N} \}$ relativ kompakt ist. Sei $(\mu'_n)_n$ eine beliebige Teilfolge von $(\mu_n)_n$. Diese hat ihrerseits eine konvergente Teilfolge $(\mu''_n)_n$ mit $\mu''_n \xrightarrow{w} \mu''$ für ein $\mu'' \in \mathbb{P}(\mathbb{R}^d)$. 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

$$\widehat{\mu}(x) = \int e^{i\langle x,y \rangle} \mu(dy) = \int e^{it} (\mu \pi_x^{-1})(dt) = \int e^{it} (\mu'' \pi_x^{-1})(dt) = \widehat{\mu''}(x).$$

Aus Satz 2.49 folgt $\mu = \mu''$, und Lemma 6.18 läßt die Behauptung $\mu_n \xrightarrow{w} \mu$ folgen. $\square$

Mit Satz 6.19 läßt sich leicht eine mehrdimensionale Version des zentralen Grenzwertsatzes beweisen. Mit $|\cdot|$ bezeichnen wir die Euklid'sche Norm im $\mathbb{R}^d$.

Satz 6.20. *Es sei $(X_n)_n$ eine Folge unabhängiger, identisch verteilter, d-dimensionaler Zufallsvektoren. Es gelte $E|X_i|^2 < \infty$. Seien $a = EX_1$ und Σ die Kovarianzmatrix der X_i. Dann gilt $\mathcal{L}\left(\sum_{i=1}^n (X_i - a)/\sqrt{n}\right) \xrightarrow{w} \mu$, wobei μ die d-dimensionale Normalverteilung mit Mittel 0 und Kovarianzmatrix Σ ist.*

Beweis. Sei $T_n := \sum_{i=1}^n (X_i - a)/\sqrt{n}$. Nach Satz 6.19 genügt es zu zeigen, daß für jedes $x \in \mathbb{R}^d$ gilt: $\mathcal{L}(\langle x, T_n\rangle) \xrightarrow{w} \mu\pi_x^{-1}$. Es ist $\langle x, T_n\rangle = \sum_{i=1}^n (\langle x, X_i\rangle - \langle x, a\rangle)/\sqrt{n}$.

Die $\langle x, X_i\rangle$, $i \in \mathbb{N}$, sind unabhängige, identisch verteilte eindimensionale Zufallsgrößen mit Erwartungswert $\langle x, a\rangle$ und Varianz $\sigma_x^2 = E(\langle x, X_i - a\rangle^2) = x^t\Sigma x$, wenn x als Spaltenvektor geschrieben wird.

Ist $\sigma_x^2 > 0$, so konvergiert $\mathcal{L}(\langle x, T_n/\sigma_x\rangle)$ nach Satz 6.6 gegen die Standardnormalverteilung, also konvergiert $\mathcal{L}(\langle x, T_n\rangle)$ nach Lemma 6.12 (mit $h\colon \mathbb{R} \to \mathbb{R}$, $h(y) = \sigma_x y$) gegen die Normalverteilung mit Mittel 0 und Varianz σ_x^2. Gilt $\sigma_x^2 = 0$, so ist $\langle x, T_n\rangle = 0$ fast sicher nach Lemma 2.43(c), und somit gilt trivialerweise $\mathcal{L}(\langle x, T_n\rangle) \xrightarrow{w} \delta_0$, und δ_0 ist nach unserer Konvention die Normalverteilung mit Mittel 0 und Varianz 0.

Nun ist aber $\mu\pi_x^{-1}$ die Normalverteilung mit Mittel 0 und Varianz σ_x^2. Damit ist Satz 6.20 bewiesen. □

Aus Satz 6.20 und Lemma 6.12 (oder Satz 6.14) können leicht neue Grenzwertsätze hergeleitet werden. Beispiele dafür werden in den Übungen diskutiert.

6c. Schwache Konvergenz und Konvergenz in Wahrscheinlichkeit von Zufallsgrößen

Wie bisher sei (S, d) ein metrischer Raum. Es soll nun definiert werden, was es heißt, daß eine Folge $(X_n)_n$ von $(S, \mathcal{B}_S)$-wertigen Zufallsgrößen in Wahrscheinlichkeit gegen eine Zufallsgröße X konvergiert. Naheliegend ist es, die reellwertige Zufallsgröße $d(X_n, X)$ zu betrachten. Ist das stets eine Zufallsgröße? Leider nicht in jedem Fall! Man muß dazu voraussetzen, daß S separabel ist.

Ist (S, d) ein metrischer Raum, so betrachten wir den Produktraum $(S \times S,\ d')$, wobei $d'((x_1, x_2),\ (y_1, y_2)) = (d(x_1, y_1)^2 + d(x_2, y_2)^2)^{1/2}$ sei.

Lemma 6.21. *Ist S separabel, so ist $\mathcal{B}_{S\times S} = \mathcal{B}_S \otimes \mathcal{B}_S$.*

Beweisskizze. Jede Produktmenge $A \times B$, wobei $A \subset S$ und $B \subset S$ offen sind, ist offen in $S \times S$, das heißt sie liegt 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 wird 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, das heißt: $\mathcal{B}_{S\times S} \subset \mathcal{B}_S \otimes \mathcal{B}_S$. □

Sind nun X und Y zwei $(S, \mathcal{B}_S)$-wertige Zufallsgrößen, so ist (X, Y) eine $(S \times S,\ \mathcal{B}_S \otimes \mathcal{B}_S)$-wertige Zufallsgröße. Die Abbildung $d\colon S \times S \to \mathbb{R}$ ist aber d'-stetig, also $\mathcal{B}_{S\times S}$-$\mathcal{B}$-meßbar. Somit ist die folgende Definition sinnvoll:

Definition 6.22. Es sei S separabel und X sowie X_n für $n \in \mathbb{N}$ Zufallsgrößen mit Werten in $(S, \mathcal{B}_S)$. Wir sagen, die Folge $(X_n)_n$ *konvergiert in Wahrscheinlichkeit* gegen X, falls $d(X_n, X)$ in Wahrscheinlichkeit gegen 0 konvergiert, das heißt wenn für alle $\varepsilon > 0$ gilt: $\lim_{n\to\infty} P(d(X_n, X) \geq \varepsilon) = 0$.

Satz 6.23. *Konvergiert X_n in Wahrscheinlichkeit gegen X, so gilt $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$.*

Beweis. Sei $A \in \mathcal{B}_S$ mit $P(X \in \partial A) = 0$. Für $\varepsilon > 0$ gilt

$$P(X_n \in A,\, X \notin A) \leq P(d(X_n, X) \geq \varepsilon) + P(d(X, A) < \varepsilon,\, X \notin A).$$

Dies und dasselbe mit A^c ergibt aufgrund von $\lim_{n\to\infty} P(d(X_n, X) \geq \varepsilon) = 0$ die Abschätzung:

$$\limsup_{n\to\infty} P(\{X_n \in A\} \triangle \{X \in A\}) \leq P(d(X, A) < \varepsilon,\, X \notin A) + P(d(X, A^c) < \varepsilon,\, X \in A).$$

Für den Grenzübergang $\varepsilon \downarrow 0$ gelten $\{\, d(X, A) < \varepsilon,\, X \notin A \,\} \downarrow \{\, X \in \partial A \cap A^c \,\}$ und $\{\, d(X, A^c) < \varepsilon,\, X \in A \,\} \downarrow \{\, X \in A \cap \partial A \,\}$. Wegen $P(X \in \partial A) = 0$ folgt $\lim_{n\to\infty} P(\{X_n \in A\} \triangle \{X \in A\}) = 0$, das heißt $\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$. Nach Satz 6.7(d) gilt $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$. $\square$

Die Umkehrung dieses Satzes ist nicht richtig (Übungsaufgabe).

Ein anderes nützliches Ergebnis in diesem Zusammenhang ist das folgende:

Lemma 6.24. *Es sei S separabel, und $(X_n)_n$ und $(X'_n)_n$ seien zwei Folgen von $(S, \mathcal{B}_S)$-wertigen Zufallsgrößen. 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 $F_\varepsilon := \{\, x \colon d(x, F) \leq \varepsilon \,\}$. Dann gilt

$$\limsup_{n\to\infty} P(X'_n \in F) \leq \limsup_{n\to\infty} P(d(X_n, X'_n) \geq \varepsilon) + \limsup_{n\to\infty} P(X_n \in F_\varepsilon) \leq \mu(F_\varepsilon).$$

Für $\varepsilon \downarrow 0$ gilt $\mu(F_\varepsilon) \downarrow \mu(F)$. $\square$

6d. Der Satz von Donsker

Wir betrachten $S = C[0,1]$, die Menge der stetigen Funktionen $[0,1] \to \mathbb{R}$. Für $f, g \in C[0,1]$ sei $d(f, g) = \sup_{t\in[0,1]} |f(t) - g(t)|$.

Bekanntlich ist $(C[0,1], d)$ vollständig und separabel. (Wir schreiben auch C statt $C[0,1]$.) Für $n \in \mathbb{N}$ und $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$ sei $\pi_{t_1,\dots,t_n} : C \to \mathbb{R}^n$ die endlich dimensionale Projektion $f \mapsto (f(t_1),\, f(t_2), \dots, f(t_n))$.

Lemma 6.25. *Es gilt*

$$\mathcal{B}_C = \bigvee_{t\in[0,1]} \pi_t^{-1}(\mathcal{B}).$$

Beweis. Sei $\mathcal{B}' := \bigvee_{t\in[0,1]} \pi_t^{-1}(\mathcal{B})$. Wir zeigen $\mathcal{B}_C = \mathcal{B}'$.

Da π_t stetig ist, ist für offenes $U \subset \mathbb{R}$ auch $\pi_t^{-1}(U)$ offen, liegt also in $\mathcal{B}_C$. Daraus folgt $\mathcal{B}' \subset \mathcal{B}_C$.

Wir zeigen nun die umgekehrte Inklusion. Für $f \in C[0,1]$ und $\varepsilon > 0$ sei $B_\varepsilon(f) := \{\, g \in C : d(f,g) \le \varepsilon \,\}$. Dann ist

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

Da $C[0,1]$ separabel ist, ist jede offene Menge als abzählbare Vereinigung von derartigen „Kugeln“ darstellbar. Somit folgt $\mathcal{B}_C \subset \mathcal{B}'$. □

Wir streben nun langsam der Formulierung des Satzes von Donsker zu. Dieser ist eine Verallgemeinerung des zentralen Grenzwertsatzes, indem nicht nur die Asymptotik der Verteilung von $S_n/\sqrt{n}$ (wobei $S_n = \sum_{i=1}^n X_i$ mit unabhängig identisch verteilten X_i) untersucht wird, sondern die Verteilung des gesamten Pfades. Aus rein formalen Gründen ist es bequem, den Pfad als stetige Funktion darzustellen.

Es sei also $(X_i)_{i\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter, eindimensionaler Zufallsgrößen, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Es sollen $EX_i = 0$ gelten (falls nicht, ersetzen wir X_i durch $X_i - EX_i$) und $\sigma^2 := EX_i^2 \in (0,\infty)$. Wir setzen $S_0 := 0$ und $S_n := \sum_{i=1}^n X_i$. Für $n \in \mathbb{N}$ und $\omega \in \Omega$ definieren wir die Abbildung $Y_n(\omega, .) \colon [0,1] \to \mathbb{R}$ durch

$$Y_n\Big(\omega, \frac{k}{n}\Big) := \frac{S_k(\omega)}{\sqrt{n}\,\sigma} \quad (k = 0,1,\dots,n) \quad \text{und lineare Interpolation.} \tag{6.26}$$

Man kann schreiben $Y_n(\omega,t) = (S_{[nt]}(\omega) + (nt - [nt])X_{[nt]+1}(\omega))/(\sigma\sqrt{n})$, wobei $[x]$ der ganzzahlige Anteil der reellen Zahl x ist. Wir können Y_n als Abbildung $\Omega \to C[0,1]$ auffassen. Für ein festes $t \in [0,1]$ ist die reelle Zufallsgröße $Y_n(\,.\,,t)$ offenbar $\mathcal{F}$-$\mathcal{B}$-meßbar. Nach Lemma 6.25 folgt, daß $Y_n : \Omega \to C[0,1]$ eine $(C, \mathcal{B}_C)$-wertige Zufallsgröße ist. Ohne Einschränkung der Allgemeinheit werden wir in Zukunft annehmen, daß $\sigma^2 = 1$ gilt (sonst ersetze X_i durch X_i/σ).

Der Satz von Donsker besagt, daß $\mathcal{L}(Y_n)$ schwach gegen ein Wahrscheinlichkeitsmaß auf $(C, \mathcal{B}_C)$ konvergiert.

Da $\pi_{t_1,\dots,t_m}$ stetig ist, ist nach Lemma 6.12 für die Konvergenz von $\mu_n := PY_n^{-1}$ notwendig, daß $\mu_n\pi_{t_1,\dots,t_m}^{-1}$ für $n \to \infty$ auf $(\mathbb{R}^m, \mathcal{B}^m)$ konvergiert.

Satz 6.27. *Für jedes $m \in \mathbb{N}$ und $0 \le t_1 < t_2 < \cdots < t_m \le 1$ konvergiert $\mu_n\pi_{t_1,\dots,t_m}^{-1}$ schwach auf $(\mathbb{R}^m, \mathcal{B}^m)$ gegen die m-dimensionale Normalverteilung mit Erwartungswert 0 und Kovarianzmatrix* $(\min\{t_i, t_j\})_{i,j}$.

Wir betrachten zunächst zwei Spezialfälle: Für $m = 1$, $t_1 = 1$ ist $\mu_n \pi_1^{-1} = \mathcal{L}(Y_n(1)) = \mathcal{L}(S_n/\sqrt{n})$, was nach dem zentralen Grenzwertsatz gegen die Standard-Normalverteilung konvergiert. Für $m = 1$ und $t_1 = 0$ ist $\mu_n \pi_0^{-1} = \mathcal{L}(Y_n(0)) = \delta_0$.

Als Vorbereitung zum Beweis von Satz 6.27 benötigen wir das folgende

Lemma 6.28. *Sei $d \in \mathbb{N}$, und für $j = 1, \dots, d$ sei $(\mu_n^{(j)})_n$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$ mit $\mu_n^{(j)} \xrightarrow{w} \mu^{(j)} \in \mathbb{P}(\mathbb{R})$ für $n \to \infty$. 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. Sei $A_j = \{ x \in \mathbb{R} \colon \mu^{(j)}(\{x\}) = 0 \}$. Das Komplement 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 $\{ (a_j, b_j) \colon a_j, b_j \in B_j \}$ eine abzählbare Basis der Topologie von $\mathbb{R}$, also ist $\mathcal{U} := \{ (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d) \colon a_j, b_j \in B_j$ für $j = 1, \dots, d \}$ eine Basis der Topologie von $\mathbb{R}^d$. $\mathcal{U}$ ist durchschnittstabil, und für $(a_1, b_1) \times \cdots \times (a_d, b_d) \in \mathcal{U}$ gilt wegen des Satzes 6.7(d)

$$\begin{aligned}
\mu_n^{(1)} \otimes \cdots \otimes \mu_n^{(d)}((a_1, b_1) \times \cdots \times (a_d, b_d)) &= \prod_{j=1}^{d} \mu_n^{(j)}((a_j, b_j)) \to \prod_{j=1}^{d} \mu^{(j)}((a_j, b_j)) \\
&= \mu^{(1)} \otimes \cdots \otimes \mu^{(d)}((a_1, b_1) \times \cdots \times (a_d, b_d)).
\end{aligned}$$

Lemma 6.28 folgt nun aus Lemma 6.8. □

Beweis von Satz 6.27. Man kann voraussetzen, daß $t_1 > 0$ gilt. Seien $Z_1^{(n)} := \sum_{i=1}^{[nt_1]} X_i/\sqrt{n}$, $Z_2^{(n)} := \sum_{i=[nt_1]+1}^{[nt_2]} X_i/\sqrt{n}, \dots,$ $Z_m^{(n)} := \sum_{i=[nt_{m-1}]+1}^{[nt_m]} X_i/\sqrt{n}$. (Für $x \in \mathbb{R}$ sei $[x] := \max\{ n \in \mathbb{Z} \colon n \leq x \}$, per Konvention ist $\sum_{i=1}^{0} := 0$.) Die Zufallsgrößen $Z_1^{(n)}, \dots, Z_m^{(n)}$ sind für jedes $n \in \mathbb{N}$ offenbar unabhängig. Es genügt daher nach Lemma 6.28, das Konvergenzverhalten von $(Z_j^{(n)})_n$ für festes j zu untersuchen. Es gilt $\mathcal{L}(Z_j^{(n)}) = \mathcal{L}\left(\sum_{i=1}^{k(n)} X_i/\sqrt{n}\right)$, wobei wir $t_0 := 0$ und $k(n) := [nt_j] - [nt_{j-1}]$ setzen. Wegen $\lim_{n\to\infty} k(n) = \infty$ folgt aus dem zentralen Grenzwertsatz 6.6 beziehungsweise aus Satz 6.10 für alle $s \in \mathbb{R}$:

$$\lim_{n\to\infty} P\Big(\sum_{j=1}^{k(n)} \frac{X_j}{\sqrt{k(n)}} \leq s\Big) = \Phi(s) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-x^2/2}\, dx.$$

Nun gilt aber $\lim_{n\to\infty} \frac{k(n)}{n} = t_j - t_{j-1}$. Daraus folgt für alle $\varepsilon > 0$ und $s \in \mathbb{R}$:

$$\overline{\lim_{n\to\infty}}\, P\Big(\sum_{j=1}^{k(n)} \frac{X_j}{\sqrt{n}} \leq s\Big) \leq \lim_{n\to\infty} P\Big(\sum_{j=1}^{k(n)} \frac{X_j}{\sqrt{k(n)}} \leq \frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big) = \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} + \varepsilon\Big)$$

und

$$\underline{\lim_{n\to\infty}}\, P\Big(\sum_{j=1}^{k(n)} \frac{X_j}{\sqrt{n}} \leq s\Big) \geq \lim_{n\to\infty} P\Big(\sum_{j=1}^{k(n)} \frac{X_j}{\sqrt{k(n)}} \leq \frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big) = \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}} - \varepsilon\Big).$$

Da $\varepsilon > 0$ beliebig ist, folgt

$$\lim_{n\to\infty} P\Big(\sum_{j=1}^{k(n)} \frac{X_j}{\sqrt{n}} \le s\Big) = \Phi\Big(\frac{s}{\sqrt{t_j - t_{j-1}}}\Big).$$

Die Abbildung $s \mapsto \Phi\left(\frac{s}{\sqrt{t_j - t_{j-1}}}\right)$ ist aber die Verteilungsfunktion der eindimensionalen Normalverteilung mit Erwartungswert 0 und Varianz $t_j - t_{j-1}$.

Nach Lemma 6.28 folgt also, daß $\mathcal{L}(Z_1^{(n)}, \dots, Z_m^{(n)})$ für $n \to \infty$ gegen die Produktwahrscheinlichkeit dieser Normalverteilungen konvergiert, was die m-dimensionale Normalverteilung ν mit Erwartungswert 0 und Kovarianzmatrix $(\delta_{ij}(t_j - t_{j-1}))_{i,j}$ ist.

Wir definieren jetzt eine lineare Abbildung $f\colon \mathbb{R}^m \to \mathbb{R}^m$ durch $f(x_1, \dots, x_m) := (x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + \cdots + x_m)$. Nach Lemma 6.12 konvergiert die Verteilung von

$$f(Z_1^{(n)}, \dots, Z_m^{(n)}) = \Big(\sum_{j=1}^{[nt_1]} \frac{X_j}{\sqrt{n}}, \sum_{j=1}^{[nt_2]} \frac{X_j}{\sqrt{n}}, \dots, \sum_{j=1}^{[nt_m]} \frac{X_j}{\sqrt{n}}\Big)$$

gegen νf^{-1}. Sei $(U_1, \dots, U_m)$ eine Zufallsgröße mit Verteilung ν, dann besitzt die Normalverteilung νf^{-1} den Erwartungswert 0 und die Kovarianzmatrix mit Komponenten

$$\begin{aligned} E\Big(\sum_{k=1}^{i} U_k \sum_{s=1}^{j} U_s\Big) &= \sum_{k=1}^{\min\{i,j\}} E(U_k^2) + \sum_{\substack{k=1\\k\neq s}}^{i} \sum_{s=1}^{j} E(U_k U_s) \\ &= \sum_{k=1}^{\min\{i,j\}} (t_k - t_{k-1}) = \min\{t_i, t_j\}. \end{aligned}$$

Sei nun $W_j^{(n)} := \sum_{i=1}^{[nt_j]} X_i/\sqrt{n} - Y_n(t_j)$. Offenbar ist $|W_j^{(n)}| \le |X_{[nt_j]+1}|/\sqrt{n}$, falls $t_j < 1$, und sonst $W_j^{(n)} = 0$. Damit ist für $\varepsilon > 0$:

$$\begin{aligned} P(|(W_1^{(n)}, \dots, W_m^{(n)})| \ge \varepsilon) &\le P\Big(\bigcup_{j=1}^{m} \Big\{|W_j^{(n)}| \ge \frac{\varepsilon}{m}\Big\}\Big) \\ &\le \sum_{j=1}^{m} P\Big(|X_{[nt_j]+1}| \ge \sqrt{n}\,\frac{\varepsilon}{m}\Big) = mP\Big(|X_1| \ge \sqrt{n}\,\frac{\varepsilon}{m}\Big) \to 0 \end{aligned}$$

für $n \to \infty$, das heißt, $(W_1^{(n)}, \dots, W_m^{(n)})$ konvergiert in Wahrscheinlichkeit gegen 0. Nach Lemma 6.24 konvergiert dann auch $\mathcal{L}(Y_n(t_1), \dots, Y_n(t_m))$ gegen νf^{-1}. □

Konvergiert, wie behauptet, $\mu_n = \mathcal{L}(Y_n)$ gegen ein Wahrscheinlichkeitsmaß μ auf $(C, \mathcal{B}_C)$, so konvergiert für alle $m \in \mathbb{N}$ und $0 \le t_1 < t_2 < \cdots < t_m \le 1$ auch $\mu_n \pi_{t_1,\dots,t_m}^{-1}$ gegen $\mu \pi_{t_1,\dots,t_m}^{-1}$. Dieses Wahrscheinlichkeitsmaß muß also das in Satz 6.27 angegebene Wahrscheinlichkeitsmaß auf $(\mathbb{R}^m, \mathcal{B}^m)$ sein. Gibt es ein derartiges Maß μ?

Satz 6.29. *Es gibt genau ein Wahrscheinlichkeitsmaß μ auf $(C, \mathcal{B}_C)$ derart, daß für alle $m \in \mathbb{N}$ und alle $0 \leq t_1 < \cdots < t_m \leq 1$ das Maß $\mu\pi_{t_1,\ldots,t_m}^{-1}$ die m-dimensionale Normalverteilung mit Erwartungswert 0 und Kovarianzmatrix $(\min\{t_i, t_j\})_{i,j}$ ist.*

Dieses Maß μ ist das sogenannte *Wiener-Maß* oder die Verteilung der eindimensionalen Brownschen Bewegung. Die Eindeutigkeit von μ folgt unmittelbar aus der Bemerkung 1.9 und der Tatsache, daß die Menge $\{\, \pi_{t_1,\ldots,t_m}^{-1}(A) : m \in \mathbb{N},\ 0 \leq t_1 < t_2 < \cdots < t_m \leq 1,\ A \in \mathcal{B}^m \,\}$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_C$ ist.

Die Existenz des Wiener-Maßes μ wird simultan mit dem folgenden Satz bewiesen werden (Es sei daran erinnert, daß μ_n die Verteilung der durch (6.26) definierten Funktion Y_n bezeichnet.):

Satz 6.30 (SATZ VON DONSKER). *Es gilt $\mu_n \xrightarrow{w} \mu$ auf $(C, \mathcal{B}_C)$.*

Die Existenz in Satz 6.29 und Satz 6.30 folgen sofort aus der folgenden Aussage:

$$\{\, \mu_n : n \in \mathbb{N} \} \quad \text{ist straff.} \tag{6.31}$$

Aus (6.31) folgt nämlich (mit Satz 6.17), daß $(\mu_n)_n$ konvergente Teilfolgen enthält. Jedes Grenzelement μ einer derartigen Teilfolge hat aber nach Satz 6.27 die richtigen endlichdimensionalen Randverteilungen $\mu\pi_{t_1,\ldots,t_m}^{-1}$. Damit ist die Existenz eines Maßes mit den in Satz 6.29 geforderten Eigenschaften bewiesen, wenn man (6.31) hat.

Weiter folgt aus (6.31), daß jede Teilfolge von $(\mu_n)_n$ ihrerseits eine konvergente Teilfolge hat. Deren Grenzelement stimmt wegen Satz 6.27 mit dem Wiener-Maß überein. Aus Lemma 6.18 folgt dann $\mu_n \xrightarrow{w} \mu$. Damit ist Satz 6.30 bewiesen. Wir verschieben den Beweis von (6.31) auf den nächsten Abschnitt.

Aus dem Satz von Donsker und Lemma 6.13 folgt sofort

Satz 6.32. *Ist $h : C[0,1] \to \mathbb{R}$ eine Borel-meßbare Abbildung mit $\mu(D_h) = 0$ und ist $(X_i)_i$ eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit $EX_i = 0$ und $EX_i^2 = 1$, so gilt $\mathcal{L}(h(Y_n)) \xrightarrow{w} \mu h^{-1}$, wobei Y_n die durch (6.26) definierte $(C, \mathcal{B}_C)$-wertige Zufallsgröße sei.*

Dieser Satz hat zwei Aspekte. Einerseits liefert er die asymptotische Verteilung von $h(Y_n)$, sofern 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} herangezogen werden, wenn man die Verteilung der $h(Y_n)$ kennt. Man kann dazu die Verteilung der X_i beliebig wählen, solange $EX_i = 0$ und $E(X_i)^2 = 1$ 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*.

Anders ausgedrückt: Sind $(X_i)_{i\in\mathbb{N}}$ und $(X_i')_{i\in\mathbb{N}}$ zwei Folgen unabhängig identisch verteilter Zufallsgrößen mit $EX_i = EX_i' = 0$ und $EX_i^2 = E(X_i')^2 = 1$, und sind Y_n

und Y_n' die dazugehörigen interpolierten Irrfahrten, so gilt für jede stetige Funktion $h\colon C[0,1] \to \mathbb{R}$ mit $\mu(D_h) = 0$

$$\lim_{n\to\infty} \mathcal{L}(h(Y_n)) = \lim_{n\to\infty} \mathcal{L}(h(Y_n')). \tag{6.33}$$

Aus Satz 6.27 wissen wir schon, daß für $0 < t_1 < \cdots < t_m = 1$ gilt:

$$\lim_{n\to\infty} \mathcal{L}((Y_n(t_1), \ldots, Y_n(t_m))) = \lim_{n\to\infty} \mathcal{L}((Y_n'(t_1), \ldots, Y_n'(t_m))), \tag{6.34}$$

und demzufolge (siehe Lemma 6.12) für jede stetige Funktion $h\colon \mathbb{R}^m \to \mathbb{R}$ gilt:

$$\lim_{n\to\infty} \mathcal{L}(h(Y_n(t_1), \ldots, Y_n(t_m))) = \lim_{n\to\infty} \mathcal{L}(h(Y_n'(t_1), \ldots, Y_n'(t_m))). \tag{6.35}$$

Als Motivation für den Beweis von (6.31) im nächsten Abschnitt diskutieren wir (6.33) und (6.35) in einem Spezialfall, nämlich wo $h\colon C[0,1] \to \mathbb{R}$ die Maximumfunktion ist: $h(f) = \max_{0\le t\le 1} f(t)$, beziehungsweise $h\colon \mathbb{R}^m \to \mathbb{R}$, $h(x_1, \ldots, x_m) = \max_{i=1}^m x_i$. Offensichtlich ist alles, was man für die Herleitung von (6.33) aus (6.35) zu tun hat, eine Limesvertauschung von $m \to \infty$, $n \to \infty$. Genauer: Sei $(t_0^{(m)}, \ldots, t_m^{(m)})_{m\in\mathbb{N}}$ eine Folge von Einteilungen $0 < t_0^{(m)} < t_1^{(m)} < \cdots < t_m^{(m)} = 1$ des Einheitsintervalls, wobei wir der Einfachheit halber annehmen, daß man die $(m+1)$-te Einteilung durch Hinzunahme eines Punktes aus der m-ten gewinnt. Ferner gelte $\lim_{m\to\infty} \max_{1\le i\le m}(t_i^{(m)} - t_{i-1}^{(m)}) = 0$. Offensichtlich gilt für jedes $n \in \mathbb{N}$

$$\lim_{m\to\infty} \max_{1\le i\le m} Y_n(t_i^{(m)}) = \max_{0\le t\le 1} Y_n(t),$$

und demzufolge

$$\mathcal{L}(\max_{0\le t\le 1} Y_n(t)) = \lim_{m\to\infty} \mathcal{L}(\max_{1\le i\le m} Y_n(t_i^{(m)})).$$

Die Aussage (6.33) folgt also aus (6.35), sofern wir

$$\lim_{n\to\infty}\lim_{m\to\infty} \mathcal{L}(\max_i Y_n(t_i^{(m)})) = \lim_{m\to\infty}\lim_{n\to\infty} \mathcal{L}(\max_i Y_n(t_i^{(m)})) \tag{6.36}$$

zeigen (die Existenz der Limiten einmal vorausgesetzt). Es mag hier nützlich sein, sich an das Konvergenzverhalten von Doppelfolgen $(a_{nm})_{n,m\in\mathbb{N}}$ reeller Zahlen zu erinnern. 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: Unter welchen Bedingungen gilt $b = \lim_{m\to\infty} c_m$? Eine hinreichende (aber nicht notwendige) Bedingung ist bekanntlich, daß die Konvergenz von $a_{nm} \to b_n$ für $m \to \infty$ gleichmäßig in n erfolgt, oder, was auf das gleiche hinausläuft, daß $(a_{nm})_m$ eine in n gleichmäßige Cauchyfolge ist, das heißt

$$\sup_n \lim_{m,m'\to\infty} |a_{nm} - a_{nm'}| = 0.$$

Ganz analog diskutiert man nun (6.36). Man zeigt also, daß $\mathcal{L}(\max_{i=0}^m Y_n(t_i^{(m)}))$ für große m, m' nahe bei $\mathcal{L}(\max_{i=0}^{m'} Y_n(t_i^{(m')}))$ liegt, gleichmäßig in n. Ein naheliegender

Ansatz, dies zu diskutieren, besteht darin, die Fluktuationen von $Y_n(t)$, $0 \le t \le 1$, zu untersuchen. Für $f \in C[0,1]$ und $\delta > 0$ sei

$$w_\delta(f) := \sup\{\, |f(s) - f(t)| : s, t \in [0,1],\ |s-t| \le \delta \,\}.$$

Offensichtlich ist

$$|\max_i Y_n(t_i^{(m)}) - \max_i Y_n(t_i^{(m')})| \le w_\delta(Y_n)$$

für $m' \ge m$, falls m so groß ist, daß $\max_i(t_i^{(m)} - t_{i-1}^{(m)}) \le \delta$. Als Übungsaufgabe möge der Leser zeigen, daß, wenn $w_\delta(Y_n)$ in Wahrscheinlichkeit gegen 0 konvergiert für $\delta \to 0$, gleichmäßig in n, das heißt, wenn

$$\lim_{\delta \to \infty} \sup_n P(w_\delta(Y_n) \ge \varepsilon) = 0 \tag{6.37}$$

für alle $\varepsilon > 0$ gilt, (6.36) folgt.

Es ist erstaunlich, daß der im Vergleich zu der obigen Diskussion sehr viel allgemeinere Satz von Donsker sich ebenfalls aus (6.37) ergibt und diese Aussage die wirklich entscheidende für (6.31) ist. Wir diskutieren das zusammen mit dem Beweis von (6.37) im nächsten Abschnitt.

Es ist wichtig zu betonen, daß (6.33) nicht automatisch aus (6.35) folgt, daß also (6.36) nicht für beliebige $C[0,1]$-wertige Zufallsgrößen Y_n gilt. Als Gegenbeispiel dazu betrachten wir exponentiell verteilte, unabhängige Zufallsgrößen $X_1, X_2, \ldots$ und definieren

$$Z_n\Big(\frac{k}{n}\Big) := \frac{X_k}{\log n}, \quad k = 1, \ldots, n,$$

und $Z_n(0) = 0$. Wir definieren $Z_n(t)$ für $t \in [0,1]$ durch lineare Interpolation. Für jedes $t \in [0,1]$ und $\varepsilon > 0$ gilt dann

$$\lim_{n\to\infty} P(Z_n(t) \ge \varepsilon) \le \lim_{n\to\infty} P(\{\, X_{[nt]/n} \ge \varepsilon \log n \,\} \cup \{\, X_{([nt]+1)/n} \ge \varepsilon \log n \,\}) = 0,$$

woraus sofort folgt:

$$\lim_{n\to\infty} \mathcal{L}(\max_i Z_n(t_i^{(m)})) = \delta_0,$$

das heißt (6.35) gilt mit Z_n anstelle von Y_n, $Y_n' \equiv 0$ und $h(x_1, \ldots, x_m) = \max_{i=1}^m x_i$.

Andererseits ist aber für $x > 0$

$$\begin{aligned} P(\max_{t\in[0,1]} Z_n(t) \ge x) &= 1 - P(\max_{t\in[0,1]} Z_n(t) < x) \\ &= 1 - P(\max_{1\le i\le n} X_i < x \log n) \\ &= 1 - (1 - e^{-x \log n})^n = 1 - (1 - n^{-x})^n, \end{aligned}$$

und demzufolge

$$\lim_{n\to\infty} P(\max_t Z_n(t)) = \begin{cases} 1 & \text{für } x \le 1, \\ 0 & \text{für } x > 1, \end{cases}$$

also $\mathcal{L}(\max_t Z_n(t)) = \delta_1$. Somit ist (6.33) (mit Z_n anstelle von Y_n und mit $Y_n' \equiv 0$) falsch, und demzufolge auch (6.36) beziehungsweise (6.37) für Z_n.

6e. Beweis der Straffheit von (μ_n)

Die Grundlage des Beweises von (6.31) (und aller Beweise der Straffheit von Wahrscheinlichkeitsmaßen auf $C[0,1]$) bildet die folgende Charakterisierung relativkompakter Teilmengen in $C[0,1]$:

Für $f \in C[0,1]$ und $\delta < 0$ sei wie oben

$$w_\delta(f) := \sup\{\, |f(s) - f(t)| : s, t \in [0,1],\ |s-t| \le \delta \,\}.$$

Satz 6.38 (SATZ VON ARZELÀ-ASCOLI). *Eine Teilmenge $A \subset C[0,1]$ hat genau dann kompakten Abschluß, wenn*

(a) $\sup\{\, |f(0)| : f \in A \,\} < \infty$ *ist und*
(b) $\lim_{\delta \downarrow 0} \sup_{f \in A} w_\delta(f) = 0$ *gelten.*

Der Satz wird als bekannt vorausgesetzt. Er kann in ein Kriterium für die Straffheit einer Folge von Wahrscheinlichkeitsmaßen auf C übersetzt werden:

Satz 6.39. *Eine Folge $(\nu_n)_n$ von Wahrscheinlichkeitsmaßen auf $(C, \mathcal{B}_C)$ ist genau dann straff, wenn*

(a) $\lim_{a \uparrow \infty} \sup_n \nu_n(\{\, f : |f(0)| \ge a \,\}) = 0$ *und*
(b) $\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \nu_n(\{\, f : w_\delta(f) \ge \varepsilon \,\}) = 0$ *für alle $\varepsilon > 0$ gelten.*

Bemerkung 6.40.

(a) Für ein festes $\delta > 0$ ist die Abbildung $f \mapsto w_\delta(f)$ stetig auf $C[0,1]$. Somit ist $\{\, f : w_\delta(f) \ge \varepsilon \,\} \in \mathcal{B}_C$.
(b) Die Bedingungen (a) und (b) aus 6.39 können wie folgt übersetzt werden:

$$\forall \eta > 0 \ \ \exists a > 0 \ \forall n \in \mathbb{N} : \ \ \nu_n(\{\, f : |f(0)| \ge a \,\}) \le \eta, \tag{6.41}$$

$$\forall \varepsilon, \eta > 0 \ \exists \delta > 0 \ \exists n_0 \in \mathbb{N} \ \forall n \ge n_0 : \ \ \nu_n(\{\, f : w_\delta(f) \ge \varepsilon \,\}) \le \eta. \tag{6.42}$$

(c) Da $C[0,1]$ vollständig und separabel ist, ist jedes einzelne Wahrscheinlichkeitsmaß ν straff. Für jedes $\eta > 0$ existiert also eine kompakte Menge K mit $\nu(K) \ge 1 - \eta$. Insbesondere folgt, daß für $\varepsilon > 0$ ein $\delta > 0$ existiert mit $\nu(\{\, f : w_\delta(f) \ge \varepsilon \,\}) \le \eta$. Damit folgt sofort, daß (6.42) äquivalent ist zu

$$\forall \varepsilon, \eta > 0 \ \exists \delta > 0 \ \forall n \in \mathbb{N} : \ \ \nu_n(\{\, f : w_\delta(f) \ge \varepsilon \,\}) \le \eta. \tag{6.43}$$

Beweis von Satz 6.39. Sei $\{\, \nu_n : n \in \mathbb{N} \}$ straff. Für $\eta > 0$ sei K eine kompakte Menge mit $\nu_n(K) \ge 1 - \eta$ für alle n. Daraus folgen mit dem Satz von Arzelà-Asoli (Satz 6.38) die Aussagen (6.41) und (6.43).

Um die Umkehrung zu beweisen, sei $(\nu_n)_n$ eine Folge, die (6.41) und (6.43) erfüllt.

Sei $\eta > 0$ vorgegeben. Nach (6.41) existiert ein $a \in \mathbb{R}$, so daß die Menge $A := \{ f : |f(0)| \le a \}$ erfüllt: $\nu_n(A) \ge 1 - \eta/2$ für alle n. Für $k \in \mathbb{N}$ sei δ_k so gewählt, daß $\nu_n(\{ f : w_{\delta_k}(f) < 1/k \}) \ge 1 - \eta/2^{k+1}$ für alle n gilt. Nach Satz 6.38 hat $K := A \cap \bigcap_{k=1}^{\infty} \{ f : w_{\delta_k}(f) < 1/k \}$ kompakten Abschluß, und es gilt

$$\nu_n(\bar{K}^c) \le \nu_n(K^c) \le \frac{\eta}{2} + \sum_{k=1}^{\infty} \frac{\eta}{2^{k+1}} = \eta \quad \text{für alle } n \in \mathbb{N}.$$

□

Bemerkung 6.44. Hinreichend für die Aussage (6.41) ist natürlich die Bedingung $\nu_n(\{ f : f(0) = 0 \}) = 1$, was bei unseren μ_n erfüllt ist.

Lemma 6.45. *Die folgende Bedingung ist hinreichend für (6.42):*

$$(6.46) \qquad \begin{cases} \forall \varepsilon > 0 \ \forall \eta > 0 \ \exists \delta \in (0,1) \ \exists n_0 \in \mathbb{N} \quad \forall n \ge n_0 \ \forall t \in [0, 1-\delta] : \\ \frac{1}{\delta} \nu_n(\{ f : \sup_{t \le s \le t+\delta} |f(s) - f(t)| \ge \varepsilon \}) \le \eta. \end{cases}$$

Beweis. Es gelte (6.46), und es soll (6.42) nachgewiesen werden. Seien $\varepsilon, \eta > 0$. Dann wählen wir $\delta_0 \in (0,1)$ und $n_0 \in \mathbb{N}$ nach (6.46) zu $\varepsilon/2$, $\eta/3 > 0$ (anstelle von ε, η). Sei m die kleinste natürliche Zahl mit $1/m < \delta_0$. Wir setzen $\delta := 1/(2m)$.

Ist $f \in C[0,1]$ mit $w_\delta(f) \ge \varepsilon$, so existieren Punkte $t < s$ mit $|f(t) - f(s)| \ge \varepsilon$ und $|t - s| \le \delta$. Zu t, s existiert $k \in \mathbb{N}_0$ mit $k \le 2m-2$ und $k/(2m) \le t < s \le k/(2m) + 2\delta = k/(2m) + 1/m$. Dann ist $|f(t) - f(k/(2m))| \ge \varepsilon/2$ oder $|f(s) - f(k/(2m))| \ge \varepsilon/2$. Somit ist

$$\{ f : w_\delta(f) \ge \varepsilon \} \subset \bigcup_{k=0}^{2m-2} \Big\{ f : \sup_{k/(2m) \le s \le k/(2m) + \delta_0} \Big| f(s) - f\Big(\frac{k}{2m}\Big) \Big| \ge \frac{\varepsilon}{2} \Big\},$$

und somit gilt für $n \ge n_0$:

$$\begin{aligned} \nu_n(\{ f : w_\delta(f) \ge \varepsilon \}) &\le \sum_{k=0}^{2m-2} \nu_n\Big(\Big\{ f : \sup_{k/(2m) \le s \le k/(2m) + \delta_0} \Big| f(s) - f\Big(\frac{k}{2m}\Big) \Big| \ge \frac{\varepsilon}{2} \Big\} \Big) \\ &\le (2m-1)\, \delta_0 \, \frac{\eta}{3} \le (2 + \delta_0) \, \frac{\eta}{3} \le \eta. \end{aligned}$$

Damit ist (6.42) gezeigt. □

Bemerkung 6.47.

(a) Die Bedingung (6.46) folgt aus der Aussage, daß für alle $\varepsilon > 0$ gilt:

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{t \in [0, 1-\delta]} \frac{1}{\delta} \nu_n(\{ f : \sup_{t \le s \le t+\delta} |f(t) - f(s)| \ge \varepsilon \}) = 0.$$

(b) Die Bedingung in Lemma 6.45 ist hinreichend, aber nicht notwendig für die Straffheit. Der Leser möge als Übungsaufgabe dafür ein Beispiel angeben. Das Vorgehen, (6.42) durch (6.46) aus Lemma 6.45 zu ersetzen, erscheint auf den ersten Blick schon grob. Tatsächlich geht jedoch für die Diskussion von Y_n beziehungsweise $\mu_n = PY_n^{-1}$ dabei nichts Wesentliches verloren. Für eine etwas ausführliche Diskussion dieses Punktes siehe das Buch von Billingsley: 'Convergence of Probability Measures', S. 57.

Die Bedingung in Bemerkung 6.47(a) soll nun für $\mu_n = PY_n^{-1}$ untersucht werden. Natürlich ist für $\delta \in (0,1)$ und $t \in [0, 1-\delta]$

$$\mu_n(\{ f\colon \sup_{t\le s\le t+\delta} |f(s) - f(t)| \ge \varepsilon \}) = P(\sup_{t\le s\le t+\delta} |Y_n(s) - Y_n(t)| \ge \varepsilon).$$

Ist speziell $t = k/n$ und $t + \delta = j/n$ (mit $k < j$), so ist:

$$\sup_{t\le s\le t+\delta} |Y_n(s) - Y_n(t)| = \max_{1\le i\le n\delta} \frac{|S_{k+i} - S_k|}{\sqrt{n}}.$$

Für allgemeine $t \in [0,1]$ und $\delta \in (0,1)$ mit $t + \delta \le 1$ kann die linke Seite der obigen Gleichung wie folgt abgeschätzt werden: Es existieren $j, k \in \{0, 1, \ldots, n\}$ mit $k < j$ und $k/n \le t < (k+1)/n$ sowie $(j-1)/n < t+\delta \le j/n$. Dann gilt für jedes $s \in [t, t+\delta]$:

$$\begin{aligned}|Y_n(s) - Y_n(t)| &\le \Big|Y_n(t) - Y_n\Big(\frac{k}{n}\Big)\Big| + \max_{1\le i\le j-k}\Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big| \\ &\le 2\max_{1\le i\le j-k}\Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big|.\end{aligned}$$

Demzufolge ist

$$\sup_{t\le s\le t+\delta} |Y_n(s) - Y_n(t)| \le 2\max_{1\le i\le j-k}\Big|Y_n\Big(\frac{k+i}{n}\Big) - Y_n\Big(\frac{k}{n}\Big)\Big| = 2\max_{1\le i\le j-k}\Big|\sum_{r=k+1}^{k+i} X_r\Big|/\sqrt{n}.$$

Nun ist $\frac{j-k-2}{n} \le \delta$ und für $n \ge \frac{1}{\delta}$ also $j - k \le 3n\delta$. Somit ist für $n \ge \frac{1}{\delta}$ die rechte Seite der obigen Ungleichung nicht größer als $2\max_{1\le i\le 3n\delta}\Big|\sum_{r=k+1}^{k+i} X_r\Big|/\sqrt{n}$. Die Verteilung dieser Zufallsgröße hängt aber nicht von k ab. Somit ergibt sich für $n \ge 1/\delta$:

$$\sup_{t\in[0,1-\delta]} P(\sup_{t\le s\le t+\delta} |Y_n(s) - Y_n(t)| \ge \varepsilon) \le P\Big(\max_{1\le i\le 3n\delta} \frac{|S_i|}{\sqrt{n}} \ge \frac{\varepsilon}{2}\Big).$$

Ersetzt man den ganzzahligen Anteil von $3n\delta$ durch m, so ist $\sqrt{n} \ge \sqrt{m/(3\delta)}$. Damit ist

$$P\Big(\max_{1\le i\le 3n\delta} \frac{|S_i|}{\sqrt{n}} \ge \frac{\varepsilon}{2}\Big) \le P\Big(\max_{1\le i\le m} \frac{|S_i|}{\sqrt{m}} \ge \frac{\varepsilon}{\sqrt{12\,\delta}}\Big).$$

Für jedes feste $\delta > 0$ geht $m \to \infty$ für $n \to \infty$.

Um Bemerkung 6.47(a) nachzuweisen, genügt es also für jedes $\varepsilon > 0$ zu zeigen, daß

$$\lim_{\delta\downarrow 0}\limsup_{m\to\infty}\frac{1}{\delta}P\Big(\max_{1\le i\le m}\frac{|S_i|}{\sqrt{m}}\ge\frac{\varepsilon}{\sqrt{\delta}}\Big)=0 \tag{6.48}$$

gilt. Man kann nun versucht sein, $P(\max_{1\le i\le m}|S_i|/\sqrt{m}\ge\varepsilon/\sqrt{\delta})$ abzuschätzen durch $\sum_{i=1}^m P(|S_i|/\sqrt{m}\ge\varepsilon/\sqrt{\delta})$ analog zum Vorgehen in Lemma 6.45 mit der Ersetzung von (6.42) durch (6.46), gemäß der Devise: Was einmal gut geht, geht auch ein zweites Mal. Auf diesem Weg kann (6.48) jedoch nicht bewiesen werden. Tatsächlich muß man $P(\max_{1\le i\le m}|S_i|/\sqrt{m}\ge\lambda)$ für $\lambda>0$ wesentlich genauer abschätzen:

Lemma 6.49. *Für alle $\lambda > 0$ und $m\in\mathbb{N}$ gilt*

$$P(\max_{1\le i\le m}|S_i|\ge\lambda\sqrt{m})\le 2P(|S_m|\ge(\lambda-\sqrt{2})\sqrt{m}).$$

Beweis. Die Aussage ist trivial für $\lambda\le\sqrt{2}$. Sei also $\lambda>\sqrt{2}$. Für $1\le i\le m$ sei

$$A_i=\bigcap_{j=1}^{i-1}\{\,|S_j|<\lambda\sqrt{m}\,\}\cap\{\,|S_i|\ge\lambda\sqrt{m}\,\}.$$

Die A_i sind paarweise disjunkt, und es gilt

$$A:=\{\max_{1\le i\le m}|S_i|\ge\lambda\sqrt{m}\,\}=\bigcup_{i=1}^m A_i.$$

Wir haben

$$\begin{aligned}P(A)&=P(A\cap\{\,|S_m|\ge(\lambda-\sqrt{2})\sqrt{m}\,\})\\&\qquad+P(A\cap\{\,|S_m|<(\lambda-\sqrt{2})\sqrt{m}\,\})\\&\le P(|S_m|\ge(\lambda-\sqrt{2})\sqrt{m})\\&\qquad+\sum_{j=1}^{m-1}P(A_j\cap\{\,|S_m|<(\lambda-\sqrt{2})\sqrt{m}\,\})\end{aligned}\tag{6.50}$$

wegen $A_m\cap\{\,|S_m|<(\lambda-\sqrt{2})\sqrt{m}\,\}=\emptyset$. Für $j=1,\dots,m-1$ gilt

$$A_j\cap\{\,|S_m|<(\lambda-\sqrt{2})\sqrt{m}\,\}\subset A_j\cap\{\,|S_m-S_j|\ge\sqrt{2}\sqrt{m}\,\}.$$

Da die Ereignisse A_j und $\{\,|S_m-S_j|\ge\sqrt{2}\sqrt{m}\,\}$ unabhängig sind, ist die rechte Seite von (6.50) nicht größer als

$$P(|S_m|\ge(\lambda-\sqrt{2})\sqrt{m})+\sum_{j=1}^{m-1}P(A_j)P(|S_m-S_j|\ge\sqrt{2m}).\tag{6.51}$$

Wegen

$$P(|S_m - S_j| \geq \sqrt{2m}) \leq \frac{1}{2m} E\Big(\Big(\sum_{k=j+1}^{m} X_k\Big)^2\Big) = \frac{1}{2m} \sum_{k=j+1}^{m} E(X_k^2) \leq \frac{1}{2}$$

kann (6.51) weiter nach oben abgeschätzt werden durch

$$P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}) + \frac{1}{2}\sum_{j=1}^{m} P(A_j) = P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m}) + \frac{1}{2}P(A).$$

Demzufolge ist $P(A) \leq 2P(|S_m| \geq (\lambda - \sqrt{2})\sqrt{m})$. $\square$

Beweis von (6.48). Nach Lemma 6.49 und nach dem zentralen Grenzwertsatz ist (es sei Φ die Verteilungsfunktion der Standardnormalverteilung)

$$\begin{aligned}
\limsup_{m\to\infty} \frac{1}{\delta}\Big(\max_{1\leq i\leq m} \frac{|S_i|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}}\Big) &\leq \limsup_{m\to\infty} \frac{2}{\delta} P\Big(\frac{|S_m|}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big) \\
&\leq \limsup_{m\to\infty} \frac{2}{\delta}\Big\{P\Big(\frac{S_m}{\sqrt{m}} \geq \frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big) \\
&\qquad + P\Big(\frac{S_m}{\sqrt{m}} \leq -\frac{\varepsilon}{\sqrt{\delta}} + \sqrt{2}\Big)\Big\} \\
&= \frac{2}{\delta}\Big(\Big(1 - \Phi\Big(\frac{\varepsilon}{\sqrt{\delta}} - \sqrt{2}\Big)\Big) + \Phi\Big(-\frac{\varepsilon}{\sqrt{\delta}} + \sqrt{2}\Big)\Big) \\
&= \frac{4}{\delta}\Phi\Big(-\frac{\varepsilon}{\sqrt{\delta}} + \sqrt{2}\Big).
\end{aligned}$$

Eine einfache Anwendung der Markoff-Ungleichung (Satz 2.39) ergibt, daß für $\varepsilon > 0$

$$\lim_{\delta\downarrow 0} \frac{1}{\delta}\Phi\Big(-\frac{\varepsilon}{\sqrt{\delta}} + \sqrt{2}\Big) = 0$$

gilt. Damit ist (6.48) bewiesen und somit die Straffheit der Folge $(\mu_n)_n$. $\square$

6f. Anwendungen

Der Satz 6.39 soll auf zwei Beispiele angewendet werden. Zunächst betrachten wir $h\colon C[0,1] \to \mathbb{R}$, $h(f) := \sup_{0\leq t\leq 1} f(t)$. Es ist leicht zu sehen, daß h stetig ist. Die X_i seien zunächst unabhängige Zufallsgrößen mit $P(X_i = 1) = P(X_i = -1) = 1/2$. Seien $S_0 := 0$, $S_n := \sum_{i=1}^{n} X_i$ und $M_n := \max_{1\leq i\leq n} S_i$ für $n \in \mathbb{N}$. Die Verteilung von M_n kann leicht mit dem sogenannten Spiegelungsprinzip gewonnen werden (siehe „Einführung in die Stochastik“, Lemma 4.1). Für $l, k \in \mathbb{N}_0$ gilt

$$P(S_n = l + k) = P(S_n = l + k,\ M_n \geq k) = P(M_n \geq k,\ S_n = k - l).$$

Damit ist

$$\begin{aligned}P(M_n \geq k) &= \sum_{l=-\infty}^{\infty} P(M_n \geq k,\ S_n = l+k) \\ &= \sum_{l=-\infty}^{-1} P(M_n \geq k,\ S_n = k+l) + \sum_{l=1}^{\infty} P(S_n = k+l) + P(S_n = k) \\ &= 2P(S_n > k) + P(S_n = k).\end{aligned}$$

Wir überlegen uns kurz, daß $P(S_n = k)$ für $n \to \infty$ gegen 0 gleichmäßig in k konvergiert: Zunächst gilt $P(S_n = k) \leq P(S_n = 0)$, wenn n gerade ist und $P(S_n = k) \leq P(S_n = 1)$, wenn n ungerade ist. Nach dem lokalen zentralen Grenzwertsatz ist für gerades n aber $P(S_n = 0) \sim 1/\sqrt{2\pi n}$, und für ungerades n ist $P(S_n = 1) \sim 1/\sqrt{2\pi n}$ für $n \to \infty$.

Seien nun $t \geq 0$ und $\varepsilon > 0$. Für genügend großes n existieren Zahlen $k_1 = k_1(n)$, und $k_2 = k_2(n) \in \mathbb{N}_0$ mit $t - \varepsilon < k_1/\sqrt{n} \leq t \leq k_2/\sqrt{n} \leq t + \varepsilon$. Damit gilt nach dem zentralen Grenzwertsatz:

$$\begin{aligned}\varliminf_{n\to\infty} P\Big(\frac{M_n}{\sqrt{n}} > t\Big) &\geq \varliminf_{n\to\infty} P\Big(\frac{M_n}{\sqrt{n}} \geq \frac{k_2(n)}{\sqrt{n}}\Big) \\ &= 2 \varliminf_{n\to\infty} P\Big(\frac{S_n}{\sqrt{n}} \geq \frac{k_2(n)}{\sqrt{n}}\Big) \geq 2 \lim_{n\to\infty} P\Big(\frac{S_n}{\sqrt{n}} > t+\varepsilon\Big) \\ &= 2(1 - \Phi(t+\varepsilon)).\end{aligned}$$

(Hierbei ist wie bisher Φ die Verteilungsfunktion der Standardnormalverteilung.)

Analog (mit $k_1(n)$) zeigt man

$$\varlimsup_{n\to\infty} P\Big(\frac{M_n}{\sqrt{n}} > t\Big) \leq 2(1 - \Phi(t-\varepsilon)).$$

Da $\varepsilon > 0$ beliebig und Φ stetig ist, folgt

$$\lim_{n\to\infty} P\Big(\frac{M_n}{\sqrt{n}} \leq t\Big) = 2\Phi(t) - 1.$$

Für $t < 0$ gilt natürlich $P(M_n/\sqrt{n} \leq t) = 0$.

Aus Satz 6.39 folgt für das Wiener-Maß μ und obiges h:

Satz 6.52. *Das Maß μh^{-1} hat die Verteilungsfunktion $t \mapsto \max\{2\Phi(t) - 1, 0\}$.* □

Da $\sup_{0\leq t\leq 1} Y_n(t) = \max_{0\leq i\leq n} S_i/\sqrt{n}$ ist, so folgt nun aus Satz 6.39

Satz 6.53. *Erfüllen die X_i die Voraussetzungen von 6.39, so gilt für alle $t \in \mathbb{R}$*

$$\lim_{n\to\infty} P\Big(\max_{0\leq i\leq n} \frac{S_i}{\sqrt{n}} \leq t\Big) = \max\{2\Phi(t) - 1, 0\}.$$

□

Wir betrachten nun eine andere Abbildung $C[0,1] \to \mathbb{R}$, die mit g bezeichnet wird. Für $f \in C$ sei $g(f) = \lambda(\{\, t \in [0,1] \colon f(t) \geq 0 \,\})$, wobei λ das Lebesgue-Maß sei. Die Funktion g ist nicht überall stetig; zum Beispiel ist g unstetig in $f \equiv 0$. Es gilt jedoch (zur Definition von D_g siehe Lemma 6.13):

Lemma 6.54.

(a) *g ist $\mathcal{B}_C$-$\mathcal{B}$-meßbar.*
(b) $\mu(D_g) = 0$.

Beweis. Wir definieren die Abbildung $\psi \colon C[0,1] \times [0,1] \to \mathbb{R}$ durch $\psi(f,t) = f(t)$. Leicht ist zu sehen, daß ψ stetig ist, also $\mathcal{B}_{C\times[0,1]}$-$\mathcal{B}$-meßbar. Da C und $[0,1]$ separabel sind, so folgt völlig analog wie in Lemma 6.21, daß $\mathcal{B}_{C\times[0,1]} = \mathcal{B}_C \otimes \mathcal{B}_{[0,1]}$ gilt. Damit ist ψ $\mathcal{B}_C \otimes \mathcal{B}_{[0,1]}$-$\mathcal{B}$-meßbar.

Beweis von 6.54(a). Sei $A = \{\, (f,t) \colon f(t) \geq 0 \,\} = \psi^{-1}([0,\infty)) \in \mathcal{B}_C \otimes \mathcal{B}_{[0,1]}$. Für $f \in C$ ist $g(f) = \lambda(\{\, t \colon (f,t) \in A \,\})$. Nach Satz 1.53 ist dann die Abbildung $f \mapsto g(f)$ $\mathcal{B}_C$-$\mathcal{B}$-meßbar.

Beweis von 6.54(b). Man bemerkt zunächst, daß $g(f) = \int_0^1 1_{[0,\infty]}(f(t))\,dt$ ist. Ist $f \in C$ mit $\lambda(\{t \colon f(t) = 0\}) = 0$, und ist $(f_n)_n$ eine Folge in C mit $d(f_n, f) \to 0$, so gilt $1_{[0,\infty)}(f_n(t)) \to 1_{[0,\infty)}(f(t))$ für λ-fast alle $t \in [0,1]$. Nach dem beschränkten Konvergenzsatz folgt dann $g(f_n) \to g(f)$. Es ist somit gezeigt, daß die Inklusion $D_g \subset \{\, f \colon \lambda(\{\, t \colon f(t) = 0 \,\}) > 0 \,\}$ gilt.

Es soll nun gezeigt werden, daß $\mu(\{\, f \colon \lambda(\{\, t \colon f(t) = 0 \,\}) > 0 \,\}) = 0$ gilt. (Analog wie oben zeigt man, daß $f \mapsto \lambda(\{\, t \colon f(t) = 0 \,\})$ meßbar ist.) Die obige Aussage ist äquivalent dazu, daß

$$0 = \int_C \lambda\big(\{\, t \colon f(t) = 0 \,\}\big)\mu(df) = \int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f,t)\,dt\,\mu(df)$$

ist. Nach dem Satz von Fubini ist jedoch

$$\begin{aligned} \int_C \int_{[0,1]} (1_{\{0\}} \circ \psi)(f,t)\,dt\,\mu(df) &= \int_{[0,1]} \int_C (1_{\{0\}} \circ \psi)(f,t)\,\mu(df)\,dt \\ &= \int_{[0,1]} \mu(\{\, f \colon f(t) = 0 \,\})\,dt = \int_{[0,1]} \mu\pi_t^{-1}(\{0\})\,dt, \end{aligned}$$

was in der Tat gleich null ist, da $\mu\pi_t^{-1}$ für $t > 0$ die Normalverteilung mit Erwartungswert 0 und Varianz t ist. Damit ist Lemma 6.54 bewiesen. $\square$

Es ist somit gezeigt, daß g die Voraussetzungen von Satz 6.32 erfüllt. Die Berechnung von $\mathcal{L}(g(Y_n))$ im Spezialfall, wo $P(X_i = \pm 1) = 1/2$ ist, ist elementar und wurde in „Einführung in die Stochastik“ durchgeführt. Es gilt das sogenannte Arcussinus-Gesetz: Für $t \in [0,1]$ gilt $\lim_{n\to\infty} P(g(Y_n) \leq t) = 2/\pi \cdot \arcsin \sqrt{t}$. Dies ist somit die Verteilungsfunktion von μg^{-1}. Es ergeben sich wiederum zwei Sätze:

Satz 6.55 (ARCUSSINUS-GESETZ). *Die auf* $(C, \mathcal{B}_C, \mu)$ *definierte Zufallsgröße* $f \mapsto \lambda(\{\, t\colon f(t) \geq 0 \,\})$ *hat die Verteilungsfunktion* $t \mapsto 2/\pi \cdot \arcsin \sqrt{t}$ *(für* $t \in [0,1]$*).*

Satz 6.56. *Erfüllen die* X_i *die Voraussetzungen von Satz 6.39, so gilt für* $t \in [0,1]$

$$\lim_{n\to\infty} P(g(Y_n) \leq t) = \frac{2}{\pi} \cdot \arcsin \sqrt{t}.$$

Meist wird dieser Satz für etwas andere Zufallsgrößen formuliert: Es ist nicht sehr schwer zu zeigen, daß $g(Y_n) - \frac{1}{n}|\{\, m \leq n\colon S_m > 0 \,\}|$ in Wahrscheinlichkeit gegen 0 konvergiert. Das soll hier nicht bewiesen werden. Nach Lemma 6.24 folgt dann, daß auch $\mathcal{L}\left(\frac{1}{n}|\{\, m \leq n\colon S_m > 0 \,\}|\right)$ asymptotisch nach der Arcussinus-Verteilung verteilt ist.

Anhang

7a. Beweis des Satzes von Ionescu-Tulcea (Satz 2.27)

Wir verwenden die Notationen aus Kapitel 2 und setzen zur Abkürzung $\Omega := E^{\mathbb{N}}$. Für $n \in \mathbb{N}$ sei $\mathcal{F}_n := \pi^{(n)-1}(\mathcal{E}^n)$. Die $\mathcal{F}_n$ sind Teil-σ-Algebren von $\mathcal{E}^{\mathbb{N}}$ mit

$$\mathcal{E}^{\mathbb{N}} = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$$
$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \ldots$$

Da $\pi^{(n)}$ surjektiv ist, ist die Zuordnung $\mathcal{E}^n \ni A \mapsto \pi^{(n)-1}(A) \in \mathcal{F}_n$ bijektiv. Demzufolge entsprechen sich Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{F}_n)$ und auf $(E^n, \mathcal{E}^n)$ eineindeutig über die Zuordnung: $\mu \in \mathbb{P}(E^n, \mathcal{E}^n) \mapsto \bar{\mu} \in \mathbb{P}(\Omega, \mathcal{F}_n)$ mit $\bar{\mu}(\pi^{(n)-1}(A)) := \mu(A)$. Die Wahrscheinlichkeitsmaße $\bar{Q}^{(n)}$ auf $(\Omega, \mathcal{F}_n)$ seien auf diese Weise aus den Wahrscheinlichkeitsmaßen $Q^{(n)}$ von Kapitel 2 konstruiert. Die Verträglichkeit der $Q^{(n)}$ ist gleichbedeutend damit, daß für $n < m$

$$\bar{Q}^{(m)}|_{\mathcal{F}_n} = \bar{Q}^{(n)}$$

gilt. $\bar{Q}^{(m)}|_{\mathcal{F}_n}$ ist dabei die Einschränkung von $\bar{Q}^{(m)}$ auf $\mathcal{F}_n$. Demzufolge definieren die $\bar{Q}^{(n)}$ eindeutig eine Abbildung $\bar{Q}\colon \bigcup_{n\in\mathbb{N}} \mathcal{F}_n \to [0,1]$, deren Einschränkungen auf die $\mathcal{F}_n$ gleich $\bar{Q}^{(n)}$ sind. $\mathcal{A} := \bigcup_n \mathcal{F}_n$ ist offenbar eine Algebra und $\bar{Q}$ ein Inhalt, der $\bar{Q}(\Omega) = 1$ erfüllt. Wir zeigen, daß $\bar{Q}$ auf $\mathcal{A}$ σ-additiv ist. Mit dem Satz von Carathéodory läßt sich dann $\bar{Q}$ eindeutig zu einem Wahrscheinlichkeitsmaß Q auf $\sigma(\mathcal{A}) = \mathcal{E}^{\mathbb{N}}$ erweitern, und natürlich gilt dann $Q|_{\mathcal{F}_n} = \bar{Q}^{(n)}$, also $Q\pi^{(n)-1} = Q^{(n)}$. Wegen $\bar{Q}(\Omega) < \infty$ ist die σ-Additivität äquivalent zu der Bedingung

$$(7.1) \qquad A_n \in \mathcal{A}, \quad n \in \mathbb{N}, \quad A_n \downarrow \emptyset \Rightarrow \bar{Q}(A_n) \downarrow 0.$$

Wir definieren Kerne $\overline{K}_{n,m}$ von $(\Omega, \mathcal{F}_n)$ nach $(\Omega, \mathcal{F}_m)$ für $n < m$ folgendermaßen: Für $x \in E^n$ ist $K_n(x, \cdot)$ ein Wahrscheinlichkeitsmaß auf $(E, \mathcal{E})$ und demzufolge $\delta_x \otimes K_n(x, \cdot)$ ein Wahrscheinlichkeitsmaß auf $(E^{n+1}, \mathcal{E}^{n+1})$, das wir wie oben ausgeführt zu einem Wahrscheinlichkeitsmaß $\overline{\delta_x \otimes K_n(x, \cdot)}$ auf $(\Omega, \mathcal{F}_{n+1})$ hochheben können. Wir definieren den Kern $\overline{K}_{n,n+1}$ von $(\Omega, \mathcal{F}_n)$ nach $(\Omega, \mathcal{F}_{n+1})$ durch

$$\overline{K}_{n,n+1}(x, \cdot) = \overline{\delta_{\pi^{(n)}(x)} \otimes K_n(\pi^{(n)}(x), \cdot)} \quad \text{für} x \in \Omega \text{ und } n \in \mathbb{N}$$

und rekursiv:

$$\overline{K}_{n,m+1}(x,A) = \int \overline{K}_{n,m}(x,dy)\overline{K}_{m,m+1}(y,A). \tag{7.2}$$

Man beachte, daß für $x \in \Omega$ und $n < m$ gilt:

$$K_{n,m}(x, \Omega_x^{(n)}) = 1, \tag{7.3}$$

wobei

$$\Omega_x^{(n)} = \{\, y \colon \pi^{(n)}(x) = \pi^{(n)}(y) \,\}.$$

Der Leser möge sich von den folgenden Sachverhalten selbst überzeugen: Für $n < l < m$ gilt

$$\overline{K}_{n,m}(x,\cdot)|_{\mathcal{F}_l} = \overline{K}_{n,l}(x,\cdot) \tag{7.4}$$

und

$$\bar{Q}_m = \bar{Q}_n \overline{K}_{n,m} \qquad \text{für } n < m. \tag{7.5}$$

Wegen (7.4) läßt sich für $x \in \Omega$, $n \in \mathbb{N}$, eindeutig ein Inhalt $\widehat{K}_n(x,\cdot)$ auf $\mathcal{A}$ definieren mit $\widehat{K}_n(x,\cdot)|_{\mathcal{F}_m} = \overline{K}_{n,m}(x,\cdot)$, und es gelten wegen (7.5) für $A \in \mathcal{A}$

$$\bar{Q}(A) = \int_\Omega \bar{Q}_n(dx)\widehat{K}_n(x,A) \tag{7.6}$$

und

$$\begin{aligned} \widehat{K}_n(x,A) &= \int_\Omega K_{n,n+1}(x,dy)\widehat{K}_{n+1}(y,A) \\ &= \int_{\Omega_x^{(n)}} K_{n,n+1}(x,dy)\widehat{K}_{n+1}(y,A). \end{aligned} \tag{7.7}$$

Die erste Gleichung in (7.7) ergibt sich aus (7.2) und die zweite wegen (7.3). Wir zeigen nun (7.1). Seien $A_n \in \mathcal{A}$, $A_n \downarrow$, $\lim_{n\to\infty} \bar{Q}(A_n) > 0$. Wir zeigen, daß $x \in \Omega$ existiert mit $x \in A_n$ für alle n. Mit Induktion nach m konstruieren wir eine Folge (x_m) mit

$$\lim_{n\to\infty} \widehat{K}_m(x_m, A_n) > 0 \tag{7.8}$$

und

$$x_m \in \Omega_{x_{m-1}}^{(m-1)}, \qquad m \geq 2. \tag{7.9}$$

$m = 1$: Wir wenden (7.6) für $n = 1$ an. $\widehat{K}_1(\cdot, A_n)$ ist eine monoton fallende Folge $\mathcal{F}_1$-meßbarer, positiver, beschränkter Funktionen. Aus $\lim_{n\to\infty} \bar{Q}(A_n) > 0$ folgt, daß der Limes dieser Folge nicht identisch verschwindet. Demzufolge existiert $x_1 \in \Omega$ mit

$$\lim_{n\to\infty} \widehat{K}_1(x_1, A_n) > 0.$$

Induktionsschluß $m \Rightarrow m+1$: Wir wenden (7.7) an. Aus $\lim_{n\to\infty} \widehat{K}_m(x_m, A_n) > 0$ folgt ganz gleich wie oben, daß $x_{m+1} \in \Omega_{x_m}^{(m)}$ existiert mit $\lim_{n\to\infty} \widehat{K}_{m+1}(x_{m+1}, A_n) > 0$.

Damit ist eine Folge $(x_m)_m$ konstruiert worden, die (7.8) und (7.9) erfüllt. (7.9) impliziert, daß $x \in \Omega$ existiert mit $x \in \Omega_{x_m}^{(m)}$ für alle m. (7.8) impliziert, daß $\widehat{K}_m(x_m, A_n) > 0$ für alle n, m gilt. Ist $A_n \in \mathcal{F}_k$ und $m > k$, so ist $\widehat{K}_m(x_m, A_n) = 1_{A_n}(x_m)$, also $x_m \in A_n$, und wegen $A_n = \pi^{(k)-1}(B)$ für ein $B \in \mathcal{E}^k$ folgt sofort $x \in A_n$. Damit ist gezeigt, daß x in allen A_n liegt. □

7b. Existenz regulärer, bedingter Wahrscheinlichkeiten

Satz 7.10. *Es sei (Ω, d) ein vollständiger, separabler, metrischer Raum und $\mathcal{B}_\Omega$ die Borel-σ-Algebra. P sei ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B}_\Omega)$ und $\mathcal{F}$ eine beliebige Teil-σ-Algebra von $\mathcal{B}_\Omega$. Dann existiert ein Markoff-Kern K von $(\Omega, \mathcal{F})$ nach $(\Omega, \mathcal{B}_\Omega)$ mit*

$$\int_A K(\omega, B) P(d\omega) = P(A \cap B)$$

für alle $A \in \mathcal{F}$ und alle $B \in \mathcal{B}_\Omega$, das heißt, $K(\cdot, B)$ ist für jedes $B \in \mathcal{B}_\Omega$ eine Version der bedingten Wahrscheinlichkeit $P(B|\mathcal{F})$.

Wir benötigen einige Ergebnisse über metrische Räume, die hier nicht bewiesen werden.

Ist (K, d) ein kompakter, metrischer Raum, so ist die Metrik d offenbar beschränkt, das heißt, es gilt $\sup_{x,y\in K} d(x,y) < \infty$. Wir definieren auf dem unendlichen Produktraum $K^{\mathbb{N}}$ die Produktmetrik $\bar{d}$ durch

$$\bar{d}(x, y) = \sum_{i=1}^{\infty} 2^{-i} d(x_i, y_i),$$

$x = (x_i)_{i\in\mathbb{N}}$, $y = (y_i)_{i\in\mathbb{N}}$. Man überzeugt sich leicht davon, daß $\bar{d}$ eine Metrik ist und daß eine Folge in $(K^{\mathbb{N}}, \bar{d})$ genau dann konvergiert, wenn alle ihre Komponenten konvergieren.

Satz 7.11 (Satz von Tychonoff). *$(K^{\mathbb{N}}, \bar{d})$ ist kompakt.*

Beweisskizze. Sei $(x^n)_{n\in\mathbb{N}} = ((x_i^n)_i)_{n\in\mathbb{N}}$ eine Folge in $K^{\mathbb{N}}$. Wir wählen eine Teilfolge $(x^{n_{1,m}})_{m\in\mathbb{N}}$, so daß $(x_1^{n_{1,m}})_m$ konvergiert, dann eine Teilfolge $(x^{n_{2,m}})_{m\in\mathbb{N}}$ dieser Folge, so daß $(x_2^{n_{2,m}})_m$ konvergiert, etc. Dann konvergiert $(x^{n_{m,m}})_{m\in\mathbb{N}}$. □

Ist K kompakt, so bezeichnen wir mit $C(K)$ die Menge der stetigen Funktionen $f\colon K \to \mathbb{R}$. Auf $C(K)$ betrachten wir die Supremumnorm $\|f\|_\infty = \sup_{x\in K} f(x)$.

Satz 7.12. *Ist (K, d) ein kompakter, metrischer Raum, so ist $(C(K), \| \ \|_\infty)$ separabel.*

(Beweis siehe Dieudonné: Foundation of Modern Analysis (7.4.4).)

Satz 7.13 (DARSTELLUNGSSATZ VON RIESZ). *Sei (K,d) ein kompakter, metrischer Raum und ψ eine lineare Abbildung $C(K) \to \mathbb{R}$, die die folgenden Eigenschaften hat:*

(i) $f(x) \geq 0 \ \forall x \in K \Rightarrow \psi(f) \geq 0$
(ii) $\psi(1) = 1$.

Dann existiert genau ein Wahrscheinlichkeitsmaß μ auf $(K, \mathcal{B}_K)$ mit

$$\psi(f) = \int f(x)\,\mu(dx) \qquad \text{für } f \in C(K).$$

(Beweis siehe z. B. Cohn: Measure Theory, Chapter 8.2.)

Lemma 7.14. *Sei (Ω, d) separabel. Dann existieren ein kompakter, metrischer Raum (K, ϱ) und eine injektive Abbildung $\psi\colon \Omega \to K$, so daß für x_n, $x \in \Omega$, $\lim_{n\to\infty} \psi(x_n) = x$ genau dann gilt, wenn (x_n) gegen x konvergiert.*

Beweis. Sei $\{\xi_n\}_{n\in\mathbb{N}}$ eine abzählbare, dichte Teilmenge in Ω. Für $n, m \in \mathbb{N}$ sei $\varphi_{n,m}\colon \Omega \to [0,1]$ durch

$$\varphi_{n,m}(x) = ((2 - m\,d(x,\xi_n)) \wedge 1) \vee 0$$

definiert. (Wir benutzen die Schreibweise $y \wedge z = \min\{y,z\}$ und $y \vee z = \max\{y,z\}$ für $y, z \in \mathbb{R}$.) Die $\varphi_{n,m}$ sind offensichtlich stetig. $\varphi_{n,m}(x) = 1$ gilt genau dann, wenn $d(x,\xi_n) \leq 1/m$ ist, und $\varphi_{n,m}(x) = 0$ ist äquivalent zu $d(x,\xi_n) \geq 2/m$. Mit Hilfe dieser Funktionen definieren wir die Abbildung

$$\psi\colon \Omega \to [0,1]^{\mathbb{N}\times\mathbb{N}}$$

durch $\psi(x) = (\varphi_{n,m}(x))_{n,m\in\mathbb{N}}$. $[0,1]$ mit der üblichen Abstandsmetrik ist kompakt. Nach 7.11 ist $[0,1]^{\mathbb{N}\times\mathbb{N}}$ mit der Produktmetrik, die wir mit ϱ bezeichnen, ebenfalls kompakt.

Wir zeigen:

(a) ψ ist injektiv.
(b) Seien $x^{(n)}$, $x \in \Omega$. $\lim_{n\to\infty} x^{(n)} = x$ gilt dann und nur dann, wenn für alle i, j $\lim_{n\to\infty} \varphi_{i,j}(x^{(n)}) = \varphi_{i,j}(x)$ gilt, das heißt wenn $\lim_{n\to\infty} \psi(x^{(n)}) = \psi(x)$ in der Produktmetrik gilt.

(a) folgt unmittelbar aus (b).

(b): Aus $x^{(n)} \to x$ folgt $\varphi_{ij}(x^{(n)}) \to \varphi_{ij}(x)$ wegen der Stetigkeit der φ_{ij}.

Wir zeigen die Umkehrung: Sei $m \in \mathbb{N}$ beliebig. Wir wählen ξ_i mit $d(x,\xi_i) < 1/m$. Dann ist $\varphi_{i,m}(x) = 1$. Aus $\lim_{n\to\infty} \varphi_{i,m}(x^{(n)}) = 1$ folgt, daß für genügend große n $d(x^{(n)},\xi_i) \leq 2/m$ und somit $d(x^{(n)},x) \leq 3/m$ gilt. Somit folgt

$$\lim_{n\to\infty} x^{(n)} = x.$$

□

Wir können die Metrik ϱ mit Hilfe von ψ natürlich auf Ω übertragen. Der Einfachheit halber bezeichnen wir diese übertragene Metrik ebenfalls mit ϱ. Aus 7.14 folgt sofort, daß (Ω, ϱ) dieselben offenen Mengen hat wie (Ω, d) und demzufolge auch dieselbe Borel-σ-Algebra. ϱ hat gegenüber d jedoch einen Vorteil, nämlich:

Lemma 7.15. *$U_\varrho(\Omega)$ ist separabel bezüglich $\| \ \|_\infty$.*

Beweis. Sei $\widehat{K}$ der Abschluss von $\psi(\Omega)$ in (K, ϱ). $\widehat{K}$ ist kompakt. Offensichtlich sind die gleichmäßig stetigen, beschränkten Funktionen $\psi(\Omega) \to \mathbb{R}$ genau die Einschränkungen von Funktionen aus $C(\widehat{K})$ auf $\psi(\Omega)$. Da $C(\widehat{K})$ nach 7.12 separabel ist, folgt, daß $U_\varrho(\psi(\Omega))$ separabel ist. Daraus ergibt sich die Behauptung unmittelbar. $\square$

Bemerkung 7.16. $\widehat{K}$ ist natürlich die Vervollständigung von $(\psi(\Omega), \varrho)$. Im allgemeinen ist $\widehat{K} \neq \psi(\Omega)$, das heißt, $(\psi(\Omega), \varrho)$ ist nicht vollständig und demzufolge auch (Ω, ϱ) nicht. Es ergibt sich jedoch unmittelbar aus den obigen Überlegungen, daß die Vervollständigung $(\widehat{\Omega}, \varrho)$ von (Ω, ϱ) kompakt ist.

Beweis von 7.10. Wir benutzen die Metrik ϱ. Nach Lemma 7.16 ist $U_\varrho(\Omega)$ separabel. Wir konstruieren eine abzählbare, dichte Teilmenge in $U_\varrho(\Omega)$ mit etwas speziellen Eigenschaften. Sei zunächst $\{g_i\}_{i \in \mathbb{N}}$ eine beliebige abzählbare, dichte Teilmenge, wobei wir $g_1 \equiv 1$ voraussetzen dürfen. Falls notwendig, streichen wir in dieser Liste Elemente, so daß nachher je endlich viele linear unabhängig sind ($U_\varrho(\Omega)$ ist ein $\mathbb{R}$-VR): Wir streichen g_2, wenn $(1, g_2)$ linear abhängig ist und fahren so fort: g_n wird gestrichen, wenn es zusammen mit den bislang Ungestrichenen aus $g_1, \ldots, g_{n-1}$ linear abhängig ist. Die verbleibende Menge sei $\{f_i\}_{i \in \mathbb{N}}$. Im Fall $|\Omega| < \infty$ (und nur dann) bleiben nur endlich viele übrig, was jedoch nicht weiter stört. (Für $|\Omega| < \infty$ ist der Satz ohnehin trivial.) Die Folge (f_i) braucht nicht mehr dicht zu liegen, evidenterweise jedoch die abzählbare Menge

$$\Gamma = \Big\{ \sum_{i=1}^{n} q_i f_i : n \in \mathbb{N};\ q_1, \ldots, q_n \in \mathbb{Q} \Big\}.$$

Es sei $\widehat{f_i}$ eine Version von $E(f_i | \mathcal{A})$.

Lemma 7.17. *Es existiert eine Abbildung*

$$\Phi : \Omega \times U_\varrho(\Omega) \to \mathbb{R}$$

mit folgenden Eigenschaften:

(a) *Für alle $f \in U_\varrho(\Omega)$ ist $\omega \to \Phi(\omega, f)$ $\mathcal{F}$-meßbar.*
(b) $\int_A \Phi(\omega, f) P(d\omega) = \int_A f\, dP \ \ \forall f \in U_\varrho(\Omega),\ \forall A \in \mathcal{F}.$
(c) $\Phi(\omega, 1) = 1 \ \ \forall \omega.$
(d) $\Phi(\omega, f) \geq 0 \ \ \forall \omega,\ \forall f \geq 0.$
(e) *$f \to \Phi(\omega, f)$ ist $\mathbb{R}$-linear $\forall \omega$.*

Beweis. Für $f \in \Gamma$ ist die Darstellung $f = \sum_{i=1}^{n} q_i f_i$ (mit $n \in \mathbb{N}$ und $q_1, \ldots, q_n \in \mathbb{Q}$) eindeutig und demzufolge $\widehat{f} = \sum_{i=1}^{n} q_i \widehat{f_i}$ eindeutig definiert und $\mathcal{F}$-meßbar. Für jedes $\omega \in \Omega$ ist $\Gamma \ni f \mapsto \widehat{f}(\omega)$ eine $\mathbb{Q}$-lineare Abbildung. (Γ ist ein $\mathbb{Q}$-Vektorraum.)

Sei $\Gamma^+ = \{ f \in \Gamma \colon f(\omega) \geq 0 \ \forall \omega \}$. Wir definieren

$$F := \{ \omega \in \Omega \colon \exists f \in \Gamma^+ \text{ mit } \widehat{f}(\omega) < 0,\ \widehat{f_1}(\omega) \neq 1 \}$$

(für f_1 war die Funktion identisch 1). F ist offensichtlich eine P-Nullmenge in $\mathcal{F}$.

Sind $f \in \Gamma$, $\|f\|_\infty \leq \varepsilon$, ε rational und $\omega \notin F$, so liegen $f + \varepsilon f_1$ und $-f + \varepsilon f_1$ in Γ^+, und demzufolge gelten $\widehat{f}(\omega)$, $-\widehat{f}(\omega) \leq \varepsilon$, das heißt $|\widehat{f}(\omega)| \leq \varepsilon$. Mit dieser Überlegung können wir für $\omega \notin F$ die Abbildung $\Gamma \ni f \mapsto \widehat{f}(\omega)$ auf ganz U_ϱ ausdehnen:

Seien $f \in U_\varrho(\Omega)$ und $g_n \in \Gamma$ eine Folge, die gleichmäßig gegen f konvergiert. Nach der obigen Überlegung ist für $\omega \notin \Gamma$ $\widehat{g}_n(\omega)$ eine Cauchy-Folge, und wir setzen

$$\widehat{f}(\omega) := \lim_{n \to \infty} \widehat{g}_n(\omega).$$

Wir definieren

$$\Phi(\omega, f) = \begin{cases} \widehat{f}(\omega) & \text{für } \omega \notin F, \\ f(\omega_0) & \text{für } \omega \in F, \end{cases}$$

wobei ω_0 ein bliebiger, fest gewählter Punkt in Ω ist.

(a)–(e) sind nun leicht zu verifizieren: (a) folgt aus $F \in \mathcal{F}$, $\widehat{f}$ $\mathcal{F}$-meßbar für $f \in \Gamma$, und der Tatsache, daß punktweise Limiten von meßbaren Funktionen meßbar sind. (b) ist per Definition richtig für $f \in \Gamma$ und mit dem beschränkten Konvergenzsatz dann auch für $f \in U_\varrho(\Omega)$. (c) gilt per Definition. (d) folgt unmittelbar aus $\widehat{f}(\omega) \geq 0$ für $\omega \notin F$, $f \in \Gamma$. (e) folgt aus der $\mathbb{Q}$-Linearität von $\Gamma \ni f \mapsto \widehat{f}(\omega)$ und sei dem Leser überlassen. $\square$

Es bleibt nun noch zu zeigen, daß eine Darstellung

$$\Phi(\omega, f) = \int f(\omega') K(\omega, d\omega')$$

existiert, mit einem Kern K, der die richtigen Eigenschaften hat.

Sei $\widehat{\Omega}$ die Vervollständigung von (Ω, ϱ). $\widehat{\Omega}$ ist kompakt. Da $U_\varrho(\Omega)$ und $C(\widehat{\Omega})$ isomorph sind, läßt sich Φ als Abbildung $\Omega \times C(\widehat{\Omega}) \to \mathbb{R}$ auffassen. Nach dem Rieszschen Darstellungssatz existiert für jedes $\omega \in \Omega$ ein Wahrscheinlichkeitsmaß $K(\omega, \cdot)$ auf $(\widehat{\Omega}, \mathcal{B}_{\widehat{\Omega}})$ mit

$$\Phi(\omega, f) = \int f(\omega') K(\omega, d\omega').$$

Für $\omega \in F$ ist natürlich $K(\omega, \cdot) = \delta_{\omega_0}$.

Man überzeugt sich zunächst davon, daß K ein Markoff-Kern von $(\Omega, \mathcal{F})$ nach $(\widehat{\Omega}, \mathcal{B}_{\widehat{\Omega}})$ ist: Da $\{ A \in \mathcal{B}_{\widehat{\Omega}} \colon K(\cdot, A) \text{ ist } \mathcal{F}\text{-meßbar} \}$ ein Dynkin-System ist, genügt es zu zeigen, daß $K(\cdot, A)$ für abgeschlossene $A \subset \widehat{\Omega}$ $\mathcal{F}$-meßbar ist. Für abgeschlossene A existiert eine Folge $f_n \in C(\widehat{\Omega})$, $0 \leq f_n \leq 1$ mit $f_n \downarrow 1_A$. Somit gilt

$$K(\omega, A) = \lim_{n\to\infty} \Phi(\omega, f_n)$$

für alle ω. Die linke Seite ist also $\mathcal{F}$-meßbar als Funktion von $\omega \in \Omega$.

Wir müssen nun zeigen, daß K in Tat und Wahrheit (nach einer geringfügigen Korrektur) ein Markoff-Kern von $(\Omega, \mathcal{F})$ nach $(\Omega, \mathcal{B}_\Omega)$ ist. Dazu verwenden wir einen Spezialfall des Satzes von Prohorov, der in 7c bewiesen wird: Jedes Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B}_\Omega)$ — also auch P — ist straff, das heißt, es existiert eine aufsteigende Folge (K_n) von kompakten Teilmengen von Ω mit $P(K_n) \uparrow 1$. Da die Einbettung von Ω in $\widehat{\Omega}$ stetig ist, ist K_n kompakt in $\widehat{\Omega}$. Es existieren also $f_{n,m} \in C(\widehat{\Omega})$ mit $0 \leq f_{n,m} \leq 1$, $f_{n,m} \downarrow 1_{K_n}$ für $m \to \infty$. Die Einschränkungen der $f_{n,m}$ auf Ω, die wir ebenfalls mit $f_{n,m}$ bezeichnen, sind in $U_\varrho(\Omega)$, und wir hatten schon gesehen, daß

$$\int \Phi(\omega, f_{n,m}) P(d\omega) = E f_{n,m}$$

gilt. Daraus folgt sofort

$$P(K_n) = \int K(\omega, K_n) P(d\omega)$$

und mit $D := \bigcup_n K_n$

$$1 = \int K(\omega, D) P(d\omega),$$

und somit existiert $F' \in \mathcal{F}$ mit $P(F') = 0$ und $K(\omega, D) = 1$ für $\omega \notin F'$. D ist jedoch eine Borel-Menge in Ω. Modifizieren wir $K(\omega, \cdot)$ auf F', indem wir nochmals $K(\omega, \cdot) = \delta_{\omega_0}$ für $\omega \in F'$ setzen, so ist K ein Markoff-Kern von $(\Omega, \mathcal{F})$ nach $(\Omega, \mathcal{B}_\Omega)$.

$$\int_A K(\omega, B) P(d\omega) = P(A \cap B) \quad \text{für } A \in \mathcal{F} \text{und } B \in \mathcal{B}_\Omega$$

folgt aus 7.17(b), dem beschränkten Konvergenzsatz und einem nochmaligen Dynkin-Argument. Damit ist Satz 7.10 bewiesen. □

7c. Beweis des Satzes von Prohorov

Ein wichtiger Aspekt der schwachen Konvergenz von Wahrscheinlichkeitsmaßen auf einem vollständigen, separablen, metrischen Raum ist ihre Metrisierbarkeit.

Satz 7.18. *Es sei (S, d) vollständig und separabel. Dann existiert eine Metrik Δ auf $\mathbb{P}(S)$ mit der Eigenschaft, daß für $\mu_n, \mu \in \mathbb{P}(S)$ $\lim_{n\to\infty} \Delta(\mu_n, \mu)$ genau dann gilt, wenn μ_n schwach gegen μ konvergiert.*

Zum Beweis benötigen wir eine kleine Ergänzung zum Portmenteau-Theorem (Satz 6.7).

Lemma 7.19. *Sei $U_d(S)$ die Menge der beschränkten, gleichmäßig stetigen Abbildungen $S \to \mathbb{R}$. $\mu_n \xrightarrow{w} \mu$ gilt genau dann, wenn*

$$\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu \tag{7.20}$$

für alle $f \in U_d(S)$ gilt.

Beweis. Die Bedingung (7.20) ist offensichtlich notwendig. Im Beweisschritt (a)⇒(b) im Satz 6.8 wird (7.20) nur für gleichmäßig stetige Funktionen verwendet. Dies beweist die Umkehrung. □

Beweis von Satz 7.18. Offensichtlich können wir anstelle der Metrik d die in 7b eingeführte Metrik ϱ verwenden, die den Vorteil hat, daß $U_\varrho(S)$ nach 7.15 separabel ist. (Allerdings ist im allgemeinen (S, ϱ) nicht vollständig.) Wir können annehmen, daß $\|\varphi_k\|_\infty \neq 0$ für alle k gilt. Wir definieren für $\mu, \nu \in \mathbb{P}(S)$:

$$\Delta(\mu, \nu) = \sum_{k=1}^{\infty} \frac{1}{2^k \|\varphi_k\|_\infty} \left| \int \varphi_k\, d\mu - \int \varphi_k\, d\nu \right|.$$

Δ ist eine Metrik auf $\mathbb{P}(S)$, was der Leser selbst verifizieren möge. Seien $\mu_n, \mu \in \mathbb{P}(S)$. $\Delta(\mu_n, \mu) \to 0$ gilt genau dann, wenn $\int \varphi_k\, d\mu_n \to \int \varphi_k\, d\mu$ für alle k gilt. Wegen der Dichtheit von $\{\varphi_1, \varphi_2, \dots\}$ in $U_\varrho(S)$ ist dies äquivalent zu $\int f\, d\mu_n \to \int f\, d\mu$ für alle $f \in U_\varrho(S)$, was wegen 7.19 äquivalent zu $\mu_n \xrightarrow{w} \mu$ ist. □

Da die schwache Topologie metrisierbar ist ((S, d) Polnisch vorausgesetzt), ist eine Menge $\Gamma \subset \mathbb{P}(S)$ genau dann kompakt, wenn sie folgenkompakt ist, das heißt wenn jede Folge eine konvergente Teilfolge hat. Wir brauchen also nicht zwischen diesen Begriffen zu unterscheiden. Wir benötigen eine Charakterisierung kompakter Teilmengen von vollständigen, metrischen Räumen.

Definition 7.21. Eine Teilmenge A eines metrischen Raumes (X, d) heißt *total beschränkt,* wenn zu jedem $\varepsilon > 0$ eine endliche Überdeckung von A mit Mengen von d-Durchmesser $< \varepsilon$ existiert (d-Durchmesser einer Menge $B := \sup_{x,y\in B} d(x, y)$).

Lemma 7.22. *Sei (X, d) vollständig. Eine Teilmenge $A \subset X$ ist genau dann kompakt, wenn sie abgeschlossen und total beschränkt ist.*

Beweisskizze. A kompakt ⇒ A abgeschlossen und total beschränkt, ist offensichtlich.

Die Umkehrung verwendet das übliche Heine-Borel Argument: Sei (x_n) eine Folge in A. Zu jedem $m > 0$ nehme man eine endliche Überdeckung von A mit Mengen von Durchmesser $< 1/m$. Dann existiert für jedes m eine Teilfolge von (x_n), deren Elemente alle bis auf endlich viele in *einer* Menge der zu m gehörigen Überdeckung liegen. Mit dem üblichen Diagonalargument (siehe 7.11) konstruiert man eine Teilfolge, für die das für jedes m gilt. Diese ist eine Cauchyfolge, also konvergiert sie in X. Da A abgeschlossen ist, liegt der Grenzpunkt in A. □

Beweis des Satzes von Prohorov (Satz 6.17).

I) Es sei $\Gamma \subset \mathbb{P}(S)$ relativ kompakt. Wir zeigen, daß zu $\varepsilon > 0$ eine kompakte Menge K existiert mit

$$\mu(K) \geq 1 - \varepsilon$$

für alle $\mu \in \Gamma$. Wir benützen die Metrik d. Wir können voraussetzen, daß Γ kompakt ist, andernfalls ersetzen wir Γ durch seinen Abschluß. Sei $\{ s_j : j \in \mathbb{N} \}$ eine abzählbare, dichte Teilmenge von S. $U_\varepsilon(s)$ bezeichne die offene k-Umgebung von s in der Metrik d. Wir setzen

$$G_{k,n} = \bigcup_{j=1}^{n} U_{1/k}(s_j).$$

Die Funktionen $\lambda_{k,n} : \Gamma \to [0,1]$

$$\lambda_{k,n}(\mu) := 1 - \mu(G_{k,n})$$

sind nach Satz 6.7(c) oberhalb halbstetig, da $G_{k,n}$ offen ist. Für jedes k gilt $G_{k,n} \uparrow S$ für $n \to \infty$ und demzufolge $\lambda_{k,n}(\mu) \downarrow 0$ für jedes $k \in \mathbb{N}$ und $\mu \in \Gamma$. Da Γ kompakt ist, folgt aus dem Satz von Dini, daß für jedes $\varepsilon > 0$ ein n_k existiert mit

$$\lambda_{k,n_k}(\mu) \leq \varepsilon/2^k \tag{7.23}$$

für alle $\mu \in \Gamma$. Wir setzen

$$K := \bigcap_{k=1}^{\infty} \bar{G}_{k,n_k}.$$

K ist abgeschlossen und offensichtlich total beschränkt und demzufolge (nach 7.22) kompakt. $\mu(K) \geq 1 - \varepsilon$ für alle $\mu \in \Gamma$ folgt unmittelbar aus (7.23).

II) Es sei $\Gamma \in \mathbb{P}(S)$ straff.

(i) Wir diskutieren zunächst den Spezialfall, daß (S, d) kompakt ist.

Wir haben zu zeigen, daß $\mathbb{P}(S)$ kompakt ist. Dies ist ein bekanntes Ergebnis aus der Funktionalanalysis, dessen Beweis kurz skizziert werden soll:

Sei $\{f_1, f_2, f_3, \dots\}$ eine abzählbare, dichte Teilmenge von $C(S)$ (Weierstraßscher Approximationssatz 7.12). Ist $(\mu_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{P}(S)$, so ist für jedes j die reelle Zahlenfolge $(\int f_j \, d\mu_n)_{n\in\mathbb{N}}$ beschränkt. Mit dem üblichen Diagonalverfahren (siehe Beweis von 7.11) zeigt man, daß eine Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ existiert, so daß

$$\alpha_j = \lim_{k\to\infty} \int f_j \, d\mu_{n_k}$$

für alle $j \in \mathbb{N}$ existiert. Die folgenden Aussagen (a)–(c) sind einfache Routineüberlegungen. Ihr Beweis sei dem Leser überlassen.

(a) Ist $f \in C(S)$ und gilt $f_{j_k} \to f$, $k \to \infty$, in $(C(S), \| \, \|_\infty)$, so existiert $\alpha(f) = \lim_{k\to\infty} \alpha_{j_k}$ und ist eindeutig durch f definiert (hängt also nicht von der speziell gewählten Teilfolge ab).

(b) $f \mapsto \alpha(f)$ ist linear und erfüllt $\alpha(f) \geq 0$ für $f \geq 0$ und $\alpha(1) = 1$. Nach dem Rieszschen Darstellungssatz existiert $\mu \in \mathbb{P}(S)$ mit $\alpha(f) = \int f \, d\mu$, $f \in C(S)$.

(c) $\mu_{n_k} \xrightarrow{w} \mu$ für $k \to \infty$.

Damit ist gezeigt, daß $\mathbb{P}(S)$ kompakt ist.

(ii) Allgemeiner Fall: Wir verwenden die Metrik ϱ. Nach Bemerkung 7.16 ist die Vervollständigung $(\widehat{S}, \varrho)$ von (S, ϱ) kompakt. Die Einbettung $\iota \colon S \to \widehat{S}$ ist stetig. Für $\mu \in \mathbb{P}(S)$ sei $\widehat{\mu} := \mu\iota^{-1}$. Sei (μ_n) eine Folge in Γ. Nach (i) hat $(\widehat{\mu}_n)$ eine konvergente Teilfolge in $\mathbb{P}(\widehat{S})$: $\widehat{\mu}_{n_k} \to \nu \in \mathbb{P}(\widehat{S})$. Wir zeigen:

(a) $\nu = \widehat{\mu}$ für ein $\mu \in \mathbb{P}(S)$.
(b) $\mu_{n_k} \xrightarrow{w} \mu$ in $\mathbb{P}(S)$.

Damit ist gezeigt, daß Γ relativ kompakt ist.

(a) Da Γ straff ist, existieren kompakte Mengen $K_m \subset S$ mit $K_m \uparrow$ und $\mu_{n_k}(K_m) \geq 1 - 1/m$ für alle k. Die Mengen K_m sind ebenfalls kompakt in $\widehat{S}$, also insbesondere abgeschlossen. Nach dem Portmenteau-Theorem (Satz 6.7(b)) folgt:

$$\nu\Big(\bigcup_m K_m\Big) = \lim_{m\to\infty} \nu(K_m) \geq \lim_{m\to\infty} \limsup_{k\to\infty} \widehat{\mu}_{n_k}(K_m) \geq \lim_{m\to\infty} (1 - 1/m) = 1.$$

Die Einschränkung μ von ν auf $\bigcup_m K_m$ ist also ein Wahrscheinlichkeitsmaß auf S, und es gilt $\nu = \widehat{\mu}$.

(b) Wegen $\lim_{n\to\infty} \int f \, d\widehat{\mu}_n = \int f \, d\nu$ für $f \in C(\widehat{S})$ folgt $\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu$ für alle $f \in U_\varrho(S)$. Nach 7.19 folgt (b) $\square$

Stichwortverzeichnis

G

H

I

K

L

M