Wahrscheinlichkeitstheorie

Fakultät für Mathematik
Bergische Universität Wuppertal

Martin Friesen[1]

Letzte Version vom
18. Januar 2018

[1] friesen@math.uni-wuppertal.de

Inhaltsverzeichnis

1 Grundbegriffe der Maß und Integrationstheorie

1.1 Maßräume

Die folgenden Begriffe, Definitionen und Konstruktionen werden als bekannt vorausgesetzt.

Definition 1.1. *Sei Ω eine nicht-leere Menge. Ein Dynkin-System ist ein Mengensystem $\mathcal{D}$ mit den folgenden Eigenschaften*

- $\Omega \in \mathcal{D}$.
- *Ist $A \in \mathcal{D}$, so folgt $A^c \in \mathcal{D}$.*
- *Ist $(A_n)_{n\in\mathbb{N}} \subset \mathcal{D}$ mit $A_n \cap A_m = \varnothing$ (paarweise disjunkt), so folgt $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$.*

Beispiel 1.2. *Sei Ω eine nicht-leere Menge. Dann ist $2^\Omega = \{A \subset \Omega\}$ ein Dynkin-System. Ist $\mathcal{A}$ ein beliebiges Mengensystem auf Ω, so ist*

$$d(\mathcal{A}) = \bigcap_{\mathcal{A}' \text{ ist ein Dynkin-System mit } \mathcal{A}\subset\mathcal{A}'} \mathcal{A}'$$

das kleinste Dynkin-System welches $\mathcal{A}$ enthält.

Die nächste Definition ist zentral für die Maß und Integrationstheorie.

Definition 1.3. *Sei Ω eine nicht-leere Menge. Eine σ-Algebra ist ein Mengensystem $\mathcal{F}$ mit den folgenden Eigenschaften*

- $\Omega \in \mathcal{F}$.
- *Ist $A \in \mathcal{F}$, so ist auch $A^c \in \mathcal{F}$.*
- *Ist $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$, so ist auch $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.*

Man beachte, dass jede σ-Algebra somit auch ein Dynkin-System ist.

Beispiel 1.4. *Sei Ω eine nicht-leere Menge. Dann heisst $2^\Omega = \{A \subset \Omega\}$ Potzenz-σ-Algebra. Ist $\mathcal{A}$ ein Mengensystem auf Ω, so ist*

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{F} \text{ ist eine } \sigma-\text{Algebra mit } \mathcal{A}\subset\mathcal{F}} \mathcal{F}$$

die kleinste σ-Algebra welche $\mathcal{A}$ enthält.

Ein Mengensystem $\mathcal{A}$ heisst $\cap$-stabil, falls $A, B \in \mathcal{A}$ bereits $A \cap B \in \mathcal{A}$ impliziert. Der nächste Satz ist zentral für viele Maßtheoretische Argumente.

Satz 1.5. *Sei Ω eine nicht-leere Menge und $\mathcal{A}$ ein $\cap$-stabiles Mengensystem. Dann gilt*

$$d(\mathcal{A}) = \sigma(\mathcal{A}).$$

Definition 1.6. *Sei Ω eine nicht-leere Menge. Ein messbarer Raum ist ein Tupel $(\Omega, \mathcal{F})$ wo $\mathcal{F}$ eine σ-Algebra auf Ω ist. Ein Maß auf einem messbarem Raum $(\Omega, \mathcal{F})$ ist eine Abbildung $\mu : \mathcal{F} \longrightarrow [0, \infty]$ mit den folgenden Eigenschaften*

- $\mu(\varnothing) = 0$.
- *μ ist σ-additiv, d.h. ist $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ eine Folge paarweise disjunkter Mengen, so gilt*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

In diesem Fall heisst $(\Omega, \mathcal{F}, \mu)$ Maßraum.

Bemerkung 1.7. *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und sei $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ eine Folge von Mengen. Dann gilt*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

Zum Schluss wollen wir die Konstruktion von Maßen mittels dem Satz von Caratheodory angeben.

Satz 1.8. *Sei Ω ein nicht-leere Menge und sei $\mathcal{A}$ ein Mengensystem mit den Eigenschaften*

- $\varnothing \in \mathcal{A}$.
- *Ist $A, B \in \mathcal{A}$, so gilt $A \cup B \in \mathcal{A}$ und $A \backslash B \in \mathcal{A}$.*

Sei $\mu : \mathcal{A} \longrightarrow [0, \infty]$ eine Abbildung mit den Eigenschaften

- $\mu(\varnothing) = 0$.
- *Ist $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ eine Folge von paarweise disjunkten Mengen mit $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$, so folgt*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

Definiere eine Abbildung $\mu^* : 2^\Omega \longrightarrow [0, \infty]$ durch

$$\mu^*(A) = \inf\left\{ \sum_{n=1}^{\infty} \mu(A_n) \;\middle|\; (A_n)_{n\in\mathbb{N}} \subset \mathcal{A},\ A \subset \bigcup_{n\in\mathbb{N}} A_n \right\}$$

mit $\inf(\varnothing) = +\infty$ und sei

$$\mathcal{F}_\mu = \{A \subset \Omega \mid \mu^*(B) = \mu^*(A) + \mu^*(B \cap (\Omega \backslash A)), \quad \forall B \subset \Omega\}.$$

Dann gelten die folgenden Eigenschaften

- *$\mu^*(A) = \mu(A)$ für alle $A \in \mathcal{A}$ und*

$$\mu^*(B) = \mu^*(A) + \mu^*(B \cap (\Omega \backslash A)), \quad B \subset \Omega, \quad A \in \mathcal{A}.$$

- *$\mathcal{F}_\mu$ ist eine σ-Algebra mit $\mathcal{A} \subset \mathcal{F}_\mu$.*
- *μ^* ist ein Maß auf $(\Omega, \mathcal{F}_\mu)$.*

Gilt zuätzlich: Es gibt eine Folge $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ mit $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$, so ist $\mu^*|_{\mathcal{F}_\mu}$ eindeutig durch die Forderung $\mu^*(A) = \mu(A)$ für alle $A \in \mathcal{A}$ festgelegt.

Das wichtigste Beispiel ist das Lebesguemaß auf $\mathbb{R}^d$.

Beispiel 1.9. *Setze*

$$m^*(A) = \inf \left\{ \sum_{n\in\mathbb{N}} m(Q_n) \ : \ A \subset \bigcup_{n\in\mathbb{N}} Q_n, \ Q_n = (a_1^{(n)}, b_1^{(n)}] \times \ldots (a_d^{(n)}, b_d^{(n)}], \quad a_i^{(n)} < b_i^{(n)} \right\}$$

wo $m(Q_n) = \prod_{j=1}^{d} (b_j^{(n)} - a_j^{(n)})$. Eine Menge $A \subset \mathbb{R}^d$ heisst Lebesgue messbar, falls

$$m^*(A) = m^*(A \cap B) + m^*(A \backslash B), \quad \forall B \subset \mathbb{R}^d.$$

Sei $\mathcal{F}_m$ das Mengensystem aller Lebesgue messbaren Mengen. Dann ist $\mathcal{F}_m$ eine σ-Algebra und m^* ein Maß auf $\mathcal{F}_m$ (das Lebesgue Maß).

Sei $\mathcal{B}(\mathbb{R}^d)$ die Borel-σ-Algebra auf $\mathbb{R}^d$ erzeugt durch die offenen Mengen, d.h.

$$\mathcal{B}(\mathbb{R}^d) = \sigma(A \subset \mathbb{R}^d \mid A \text{ ist offen }).$$

Es lässt sich zeigen, dass $\mathcal{B}(\mathbb{R}^d) \subset \mathcal{F}_m$ und dass $\mathcal{F}_m$ echt grösser ist als $\mathcal{B}(\mathbb{R}^d)$. Die Einschränkung vom m^* auf $\mathcal{B}(\mathbb{R}^d)$ wird häufig als Lebesgue-Borel-Maß bezeichnet. Für unsere Zwecke wird eine solche Einschräkung häufig ausreichen.

1.2 Messbare Abbildungen und Elementarfunktionen

Definition 1.10. *Seien $(\Omega, \mathcal{F})$ sowie $(\Omega', \mathcal{F}')$ zwei messbare Räume. Eine Abbildung $f : \Omega \longrightarrow \Omega'$ heisst messbar (bezüglich $\mathcal{F} - \mathcal{F}'$), falls*

$$f^{-1}(A) = \{\omega \in \Omega \mid f(\omega) \in A\} \in \mathcal{F}, \quad \forall A \in \mathcal{F}'.$$

Auf $\mathbb{R}$ betrachten wir die σ-Algebra der Borel-Mengen $\mathcal{B}(\mathbb{R})$. Es sei $[-\infty, \infty] := \mathbb{R} \cup \{\infty, -\infty\}$ die erweiterte Zahlengerade. Diese ist ausgestattet mit der σ-Algebra

$$\mathcal{B}(\overline{\mathbb{R}}) := \{B \subset \overline{\mathbb{R}} \mid B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\}.$$

Wir setzen $\pm\infty \cdot 0 := 0$ sowie $\pm\infty \cdot \infty := \pm\infty$ und $a + \infty = \infty$ für $a \in \mathbb{R}$ etc. Auf diese Weise lassen sich messbare Funktionen $f : \Omega \longrightarrow \overline{\mathbb{R}}$ definieren.

Bemerkung 1.11. *Sei* $(\Omega, \mathcal{F})$ *ein messbarer Raum. Eine Funktion* $f : \Omega \longrightarrow \overline{\mathbb{R}}$ *ist messbar, falls*

$$\{\omega \in \Omega \mid X(\omega) < a\} \in \mathcal{F}, \quad \forall a \in \overline{\mathbb{R}}.$$

Eine analoge Aussage gilt für eine Funktion $f : \Omega \longrightarrow \mathbb{R}$.

Satz 1.12. *Seien* $f, g : \Omega \longrightarrow \mathbb{R}$ *messbar. Dann gilt*

$$\{f < g\},\ \{f \leq g\},\ \{f = g\},\ \{f \neq g\} \in \mathcal{F}.$$

Satz 1.13. *Seien* $f, g : \Omega \longrightarrow \overline{\mathbb{R}}$ *messbar. Dann gilt*

- *Ist* $\{f = +\infty\} \cap \{g = -\infty\} = \varnothing = \{f = -\infty\} \cap \{g = +\infty\}$, *dann ist* $f + g$ *wohldefiniert und messbar.*
- $f \cdot g$ *ist messbar. Insbesondere ist* $f \cdot a$ *messbar für alle* $a \in \overline{\mathbb{R}}$.

Satz 1.14. *Seien* $f_n : \Omega \longrightarrow \overline{\mathbb{R}}$ *messbar. Dann gilt*

- $\sup_{n\in\mathbb{N}} f_n$, $\inf_{n\in\mathbb{N}} f_n$, $\limsup_{n\to\infty} f_n$, $\liminf_{n\to\infty} f_n$ *sind messbar.*
- *Existiert der Grenzwert* $\lim_{n\to\infty} f_n(\omega) =: f(\omega) \in \mathbb{R}$ *für alle* $\omega \in \Omega$, *so ist* f *messbar.*

Korollar 1.15. *Ist* $f : \Omega \longrightarrow \overline{\mathbb{R}}$ *messbar, so sind insbesondere auch*

$$\begin{aligned} |f| &:= \sup\{f, -f\}, \\ f^+ &:= \sup\{f, 0\}, \\ f^- &:= -\inf\{f, 0\} \end{aligned}$$

messbar.

Für die Konstruktion des Lebesgue Integrales ist die Klasse der Elementarfunktionen von besonderer Bedeutung.

Definition 1.16. *Eine messbare Funktion* $f : \Omega \longrightarrow \overline{\mathbb{R}}$ *heisst Elementarfunktion, falls*

$$f = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}$$

wo $a_k \in [0, +\infty]$ *und* $A_k \in \mathcal{F}$ *paarweise disjunkt mit* $\Omega = \bigcup_{k=1}^{n} A_k$.

Man beachte, dass eine solche Darstellung nicht notwendigerweise eindeutig ist.

Lemma 1.17. *Sei* $f : \Omega \longrightarrow [0, +\infty]$. *Dann ist* f *genau dann messbar, wenn es eine Folge von Elementarfunktionen* f_n *gibt mit*

$$f_n(\omega) \leq f_{n+1}(\omega), \quad \lim_{n\to\infty} f_n(\omega) = f(\omega), \quad \forall \omega \in \Omega.$$

Wir schreiben kurz: $f_n \searrow f$. *Ist* f *zusätzlich beschränkt, so konvergiert* f_n *gleichmässig gegen* f.

Beweis. Idee: Ist f Grenzwert von Elementarfunktionen, so ist es messbar nach vorhergehendem Satz.
Ist f gegeben, so erfüllt

$$f_n := n\mathbb{1}_{f\geq n} + \sum_{i=0}^{n2^n-1} \frac{i}{2^n} \mathbb{1}_{\frac{i}{2^n} \leq f < \frac{i+1}{2^n}}$$

das gewünschte. □

1.3 Lebesgue Integral

In diesem Abschnitt geben wir eine kurze Skizze zur Konstruktion des Lebesgue Integrales. Diese Skizze ist ohne Beweise wobei einige technische Feinheiten unterschlagen werden. Die Grundidee dieser Konstruktion wird später immer wieder in Beweisen verwendet werden. Hier und im folgenden sei $(\Omega, \mathcal{F}, \mu)$ ein Massraum

Schritt 1. Indikatiorfunktionen

Für eine Funktion der Form $f = \mathbb{1}_A$, wo $A \in \mathcal{F}$ definieren wir

$$\int_\Omega f d\mu := \mu(A).$$

Schritt 2. Elementarfunktionen

Sei f eine Elementarfunktion der Form

$$f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$$

wo $a_k \in [0, +\infty]$ und $A_k \in \mathcal{F}$ paarweise disjunkt mit $\Omega = \bigcup_{k=1}^n A_k$. Dann definieren wir

$$\int_\Omega f d\mu := \sum_{k=1}^n a_k \mu(A_k).$$

Diese Definition ist sinvoll in dem Sinne, dass wir Schritt 1 linear auf Summen erweitern. In diesem Zusammenhang ist es wichtig die folgenden Punkte zu überprüfen:

- Dieses Integral ist wohldefiniert, d.h. es hängt nicht ab von der genauen Wahl der Darstellung der Elementarfunktion ab.
- Das so definierte Integral ist linear auf Elementarfunktionen.

Schritt 3. Erweiterung auf nicht-negative Funktionen

Sei $f : \Omega \longrightarrow [0, +\infty]$ messbar. Dann definieren wir

$$\int_\Omega f d\mu := \sup \left\{ \int_\Omega g d\mu \ : \ g \leq f \text{ Elementarfunktion} \right\}.$$

Das so definierte Integral sieht nicht notwendigerweise linear aus, dieses muss daher bewiesen werden.

Satz 1.18. *Seien $f, g : \Omega \longrightarrow [0, +\infty]$ messbar. Dann gelten die folgenden Eigenschaften*

- *Für alle $a, b \in [0, +\infty]$ ist*

$$\int_\Omega (af + bg) d\mu = a \int_\Omega f d\mu + b \int_\Omega g d\mu.$$

- *Ist $f \le g$, so gilt*

$$\int_\Omega f d\mu \le \int_\Omega g d\mu.$$

Für das Rechnen mit Integralen ist der nächste Satz zentral.

Satz 1.19. *(Beppo-Levi, Satz der monotonen Konvergenz) Seien $f_n : \Omega \longrightarrow [0, +\infty]$ messbar und monoton wachsend, d.h. $f_n(\omega) \le f_{n+1}(\omega)$ für alle $\omega \in \Omega$. Definiere $f(\omega) = \lim_{n\to\infty} f_n(\omega) = \sup_{n\in\mathbb{N}} f_n(\omega)$. Dann ist f messbar und es gilt*

$$\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu = \sup_{n\to\infty} \int_\Omega f_n d\mu.$$

Satz 1.20. *(Lemma von Fatou) Seien $f_n : \Omega \longrightarrow [0, +\infty]$ messbar. Dann gilt*

$$\int_\Omega \liminf_{n\to\infty} f_n d\mu \le \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

Schritt 4. Erweiterung auf integrierbare Funktionen

Bisher haben wir das Integral lediglich für nicht-negative Funktionen definiert. Im letzten Schritt erweitern wir dieses messbare Funktionen $f : \Omega \longrightarrow \overline{\mathbb{R}}$.

Definition 1.21. *Eine messbare Funktion $f : \Omega \longrightarrow \overline{\mathbb{R}}$ heisst integrierbar, falls*

$$\int_\Omega |f| d\mu < \infty.$$

Schreiben wir $|f| = f^+ + f^-$ so sehen wir, dass f genau dann integrierbar ist, falls

$$\int_\Omega f^+ d\mu, \quad \int_\Omega f^- d\mu < \infty.$$

Sei f integrierbar. Dann definieren wir

$$\int_\Omega f d\mu := \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu.$$

Für $A \in \mathcal{F}$ setzen wir

$$\int_A f d\mu = \int_\Omega \mathbb{1}_A f d\mu.$$

Häufig schreiben wir

$$\int_\Omega f d\mu := \int_\Omega f(\omega) d\mu(\omega)$$

sofern wir hervorheben wollen bezüglich welcher Variable integriert wird.

Bemerkung 1.22. *Ist $f : \Omega \longrightarrow [0, +\infty]$, so ist f integrierbar genau dann, wenn*

$$\int_\Omega f d\mu < \infty.$$

Satz 1.23. *Seien f, g integrierbare Funktionen. Dann gelten die folgenden Aussagen*

- *af ist für alle $a \in \mathbb{R}$ integrierbar und es gilt*

$$\int_\Omega a f d\mu = a \int_\Omega f d\mu.$$

- *Ist $f + g$ wohldefiniert, d.h. $\{f = +\infty\} \cap \{g = -\infty\} = \emptyset = \{f = -\infty\} \cap \{g = +\infty\}$, so ist auch $f + g$ integrierbar und es gilt*

$$\int_\Omega (f + g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu.$$

- *Ist $f \leq g$, so gilt*

$$\int_\Omega f d\mu \leq \int_\Omega g d\mu.$$

- *Es gilt die Dreiecksungleichung*

$$\left| \int_\Omega f d\mu \right| \leq \int_\Omega |f| d\mu.$$

- $\sup\{f, g\}$ *sowie* $\inf\{f, g\}$ *sind integrierbar.*

Die Bedingung, dass $f + g$ wohldefiniert ist, kann fallengelassen werden. Dieses werden wir später für den Fall von "Zufallsvariablen" zeigen.

1.4 Der Satz von Fubini und Tonelli

1.5 Der Satz von Radon-Nikodym

2 Grundbegriffe der Wahrscheinlichkeitstheorie

2.1 Wahrscheinlichkeitsräume und Verteilungen

Definition 2.1. *Ein Wahrscheinlichkeitsraum ist ein Maßraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathbb{P}(\Omega) = 1$.*

Hier und im Folgenden bezeichnet $(\Omega, \mathcal{F}, \mathbb{P})$ einen Wahrscheinlichkeitsraum.
Notation:

- Ω – Ereignisraum
- $\omega \in \Omega$ – Elementarereignis
- $A \in \mathcal{F}$ – Ereignis
- $\mathbb{P}$ – Wahrscheinlichkeitsmaß
- $\mathbb{P}(A)$ – Wahrscheinlichkeit von A.
- A^c – Ereignis A tritt nicht ein.
- $A \cup B$ – A oder B treten ein.
- $A \cap B$ – A und B treten ein.
- $A \subset B$ – A impliziert B.

Man beachte

$$0 \leq \mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = 1$$

sowie

$$\mathbb{P}(A^c) = \mathbb{P}(\Omega \backslash A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A).$$

Bemerkung 2.2. *Für eine Folge von Mengen $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$*

- $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.
- $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{N\to\infty} \mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right)$.
- $\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{N\to\infty} \mathbb{P}\left(\bigcap_{n=1}^{N} A_n\right)$.

Die Folgende Definition ist zentral für die Wahrscheinlichkeitstheorie.

Definition 2.3. *Eine Zufallsvariable, ist eine messbare Abbildung $X : \Omega \longrightarrow \overline{\mathbb{R}}$.*

Definition 2.4. *Sei X eine Zufallsvariable. Die Verteilung von X ist definiert durch*

$$\mu_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

Wir schreiben häufig auch $\mu_X = \mathbb{P} \circ X^{-1}$.

Satz 2.5. *Sei X eine Zufallsvariable. Dann gilt:*

(a) μ_X ist ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\overline{\mathbb{R}})$.

(b) Ist $h : \overline{\mathbb{R}} \longrightarrow [0, \infty]$ messbar, so gilt

$$\int\limits_{\Omega} h(X) d\mathbb{P} = \int\limits_{\overline{\mathbb{R}}} h(x) d\mu_X(x). \tag{2.1}$$

Hierbei ist die linke Seite endlich genau dann, wenn die rechte Seite endlich ist.

(c) Ist $h : \overline{\mathbb{R}} \longrightarrow \overline{\mathbb{R}}$ messbar und ist eine der Seiten in (b) endlich für $|h|$ so gilt (b) immernoch.

Diesselbe Aussage gilt auch wenn wir $\overline{\mathbb{R}}$ durch $\mathbb{R}$ ersetzen.

Beweis. (a) Übung.
(b) Sei zunächst $h = a_\infty \mathbb{1}_\infty + \sum_{k=1}^N a_k \mathbb{1}_{A_k}, \quad a_k, a_\infty \in [0, +\infty]$ und $A_k \in \mathcal{F}$. Dann gilt

$$\begin{aligned}\int\limits_{\Omega} h(X) d\mathbb{P} &= a_\infty \mathbb{P}(X = +\infty) + \sum_{k=1}^N a_k \mathbb{P}(X \in A_k) \\ &= a_\infty \mu_X(\{+\infty\}) + \sum_{k=1}^N a_k \mu_X(A_k) = \int\limits_{\overline{R}} h(x) d\mu_X(x),\end{aligned}$$

wobei die linke Seite endlich ist genau dann, wenn die rechte Seite endlich ist. Ist $h \geq 0$ beliebig messbar, so finden wir eine Folge von Elementarfunktionen h_n wie oben mit $h_n \nearrow h$ für $n \to \infty$. Dann gilt $\int_\Omega h_n(X) d\mathbb{P} = \int_{\overline{\mathbb{R}}} h_n(x) d\mu_X(x)$ und die linke Seite ist endlich genau dann, wenn die rechte Seite endlich ist. Wegen $h_n(X) \nearrow h(X)$ fast sicher folgt mit dem Satz der monotonen Konvergenz (2.1).
(c) Ist $h : \overline{\mathbb{R}} \longrightarrow \overline{\mathbb{R}}$ messbar, so ist auch $|h| \geq 0$ messbar und es gilt

$$\int\limits_{\Omega} |h(X)| d\mathbb{P} = \int\limits_{\overline{\mathbb{R}}} |h(x)| d\mu_X(x).$$

Sind diese Integrale endlich, so erhalten wir die Behauptung durch Zerlegen in Positiv und Negativteil. □

Folglich können wir zu jeder Zufallsvariable eine Verteilung μ_X assozieren. Die Umkehrung ist im folgenden Satz angegeben.

Satz 2.6. *Sei μ ein Wahrscheinlichkeitsmaß auf $\overline{\mathbb{R}}$. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Zufallsvariable X mit $\mu_X = \mu$.*

Beweis. Sei $\Omega = \overline{\mathbb{R}}$, $\mathcal{F} = \mathcal{B}(\overline{R})$ sowie $\mathbb{P} = \mu$. Dann definiert

$$X : \Omega \longrightarrow \overline{\mathbb{R}}, \quad \omega \longmapsto X(\omega) = \omega$$

eine Zufallsvariable mit $\mu_X = \mathbb{P} \circ X^{-1} = \mathbb{P} = \mu$. □

Man beachte, dass die Wahl von X im obigen Beweis nicht eindeutig ist. Folglich enthält eine Zufallsvariable mehr Informationen als lediglich deren Verteilung μ_X. Um Verteilungen auf $\mathbb{R}$ zu untersuchen ist das nachfolgende Konzept nützlich.

Definition 2.7. *Sei μ ein Wahrscheinlichkeitsmaß auf $\overline{\mathbb{R}}$. Die Verteilungsfunktion F_μ ist definiert durch*

$$F_\mu(t) := \mu([-\infty, t]), \quad t \in \overline{\mathbb{R}}.$$

Satz 2.8. *(a) Sei μ ein Wahrscheinlichkeitsmaß auf $\overline{\mathbb{R}}$. Dann ist die Verteilungsfunktion F_μ monoton wachsend, rechtsstetig und es gilt*

$$\lim_{t \to -\infty} F_\mu(t) = 0, \quad \lim_{t \to \infty} F_\mu(t) = 1. \tag{2.2}$$

(b) Sei $F : \overline{\mathbb{R}} \longrightarrow [0,1]$ monoton wachsend, rechtsstetig mit (2.2). Dann gibt es genau ein Wahrscheinlichkeitsmaß μ auf $\overline{\mathbb{R}}$ mit $F_\mu = F$.

Diskreter Wahrscheinlichkeitsräume

Wir beginnen mit dem einfachsten Beispiel einer "diskreten" Wahrscheinlichkeitsverteilung.

Beispiel 2.9. *(Punktmaß) Es sei Ω eine nicht-leere Menge und $\omega_0 \in \Omega$ fest. Dann defininiert*

$$\delta_{\omega_0}(A) := \begin{cases} 1, & \omega_0 \in A \\ 0, & \omega_0 \notin A \end{cases}$$

ein Wahrscheinlichkeitsmaß auf Ω. Die σ-Algebra $\mathcal{F}$ kann hierbei beliebig gewählt werden. Für eine Zufallsvariable X mit $X(\omega_0) \in \mathbb{R}$ gilt

$$\int_\Omega X(\omega)\delta_{\omega_0}(d\omega) = X(\omega_0).$$

Der folgende Satz erweitert obiges Beispiel auf abzählbar viele Punktmaße.

Satz 2.10. *Sei $(\omega_n)_{n\in\mathbb{N}} \subset \Omega$ und $(p_n)_{n\in\mathbb{N}} \subset [0,1]$. Dann definiert für jede σ-Algebra $\mathcal{F}$ auf Ω*

$$\mathbb{P} = \sum_{n=1}^{\infty} p_n \delta_{\omega_n}, \tag{2.3}$$

ein Maß auf Ω. Dieses ist ein Wahrscheinlichkeitsmaß genau dann, wenn $\sum_{n=1}^\infty p_n = 1$. Es gelte $\sum_{n=1}^\infty p_n = 1$. Sei X eine Zufallsvariable mit $\sum_{n=1}^\infty |X(\omega_n)| p_n < \infty$. Dann ist

$$\int_\Omega X d\mathbb{P} = \sum_{n=1}^{\infty} X(\omega_n) p_n. \tag{2.4}$$

Beweis. $\mathbb{P}$ ist eine positive Linearkombination von Maßen und damit wieder ein Maß. Die zweite Behauptung ist offensichtlich. Für die letzte Behauptung zeigen wir zuerst (2.4) für alle nichtnegativen, messbaren Funktionen.

Fall 1: Sei $X = \sum\limits_{k=1}^{N} a_k \mathbb{1}_A$ mit $A_k \in \mathcal{F}$ und $a_k \geq 0$. Dann gilt

$$\begin{aligned}\int\limits_\Omega X d\mathbb{P} &= \sum_{k=1}^{N} a_k \int\limits_\Omega \mathbb{1}_{A_k} d\mathbb{P} = \sum_{k=1}^{N} a_k \mathbb{P}(A_k) \\ &= \sum_{k=1}^{N} a_k \sum_{n=1}^{\infty} p_n \delta_{\omega_n}(A_k) = \sum_{n=1}^{\infty} p_n \sum_{k=1}^{N} a_k \mathbb{1}_{A_k}(\omega_n) = \sum_{n=1}^{\infty} X(\omega_n) p_n.\end{aligned}$$

Fall 2: Sei $X \geq 0$ messbar, dann gibt es eine Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ wie in Fall 1 mit $X_n \nearrow X$ für $n \to \infty$. Folglich gilt

$$\int\limits_\Omega X_n d\mathbb{P} = \sum_{k=1}^{\infty} X_n(\omega_k) p_k, \quad n \geq 1.$$

Da beide Seiten monoton steigend in n sind ist folgt aus dem Satz der monotonen Konvergenz

$$\int\limits_\Omega X d\mathbb{P} = \lim_{n\to\infty} \int\limits_\Omega X_n d\mathbb{P} = \lim_{n\to\infty} \sum_{k=1}^{\infty} X_n(\omega_k) p_k \leq \sum_{k=1}^{\infty} X(\omega_k) p_k.$$

Sei $\varepsilon > 0$ und $N \in \mathbb{N}$ derart, dass $\sum_{k=N+1}^{\infty} X(\omega_k) p_k < \frac{\varepsilon}{2}$. Ferner gibt es $n_0(\varepsilon) \in \mathbb{N}$ mit $\sum_{k=1}^{N} X(\omega_k) p_k \leq \sum_{k=1}^{N} X_n(\omega_k) p_k + \frac{\varepsilon}{2}$ für $n \geq n_0$. Daraus folgt

$$\sum_{k=1}^{\infty} X(\omega_k) p_k \leq \sum_{k=1}^{N} X_n(\omega_k) p_k + \varepsilon \leq \varepsilon + \sum_{k=1}^{\infty} X_n(\omega_k) p_k = \varepsilon + \int\limits_\Omega X_n d\mathbb{P}, \quad n \geq n_0(\varepsilon).$$

Bilden wir den Grenzwert $n \to \infty$ auf beiden Seiten so ergibt sich

$$\sum_{k=1}^{\infty} X(\omega_k) p_k \leq \varepsilon + \int\limits_\Omega X d\mathbb{P}, \quad \varepsilon > 0.$$

Folglich gilt (2.4) für alle Zufallsvariablen $X \geq 0$. Insbesondere ist die linke Seite endlich genau dann, wenn die rechte Seite endlich ist.
Fall 3: Es sei X eine beliebige Zufallsvariable. Zerlegen wir $X = X^+ - X^-$ mit $X^\pm \geq 0$ in Positiv- und Negativteil, so folgt

$$\int\limits_\Omega X d\mathbb{P} = \int\limits_\Omega X^+ d\mathbb{P} - \int\limits_\Omega X^- d\mathbb{P} = \sum_{n=1}^{\infty} X^+(\omega_n) p_n - \sum_{n=1}^{\infty} X^-(\omega_n) p_n = \sum_{n=1}^{\infty} X(\omega_n) p_n.$$

□

Das folgende Beispiel ist ein Spezialfall von (2.3).

Beispiel 2.11. *(diskrete Gleichverteilung) Sei Ω ist eine endliche Menge mit Potenz-σ-Algebra $\mathcal{F} = 2^{\Omega}$. Die diskrete Gleichverteilung ist dadurch festgelegt, dass jedes Elementarereignis dieselbe Wahrscheinlichkeit hat, d.h. $\mathbb{P}(\{\omega\}) = a$ für ein geeignetes $a \in [0,1]$ und alle $\omega \in \Omega$ gilt. Wegen $\mathbb{P}(\Omega) = 1$ folgt $a = \frac{1}{|\Omega|}$, d.h.*

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}, \quad \omega \in \Omega,$$

wo $|\Omega| := \sum_{\omega \in \Omega} 1$ die Anzahl der Elemente in Ω bezeichnet. Für eine Menge $A \subset \Omega$ gilt demnach

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

Das so definierte Maß hat die Darstellung

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

Für eine Zufallsvariable X gilt demnach

$$\int_{\Omega} X d\mathbb{P} = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} X(\omega).$$

Das Nachfolgende ist eine Verallgemeinerung der Gleichverteilung und folgt aus Satz 2.10.

Korollar 2.12. *Sei Ω eine höchstens abzählbare Menge und $\mathcal{F}$ die Potenz-σ-Algebra of Ω. Dann gelten die folgenden Aussagen.*

(a) Sei $p : \Omega \longrightarrow [0,1]$ mit $\sum_{\omega \in \Omega} p(\omega) = 1$. Dann definiert

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega), \quad A \subset \Omega \tag{2.5}$$

ein Wahrscheinlichkeitsmaß auf Ω.

(b) Jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf Ω ist von der Form (2.5) für eine geeignete Funktion $p : \Omega \longrightarrow [0,1]$.

Für eine Zufallsvariable X mit $\sum_{\omega \in \Omega} |X(\omega)| p(\omega) < \infty$ gilt

$$\int_{\Omega} X d\mathbb{P} = \sum_{\omega \in \Omega} X(\omega) p(\omega). \tag{2.6}$$

Beweis. (a) $\mathbb{P}$ gegeben durch (2.5) hat die Darstellung

$$\mathbb{P}(A) = \sum_{\eta \in \Omega} p(\eta) \delta_{\eta}(A).$$

Die Behauptung folgt aus Satz 2.10
(b) Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf Ω und $A \subset \Omega$. Dann gilt

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

Also gilt (2.5) mit $p(\omega) := \mathbb{P}(\{\omega\})$. Die Identität (2.6) folgt aus Satz 2.10. □

Beispiel 2.13. *(Münzwurf) Sei $\Omega = \{0,1\}$ mit $\mathcal{F} = 2^\Omega$. Hierbei steht 0 für Kopf und 1 für Zahl. Unter der Annahme, dass die Münze fair ist (beide Ausgänge sind gleich wahrscheinlich), wählen wir $\mathbb{P}$ als diskrete Gleichverteilung auf Ω, d.h.*

$$\mathbb{P} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1.$$

Beispiel 2.14. *(n Münzwürfe) Führen wir $n \geq 1$ Münzwürfe mit derselben Münze aus, so wählen wir $\mathcal{F} = 2^\Omega$ mit*

$$\Omega = \{0,1\}^n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_k \in \{0,1\},\ k = 1, \dots, n\}.$$

Das Wahrscheinlichkeitsmaß ist demnach gegeben durch die Gleichverteilung auf Ω, d.h.

$$\mathbb{P}(\{\omega\}) = \frac{1}{2^n},$$

siehe (2.5). Wir betrachten die Zufallsvariable

$$X(\omega) = \sum_{k=1}^{n} \omega_k,$$

welche die absolute Häufigkeit angibt Zahl zu werfen. Diese definiert die disjunkte Zerlegung

$$\Omega = \bigcup_{k=0}^{n} A_k, \quad A_k = \{\omega \in \Omega \mid X(\omega) = k\}.$$

Es gilt $|A_k| = \binom{n}{k}$ und demnach $\mathbb{P}(A_k) = \frac{1}{2^n}\binom{n}{k}$. Die Verteilung von X ist somit gegeben durch $\mu_X = \sum_{k=0}^{n} \mathbb{P}(A_k)\delta_k$. Das Integral von X bezüglich $\mathbb{P}$ lässt sich wie folgt berechnen:

$$\int_\Omega X d\mathbb{P} = \frac{1}{2^n}\sum_{\omega \in \Omega} X(\omega) = \frac{1}{2^n}\sum_{k=0}^{n}\sum_{\omega \in A_k} X(\omega) = \frac{1}{2^n}\sum_{k=0}^{n} k\binom{n}{k} = \frac{n}{2}.$$

Diesselbe Rechnung lässt sich auch in einer anderen Notation kürzer führen

$$\int_\Omega X d\mathbb{P} = \sum_{k=0}^{n} \mathbb{P}(X = k)k = \frac{n}{2}.$$

Beispiel 2.15. *(unfairer Münzwurf) Führen wir $n \geq 1$ Münzwürfe mit derselben Münze aus, so wählen wir $\mathcal{F} = 2^\Omega$ mit*

$$\Omega = \{0,1\}^n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_k \in \{0,1\},\ k = 1, \dots, n\}.$$

Wir nehmen an, dass die Wahrscheinlichkeit Kopf (also 0) zu werfen gegeben ist durch $p \in [0,1]$. Die Wahrscheinlichkeit Zahl (also 1) zu werfen, ist demnach $1-p$. Das Wahrscheinlichkeitsmaß ist gegeben durch

$$\mathbb{P}(\{\omega\}) = p^{\sum\limits_{k=1}^{n}(1-\omega_k)}(1-p)^{\sum\limits_{k=1}^{n}\omega_k}, \quad \omega \in \Omega.$$

Wir betrachten die Ereignisse

$$A_k = \left\{\omega \in \Omega \mid \sum_{j=1}^{n} \omega_j = k\right\}, \quad k = 1, \dots, n.$$

Dann gilt $|A_k| = \binom{n}{k}$. Die Wahrscheinlichkeit k-mal Zahl zu werfen ist gegeben durch

$$\mathbb{P}(A_k) = \sum_{\omega \in A_k} p^{\sum\limits_{j=1}^{n}(1-\omega_j)}(1-p)^{\sum\limits_{j=1}^{n}\omega_j} = \binom{n}{k} p^{n-k}(1-p)^k.$$

Sei $X(\omega) := \sum\limits_{j=1}^{n} \omega_j$. Dann ist die Verteilung von X gegeben durch

$$\mu_X(\{k\}) := \mathbb{P}(X = k) = \binom{n}{k} p^{n-k}(1-p)^k, \quad k \in \{1, \dots, n\}.$$

Diese ist die sogenannte Binomialverteilung.

Beispiel 2.16. *(Poisson Verteilung) Sei $\Omega = \mathbb{R}_+$ mit $\mathcal{F} = \mathcal{B}(\mathbb{R}_+)$. Die Poisson Verteilung mit Parameter $\lambda > 0$ ist gegeben durch*

$$\mathbb{P} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta_k.$$

Kontinuierliche Wahrscheinlichkeitsräume

Wir wollen uns einige wichtige Beispiele von Verteilungen auf $\mathbb{R}^d$ ansehen.

Beispiel 2.17. *(uniforme Verteilung) Es sei $V \subset \mathbb{R}$ messbar mit $0 < m(V) < \infty$ und*

$$\mathcal{F} = \{A \in \mathcal{B}(\mathbb{R}) \mid A \subset V\}$$

die Menge aller Borel messbaren Mengen in V. Die uniforme Verteilung auf V ist gegeben durch

$$\mu(A) := \frac{1}{m(V)} \int_V \mathbb{1}_A(x) m(dx) = \frac{m(A)}{m(V)}.$$

Obige Definition kürzen wir ab mit $\mu(dx) = \frac{1}{m(V)} \mathbb{1}_V(x) m(dx)$.

Beispiel 2.18. *(Cauchy Verteilung) Es sei* $\mu(dx) = p(x)m(dx)$ *mit* $c > 0$, $a \in \mathbb{R}$ *und*

$$p(x) = \frac{1}{\pi} \frac{c}{c^2 + (x-a)^2}, \quad x \in \mathbb{R},$$

d.h. $\mu(A) = \int\limits_A p(x)m(dx)$ *für* $A \in \mathcal{B}(\mathbb{R})$. *Dann ist* μ *ein Wahrscheinlichkeitsmaß und es gilt*

$$\int\limits_{\mathbb{R}} |x|^k \mu(dx) = \infty, \quad k \geq 1.$$

Beispiel 2.19. *(Normalverteilung auf* $\mathbb{R}$*) Es sei* $\nu(dx) = p(x)m(dx)$ *mit* $\mu \in \mathbb{R}$, $\sigma^2 > 0$ *und*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Dann ist ν *ein Wahrscheinlichkeitsmaß mit*

$$\int\limits_{\mathbb{R}} x p(x) \nu(dx) = \mu$$

und

$$\int\limits_{\mathbb{R}} (x-\mu)^2 p(x) \nu(dx) = \sigma^2.$$

Man beachte dass obige Beispiele absolut stetig bezüglich dem Lebesgue-Maß auf $\mathbb{R}$ sind. In diesem Fall sagen wir auch, dass μ eine Dichte hat.

Definition 2.20. *(multivariate Normalverteilung)*

(a) Die Standardnormalverteilung auf $\mathbb{R}^d$ *ist gegeben durch ein Wahrscheinlichkeitsmaß* μ *auf* $\mathbb{R}^d$ *mit Dichte*

$$p(x) := (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}}, \quad x \in \mathbb{R}^d.$$

(b) Es sei A *eine reelle* $d \times d$*-Matrix und* $b \in \mathbb{R}^d$. *Sei* $T(x) := Ax + b$ *die dazugehörige affine Abbildung. Dann heißt* $\mu \circ T^{-1}$ *(multivariate) Normalverteilung, wo* μ *gegeben ist wie in (a).*

Satz 2.21. *Es gelten die folgenden Aussagen.*

(a) Sei A *invertierbar. Dann hat* $\mu \circ T^{-1}$ *eine Dichte gegeben durch*

$$\frac{d(\mu \circ T^{-1})}{dm}(x) = (2\pi)^{-\frac{d}{2}} (\det(AA^t))^{-\frac{1}{2}} e^{-\frac{1}{2}\langle (x-b), (AA^t)^{-1}(x-b) \rangle}, \quad x \in \mathbb{R}^d.$$

Hierbei bezeichnet A^t *die zu* A *transponierte Matrix und* $\langle \cdot, \cdot \rangle$ *das Skalarprodukt im* $\mathbb{R}^d$.

(b) Ist A *nicht invertierbar, so ist* $\mathrm{Im}(T)$ *eine Nullmenge bezüglich* m *und* $\mu \circ T^{-1}$ *ist nicht absolut stetig bezüglich* m.

(c) *Die Klasse der Normalverteilungen ist invariant unter affin linearen Transformationen $\mathbb{R}^d \longrightarrow \mathbb{R}^d$. Genauer, sei $T(x) = Ax + b$ mit A eine $d \times d$-Matrix, $b \in \mathbb{R}^d$ und sei ν eine (multivariate) Normalverteilung. Dann ist $\nu \circ T^{-1}$ eine (multivariate) Normalverteilung auf $\mathbb{R}^d$.*

Beweis. (a) Sei A invertierbar. Dann ist auch $\Sigma := AA^t$ invertierbar mit $\Sigma^{-1} = (A^t)^{-1}A^{-1}$ und es gilt $T^{-1}(y) = A^{-1}(y-b) = A^{-1}y - A^{-1}b$. Aus der Translationsinvarianz von m und dem Transformationssatz folgt für jedes Rechteck im $\mathbb{R}^d$

$$(m \circ T^{-1})(R) = m(T^{-1}(R)) = \det(A^{-1})m(R) = \det(A)^{-1}m(R).$$

Da Rechtecke einen $\cap$-stabilen Erzeuger der Borel-σ-Algebra bilden folgt

$$(m \circ T^{-1})(B) = \det(A)^{-1}m(B), \quad B \in \mathcal{B}(\mathbb{R}^d).$$

Also ist $m \circ T^{-1}$ absolut stetig bezüglich m und es gilt mit dem Determinanten Produktsatz

$$\frac{d(m \circ T^{-1})}{dm} = \det(A^{-1}) = (\det(\Sigma))^{-\frac{1}{2}}.$$

Die Behauptung folgt aus

$$\begin{aligned}
(\mu \circ T^{-1})(B) &= \int\limits_{\mathbb{R}^d} \mathbb{1}_B(x)(\mu \circ T^{-1})(dx) = \int\limits_{\mathbb{R}^d} \mathbb{1}_B(T(x))\mathbb{P}(dx) \\
&= \int\limits_{\mathbb{R}^d} \mathbb{1}_B(T(x))(2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|T^{-1}Tx|^2} m(dx) \\
&= \int\limits_{\mathbb{R}^d} \mathbb{1}_B(y)(2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|T^{-1}y|^2} (m \circ T^{-1})(dy) \\
&= \int\limits_{B} (2\pi)^{-\frac{d}{2}} e^{-\langle y-b, \Sigma^{-1}(y-b)\rangle} (\det(\Sigma))^{-\frac{1}{2}} m(dy)
\end{aligned}$$

wo wir $T^{-1}(y) = A^{-1}(y-b)$ und

$$|T^{-1}y|^2 = \langle T^{-1}y, T^{-1}y\rangle = \langle A^{-1}(y-b), A^{-1}(y-b)\rangle = \langle y-b, \sigma^{-1}(y-b)\rangle$$

benutzt haben.

(b) Ist A nicht invertierbar, so ist $B := \mathrm{Im}(T) \subset \mathbb{R}^d$ ein höchstens $d-1$-dimensionaler Unterraum und folglich eine Nullmenge bezüglich m, d.h. $m(B) = 0$. Wegen $\mathbb{1}_B(T(\mathbb{R}^d)) = 1$ erhalten wir

$$\mu \circ T^{-1}(B) = \int\limits_{\mathbb{R}^d} \mathbb{1}_B(x)(\mu \circ T^{-1})(dx) = \int\limits_{\mathbb{R}^d} \mathbb{1}_B(T(x))\mu(dx) = \mu(\mathbb{R}^d) = 1.$$

(c) Verknüpfungen affin linearer Funktionen sind affin linear. $\square$

2.2 Erwartungswert und Varianz

Definition 2.22. *Für eine nicht-negative Zufallsvariable ist der Erwartungswert definiert über*

$$\mathbb{E}(X) := \int_\Omega X d\mathbb{P} := \int_\Omega X(\omega) d\mathbb{P}(\omega) \in [0, \infty]. \tag{2.7}$$

Eine Zufallsvariable X heißt integrierbar, falls

$$\mathbb{E}(|X|) = \int_\Omega |X| d\mathbb{P} < \infty.$$

Für eine integrierbare Zufallsvariable definieren wir den Erwartungswert durch (2.7)*.*

Man beachte, dass für $X = 1$ gilt $\mathbb{E}(1) = \mathbb{P}(\Omega) = 1$.

Bemerkung 2.23. *Sei X eine integrierbare Zufallsvariable und μ_X deren Verteilung. Dann gilt*

$$\mathbb{E}(X) = \int_{\overline{\mathbb{R}}} x d\mu_X(x).$$

Allgemeiner, ist $\mathbb{E}(|X|^n) < \infty$ für ein $n \geq 1$, so gilt

$$\mathbb{E}(X^n) = \int_{\overline{\mathbb{R}}} x^n d\mu_X(x).$$

Um Erwartungswerte und Momente zu berechnen reicht es also aus die Verteilung der Zufallsvariable zu bestimmen.

Die Nachfolgende Ungleichung ist hilfreich für verschiedene Konvergenzargumente.

Lemma 2.24. *(Markovsche Ungleichung) Sei X eine Zufallsvariable und $h \geq 0$ monoton steigend. Dann gilt für alle $c > 0$*

$$h(c)\mathbb{P}(X \geq c) \leq \mathbb{E}(h(X)). \tag{2.8}$$

Insbesondere gilt die Markovsche Ungleichung für $p > 0$

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}(|X|^p), \quad \varepsilon > 0. \tag{2.9}$$

Beweis. Die Ungleichung (2.8) folgt aus

$$h(c)\mathbb{P}(X \geq c) \leq h(c)\mathbb{P}(h(X) \geq h(c)) = \mathbb{E}(h(c)\mathbb{1}_{h(X)\geq h(c)}) \leq \mathbb{E}(h(X)).$$

Die Ungleichung (2.9) ist eine Folgerung für den Spezialfall $h(x) = |x|^p$. $\square$

Eine Menge $N \in \mathcal{F}$ heißt Nullmenge, falls $\mathbb{P}(N) = 0$ gilt. Wir sagen, dass eine Eigenschaft fast sicher gilt, falls es eine Menge Nullmenge $N \in \mathcal{F}$ gibt und die Eigenschaft für alle $\omega \notin N$ erfüllt ist. Daraus lassen sich einige einfache Folgerungen ableiten.

Satz 2.25. *Sei X eine Zufallsvariable.*

(a) Ist $X \geq 0$ fast sicher und gilt $\mathbb{E}(X) = 0$, so gilt auch $X = 0$ fast sicher.

(b) Ist X eine integrierbare Zufallsvariable, so ist X fast sicher endlich.

(c) Sei N eine Nullmenge. Dann ist $\mathbb{1}_N X$ integrierbar und es gilt

$$\mathbb{E}(\mathbb{1}_N X) = \mathbb{E}(\mathbb{1}_N |X|) = 0.$$

Beweis. (a) Aus (2.9) mit $p = 1$ und $\varepsilon = \frac{1}{n}$ folgt für alle $n \geq 1$

$$\mathbb{P}\left(X \geq \frac{1}{n}\right) \leq n\mathbb{E}(X) = 0.$$

Daraus folgt

$$\mathbb{P}(X > 0) = \mathbb{P}\left(\bigcup_{n \geq 1} \left\{X \geq \frac{1}{n}\right\}\right) \leq \sum_{n=1}^{\infty} \mathbb{P}\left(X \geq \frac{1}{n}\right) = 0$$

und somit $\mathbb{P}(X = 0) = 1$.
(b) Es gilt $\mathbb{P}(|X| \geq n) \leq \frac{1}{n}\mathbb{E}(|X|)$ für alle $n \geq 1$. Lassen wir $n \to \infty$ folgt mit der Stetigkeit des Maßes von oben

$$\mathbb{P}(X \in \{\pm\infty\}) = \mathbb{P}(|X| = +\infty) = 0.$$

(c) Es reicht $X \geq 0$ zu betrachten, ansonsten betrachte die Zerlegung $X = X^+ - X^-$.
Fall 1: Ist $X = \sum_{k=1}^{N} a_k \mathbb{1}_{A_k}$ mit $a_k \geq 0$ und $A_k \in \mathcal{F}$, so folgt aus $\mathbb{P}(A_k \cap N) \leq \mathbb{P}(N) = 0$

$$\mathbb{E}(\mathbb{1}_N X) = \sum_{k=1}^{N} a_k \mathbb{P}(A_k \cap N) = 0.$$

Fall 2: Ist $X \geq 0$ eine beliebige Zufallsvariable, so finden wir eine Folge von Zufallsvariablen X_n wie in Fall 1 mit $X_n \searrow X$. Dann ist $\mathbb{E}(\mathbb{1}_N X) = \lim_{n\to\infty} \mathbb{E}(\mathbb{1}_N X_n) = 0$ nach dem Satz der monotonen Konvergenz. □

Dieses liefert uns das nächste Korollar.

Korollar 2.26. *(a) Sei X eine Zufallsvariable mit $X \geq 0$ fast sicher. Dann gilt $\mathbb{E}(X) \geq 0$.*

(b) Seien X, Y integrierbare Zufallsvariablen mit $X \leq Y$ fast sicher. Dann gilt $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

(c) Sind X, Y integrierbar und $a, b \in \mathbb{R}$, so ist auch $aX + bY$ integrierbar und es gilt

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Beweis. (a) Sei N eine Nullmenge mit $X(\omega) \geq 0$ für alle $\omega \in \Omega \backslash N$. Definiere $X' := \mathbb{1}_{\Omega\backslash N} X$. Dann ist $X' \geq 0$ eine Zufallsvariable und es gilt

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{1}_{\Omega\backslash N} X) = \mathbb{E}(X') \geq 0.$$

(b) Folgt direkt aus (a), da $Y - X \geq 0$ fast sicher.
(c) Wegen $|aX+bY| \leq |a||X|+|b||Y|$ ist $aX+bY$ integrierbar. Folglich ist $\mathbb{P}(|aX+bY| < \infty) = 1$ und somit

$$\begin{aligned}\mathbb{E}(aX + bY) &= \mathbb{E}(\mathbb{1}_{|aX+bY|<\infty}(aX + bY)) \\ &= a\mathbb{E}(\mathbb{1}_{|aX+bY|<\infty}X) + b\mathbb{E}(\mathbb{1}_{|aX+bY|<\infty}Y) = a\mathbb{E}(X) + b\mathbb{E}(Y).\end{aligned}$$

□

Definition 2.27. *Seien X, Y integrierbar derart, dass $X \cdot Y$ integrierbar sind. Die Kovarianz ist definiert durch*

$$\operatorname{cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(X))).$$

Für $X = Y$ bezeichnen wir mit $\operatorname{var}(X) := \operatorname{cov}(X, X)$ die Varianz von X.

Satz 2.28. *Es seien X, Y Zufallsvariablen mit $\mathbb{E}(|X|^2), \mathbb{E}(|Y|^2) < \infty$. Dann gelten folgenden Eigenschaften.*

1. $\operatorname{var}(X) \geq 0$ *und* $\operatorname{var}(X) = 0$ *genau dann wenn X fast sicher konstant ist.*
2. $\operatorname{cov}(X, Y)$ *ist symmetrisch und bilinear.*
3. $\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X, Y)$.

Beweis. Übung. □

2.3 Räume integrierbarer Zufallsvariablen

Wir definieren die Räume der integrierbaren Funktionen. Es sei $p \in [1, \infty)$ und setze

$$\|X\|_{\mathcal{L}^p} := (\mathbb{E}(|X|^p))^{\frac{1}{p}}.$$

Dann bezeichnet

$$\mathcal{L}^p := \mathcal{L}^p(\Omega) = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \longrightarrow \mathbb{R} \mid X \text{ Zufallsvariable mit } \|X\|_{\mathcal{L}^p} < \infty\}$$

den Raum der p-fach integrierbaren Zufallsvariablen. Für $p = \infty$ sei $\mathcal{L}^\infty$ der Raum der wesentlich beschränkten Zufallsvariablen. Eine Zufallsvariable heißt wesentlich beschränkt falls

$$\|X\|_{\mathcal{L}^\infty} := \operatorname*{ess\,sup}_{\omega \in \Omega} |X(\omega)| := \inf_{N \in \mathcal{F},\ \mathbb{P}(N)=0} \sup_{\omega \in \Omega \setminus N} |X(\omega)| < \infty.$$

Beispiel 2.29. *Sei $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, $\mathbb{P} = m$ das Lebesgue Maß und für $r \in \mathbb{R}$ sei*

$$X(\omega) := \begin{cases} \frac{1}{\omega^r}, & \omega \neq 0 \\ 0, & \omega = 0 \end{cases}.$$

$r \leq 0$: Dann ist $X \in \mathcal{L}^p$ für alle $p \in [1, \infty]$.
$r > 0$: Dann ist $X \notin \mathcal{L}^\infty$. Für $p \in [1, \infty)$ gilt

$$\mathbb{E}(|X|^p) = \int_{(0,1]} \omega^{-rp} m(d\omega).$$

Das rechte Integral ist endlich genau dann, wenn $rp < 1$, d.h. $X \in \mathcal{L}^p$ falls $rp < 1$.

Lemma 2.30. *Sei $h : \mathbb{R} \longrightarrow \mathbb{R}$ eine konvexe Funktion mit $h(X) \in \mathcal{L}^1$. Dann gilt $X \in \mathcal{L}^1$ sowie*

$$h(\mathbb{E}(X)) \leq \mathbb{E}(h(X)).$$

Beweis. Wir betrachten zuerst $X \geq 0$. Da h konvex ist, gibt es eine affin lineare Funktion g mit $g \leq h$ und $g(\mathbb{E}(X)) = h(\mathbb{E}(X))$ (Stützgerade). Es folgt

$$h(\mathbb{E}(X)) = g(\mathbb{E}(X)) = \mathbb{E}(g(X)) \leq \mathbb{E}(h(X))$$

welches in diesem Fall die Behauptung zeigt. Ist X nun beliebig, so ergibt obiger Fall angewand auf $|X|$, dass $X \in \mathcal{L}^1$. Diesselbe Argumentation zeigt dann die Behauptung. □

Korollar 2.31. *Für alle $1 \leq p \leq q \leq \infty$ gilt*

$$\|X\|_{\mathcal{L}^p} \leq \|X\|_{\mathcal{L}^q}$$

und somit ist $\mathcal{L}^q \subset \mathcal{L}^p$ ein Teilvektorraum.

Beweis. Die Funktion $h(x) := |x|^{\frac{q}{p}}$ ist für $1 \leq p \leq q < \infty$ konvex. Es folgt mit $Y := |X|^p$

$$(\mathbb{E}(|X|^p))^{\frac{q}{p}} = h(\mathbb{E}(Y)) \leq \mathbb{E}(h(Y)) = \mathbb{E}(|Y|^{\frac{q}{p}}) = \mathbb{E}(|X|^q)$$

und somit die Behauptung im Fall $q < \infty$. Für $1 \leq p < q = \infty$ folgt aus der Monotonie des Erwartungswertes

$$\|X\|_{\mathcal{L}^p} = (\mathbb{E}(|X|^p))^{\frac{1}{p}} \leq \|X\|_{\mathcal{L}^\infty}.$$

□

Lemma 2.32. *Sei $1 \leq p \leq \infty$. Dann ist $\mathcal{L}^p$ ein Vektorraum und $\|\cdot\|_{\mathcal{L}^p}$ eine Halbnorm auf $\mathcal{L}^p$. D.h. für $X, Y \in \mathcal{L}^p$ und $a \in \mathbb{R}$ gilt*

(a) $\|aX\|_{\mathcal{L}^p} = |a|\|X\|_{\mathcal{L}^p}$

(b) $\|X + Y\|_{\mathcal{L}^p} \leq \|X\|_{\mathcal{L}^p} + \|Y\|_{\mathcal{L}^p}$.

(c) Ist $\|X\|_{\mathcal{L}^p} = 0$ so ist $X = 0$ fast sicher.

Beweis. Eigenschaft (a) folgt aus der Linearität des Erwartungswerts. Eigenschaft (b) ist die Minkowski Ungleichung, siehe Maßtheorie oder Funktionalanalysis. Aus Teil (b) lässt sich insbesondere leicht zeigen, dass $\mathcal{L}^p$ ein Vektorraum ist. Wir zeigen (c).
$p < \infty$: In diesem Fall folgt aus $\|X\|_{\mathcal{L}^p} = 0$ auch $\mathbb{E}(|X|^p) = 0$. Damit folgt $|X|^p = 0$ fast sicher und somit $X = 0$ fast sicher.
$p = \infty$: In diesem Fall gibt es zu jedem $n \in \mathbb{N}$ ein $N_n \in \mathcal{F}$ mit $\mathbb{P}(N_n) = 0$ und

$$\sup_{\omega \in \Omega \setminus N_n} |X(\omega)| < \frac{1}{n}.$$

Es sei $N := \bigcup_{n \in \mathbb{N}} N_n$, dann ist $N \in \mathcal{F}$ und es gilt $\mathbb{P}(N) \leq \sum_{n=1}^{\infty} \mathbb{P}(N_n) = 0$. Ferner gilt für alle $\omega \in \Omega \setminus N$ bereits $|X(\omega)| < \frac{1}{n}, \quad \forall n \geq 1$ und damit $X = 0$ fast sicher. □

Also nächstes wollen wir aus $\mathcal{L}^p$ normierte Räume definieren. Es sei

$$\mathcal{N} := \{X \text{ Zufallsvariable mit } X = 0 \text{ fast sicher } \}.$$

Dann definiert

$$X \sim Y \Longleftrightarrow X - Y \in \mathcal{N}$$

eine Äquivalenzrelation auf $\mathcal{L}^p$ für alle $1 \leq p \leq \infty$ (Übung). Es sei $L^p := \mathcal{L}^p/\mathcal{N}$ der Raum der Äquivalenzklassen von Zufallsvariablen. Jedes $[X] \in L^p$ hat demnach einen Repräsentanten $X \in \mathcal{L}^p$ und es ist

$$[X] = \{Y \in \mathcal{L}^p \mid Y \sim X\} = \{Y \in \mathcal{L}^p \mid X = Y \text{ fast sicher }\}.$$

Warnung: Eine Äquivalenzklasse $[X]$ kann (im Allgemeinen) nicht punktweise ausgewertet werden. D.h. $[X](\omega)$ macht (im Allgemeinen) keinen Sinn.

Da der Erwartungswert invariant gegenüber Nullmengen ist, sind die folgenden Definitionen sinvoll

$$\mathbb{E}([X]) := \mathbb{E}(X)$$

und

$$\|[X]\|_{L^p} := \|X\|_{\mathcal{L}^p}.$$

Da keine Verwechslungsgefahr besteht, schreiben wir oft auch X anstelle von $[X]$, wo X ein Repräsentant von X ist.

Satz 2.33. *Sei $1 \leq p \leq \infty$. Dann ist $(L^p, \|\cdot\|_{L^p})$ ein Banachraum. Ist $p = 2$, so ist L^2 ein Hilbertraum mit Skalarprodukt*

$$\langle [X], [Y] \rangle_{L^2} := \mathbb{E}([X \cdot Y]).$$

Satz 2.34. *(Cauchy-Schwartz Ungleichung) Sind $X, Y \in \mathcal{L}^2$, so gilt $X \cdot Y \in \mathcal{L}^1$ und*

$$\mathbb{E}(|X \cdot Y|) \leq \sqrt{\mathbb{E}(|X|^2)}\sqrt{\mathbb{E}(|Y|^2)}.$$

Diesselbe Aussage gilt für L^2 und L^1 anstelle von $\mathcal{L}^2$ und $\mathcal{L}^1$.

Satz 2.35. *(Hölder Ungleichung) Seien $1 \leq p, q \leq \infty$ mit $\frac{1}{p} + \frac{1}{q} = 1$ (wo $\frac{1}{\infty} := 0$) und $X \in \mathcal{L}^p, Y \in \mathcal{L}^q$. Dann gilt $X \cdot Y \in \mathcal{L}^1$ und*

$$\|X \cdot Y\|_{\mathcal{L}^1} \leq \|X\|_{\mathcal{L}^p} \|Y\|_{\mathcal{L}^q}.$$

Diesselbe Aussage gilt für L^p, L^q, L^1 anstelle von $\mathcal{L}^p, \mathcal{L}^q, \mathcal{L}^1$.

2.4 Zufallsvariablen mit Werten in $\mathbb{R}^d$

Es sei $|a| := \sqrt{\sum_{k=1}^d a_k^2}$ die euklidische Norm von $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$. Eine messbare Abbildung $X : \Omega \longrightarrow \mathbb{R}^d$ heißt $\mathbb{R}^d$-werige Zufallsvariable. Häufig lassen wir den Zusatz $\mathbb{R}^d$-wertig weg, sofern dieses aus dem Zusammenhang hervorgeht. Eine solche Abbildung hat eine Darstellung in Komponenten über

$$X(\omega) = (X_1(\omega), \ldots, X_d(\omega)),$$

wo $X_j : \Omega \longrightarrow \mathbb{R}$ Abbildungen sind.

Bemerkung 2.36. *Sei* $X = (X_1, \dots, X_d) : \Omega \longrightarrow \mathbb{R}^d$ *eine Abbildung. Dann sind äquivalent:*

(a) X *ist eine* $\mathbb{R}^d$*-wertige Zufallsvariable.*

(b) $X_1, \dots, X_d$ *sind Zufallsvariablen.*

Lemma 2.37. *Es sei* $X = (X_1, \dots, X_d)$ *eine* $\mathbb{R}^d$*-wertige Zufallsvariable und* $p \in [1, \infty]$*. Dann gilt*

$$|X| \in \mathcal{L}^p \Longleftrightarrow X_k \in \mathcal{L}^p, \quad k = 1, \dots, d.$$

Beweis. Es reicht zu zeigen, dass

$$\int_\Omega |X|^p d\mathbb{P} < \infty \Longleftrightarrow \int_\Omega |X_k|^p d\mathbb{P} < \infty, \quad k = 1, \dots, d.$$

Es gilt $|X|^p = \left(\sum_{k=1}^d |X_k|^2\right)^{\frac{p}{2}}$. Daraus folgt

$$|X_k|^p = (|X_k|^2)^{\frac{p}{2}} \leq \left(\sum_{k=1}^d |X_k|^2\right)^{\frac{p}{2}} = |X|^p.$$

Andererseits gilt auch

$$|X|^p \leq \left(d\left(\max_{k=1,\dots,d} |X_k|\right)^2\right)^{\frac{p}{2}} = d^{\frac{p}{2}} \left(\max_{k=1,\dots,d} |X_k|\right)^p \leq d^{\frac{p}{2}} \sum_{k=1}^d |X_k|^p.$$

□

Für $p \in [1, \infty)$ sei $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) = \mathcal{L}^p(\Omega; \mathbb{R}^d)$ der Vektorraum der $\mathbb{R}^d$-wertigen Zufallsvariablen mit

$$\|X\|_{\mathcal{L}^p} := \mathbb{E}(|X|^p) < \infty.$$

Für $p = \infty$ sei $\mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) = \mathcal{L}^\infty(\Omega; \mathbb{R}^d)$ der Vektorraum der $\mathbb{R}^d$-wertigen Zufallsvariablen mit

$$\|X\|_{\mathcal{L}^\infty} := \inf_{N \in \mathcal{F},\ \mathbb{P}(N)=0} \sup_{\omega \in \Omega \setminus N} |X(\omega)| < \infty.$$

Analog zu dem eindimensionalen Fall lassen sich die Räume $L^p(\Omega; \mathbb{R}^d)$ definieren. Diese sind Banachräume.

Bemerkung 2.38. *Es sei* $X \in \mathcal{L}^2(\Omega; \mathbb{R}^d)$*. Dann gilt*

$$|X_k \cdot X_j| \leq \frac{|X_k|^2 + |X_j|^2}{2} \leq \frac{1}{2}\sqrt{\sum_{k=1}^d |X_k|^2} = \frac{1}{2}|X|^2 \in \mathcal{L}^1.$$

Also gilt $X_j \cdot X_k \in \mathcal{L}^1$ *für alle* $j, k = 1, \dots, d$*. Umgekehrt, gilt* $X_k \cdot X_j \in \mathcal{L}^1$ *für alle* $j, k = 1, \dots, d$*, so ist offensichtlich* $X \in \mathcal{L}^2(\Omega; \mathbb{R}^d)$*.*

Definition 2.39. *Sei $X = (X_1, \dots X_d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable mit $X \in \mathcal{L}^1(\Omega; \mathbb{R}^d)$. Dann ist der Erwartungswert definiert durch*

$$\mathbb{E}(X) := (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d)).$$

Ist $X \in \mathcal{L}^2(\Omega; \mathbb{R}^d)$ so ist die Kovarianzmatrix $\Sigma = (\sigma_{ij})_{i,j=1,\dots,d}$ ist definiert durch

$$\sigma_{ij} = \operatorname{cov}(X_i, X_j) := \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))).$$

Der nächste Satz zeigt dass die Eigenwerte von Σ reellwertig und nicht-negativ sind.

Satz 2.40. *Σ ist symmetrisch und es gilt*

$$\sum_{i,j=1}^{d} \overline{\lambda}_i \lambda_j \sigma_{ij} = \mathbb{E}\left(\left|\sum_{i=1}^{d} \lambda_i (X_i - \mathbb{E}(X_i))\right|^2\right) \geq 0$$

für alle $\lambda = (\lambda_1, \dots, \lambda_d) \in \mathbb{C}^d$.

Beweis. Übung. □

2.5 Konvergenz von Zufallsvariablen

Wir wollen einige Resultate aus der Integrationstheorie bezüglich der Konvergenz wiederholen. Im folgenden Betrachten wir Eigenschaften welche fast sicher gelten. Daher können wir uns ohne Einschränkung der Allgemeinheit auf Zufallsvariablen mit Werten in $\mathbb{R}$ zurückziehen.

Definition 2.41. *Seien $(X_n)_{n\in\mathbb{N}}$ und X Zufallsvariablen.*

(a) $(X_n)_{n\in\mathbb{N}}$ konvergiert fast sicher gegen X, falls es eine Nullmenge N gibt mit

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega \backslash N.$$

(b) $(X_n)_{n\in\mathbb{N}}$ konvergiert in Wahrscheinlichkeit gegen X, falls für alle $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0.$$

(c) $(X_n)_{n\in\mathbb{N}}$ konvergiert im p-ten Mittel (bzw. in $\mathcal{L}^p$) gegen X (wo $p \in (0, \infty]$), falls

$$\|X_n - X\|_{\mathcal{L}^p} \longrightarrow 0, \quad n \to \infty.$$

Bemerkung 2.42. *(a) Gilt $X_n \longrightarrow X$ in $\mathcal{L}^1$, so konvergiert auch die Folge der Erwartungswerte $(\mathbb{E}(X_n))_{n\in\mathbb{N}}$ gegen $\mathbb{E}(X)$. Die Umkehrung ist im Allgemeinen falsch.*

(b) Gilt $X_n \longrightarrow X$ in $\mathcal{L}^q$, so gilt auch $X_n \longrightarrow X$ in $\mathcal{L}^p$ für alle $1 \leq p \leq q \leq \infty$.

Der nächste Satz zeigt, dass unter der zusätzlichen Bedingung, dass die Folge $(X_n)_{n\in\mathbb{N}}$ eine Majorante besitzt, fast sichere Konvergenz die Konvergenz in $\mathcal{L}^1$ impliziert.

Satz 2.43. *(Satz von Lebesgue, Satz der dominierten Konvergenz) Sei* $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1$ *eine Folge von Zufallsvariablen und* $X \in \mathcal{L}^1$*. Es gebe* $Y \in \mathcal{L}^1$ *mit*

$$|X_n| \leq Y, \quad n \geq 1 \quad \text{fast sicher}$$

und $X_n \longrightarrow X$ *fast sicher. Dann gilt* $X_n \longrightarrow X$ *in* $\mathcal{L}^1$*.*

Korollar 2.44. *(Satz von Lebesgue für* $\mathcal{L}^p$*) Sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge von Zufallsvariblen mit* $X_n \longrightarrow X$ *fast sicher. Es gebe ein* $Y \in \mathcal{L}^p$ *mit* $p \in [1,\infty)$ *und* $|X_n| \leq Y$ *fast sicher. Dann gilt* $X_n \longrightarrow X$ *in* $\mathcal{L}^p$*.*

Beweis. Es gilt $|X_n - X|^p \longrightarrow 0$ fast sicher. Folglich gibt es $M \in \mathcal{F}$ mit $\mathbb{P}(M) = 1$ und $n_0 \in \mathbb{N}$ mit

$$|X(\omega)| \leq |X(\omega) - X_n(\omega)| + |X_n(\omega)| \leq 1 + Y(\omega), \quad \omega \in M, \quad n \geq n_0.$$

Somit gilt für $n \geq n_0$, $\omega \in M$ mit $(a+b)^p \leq 2^{p-1}(a^p + b^p)$

$$|X_n - X|^p \leq (|X_n| + |X|)^p \leq (1 + 2Y)^p \leq 2^{p-1}(1 + 2^p Y^p) \in \mathcal{L}^1.$$

Die Behauptung folgt aus dem Satz von Lebesgue. □

Satz 2.45. *(Satz von Beppo-Levi, Satz der monotonen Konvergenz) Sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge nicht-negativer Zufallsvariablen mit* $X_n \leq X_{n+1}$ *und* $X_n \nearrow X$ *fast sicher, wo* X *eine weitere Zufallsvariable ist. Dann gilt*

$$\mathbb{E}(X_n) \longrightarrow \mathbb{E}(X), \quad n \to \infty.$$

Korollar 2.46. *Sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge nicht-negativer Zufallsvariablen. Dann gilt*

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n). \tag{2.10}$$

Hierbei ist $\infty = \infty$ *zugelassen und die linke Seite ist endlich genau dann, wenn die rechte Seite endlich ist. Ist* $(X_n)_{n\in\mathbb{N}}$ *eine Folge reellwertiger Zufallsvariablen mit* $\sum_{n=1}^{\infty} \mathbb{E}(|X_n|) < \infty$ *so gilt* (2.10) *und beide Seiten sind endlich.*

Beweis. Übung. □

Lemma 2.47. *(Lemma von Fatou) Sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge nicht-negativer Zufallsvariablen. Dann gilt*

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

Satz 2.48. *Seien* $(X_n)_{n\in\mathbb{N}}$ *und* X, Y *Zufallsvariablen. Dann gelten die folgenden Aussagen:*

(a) Angenommen $X_n \longrightarrow X$ *fast sicher, dann gilt* $X_n \longrightarrow X$ *in Wahrscheinlichkeit.*

(b) Angenommen $X_n \longrightarrow X$ *in* $\mathcal{L}^p$ *für* $p \in [1,\infty]$*, dann gilt* $X_n \longrightarrow X$ *in Wahrscheinlichkeit.*

(c) Angenommen $X_n \longrightarrow X$ *und* $X_n \longrightarrow Y$ *in Wahrscheinlichkeit. Dann gilt* $X = Y$ *fast sicher.*

Beweis. (a) Sei $Y_n := \mathbb{1}_{|X_n - X| \geq \varepsilon}$ mit $\varepsilon > 0$ fest. Dann gilt $Y_n \leq 1$ und $Y_n \longrightarrow 0$ fast sicher. Aus dem Satz von Lebesgue folgt

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}(Y_n) \longrightarrow 0, \quad n \to \infty.$$

(b) Die Markovsche Ungleichung impliziert

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}(|X_n - X|^p) \longrightarrow 0, \quad n \to \infty$$

sofern $p \in [1, \infty)$. Für $p = \infty$ betrachten wir

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}(|X_n - X|) \leq \frac{1}{\varepsilon}\|X_n - X\|_{\mathcal{L}^\infty}.$$

(c) Für alle $\varepsilon > 0$ gilt

$$\begin{aligned}
\mathbb{P}(|X - Y| \geq \varepsilon) &\leq \mathbb{P}(|X_n - X| + |X_n - Y| \geq \varepsilon) \\
&\leq \mathbb{P}\left(\left\{|X_n - X| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|X_n - Y| \geq \frac{\varepsilon}{2}\right\}\right) \\
&\leq \mathbb{P}\left(|X_n - X| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - X| \geq \frac{\varepsilon}{2}\right) \longrightarrow 0, \quad n \to \infty.
\end{aligned}$$

□

Beispiel 2.49. *Alle anderen Implikationen sind im Allgemeinen falsch. Sei hierzu* $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), m(dx))$.

(a) Sei $X_n = n^{\frac{1}{p}}\mathbb{1}_{[0,\frac{1}{n}]}$. *Dann ist* $X_n \in \mathcal{L}^p$ *mit* $X_n \longrightarrow 0$ *fast sicher und damit auch in Wahrscheinlichkeit. Aber* $\mathbb{E}(|X_n|^p) = \frac{1}{n}(n^{\frac{1}{p}})^p = 1$ *zeigt, dass* $(X_n)_{n\in\mathbb{N}}$ *nicht in* $\mathcal{L}^p$ *konvergiert.*

(b) Sei

$$X_n := \mathbb{1}_{[k2^{-m},(k+1)2^{-m})}, \quad n = 2^m + k, \quad m \in \mathbb{N}_0,\ 0 \leq k < 2^m.$$

Dann gilt

$$\mathbb{P}(|X_n| \geq \varepsilon) \leq \mathbb{P}(|X_n| > 0) = 2^{-m} \longrightarrow 0, \quad n \to \infty$$

und

$$\mathbb{E}(|X_n|^p) = \mathbb{P}(|X_n| > 0) = 2^{-m}.$$

Folglich gilt $X_n \longrightarrow 0$ *in* $\mathcal{L}^p$ *und in Wahrscheinlichkeit. Wir zeigen, dass* $(X_n(\omega))_{n\in\mathbb{N}}$ *für kein* $\omega \in [0,1)$ *konvergiert. Sei* $\omega \in [0,1)$ *fixiert. Für jedes* $m \in \mathbb{N}$ *finden wir genau ein* $k = k(\omega, m)$ *mit* $\omega \in [k2^{-m}, (k+1)2^{-m})$. *Also gilt* $X_{k(\omega,m)+2^m}(\omega) = 0$. *Für alle anderen* $0 \leq k' < 2^m$ *ist* $X_{k'+2^m}(\omega) = 0$. *Damit kann* $(X_n(\omega))_{n\in\mathbb{N}}$ *nicht konvergieren.*

Im Folgenden betrachten wir die Konvergenz in Wahrscheinlihkeit und fast sichere Konvergenz genauer.

Lemma 2.50. *(Erstes Borel-Cantelli Lemma) Sei $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ mit $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Dann gilt*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 0.$$

Hierbei ist $\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$.

Beweis. Seien $B_n := \bigcup_{k\geq n} A_k$. Dann ist B_n eine fallende Folge (d.h. $B_{n+1} \subset B_n$ für alle $n \geq 1$) und es gilt $\bigcap_{n\geq 1} B_n = \limsup_{n\to\infty} A_n$, d.h. $B_n \searrow \limsup_{n\to\infty} A_n$. Es folgt

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

□

Satz 2.51. *Sei $(X_n)_{n\in\mathbb{N}}$ mit $X_n \longrightarrow X$ in Wahrscheinlichkeit. Dann gibt es eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $X_{n_k} \longrightarrow X$ fast sicher.*

Beweis. Da $X_n \longrightarrow X$ in Wahrscheinlichkeit, finden wir zu jedem $k \in \mathbb{N}$ ein $n_k \in \mathbb{N}$ mit

$$\mathbb{P}\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \frac{1}{k^2}.$$

Ohne Einschränkung der Allgemeinheit sei $n_k < n_{k+1}$ für alle k. Setze $M := \limsup_{k\to\infty} A_k$ mit $A_k := \left\{|X_{n_k} - X| \geq \frac{1}{k}\right\}$. Nach dem ersten Borel-Cantelli Lemma folgt $\mathbb{P}(M) = 0$. Für $\omega \in M^c = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k^c$, d.h. es gibt ein $n(\omega) \geq 1$ sodass für alle $k \geq n(\omega)$ gilt: $|X_{n_k}(\omega) - X(\omega)| < \frac{1}{k}$. Also gilt

$$\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega), \quad \omega \in M^c.$$

□

Satz 2.52. *(Vollständigkeit bei fast sicherer Konvergenz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit*

$$|X_n - X_m| \longrightarrow 0, \quad n, m \to \infty \quad \textit{fast sicher.} \tag{2.11}$$

Dann gibt es eine Zufallsvariable X mit $X_n \longrightarrow X$ fast sicher.

Beweis. Sei $M \in \mathcal{F}$ mit $\mathbb{P}(M) = 1$ derart, dass (2.11) für jedes $\omega \in M$ gilt. Da $\mathbb{R}$ vollständig ist, gibt es für jedes $\omega \in M$ einen Grenzwert $X(\omega) := \lim_{n\to\infty} X_n(\omega)$. Für $\omega \notin M$ setze $X(\omega) := 0$. Wir zeigen, dass X messbar ist, dann folgt $X_n \longrightarrow X$ fast sicher und somit die Behauptung.

Sei $X|_M : M \longrightarrow \mathbb{R}$, $\omega \longrightarrow X(\omega) = \lim\limits_{n\to\infty} X_n(\omega)$. Dann ist $X|_M$ als Grenzwert messbarer Abbildungen messbar bezüglich

$$\mathcal{F} \cap M := \{A \cap M \mid A \in \mathcal{F}\}.$$

Wegen $M \in \mathcal{F}$ gilt $\mathcal{F} \cap M \subset \mathcal{F}$ und damit ist $X|_M$ auch messbar bezüglich $\mathcal{F}$. Die Darstellung

$$X = \mathbb{1}_M X|_M + \mathbb{1}_{M^c} \cdot 0$$

zeigt, dass X messbar ist. □

Satz 2.53. *(Vollständigkit bei Konvergenz in Wahrscheinlichkeit) Sei* $(X_n)_{n\in\mathbb{N}}$ *mit*

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

Dann gibt es eine Zufallsvariable X *mit* $X_n \longrightarrow X$ *in Wahrscheinlichkeit.*

Beweis. Sei $(X_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge mit

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

Setze $A_k = \{|X_{n_k} - X_{n_{k+1}}| \geq \frac{1}{k^2}\}$ und $M := \limsup_{n\to\infty} A_n$. Nach dem ersten Borel-Cantelli Lemma gilt $\mathbb{P}(M) = 0$, also $\mathbb{P}(M^c) = 1$. Für jedes $\omega \in M^c = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k^c$ gibt es ein $n(\omega) \geq 1$ sodass für alle $k \geq n(\omega)$

$$|X_{n_k}(\omega) - X_{n_{k+1}}(\omega)| < \frac{1}{k^2}.$$

Damit ist $(X_{n_k}(\omega))_{k\in\mathbb{N}}$ eine Cauchy Folge in $\mathbb{R}$. Sei $X(\omega) := \lim_{k\to\infty} X_{n_k}(\omega)$ für alle $\omega \in M^c$. Für $\omega \in M$ setze $X(\omega) := 0$. Dann ist X eine Zufallsvariable mit $X_{n_k} \longrightarrow X$ fast sicher. Es bleibt zu zeigen, dass $X_n \longrightarrow X$ in Wahrscheinlichkeit konvergiert. Sei $n \in \mathbb{N}$ und $n_k \geq n$, dann ist

$$\begin{aligned}
\mathbb{P}(|X_n - X| \geq \varepsilon) &\leq \mathbb{P}(|X_n - X_{n_k}| + |X_{n_k} - X| \geq \varepsilon) \\
&\leq \mathbb{P}\left(\left\{|X_n - X_{n_k}| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|X_{n_k} - X| \geq \frac{\varepsilon}{2}\right\}\right) \\
&\leq \mathbb{P}\left(|X_{n_k} - X| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - X_{n_k}| \geq \frac{\varepsilon}{2}\right).
\end{aligned}$$

Mit $n, k \to \infty$ folgt die Behauptung. □

2.6 Stochastische Unabhängigkeit

Definition 2.54. *Sei* I *eine beliebige Indexmenge und* $\{A_i \mid i \in I\} \subset \mathcal{F}$ *eine Familie von Ereignissen.*

(a) Die Familie heißt unabhängig, wenn

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j)$$

für alle endlichen Teilmengen $J \subset I$ *gilt.*

(b) Die Familie heißt paarweise unabhängig, wenn

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j), \quad \forall i, j \in I, \quad i \neq j.$$

Bemerkung 2.55. *Unabhängigkeit impliziert paarweise unabhängigkeit. Die Umkehrung ist im Allgemeinen falsch.*

Beispiel 2.56. *(a) Sei $\Omega = \{1,2,3,4\}$ mit $\mathcal{F} = 2^\Omega$ und Gleichverteilung $\mathbb{P}$ auf Ω. Seien*

$$A = \{1,2\}, \quad B = \{2,3\}, \quad C = \{1,3\}.$$

Dann gilt

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$$

und

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}.$$

D.h. die Ereignisse sind paarweise unabhängig. Die Ereignisse sind nicht unabhängig, da

$$\mathbb{P}(A \cap B \cap C) = 0 \neq \left(\frac{1}{2}\right)^3 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

(b) Sei $I = \{1,2\}$ mit Mengen A_1, A_2. Dann ist Unabhängigkeit äquivalent zu

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

(c) Sei $I = \{1,2,3\}$ mit Mengen A_1, A_2, A_3. Dann ist Unabhängigkeit äquivalent zu $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$, $\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_3)$, $\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2)\mathbb{P}(A_3)$ und

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

Definition 2.57. *Sei I eine Indexmenge und für $i \in I$ seien Mengensysteme $\mathcal{E}_i \subset \mathcal{F}$ gegeben. Die Familie $\{\mathcal{E}_i \mid i \in I\}$ heißt unabhängig, wenn für jede Wahl $A_i \in \mathcal{E}_i$ die Familie $\{A_i \mid i \in I\}$ von Ereignissen unabhängig ist.*
Äquivalent: *Für jede endliche Teilmenge $J \subset I$ und jede Wahl $A_j \in \mathcal{E}_j$ mit $j \in J$ gilt*

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Bemerkung 2.58.

(a) Seien $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ mit $i = 1, \dots, n$ Wahrscheinlichkeitsräume. Setze $\Omega = \Omega_1 \times \cdots \times \Omega_n$ mit $\mathcal{F} = \bigotimes_{i=1}^n \mathcal{F}_i$ und $\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i$. Seien Mengen A_i gegeben durch

$$\begin{aligned} &A_1 := B_1 \times \Omega_2 \times \cdots \times \Omega_n, \ B_1 \in \mathcal{F}_1 \\ &\cdots \\ &A_n := \Omega_1 \times \cdots \times B_n, \ B_n \in \mathcal{F}_n. \end{aligned}$$

Dann ist die Familie $\{A_1, \dots, A_n\}$ unabhängig unter $\mathbb{P}$.

(b) $\{\mathcal{E}_i \mid i \in I\}$ ist unabhängig genau dann, wenn $\{\mathcal{E}_i \mid i \in J\}$ für alle endlichen Teilmengen $J \subset I$ unabhängig ist.

(c) *Ist $\mathcal{A}_i \subset \mathcal{E}_i$ für alle $i \in I$, so gilt*

$$\{\mathcal{E}_i \mid i \in I\} \text{ ist unabhängig } \Longrightarrow \{\mathcal{A}_i \mid i \in I\} \text{ ist unabhängig.}$$

Die Umkehrung ist im Allgemeinen falsch.

Satz 2.59. *Sei $\{\mathcal{E}_i \mid i \in I\}$ eine unabhängige Familie. Dann ist auch die Familie der erzeugten Dynkin-Systeme $\{d(\mathcal{E}_i) \mid i \in I\}$ unabhängig.*

Beweis. Wegen Bemerkung 2.58 sei ohne Einschränkung I endlich mit $I = \{1, \ldots, n\}$. Für $i_0 \in I$ sei

$$\mathcal{D}_{i_0} := \{A \in \mathcal{F} \mid \{\mathcal{E}_1, \ldots, \mathcal{E}_{i_0-1}, \{A\}, \mathcal{E}_{i_0+1}, \ldots, \mathcal{E}_n\} \text{ ist unabhängige Familie}\}.$$

Wir zeigen, dass $\mathcal{D}_{i_0}$ ein Dynkin-System ist. Hierfür sei $\{i_1, \ldots, i_k\} \subset I \backslash \{i_0\}$ beliebig und wähle $A_{i_\nu} \in \mathcal{E}_{i_\nu}$, wo $\nu = 1, \ldots, k$.

1. Da die kleinere Auswahl von Mengen auch unabhängig ist, folgt

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k} \cap \Omega) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}) \cdot 1.$$

   Also ist $\Omega \in \mathcal{D}_{i_0}$.

2. Sei $A \in \mathcal{D}_{i_0}$. Dann ist

$$\begin{aligned}\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k} \cap A^c) &= \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k} \cap \Omega) - \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k} \cap A) \\ &= \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})(1 - \mathbb{P}(A)) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})\mathbb{P}(A^c)\end{aligned}$$

   und folglich $A^c \in \mathcal{D}_{i_0}$.

3. Seien $B_l \in \mathcal{D}_{i_0}$ mit $l \in \mathbb{N}$ disjunkt. Dann ist

$$\begin{aligned}\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_k} \cap \bigcup_{l=1}^{\infty} B_l\right) &= \mathbb{P}\left(\bigcup_{l=1}^{\infty} A_{i_1} \cap \cdots \cap A_{i_k} \cap B_l\right) \\ &= \sum_{l=1}^{\infty} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k} \cap B_l) = \sum_{l=1}^{\infty} \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}) \cdot \mathbb{P}(B_l) \\ &= \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})\mathbb{P}\left(\bigcup_{l=1}^{\infty} B_l\right)\end{aligned}$$

   und folglich $\bigcup_{l=1}^{\infty} B_l \in \mathcal{D}_{i_0}$.

Damit ist $\mathcal{D}_{i_0}$ ein Dynkin System und $\{\mathcal{E}_1, \ldots, \mathcal{E}_{i_0-1}, \mathcal{D}_{i_0}, \mathcal{E}_{i_0+1}, \ldots, \mathcal{E}_n\}$ ist eine unabhängige Familie. Es gilt $\mathcal{E}_{i_0} \subset \mathcal{D}_{i_0}$ und da $\mathcal{D}_{i_0}$ ein Dynkin System ist, folgt $\mathcal{E}_{i_0} \subset d(\mathcal{E}_{i_0}) \subset \mathcal{D}_{i_0}$. Nach Bemerkung 2.58 ist $\{\mathcal{E}_1, \ldots, \mathcal{E}_{i_0-1}, d(\mathcal{E}_{i_0}), \mathcal{E}_{i_0+1}, \ldots, \mathcal{E}_n\}$ eine unabhängige Familie. Da $i_0 \in I$ beliebig war, folgt die Behauptung durch Induktion. □

Korollar 2.60. *Sei $\{\mathcal{E}_i \mid i \in I\}$ eine unabhängige Familie von $\cap$-stabilen Mengensystemen mit $\mathcal{E}_i \subset \mathcal{F}$. Dann ist $\{\sigma(\mathcal{E}_i) \mid i \in I\}$ eine unabhängige Familie von σ-Algebren.*

Korollar 2.61. *Sei $\{A_i \mid i \in I\}$ unabhängig mit $A_i \in \mathcal{F}$. Für jedes $i \in I$ seien $A_i^* \in \{\emptyset, A_i, A_i^c, \Omega\}$. Dann ist $\{A_i^* \mid i \in I\}$ auch unabhängig.*

Beweis. Sei $\mathcal{E}_i = \{A_i\}$, dann ist nach Voraussetzung $\{\mathcal{E}_i \mid i \in I\}$ eine unabhängige Familie von $\cap$-stabilen Mengensystemen. Nach Korollar 2.60 ist auch $\{\sigma(\mathcal{E}_i) \mid i \in I\}$ unabhängig. Die Behauptung folgt aus $\sigma(\mathcal{E}_i) = \{\emptyset, A_i, A_i^c, \Omega\}$. $\square$

Korollar 2.62. *(Gruppieren) Sei $\{\mathcal{E}_i \mid i \in I\}$ unabhängig und $\cap$-stabil mit $\mathcal{E}_i \subset \mathcal{F}$. Sei $I = \bigcup\limits_{j \in J} I_j$ eine disjunkte Zerlegung und $\mathcal{A}_j = \sigma\left(\bigcup\limits_{i \in I_j} \mathcal{E}_j\right)$. Dann ist $\{\mathcal{A}_j \mid j \in J\}$ eine unabhängige Familie.*

Beweis. Für $k \in J$ sei

$$\mathcal{D}_k = \{A_{i_1} \cap \cdots \cap A_{i_n} \mid A_{i_l} \in \mathcal{E}_{i_l}, \quad i_1, \dots, i_n \in I_k, \ n \in \mathbb{N}\}$$

die Menge aller endlichen Schnitte von Mengen aus $\bigcup\limits_{i \in I_k} \mathcal{E}_i$. Dann ist $\mathcal{D}_k$ per Definition $\cap$-stabil. Weiterhin ist $\{\mathcal{D}_k \mid k \in J\}$ unabhängig, da die $\mathcal{E}_i$ unabhängig sind. Wegen

$$\mathcal{A}_k = \sigma\left(\bigcup_{i \in I_k} \mathcal{E}_i\right) = \sigma(\mathcal{D}_k), \quad k \in J$$

und Korollar 2.60 ist $\{\mathcal{A}_k \mid k \in J\}$ unabhängig. $\square$

Satz 2.63. *(Kolmogorovsches 0-1-Gesetz) Sei $(\mathcal{F}_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger σ-Algebren mit $\mathcal{F}_n \subset \mathcal{F}$. Sei*

$$\mathcal{A}_n := \bigvee_{k \geq n} \mathcal{F}_k := \sigma\left(\bigcup_{k \geq n} \mathcal{F}_k\right)$$

und $\mathcal{T}_\infty := \bigcap\limits_{n \geq 1} \mathcal{A}_n$ die terminale σ-Algebra. Dann gilt

$$A \in \mathcal{T}_\infty \Longrightarrow \mathbb{P}(A) \in \{0, 1\}.$$

Notation: $\mathcal{E}_1 \perp \mathcal{E}_2 \Longleftrightarrow \{\mathcal{E}_1, \mathcal{E}_2\}$ unabhängige Familie.

Beweis. **Idee:** $A \in \mathcal{T}_\infty$, dann ist A unabhängig von A, d.h. $\mathbb{P}(A)^2 = \mathbb{P}(A)$.
Nach Korollar 2.60 und Korollar 2.62 gilt für $n \in \mathbb{N}$

$$\mathcal{A}_{n+1} \perp \bigvee_{k=1}^{n} \mathcal{F}_k.$$

Insbesondere wegen $\mathcal{A}_{n+1} \subset \mathcal{A}_{n+2}$ folgt $\mathcal{T}_\infty = \bigcap\limits_{k \geq n+1} \mathcal{A}_k \perp \bigvee\limits_{k=1}^{n} \mathcal{F}_k$. Daraus folgt

$$\mathcal{T}_\infty \perp \bigcup_{n \geq 1} \bigvee_{k=1}^{n} \mathcal{F}_k,$$

da zu jeder Menge aus $\bigcup_{n\geq 1}\bigvee_{k=1}^{n}\mathcal{F}_k$ es ein $n \in \mathbb{N}$ gibt, sodasss diese in $\bigvee_{k=1}^{n}\mathcal{F}_k$ liegt. Korollar 2.60 impliziert

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

Es ist $\mathcal{T}_\infty \subset \mathcal{A}_k \subset \mathcal{A}_1 = \bigvee_{k=1}^{\infty}\mathcal{F}_k$ und somit $\mathcal{T}_\infty \perp \mathcal{T}_\infty$. $\square$

Satz 2.64. *Sei $\mathcal{T} \subset \mathcal{F}$ eine σ-Algebra mit $\mathbb{P}(A) \in \{0,1\}$ für alle $A \in \mathcal{T}$. Sei $Z : \Omega \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ eine $\mathcal{T}$ messbare Zufallsvariable. Dann gibt es ein $c \in \mathbb{R} \cup \{\pm\infty\}$ mit $\mathbb{P}(Z = c) = 1$.*

Beweis. Nach Voraussetzung ist $\{Z \leq t\} \in \mathcal{T}$ für alle $t \in \mathbb{R}$. Also ist $F(t) := \mathbb{P}(Z \leq t)$ monoton wachsend mit $F(t) \in \{0,1\}$ für alle t.

Fall 1. $F(t) = 0$ für alle $t \in \mathbb{R}$. Dann ist $\mathbb{P}(Z > n) = 1$ für alle $n \in \mathbb{N}$ und somit

$$\mathbb{P}(Z = +\infty) = \lim_{n\to\infty}\mathbb{P}(Z > n) = 1.$$

Also $c = +\infty$.

Fall 2. $F(t) = 1$ für alle $t \in \mathbb{R}$. Dann ist

$$\mathbb{P}(Z = -\infty) = \lim_{n\to\infty}\mathbb{P}(Z \leq -n) = 1.$$

Also $c = -\infty$.

Fall 3. Es gibt ein $t_0 \in \mathbb{R}$ mit

$$F(t) = \begin{cases} 0, & t < t_0 \\ 1, & t \geq t_0 \end{cases}.$$

Dann ist

$$1 = F\left(t_0 + \frac{1}{n}\right) - F\left(t_0 - \frac{1}{n}\right) = \mathbb{P}\left(t_0 - \frac{1}{n} < Z \leq t_0 + \frac{1}{n}\right), \quad n \in \mathbb{N}.$$

Also gilt

$$\mathbb{P}(Z = t_0) = \lim_{n\to\infty}\mathbb{P}\left(t_0 - \frac{1}{n} < Z \leq t_0 + \frac{1}{n}\right) = 1.$$

$\square$

Definition 2.65. *Sei I eine Indexmenge und seien $X_i : (\Omega, \mathcal{F}) \longrightarrow (E_i, \mathcal{E}_i)$ messbar, wo $(E_i, \mathcal{E}_i)$ Massräume sind für alle $i \in I$. Die Familie $\{X_i \mid i \in I\}$ heißt unabhängig, falls die σ-Algebren*

$$\sigma(X_i) := X_i^{-1}(\mathcal{E}_i) = \{X_i^{-1}(A) \mid A \in \mathcal{E}_i\}$$

eine unabhängige Familie $\{\sigma(X_i) \mid i \in I\}$ bilden.

Bemerkung 2.66. *(a) Falls X_i messbar bezüglich $\mathcal{D}_i$-$\mathcal{E}_i$ sind und $\{\mathcal{D}_i \mid i \in I\}$ unabhängig sind, so ist $\{X_i \mid i \in I\}$ unabhängig.*

(b) Seien $\{X_i \mid i \in I\}$ unabhängig und $\varphi_i : E_i \longrightarrow W_i$ messbar. Dann ist $\{\varphi_i \circ X_i \mid i \in I\}$ unabhängig.

(c) $\{X_i \mid i \in I\}$ ist eine Familie von unabhängigen Zufallsvariablen genau dann, wenn für alle $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ paarweise verschieden und $A_{i_1} \in \mathcal{E}_{i_1}, \ldots, A_{i_n} \in \mathcal{E}_{i_n}$ gilt

$$\mathbb{P}(X_{i_1} \in A_{i_1}, \ldots, X_{i_n} \in A_{i_n}) = \prod_{j=1}^{n} \mathbb{P}(X_{i_j} \in A_{i_j}). \tag{2.12}$$

Wegen $\mathbb{P}(X_{i_1} \in A_{i_1}, \ldots, X_{i_n} \in A_{i_n}) = \mathbb{P} \circ (X_{i_n}, \ldots, X_{i_n})^{-1}(A_{i_1} \times \cdots \times A_{i_n})$ ist folglich $\{X_i \mid i \in I\}$ genau dann unabhängig, wenn alle endlich dimensionalen Verteilungen $\mathbb{P} \circ (X_{i_n}, \ldots, X_{i_n})^{-1}$ Produktmaße sind, d.h.

$$\mathbb{P} \circ (X_{i_n}, \ldots, X_{i_n})^{-1} = \mathbb{P} \circ X_{i_1}^{-1} \otimes \cdots \otimes \mathbb{P} \circ X_{i_n}^{-1}.$$

Satz 2.67. *Seien X_i Zufallsvariablen. Genau dann ist $\{X_i \mid i \in I\}$ unabhängig, wenn für alle $n \in \mathbb{N}$, alle paarweise verschiedenen $i_1, \ldots, i_n \in I$ und alle $t_1, \ldots, t_n \in \mathbb{R}$*

$$\mathbb{P}(X_{i_1} \le t_1, \ldots, X_{i_n} \le t_n) = \prod_{j=1}^{n} \mathbb{P}(X_{i_j} \le t_j). \tag{2.13}$$

Beweis. Ist $\{X_i \mid i \in I\}$ unabhängig, so folgt (2.13) direkt aus der Definition. Es gelte also (2.13). Sei $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ paarweise verschieden. Es reicht (2.12) zu zeigen. Diese gilt wegen (2.13) für alle Mengen der Form $A_{i_k} = [-\infty, t_k]$ mit $t_k \in \mathbb{R}$ und $k = 1, \ldots, n$. Da Mengen dieser Form ein $\cap$-stabiles Erzeugendensystem bilden, folgt die Behauptung. □

Lemma 2.68. *Seien $X = \sum_{n=1}^{N} a_n \mathbb{1}_{A_n}$ sowie $Y = \sum_{m=1}^{M} b_n \mathbb{1}_{B_n}$ mit $a_n, b_n \ge 0$ und $A_n, B_m \in \mathcal{F}$ derart dass die a_n sowie b_m paarweise verschieden sind und A_n sowie B_n disjunkt. Dann sind äquivalent*

- *X, Y sind unabhängig.*
- *$\mathbb{P}(A_n \cap B_m) = \mathbb{P}(A_n)\mathbb{P}(B_m)$ für alle $1 \le n \le N$ sowie $1 \le m \le M$.*

Beweis. Übung. □

Lemma 2.69. *Seien X, Y unabhängige Zufallsvariablen. Sei $X = X^+ - X^-$ und $Y = Y^+ - Y^-$ die Zerlegung in Positiv und Negativteil. Dann sind unabhängig:*

- *X^+, Y^+.*
- *X^-, Y^-.*
- *X^+, Y^-.*
- *X^-, Y^+.*

Beweis. Wegen $X^+ = \max\{0, X\}$ und $X^- = \max\{0, -X\}$ sind $X^\pm$ messbar ist bezüglich $\sigma(X)$. Also gilt $\sigma(X^\pm) \subset \sigma(X)$. Analog sehen wir $\sigma(Y^\pm) \subset \sigma(Y)$. Da $\sigma(X) \perp \sigma(Y)$ folgt die Behauptung. $\square$

Satz 2.70. *Seien $X, Y \in \mathcal{L}^1$.*

(a) Sind X, Y unabhängig, so gilt $X \cdot Y \in \mathcal{L}^1$ und $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

(b) Sind X, Y unabhängig mit $X, Y \in \mathcal{L}^2$, so gilt $\operatorname{cov}(X, Y) = 0$ und

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y).$$

(c) X, Y sind genau dann unabhängig, wenn für alle messbaren, beschränkten Funktionen $f, g : \overline{\mathbb{R}} \longrightarrow \mathbb{C}$ gilt

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)). \tag{2.14}$$

Beweis. Teil (b) folgt direkt aus (a). Wir zeigen zunächst (a). Wir zeigen die Behauptung für $X, Y \geq 0$. Der allgemeine Fall ist eine Konsequenz aus Lemma 2.69 und der Zerlegung in Positiv- und Negativteil.
Fall 1. $X = \mathbb{1}_A$ und $Y = \mathbb{1}_B$. Dann gilt $0 \leq X \cdot Y \leq 1$, also $X \cdot Y \in \mathcal{L}^1$. Ferner gilt

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(\mathbb{1}_{A \cap B}) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}(X)\mathbb{E}(Y).$$

Fall 2. $X = \sum_{n=1}^{N} a_n \mathbb{1}_{A_n}$ und $Y = \sum_{m=1}^{M} b_m \mathbb{1}_{B_m}$ $a_n, b_n \geq 0$ und $A_n, B_m \in \mathcal{F}$ derart dass die a_n sowie b_m paarweise verschieden sind und A_n sowie B_n disjunkt. Dann gilt

$$X \cdot Y = \sum_{n=1}^{N} \sum_{m=1}^{M} a_n b_m \mathbb{1}_{A_n \cap B_m} \in \mathcal{L}^1.$$

Ferner gilt

$$\begin{aligned} \mathbb{E}(X \cdot Y) &= \sum_{n=1}^{N} \sum_{m=1}^{M} a_n b_m \mathbb{P}(A_n \cap B_m) = \sum_{n=1}^{N} \sum_{m=1}^{M} a_n b_m \mathbb{P}(A_n)\mathbb{P}(B_m) \\ &= \left(\sum_{n=1}^{N} a_n \mathbb{P}(A_n) \right) \left(\sum_{m=1}^{M} b_m \mathbb{P}(B_m) \right) = \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

Fall 3. Seien $0 \leq X, Y \in \mathcal{L}^1$, X_n, Y_n Folgen von Elementarfunktionen wie in Fall 2 mit $X_n \nearrow X$ und $Y_n \nearrow Y$. Dann gilt $X_n \cdot Y_n \in \mathcal{L}^1$, $X_n \cdot Y_n \nearrow X \cdot Y$ und $\mathbb{E}(X_n \cdot Y_n) = \mathbb{E}(X_n)\mathbb{E}(Y_n)$. Aus dem Satz der monotonen Konvergenz folgt

$$\mathbb{E}(X \cdot Y) = \lim_{n \to \infty} \mathbb{E}(X_n \cdot Y_n) = \lim_{n \to \infty} \mathbb{E}(X_n)\mathbb{E}(Y_n) = \mathbb{E}(X)\mathbb{E}(Y) < \infty.$$

Fall 4. Sind $X, Y \in \mathcal{L}^1$ unabhängig, so zerlegen wir $X = X^+ - X^-$ sowie $Y = Y^+ - Y^-$. Dann gilt

$$X \cdot Y = X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+$$

und wegen $X^+Y^+ + X^-Y^- + X^+Y^- + X^-Y^+ = |X \cdot Y| \in \mathcal{L}^1$ zusammen mit Schritt 3 sind alle Terme integrierbar. Die Behauptung folgt in diesem Fall aus einer einfachen Rechnung.

Es bleibt (c) zu zeigen. Sind X, Y unabhängig, so sind auch $f(X), g(Y)$ unabhängig und folglich gilt (2.14). Umgekehrt gelte (2.14). Seien $A \in \sigma(X)$ und $B \in \sigma(Y)$, dann gibt es A' und B' mit $A = X^{-1}(A')$ und $B = X^{-1}(B')$. Es gilt $\mathbb{1}_{A'}(X) = \mathbb{1}_A$ und $\mathbb{1}_{B'}(Y) = \mathbb{1}_B$ und folglich

$$\mathbb{P}(A \cap B) = \mathbb{E}(\mathbb{1}_{A'}(X)\mathbb{1}_{B'}(Y)) = \mathbb{E}(\mathbb{1}_{A'}(X))\mathbb{E}(\mathbb{1}_{B'}(Y)) = \mathbb{P}(A)\mathbb{P}(B).$$

□

2.7 Bedingte Erwartungswerte

Im folgenden betrachten wir stets $\mathbb{R}$-wertige Zufallsvariablen. Für $\mathbb{R}^d$-wertige Zufallsvariablen lassen sich alle Resultate durch komponentenweise Anwendung übertragen.

Definition 2.71. *Seien $A, B \in \mathcal{F}$ und $\mathbb{P}(B) > 0$. Die bedingte Wahrscheinlichkeit von A gegeben B ist definiert über*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Bemerkung 2.72. *$A \longmapsto \mathbb{P}(A|B)$ ist für festes B mit $\mathbb{P}(B) > 0$ ein Wahrscheinlichkeitsmaß. Für eine Zufallsvariable $X \in \mathcal{L}^1$ ist der bedingte Erwartungswert von X gegeben B definiert über*

$$\mathbb{E}(X|B) := \int_\Omega X(\omega)\mathbb{P}(d\omega|B).$$

Satz 2.73. *(Formel der totalen Wahrscheinlichkeit) Sei $(B_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ eine Folge disjunkter Mengen mit $\bigcup_{n\geq 1} B_n = \Omega$ und $\mathbb{P}(B_n) > 0$ für alle $n \geq 1$. Dann gilt für alle $A \in \mathcal{F}$*

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

Beweis. Es gilt

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n\geq 1} A \cap B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A \cap B_n) = \sum_{n=1}^{\infty} \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

□

Definition 2.74. *Sei X eine Zufallsvariable mit höchstens abzählbar vielen Werten $\{x_k\}_{k\in\mathbb{N}}$. Definiere*

$$\mathbb{P}(A|X = x_k) = \begin{cases} \frac{\mathbb{P}(A \cap \{X = x_k\})}{\mathbb{P}(X = x_k)}, & \mathbb{P}(X = x_k) > 0 \\ 0, & \text{sonst} \end{cases}.$$

Der bedingte Erwartungswert einer weiteren Zufallsvariable Y gegeben $X = x_k$ ist in diesem Fall definiert als

$$\mathbb{E}(Y|X = x_k) = \int_\Omega Y(\omega)\mathbb{P}(d\omega|X = x_k),$$

sofern Y bezüglich $\mathbb{P}(\cdot|X = x_k)$ integrierbar ist.

Wir wollen diese Definition auf Zufallsvariablen mit kontinuierlichem Bild erweitern. Ist $\mathbb{P}(X \in B) > 0$, so ist

$$\mathbb{P}(A|X \in B) = \frac{\mathbb{P}(A \cap \{X \in B\})}{\mathbb{P}(X \in B)}$$

wohldefiniert. Für $B = \{x\}$ mit $x \in \mathbb{R}^d$ ist jedoch in vielen Fällen $\mathbb{P}(X = x) = 0$. Die Idee ist

$$\mathbb{P}(A|X = x) = \lim_{h \to 0} \frac{\mathbb{P}(A \cap \{X \in (x-h, x+h)\})}{\mathbb{P}(X \in (x-h, x+h))}.$$

Problem: Der Limes muss nicht existieren!
Lösung: Wir benutzen den Satz von Radon-Nikodym.

Definition 2.75. *Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra und $X \in \mathcal{L}^1$. Der bedingte Erwartungswert von X gegeben $\mathcal{A}$ wird mit $\mathbb{E}(X|\mathcal{A})$ bezeichnet. Dieser ist definiert über die folgenden Eigenschaften*

- $\mathbb{E}(X|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$.
- *Für alle $A \in \mathcal{A}$ gilt*

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X|\mathcal{A})) \tag{2.15}$$

Kurz: Der bedingte Erwartungswert ist die bezüglich $\mathcal{A}$ messbare Zufallsvariable mit (2.15).

Satz 2.76. *(Existenz und Eindeutigkeit vom bedingten Erwartungswert) Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra und $X \in \mathcal{L}^1$. Dann gibt es den bedingten Erwartungswert $\mathbb{E}(X|\mathcal{A})$. Dieser ist eindeutig in dem folgenden Sinn:*

Ist $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ gegeben mit

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot g), \quad A \in \mathcal{A},$$

so gilt $g = \mathbb{E}(X|\mathcal{A})$ fast sicher.

Beweis. Wir zeigen zunächst die Existenz.
Fall 1: Sei $X \geq 0$ fast sicher. Dann definiert

$$\mathcal{A} \ni A \longmapsto \int_A X d\mathbb{P} = \mathbb{E}(\mathbb{1}_A \cdot X) \geq 0$$

ein Maß auf $\mathcal{A}$. Ist A derart, dass $\mathbb{P}(A) = 0$. So gilt auch $\mathbb{E}(\mathbb{1}_A \cdot X) = 0$. Der Satz von Radon-Nikodym liefert die Existenz einer Zufallsvariable $\mathbb{E}(X|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ mit

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot \mathbb{E}(X|\mathcal{A})), \quad A \in \mathcal{A}.$$

Dieses zeigt die Existenz in diesem Fall.
Fall 2: Sei $X \in \mathcal{L}^1$ beliebig. Betrachte die Zerlegung $X = X^+ - X^-$ in Positiv- und Negativteil. Dann sind $X^\pm \in \mathcal{L}^1$ nicht-negativ. Aus Fall 1 folgt die Existenz vom bedingten Erwartungswert

$\mathbb{E}(X^{\pm}|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Setze $\mathbb{E}(X|\mathcal{A}) := \mathbb{E}(X^+|\mathcal{A}) - \mathbb{E}(X^-|\mathcal{A})$. Dann gilt $\mathbb{E}(X|\mathcal{A}) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ und für alle $A \in \mathcal{A}$

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot X^+) - \mathbb{E}(\mathbb{1}_A \cdot X^-) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X^+|\mathcal{A})) - \mathbb{E}(\mathbb{1}_A \cdot \mathbb{E}(X^-|\mathcal{A})) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X|\mathcal{A})).$$

Dieses zeigt die Existenz im allgemeinen Fall. Wir zeigen die Eindeutigkeit. Seien dazu $g_1, g_2 \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ gegeben mit

$$\mathbb{E}(\mathbb{1}_A \cdot g_1) = \mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot g_2), \quad A \in \mathcal{A}. \tag{2.16}$$

Dann ist $A_1 := \{g_1 - g_2 > 0\} \in \mathcal{A}$ und $A_2 := \{g_2 - g_1 > 0\} \in \mathcal{A}$. Aus (2.16) folgt $\mathbb{E}(\mathbb{1}_{A_j} \cdot g_1) = \mathbb{E}(\mathbb{1}_{A_j} \cdot g_2)$ mit $j = 1, 2$. Daraus folgt

$$\mathbb{E}(\mathbb{1}_{A_1}(g_1 - g_2)) = 0 = \mathbb{E}(\mathbb{1}_{A_2}(g_2 - g_1)).$$

Da die Integranden nicht-negativ sind, sind diese fast sicher Null. Wegen der Wahl der Mengen A_j ist $\mathbb{1}_{A_1} = 0 = \mathbb{1}_{A_2}$ fast sicher, d.h. $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 0$. Insbesondere ist

$$\mathbb{P}(g_1 \neq g_2) = \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = 0.$$

□

Bemerkung 2.77. *Die Eindeutigkeit folgt aus dem Satz von Radon-Nikodym und hätte daher nicht gesondert gezeigt werden müssen.*

Das nächste Lemma sammelt einige wichtige Eigenschaften des bedingten Erwartungswertes.

Lemma 2.78. *Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra und $X \in \mathcal{L}^1$. Es gelten die folgenden Eigenschaften:*

(a) Ist X messbar bezüglich $\mathcal{A}$, so gilt $\mathbb{E}(X|\mathcal{A}) = X$.

(b) Ist X unabhängig von $\mathcal{A}$, so gilt $\mathbb{E}(X|\mathcal{A}) = \mathbb{E}(X)$.

(c) Es gilt $\mathbb{E}(X|\{\varnothing, \Omega\}) = \mathbb{E}(X)$.

(d) Für $A \in \mathcal{F}$ mit $\mathbb{P}(A) \in (0,1)$ sei $\mathcal{G} = \{A, A^c, \varnothing, \Omega\}$. Dann gilt

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X|A)\mathbb{1}_A + \mathbb{E}(X|A^c)\mathbb{1}_{A^c}.$$

(e) Ist $Y \in \mathcal{L}^1$ und $a, b \in \mathbb{R}$ so gilt

$$\mathbb{E}(aX + bY|\mathcal{A}) = a\mathbb{E}(X|\mathcal{A}) + b\mathbb{E}(Y|\mathcal{A}).$$

(f) Sei $\mathcal{G} \subset \mathcal{A}$ eine weitere Teil-σ-Algebra. Dann gilt

$$\mathbb{E}(\mathbb{E}(X|\mathcal{A})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}).$$

(g) Es gilt $\mathbb{E}(\mathbb{E}(X|\mathcal{A})) = \mathbb{E}(X)$.

(h) Sei $Y \in \mathcal{L}^1$ mit $Y \leq X$. Dann gilt $\mathbb{E}(Y|\mathcal{A}) \leq \mathbb{E}(X|\mathcal{A})$.

(i) Es gilt $|\mathbb{E}(X|\mathcal{A})| \leq \mathbb{E}(|X||\mathcal{A})$.

Beweis. Übung. $\square$

Wir schreiben $\mathbb{E}(X|Y)$ für $\mathbb{E}(X|\sigma(Y))$ für zwei Zufallsvariablen X, Y.

Satz 2.79. *Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra. Dann gelten die folgenden Aussagen.*

(a) Sei $0 \leq X_n \in \mathcal{L}^1$ und $0 \leq X \in \mathcal{L}^1$. Es gelte $X_n \nearrow X$ fast sicher für $n \to \infty$. Dann gilt auch

$$\mathbb{E}(X_n|\mathcal{A}) \nearrow \mathbb{E}(X|\mathcal{A}), \quad n \to \infty$$

fast sicher.

(b) Sei $X_n \in \mathcal{L}^1$ und $X \in \mathcal{L}^1$. Sei $\varphi \in \mathcal{L}^1$ mit $|X_n| \leq \varphi$ für alle $n \geq 1$ und es gelte $X_n \longrightarrow X$ fast sicher. Dann gilt

$$\mathbb{E}(X_n|\mathcal{A}) \longrightarrow \mathbb{E}(X|\mathcal{A}), \quad n \to \infty$$

fast sicher.

(c) Seien X, Y Zufallsvariablen mit $X \cdot Y, Y \in \mathcal{L}^1$. Ferner sei X messbar bezüglich $\mathcal{A}$. Dann gilt

$$\mathbb{E}(X \cdot Y|\mathcal{A}) = X\mathbb{E}(Y|\mathcal{A})$$

fast sicher.

Beweis. (a) Aus der Monotonie des bedingten Erwartungswertes folgt $0 \leq \mathbb{E}(X_n|\mathcal{A}) \leq \mathbb{E}(X_{n+1}|\mathcal{A})$. Es sei $Y := \lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{A}) = \sup_{n\in\mathbb{N}} \mathbb{E}(X_n|\mathcal{A}) \in \mathbb{R} \cup \{\pm\infty\}$. Dann ist Y messbar bezüglich $\mathcal{A}$. Für jedes $A \in \mathcal{A}$ gilt

$$\mathbb{E}(\mathbb{1}_A \cdot X_n) = \mathbb{E}(\mathbb{1}_A \cdot \mathbb{E}(X_n|\mathcal{A})), \quad n \geq 1.$$

Aus dem Satz der monotonen Konvergenz folgt

$$\mathbb{E}(\mathbb{1}_A \cdot X) = \mathbb{E}(\mathbb{1}_A \cdot Y), \quad A \in \mathcal{A}.$$

(b) Wir betrachten zuerst den Fall $X_n \geq 0$ fast sicher. In diesem Fall ist auch $X \geq 0$ fast sicher. Angenommen X_n ist monoton. Wir betrachten den Fall, dass X_n monoton steigend ist. Der Fall einer monoton fallenden Folge geht analog. Ist X_n monoton steigend, so folgt aus (a) $\mathbb{E}(X_n|\mathcal{A}) \nearrow \mathbb{E}(X|\mathcal{A})$ für $n \to \infty$. Ist X_n nicht monoton, so betrachten wir $X_n' := \inf_{m\geq n} X_m$ und $X_n'' = \sup_{m\geq n} X_m$. Dann gilt $X_n', X_n'' \leq \varphi$. Diese Folgen sind monoton mit $X_n' \nearrow \liminf_{n\to\infty} X_n = X$ und $X_n'' \searrow \limsup_{n\to\infty} X_n = X$ fast sicher. Aus dem ersten Fall folgt

$$\lim_{n\to\infty} \mathbb{E}(X_n'|\mathcal{A}) = \mathbb{E}(X|\mathcal{A}) = \lim_{n\to\infty} \mathbb{E}(X_n''|\mathcal{A}).$$

Wegen $X_n' \leq X_n \leq X_n''$ folgt in diesem Fall die Behauptung aus

$$\mathbb{E}(X_n'|\mathcal{A}) \leq \mathbb{E}(X_n|\mathcal{A}) \leq \mathbb{E}(X_n''|\mathcal{A}), \quad n \geq 1.$$

Im allgemeinen Fall, sei $X_n = X_n^+ - X_n^-$ sowie $X = X^+ - X^-$ die Zerlegung in Positiv- und Negativteil. Dann folgt

$$\mathbb{E}(X_n|\mathcal{A}) = \mathbb{E}(X_n^+|\mathcal{A}) - \mathbb{E}(X_n^-|\mathcal{A}) \longrightarrow \mathbb{E}(X^+|\mathcal{A}) - \mathbb{E}(X^-|\mathcal{A}) = \mathbb{E}(X|\mathcal{A})$$

fast sicher.
(c) *Fall 1.* Sei $X = \mathbb{1}_B$ mit $B \in \mathcal{A}$. Dann gilt für jede Menge $A \in \mathcal{A}$

$$\mathbb{E}(\mathbb{1}_A X \cdot Y) = \mathbb{E}(\mathbb{1}_{A\cap B} Y) = \mathbb{E}(\mathbb{1}_{A\cap B}\mathbb{E}(Y|\mathcal{A})) = \mathbb{E}(\mathbb{1}_A X \mathbb{E}(Y|\mathcal{A})).$$

Fall 2. Aufgrund der Linearität obiger Gleichung gilt die Behauptung für alle Elementarfunktionen $X = \sum_{n=1}^{N} a_n \mathbb{1}_{B_n}$ mit $a_n \in [0,\infty)$ und $B_n \in \mathcal{A}$.
Fall 3. Seien $X, Y \geq 0$. Dann gibt es eine Folge von Elementarfunktionen X_n mit $X_n \nearrow X$. Dann gilt $X_n \cdot Y \nearrow X \cdot Y$ und wir erhalten aus (a)

$$\mathbb{E}(X \cdot Y|\mathcal{A}) = \lim_{n\to\infty} \mathbb{E}(X_n \cdot Y|\mathcal{A}) = \lim_{n\to\infty} X_n \mathbb{E}(Y|\mathcal{A}) = X\mathbb{E}(Y|\mathcal{A}).$$

Fall 4. Seien X, Y wie in der Behauptung. Durch Zerlegung in Positiv- und Negativteil erhalten wir

$$\mathbb{E}(X \cdot Y|\mathcal{A}) = X\mathbb{E}(Y|\mathcal{A}).$$

□

Satz 2.80. *(Jensen Ungleichung für bedingte Erwartungswerte, [Bil99]) Sei $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ konvex und X eine Zufallsvariable mit $X, \varphi \circ X \in \mathcal{L}^1$. Sei $\mathcal{A} \subset \mathcal{F}$ eine Teil-σ-Algebra. Dann gilt*

$$\varphi(\mathbb{E}(X|\mathcal{A})) \leq \mathbb{E}(\varphi \circ X|\mathcal{A}).$$

3 Wahrscheinlichkeitsmaße auf metrischen Räumen

Bisher haben wir lediglich Zufallsvariablen mit Werten in $\mathbb{R}$ bzw. $\overline{\mathbb{R}}$ betrachtet. In vielen Anwendungen ist es jedoch notwendig den Wertebereich allgemeiner zu wählen, z.B.

- Matrizen (Zufällige Matrizen)
- Räume von Funktionen (stochastische Prozesse)
- Räume von (Punkt-)Maßen (Interagierende Teilchensysteme).

Um solchen Anforderungen gerecht zu werden betrachten wir im Folgenden metrische Räume.

Definition 3.1. *Sei E eine nicht-leere Menge. Eine Metrik auf E ist eine Abbildung $d : E \times E \longrightarrow \mathbb{R}_+$ mit den Eigenschaften*

- *$d(x, y) = 0$ genau dann, wenn $x = y$.*
- *$d(x, y) = d(y, x)$.*

- $d(x, y) \leq d(x, z) + d(z, y)$.

Ist d eine Metrik auf E, so heisst (E, d) metrischer Raum.

Im folgenden ist (E, d) stets ein metrischer Raum. Für $x \in E$ und $\varepsilon > 0$ setze

$$U_\varepsilon(x) = \{y \in E \mid d(x, y) < \varepsilon\}, \qquad B_\varepsilon(x) = \{y \in E \mid d(x, y) \leq \varepsilon\}.$$

Eine Menge $A \subset E$ heißt offen, falls für alle $x \in A$ es ein $\varepsilon > 0$ gibt mit $U_\varepsilon(x) \subset A$. Eine Menge B heisst abgeschlossen, falls B^c offen ist. Die Borel-σ-Algebra ist definiert als die kleinste σ-Algebra, welche die offenen Mengen enthält, d.h.

$$\mathcal{B}(E) = \sigma\left(A \subset E \mid A \text{ ist offen }\right).$$

Wir bezeichnen mit $\mathcal{P}(E)$ den Raum aller (Borel-)Wahrscheinlichkeitsmaße auf $(E, \mathcal{B}(E))$.

Definition 3.2. *Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und (E, d) ein metrischer Raum. Für eine messbare Abbildung $X : \Omega \longrightarrow E$ ist das Bildmaß definiert durch*

$$\mu_X(A) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}), \quad A \in \mathcal{B}(E).$$

Dann heißt μ_X auch Verteilung von X.

Wir benutzen häufig die folgende Notation

$$\mu(A) = (\mathbb{P} \circ X^{-1})(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A).$$

Der nächste Satz bildet das Kernstück für das Rechnen mit Bildmaßen.

Satz 3.3. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, (E, d) ein metrischer Raum $X : \Omega \longrightarrow E$ eine messbare Abbildung. Dann gelten die folgenden Aussagen.*

(a) μ_X ist ein Wahrscheinlichkeitsmaß auf $(E, \mathcal{B}(E))$.

(b) Für jede messbare Abbildung $h : E \longrightarrow [0, \infty]$ gilt

$$\int_E h(x)\mu_X(dx) = \mathbb{E}\left(h(X)\right) \tag{3.1}$$

wobei $+\infty = +\infty$ möglich ist und die linke Seite endlich ist genau dann wenn die rechte Seite endlich ist.

(c) Sei $h : E \longrightarrow \overline{\mathbb{R}}$ messbar derart dass eine der beiden Seiten in (3.1) endlich ist für $|h|$. Dann gilt (3.1) auch für h.

Beweis. Übung. □

Beispiel 3.4. *Sei $E = \mathbb{R}$, X eine Zufallsvariable und μ_X die Verteilung von X. Dann ist*

$$\mathbb{E}(X^k) = \int_{\mathbb{R}} x^k d\mu_X(x)$$

sofern die rechte Seite endlich ist für $|x|^k$.

Der nächste Satz zeigt, dass jede Verteilung realisiert werden kann durch eine Zufallsvariable.

Satz 3.5. *Sei μ ein Wahrscheinlichkeitsmaß auf einem metrischen Raum (E, d). Dann gibt es einen Wahrscheinlichkeitsraum und eine Zufallsvariable X derart, dass $\mu_X = \mu$.*

Beweis. Übung. □

Aus diesem Grund beschränken wir uns in diesem Kapitel darauf Eigenschaften von Wahrscheinlichkeitsmaßen auf einem metrischen Raum E genauer zu studieren.

3.1 Reguläre Wahrscheinlichkeitsmaße

Der Begriff der Kompaktheit ist für die nächsten zwei Kapitel von zentraler Bedeutung.

Definition 3.6. *Sei (E, d) ein metrischer Raum und $K \subset E$.*

(i) K heisst kompakt, falls es für jede offene Überdeckung eine endliche Teilüberdeckung gibt. D.h. für jede Indexmenge I und jede Wahl von offenen Mengen $(U_i)_{i \in I} \subset E$ mit $K \subset \bigcup_{i \in I} U_i$, gibt es $n \in \mathbb{N}$ und $i_1, \ldots, i_n \in I$ mit $K \subset \bigcup_{j=1}^{n} U_{i_j}$.

(ii) K heisst folgenkompakt, falls jede Folge in K eine konvergente Teilfolge in K hat. D.h. für jede Folge $(x_n)_{n \in \mathbb{N}} \subset K$ gibt es $x \in K$ und eine Teilfolge $(x_{n_k})_{k \in \mathbb{N}}$ derart dass $x_{n_k} \longrightarrow x$ für $k \to \infty$.

Bemerkung 3.7. *$K \subset E$ ist genau dann kompakt wenn, es folgenkompakt ist.*

Der nächste Satz charakterisiert die kompakten Mengen in vollständigen metrischen Räumen.

Lemma 3.8. *Sei (E, d) ein vollständiger metrischer Raum und $K \subset E$ abgeschlossen. Dann ist K genau dann kompakt, wenn es total beschränkt ist, d.h. für jedes $\varepsilon > 0$ gibt es endlich viele Punkte $x_1, \ldots, x_m \in E$ mit*

$$K \subset \bigcup_{k=1}^{m} B_\varepsilon(x_k). \tag{3.2}$$

Beweis. Sei K kompakt und $\varepsilon > 0$. Dann ist

$$K \subset \bigcup_{x \in K} U_\varepsilon(x)$$

eine offene Überdeckung von K. Es gibt daher endlich viele Punkte $x_1, \ldots, x_m \in K$ mit (3.2).

Umgekehrt gelte (3.2) für jedes $\varepsilon > 0$. Wir zeigen, dass jede Folge in K eine konvergente Teilfolge enthält. Sei $(x_n)_{n \in \mathbb{N}} \subset K$. Für jedes $m \geq 1$ gibt es endlich viele Bälle mit Radius $\frac{1}{m}$, welche K überdecken. Mindestens einer dieser Bälle enthält unendlich viele Folgenglieder von $(x_n)_{n \in \mathbb{N}}$. Für $m = 1$ sei B_1 der Ball mit der Eigenschaft, dass $N_1 := \{n \mid x_n \in B_1$ unendlich viele Elemente hat. Sei $n_1 \in N_1$ beliebig. Für $m = 2$ sei B_2 der Ball mit der Eigenschaft, dass

$N_2 := \{n > n_1 \mid x_n \in B_1 \cap B_2\}$ unendlich viele Elemente hat. Wähle $n_2 \in N_2$ beliebig. Durch Iteration erhalten wir Folgen B_m, N_m und $(n_m)_{m\in\mathbb{N}}$ sodass

$$N_{m+1} = \left\{ n > n_m \mid x_n \in \bigcap_{k=1}^{m} B_k \right\}$$

unendlich viele Elemente enthält und $n_{m+1} \in N_{m+1}$. Dann ist $(x_{n_m})_{m\in\mathbb{N}}$ eine Teilfolge und es gilt $x_{n_k} \in B_m$ für alle $k \geq m$. Insbesondere gilt

$$d(x_{n_k}, x_{n_l}) \leq \frac{1}{m}, \quad k, l \geq m,$$

d.h. $(x_{n_m})_{m\geq 1}$ ist eine Cauchy Folge. Da E vollständig ist, hat diese einen Grenzwert $x \in E$. Da K abgeschlossen ist, liegt der Grenzwert in K. □

Als nächstes definieren wir die zentrale Eigenschaft für diesen Abschnitt.

Definition 3.9. *$\mu \in \mathcal{P}(E)$ heißt regulär, falls für jedes $A \in \mathcal{B}(E)$*

$$\mu(A) = \sup\{\mu(K) \mid K \subset A \text{ kompakt}\} = \inf\{\mu(U) \mid A \subset U \text{ offen }\}.$$

Eine äquivalente Definition ist gegeben durch: Für jedes $A \in \mathcal{B}(E)$ und jedes $\varepsilon > 0$ gibt es eine kompakte Menge K_ε und eine offene Menge U_ε mit $K_\varepsilon \subset A \subset U_\varepsilon$ und

$$\mu(U_\varepsilon \backslash K_\varepsilon) < \varepsilon. \tag{3.3}$$

Satz 3.10. *Sei E ein metrischer Raum. Dann gilt für jedes $\mu \in \mathcal{P}(E)$*

$$\mu(A) = \sup\{\mu(K) \mid K \subset A \text{ abgeschlossen }\} = \inf\{\mu(U) \mid A \subset U \text{ offen }\}.$$

Für den Beweis ist die folgende Funktion hilfreich. Sei $A \subset E$. Der Abstand von $x \in E$ zu A ist definiert über

$$d(x, A) := \inf\{d(x, a) \mid a \in A\}.$$

Es gilt

$$|d(x, A) - d(y, A)| \leq d(x, y), \quad x, y \in E$$

und $d(x, A) = 0$ genau dann, wenn $x \in \overline{A}$. Hierbei bezeichnet $\overline{A}$ den Abschluss von A.

Beweis. Es sei

$$\mathcal{K} = \{A \in \mathcal{B}(E) \mid \forall \varepsilon > 0 \ \exists U_\varepsilon \text{ offen, } \exists K_\varepsilon \text{ abgeschlossen mit } K_\varepsilon \subset A \subset U_\varepsilon \text{ und (3.3)}\}.$$

Dann enthält $\mathcal{K}$ alle abgeschlossenen Mengen. Denn ist A abgeschlossen, so wählen wir

$$U_n := \left\{ x \in E \mid d(x, A) < \frac{1}{n} \right\}.$$

Dann ist $A \subset U_n$ und $A = \bigcap_{n\geq 1} U_n$. Folglich gilt $\lim_{n\to\infty} \mu(U_n) = \mu(A)$, also $A \in \mathcal{K}$.

Wir müssen zeigen, dass $\mathcal{K} = \mathcal{B}(E)$. Hierfür reicht es zu zeigen, dass $\mathcal{K}$ ein Dynkin-System ist.

- $E \in \mathcal{K}$. Da E offen und abgeschlossen ist können wir $U_\varepsilon = E = K_\varepsilon$ wählen.
- Sei $A \in \mathcal{K}$ und $\varepsilon > 0$. Seien $K_\varepsilon, U_\varepsilon$ die dazugehörigen Mengen. Dann ist $K' := E \backslash K$ offen, $U' := E \backslash U$ abgeschlossen und es gilt $U' \subset E \backslash A \subset K'$ und
$$\mu(K' \backslash U') = \mu(U \backslash K) < \varepsilon.$$
- Sei $(A_n)_{n \in \mathbb{N}} \subset \mathcal{K}$ eine Folge disjunkter Mengen. Sei $\varepsilon > 0$. Dann gilt
$$0 \leq \sum_{n=1}^{\infty} \mu(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \mu(E) \leq 1,$$
d.h. $(\mu(A_n))_{n \geq 1}$ ist summierbar. Folglich gibt es ein $n_0 \geq 2$ mit $\sum_{n=n_0}^{\infty} \mu(A_n) < \frac{\varepsilon}{2}$. Wähle abgeschlossene Mengen K_e^n und offene Mengen U_ε^n mit $K_\varepsilon^n \subset A_n \subset U_\varepsilon^n$ und
$$\mu(U_\varepsilon^n \backslash K_\varepsilon^n) < \frac{\varepsilon}{2^{n+1}}, \quad n \in \mathbb{N}.$$
Dann ist $K_\varepsilon := \bigcup_{n=1}^{n_0-1} K_\varepsilon^n$ abgeschlossen und $U_\varepsilon := \bigcup_{n=1}^{\infty} U_\varepsilon^n$ offen. Ferner gilt
$$\mathcal{K}_\varepsilon \subset \bigcup_{n=1}^{\infty} K_\varepsilon^n \subset \bigcup_{n=1}^{\infty} A_n \subset U_\varepsilon$$
und
$$\begin{aligned}
\mu(U_\varepsilon \backslash K_\varepsilon) &\leq \sum_{n=1}^{\infty} \mu(U_\varepsilon^n \backslash K_\varepsilon) = \sum_{n=1}^{n_0-1} \mu(U_\varepsilon^n \backslash K_\varepsilon) + \sum_{n=n_0}^{\infty} \mu(U_\varepsilon^n \backslash K_\varepsilon) \\
&\leq \sum_{n=1}^{n_0-1} \mu(U_\varepsilon^n \backslash K_\varepsilon^n) + \sum_{n=n_0}^{\infty} \mu(U_\varepsilon^n) \leq \varepsilon \sum_{n=1}^{n_0-1} \frac{1}{2^{n+1}} + \sum_{n=n_0}^{\infty} \left(\frac{\varepsilon}{2^{n+1}} + \mu(K_\varepsilon^n)\right) \\
&\leq \frac{\varepsilon}{2} \sum_{n=1}^{\infty} \frac{1}{2^n} + \sum_{n=n_0}^{\infty} \mu(A_n) \leq \varepsilon.
\end{aligned}$$
Damit ist $\bigcup_{n=1}^{\infty} A_n \in \mathcal{K}$.

□

Unser Ziel ist es zu zeigen, dass jedes $\mu \in \mathcal{P}(E)$ regulär ist.

Definition 3.11. *Ein metrischer Raum E heisst seperabel, falls er eine abzählbar dichte Teilmenge enthält; d.h. es gibt eine Folge $(a_n)_{n \in \mathbb{N}} \subset E$ derart dass für jedes $x \in E$ finden wir eine Teilfolge $(n_k)_{k \in \mathbb{N}}$ mit $d(a_{n_k}, x) \longrightarrow 0$ für $k \to \infty$.*

Satz 3.12. *Sei E ein vollständiger, seperabler metrischer Raum. Dann ist jedes $\mu \in \mathcal{P}(E)$ regulär.*

Beweis. Sei $(a_n)_{n\in\mathbb{N}} \subset E$ eine abzählbar dichte Teilmenge. Dann gilt für jedes $\delta > 0$

$$\bigcup_{n\geq 1} B_\delta(a_n) = E.$$

Damit folgt

$$1 = \mu(E) = \mu\left(\bigcup_{n\geq 1} B_\delta(a_n)\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} B_\delta(a_k)\right).$$

Sei $\varepsilon > 0$. Dann gibt es für jedes $m \geq 1$ ein $n_m \in \mathbb{N}$ mit

$$\mu\left(\bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)\right) \geq 1 - 2^{-m}\varepsilon.$$

Sei

$$K := \bigcap_{m\geq 1} \bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k).$$

Dann ist K abgeschlossen. Für jedes $\delta > 0$ sei $m \geq 1$ mit $\frac{1}{m} \leq \delta$. Dann gilt

$$K \subset \bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k) \subset \bigcup_{k=1}^{n_m} B_\delta(a_k).$$

Also ist K nach Lemma 3.8 kompakt. Dann gilt

$$\begin{aligned}\mu(K^c) &= \mu\left(\bigcup_{m\geq 1} \bigcap_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)^c\right) \leq \sum_{m=1}^{\infty} \mu\left(\bigcap_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)^c\right) \\ &= \sum_{m=1}^{\infty}\left(1 - \mu\left(\bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)\right)\right) \leq \varepsilon \sum_{m=1}^{\infty} 2^{-m} = \varepsilon,\end{aligned}$$

wo wir

$$\bigcap_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)^c = \bigcap_{k=1}^{n_m} (E \backslash B_{\frac{1}{m}}(a_k)) = E \backslash \left(\bigcup_{k=1}^{n_m} B_{\frac{1}{m}}(a_k)\right)$$

benutzt haben.

Wir haben gezeigt, dass für jedes $\varepsilon > 0$ es eine kompakte Menge $K_\varepsilon \subset E$ gibt mit $\mu(K_\varepsilon^c) < \varepsilon$. Daraus folgern wir jetzt die Behauptung. Sei $\varepsilon > 0$ fest und wähle U_ε offen, A_ε abgeschlossen mit $A_\varepsilon \subset A \subset U_\varepsilon$ und

$$\mu(U_\varepsilon \backslash A_\varepsilon) < \varepsilon.$$

Wähle K_ε wie oben und setze $A'_\varepsilon = A_\varepsilon \cap K_\varepsilon$. Dann ist A'_ε kompakt, es gilt

$$A'_\varepsilon \subset A_\varepsilon \subset A \subset U_\varepsilon$$

sowie

$$\mu(U_\varepsilon \backslash A'_\varepsilon) \leq \mu(U_\varepsilon \backslash A_\varepsilon) + \mu(U_\varepsilon \backslash K_\varepsilon) \leq \varepsilon + \mu(K_\varepsilon^c) \leq 2\varepsilon.$$

□

Zum Schluss geben wir eine hinreichende Bedingung an, dass ein metrischer Raum seperabel ist.

Definition 3.13. *Ein metrischer Raum (E, d) heisst σ-kompakt, falls es eine Folge kompakter Mengen $K_n \subset E$ mit $\bigcup_{n\geq 1} K_n = E$.*

Lemma 3.14. *Sei (E, d) ein σ-kompakter metrischer Raum. Dann ist (E, d) seperabel.*

Beweis. Schritt 1. Angenommen (E, d) ein ein kompakter metrischer Raum. Für jedes $n \geq 1$ gilt $E = \bigcup_{x\in E} U_{\frac{1}{n}}(x)$ und wegen kompaktheit gibt es $x_1^n, \ldots, x_{N^n}^n \in E$ mit $N^n \in \mathbb{N}$ und

$$E = \bigcup_{k=1}^{N^n} U_{\frac{1}{n}}(x_k^n).$$

Dann ist $D := \{x_k^n \mid k = 1, \ldots, N^n,\ n \geq 1\}$ abzählbar und erfüllt das gewünschte.
Schritt 2. Sei (E, d) ein σ-kompakter Raum. Wähle kompakte Mengen $K_n \subset E$ mit $E = \bigcup_{n\geq 1} K_n$. Dann ist (K_n, d_n), wo $d_n : K_n \times K_n \longrightarrow \mathbb{R}_+$ die Einschränkung auf K_n bezeichnet, ein kompakter metrischer Raum. Nach Schritt 1 gibt es eine abzählbar dichte Teilmenge $D_n \subset K_n$. Dann ist $D := \bigcup_{n\geq 1}$ ebenfalls abzählbar. Es ist nicht schwer zu zeigen, dass D dicht in E liegt. □

3.2 Polnische Räume

Letzer Abschnitt zeigt, dass Wahrscheinlichkeitsmaße auf vollständigen metrischen Räumen stets regulär sind. In vielen Anwendungen begegnen wir jedoch Räumen wo nicht direkt ersichtlich ist, wie diese metrisiert werden können. Hierfür hat sich das nachfolgende Konzept als nützlich erwiesen.

Definition 3.15. *Sei E eine nicht-leere Menge. Eine Topologie auf E ist ein Mengensystem $\mathcal{T}$ auf E mit den Eigenschaften*

- $\emptyset, E \in \mathcal{T}$.
- *Sind $A_1, \ldots, A_n \in \mathcal{T}$, so ist auch $\bigcap_{k=1}^n A_k \in \mathcal{T}$.*
- *Sind $(A_i)_{i\in I} \subset \mathcal{T}$ wo I eine beliebige Indexmenge ist, so ist auch $\bigcup_{i\in I} A_i \in \mathcal{T}$.*

Das Paar $(E, \mathcal{T})$ heisst dann topologischer Raum und Elemente $A \in \mathcal{T}$ heissen offene Mengen.

Jeder metrische Raum (E, d) ist auch ein topologischer Raum mit

$$\mathcal{T}_d := \{A \subset E \mid A \text{ ist offen bezüglich der Metrik } d\}.$$

Definition 3.16. *Ein Polnischer Raum E ist ein topologischer Raum $(E, \mathcal{T})$ derart dass eine Metrik d auf E gibt mit*

- *Es gilt $\mathcal{T}_d = \mathcal{T}$.*
- *E ist seperabel und vollständig bezüglich dieser Metrik.*

Beispiel 3.17. *Kontinuierliche Polnische Räume sind* $E = \mathbb{R}_+$ *und* $E = \mathbb{R}^d$. *Diskrete Polnische Räume sind* $E = \mathbb{Z}_+$ *oder* $E = \mathbb{Z}^d$ *mit* $d \geq 1$. *Die Metrik ist in allen Fällen gegeben durch*

$$d(x, y) := \sqrt{\sum_{k=1}^{d} |x_k - y_k|^2}, \quad x, y \in E.$$

Beachte, dass im diskreten Fall jede Teilmenge $A \subset E$ *offen ist. Abzählbare Produkte von Polnischen Räumen sind wieder Polnisch, sofern diese mit der Produkttopologie versehen sind. Insbesondere ist also*

$$\mathbb{R}^{\mathbb{Z}_+} := \{(x_n)_{n \geq 0} \mid x_n \in \mathbb{R}\}$$

ein Polnischer Raum.

In diesem Abschnitt betrachten wir wichtige Beispiele für Polnische Räume.

Räume stetiger Funktionen

Von besonderem Interesse sind Funktionenräume. Zufallsvariablen mit Werten in solchen Räumen können als stochastische Prozesse bzw. stochastische Felder interpretiert werden. Im folgenden betrachten wir als Beispiel verschiedene Räume stetiger Funktionen.

Sei (E, d) ein kompakter metrischer Raum. Dann ist

$$C(E) = \{f : E \longrightarrow \mathbb{R} \mid f \text{ ist stetig }\}$$

ein Vektorraum. Da jedes $f \in C(E)$ beschränkt ist, ist

$$\|f\|_\infty := \sup_{x \in E} |f(x)|$$

wohldefiniert. Dieses definiert eine Norm auf $C(E)$.

Satz 3.18. *Es gelten die folgenden Aussagen.*

- $(C(E), \|\cdot\|_\infty)$ *ein ein vollständiger, normierter Raum, d.h. ein Banachraum.*
- $(C(E), d)$ *ist bezüglich der Metrik*

  $$d(f, g) = \|f - g\|_\infty$$

  ein Polnischer Raum.

Beweis. Siehe Funktionalanalysis, z.b. [Wer00]. □

Beispiel 3.19. *Man wähle zum Beispiel* $E = [0, 1]$. *Zufallsvariablen mit Werten in* $C([0, 1])$, *d.h.* $X : \Omega \longrightarrow C([0, 1])$ *können somit als zufällige Funktionen* $X_t : \Omega \longrightarrow \mathbb{R}$ *aufgefasst werden.*

Beispiel 3.20. *Der Raum* $C([0, \infty))$ *mit der durch die Metrik*

$$d(f, g) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{d_k(f, g)}{1 + d_k(f, g)}, \quad d_k(f, g) := \sup_{x \in [0,k]} |f(x) - g(x)|$$

erzeugten Topologie ist ein Polnischer Raum.

Später werden wir weiter Räume stetiger Funktionen einführen und betrachten.

Raum endlicher Punktmaße

Zufallsvariablen mit Werten in Räumen von Maßen treten in der Theorie der Punktprozesse häufig auf. Genauer betrachten wir den Raum der endlichen Punktmaße (bzw. der Raum aller endlichen Punktkonfigurationen) auf $\mathbb{R}^d$ definiert durch

$$\ddot{\Gamma}_0 := \{\eta : \mathcal{B}(\mathbb{R}^d) \longrightarrow \mathbb{N}_0 \mid \eta \text{ ist endliches Maß}\}.$$

Satz 3.21. *Jedes $\eta \in \ddot{\Gamma}_0$ hat eine eindeutige Darstellung der Form*

$$\eta = \sum_{k=1}^{N} n_k \delta_{x_k},$$

wo $N \in \mathbb{N}$, $n_1, \dots, n_N \in \mathbb{N}$ und $x_1, \dots, x_N \in \mathbb{R}^d$ paarweise verschieden sind.

Die Punkte $x_1, \dots, x_N$ heissen Atome von η.

Beweis. Sei $A := \{x \in \mathbb{R}^d \mid \eta(\{x\}) \geq 1\}$. Angenommen A enthält unendlich viele Elemente. Dann gibt es eine Folge paarweise verschiedener Elemente $(x_n)_{n\in\mathbb{N}} \subset A$. Für diese Folge gilt

$$\infty = \sum_{n=1}^{\infty} \eta(\{x_n\}) = \eta\{\{x_n \mid n \in \mathbb{N}\}) \leq \eta(\mathbb{R}^d) < \infty.$$

Also enthält A nur endlich viele Elemente. Sei $A = \{x_1, \dots, x_N\}$ für paarweise verschiedene $x_1, \dots, x_N \in \mathbb{R}^d$ und $N \in \mathbb{N}_0$. Definiere

$$\mu := \eta - \sum_{k=1}^{N} n_k \delta_{x_k}$$

mit $n_k := \eta(\{x_k\})$. Dann folgt für $B \in \mathcal{B}(\mathbb{R}^d)$ aus $\eta(B \cap \{x_k\}) = n_k \delta_{x_k}(B)$ bereits

$$\begin{aligned}
\mu(B) = \eta(B) - \sum_{k=1}^{N} n_k \delta_{x_k}(B) &\geq \eta(A \cap B) - \sum_{k=1}^{N} n_k \delta_{x_k}(B) \\
&= \sum_{k=1}^{N} \eta(B \cap \{x_k\}) - \sum_{k=1}^{N} n_k \delta_{x_k}(B) = 0.
\end{aligned}$$

Damit ist $\mu \geq 0$ und folglich ein Maß mit Werten in $\mathbb{N}_0$ derart dass $\mu(\{x\}) = 0$ für alle $x \in \mathbb{R}^d$ gilt. Wir zeigen

$$\mu(B) = 0, \quad B \in \mathcal{B}(\mathbb{R}^d)$$

woraus die Behauptung folgt. Es gilt $\mu(B) = \lim_{R\to\infty} \mu(B \cap B_R)$, wo $B_R = \{x \in \mathbb{R}^d \mid |x| \leq R\}$. Da $\mu(B \cap B_R) \in \mathbb{N}_0$, gibt es ein $R > 0$ derart dass $\mu(B) = \mu(B \cap B_R)$. Für jedes $x \in \mathbb{R}^d$ gilt $\mu(B_\varepsilon(x)) \searrow \mu(\{x\}) = 0$ für $\varepsilon \to 0$. Wegen $\mu(B_\varepsilon(x)) \in \mathbb{N}_0$ gibt es ein $\varepsilon(x) > 0$ derart, dass $\mu(B_{\varepsilon(x)}(x)) = 0$ gilt. Da $\overline{B \cap B_R}$ beschränkt und abgeschlossen ist, ist diese Menge auch kompakt. Wegen

$$\overline{B \cap B_R} \subset \bigcup_{x \in \overline{B \cap B_R}} U_{\varepsilon(x)}(x)$$

mit $U_{\varepsilon(x)}(x) = \{y \in \mathbb{R}^d \mid |y-x| < \varepsilon(x)\}$ gibt es endlich viele Punkte $x_1, \dots, x_m$ derart, dass

$$\overline{B \cap B_R} \subset \bigcup_{j=1}^{m} U_{\varepsilon(x_j)}(x_j) \subset \bigcup_{j=1}^{m} B_{\varepsilon(x_j)}(x_j).$$

Daraus folgt

$$\mu(B) = \mu(B \cap B_R) \leq \mu(\overline{B \cap B_R}) \leq \sum_{j=1}^{m} \mu(B_{\varepsilon(x_j)}(x_j)) = 0.$$

□

Der Raum der endlichen Teilmengen von $\mathbb{R}^d$ (bzw. der Raum der endlichen Konfigurationen) ist gegeben durch

$$\Gamma_0 := \left\{ \eta \subset \mathbb{R}^d \mid |\eta| := \sum_{x \in \eta} 1 < \infty \right\}.$$

Jedes $\eta \in \Gamma_0$ kann vermöge $\sum_{x\in\eta} \delta_x$ mit einem Element in $\ddot{\Gamma}_0$ identifiziert werden.

Raum der lokal endlichen Punktmaße

Es sei $C_c(\mathbb{R}^d)$ der Raum der stetigen Funktionen $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ mit kompaktem Träger. Eine Funktion $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ hat kompakten Träger, falls es ein $R > 0$ gibt mit $f(x) = 0$ für alle $|x| \geq R$. Der Raum aller lokal endlichen Teilmengen von $\mathbb{R}^d$ (bzw. lokal endlichen Konfigurationen) ist definiert durch

$$\Gamma := \{\gamma \subset \mathbb{R}^d \mid |\gamma \cap K| < \infty \quad K \subset \mathbb{R}^d \text{ kompakt } \}.$$

Jedes $\gamma \in \Gamma$ kann vermöge $\sum_{x\in\gamma} \delta_x$ mit einem lokal endlichen Punktmaß auf $\mathbb{R}^d$ identifiziert werden. Für jedes $f \in C_c(\mathbb{R}^d)$ sei

$$I_f(\gamma) := \sum_{x\in\gamma} f(x) = \int_{\Gamma} f(x)\gamma(dx)$$

das Integral von f bezüglich des Punktmaßes γ. Die Topologie auf Γ ist definiert als die kleinste Topologie derart, dass alle Abbildungen I_f stetig sind. Es lässt sich zeigen, dass Γ bezüglich dieser Topologie ein Polnischer Raum ist, siehe [KK06].

Die Erweiterung auf alle lokal endlichen Punktmaße wird im nächsten Beispiel behandelt. Der Raum aler lokal endlichen Punktmaße ist gegeben durch

$$\ddot{\Gamma} := \{\hat{\gamma} : \mathcal{B}(\mathbb{R}^d) \longrightarrow \mathbb{N}_0 \cup \{\infty\} \mid \hat{\gamma}(K) < \infty, \quad K \subset \mathbb{R}^d \text{ kompakt } \}.$$

Satz 3.22. *Jedes $\hat{\gamma} \in \ddot{\Gamma}$ hat die Darstellung*

$$\hat{\gamma} = \sum_{x\in\gamma} n_x \delta_x,$$

wo $n_x \in \mathbb{N}_0$ und $\gamma \in \Gamma$.

Beweis. Sei $B_N := \{x \in \mathbb{R}^d \mid |x| \leq N\}$ wo $N \in \mathbb{N}$. Dann ist

$$\mathbb{R}^d = \bigcup_{N=0}^{\infty} (B_{N+1} \backslash B_N)$$

eine disjunkte Zerlegung von $\mathbb{R}^d$, wo $B_0 := \varnothing$. Folglich gilt für jedes $B \in \mathcal{B}(\mathbb{R}^d)$

$$\widehat{\gamma}(B) = \sum_{N=0}^{\infty} \widehat{\gamma}((B_{N+1} \backslash B_N) \cap B).$$

Sei $\widehat{\gamma}_N(B) := \widehat{\gamma}((B_{N+1} \backslash B_N) \cap B)$. Dann ist $\widehat{\gamma} = \sum_{N=0}^{\infty} \widehat{\gamma}_N$ und es ist nicht schwer zu sehen, dass $\widehat{\gamma}_N \in \ddot{\Gamma}_0$ für alle $N \geq 0$. Folglich gibt es $n_1^{(N)}, \ldots, n_{k(N)}^{(N)} \in \mathbb{N}_0$ sowie $x_1^{(N)}, \ldots, x_{k(N)}^{(N)} \in B_{N+1} \backslash B_N$ mit $k(N) \in \mathbb{N}$ und $N \in \mathbb{N}_0$ derart, dass

$$\widehat{\gamma}_N = \sum_{j=1}^{k(N)} n_j^{(N)} \delta_{x_j^{(N)}}, \quad N \geq 0.$$

□

4 Konvergenz von Wahrscheinlichkeitsmaßen

Sei (E, d) ein metrischer Raum. In diesem Kapitel widmen wir uns der Konvergenz von Folgen von Wahrscheinlichkeitsmaßen auf E.
Idee: Mengenweise Konvergenz, d.h.

$$\mu_n(A) \longrightarrow \mu(A), \quad A \in \mathcal{B}(E), \; n \to \infty. \tag{4.1}$$

Problem: Für viele praktische Anwendungen ist diese Art der Konvergenz zu stark. Dazu betrachten wir die Folge $\mu_n(dx) = n\mathbb{1}_{[0,\frac{1}{n}]}(x)m(dx)$ und $A = (0, 1]$. Dann gilt $\mu_n(A) = 1$. Es liegt nahe anzunehmen, dass $\mu_n \longrightarrow \delta_0$ konvergiert. Es gilt jedoch $\delta_0(A) = 0$. Eine mengenweise Konvergenz liegt in diesem Beispiel nicht vor.

Wir brauchen daher eine abgeschwächte Form von (4.1). Die Idee ist es Konvergenz über die Konvergenz der dazugehörigen Momente bzw. Erwartungswerte zu beschreiben. Genauer sei $M(E)$ eine Familie von messbaren Funktionen auf E. Dann sagen wir $\mu_n \longrightarrow \mu$ in $M(E)$, falls

$$\int_E f(x) d\mu_n(x) \longrightarrow \int_E f(x) d\mu(x), \quad n \to \infty \tag{4.2}$$

für alle $f \in M(E)$.

Bemerkung 4.1. *Es sei $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Genau dann gilt (4.1), wenn (4.2) für alle beschränkten messbaren Funktionen $f : E \longrightarrow \mathbb{R}$ gilt.*

Beweis. Gilt (4.2), so wählen wir $f = \mathbb{1}_A$ für $A \in \mathcal{B}(E)$. Umgekehrt gelte (4.1). Dann gilt (4.2) für alle Indikatiorfunktionen $f = \mathbb{1}_A$ und mittels Linearität auch für alle Elementarfunktionen. Es reicht (4.2) für alle $f \geq 0$ zu zeigen. Sei dazu $f \geq 0$ beschränkt und messbar und $\varepsilon > 0$ beliebig. Dann gibt es eine Elementarfunktion g mit $0 \leq g \leq f$ und $\sup\limits_{x \in E}(f(x) - g(x)) < \varepsilon$. Daraus folgt

$$\begin{aligned}
&\left| \int\limits_E f(x) d\mu(x) - \int\limits_E f(x) d\mu_n(x) \right| \leq \left| \int\limits_E f(x) d\mu(x) - \int\limits_E g(x) d\mu(x) \right| \\
&+ \left| \int\limits_E g(x) d\mu(x) - \int\limits_E g(x) d\mu_n(x) \right| + \left| \int\limits_E g(x) d\mu_n(x) - \int\limits_E f(x) d\mu_n(x) \right| \\
&\leq 2 \sup_{x \in E} |f(x) - g(x)| + \left| \int\limits_E g(x) d\mu(x) - \int\limits_E g(x) d\mu_n(x) \right| \\
&\leq 2\varepsilon + \left| \int\limits_E g(x) d\mu(x) - \int\limits_E g(x) d\mu_n(x) \right|
\end{aligned}$$

welches mit $n \to \infty$ die Behauptung zeigt. □

Es liegt daher nahe die Menge $M(E)$ geeignet einzuschränken. Je nach Wahl von $M(E)$ ergeben sich verschiedene Konvergenzbegriffe. Einige Wahlen von $M(E)$ werden wir genauer untersuchen.

4.1 Schwache Konvergenz

In diesem Kapitel betrachten wir den Fall $M(E) = C_b(E)$, wo $C_b(E)$ der Raum der stetigen, beschränkten Funktionen $f : E \longrightarrow \mathbb{R}$ ist. Dieser ist ein Banachraum mit der Supremumsnorm

$$\|f\|_\infty := \sup_{x \in E} |f(x)|.$$

Definition 4.2. *Sei $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Dann konvergiert μ_n schwach gegen μ ($\mu_n \longrightarrow \mu$ schwach), falls für alle $f \in C_b(E)$ gilt*

$$\int\limits_E f(x) d\mu_n(x) \longrightarrow \int\limits_E f(x) d\mu(x), \quad n \to \infty.$$

Das nachfolgende Lemma zeigt, dass der Grenzwert bezüglich schwacher Konvergenz eindeutig festgelegt ist.

Lemma 4.3. *Seien $\mu, \nu \in \mathcal{P}(E)$ derart, dass*

$$\int\limits_E f(x) d\mu(x) = \int\limits_E f(x) d\nu(x), \quad \forall f \in C_b(E).$$

Dann gilt $\mu = \nu$.

Beweis. Es reicht zu zeigen, dass $\mu(K) = \nu(K)$ für alle abgeschlossenen Mengen $K \subset E$ gilt. Es sei $K \subset E$ abgeschlossen. Definiere $\varphi(t) := \begin{cases} 1-t, & t \in [0,1] \\ 0, & t \geq 1 \end{cases}$ und sei

$$f_n(x) := \varphi(nd(x,K)), \quad x \in E.$$

Dann ist $0 \leq f_n \leq 1$ und f_n ist stetig für jedes $n \geq 1$. Ferner gilt $f_n(x) \longrightarrow \mathbb{1}_K(x)$ für $n \to \infty$ und jedes $x \in E$. Damit folgt

$$\mu(K) = \lim_{n\to\infty} \int_E f_n(x) d\mu(x) = \lim_{n\to\infty} \int_E f_n(x) d\nu(x) = \nu(K).$$

□

Beispiel 4.4. *(a) Sei $(x_n)_{n\in\mathbb{N}} \subset E$ und $x \in E$. Falls $d(x_n, x) \longrightarrow 0$ für $n \to \infty$ gilt, so konvergiert δ_{x_n} schwach gegen δ_x.*

(b) Sei $\mu_n(dx) = p_n(x)m(dx)$ mit

$$p_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{n}{2}x^2}, \quad x \in \mathbb{R}.$$

Dann konvergiert μ_n schwach gegen δ_0.

(c) Sei $\mu_n(dx) = n\mathbb{1}_{[0,\frac{1}{n}]}(x)m(dx)$. Dann konvergiert μ_n schwach gegen δ_0.

Bemerkung 4.5. *Sei $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Dann sind äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) Für jede Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ von $(\mu_n)_{n\in\mathbb{N}}$ gibt es eine weitere Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ mit $\mu_{n_{k_l}} \longrightarrow \mu$ schwach für $l \to \infty$.

Beweis. $(a) \Longrightarrow (b)$: Konvergiert eine Folge, so konvergiert auch jede Teilfolge gegen denselben Grenzwert.
$(b) \Longrightarrow (a)$: Angenommen (a) gilt nicht. Dann gibt es $f \in C_b(E)$, $\varepsilon > 0$ und eine Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ mit

$$\left| \int_E f(x) d\mu_{n_k}(x) - \int_E f(x) d\mu(x) \right| \geq \varepsilon, \quad k \in \mathbb{N}. \tag{4.3}$$

Nach Voraussetzung gibt es eine weitere Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ mit $\mu_{n_{k_l}} \longrightarrow \mu$ schwach für $l \to \infty$. Für diese Teilfolge gilt

$$\int_E f(x) d\mu_{n_{k_l}}(x) \longrightarrow \int_E f(x) d\mu(x), \quad l \to \infty$$

welches im Wiederspruch zu (4.3) steht. □

Sei $C_u(E)$ der Raum aller gleichmässig stetigen beschränkten Funktionen und sei $BL(E)$ der Raum aller Funktionen wo

$$\|f\|_{BL} = \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)} < \infty.$$

Der nächste Satz liefert eine alternative Charakterisierung der schwachen Konvergenz.

Satz 4.6. *(Portmanteau-Theorem) Sei (E,d) ein metrischer Raum. Sei $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Dann sind die folgenden Aussagen äquivalent.*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) Für alle $f \in C_u(E)$ gilt

$$\int_E f(x) d\mu_n(x) \longrightarrow \int_E f(x) d\mu(x), \quad n \to \infty.$$

(c) Für alle $f \in BL(E)$ gilt

$$\int_E f(x) d\mu_n(x) \longrightarrow \int_E f(x) d\mu(x), \quad n \to \infty.$$

(d) Für alle abgeschlossenen Mengen $F \subset E$ gilt

$$\limsup_{n \to \infty} \mu_n(F) \leq \mu(F).$$

(e) Für alle offenen Mengen $O \subset E$ gilt

$$\liminf_{n \to \infty} \mu_n(O) \geq \mu(O).$$

(f) Für alle Mengen $A \in \mathcal{B}(E)$ mit $\mu(\partial A) = 0$ gilt

$$\lim_{n \to \infty} \mu_n(A) = \mu(A).$$

Bevor wir den Beweis führen brauchen wir das folgende Lemma.

Lemma 4.7. *Sei $g : E \longrightarrow \mathbb{R}$ eine stetige Funktion und $a \in \mathbb{R}$. Dann gilt*

$$\partial\{x \in E \mid g(x) \leq a\} \subset \{x \in E \mid g(x) = a\}.$$

Beweis. Sei $x \in \partial\{x \in R \mid g(x) \leq a\}$. Dann gilt für alle $\varepsilon > 0$

$$U_\varepsilon(x) \cap \{y \in E \mid g(y) > a\} \neq \emptyset$$

sowie

$$U_\varepsilon(x) \cap \{y \in E \mid g(y) \leq a\} \neq \emptyset.$$

Hierbei bezeichnet $U_\varepsilon(x) := \{y \in E \mid d(x,y) < \varepsilon\}$ den offenen Ball mit Radius $\varepsilon > 0$ und Mittelpunkt x. Für $\varepsilon = \frac{1}{n}$ mit $n \in \mathbb{N}$ erhalten wir zwei Folgen $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ mit $d(x,x_n) < \frac{1}{n}$ sowie $d(x,y_n) < \frac{1}{n}$ derart, dass $g(x_n) > a$ und $g(y_n) \le a$. Weiterhin gilt $x_n, y_n \longrightarrow x$ und somit

$$a \le \lim_{n\to\infty} g(x_n) = g(x) = \lim_{n\to\infty} g(y_n) \le a.$$

□

Wir kommen zum Beweis von Satz 4.6.

Beweis. $(d) \iff (e)$: klar durch Komplementbildung.
$(a) \Longrightarrow (b) \Longrightarrow (c)$: trivial da $BL(E) \subset C_u(E) \subset C_b(E)$.
$(c) \Longrightarrow (d)$: Sei $F \subset E$ abgeschlossen und setze

$$G_m := \left\{x \in E \mid d(x,F) < \frac{1}{m}\right\}.$$

Dann ist G_m offen, $G_{m+1} \subset G_m$ und $F = \bigcap_{m\ge1} G_m$. Folglich gilt $\mu(G_m) \longrightarrow \mu(F)$ für $m \to \infty$. Sei

$$\varphi(t) := \begin{cases} 1-t, & t \in [0,1] \\ 0, & t \ge 1 \end{cases}$$

und setze $f_m(x) := \varphi(md(x,F))$. Dann gilt $0 \le f_m \le 1$, $f_m|_{G_m^c} = 0$, $f_m|_F = 1$ und f_m ist Lipschitz stetig mit $\mathbb{1}_F \le f_m \le \mathbb{1}_{G_m}$. Daraus folgt

$$\limsup_{n\to\infty} \mu_n(F) \le \limsup_{n\to\infty} \int_E f_m d\mu_n = \int_E f_m d\mu \le \mu(G_m) \longrightarrow \mu(F).$$

$(d) \Longrightarrow (f)$: Sei $A \in \mathcal{B}(E)$ mit $\mu(\partial A) = 0$. Dann gilt

$$\begin{aligned}\mu(A) = \mu(\overset{\circ}{A}) &\le \liminf_{n\to\infty} \mu_n(\overset{\circ}{A}) \le \liminf_{n\to\infty} \mu_n(A) \le \limsup_{n\to\infty} \mu_n(A) \\ &\le \limsup_{n\to\infty} \mu_n(\overline{A}) \le \mu(\overline{A}) = \mu(A).\end{aligned}$$

$(f) \Longrightarrow (d)$: Sei $F \subset E$ abgeschlossen. Auf Lemma 4.7 folgt für alle $\delta > 0$

$$\partial\{x \in E \mid d(x,F) \le \delta\} \subset \{x \in E \mid d(x,F) = \delta\}. \tag{4.4}$$

Da $\{x \in E \mid d(x,F) = \delta\}$ für alle $\delta > 0$ eine abgeschlossene Menge ist, liegt sie in $\mathcal{B}(E)$. Weiterhin gilt

$$\{x \in E \mid d(x,F) = \delta\} \cap \{x \in E \mid d(x,F) = \delta'\} = \varnothing, \quad \delta \ne \delta'. \tag{4.5}$$

Für $n \in \mathbb{N}$ sei

$$D_n := \left\{\delta > 0 \mid \mu\left(\{x \in E \mid d(x,F) = \delta\}\right) \ge \frac{1}{n}\right\}.$$

Behauptung: D_n enthält für jedes $n \in \mathbb{N}$ nur endlich viele Elemente.

Beweis. Angenommen nicht. Dann gibt es ein $n \in \mathbb{N}$ und eine Folge paarweise verschiedener Elemente $(\delta_k)_{k\in\mathbb{N}} \subset D_n$. Dann folgt aus (4.5)

$$1 \geq \mu\left(\bigcup_{k\geq 1}\{x \in E \mid d(x,F) = \delta_k\}\right) = \sum_{k=1}^{\infty} \mu(\{x \in E \mid d(x,F) = \delta_k\}) \geq \sum_{k=1}^{\infty} \frac{1}{n} = \infty.$$

□

Also ist $D = \bigcup_{n\geq 1} D_n$ abzählbar. Folglich gibt es eine Nullfolge $(\delta_k)_{k\in\mathbb{N}} \subset [0,\infty)\backslash D$. Aus (4.4) folgt

$$\mu(\partial F_k) \leq \mu(\{x \in E \mid d(x,F) = \delta_k\}) = 0$$

für $F_k := \{x \in E \mid d(x,F) \leq \delta_k\}$ und alle $k \in \mathbb{N}$. Wegen $F_k \searrow \{x \in E \mid d(x,F) = 0\} = F$ folgt

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F_k) = \mu(F_k) \longrightarrow \mu(F), \quad k \to \infty.$$

$(d) \Longrightarrow (a)$: Sei $f \in C_b(E)$. Es reicht zu zeigen, dass

$$\limsup_{n\to\infty} \int_E f(x) d\mu_n(x) \leq \int_E f(x) d\mu(x).$$

Denn, daraus folgt für $(-f)$ anstelle von f

$$-\liminf_{n\to\infty} \int_E f(x) d\mu_n(x) = \limsup_{n\to\infty} \int_E (-f(x)) d\mu_n(x) \leq \int_E (-f(x)) d\mu(x)$$

und somit

$$\liminf_{n\to\infty} \int_E f(x) d\mu_n(x) \geq \int_E f(x) d\mu(x).$$

Für $f = 0$ ist nichts zu zeigen. Sei also $f \neq 0$. Wir können ohne Einschränkung annehmen, dass $0 \leq f \leq 1$. Anderenfalls betrachten wir $\frac{f+\|f\|_\infty}{2\|f\|_\infty}$. Sei $k \in \mathbb{N}$ fest und für $i \in \mathbb{N}$ definiere

$$F_i := \left\{f \geq \frac{i}{k}\right\}.$$

Dann ist F_i abgeschlossen, $F_{i+1} \subset F_i$ und es gilt

$$F_i \backslash F_{i+1} = \left\{\frac{i}{k} \leq f < \frac{i+1}{k}\right\}. \tag{4.6}$$

Lemma 4.8. *Es gilt*

$$\frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{F_i} = \sum_{i=1}^{k} \frac{i}{k} \mathbb{1}_{F_i \backslash F_{i+1}} \leq f \leq \sum_{i=0}^{k} \frac{i+1}{k} \mathbb{1}_{F_i \backslash F_{i+1}} = \frac{1}{k} + \frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{F_i}.$$

Beweis. Es gilt $E = \bigcup_{i=0}^{k} F_i \backslash F_{i+1}$ mit $F_0 = E$ und $F_{k+1} = \varnothing$, wobei die Vereinigung disjunkt ist. Daraus, und aus (4.6) ergibt sich die mittlere Ungleichung. Für die beiden Identitäten beachte

$$\begin{aligned}\sum_{i=0}^{k}(i+1)\mathbb{1}_{F_i\backslash F_{i+1}} &= \sum_{i=0}^{k}(i+1)\mathbb{1}_{F_i} - \sum_{i=0}^{k-1}(i+1)\mathbb{1}_{F_{i+1}} \\ &= \sum_{i=0}^{k}\mathbb{1}_{F_i} + \sum_{i=0}^{k} i\mathbb{1}_{F_i} - \sum_{i=1}^{k} i\mathbb{1}_{F_i} = 1 + \sum_{i=1}^{k}\mathbb{1}_{F_i}.\end{aligned}$$

Dieses impliziert die rechte Identität. Für die linke Identität beachte $1 = \sum_{i=0}^{k} \mathbb{1}_{F_i\backslash F_{i+1}}$ woraus mit obiger Rechnung $\sum_{i=1}^{k} i\mathbb{1}_{F_i\backslash F_{i+1}} = \sum_{i=1}^{k}\mathbb{1}_{F_i}$ folgt. □

Die Behauptung folgt aus

$$\begin{aligned}\limsup_{n\to\infty}\int_E f d\mu_n - \frac{1}{k} &\le \frac{1}{k}\limsup_{n\to\infty}\sum_{i=1}^{k}\mu_n(F_i) \\ &\le \frac{1}{k}\sum_{i=1}^{k}\limsup_{n\to\infty}\mu_n(F_i) \le \frac{1}{k}\sum_{i=1}^{k}\mu(F_i) \le \int_E f d\mu.\end{aligned}$$

□

Das folgende einfache Korollar zeigt, dass Einschränkungen auf Teilmengen die schwache Konvergenz erhalten.

Korollar 4.9. *Sei (E, d) ein metrischer Raum und $E' \subset E$ abgeschlossen. Dann ist (E', d') ein metrischer Raum mit Borel-σ-algebra*

$$\mathcal{B}(E') = \sigma(A \subset E' \mid A \textit{ ist offen }) = \mathcal{B}(E) \cap E' := \{A \cap E' \mid A \in \mathcal{B}(E)\}.$$

Sei $\mu_n \longrightarrow \mu$ schwach in E. Dann gilt auch $\mu_n|_{\mathcal{B}(E')} \longrightarrow \mu|_{\mathcal{B}(E')}$ schwach in E'.

Beweis. Der erste Teil der Behauptung ist Übung. Wir zeigen nur die schwache Konvergenz. Sei $F \subset E'$ abgeschlossen, dann ist auch $F \subset E$ abgeschlossen in E. Folglich gilt

$$\limsup_{n\to\infty}\mu_n|_{\mathcal{B}(E')}(F) = \limsup_{n\to\infty}\mu_n(F) \le \mu(F) = \mu_{\mathcal{B}(E')}(F).$$

□

Satz 4.10. *Seien (E, d) und (E', d') metrische Räume, $\mu_n, \mu \in \mathcal{P}(E)$ und $h : E \longrightarrow E'$ stetig. Angenommen $\mu_n \longrightarrow \mu$ schwach, dann ist $\mu_n \circ h^{-1} \longrightarrow \mu \circ h^{-1}$ schwach auf E'.*

Beweis. Sei $f \in C_b(E')$. Dann ist $f \circ h \in C_b(E)$. □

Beispiel 4.11. *Auf die Voraussetzung, dass h stetig ist kann im Allgemeinen nicht verzichtet werden. Denn sei $(x_n)_{n\in\mathbb{N}} \subset E$ mit $x_n \longrightarrow x$ und $x_n \neq x$. Dann gilt $\delta_{x_n} \longrightarrow \delta_x$ schwach. Wähle $h(y) = \mathbb{1}_{\{x\}}(y)$, dann gilt*

$$\delta_{x_n} \circ h^{-1}(A) = \delta_{x_n}(h^{-1}(A)) = \begin{cases} 1, & h(x_n) \in A \\ 0, & h(x_n) \notin A \end{cases} = \delta_{h(x_n)}(A) = \delta_0(A).$$

Analog zeigen wir $\delta_x \circ h^{-1} = \delta_{h(x)} = \delta_1$.

Wir müssen daher eine Bedingung an die Stetigkeitsstellen von h stellen.

Satz 4.12. *Seien (E, d), (E', d') metrische Räume, $h : E \longrightarrow E'$ messbar und $\mu_n, \mu \in \mathcal{P}(E)$ mit $\mu_n \longrightarrow \mu$ schwach. Sei*

$$D_h := \{x \in E \mid h \text{ ist nicht stetig in } x\}.$$

Gilt $\mu(D_h) = 0$, so folgt $\mu_n \circ h^{-1} \longrightarrow \mu \circ h^{-1}$ schwach.

Beweis. Wegen

$$D_h := \bigcup_{n\geq 1} \bigcap_{m\geq 1} \left\{ x \in E \mid \exists y \in E : \ d(x, y) < \frac{1}{m} \ d'(h(y), h(x)) \geq \frac{1}{n} \right\}$$

folgt $D_h \in \mathcal{B}(E)$. Sei $F \subset E'$ abgeschlossen. Dann gilt

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

Sei $x \in \overline{h^{-1}(F)}$, dann gibt es eine Folge $(x_n)_{n\in\mathbb{N}} \subset h^{-1}(F)$ mit $x_n \longrightarrow x$. Ist h stetig in x, so folgt $h(x_n) \longrightarrow h(x)$ und da F abgeschlossen ist, folgt aus $h(x_n) \in F$ bereits $h(x) \in F$. In diesem Fall ist $x \in h^{-1}(F)$. Falls h nicht stetig in x ist, so ist $x \in D_h$ und wir erhalten $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$. Daraus folgt aus der schwachen Konvergenz $\mu_n \longrightarrow \mu$ und dem Satz von Portmanteau

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

Nach dem Satz von Portmanteau impliziert dieses die Behauptung. □

Die nächste Aussage (ohne Beweis) ist gelegentlich nützlich.

Bemerkung 4.13. *Sei $(\mu_n)_{n\geq 1} \subset \mathcal{P}(E)$ und $\mu \in \mathcal{P}(E)$. Dann ist $\mu_n \longrightarrow \mu$ schwach genau dann, wenn*

$$\int_E h(x) d\mu_n(x) \longrightarrow \int_E h(x) d\mu(x), \quad n \to \infty$$

für alle beschränkten messbaren Funktionen h mit

$$\mu\left(\{x \in E \mid h \text{ ist stetig in } x\}\right) = 1.$$

Der Beweis vom nachfolgenden Satz basiert auf [KS07, Theorem 8.5].

Satz 4.14. *Seien $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R})$ und $\mu \in \mathcal{P}(\mathbb{R})$ gegeben und bezeichne mit $F_n(t) = \mu_n((-\infty, t])$ und $F(t) = \mu((-\infty, t])$ die dazugehörigen Verteilungsfunktionen. Dann sind äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) $F_n(t) \longrightarrow F(t)$ für $n \to \infty$ und alle Stetigkeitsstellen t von F.

Beweis. (a) $\Longrightarrow$ (b) Sei $t \in \mathbb{R}$ ein Stetigkeitspunkt von F. Definiere $f(s) = \mathbb{1}_{(-\infty,t]}(s)$ sowie für $\delta > 0$

$$f_\delta^+(s) = \begin{cases} 1, & s \leq t \\ 1 - \frac{s-t}{\delta}, & t < s \leq t + \delta \\ 0, & t + \delta < s \end{cases}$$

$$f_\delta^-(s) = \begin{cases} 1, & s \leq t - \delta \\ 1 - \frac{s-t+\delta}{\delta}, & t - \delta < s \leq t \; . \\ 0, & t < s \end{cases}$$

Dann sind $f_\delta^\pm \in C_b(\mathbb{R})$ und es gilt $f_\delta^- \leq f \leq f_\delta^+$. Sei $\varepsilon > 0$ fest. Wegen (a) finden wir $n_0 \in \mathbb{N}$ derart, dass für $n \geq n_0$

$$F_n(t) = \int_{\mathbb{R}} f(s) d\mu_n(s) \leq \int_{\mathbb{R}} f_\delta^+(s) d\mu_n(s) \leq \int_{\mathbb{R}} f_\delta^+(s) d\mu(s) + \frac{\varepsilon}{2}.$$

Da t ein Stetigkeitspunkt von F ist gibt es ein $\delta > 0$ mit $|F(t \pm \delta) - F(t)| \leq \frac{\varepsilon}{2}$ und somit gilt

$$\int_{\mathbb{R}} f_\delta^+(s) d\mu(s) \leq F(t + \delta) \leq F(t) + \frac{\varepsilon}{2}.$$

Wir haben gezeigt $F_n(t) \leq F(t) + \varepsilon$ für $n \geq n_0$. Analog erhalten wir

$$F_n(t) = \int_{\mathbb{R}} f(s) d\mu_n(s) \geq \int_{\mathbb{R}} f_\delta^-(s) d\mu_n(s) \geq \int_{\mathbb{R}} f_\delta^-(s) d\mu(s) - \frac{\varepsilon}{2} \geq F(t - \delta) - \frac{\varepsilon}{2} \geq F(t) - \varepsilon.$$

Damit folgt $|F_n(t) - F(t)| \leq \varepsilon$ für $n \geq n_0$.
(b) $\Longrightarrow$ (a) Sei $f \in C_b(\mathbb{R})$ und $\varepsilon > 0$ beliebig. Ohne Einschränkung nehmen wir an, dass $\|f\|_\infty \neq 0$ gilt. Da F rechtsstetig ist, hat es nur abzählbar viele Unstetigkeitsstellen. Seien x_*, x^* zwei Punkte in denen F stetig ist mit

$$F(x_*) \leq \frac{\varepsilon}{10\|f\|_\infty}, \qquad F(x^*) \geq 1 - \frac{\varepsilon}{10\|f\|_\infty}. \tag{4.7}$$

Nach Voraussetzung (b) finden wir $n_0 \in \mathbb{N}$ sodass für alle $n \geq n_0$

$$F_n(x_*) \leq |F_n(x_*) - F(x_*)| + F(x_*) \leq \frac{\varepsilon}{5\|f\|_\infty} \tag{4.8}$$

und analog $F_n(x^*) \geq 1 - \frac{\varepsilon}{5\|f\|_\infty}$ gilt. Da f stetig ist, ist es gleichmässig stetig auf $[x_*, x^*]$ und folglich können wir $(x_*, x^*]$ in endlich viele Intervalle $(x_0, x_1], (x_1, x_2], \ldots, (x_{m-1}, x_m]$ unterteilen derart, dass $x_0 = x_*$, $x_m = x^*$,

$$|f(s) - f(x_j)| \leq \frac{\varepsilon}{10}, \quad s \in (x_j, x_{j+1}], \quad j = 1, \ldots, m-1 \tag{4.9}$$

und die x_j Stetigkeitspunkte von F sind. Nun betrachten wir

$$\begin{aligned}\left|\int_{\mathbb{R}} f(s)d\mu_n(s) - \int_{\mathbb{R}} f(s)d\mu(s)\right| &\leq \int_{(-\infty,x_*]} |f(s)|d\mu_n(s) + \int_{(-\infty,x_*]} |f(s)|d\mu(s)\\ &+ \int_{(x^*,\infty)} |f(s)|d\mu_n(s) + \int_{(x^*,\infty)} |f(s)|d\mu(s) + \left|\int_{(x_*,x^*]} f(s)d\mu_n(s) - \int_{(x_*,x^*]} f(s)d\mu(s)\right|.\end{aligned}$$

Aus (4.7) folgt

$$\int_{(-\infty,x_*]} |f(s)|d\mu(s) + \int_{(x^*,\infty)} |f(s)|d\mu(s) \leq \|f\|_\infty \left(F(x_*) + 1 - F(x^*)\right) \leq \frac{2}{5}\varepsilon.$$

Ist $n \geq n_0$ hinreichend gross so erhalten wir aus (4.8)

$$\int_{(-\infty,x_*]} |f(s)|d\mu_n(s) + \int_{(x^*,\infty)} |f(s)|d\mu_n(s) \leq \|f\|_\infty \left(F_n(x_*) + 1 - F_n(x^*)\right) \leq \frac{2}{5}\varepsilon.$$

Für den letzten Term erhalten wir aus (4.9)

$$\left|\int_{(x_*,x^*]} f(s)d\mu_n(s) - \int_{(x_*,x^*]} f(s)d\mu(s)\right| \leq \frac{\varepsilon}{5} + \left|\int_{(x_*,x^*]} f_\varepsilon(s)d\mu_n(s) - \int_{(x_*,x^*]} f_\varepsilon(s)d\mu(s)\right|$$

wo $f_\varepsilon(s) = \sum_{j=1}^{m} f(x_j)\mathbb{1}_{(x_j,x_{j+1}]}(s)$. Dann gilt

$$\begin{aligned}\left|\int_{(x_*,x^*]} f_\varepsilon(s)d\mu_n(s) - \int_{(x_*,x^*]} f_\varepsilon(s)d\mu(s)\right| &\leq \sum_{j=1}^{m-1}\left|\int_{(x_j,x_{j+1}]} f_\varepsilon(s)d\mu_n(s) - \int_{(x_j,x_{j+1}]} f_\varepsilon(s)d\mu(s)\right|\\ &= \sum_{j=1}^{m-1} |f(x_j)|\, |F_n(x_{j+1}) - F_n(x_j) - F(x_{j+1}) + F(x_j)|\end{aligned}$$

und somit

$$\lim_{n\to\infty}\left|\int_{(x_*,x^*]} f_\varepsilon(s)d\mu_n(s) - \int_{(x_*,x^*]} f_\varepsilon(s)d\mu(s)\right| = 0.$$

Insgesamt haben wir gezeigt

$$\lim_{n\to\infty}\left|\int_{\mathbb{R}} f(s)d\mu_n(s) - \int_{\mathbb{R}} f(s)d\mu(s)\right| \leq \varepsilon.$$

Da ε beliebig war, folgt die Behauptung. □

4.2 Charakterisierung der kompakten Mengen in $\mathcal{P}(E)$

Als nächstes wollen wir eine Charakterisierung der kompakten Mengen in $\mathcal{P}(E)$ geben. Man beachte, dass wir keine Metrik auf $\mathcal{P}(E)$ definiert haben und folglich keinerlei Kompaktheit definiert ist.

Definition 4.15. *Sei $\Gamma \subset \mathcal{P}(E)$ eine Familie von Wahrscheinlichkeitsmaßen.*

(a) Γ heißt relativ kompakt, falls jede Folge in Γ eine schwach konvergente Teilfolge hat.

(b) Γ heißt straff, falls für alle $\varepsilon > 0$ es eine kompakte Menge $K_\varepsilon \subset E$ gibt mit

$$\mu(K_\varepsilon) \geq 1 - \varepsilon, \quad \forall \mu \in \Gamma.$$

Bemerkung 4.16. *(i) $\{\mu\}$ ist genau dann straff, wenn μ regulär ist.*

(ii) Es gelte eine der Eigenschaften

- *E ist kompakt*
- *Es gibt eine Folge von kompakten Mengen $(K_n)_{n\in\mathbb{N}} \subset E$ mit $\bigcup_{n\geq 1} K_n = E$*
- *E ist eine vollständiger, seperabler metrischer Raum.*

Dann ist $\Gamma = \{\mu\}$ für jedes $\mu \in \mathcal{P}(E)$ straff.

(iii) $\Gamma = \mathcal{P}(\mathbb{R}^d)$ ist nicht straff.

Das nächste Lemma bietet ein Kriterium für Straffheit.

Lemma 4.17. *Sei E ein Polnischer Raum mit Metrik d. Dann ist $\Gamma \subset \mathcal{P}(E)$ genau dann straff, wenn es für jedes $\varepsilon > 0$ und jedes $r > 0$ ein $n \in \mathbb{N}$ und $a_1, \dots, a_n \in E$ gibt mit*

$$\mu\left(\bigcup_{i=1}^{n} B_r(a_i)\right) \geq 1 - \varepsilon, \quad \mu \in \Gamma. \tag{4.10}$$

Beweis. Sei Γ straff, $\varepsilon > 0$ und $r > 0$. Sei $K \subset E$ kompakt mit $\mu(K) \geq 1 - \varepsilon$ für alle $\mu \in \Gamma$. Nach Lemma 3.8 gibt es $n \in \mathbb{N}$ sowie $a_1, \dots, a_n \in E$ mit

$$K \subset \bigcup_{i=1}^{n} B_r(a_i).$$

Daraus folgt (4.10).

Es gelte Umgekehrt (4.10). Wähle für jedes $k \geq 1$ Punkte a_i^k mit $i = 1, \ldots, n_k$ derart dass

$$\mu\left(\bigcup_{i=1}^{n_k} B_{\frac{1}{k}}(a_i^k)\right) \geq 1 - \frac{\varepsilon}{2^k}.$$

Sei $A = \bigcap_{k=1}^{\infty} \bigcup_{i=1}^{n_k} B_{\frac{1}{k}}(a_i^k)$. Analog zum Beweis von Lemma 3.12 lässt sich zeigen, dass A abgeschlossen sowie total beschränkt, also nach Lemma 3.8 kompakt ist. Ferner lässt sich zeigen, dass $\mu(\overline{A}) \geq \mu(A) \geq 1 - \varepsilon$ für $\mu \in \Gamma$. □

Der nächste Satz stellt den Zusammenhang zu Straffheit und relativer Kompaktheit her.

Satz 4.18. *(Satz von Prokhorov) Sei E ein metrischer Raum. Ist $\Gamma \subset \mathcal{P}(E)$ straff, so ist Γ relativ kompakt. Ist E zusätzlich ein Polnischer Raum, so ist Γ genau dann straff, wenn es relativ kompakt ist.*

Beweis. Sei zunächst Γ straff und $(\mu_n)_{n\in\mathbb{N}} \subset \Gamma$ beliebig. Wir zeigen, dass $(\mu_n)_{n\geq 1}$ eine schwach konvergente Teilfolge besitzt.

Fall 1. Sei E kompakt. Dann ist $C_b(E) = C(E)$ seperabel. Folglich gibt es eine dichte Folge $(f_n)_{n\geq 1} \subset C(E)$. Dann ist wegen

$$\left|\int_E f_1(x) d\mu_n(x)\right| \leq \|f_1\|_\infty$$

die Folge $\left(\int_E f_1 d\mu_n\right)_{n\geq 1}$ beschränkt und folglich gibt es eine Teilfolge $(n_k^{(1)})_{k\in\mathbb{N}}$ sodass

$$\int_E f_1(x) d\mu_{n_k^{(1)}}(x) \longrightarrow I(f_1), \quad k \to \infty.$$

Auf diesselbe Weise lässt sich zeigen, das auch $\left(\int_E f_2 d\mu_{n_k^{(1)}}\right)_{k\geq 1}$ beschränkt ist. Folglich gibt es eine Teilfolge $(n_k^{(2)})_{k\geq 1}$ sodass

$$\int_E f_2(x) d\mu_{n_k^{(2)}}(x) \longrightarrow I(f_2), \quad k \to \infty.$$

Iteration dieses Verfahrens liefert Folgen $(n_k^{(l)})_{k\geq 1}$ mit $(n_k^{(l+1)})_{k\geq 1} \subset (n_k^{(l)})_{k\geq 1}$ sodass

$$\int_E f_l(x) d\mu_{n_k^{(l)}}(x) \longrightarrow I(f_l), \quad k \to \infty, \quad \forall l \geq 1.$$

Setze $m_k := n_k^{(k)}$, dann gilt per Konstruktion

$$\int_E f_l(x) d\mu_{m_k}(x) \longrightarrow I(f_l), \quad k \to \infty, \quad \forall l \geq 1.$$

Lemma 4.19. *Es gibt eine eindeutige lineare Fortsetzung von I auf $C(E)$ welche wir wieder mit I bezeichnen und welche die Bedingungen $|I(f)| \leq \|f\|_\infty$ sowie*

$$\int_E f(x) d\mu_{m_k}(x) \longrightarrow I(f), \quad k \to \infty$$

für alle $f \in C(E)$ erfüllt. Ferner erfüllt diese Fortsetzung $I(f) \geq 0$ für alle $0 \leq f \in C(E)$.

Beweis. Für $f \in C(E)$ setze

$$I_k(f) := \int_E f(x) d\mu_{m_k}(x).$$

Dann gilt $\lim_{k\to\infty} I_k(f_l) = I(f_l)$ für $l \geq 1$. Seien $f \in C(E)$ und $\varepsilon > 0$ beliebig und wähle $l \geq 1$ derart dass $\|f - f_l\|_\infty \leq \varepsilon$. Dann gilt

$$\begin{aligned} |I_k(f) - I_j(f)| &\leq |I_k(f) - I_k(f_l)| + |I_k(f_l) - I_j(f_l)| + |I_j(f_l) - I_j(f)| \\ &\leq 2\|f - f_l\|_\infty + |I_k(f_l) - I_j(f_l)| \\ &\leq 2\varepsilon + |I_k(f_l) - I_j(f_l)| \longrightarrow 0, \quad k, j \to \infty. \end{aligned}$$

Also ist $(I_k(f))_{k\in\mathbb{N}}$ eine Cauchy Folge for alle $f \in C(E)$. Dann definiert $I(f) := \lim_{k\to\infty} I_k(f)$ eine lineare Fortsetzung von I. Diese erfüllt

$$|I(f)| = |\lim_{k\to\infty} I_k(f)| \leq \|f\|_\infty.$$

Für die Eindeutigkeit seien I, I' zwei solche Fortsetzungen, $f \in C(E)$ und $\varepsilon > 0$. Wähle $l \geq 1$ derart dass $\|f - f_l\|_\infty \leq \varepsilon$ gilt und sei $k \geq 1$ derart dass

$$|I(f_l) - I_k(f_l)| + |I_k(f_l) - I'(f_l)| \leq \varepsilon.$$

. Daraus folgt

$$|I(f) - I'(f)| \leq |I(f) - I(f_l)| + |I(f_l) - I_k(f_l)| + |I_k(f_l) - I'(f_l)| + |I'(f_{l_j}) - I'(f)| \leq 3\varepsilon$$

welches die Eindeutigkeit zeigt. Ist $f \geq 0$, so gilt

$$I(f) = \lim_{k\to\infty} I_k(f) \geq 0.$$

$\square$

Der Rieszsche-Darstellungssatz (siehe Appendix) liefert die Existenz von einem endliche Maß μ auf E mit

$$I(f) = \int_E f(x) d\mu(x), \quad f \in C(E).$$

Folglich gilt

$$\lim_{k\to\infty} \int_E f(x) d\mu_{m_k}(x) = I(f) = \int_E f(x) d\mu(x), \quad f \in C(E).$$

und da die konstante Funktion $f = 1$ stetig ist, gilt auch $\mu(E) = I(1) = \lim_{k\to\infty} \mu_{m_k}(E) = 1$.
Schritt 2. Angenommen es gibt einen kompakten metrischen Raum (E', d') mit $E \in \mathcal{B}(E')$ derart dass E die Unterraumtopologie trägt. D.h. $A \subset E$ ist offen genau dann, wenn es eine offene Menge $B \subset E'$ gibt mit $A = B \cap E$. Wir erweitern μ_n auf E' durch

$$\mu_n(E' \backslash E) = 0, \quad n \geq 1.$$

Dann ist $(\mu_n)_{n\geq 1} \subset \mathcal{P}(E')$ und nach Schritt 1 gibt es eine Teilfolge $(n_k)_{k\geq 1}$ und ein $\mu \in \mathcal{P}(E')$ mit $\mu_{n_k} \longrightarrow \mu$ schwach bezüglich (E', d'). Nach Voraussetzung gibt es zu jedem $m \geq 1$ eine kompakte Menge $K_m \subset E$ mit $\mu_n(K_m) \geq 1 - \frac{1}{m}$, für $n \geq 1$. Es lässt sich zeigen, dass K_m auch kompakt in (E', d') ist. Nach dem Satz von Portmanteau folgt

$$\mu(K_m) \geq \limsup_{k\to\infty} \mu_{n_k}(K_m) \geq 1 - \frac{1}{m}.$$

Daraus folgt $1 = \mu\left(\bigcup_{m\geq 1} K_m\right) \leq \mu(E) \leq 1$, also $\mu(E) = 1$. Sei $G \subset E$ offen und sei $B \subset E'$ offen mit $G = B \cap E$. Dann gilt

$$\liminf_{k\to\infty} \mu_{n_k}(G) = \liminf k \to \infty \mu_{n_k}(B) \geq \mu(B) = \mu(G)$$

und aus dem Satz von Portmanteau folgt $\mu_{n_k} \longrightarrow \mu$ schwach in E.
Schritt 3. Sei (E, d) ein allgemeiner metrischer Raum. Nach vorhergehender Argumentation gibt es zu jedem $m \geq 1$ eine kompakte Menge $K_m \subset E$ mit $\mu_n(K_m) \geq 1 - \frac{1}{m}$, für $n \geq 1$. Folglich gilt

$$\mu_n\left(\bigcup_{m\geq 1} K_m\right) = 1, \quad n \geq 1.$$

Also können wir ohne Einschränkung der Allgemeinheit annehmen, dass (E, d) σ-kompakt und somit auch seperabel ist. Um Schritt 2 anzuwenden müssen wir eine Einbettung in einen kompakten metrischen Raum (E', d') konstruieren. Ersetzen wir d durch $d \wedge 1$ so ändert dieses nicht die offenen Mengen in E; ohne Einschränkung sei also $d \leq 1$. Sei $E' = [0, 1]^{\mathbb{N}}$ und

$$d'(a, b) = \sum_{j=1}^{\infty} 2^{-j} \frac{|a_j - b_j|}{1 + |a_j - b_j|}, \quad a = (a_j)_{j\geq 1}, b = (b_j)_{j\geq 1} \in E'.$$

Da $[0, 1]$ kompakt ist und d' die Produkttopologie erzeugt, lässt sich zeigen, dass (E', d') ein kompakter metrischer Raum ist. Sei $(z_k)_{k\geq 1} \subset E$ eine dichte Teilmenge und definiere

$$I : E \longrightarrow E', \quad x \longmapsto (d(x, z_k))_{k\geq 1}.$$

Dann hat I die folgenden Eigenschaften

- I ist stetig und injektiv.
- I bildet offene Mengen auf offene Mengen ab.

Definiere $\mu_n' := \mu_n \circ I^{-1}$ auf (E', d'). Sei $(K_m)_{m\geq 1}$ wie zu Beginn von Schritt 3 und setze $K_m' := I(K_m)$. Dann sind $K_m' \subset E'$ kompakt und es gilt

$$\mu_n'(K_m') = \mu_n(K_m) \geq 1 - \frac{1}{m}, \quad n \geq 1.$$

Nach Schritt 2 gibt es eine Teilfolge $(\mu_{n_k}')_{k\geq 1}$ und $\mu' \in \mathcal{P}(E')$ mit $\mu_{n_k}' \longrightarrow \mu'$ schwach in E'. Aus Portmanteau folgt

$$\mu'\left(K_m'\right) \geq \limsup_{k\to\infty} \mu_{n_k}'(K_m') \geq 1 - \frac{1}{m}$$

und somit wegen $\bigcup_{m\geq 1} K_m' \subset I(E)$ auch $1 \leq \mu'\left(\bigcup_{m\geq 1} K_m'\right) \leq \mu'(I(E)) \leq 1$ bereits $\mu'(I(E)) = 1$. Folglich gibt es ein $\mu \in \mathcal{P}(E)$ mit $\mu' = \mu \circ I^{-1}$. Sei $f \in C_b(E)$, dann ist

$$\int_E f(x) d\mu_{n_k}(x) = \int_{E'} f(I^{-1}(x)) d\mu_{n_k}'(x) \longrightarrow \int_{E'} f(I^{-1}(x)) d\mu'(x) = \int_E f(x) d\mu(x)$$

wobei wir benutzt haben, dass $f \circ I^{-1}$ stetig ist. Dieses zeigt den ersten Teil der Behauptung.

Umgekehrt sei Γ relativ kompakt, wir zeigen dass Γ straff ist. Angenommen Γ ist nicht straff. Dann gibt es nach Lemma 4.17 ein $\varepsilon > 0$ und $r > 0$ sodass für alle $n \geq 1$ und alle $a_1, \dots, a_n \in E$ es ein $\mu \in \Gamma$ gibt mit

$$\mu\left(\bigcup_{i=1}^n U_r(a_i)\right) \leq \mu\left(\bigcup_{i=1}^n B_r(a_i)\right) \leq 1 - \varepsilon.$$

Da E seperabel ist, gibt es es eine Folge $(a_i)_{i\in\mathbb{N}} \subset E$ mit $E = \bigcup_{i\geq 1} U_r(a_i)$. Setze $A_n = \bigcup_{i=1}^n U_r(a_i)$. Wir finden daher eine Folge $(\mu_n)_{n\in\mathbb{N}} \subset \Gamma$ mit

$$\mu_n(A_n) \leq 1 - \varepsilon, \quad n \geq 1.$$

Da Γ relativ kompakt ist gibt es eine Teilfolge μ_{n_k} welche schwach gegen ein μ konvergiert. Da A_n offen ist und da $A_n \subset A_{n_k}$ für hinreichend grosse k, folgt aus Portmanteau

$$\mu(A_n) \leq \liminf_{k\to\infty} \mu_{n_k}(A_n) \leq \liminf_{k\to\infty} \mu_{n_k}(A_{n_k}) \leq 1 - \varepsilon.$$

Dieses wiederspricht $\bigcup_{n\geq 1} A_n = E$. □

Als Korollar können wir zeigen, dass

Korollar 4.20. *Sei (E, d) ein seperabler metrischer Raum.*

(a) Für $f \in C_b(E \times E)$ ist die Abbildung

$$\mathcal{P}(E) \times E \ni (\mu, y) \longmapsto I(\mu, y) = \int_E f(x, y) d\mu(x)$$

stetig und beschränkt, d.h. $I \in C_b(\mathcal{P}(E) \times E)$.

(b) Sei $f \in C_b(\mathcal{P}(E) \times E)$. Dann ist die Abbildung

$$J(\nu) = \int_E f(\mu, x) d\mu(x), \quad \mu \in \mathcal{P}(E)$$

stetig und beschränkt, d.h. $J \in C_b(\mathcal{P}(E))$.

Beweis. (a) Sei $\mu_n \longrightarrow \mu$ schwach und $y_n \longrightarrow y$. Wir müssen zeigen dass $I(\mu_n, y_n) \longrightarrow I(\mu, y)$. Es gilt $I(\mu_n, y) \longrightarrow I(\mu, y)$, folglich reicht es wegen

$$|I(\mu_n, y_n) - I(\mu, y)| \leq |I(\mu_n, y_n) - I(\mu_n, y)| + |I(\mu_n, y) - I(\mu, y)| \tag{4.11}$$

zu zeigen, dass $|I(\mu_n, y_n) - I(\mu_n, y)| \longrightarrow 0$ gilt. Da $\mu_n \longrightarrow \mu$ schwach, ist $(\mu_n)_{n\geq 1}$ nach dem Satz von Prokhorov straff. Sei $\varepsilon > 0$ und $K_\varepsilon \subset E$ kompakt mit $\mu_n(K_\varepsilon) \geq 1 - \varepsilon$ für $n \geq 1$. Da f stetig ist gibt es $n_0(\varepsilon) \in \mathbb{N}$ mit

$$\sup_{x \in K_\varepsilon} |f(x, y_n) - f(x, y)| < \varepsilon, \quad n \geq n_0(\varepsilon).$$

Für $n \geq n_0$ erhalten wir folglich

$$\begin{aligned} |I(\mu_n, y_n) - I(\mu_n, y)| &\leq \left| \int_E \mathbb{1}_{K_\varepsilon}(x)(f(x, y_n) - f(x, y)) d\mu_n(x) \right| \\ &\quad + \left| \int_E \mathbb{1}_{K_\varepsilon^c}(x)(f(x, y_n) - f(x, y)) d\mu_n(x) \right| \\ &\leq \sup_{x \in K_\varepsilon} |f(x, y_n) - f(x, y)| + 2\|f\|_\infty \mu_n(K_\varepsilon^c) \leq (1 + 2\|f\|_\infty)\varepsilon. \end{aligned}$$

(b) Betrachte eine ähliche Zerlegung zu (4.11). Die Details sind Übung. □

4.3 Levy-Prokhorov Metrik

Definition 4.21. *Es sei $\rho : \mathcal{P}(E) \times \mathcal{P}(E) \longrightarrow [0, \infty)$ gegeben durch*

$$\rho(\mu, \nu) := \inf\{\varepsilon > 0 \mid \mu(F) \leq \nu(F^\varepsilon) + \varepsilon, \quad F \text{ ist abgeschlossen } \}$$

mit $F^\varepsilon := \{y \in E \mid d(y, F) < \varepsilon\}$. Die Abbildung ρ wird Levy-Prokhorov Metrik genannt.

Lemma 4.22. *[EK86] Die Prokhorov Metrik ρ definiert eine Metrik auf E.*

Satz 4.23. *[EK86] Sei E ein seperabler metrischer Raum und $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$, $\mu \in \mathcal{P}(E)$. Dann sind die folgenden Aussagen äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) $\rho(\mu_n, \mu) \longrightarrow 0$ für $n \to \infty$.

Satz 4.24. *[EK86] Es gelten die folgenden Aussagen.*

(a) Ist E seperabel, so ist auch $\mathcal{P}(E)$ seperabel.

(b) Ist E vollständig, so ist auch $\mathcal{P}(E)$ volltändig.

Insbesondere, ist E ein Polnischer Raum, so ist auch $\mathcal{P}(E)$ ein Polnischer Raum.

4.4 Beschräkte Lipschitz Metrik

Sei $BL(E)$ der Raum aller Funktionen f mit endlicher Norm

$$\|f\|_{BL} = \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}.$$

Dann ist $BL(E)$ ein Banachraum und es gilt $BL(E) \subset C_b(E)$.

Definition 4.25. *Definiere eine Abbildung $\rho : \mathcal{P}(E) \times \mathcal{P}(E) \longrightarrow [0, \infty)$ durch*

$$\rho(\mu, \nu) = \sup \left\{ \left| \int_E f(x) d\mu(x) - \int_E f(x) d\nu(x) \right| \ : \ f \in BL(E), \ \|f\|_{BL} \leq 1 \right\}.$$

Es ist nicht schwer zu zeigen, dass ρ symmetrisch ist und die Dreiecksungleichung erfüllt. Ferner ist $\rho(\mu, \nu) = 0$ genau dann, wenn

$$\int_E f(x) d\mu(x) = \int_E f(x) d\nu(x), \quad \forall f \in C_b(E).$$

Also genau dann, wenn $\mu = \nu$. Der nächste Satz zeigt, dass $(\mathcal{P}(E), \rho)$ ein Polnischer Raum ist, sofern (E, d) ein Polnischer Raum war.

Satz 4.26. *Sei (E, d) ein metrischer Raum.*

- *Ist (E, d) seperabel, so ist auch $(\mathcal{P}(E), \rho)$ seperabel.*
- *Ist (E, d) vollständig, so ist auch $(\mathcal{P}(E), \rho)$ vollständig.*

Beweis. Siehe [EK86]. □

Zum Schluss wollen wir die Konvergenz bezüglich ρ besser verstehen. Klar gilt

$$\left| \int_E h(x) d\nu(x) - \int_E h(x) d\mu(x) \right| \leq \rho(\nu, \mu), \quad \forall h \in BL(E) \tag{4.12}$$

für alle $\mu, \nu \in \mathcal{P}(E)$. Folglich impliziert $\rho(\mu_n, \mu) \longrightarrow 0$ bereits $\mu_n \longrightarrow \mu$ schwach. Der nachfolgende Satz zeigt, dass für seperable metrische Räume auch die Umkehrung gilt.

Satz 4.27. *Sei (E, d) ein seperabler metrischer Raum und seien $(\mu_n)_{n \geq 1} \subset \mathcal{P}(E)$ sowie $\mu \in \mathcal{P}(E)$. Dann sind äquivalent*

- $\rho(\mu_n, \mu) \longrightarrow 0$
- $\mu_n \longrightarrow \mu$ *schwach.*

Beweis. Die erste Richtung folgt aus (4.12) und dem Satz von Portmanteau. Es sei nun $\mu_n \longrightarrow \mu$ schwach. Nach dem Satz von Prokhorov sind $(\mu_n)_{n\geq 1}$ sowie μ straff und folglich gibt es zu $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subset E$ mit

$$\mu(\mathcal{K}_\varepsilon) + \mu_n(K_\varepsilon) \geq 1 - \varepsilon, \quad n \geq 1.$$

Dann ist (K_ε, d) ein kompakter metrischer Raum. Ferner ist

$$A := \{f \in BL(K_\varepsilon) \mid \|f\|_{BL(K_\varepsilon)} \leq 1\} \subset C(K_\varepsilon)$$

abgeschlossen und nach dem Satz von Arzela-Ascoli (siehe Appendix) kompakt. Also gibt es $f_1, \dots, f_k \in A$ derart, dass für jedes $f \in C(K_\varepsilon)$ mit $\|f\|_{BL(K_\varepsilon)} \leq 1$ ein $j \in \{1, \dots, k\}$ existiert mit

$$\sup_{x \in K_\varepsilon} |f(x) - f_j(x)| < \varepsilon.$$

Sei $f \in BL(E)$ mit $\|f\|_{BL} \leq 1$. Dann ist $f|_{K_\varepsilon} \in BL(K_\varepsilon)$ und es gilt $\|f|_{K_\varepsilon}\|_{BL(K_\varepsilon)} \leq 1$ und folglich erhalten wir

$$\begin{aligned}
\left| \int_E f(x) d\mu_n(x) - \int_E f(x) d\mu(x) \right| &\leq \|f\|_\infty \left(\mu_n(K_\varepsilon^c) + \mu(K_\varepsilon^c)\right) + \left| \int_{K_\varepsilon} f(x) d\mu_n(x) - \int_{K_\varepsilon} f(x) d\mu(x) \right| \\
&\leq 2\varepsilon + \int_{K_\varepsilon} |f(x) - f_j(x)| d\mu_n(x) + \int_{K_\varepsilon} |f(x) - f_j(x)| d\mu(x) \\
&\quad + \left| \int_{K_\varepsilon} f_j(x) d\mu_n(x) - \int_{K_\varepsilon} f_j(x) d\mu(x) \right| \\
&\leq 4\varepsilon + \left| \int_{K_\varepsilon} f_j(x) d\mu_n(x) - \int_{K_\varepsilon} f_j(x) d\mu(x) \right|.
\end{aligned}$$

Nach Korollar 4.9 ist $\mu_n \longrightarrow \mu$ schwach in $\mathcal{P}(K_\varepsilon)$ und damit verschwidet das letzte Integral für $n \to \infty$. □

4.5 Konvergenz von Momenten

Ziel dieses Abschnittes ist es ein analoges Result zu dem Satz von Portmanteau für unbeschränkte Testfunktionen zu beweisen. Sei (E, d) ein metrischer Raum und sei $V : E \longrightarrow \mathbb{R}_+$ stetig. Definiere

$$\mathcal{P}_V(E) = \left\{ \mu \in \mathcal{P}(E) \mid \int_E V(x) d\mu(x) < \infty \right\}.$$

Analog sei $C_V(E)$ der Raum der stetigen Funktionen auf E wo

$$\|f\|_{C_V} = \sup_{x \in E} \frac{|f(x)|}{1 + V(x)}$$

endlich ist. Dann erfüllt jedes $f \in C_V(E)$

$$|f(x)| \leq \|f\|_{C_V}(1 + V(x)), \quad x \in E.$$

Man kann zeigen, dass $(C_V(E), \|\cdot\|_{C_V})$ ein Banachraum ist.

Definition 4.28. *Sei $(\mu_n)_{n\geq 1} \subset \mathcal{P}_V(E)$ sowie $\mu \in \mathcal{P}_V(E)$. Wir sagen $\mu_n \longrightarrow \mu$ in $\mathcal{P}_V(E)$, falls*

$$\int\limits_E f(x)d\mu_n(x) \longrightarrow \int\limits_E f(x)d\mu(x), \quad n \to \infty$$

für alle $f \in C_V(E)$.

Bemerkung 4.29. *Ist V beschränkt, so ist $C_V(E) = C_b(E)$ und $\mathcal{P}_V(E) = \mathcal{P}(E)$. In diesem Fall ist die Konvergenz in $\mathcal{P}_V(E)$ gerade die schwache Konvergenz. Von besonderem Interesse ist für uns der Fall wo V nicht beschränkt ist.*

Beispiel 4.30. *Sei $E = \mathbb{R}^d$, $V(x) = |x|^p$ mit $p \geq 1$. Dann ist $f \in C_V(\mathbb{R}^d)$ genau dann, wenn*

$$|f(x)| \leq C(1 + |x|^p), \quad x \in \mathbb{R}^d$$

für eine Konstante $C > 0$. Ferner ist $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ genau dann, wenn μ endliche Momente bis zur Ordnung p hat. Da $f(x) = |x|^k \in C_V(\mathbb{R}^d)$ für $k \in [1,p]$ impliziert Konvergenz in $\mathcal{P}_V(\mathbb{R}^d)$ die Konvergenz aller Momente bis zur Ordnung p.

Im folgenden wollen wir die folgende Erweiterung vom Satz von Portmanteau beweisen.

Satz 4.31. *Sei (E,d) ein metrischer Raum, $V \in C(E)$ sowie $(\mu_n)_{n\geq 1} \subset \mathcal{P}_V(E)$ und $\mu \in \mathcal{P}_V(E)$. Dann sind äquivalent:*

(a) $\mu_n \longrightarrow \mu$ *in* $\mathcal{P}_V(E)$.

(b) $\mu_n \longrightarrow \mu$ *schwach und*

$$\int\limits_E V(x)d\mu_n(x) \longrightarrow \int\limits_E V(x)d\mu(x), \quad n \to \infty.$$

(c) $\mu_n \longrightarrow \mu$ *schwach und*

$$\limsup_{n\to\infty} \int\limits_E V(x)d\mu_n(x) \leq \int\limits_E V(x)d\mu(x).$$

(d) $\mu_n \longrightarrow \mu$ *schwach und*

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int\limits_E \mathbb{1}_{V(x)\geq R} V(x)d\mu_n(x) = 0.$$

Beweis. *(a)* $\Longrightarrow$ *(b)* $\Longrightarrow$ *(c)*: klar.
(c) $\Longrightarrow$ *(d)* Wir zeigen zunächst

$$\liminf_{n\to\infty} \int_E \mathbb{1}_{V<R}(x)V(x)d\mu_n(x) \geq \int_E \mathbb{1}_{V<R}(x)V(x)d\mu(x). \tag{4.13}$$

Es gilt

$$\int_E \mathbb{1}_{V<R}(x)V(x)d\mu_n(x) = \int_E \mathbb{1}_{V<R}(x)(R \wedge V(x))d\mu_n(x) = a_n\mu_n'(V < R)$$

wo $\mu_n'(A) = \frac{1}{a_n}\int_A (R \wedge V(x))d\mu_n(x)$ und $a_n = \int_E (R \wedge V(x))d\mu_n(x)$. Hierbei können wir ohne Einschränkung der Allgemeinheit annehmen, dass $a_n \neq 0$ (ansonsten ist nichts zu zeigen). Wegen $R \wedge V \in C_b(E)$ folgt $a_n \longrightarrow \int_E (R \wedge V(x))d\mu(x) =: a$ Aus $V \wedge R \in C_b(E)$ erhalten wir für alle $f \in C_b(E)$

$$\int_E f(x)d\mu_n'(x) = \frac{1}{a_n}\int_E f(x)(R \wedge V(x))d\mu_n(x) \longrightarrow \frac{1}{a}\int_E f(x)(R \wedge V(x))d\mu(x)$$

und folglich $\mu_n' \longrightarrow \mu'$ schwach wo $\mu'(A) = \frac{1}{a}\int_A (R \wedge V(x))d\mu(x)$. Da V stetig ist, ist $\{x \in E \mid V(x) < R\}$ offen und nach Portmanteau gilt somit $\liminf_{n\to\infty} \mu_n'(V < R) \geq \mu'(V < R)$. Daraus folgt

$$\liminf_{n\to\infty} a_n\mu_n'(V < R) \geq a\mu'(V < R) = \int_E \mathbb{1}_{V<R}(R \wedge V(x))d\mu(x) = \int_E \mathbb{1}_{V<R}V(x)d\mu(x),$$

d.h. (4.13). Mit Annahme (c) und (4.13) erhalten wir

$$\begin{aligned}
\limsup_{n\to\infty} \int_E \mathbb{1}_{V\geq R}(x)V(x)d\mu_n(x) &= \limsup_{n\to\infty} \left(\int_E V(x)d\mu_n(x) - \int_E \mathbb{1}_{V<R}(x)V(x)d\mu_n(x) \right) \\
&\leq \limsup_{n\to\infty} \int_E V(x)d\mu_n(x) - \liminf_{n\to\infty} \int_E \mathbb{1}_{V<R}(x)V(x)d\mu_n(x) \\
&\leq \int_E V(x)d\mu(x) - \int_E \mathbb{1}_{V<R}(x)V(x)d\mu(x) \\
&= \int_E \mathbb{1}_{V\geq R}(x)V(x)d\mu(x).
\end{aligned}$$

Wegen $\mu \in \mathcal{P}_V(E)$ folgt (d).
(d) $\Longrightarrow$ *(a)* Sei zunächst $f \in C_V(E)$ nicht-negativ. Wir zeigen zunächst: Für alle $\varepsilon > 0$ gibt es ein $R_\varepsilon > 0$ sodass

$$\limsup_{n\to\infty} \left| \int_E f(x)d\mu_n(x) - \int_E (R \wedge f(x))d\mu_n(x) \right| \leq 2\|f\|_{C_V}\varepsilon, \quad R \geq R_\varepsilon. \tag{4.14}$$

Sei $\varepsilon > 0$ fest. Für $R > 0$ schreibe

$$\int_E f(x)d\mu_n(x) = \int_E \mathbb{1}_{V\geq R}(x)f(x)d\mu_n(x) + \int_E \mathbb{1}_{V<R}(x)f(x)d\mu_n(x) = I_1 + I_2.$$

Wegen Annahme (d) finden wir ein $R_\varepsilon > 0$ mit

$$\mu(V \geq R) + \limsup_{n\to\infty} \int_E \mathbb{1}_{V\geq R}(x)V(x)d\mu_n(x) < \varepsilon, \quad R \geq R_\varepsilon.$$

Da $\{V \geq R\}$ abgeschlossen ist, folgt aus Portmanteau $\limsup_{n\to\infty} \mu_n(R \geq V) \leq \mu(R \geq V)$. Also

$$\begin{aligned}\limsup_{n\to\infty} I_1 &\leq \|f\|_{C_V} \limsup_{n\to\infty} \int_E \mathbb{1}_{V\geq R}(x)(1 + V(x))d\mu_n(x)\\ &= \|f\|_{C_V} \limsup_{n\to\infty} \mu_n(V \geq R) + \|f\|_{C_V} \limsup_{n\to\infty} \int_E \mathbb{1}_{V\geq R}(x)V(x)d\mu_n(x). \leq \|f\|_{C_V}\varepsilon.\end{aligned}$$

Für den zweiten Term erhalten wir mit $R' = \|f\|_{C_V}(1 + R)$

$$I_2 = \int_E \mathbb{1}_{V<R}(x)(R' \wedge f(x))d\mu_n(x) = \int_E (R' \wedge f(x))d\mu_n(x) - \int_E \mathbb{1}_{V\geq R}(x)(R' \wedge f(x))d\mu_n(x).$$

Da $R' \wedge f \in C_b(E)$ erhalten wir $\lim_{n\to\infty} \int_E (R' \wedge f(x))d\mu_n(x) = \int_E (R' \wedge f(x))d\mu(x)$. Für den anderen Term gilt

$$\begin{aligned}&\limsup_{n\to\infty} \left| \int_E \mathbb{1}_{V\geq R}(x)(R' \wedge f(x))d\mu_n(x) \right|\\ &\leq \|f\|_{C_V} \limsup_{n\to\infty} \mu_n(V \geq R) + \|f\|_{C_V} \limsup_{n\to\infty} \int_E \mathbb{1}_{V\geq R}(x)V(x)d\mu_n(x) \leq \|f\|_{C_V}\varepsilon.\end{aligned}$$

Daraus folgt (4.14). Nun schreibe

$$\begin{aligned}\left| \int_E f(x)d\mu_n(x) - \int_E f(x)d\mu(x) \right| \leq &\left| \int_E f(x)d\mu_n(x) - \int_E (R \wedge f(x))d\mu_n(x) \right|\\ &+ \left| \int_E (R \wedge f(x))d\mu_n(x) - \int_E (R \wedge f(x))d\mu(x) \right|\\ &+ \left| \int_E (R \wedge f(x))d\mu(x) - \int_E f(x)d\mu(x) \right|.\end{aligned}$$

Wegen $R \wedge f \in C_b(E)$ folgt

$$\limsup_{n\to\infty} \left| \int_E f(x)d\mu_n(x) - \int_E f(x)d\mu(x) \right| \leq 2\|f\|_{C_V}\varepsilon + \left| \int_E (R \wedge f(x))d\mu(x) - \int_E f(x)d\mu(x) \right|.$$

Da für $R \to \infty$ der letzte Term verschwindet, folgt die Behauptung. $\square$

4.6 Wasserstein 1-Distanz

Zum Schluss betrachten wir eine Variation der Beschränkten Lipschitz Metrik wo wir darauf verzichten, dass f beschränkt ist. Genauer sei

$$\|f\|_{Lip} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}$$

und sei $Lip(E) = \{f \mid \|f\|_{Lip} < \infty\}$. Dann gilt $Lip(E) \subset C(E)$ sowie

$$|f(x)| \leq |f(x) - f(x_0)| + |f(x_0)| \leq \|f\|_{Lip} d(x, x_0) + |f(x_0)|, \quad x, x_0 \in E.$$

Beispiel 4.32. *Ist $E = \mathbb{R}^d$ und $d(x,y) = |x-y|$ so ist $Lip(\mathbb{R}^d)$ die Menge aller global Lipschitz-stetigen Funktionen. Jedes $f \in Lip(\mathbb{R}^d)$ wächst demnach höchstens linear in $|x|$ (wir wählen hierbei $x_0 = 0$).*

Definiere

$$W_1(\mu, \nu) = \sup \left\{ \left| \int_E f(x) d\mu(x) - \int_E f(x) d\nu(x) \right| \; : \; f \in Lip(E), \quad \|f\|_{Lip} \leq 1 \right\}.$$

Da $f \in Lip(E)$ diesmal nicht beschränkt ist, ist W_1 im Allgemeinen nicht wohldefiniert. Wir betrachten die Einschränkung

$$\mathcal{P}_1(E) = \left\{ \mu \in \mathcal{P}(E) \mid \int_E d(x, x_0) d\mu(x) < \infty \right\}$$

wo $x_0 \in E$ ein beliebiger fester Punkt ist.

Lemma 4.33. *$(\mathcal{P}_1(E), W_1)$ ist ein metrischer Raum.*

Beweis. Übung. □

Die Konvergenz bezüglich dieser Wasserstein Metrik wird im nachfolgenden Satz charakterisiert.

Satz 4.34. *Sei $(\mu_n)_{n \geq 1} \subset \mathcal{P}_1(E)$ und $\mu \in \mathcal{P}_1(E)$. Dann sind äquivalent*

(a) $W_1(\mu_n, \mu) \longrightarrow 0$.

(b) $\mu_n \longrightarrow \mu$ *in* $\mathcal{P}_V(E)$ *für* $V(x) = d(x, x_0)$ *wo* $x_0 \in E$ *ein beliebiger Punkt ist.*

4.7 Zusammenhang zu Zufallsvariablen

Sei (E, d) ein metrischer Raum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine E-wertige Zufallsvariable. Dann ist $\omega \longmapsto (X_n(\omega), X(\omega))$ messbar bezüglich $\mathcal{F}--\mathcal{B}(E)\otimes\mathcal{B}(E)$. Ferner ist $(x, y) \longmapsto d(x, y)$ stetig, also $\mathcal{B}(E \times E) - -\mathcal{B}(\mathbb{R})$ messbar. Beachte, dass $E \times E$ mit $d_{E\times E}((x, y), (x', y')) := d(x, y) + d(x', y')$ wieder ein metrischer Raum ist. Um die Konvergenz $X_n \longrightarrow X$ zu untersuchen, betrachten wir die Abbildung

$$\Omega \longrightarrow [0, \infty), \quad \omega \longmapsto d(X_n(\omega), X(\omega)).$$

Diese ist messbar, sofern $\mathcal{B}(E \times E) \subset \mathcal{B}(E) \otimes \mathcal{B}(E)$ gilt. Ohne weiteres ist dieses jedoch nicht der Fall.

Lemma 4.35. *Sei E seperabel. Dann ist $E \times E$ seperabel und es gilt*

$$\mathcal{B}(E \times E) = \mathcal{B}(E) \otimes \mathcal{B}(E).$$

Hier und im folgenden ist (E, d) stets ein seperabler, metrischer Raum.

Definition 4.36. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X ein weitere E-wertige Zufallsvariable.*

(a) $X_n \longrightarrow X$ is Wahrscheinlichkeit, falls $d(X_n, X) \longrightarrow 0$ in Wahrscheinlichkeit.

(b) $X_n \longrightarrow X$ fast sicher, falls $d(X_n, X) \longrightarrow 0$ fast sicher.

(c) $X_n \longrightarrow X$ in $\mathcal{L}^p$, falls $d(X_n, X) \longrightarrow 0$ in $\mathcal{L}^p$.

(d) $X_n \longrightarrow X$ in Verteilung, falls $\mu_n := \mathbb{P} \circ X_n^{-1}$ schwach gegen $\mu := \mathbb{P} \circ X^{-1}$ konvergiert.

Konvergenz in Verteilung ist, per Definition vom Bildmaß, äquivalent zu

$$\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X)), \quad n \to \infty$$

für alle $f \in C_b(E)$.

Satz 4.37. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine weitere E-wertige Zufallsvariable. Angenommmen es gilt $X_n \longrightarrow X$ in Verteilung und X ist fast sicher konstant. Dann gilt $X_n \longrightarrow X$ in Wahrscheinlichkeit.*

Beweis. Es sei $\alpha \in E$ derart, dass $\mathbb{P}(X = \alpha) = 1$. Sei $A \in \mathcal{B}(E)$ beliebig, dann gilt

$$\mathbb{P} \circ X^{-1}(A) = \mathbb{P}(\{X \in A\}) = \mathbb{P}(\{X \in A\} \cap \{X = \alpha\}) = \delta_\alpha(A).$$

Folglich ist $\mathbb{P} \circ X^{-1} = \delta_\alpha$. Sei $\varepsilon > 0$ beliebig. Dann ist $\{y \in E \mid d(y, \alpha) \geq \varepsilon\}$ abgeschlossen. Aus dem Satz von Portmanteau folgt

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(d(X_n, X) \geq \varepsilon) &= \limsup_{n\to\infty} \mathbb{P}(d(X_n, \alpha) \geq \varepsilon) \\ &= \limsup_{n\to\infty} \mathbb{P} \circ X_n^{-1} (\{y \in E \mid d(y, \alpha) \geq \varepsilon\}) \\ &= \mathbb{P} \circ X^{-1} (\{y \in E \mid d(y, \alpha) \geq \varepsilon\}) = 0,\end{aligned}$$

da $\alpha \notin \{y \in E \mid d(y, \alpha) \geq \varepsilon\}$. □

Satz 4.38. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine E-wertige Zufallsvariable. Gilt $X_n \longrightarrow X$ in Wahrscheinlichkeit, so folgt $X_n \longrightarrow X$ in Verteilung.*

Beweis. Sei $f \in C_b(E)$ gleichmässig stetig und $\varepsilon > 0$. Dann gibt es ein $\delta > 0$, sodass für alle $x, y \in E$ mit $d(x,y) < \delta$ bereits $|f(x) - f(y)| < \varepsilon$ gilt. Dann gilt

$$\begin{aligned} |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| &\leq \int\limits_{d(X_n,X)<\delta} |f(X_n) - f(X)| d\mathbb{P} + \int\limits_{d(X_n,X)\geq\delta} |f(X_n) - f(X)| d\mathbb{P} \\ &\leq \varepsilon + 2\|f\|_\infty \mathbb{P}\left(d(X_n, X) \geq \delta\right). \end{aligned}$$

□

Satz 4.39. *Seien $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ Zufallsvariablen auf E und $\mu \in \mathcal{P}(E)$. Angenommen $\mathbb{P} \circ X_n^{-1} \longrightarrow \mu$ und $d(X_n, Y_n) \longrightarrow 0$ in Wahrscheinlichkeit. Dann gilt*

$$\mathbb{P} \circ Y_n^{-1} \longrightarrow \mu, \quad n \to \infty.$$

Beweis. Sei $F \subset E$ abgeschlossen. Für $\varepsilon > 0$ sei $F^\varepsilon = \{x \in E \mid d(x,F) \leq \varepsilon\}$ und $F_0^\varepsilon = \{x \in E \mid d(x,F) < \varepsilon\}$. Dann gilt

$$\begin{aligned} \{Y_n \in F\} &= (\{Y_n \in F\} \cap \{d(X_n,Y_n) < \varepsilon\}) \cup (\{Y_n \in F\} \cap \{d(X_n,Y_n) \geq \varepsilon\}) \\ &= (\{Y_n \in F\} \cap \{d(X_n,Y_n) < \varepsilon\} \cap \{X_n \in F_0^\varepsilon\}) \cup (\{Y_n \in F\} \cap \{d(X_n,Y_n) \geq \varepsilon\}) \\ &\subset \{X_n \in F_0^\varepsilon\} \cup \{d(X_n,Y_n) \geq \varepsilon\} \subset \{X_n \in F^\varepsilon\} \cup \{d(X_n,Y_n) \geq \varepsilon\}. \end{aligned}$$

Da F^ε abgeschlossen ist folgt

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}(Y_n \in F) &\leq \limsup_{n\to\infty} \mathbb{P}(d(Y_n, X_n) \geq \varepsilon) + \limsup_{n\to\infty} \mathbb{P}(X_n \in F^\varepsilon) \\ &= \limsup_{n\to\infty} \mathbb{P} \circ X_n^{-1}(F^\varepsilon) \leq \mathbb{P} \circ X^{-1}(F^\varepsilon). \end{aligned}$$

Wegen $F^\varepsilon \searrow F$ für $\varepsilon \to 0$ folgt die Behauptung. □

Wir zeigen noch einen alternativen Beweis.

Beweis. Wegen dem Satz von Portmanteau reicht es zu zeigen, dass für jede gleichmässig stetige Funktion $f : E \longrightarrow \mathbb{R}$

$$\int\limits_E f(y)(\mathbb{P} \circ Y_n^{-1})(dy) \longrightarrow \int\limits_E f(y) d\mu(y), \quad n \to \infty$$

gilt. Sei $\varepsilon > 0$ beliebig und wähle $\delta > 0$ derart, dass

$$|f(x) - f(y)| < \frac{\varepsilon}{3}, \quad \forall x, y \in E \text{ mit } d(x,y) < \delta.$$

Wähle $n_0 \in \mathbb{N}$ derart, dass

$$\left| \int\limits_E f(y)(\mathbb{P} \circ X_n^{-1})(dy) - \int\limits_E f(y) d\mu(y) \right| < \frac{\varepsilon}{3}$$

sowie $\mathbb{P}(d(X_n, Y_n) \geq \delta) < \frac{\varepsilon}{3(2\|f\|_\infty+1)}$ für alle $n \geq n_0$ gilt. Dann folgt

$$\begin{aligned}
&\left| \int_E f(y)(\mathbb{P} \circ Y_n^{-1})(dy) - \int_E f(y) d\mu(y) \right| \\
&\leq |\mathbb{E}(f(X_n)) - \mathbb{E}(f(Y_n))| + \left| \int_E f(y)(\mathbb{P} \circ X_n^{-1})(dy) - \int_E f(y) d\mu(y) \right| \\
&\leq \int_{d(X_n,Y_n)\geq\delta} |f(X_n) - f(Y_n)| d\mathbb{P} + \int_{d(X_n,Y_n)<\delta} |f(X_n) - f(Y_n)| d\mathbb{P} + \frac{\varepsilon}{3} \\
&\leq 2\|f\|_\infty \mathbb{P}(d(X_n, Y_n) \geq \delta) + \frac{2\varepsilon}{3} \leq \varepsilon.
\end{aligned}$$

□

5 Charakteristische Funktionen

Im letzten Kapitel haben wir Eigenschaften von Wahrscheinlichkeitsmaßen auf metrischen und Polnischen Räumen untersucht. Dieses Kapitel widmet sich dem Studium von Wahrscheinlichkeitsmaßen auf $E = \mathbb{R}^d$. Für diesen Spezialfall ist es möglich jedem endlichen Maß μ auf $\mathbb{R}^d$ eine komplex-wertige Funktion (charakteristische Funktion) $\widehat{\mu}(\xi)$ zuzuordnen. Schwache Konvergenz von Wahrscheinlichkeitsmaßen sowie Existenz von Momenten können mithilfe der charakteristischen Funktion beschrieben werden.

5.1 Fouriertransformation von Funktionen

Für $\xi, x \in \mathbb{R}^d$ sei $\xi \cdot x := \sum_{k=1}^{d} \xi_k x_k = \langle \lambda, A\lambda \rangle$ das euklidische Skalarprodukt. hier und im folgenden bezeichnet $\mathcal{L}^1(\mathbb{R}^d)$ den Raum aller messbaren, integrierbaren Funktionen auf $\mathbb{R}^d$, d.h. für $g \in \mathcal{L}^1(\mathbb{R}^d)$ gilt

$$\int_{\mathbb{R}^d} |g(x)| m(dx) < \infty.$$

Definition 5.1. *Sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Die Fouriertransformierte von g ist für $\xi \in \mathbb{R}^d$ definiert durch*

$$\widehat{g}(\xi) := \int_{\mathbb{R}^d} e^{i\xi\cdot x} g(x) m(dx) := \int_{\mathbb{R}^d} \cos(\xi \cdot x) g(x) m(dx) + i \int_{\mathbb{R}^d} \sin(\xi \cdot x) g(x) m(dx).$$

Die Definition ist wegen $|e^{i\xi\cdot x}| = 1$ für alle $\xi \in \mathbb{R}^d$ wohldefiniert.

Definition 5.2. *Eine Funktion $g : \mathbb{R}^d \longrightarrow \mathbb{C}$ heißt positiv (semi-)definit, wenn für alle $n \in \mathbb{N}$ und alle $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$ die Matrix $(g(\xi_i - \xi_j))_{1\leq i,j\leq n}$ positiv (semi-)definit ist.*

Diese Definition ist äquivalent zu: Für alle $n \in \mathbb{N}$, alle $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$ und alle $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$ gilt

$$\sum_{i,j=1}^{n} \lambda_j \overline{\lambda}_i g(\xi_j - \xi_i) \geq 0. \tag{5.1}$$

Denn setze $A := (g(\xi_j - \xi_i))_{1 \leq i,j \leq n}$, d.h. $A_{ij} = g(\xi_j - \xi_i)$. Dann gilt $(A\lambda)_i = \sum\limits_{j=1}^{n} A_{ij}\lambda_j$ und folglich

$$\langle \lambda, A\lambda \rangle = \sum_{i=1}^{n} \overline{\lambda}_i (A\lambda)_i = \sum_{i,j=1}^{n} \overline{\lambda}_i \lambda_j A_{ij} = \sum_{i,j=1}^{n} \overline{\lambda}_i \lambda_j g(\xi_j - \xi_i).$$

Man beachte, dass die Matrix A insbesondere hermitsch ist, d.h. $\overline{A} = A^T$. Es sei g eine positiv (semi-)definite Funktion. Dann gilt:

- Für $n = 1$ und $\lambda \in \mathbb{C}$ folgt aus (5.1) $|\lambda|^2 g(0) \geq 0$, also $g(0) \geq 0$.
- Für $n = 2$ mit $\xi_1 = 0$ und $\xi_2 = \xi \in \mathbb{R}^d$ folgt, dass die Matrix $\begin{pmatrix} g(0) & g(\xi) \\ g(-\xi) & g(0) \end{pmatrix}$ positiv (semi-)definit ist. Daraus folgt $g(-\xi) = \overline{g(\xi)}$. Ferner muss die Determinante nicht-negativ sein, also
$$g(0)^2 - g(\xi)g(-\xi) = g(0)^2 - |g(\xi)|^2 \geq 0$$
und somit $|g(\xi)| \leq g(0)$.

Es sei $C_c^\infty(\mathbb{R}^d)$ der Raum der glatten Funktionen mit kompaktem Träger. D.h. $f \in C_c^\infty(\mathbb{R}^d)$, falls f beliebig oft differenzierbar ist und $f \in C_c(\mathbb{R}^d)$.

Lemma 5.3. *[Wer00] Für jedes $g \in \mathcal{L}^1(\mathbb{R}^d)$ und jedes $\varepsilon > 0$ gibt es eine Funktion $f_\varepsilon \in C_c^\infty(\mathbb{R}^d)$ mit*

$$\|f_\varepsilon - g\|_{\mathcal{L}^1} := \int_{\mathbb{R}^d} |f_\varepsilon(x) - g(x)| m(dx) < \varepsilon.$$

Die wichtigsten Eigenschaften der Fouriertransformierten von g sind im nächsten Satz zusammengefasst.

Satz 5.4. *Es sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Es gelten die folgenden Eigenschaften.*

(a) $|\widehat{g}(\xi)| \leq \|g\|_{\mathcal{L}^1}$ und $\widehat{g}(0) = \int_{\mathbb{R}^d} g(x) m(dx)$.

(b) $\widehat{g}$ ist gleichmässig stetig.

(c) $\widehat{g}$ verschwindet im Unendlichen, d.h. für alle $\varepsilon > 0$ gibt es eine kompakte Menge $K_\varepsilon \subset \mathbb{R}^d$ derart, dass

$$|\widehat{g}(\xi)| < \varepsilon, \quad \xi \notin K_\varepsilon.$$

(d) Ist g fast überall nicht-negativ, so ist $\widehat{g}$ positiv (semi-)definit.

Beweis. (a) Klar.
(b) Seien $\xi_1, \xi_2 \in \mathbb{R}^d$ und $\varepsilon > 0$. Wähle eine kompakte Menge $K_\varepsilon \subset \mathbb{R}^d$ derart, dass

$$\int_{K_\varepsilon^c} |g(x)| m(dx) < \frac{\varepsilon}{4}.$$

Sei $\delta > 0$ mit $\delta < \frac{\varepsilon}{2(1+\sup_{x\in K_\varepsilon} |x|)(1+\|g\|_{\mathcal{L}^1})}$. Dann gilt für alle $|\xi_1 - \xi_2| < \delta$

$$|\widehat{g}(\xi_1) - \widehat{g}(\xi_2)| \leq \int_{\mathbb{R}^d} |e^{i\xi_1 \cdot x} - e^{i\xi_2 \cdot x}||g(x)| m(dx) \leq \int_{K_\varepsilon} |e^{i\xi_1 \cdot x} - e^{i\xi_2 \cdot x}||g(x)| m(dx) + \frac{\varepsilon}{2}.$$

Es gilt

$$|e^{i\xi_1 \cdot x} - e^{i\xi_2 \cdot x}| = \left| i \int_{\xi_1 \cdot x}^{\xi_2 \cdot x} e^{it} dt \right| \leq |(\xi_2 - \xi_1) \cdot x| \leq |x| \cdot |\xi_2 - \xi_1|$$

woraus folgt

$$|\widehat{g}(\xi_1) - \widehat{g}(\xi_2)| \leq \frac{\varepsilon}{2} + \|g\|_{\mathcal{L}^1} \sup_{x \in K_\varepsilon} |x| \cdot |\xi_1 - \xi_2| \leq \varepsilon.$$

(c) Wir zeigen die Behauptung zuerst für $g \in C_c^\infty(\mathbb{R}^d)$. Wähle $R > 0$ derart, dass $g(x) = 0$ für alle $|x| \geq R$. Wähle $R' > R$ und sei $|\xi| \geq R$. Dann finden wir ein $j \in \{1, \ldots, d\}$ mit $|\xi_j| \geq \frac{R'}{\sqrt{d}}$. Mittels partieller Integration, wobei der Randterm wegen $g \in C_c^\infty(\mathbb{R}^d)$ verschwidet, erhalten wir

$$\begin{aligned}
\widehat{g}(\xi) &= \int_{\mathbb{R}^d} e^{i\xi \cdot x} g(x) m(dx) = \int_{|x| \leq R'} e^{i\xi \cdot x} g(x) m(dx) \\
&= \frac{1}{i\xi_j} \int_{|x| \leq R'} \frac{\partial e^{i\xi \cdot x}}{\partial x_j} g(x) m(dx) = -\frac{1}{i\xi_j} \int_{|x| \leq R'} e^{i\xi \cdot x} \frac{\partial g(x)}{\partial x_j} m(dx).
\end{aligned}$$

Daraus folgt

$$|\widehat{g}(\xi)| \leq \frac{1}{|\xi_j|} \left\| \frac{\partial g}{\partial x_j} \right\|_{\mathcal{L}^1} \leq \frac{\sqrt{d}}{R'} \left\| \frac{\partial g}{\partial x_j} \right\|_{\mathcal{L}^1}.$$

Die Behauptung folgt aus $R' \longrightarrow \infty$. Für den allgemeinen Fall sei $g \in \mathcal{L}^1(\mathbb{R}^d)$, $\varepsilon > 0$ und $f_\varepsilon \in C_c^\infty(\mathbb{R}^d)$ derart, dass $\|g - f_\varepsilon\|_{\mathcal{L}^1} < \varepsilon$. Die Behauptung folgt aus

$$|\widehat{g}(\xi)| \leq |\widehat{g}(\xi) - \widehat{f}_\varepsilon(\xi)| + |\widehat{f}_\varepsilon(\xi)| \leq \|g - f_\varepsilon\|_{\mathcal{L}^1} + |\widehat{f}_\varepsilon(\xi)|.$$

(d) Es gilt

$$\int_{\mathbb{R}^d} \left| \sum_{j=1}^n \lambda_j e^{i\xi_j \cdot x} \right|^2 g(x) m(dx) = \int_{\mathbb{R}^d} \sum_{i,j=1}^n \lambda_j \overline{\lambda}_i e^{ix \cdot (\xi_j - \xi_i)} g(x) m(dx) = \sum_{i,j=1}^n \lambda_j \overline{\lambda}_i \widehat{g}(\xi_j - \xi_i).$$

□

Wir betrachten das Beispiel einer Normalverteilung.

Beispiel 5.5. *Es sei* $g(x) := e^{-\alpha|x|^2}$ *mit* $\alpha > 0$. *Dann ist*

$$\widehat{g}(\xi) = \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} e^{-\frac{|\xi|^2}{4\alpha}}. \tag{5.2}$$

Beachte, für $\alpha = \frac{1}{2}$ *ist* $\widehat{g}(\xi) = (2\pi)^{\frac{d}{2}} g(\xi)$.

Beweis. Es gilt

$$\widehat{g}(\xi) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} e^{-\alpha|x|^2} m(dx) = \prod_{k=1}^{d} \int\limits_{\mathbb{R}} e^{i\xi_k\cdot x_k} e^{-\alpha x_k^2} m(dx).$$

Also reicht es aus die Behauptung für $d = 1$ zu beweisen. Für diesen Fall erhalten wir für $\xi, x \in \mathbb{R}$

$$e^{-\alpha x^2} e^{i\xi x} = e^{-\alpha\left(x^2 - 2x\frac{i\xi}{2\alpha} + \left(\frac{i\xi}{2\alpha}\right)^2 - \left(\frac{i\xi}{2\alpha}\right)^2\right)} = e^{-\frac{\xi^2}{4\alpha}} e^{-\alpha\left(x - \frac{i\xi}{2\alpha}\right)^2}.$$

Daraus folgt

$$\int\limits_{\mathbb{R}} e^{i\xi x} e^{-\alpha x^2} m(dx) = e^{-\frac{\xi^2}{4\alpha}} \int\limits_{\mathbb{R}} e^{-\alpha\left(x - \frac{i\xi}{2\alpha}\right)^2} m(dx).$$

Für das Integral erhalten wir mittels Cauchys Integralsatz (siehe [FL80]) und der Substitution $y = \sqrt{\alpha}x$

$$\int\limits_{\mathbb{R}} e^{-\alpha\left(x - \frac{i\xi}{2\alpha}\right)^2} m(dx) = \int\limits_{\mathbb{R}} e^{-\alpha x^2} m(dx) = \frac{1}{\sqrt{\alpha}} \int\limits_{\mathbb{R}} e^{-x^2} m(dx) = \sqrt{\frac{\pi}{\alpha}}.$$

□

Als nächstes wollen wir die Inverse Transformation finden.

Lemma 5.6. *Es seien* $g, h \in \mathcal{L}^1(\mathbb{R}^d)$. *Dann gilt*

$$\int\limits_{\mathbb{R}^d} \widehat{g}(\xi) h(\xi) m(d\xi) = \int\limits_{\mathbb{R}^d} g(\xi) \widehat{h}(\xi) m(d\xi).$$

Beweis. Mittels Fubini erhalten wir

$$\int\limits_{\mathbb{R}^d} \widehat{g}(\xi) h(\xi) m(d\xi) = \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} g(x) h(\xi) m(dx) m(d\xi) = \int\limits_{\mathbb{R}^d} g(\xi) \widehat{h}(\xi) m(d\xi).$$

Fubini ist anwendbar, da $(x, \xi) \longmapsto g(x)h(\xi)$ integrierbar ist bezüglich $m(dx)m(d\xi)$. □

Lemma 5.7. *Für jedes* $g \in C_b(\mathbb{R}^d)$ *und jedes* $x \in \mathbb{R}^d$ *gilt*

$$\int\limits_{\mathbb{R}^d} g(y) \frac{e^{-\frac{|x+y|^2}{4\alpha}}}{(4\pi\alpha)^{\frac{d}{2}}} m(dy) \longrightarrow g(-x), \quad \alpha \to 0.$$

Beweis. Übung. $\square$

Lemma 5.8. *Es sei $g \in C_b(\mathbb{R}^d) \cap \mathcal{L}^1(\mathbb{R}^d)$ derart, dass $\widehat{g} \in \mathcal{L}^1(\mathbb{R}^d)$. Dann gilt*

$$\widehat{\widehat{g}}(x) = (2\pi)^d g(-x), \quad x \in \mathbb{R}^d.$$

Beweis. Sei $\alpha > 0$ und setze $\varphi_x(\xi) := e^{i\xi\cdot x}e^{-\alpha|\xi|^2}$. Für diese Funktion gilt nach (5.2)

$$\widehat{\varphi}_x(\xi) = \int\limits_{\mathbb{R}^d} e^{i(\xi+x)\cdot y}e^{-\alpha|y|^2}m(dy) = \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} e^{-\frac{|\xi+x|^2}{4\alpha}}.$$

Damit erhalten wir

$$\begin{aligned}\int\limits_{\mathbb{R}^d} \widehat{g}(\xi)e^{i\xi\cdot x}e^{-\alpha|\xi|^2}m(d\xi) &= \int\limits_{\mathbb{R}^d} g(\xi)\widehat{\varphi}_x(\xi)m(d\xi)\\ &= \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} \int\limits_{\mathbb{R}^d} g(\xi)e^{-\frac{|\xi+x|^2}{4\alpha}}m(d\xi) = (2\pi)^d \int\limits_{\mathbb{R}^d} g(\xi)\frac{e^{-\frac{|\xi+x|^2}{4\alpha}}}{(4\pi\alpha)^{\frac{d}{2}}}m(d\xi).\end{aligned}$$

Betrachten wir den Grenzwert $\alpha \to 0$, so folgt die Behauptung aus Lemma 5.7. $\square$

Satz 5.9. *Sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Definiere die Abbildung*

$$\check{g}(x) := \frac{1}{(2\pi)^d}\int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x}g(\xi)m(d\xi), \quad x \in \mathbb{R}^d.$$

Dann gilt für alle $g \in C_b(\mathbb{R}^d) \cap \mathcal{L}^1(\mathbb{R}^d)$ derart, dass $\widehat{g} \in \mathcal{L}^1(\mathbb{R}^d)$

$$\check{\widehat{g}} = g = \widehat{\check{g}}.$$

Beweis. Beachte, dass $\check{g}(x) = \frac{1}{(2\pi)^d}\widehat{g}(-x)$ für alle $g \in \mathcal{L}^1(\mathbb{R}^d)$ gilt. Sei $g \in C_b(\mathbb{R}^d) \cap \mathcal{L}^1(\mathbb{R}^d)$ mit $\widehat{g} \in \mathcal{L}^1(\mathbb{R}^d)$. Aus Lemma 5.8 folgt $\check{\widehat{g}}(x) = \frac{1}{(2\pi)^d}\widehat{\widehat{g}}(-x) = g(x)$ für alle $x \in \mathbb{R}^d$. Analog lässt sich die andere Identität zeigen. $\square$

5.2 Charakteristische Funktion für Maße

Definition 5.10. *Sei μ ein endliches Maß auf $\mathbb{R}^d$. Die charakteristische Funktion von μ ist die Fouriertransformierte von μ gegeben durch*

$$\widehat{\mu}(\xi) := \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x}\mu(dx), \quad \xi \in \mathbb{R}^d.$$

Bemerkung 5.11. *(a) Sei $\mu = \sum\limits_{k=1}^{\infty} a_k\delta_{x_k}$ mit $a_k \geq 0$ und $x_k \in \mathbb{R}^d$ derart, dass $\sum\limits_{k=1}^{\infty} a_k = 1$. Dann ist μ ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$ und es gilt*

$$\widehat{\mu}(\xi) = \sum_{k=1}^{\infty} a_k e^{i\xi\cdot x_k}, \quad \xi \in \mathbb{R}^d.$$

(b) *Hat μ eine Dichte, d.h. $\mu(dx) = p(x)m(dx)$, so ist*

$$\widehat{\mu}(\xi) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} p(x)m(dx) = \widehat{p}(\xi).$$

Die wichtigsten Eigenschaften sind im nächsten Satz zusammengefasst.

Satz 5.12. *Es sei μ ein endliches Maß auf $\mathbb{R}^d$. Dann gelten die folgenden Eigenschaften.*

(a) $|\widehat{\mu}(\xi)| \leq \mu(\mathbb{R}^d) = \widehat{\mu}(0)$ *für alle* $\xi \in \mathbb{R}^d$.

(b) $\widehat{\mu}$ *ist gleichmässig stetig.*

(c) $\widehat{\mu}$ *ist positiv (semi-)definit.*

Beweis. Übung. $\square$

Wir betrachten einige Beispiele.

Beispiel 5.13. (a) *Benoulli Verteilung: $\mu = p\delta_1 + (1-p)\delta_0$. Dann ist*

$$\widehat{\mu}(\xi) = pe^{i\xi} + (1-p) = 1 + p(1 - e^{i\xi}), \quad \xi \in \mathbb{R}.$$

(b) *Binomialverteilung: $\mu = \sum\limits_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k$. Dann ist*

$$\widehat{\mu}(\xi) = (1 - p + pe^{i\xi})^n, \quad \xi \in \mathbb{R}.$$

(c) *Standardnormalverteilung: Im eindimensionalen Fall sei $\mu(dx) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} m(dx)$. Dann ist*

$$\widehat{\mu}(\xi) = e^{-\frac{\xi^2}{2}}, \quad \xi \in \mathbb{R}.$$

Für den d-dimansionalen Fall sei $\mu(dx) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}}$. Dann gilt

$$\widehat{\mu}(\xi) = e^{-\frac{|\xi|^2}{2}}, \quad \xi \in \mathbb{R}^d.$$

(d) *Normalverteilung im $\mathbb{R}^m$: Sei μ eine Normalverteilung auf dem $\mathbb{R}^m$. Dann gibt es eine $m \times d$-Matrix A und $b \in \mathbb{R}^m$ mit $\mu = \nu \circ T^{-1}$, wo $T(x) = Ax + b$ und*

$$\nu(dx) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}} m(dx).$$

Damit folgt

$$\begin{aligned}
\widehat{\mu}(\xi) &= \int\limits_{\mathbb{R}^m} e^{i\xi\cdot x} \mu(dx) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot T(x)} \nu(dx) \\
&= e^{i\xi\cdot b} \int\limits_{\mathbb{R}^d} e^{i(A^T\xi)\cdot x} \nu(dx) = e^{i\xi\cdot b} e^{-\frac{1}{2}|A^t\xi|^2} = e^{i\xi\cdot b} e^{-\frac{1}{2}\langle \xi, (AA^t)\xi\rangle}.
\end{aligned}$$

(e) *Uniforme Verteilung: Sei* $\mu(dx) = \mathbb{1}_{(-a,a)}(x)\frac{1}{2a}m(dx)$ *mit* $a > 0$. *Dann ist*

$$\widehat{\mu}(\xi) = \begin{cases} \frac{\sin(\xi a)}{\xi a}, & \xi \neq 0 \\ 1, & \textit{sonst} \end{cases}, \quad \xi \in \mathbb{R}.$$

(f) *Cauchy Verteilung: Sei* $\mu(dx) = \frac{1}{\pi}\frac{c}{c^2+x^2}m(dx)$ *mit* $c > 0$. *Dann gilt*

$$\widehat{\mu}(\xi) = e^{-c|\xi|}, \quad \xi \in \mathbb{R}.$$

Der nächste Satz zeigt, dass die charakteristische Funktion das Maß eindeutig festlegt.

Satz 5.14. *Seien* μ, ν *zwei endliche Maße auf* $\mathbb{R}^d$ *mit* $\widehat{\mu} = \widehat{\nu}$. *Dann gilt* $\mu = \nu$.

Beweis. Es gilt $\mu(\mathbb{R}^d) = \widehat{\mu}(0) = \widehat{\nu}(0) = \nu(\mathbb{R}^d)$. Ist $\mu(\mathbb{R}^d) = 0$, so ist nichts zu zeigen. Sei also $\mu(\mathbb{R}^d) > 0$. Definiere neue Maße $\mu' := \frac{1}{\mu(\mathbb{R}^d)}\mu$ sowie $\nu' := \frac{1}{\nu(\mathbb{R}^d)}\nu$. Dann gilt $\widehat{\mu'} = \widehat{\nu'}$. Es reicht daher aus die Behauptung für Wahrscheinlichkeitsmaße zu beweisen.

Die kompakten Mengen bilden ein $\cap$-stabiles Erzeugendensystem der Borel-σ-Algebra $\mathcal{B}(\mathbb{R}^d)$. Es genügt also $\mu(K) = \nu(K)$ für alle kompakten Mengen $K \subset \mathbb{R}^d$ zu zeigen. Sei $K \subset \mathbb{R}^d$ kompakt und

$$f_m(x) = \begin{cases} 1, & x \in K \\ 0, & d(x,K) \geq \frac{1}{m} \\ 1 - md(x,K), & \text{sonst} \end{cases}.$$

Dann gilt $0 \leq f_m \leq 1$, f_m ist stetig und $f_m \searrow \mathbb{1}_K$ für $m \to \infty$. Wir zeigen

$$\int\limits_{\mathbb{R}^d} f_m(x)d\mu(x) = \int\limits_{\mathbb{R}^d} f_m(x)d\nu(x), \quad m \in \mathbb{N}. \tag{5.3}$$

Dann folgt mit dem Satz von Lebesgue $\nu(K) = \mu(K)$. Allgemeiner zeigen wir (5.3) für alle $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ mit

- $0 \leq f \leq 1$.
- $f \in C_c(\mathbb{R}^d)$.

Sei $\varepsilon > 0$ beliebig. Da f kompakten Träger hat, können wir $N \in \mathbb{N}$ wählen mit

$$\overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}} \subset [-N,N]^d =: B_N$$

und $\mu(B_N^c) < \varepsilon$, $\nu(B_N^c) < \varepsilon$. Nach dem Weierstraßschen Approximationssatz (siehe [Wer00] für den Beweis in einer Dimension) gibt es eine Funktion

$$g_\varepsilon(x) = \sum_{j=1}^m c_j^\varepsilon e^{i\frac{\pi}{N}\langle \xi_j^\varepsilon, x\rangle}$$

mit $c_j^\varepsilon \in \mathbb{C}$, $\xi_j^\varepsilon \in \mathbb{Z}^d$, $\|g_\varepsilon\|_\infty \leq \|f\|_\infty$ und

$$\sup_{x \in B_N} |f(x) - g(x)| \leq \varepsilon.$$

Es gilt mit $f|_{B_N^c} = 0$ und $\mu(B_N) \leq 1$

$$\left|\int_{\mathbb{R}^d} (f - g_\varepsilon) d\mu\right| \leq \int_{B_N} |f - g_\varepsilon| d\mu + \int_{B_N^c} |f| + |g_\varepsilon| d\mu$$
$$\leq \varepsilon\mu(B_N) + \|g_\varepsilon\|_\infty \mu(B_N^c) \leq (1 + \|f\|_\infty)\varepsilon.$$

Analog lässt sich zeigen, dass

$$\left|\int_{\mathbb{R}^d} (f - g_\varepsilon) d\nu\right| \leq (1 + \|f\|_\infty)\varepsilon.$$

Insgesamt erhalten wir

$$\left|\int_{\mathbb{R}^d} f d\mu - \int_{\mathbb{R}^d} f d\nu\right| \leq \left|\int_{\mathbb{R}^d} (f - g_\varepsilon) d\mu\right| + \left|\int_{\mathbb{R}^d} g_\varepsilon d\mu - \int_{\mathbb{R}^d} g_\varepsilon d\nu\right| + \left|\int_{\mathbb{R}^d} (g_\varepsilon - f) d\nu\right|$$
$$\leq 2\varepsilon + \left|\int_{\mathbb{R}^d} g_\varepsilon d\mu - \int_{\mathbb{R}^d} g_\varepsilon d\nu\right|.$$

Die Behauptung folgt aus

$$\int_{\mathbb{R}^d} g_\varepsilon d\mu = \sum_{j=1}^m c_j^\varepsilon \int_{\mathbb{R}^d} e^{i\frac{\pi}{N}\langle \xi_j^\varepsilon, x\rangle} d\mu(x) = \sum_{j=1}^m c_j^\varepsilon \int_{\mathbb{R}^d} e^{i\frac{\pi}{N}\langle \xi_j^\varepsilon, x\rangle} d\nu(x) = \int_{\mathbb{R}^d} g_\varepsilon d\nu.$$

□

Aus dem Beweis erhalten wir direkt.

Korollar 5.15. *Seien μ, ν zwei endliche Maße mit*

$$\int_{\mathbb{R}^d} f(x) d\mu(x) = \int_{\mathbb{R}^d} f(x) d\nu(x)$$

for alle stetigen Funktionen f mit kompaktem Träger. Dann gilt $\mu = \nu$.

Der nachfolgende Satz ist ein wichtiges Mittel zur Konstruktion von Wahrscheinlichkeitsmaßen.

Satz 5.16. *[Jac01, Theorem 3.5.7] Eine Funktion $g : \mathbb{R}^d \longrightarrow \mathbb{C}$ ist genau dann die charakteristische Funktion von einem endlichen Maß μ auf $\mathbb{R}^d$, wenn die folgenden Eigenschaften gelten:*

(i) g ist stetig.

(ii) g ist positiv (semi-)definit.

In diesem Fall ist $g(0) = \mu(\mathbb{R}^d)$. Insbesondere ist μ ein Wahrscheinlichkeitsmaß genau dann, wenn $g(0) = 1$.

Beispiel 5.17. *Betrachte $d = 1$ und für $\alpha \geq 0$ setze*

$$g_\alpha(\xi) = e^{-|\xi|^\alpha}, \quad \alpha \geq 0, \ \xi \in \mathbb{R}.$$

Diese ist offensichtlich stetig und erfüllt $g_\alpha(0) = 1$.

(i) $\alpha = 0$, Punktverteilung. Wähle $\mu_0(dx) = \delta_0(dx)$, dann ist $\widehat{\mu_0} = g_0$.

(ii) $\alpha = 1$, Cauchy-Verteilung. Wähle $\mu_1(dx) = p_1(x)dx$ mit $p_1(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

(iii) $\alpha = 2$, Gauss-Verteilung. Wähle $\mu_2(dx) = p_2(x)dx$ mit $p_2(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$.

Gibt es Verteilungen μ_α für andere Werte von α? Es lässt sich zeigen (siehe [Jac01]) dass g_α für alle $\alpha \in [0, 2]$ positiv (semi-)definit ist. Das zur charakteristischen Funktion g_α mit $\alpha \in [0, 2]$ gehörende Wahrscheinlichkeitsmaß wird auch also α-stabile Verteilung bezeichnet.

Für welche Werte von α hat μ_α eine Dichte? Diese Frage wird im nachfolgenden Satz beantwortet.

Im Folgenden wollen wir weitere Eigenschaften von μ aus der charakteristischen Funktion $\widehat{\mu}$ ableiten.

Satz 5.18. *Es sei μ ein endliches Maß auf $\mathbb{R}^d$. Es gelten die folgenden Aussagen.*

(a) Ist $\widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$, so hat μ eine Dichte $0 \leq p \in C_b(\mathbb{R}^d)$, welche im Unendlichen verschwindet und gegeben ist durch

$$p(x) := \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x} \widehat{\mu}(\xi) m(d\xi), \quad x \in \mathbb{R}^d. \tag{5.4}$$

(b) Ist $\widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$ derart, dass für ein $m \geq 1$

$$\int\limits_{\mathbb{R}^d} |\xi|^m |\widehat{\mu}(\xi)| m(d\xi) < \infty.$$

Dann hat p stetige Ableitungen bis zur Ordnung m. In diese Fall gilt für alle $0 \leq k \leq m$ und alle $1 \leq j_1, \ldots, j_k \leq d$

$$\frac{\partial^k p(x)}{\partial x_{j_1} \cdots \partial x_{j_k}} = \frac{(-i)^k}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x} \xi_{j_1} \cdots \xi_{j_k} \widehat{\mu}(\xi) m(d\xi).$$

(c) Ist

$$\int\limits_{\mathbb{R}^d} |x|^m \mu(dx) < \infty$$

für ein $m \geq 1$. Dann hat $\widehat{\mu}$ stetige partielle Ableitungen bis zur Ordnung m und es gilt für alle $0 \leq k \leq m$ und $1 \leq j_1, \ldots, j_k \leq d$

$$\frac{\partial^k \widehat{\mu}(\xi)}{\partial \xi_{j_1} \cdots \partial \xi_{j_k}} = i^k \int\limits_{\mathbb{R}^d} x_{j_1} \cdots x_{j_k} e^{i\xi \cdots x} \mu(dx).$$

Beweis. (a) Offensichtlich is p definiert durch (5.4) als Fouriertransformierte einer integrierbaren Funktion selbe stetig und verschwident im unendlichen. Ferner gilt $|p(x)| \leq \|\widehat{\mu}\|_{\mathcal{L}^1}$. Wir zeigen, dass

$$\int\limits_{\mathbb{R}^d} \varphi(x) d\mu(x) = \int\limits_{\mathbb{R}^d} \varphi(x) p(x) m(dx), \quad \forall \varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d).$$

Da die stetigen Funktionen mit kompaktem Träger in $\mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ enthalten sind folgt dann $d\mu(x) = p(x)dx$.

Für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d)$ gilt $\varphi \cdot \widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$ und folglich

$$\int\limits_{\mathbb{R}^d} \varphi(x) \widehat{\mu}(x) m(dx) = \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \varphi(x) e^{i\xi \cdot x} \mu(d\xi) m(dx) = \int\limits_{\mathbb{R}^d} \widehat{\varphi}(\xi) \mu(d\xi).$$

Man beachte, dass alle Integrale existieren und somit der Satz von Fubini anwendbar ist. Wir erhalten für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ mit $\widehat{\varphi} \in \mathcal{L}^1(\mathbb{R}^d)$

$$\begin{aligned}(2\pi)^d \int\limits_{\mathbb{R}^d} \varphi(x) p(x) m(dx) &= \int\limits_{\mathbb{R}^d} \varphi(x) \widehat{\widehat{\mu}}(-x) m(dx) = \int\limits_{\mathbb{R}^d} \varphi(-x) \widehat{\widehat{\mu}}(x) m(dx) \\ &= \int\limits_{\mathbb{R}^d} \widehat{\varphi(-\cdot)}(x) \widehat{\mu}(x) m(dx) = \int\limits_{\mathbb{R}^d} \widehat{\widehat{\varphi(-\cdot)}}(x) \mu(dx) = (2\pi)^d \int\limits_{\mathbb{R}^d} \varphi(x) \mu(dx).\end{aligned}$$

Sei $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ beliebig. Definiere $p_\alpha(x) = (4\pi\alpha)^{-\frac{d}{2}} e^{-\frac{|x|^2}{4\alpha}}$ und

$$\varphi_\alpha(x) := \int\limits_{\mathbb{R}^d} \varphi(x-y) p_\alpha(y) m(dy) = \int\limits_{\mathbb{R}^d} \varphi(y) p_\alpha(x-y) m(dy).$$

Dann ist φ_α offensichtlich stetig (nach dem Satz der dominierten Konvergenz) und es gilt

$$|\varphi_\alpha(y)| \leq \|\varphi\|_\infty \int\limits_{\mathbb{R}^d} p_\alpha(y) m(dy) = \|\varphi\|_\infty.$$

Ferner lässt sich zeigen dass $\widehat{\varphi}_\alpha = \widehat{\varphi} \cdot \widehat{p}_\alpha$ gilt. Folglich gilt

$$\int\limits_{\mathbb{R}^d} |\widehat{\varphi_\alpha}(y)| m(dy) \leq \|\widehat{\varphi}\|_\infty \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \widehat{p}_\alpha(y) m(dy) < \infty$$

da $\widehat{p}_\alpha$ wieder die Dichte einer Normalverteilung ist. Damit erhalten wir

$$\int_{\mathbb{R}^d} \varphi_\alpha(x)p(x)m(dx) = \int_{\mathbb{R}^d} \varphi_\alpha(x)\mu(dx), \quad \forall \alpha > 0.$$

Da $\varphi_\alpha(x) \longrightarrow \varphi(x)$ for alle x für $\alpha \to 0$ gilt erhalten wir aus dem Satz der dominierten Konvergenz

$$\int_{\mathbb{R}^d} \varphi_\alpha(x)\mu(dx) \longrightarrow \int_{\mathbb{R}^d} \varphi(x)\mu(dx), \quad \alpha \to 0.$$

Für die linke Seite gilt mit der Substitution $x' = x - y$

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi_\alpha(x)p(x)m(dx) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(x-y)p_\alpha(y)p(x)m(dy)m(dx) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(x')p_\alpha(y)p(x'+y)dydx'.\end{aligned}$$

Das innere Integral erfüllt für alle x'

$$\int_{\mathbb{R}^d} p_\alpha(y)p(x'+y)dy \longrightarrow p(x'), \quad \alpha \to 0$$

sowie

$$\left|\int_{\mathbb{R}^d} p_\alpha(y)p(x'+y)dy\right| \leq \|p\|_\infty \leq \|\widehat{\mu}\|_{\mathcal{L}^1} < \infty.$$

Mit dem Satz der dominierten Konvergenz folgt

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(x')p_\alpha(y)p(x'+y)dydx' \longrightarrow \int_{\mathbb{R}^d} \varphi(x)p(x)dx, \quad \alpha \to 0.$$

(b) Die Behauptung folgt aus der Darstellung (5.4) und denselben Argumenten wir für Teil (c).
(c) Wir zeigen nur den Fall $m = 1$. Der allgemeine Fall folgt dann über Induktion. Sei e_j der Basisvektor im $\mathbb{R}^d$ in Richtung j. Dann ist für $h > 0$

$$\frac{\widehat{\mu}(\xi + he_j) - \widehat{\mu}(xi)}{h} = \int_{\mathbb{R}^d} \frac{e^{i(\xi+he_j)\cdot x} - e^{i\xi\cdot x}}{h}\mu(dx) = \int_{\mathbb{R}^d} e^{i\xi\cdot x}\frac{e^{ix_jh}-1}{h}\mu(dx).$$

Es gilt $\frac{e^{ix_jh}-1}{h} \longrightarrow ix_j$ und nach Lemma 7.1 (siehe Appendix) gibt es eine stetige Funktion Θ_1 mit $|\Theta_1(x)| \leq 1$ derart, dass $e^{ix_kh} = 1 + ix_jh\Theta_1(x_jh)$. Daraus folgt

$$\frac{e^{ix_jh}-1}{h} = ix_j\Theta_1(x_jh)$$

und somit $\left|\frac{e^{ix_j h}-1}{h}\right| \leq |x_j|$ für alle $h > 0$. Dieses liefert eine integrierbare Majorante, also folgt aus dem Satz von Lebesgue

$$\frac{\partial \widehat{\mu}(\xi)}{\partial \xi_j} = i \int\limits_{\mathbb{R}^d} x_j e^{i\xi_j \cdot x} \mu(dx).$$

Die Ableitung ist stetig, ebenfalls aus dem Satz von Lebesgue. $\square$

Bemerkung 5.19. *Die Momente von μ haben die Darstellung*

$$\frac{1}{i^k} \frac{\partial^k \widehat{\mu}(0)}{\partial \xi_{j_1} \cdots \partial \xi_{j_k}} = \int\limits_{\mathbb{R}^d} x_{j_1} \cdots x_{j_k} \mu(dx).$$

Für $x \in \mathbb{R}^d$ setze $g_x(y) := \langle x, y \rangle = x \cdot y$.

Satz 5.20. *(Satz von Cramer-Wald) Seien $\mu_n, \mu \in \mathcal{P}(\mathbb{R}^d)$. Dann sind äquivalent:*

(a) $\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.

(b) $\mu_n \circ g_x^{-1} \longrightarrow \mu \circ g_x^{-1}$ schwach für $n \to \infty$ und alle $x \in \mathbb{R}^d$.

Beweis. $(a) \Longrightarrow (b)$: Folgt aus der Stetigkeit von g_x.
$(b) \Longrightarrow (a)$: Die Familie $\{\mu_n \circ g_x^{-1} \mid n \in \mathbb{N}\}$ ist nach Voraussetzung für jedes $x \in \mathbb{R}^d$ relativ kompakt und folglich straff. Seien $e_1, \ldots, e_d \in \mathbb{R}^d$ die Einheitsvektoren und $\varepsilon > 0$ beliebig. Dann gibt es für jedes $i = 1, \ldots, d$ eine kompakte Menge $K_i \subset \mathbb{R}$ mit

$$(\mu_n \circ g_{e_i}^{-1})(K_i) \geq 1 - \frac{\varepsilon}{d}, \quad n \in \mathbb{N}.$$

Setze $K := \bigcap\limits_{i=1}^{d} g_{e_i}^{-1}(K_i)$. Es ist nicht schwer zu zeigen, dass K abgeschlossen und beschränkt ist, also kompakt. Ferner gilt

$$\mu_n(K^c) = \mu_n\left(\bigcup_{i=1}^{d} (g_{e_i}^{-1}(K_i))^c\right) \leq \sum_{i=1}^{d} (\mu_n \circ g_{e_i}^{-1})(K_i^c) \leq d\frac{\varepsilon}{d} = \varepsilon$$

und folglich ist $\{\mu_n \mid n \in \mathbb{N}\}$ straff. Nach dem Satz von Prokhorov ist die Folge auch relativ kompakt. Sei $(\mu_n')_{n\in\mathbb{N}}$ eine konvergente Folge in $\{\mu_n \mid n \in \mathbb{N}\}$ und sei $\mu' \in \mathcal{P}(\mathbb{R}^d)$ der Grenzwert. Dann gilt auch $\mu_n' \circ g_x^{-1} \longrightarrow \mu' \circ g_x^{-1}$ schwach für alle $x \in \mathbb{R}^d$. Da $\mu_n' \circ g_x^{-1}$ eine Teilfolge von $\mu_n \circ g_x^{-1}$ ist, konvergiert $\mu_n' \circ g_x^{-1}$ gegen denselben Grenzwertwie $\mu_n \circ g_x^{-1}$, d.h. $\mu \circ g_x^{-1} = \mu' \circ g_x^{-1}$ für alle $x \in \mathbb{R}^d$. Wegen

$$\widehat{\mu}(\xi) = \int\limits_{\mathbb{R}^d} e^{i\xi \cdot x} \mu(dx) = \int\limits_{\mathbb{R}} e^{it} \mu \circ g_\xi(dt) = \widehat{\mu'}(\xi), \quad \xi \in \mathbb{R}^d$$

folgt $\mu = \mu'$. Die Behauptung folgt aus Lemma 4.5. $\square$

Satz 5.21. *(Levys Stetigkeitssatz) Seien $(\mu_n)_{n\in\mathbb{N}}$ Wahrscheinlichkeitsmaße auf $\mathbb{R}^d$. Angenommen es gibt eine in 0 stetige Funktion $f:\mathbb{R}^d \longrightarrow \mathbb{C}$ mit*

$$\widehat{\mu}_n(\xi) \longrightarrow f(\xi), \quad \forall \xi \in \mathbb{R}^d, \quad n \to \infty.$$

Dann gibt es ein Wahrscheinlichkeitsmaß μ auf $\mathbb{R}^d$ mit $\widehat{\mu} = f$ und $\mu_n \longrightarrow \mu$ schwach.

Korollar 5.22. *Für $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$ und $\mu \in \mathcal{P}(\mathbb{R}^d)$ gilt*

$$\mu_n \longrightarrow \mu \text{ schwach} \iff \widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi), \quad \forall \xi \in \mathbb{R}^d.$$

Beweis. Übung. □

Beweis. (Levys Stetigkeitssatz) Wir zeigen nur den Fall $d = 1$.
Schritt 1: $\{\mu_n \mid n \in \mathbb{N}\}$ ist straff.
Für alle $\alpha > 0$ gilt

$$\begin{aligned}\int_{-\alpha}^{\alpha} \widehat{\mu}_n(t)dt &= \int_{-\alpha}^{\alpha}\int_{\mathbb{R}} e^{itx}\mu_n(dx)dt = \int_{\mathbb{R}}\int_{-\alpha}^{\alpha} e^{itx}dt\mu_n(dx) \\ &= \int_{\mathbb{R}} \frac{e^{i\alpha x} - e^{-i\alpha x}}{ix}\mu_n(dx) = \int_{\mathbb{R}} \frac{2\sin(\alpha x)}{x}\mu_n(dx).\end{aligned}$$

Es folgt

$$\begin{aligned}\int_{-\alpha}^{\alpha}(1-\widehat{\mu}_n(t))dt &= 2\alpha - \int_{\mathbb{R}} \frac{2\sin(\alpha x)}{x}\mu_n(dx) \\ &= 2\alpha\left(1 - \int_{\mathbb{R}} \frac{\sin(\alpha x)}{\alpha x}\mu_n(dx)\right) = 2\alpha \int_{\mathbb{R}}\left(1 - \frac{\sin(\alpha x)}{\alpha x}\right)\mu_n(dx).\end{aligned}$$

Der Integrand ist nicht-negativ und es gilt

$$1 - \frac{\sin(\alpha x)}{\alpha x} \geq \frac{1}{2}, \quad |\alpha x| \geq 2.$$

Daraus folgt

$$\int_{-\alpha}^{\alpha}(1-\widehat{\mu}_n(t))dt \geq \alpha\mu_n\left(\left[-\frac{2}{\alpha}, \frac{2}{\alpha}\right]^c\right).$$

Mit $\beta = \frac{2}{\alpha}$ folgt

$$\mu_n([-\beta,\beta]^c) \leq \frac{1}{\alpha}\int_{-\alpha}^{\alpha}(1-\widehat{\mu}_n(t))dt = \frac{\beta}{2}\int_{-\frac{2}{\beta}}^{\frac{2}{\beta}}(1-\widehat{\mu}_n(t))dt.$$

Nach Voraussetzung ist f stetig in 0 und $f(0) = \lim_{n\to\infty} \widehat{\mu}_n(0) = 1$. Also gibt es zu $\varepsilon > 0$ ein $\beta > 0$ mit

$$|1 - f(t)| \leq \frac{\varepsilon}{4}, \quad |t| \leq \frac{2}{\beta}.$$

Dann ist

$$\lim_{n\to\infty} \frac{2}{\beta} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu}_n(t))dt = \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - f(t))dt \leq \frac{\varepsilon}{4}\frac{\beta}{2}\frac{4}{\beta} = \frac{\varepsilon}{2}$$

und es existiert ein $n_0(\varepsilon) = n_0 \in \mathbb{N}$ mit

$$\left| \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu}_n(t))dt - \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - f(t))dt \right| < \frac{\varepsilon}{2}, \quad n \geq n_0.$$

Damit ist also $\frac{2}{\beta} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu}_n(t))dt < \varepsilon$ für alle $n \geq n_0$, d.h.

$$\mu_n([-\beta, \beta]^c) \leq \varepsilon, \quad n \geq n_0.$$

Für $n < n_0$ gibt es ein $\beta_n > 0$ mit $\mu_n([-\beta_n, \beta_n]^c) \leq \varepsilon$. Sei $\beta_0 := \max\{\beta, \beta_1, \ldots, \beta_{n_0-1}\}$, so gilt

$$\mu_n([-\beta_0, \beta_0]^c) \leq \varepsilon, \quad n \geq 1.$$

Schritt 2. Sei $(\mu_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge von $(\mu_n)_{n\in\mathbb{N}}$. Dann ist $\{\mu_{n_k} \mid k \in \mathbb{N}\}$ straff und damit nach dem Satz von Prokhorov auch relativ kompakt. Folglich gibt es eine Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ welche schwach gegen μ' konvergiert. Damit folgt

$$f(\xi) = \lim_{l\to\infty} \widehat{\mu}_{n_{k_l}}(\xi) = \widehat{\mu'}(\xi), \quad \xi \in \mathbb{R}.$$

Die Behauptung folgt aus Lemma 4.5. □

5.3 Charakteristische Funktion für Zufallsvariablen

Definition 5.23. *Für eine Zufallsvariable X mit Werten in $\mathbb{R}^d$ ist die charakteristische Funktion definiert durch*

$$\varphi_X(\xi) := \mathbb{E}(e^{i\xi\cdot X}) = \widehat{\mathbb{P} \circ X^{-1}}(\xi), \quad \xi \in \mathbb{R}^d.$$

Korollar 5.24. *Sei X eine Zufallsvariable mit Werten in $\mathbb{R}^d$. Dann gelten die folgenden Aussagen:*

(a) Ist $\varphi_X \in \mathcal{L}^1(\mathbb{R}^d)$, so hat X eine Dichte $p \in C_b(\mathbb{R}^d)$, welche im Unendlichen verschwindet und gegeben ist durch

$$p(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\xi\cdot x} \varphi_X(\xi) m(d\xi).$$

(b) *Ist $\varphi_X \in \mathcal{L}^1(\mathbb{R}^d)$ und gilt für ein $m \geq 1$*

$$\int_{\mathbb{R}^d} |\xi|^m |\varphi_X(\xi)| m(d\xi) < \infty,$$

so hat p stetige Ableitungen bis zur Ordnung m und für $0 \leq k \leq m$, $1 \leq j_1, \ldots, j_k \leq d$ gilt

$$\frac{\partial^k p(x)}{\partial x_{j_1} \cdots \partial x_{j_k}} = \frac{(-i)^k}{(2\pi)^d} \int_{\mathbb{R}^d} \xi_{j_1} \cdots \xi_{j_k} e^{-i\xi \cdot x} \varphi_X(\xi) m(d\xi).$$

(c) *Es gebe ein $m \in \mathbb{N}$ derart, dass*

$$\mathbb{E}(|X|^m) < \infty.$$

Dann hat φ_X stetige Ableitungen bis zur Ordnung m und es gilt für alle $0 \leq k \leq m$, $1 \leq j_1, \ldots, j_k \leq d$

$$\frac{\partial^k \varphi_X(\xi)}{\partial \xi_{j_1} \cdots \partial \xi_{j_k}} = i^k \mathbb{E}(X_{j_1} \cdots X_{j_k} e^{i\xi \cdot X}).$$

Insbesondere gilt für die Momente von X die Darstellung

$$\mathbb{E}(X_{j_1} \cdots X_{j_k}) = \frac{1}{i^k} \frac{\partial^k \varphi_X(0)}{\partial \xi_{j_1} \cdots \partial \xi_{j_k}}.$$

Satz 5.25. *Seien X, Y unabhängige Zufallsvariablen mit Werten in $\mathbb{R}^d$. Dann gilt*

$$\varphi_X(\xi)\varphi_Y(\xi) = \varphi_{X+Y}(\xi), \quad \xi \in \mathbb{R}^d.$$

Beweis. Klar. □

Satz 5.26. *Sei X eine Zufallsvariable mit Werten in $\mathbb{R}^d$ sowie Y eine Zufallsvariable mit Werten in $\mathbb{R}^m$. Seien $\varphi_X, \varphi_Y, \varphi_{(X,Y)}$ die dazugehörigen charakteristischen Funktionen. Genau dann sind X, Y unabhängig, wenn*

$$\varphi_X(\xi_1)\varphi_Y(\xi_2) = \varphi_{(X,Y)}(\xi_1, \xi_2), \quad \xi_1 \in \mathbb{R}^d,\ \xi_2 \in \mathbb{R}^m. \tag{5.5}$$

Beweis. Seien X, Y unabhängig. Dann ist $\mathbb{P} \circ (X,Y)^{-1} = \mathbb{P} \circ X^{-1} \otimes \mathbb{P} \circ Y^{-1}$ und folglich

$$\begin{aligned}
\varphi_{(X,Y)}(\xi_1, \xi_2) &= \int_{\mathbb{R}^d \times \mathbb{R}^m} e^{i\xi_1 \cdot x} e^{i\xi_2 \cdot x} \mathbb{P} \circ (X,Y)^{-1}(dx, dy) \\
&= \int_{\mathbb{R}^d} e^{i\xi_1 \cdot x} \mathbb{P} \circ X^{-1}(dx) \int_{\mathbb{R}^m} e^{i\xi_2 \cdot y} \mathbb{P} \circ Y^{-1}(dy) \\
&= \varphi_X(\xi_1)\varphi_Y(\xi_2).
\end{aligned}$$

Umgekehrt gelte (5.5). Seien X', Y' unabhängige Zufallsvariablen mit charakteristischen Funktionen φ_X sowie φ_Y. Sei $\varphi_{(X',Y')}$ die charakteristische Funktion von (X', Y'). Dann gilt

$$\varphi_{(X',Y')}(\xi_1, \xi_2) = \varphi_{X'}(\xi_1)\varphi_{Y'}(\xi_2) = \varphi_X(\xi_1)\varphi_Y(\xi_2) = \varphi_{(X,Y)}(\xi_1, \xi_2)$$

und folglich

$$\mathbb{P} \circ (X,Y)^{-1} = \mathbb{P} \circ (X',Y')^{-1} = \mathbb{P} \circ X'^{-1} \otimes \mathbb{P} \circ Y'^{-1} = \mathbb{P} \circ X^{-1} \otimes \mathbb{P} \circ Y^{-1}.$$

□

Korollar 5.27. *Sei $X = (X_1, \ldots, X_d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable. Genau dann sind $X_1, \ldots, X_d$ unabhängig, wenn*

$$\varphi_X(\xi) = \varphi_{X_1}(\xi_1) \cdots \varphi_{X_d}(\xi_d), \quad \forall \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d.$$

Bemerkung 5.28. *Sei X eine $\mathbb{R}^d$-wertige Zufallsvariable, A eine $m \times d$ Matrix und $b \in \mathbb{R}^m$. Dann hat $Y := AX + b$ die charakteristische Funktion*

$$\varphi_Y(\xi) = e^{i\xi \cdot b} \varphi_X(A^t \xi), \quad \xi \in \mathbb{R}^m$$

Beispiel 5.29. *(a) Cauchy-Verteilung: Seien X, Y unabhängig und Cauchy verteilt mit Parameter $c > 0$, d.h. diese haben die Verteilung*

$$\mu(dx) = \frac{1}{\pi} \frac{c}{c^2 + x^2} m(dx).$$

Sei $\lambda \in (0,1)$. Dann gilt

$$\varphi_{\lambda X + (1-\lambda)Y}(\xi) = \varphi_X(\lambda \xi) \varphi_Y((1-\lambda)\xi) = e^{-c\lambda|\xi|} e^{-c(1-\lambda)|\xi|} = e^{-c|\xi|}.$$

Also ist $\lambda X + (1-\lambda)Y$ wieder Cauchy verteilt mit Parameter c.

(b) Seien X, Y unabhängig und normalverteilt, d.h. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ mit $\mu_X, \mu_Y \in \mathbb{R}$ und $\sigma_X^2, \sigma_Y^2 > 0$. Dann ist

$$\varphi_{X+Y}(\xi) = \varphi_X(\xi) \varphi_Y(\xi) = e^{i\mu_X \xi - \sigma_X^2 \frac{\xi^2}{2}} e^{i\mu_Y \xi - \sigma_Y^2 \frac{\xi^2}{2}} = e^{i(\mu_X + \mu_Y)\xi - (\sigma_X^2 + \sigma_Y^2)\frac{\xi^2}{2}}.$$

Also ist $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Als nächstes zeigen wir, dass die Kovarianzmatrix und der Erwartungswert die Normalverteilung eindeutig festlegen.

Satz 5.30. *Für jedes $b \in \mathbb{R}^d$ und jede positiv (semi-)definite, symmetrische Matrix Σ gibt es genau eine Normalverteilung mit Erwartungswert b und Kovarianzmatrix Σ.*

Für den Existenzbeweis brauchen wir folgendes Lemma aus der Linearen Algebra.

Lemma 5.31. *Sei Σ eine symmetrische, positiv (semi-)definite reellwertige $d \times d$-Matrix. Dann gibt es eine reellwertige Matrix A mit $\Sigma = AA^t$.*

Beweis. Jede symmetrische, reellwertige Matrix ist orthogonal diagonalisierbar. Es gibt daher eine orthogonale Matrix S mit
$$S\Sigma S^t = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$$
wo $\lambda_1, \ldots, \lambda_d$ die Eigenwerte von Σ bezeichnen. Da Σ positiv (semi-)definit ist, sind die Eigenwerte nicht-negativ. Definiere eine neue Matrix
$$D^{\frac{1}{2}} = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d}).$$
Dann gilt $(D^{\frac{1}{2}})^t = D^{\frac{1}{2}}$ und $D^{\frac{1}{2}}(D^{\frac{1}{2}})^t = D$. Sei $A := SD^{\frac{1}{2}}S^t$. Dann gilt
$$AA^t = SD^{\frac{1}{2}}S^t(SD^{\frac{1}{2}}S^t)^t = SD^{\frac{1}{2}}S^t(S^t)^t(D^{\frac{1}{2}})^tS^t = SDS^t = \Sigma^T = \Sigma.$$
$\square$

Beweis. (Satz 5.30) Ist μ eine Normalverteilung mit Erwartungswert b und Kovarianzmatrix Σ, so gilt
$$\widehat{\mu}(\xi) = e^{i\langle \xi, b\rangle} e^{-\frac{1}{2}\langle \xi, \Sigma\xi\rangle}, \quad \xi \in \mathbb{R}^d.$$
Dieses zeigt die Eindeutigkeit. Für die Existenz wähle A mit $\Sigma = AA^t$ und benutze die Definition der Normalverteilung. $\square$

Satz 5.32. *Sei $X = (X_1, \ldots, X_d)$ eine normalverteilte Zufallsvariable mit Erwartungswert $b \in \mathbb{R}^d$ und Kovarianzmatrix Σ. Dann gelten die folgenden Aussagen:*

(a) Sei $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$ und definiere
$$Y := \sum_{k=1}^{d} a_k X_k = \langle a, X\rangle.$$
Dann ist Y eine normalverteilte Zufallsvariable mit Erwartungswert $\langle a, b\rangle$ und Varianz $\langle a, \Sigma a\rangle$.

(b) X_j sind für $j = 1, \ldots, d$ normalverteilt mit Erwartungswert b_j und Varianz Σ_{jj}.

(c) Sind die X_j paarweise unkorreliert, so ist $(X_1, \ldots, X_d)$ unabhängig.

Umgekehrt, seien $X_1, \ldots, X_d$ unabhängige, normalverteilte Zufallsvariablen mit Erwartungswert $b_j \in \mathbb{R}$ und Varianz σ_j^2. Dann ist $X = (X_1, \ldots, X_d)$ normalverteilt mit Erwartungswert $b = (b_1, \ldots, b_d) \in \mathbb{R}^d$ un Kovarianzmatrix $\Sigma_{ij} = \delta_{ij}\sigma_j^2$.

Beweis. (a) Für alle $\xi \in \mathbb{R}$ gilt
$$\varphi_Y(\xi) = \mathbb{E}(e^{i\xi Y}) = \mathbb{E}(e^{i\langle a\xi, X\rangle}) = e^{i\langle a\xi, b\rangle} e^{-\frac{1}{2}\langle a\xi, \Sigma(a\xi)\rangle} = e^{i\xi\langle a, b\rangle} e^{-\frac{1}{2}|\xi|^2\langle a, \Sigma a\rangle}.$$
(b) Betrachte $a \in \mathbb{R}^d$ gegeben durch die Koordinaten $a_k = \delta_{jk}$.
(c) Es reicht zu zeigen, dass
$$\varphi_X(\xi) = \varphi_{X_1}(\xi) \cdots \varphi_{X_d}(\xi_d), \quad \forall \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d.$$

Da X_j paarweise unkorreliert sind, folgt $\Sigma_{ij} = \operatorname{cov}(X_i, X_j) = 0$ für $i \neq j$. Daraus folgt

$$\varphi_X(\xi) = e^{i\langle \xi, b\rangle} e^{-\frac{1}{2}\langle \xi, \Sigma\xi\rangle} = e^{i\langle \xi, b\rangle} e^{-\frac{|\xi|^2}{2}}.$$

Sind $X_1, \ldots, X_d$ unabhängig, so folgt

$$\varphi_X(\xi) = \prod_{k=1}^{d} \varphi_{X_k}(\xi_k) = \prod_{k=1}^{d} e^{i\xi_k b_k} e^{-\frac{\sigma_k^2}{2}\xi_k^2} = e^{i\xi \cdot b} e^{-\frac{1}{2}\langle \xi, \Sigma\xi\rangle}.$$

□

6 Grenzwertsätze

6.1 Gesetze der großen Zahlen

Sei $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1$ eine Folge von Zufallsvariablen, $m_n := \mathbb{E}(X_n)$, $M_n := \sum\limits_{k=1}^{n} m_k$ die Erwartungswerte und $S_n := \sum\limits_{k=1}^{n} X_k$.

Definition 6.1. *Die Folge $(X_n)_{n\in\mathbb{N}}$ erfüllt das schwache Gesetz der großen Zahlen, wenn*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}|S_n - M_n| \geq \varepsilon\right) = 0, \quad \forall \varepsilon > 0.$$

Die Folge X_n erfüllt das starke Gesetz der großen Zahlen, wenn $\frac{S_n - M_n}{n} \longrightarrow 0$ fast sicher.

Sehen wir jede Variable X_n als ein Zufallsexperiment mit Ausgang X_n an, so ist $\frac{1}{n}S_n$ der Mittelwert der Ausgänge der ersten n Experimente. Folglich ist $\frac{1}{n}M_n$ der erwartete (deterministische!) Mittelwert von $\frac{1}{n}S_n$. Das Gesetz der großen Zahlen besagt, dass nach hinreichend vielen Experimenten die Ausgänge der Experimente sich im Mittel deterministisch Verhalten.

Satz 6.2. *Seien $X_n \in \mathcal{L}^2$ unkorreliert und es gelte $\operatorname{var}(X_n) \leq C < \infty$ für alle $n \in \mathbb{N}$ und eine Konstante $C > 0$. Dann gilt*

$$\frac{1}{n^2}\mathbb{E}((S_n - M_n)^2) \longrightarrow 0, \quad n \to \infty.$$

Insbesondere gilt das schwache Gesetz der großen Zahlen.

Beweis. Es gilt $\mathbb{E}(S_n) = M_n$ und folglich

$$\begin{aligned}\frac{1}{n^2}\mathbb{E}((S_n - M_n)^2) &= \frac{1}{n^2}\mathbb{E}\left((S_n - \mathbb{E}(S_n))^2\right) = \frac{1}{n}\operatorname{var}(S_n) \\ &= \frac{1}{n^2}\sum_{k=1}^{n} \operatorname{var}(X_k) \leq \frac{C}{n}.\end{aligned}$$

Aus der Tschebyscheff Ungleichung folgt

$$\mathbb{P}\left(|S_n - M_n| \geq \varepsilon n\right) \leq \frac{1}{\varepsilon^2 n^2}\mathbb{E}((S_n - M_n)^2)$$

und somit die Behauptung. □

Im folgenden wollen wir zeigen, dass unter denselben Voraussetzungen bereits das starke Gesetz der großen Zahlen gilt.

Lemma 6.3. *(stochastisch schnelle Konvergenz) Sei X_n eine Folge von Zufallsvariablen mit*

$$\sum_{k=1}^{\infty} \mathbb{P}(|X_k| \geq \varepsilon) < \infty, \quad \forall \varepsilon > 0.$$

Dann gilt $\mathbb{P}\left(\lim_{n\to\infty} X_n = 0\right) = 1$.

Beweis. Nach dem ersten Borel-Cantelli Lemma gilt

$$\mathbb{P}\left(\limsup_{n\to\infty}\{|X_n| \geq \varepsilon\}\right) = \mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{k\geq n}\{|X_k| \geq \varepsilon\}\right) = 0.$$

Insbesondere gilt

$$1 = \mathbb{P}\left(\left(\limsup_{n\to\infty}\{|X_n| \geq \varepsilon\}\right)^c\right) = \mathbb{P}\left(\bigcup_{n\geq 1}\bigcap_{k\geq n}\{|X_k| < \varepsilon\}\right).$$

D.h. für fast alle ω gibt es ein $n(\omega) \in \mathbb{N}$ sodass für alle $k \geq n(\omega)$ gilt $|X_k(\omega)| < \varepsilon$. □

Satz 6.4. *Seien $X_n \in \mathcal{L}^2$ paarweise unkorreliert mit $\mathrm{var}(X_n) \leq C < \infty$ für ein $C > 0$ und alle $n \geq 1$. Dann gilt das starke Gesetz der großen Zahlen.*

Beweis. Wir betrachten zuerst den Fall $\mathbb{E}(X_n) = 0$. In diesem Fall ist $M_n = 0$.
Schritt 1. Aus der Tschebyscheff Ungleichung folgt

$$\mathbb{P}(|S_{k^2}| \geq \varepsilon k^2) \leq \frac{1}{\varepsilon^2}\frac{1}{k^4}\mathrm{var}((S_{k^2})^2) = \frac{1}{\varepsilon^2 k^4}\sum_{j=1}^{k^2}\mathrm{var}(X_j) \leq \frac{C}{\varepsilon^2}\frac{1}{k^2}$$

Insbesondere gilt

$$\sum_{k=1}^{\infty}\mathbb{P}(|S_{k^2}| \geq k^2\varepsilon) \leq \frac{C}{\varepsilon^2}\sum_{k=1}^{\infty}\frac{1}{k^2} < \infty, \quad \forall \varepsilon > 0.$$

Nach Lemma 6.3 gibt es $N_1 \in \mathcal{F}$ mit $\mathbb{P}(N_1) = 0$ und

$$\lim_{k\to\infty} S_{k^2}(\omega) = 0, \quad \omega \in \Omega \backslash N_1.$$

Schritt 2. Definiere eine neue Zufallsvariable

$$D_{k^2} := \max_{k^2 < l \leq (k+1)^2}\left|\sum_{j=1}^{l} X_j - \sum_{j=1}^{k^2} X_j\right| = \max_{k^2 < l \leq (k+1)^2}\left|\sum_{j=k^2+1}^{l} X_j\right|.$$

Für jedes $\varepsilon > 0$ folgt aus der Tschebyscheff Ungleichung

$$\begin{aligned}
\mathbb{P}\left(\frac{D_{k^2}}{k^2} \geq \varepsilon\right) &= \mathbb{P}\left(\bigcup_{l=k^2+1}^{(k+1)^2}\left\{\left|\sum_{j=k^2+1}^{l} X_j\right| \geq \varepsilon k^2\right\}\right) \\
&\leq \sum_{l=k^2+1}^{(k+1)^2} \mathbb{P}\left(\left|\sum_{j=k^2+1}^{l} X_j\right| \geq \varepsilon k^2\right) \leq \frac{1}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2} \operatorname{var}\left(\sum_{j=k^2+1}^{l} X_j\right) \\
&\leq \frac{C}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2}(l-k^2) \leq \frac{C}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2}\left((k+1)^2-k^2\right) \\
&= \frac{C}{\varepsilon^2 k^4}\sum_{l=k^2+1}^{(k+1)^2}(2k+1) \leq \frac{4C}{\varepsilon^2 k^3}\left((k+1)^2-k^2\right) = \frac{8C}{\varepsilon^2 k^2}.
\end{aligned}$$

Nach Lemma 6.3 gibt es $N_2 \in \mathcal{F}$ mit $\mathbb{P}(N_2) = 0$ und

$$\lim_{k\to\infty} \frac{D_{k^2}(\omega)}{k^2} = 0, \quad \omega \in \Omega \backslash N_2.$$

Schritt 3. Für $n \in \mathbb{N}$ sei $k = k(n) \in \mathbb{N}$ mit $k^2 \leq n < (k+1)^2$. Dann gilt für alle $\omega \in \Omega\backslash(N_1 \cup N_2)$

$$|S_n(\omega)| \leq \frac{1}{n}\left|\sum_{j=1}^{n} X_k(\omega)\right| \leq S_{k^2}(\omega) + \frac{D_{k^2}(\omega)}{k^2}.$$

Schritt 4. Für den allgemeinen Fall sei $Y_n := X_n - \mathbb{E}(X_n)$. Dann ist $Y_n \in \mathcal{L}^2$ mit $\mathbb{E}(Y_n) = 0$ und

$$\operatorname{var}(Y_n) = \operatorname{var}(X_n) \leq C < \infty, \quad n \in \mathbb{N}.$$

Ferner gilt $\operatorname{cov}(Y_n, Y_m) = \operatorname{cov}(X_n, X_m)$ und

$$S_n - M_n = \sum_{k=1}^{n}(X_k - \mathbb{E}(X_k)) = \sum_{k=1}^{n} Y_k.$$

Die Behauptung folgt aus Schritt 1 – 3. □

Im folgenden wollen wir die Bedingung an die Momente

$$\operatorname{var}(X_n) \leq C < \infty, \quad n \in \mathbb{N}$$

weiter abschwächen. Hierfür brauchen wir das zweite Borel-Cantelli Lemma.

Lemma 6.5. *(Borel-Cantelli Lemma 2) Sei $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ eine Folge unabhängiger Mengen und es gelte $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. Dann gilt $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{k\geq n} A_k\right) = 1$.*

Beweis. Es gilt

$$\mathbb{P}((\limsup_{n\to\infty} A_n)^c) = \mathbb{P}\left(\bigcup_{n\geq 1}\bigcap_{k\geq n} A_k^c\right) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right).$$

Folglich reicht es zu zeigen, dass $\mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = 0$ für $n \geq 1$. Dieses folgt aus

$$\mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = \prod_{k=n}^{\infty}\mathbb{P}(A_k^c) = \prod_{k=n}^{\infty}(1-\mathbb{P}(A_k)) \leq \prod_{k=n}^{\infty} e^{-\mathbb{P}(A_k)} = e^{-\sum\limits_{k=n}^{\infty}\mathbb{P}(A_k)} = 0$$

wobei wir $1-x \leq e^{-x}$ für $0 \leq x \leq 1$ benutzt haben (siehe Appendix). □

Lemma 6.6. *(Kolmogorov Ungleichung) Seien $X_n \in \mathcal{L}^2$ unabhängig. Dann gilt*

$$\mathbb{P}\left(\max_{1\leq k\leq n}|S_k - M_k| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sum_{k=1}^{n}\operatorname{var}(X_k) = \frac{1}{\varepsilon^2}\operatorname{var}(S_n).$$

Beweis. Betrachte die Ereignisse

$$C_k = \left\{\left|\sum_{j=1}^{i} X_j - \sum_{j=1}^{i} m_j\right| < \varepsilon,\ 1 \leq i < k,\ \left|\sum_{j=1}^{k} X_j - \sum_{j=1}^{k} m_j\right| \geq \varepsilon\right\}.$$

Dann gilt $C_k \cap C_j = \varnothing$ für $k \neq j$ und $C = \bigcup_{k=1}^{n} C_k$ erfüllt

$$\mathbb{P}\left(\max_{1\leq k\leq n}|S_k - M_k| \geq \varepsilon\right) = \mathbb{P}(C).$$

Es gilt

$$\begin{aligned}
\operatorname{var}(S_n) &= \int_{\Omega}\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i\right)^2 d\mathbb{P} \\
&\geq \int_{C}\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i\right)^2 d\mathbb{P} = \sum_{k=1}^{n}\int_{C_k}\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i\right)^2 d\mathbb{P} \\
&= \sum_{k=1}^{n}\int_{C_k}\left(\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} m_i\right)^2 d\mathbb{P} \\
&\quad + 2\sum_{k=1}^{n}\int_{C_k}\left(\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} m_i\right)\left(\sum_{i=k+1}^{n} X_i - \sum_{i=k+1}^{n} m_i\right) d\mathbb{P} \\
&\quad + \sum_{k=1}^{n}\int_{C_k}\left(\sum_{i=k+1}^{n} X_i - \sum_{i=k+1}^{n} m_i\right)^2 d\mathbb{P}.
\end{aligned}$$

Die Zufallsvariablen $\eta_1 = \mathbb{1}_{C_k}\left(\sum\limits_{i=1}^{k} X_i - \sum\limits_{i=1}^{k} m_i\right)$ und $\eta_2 = \sum\limits_{i=k+1}^{n} X_i - \sum\limits_{i=k+1}^{n}$ sind nach Voraussetzung unabhängig. Es gilt folglich

$$\mathbb{E}(\eta_1\eta_2) = \mathbb{E}(\eta_1)\mathbb{E}(\eta_2) = 0.$$

Damit erhalten wir

$$\operatorname{var}(S_n) \geq \sum_{k=1}^{n} \int\limits_{C_k} \left(\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} m_i\right)^2 d\mathbb{P} \geq \varepsilon^2 \sum_{k=1}^{n} \mathbb{P}(C_k) = \varepsilon^2 \mathbb{P}(C).$$

□

Lemma 6.7. *Es gilt*

$$\sum_{k:\ 2^k \geq i} \frac{1}{2^{2k}} \leq \frac{16}{3}\frac{1}{i^2}, \quad i \in \mathbb{N}.$$

Beweis. Sei $q \in (0,1)$ und $n_0 \in \mathbb{N}$. Dann gilt

$$\sum_{n=n_0}^{\infty} q^n = q^{n_0} \sum_{n=n_0}^{\infty} q^{n-n_0} = \frac{q^{n_0}}{1-q}.$$

Für $q = \frac{1}{4}$ und $n_0 \in \mathbb{N}$ mit $\frac{\ln(i)}{\ln(2)} - 1 < n_0 \leq \frac{\ln(i)}{\ln(2)}$ folgt

$$\sum_{k:\ 2^k \geq i} \frac{1}{2^{2k}} \leq \sum_{k=n_0}^{\infty} 4^{-k} = \frac{4^{-n_0}}{1-\frac{1}{4}} \leq \frac{4}{3} \exp\left(-\left(\frac{\ln(i)}{\ln(2)} - 1\right) 2\ln(2)\right) = \frac{16}{3}\frac{1}{i^2}.$$

□

Satz 6.8. *(Erstes Kolmogorov Theorem) Seien $X_n \in \mathcal{L}^2$ unabhängig und es gelte*

$$\sum_{n=1}^{\infty} \frac{\operatorname{var}(X_n)}{n^2} < \infty.$$

Dann gilt das starke Gesetz der großen Zahlen.

Beweis. Wir können ohne Einschränkung der Allgemeinheit annehmen, dass $\mathbb{E}(X_n) = 0$ für alle $n \in \mathbb{N}$ gilt. Sei $\varepsilon > 0$ und

$$\begin{aligned} B(\varepsilon) &= \{\omega \in \Omega \mid \exists N(\varepsilon) \in \mathbb{N}: \quad |S_n(\omega)| < n\varepsilon, \quad \forall n \geq N(\varepsilon)\} \\ &= \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|S_n| < n\varepsilon\}. \end{aligned}$$

Setze $B_k(\varepsilon) = \left\{ \max_{2^{k-1} \le n < 2^k} \frac{1}{n}|S_n| \ge \varepsilon \right\}$. Die Kolmogorov Ungleichung impliziert

$$\begin{aligned}\mathbb{P}(B_k(\varepsilon)) &= \mathbb{P}\left(\max_{2^{k-1} \le n < 2^k} \frac{1}{n} \left| \sum_{i=1}^n X_i \right| \right) \le \mathbb{P}\left(\max_{2^{k-1} \le n < 2^k} \left| \sum_{i=1}^n X_i \right| \ge \varepsilon 2^{k-1} \right) \\ &\le \mathbb{P}\left(\max_{1 \le n \le 2^k} \left| \sum_{i=1}^n X_i \right| \ge \varepsilon 2^{k-1} \right) \le \frac{1}{\varepsilon^2 2^{2k-2}} \sum_{i=1}^{2^k} \operatorname{var}(X_i).\end{aligned}$$

Daraus folgt

$$\begin{aligned}\sum_{k=1}^\infty \mathbb{P}(B_k(\varepsilon)) &\le \sum_{k=1}^\infty \frac{1}{\varepsilon^2 2^{2k-2}} \sum_{i=1}^{2^k} \operatorname{var}(X_i) \\ &= \frac{1}{\varepsilon^2} \sum_{i=1}^\infty \operatorname{var}(X_i) \sum_{k:\ 2^k \ge i} \frac{1}{2^{2k-2}} \le \frac{32}{3\varepsilon^2} \sum_{i=1}^\infty \frac{\operatorname{var}(X_i)}{i^2}.\end{aligned}$$

Nach dem ersten Borel-Cantelli Lemma gibt es für fast alle ω ein $k_0(\omega) \in \mathbb{N}$ mit

$$\max_{2^{k-1} \le n < 2^k} \frac{1}{n}|S_n(\omega)| < \varepsilon, \quad k \ge k_0(\omega).$$

Insbesondere gilt $\mathbb{P}(B(\varepsilon)) = 1$ für alle $\varepsilon > 0$. Also gilt auch $\mathbb{P}(B(m^{-1})) = 1$ für alle $m \ge 1$ und somit $\mathbb{P}\left(\bigcap_{m \ge 1} B(m^{-1}) \right) = 1$. Für jedes $\omega \in \bigcap_{m \ge 1} B(m^{-1})$ gibt es ein $N(\omega, m) \in \mathbb{N}$ sodass für alle $n \ge N(\omega, m)$ gilt $\frac{1}{n}|S_n(\omega)| < \frac{1}{m}$. □

Eine Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ heißt identisch verteilt, wenn $\mathbb{P} \circ X_n^{-1} = \mathbb{P} \circ X_1^{-1}$ für alle $n \ge 1$ gilt. In diesem Fall setze $m := \mathbb{E}(X_n) = \mathbb{E}(X_1)$. Das starke Gesetz der großen Zahlen ist äquivalent zu

$$\frac{1}{n} \sum_{k=1}^n X_k \longrightarrow m, \quad n \to \infty$$

fast sicher. Wollen wir die Integrabilität noch weiter abschwächen, z.B. $X_n \in \mathcal{L}^1$, so müssen wir annehmen dass die Folge X_n identisch verteilt ist.

Satz 6.9. *(Zweites Kolmogorov Theorem, [KS07]) Seien $X_n \in \mathcal{L}^1$ unabhängig und identisch verteilt. Dann gilt das starke Gesetz der großen Zahlen.*

Satz 6.10. *(Satz von Etemadi, 1981, [Ete81]) Seien $X_n \in \mathcal{L}^1$ paarweise unabhängig und identisch verteilt. Dann gilt das starke Gesetz der großen Zahlen.*

6.2 Zentraler Grenzwertsatz

Sei $X_n \in \mathcal{L}^2$ eine Folge von unabhängigen Zufallsvariablen mit Erwartungswert $m_k := \mathbb{E}(X_k)$ und Varianz $\sigma_i^2 := \operatorname{var}(X_i)$. Definiere $S_n := \sum_{i=1}^n X_i$, $M_n := \sum_{i=1}^n m_i$ und $D_n^2 := \operatorname{var}(S_n) = \sum_{i=1}^n \sigma_i^2$.

Definition 6.11. *(Lindeberg Bedingung) Die Folge $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^2$ erfüllt die Lindeberg Bedingung, wenn für jedes $\varepsilon > 0$ gilt*

$$\lim_{n\to\infty} \frac{1}{D_n^2} \sum_{i=1}^{n} \int_{|X_i - m_i| \geq \varepsilon D_n} (X_i - m_i)^2 d\mathbb{P} = 0.$$

Man beachte, ist $D_n = 0$ für alle $n \in \mathbb{N}$, so ist $X_n = 0$ fast sicher für alle $n \in \mathbb{N}$. Um diesen Fall auszuschließen, können wir annehmen, dass $D_n > 0$ für hinreichend große $n \in \mathbb{N}$ gilt. Wir werden später, unter geeigneten Bedingungen, zeigen dass $D_n \longrightarrow \infty$ gilt. Vorher brauchen wir jedoch das nächste Lemma.

Lemma 6.12. *Sei* ln *die Fortsetzung des natürlichen Logarithmus auf das Gebiet*

$$G = \left\{ 1 + z \in \mathbb{C} \mid |z| < \frac{1}{4} \right\},$$

d.h. es gilt

$$\ln(1+z) = \sum_{k=1}^{n} (-1)^{k+1} \frac{z^k}{k}, \quad |z| < \frac{1}{4}.$$

Dann gibt es eine stetige Funktion $\Theta : \{z \in \mathbb{C} \mid |z| < \frac{1}{4}\} \longrightarrow \mathbb{C}$ mit $|\Theta(z)| \leq 1$ und

$$\ln(1+z) = z + \Theta(z)|z|^2, \quad |z| < \frac{1}{4}.$$

Beweis. Für $z = 0$ ist nichts zu zeigen. Wir betrachten $0 < |z| < \frac{1}{4}$. Dann gilt

$$\ln(1+z) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{z^k}{k} = z + |z|^2 \sum_{k=2}^{\infty} (-1)^{k+1} \frac{z^{k-2}}{k} \frac{z^2}{|z|^2}.$$

Setze $\Theta(z) := \sum\limits_{k=2}^{\infty} (-1)^{k+1} \frac{z^{k-2}}{k} \frac{z^2}{|z|^2}$. Wir zeigen, dass die Reihe absolut konvergiert

$$\sum_{k=2}^{\infty} \left| (-1)^{k+1} \frac{z^{k-2}}{k} \frac{z^2}{|z|^2} \right| = \sum_{k=0}^{\infty} \frac{|z|^k}{k+2} \leq \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{4^k} = \frac{2}{3} < 1.$$

□

Satz 6.13. *(zentraler Grenzwertsatz) Sei $X_n \in \mathcal{L}^2$ eine Folge von unabhängigen Zufallsvariablen und es gelte die Lindeberg Bedingung. Dann konvergiert $S_n^* := \frac{S_n - M_n}{D_n}$ schwach gegen die Normalverteilung $\mathcal{N}(0,1)$ auf $\mathbb{R}$.*

Für den Beweis können wir ohne Einschränkung der Allgemeinheit annehmen, dass $m_i = 0$. Anderseits definiere $Y_i := X_i - m_i$. Für diese Zufallsvariablen ist die Lindeberg Bedingung immernoch erfüllt. Im Beweis zeigen wir die schwache Konvergenz mittels Konvergenz der charakteristischen Funktionen. Sei φ_n die charakteristische Funktion von X_n. Das nächste Lemma liefert hierfür eine geeignete Restgliedabschätzung.

Lemma 6.14. *Es gibt Zahlen $a_k^{(n)}$ derart, dass für $1 \leq k \leq n$ gilt*

$$\varphi_k\left(\frac{\xi}{D_n}\right) = 1 - \frac{\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}$$

sowie

$$\lim_{n\to\infty}\sum_{k=1}^{n}|a_k^{(n)}(\xi)| = 0, \quad \forall \xi \in \mathbb{R}.$$

Beweis. Es gibt stetige Funktionen Θ_1, Θ_2 mit $|\Theta_1(x)|, |\Theta_2(x)| \leq 1$ sowie

$$e^{ix} = 1 + ix + \frac{\Theta_1(x)x^2}{2},$$
$$e^{ix} = 1 + ix - \frac{x^2}{2} + \frac{\Theta_2(x)x^3}{6}.$$

für alle $x \in \mathbb{R}$. Aus diesen Entwicklungen erhalten wir

$$\begin{aligned}
\varphi_k\left(\frac{\xi}{D_n}\right) &= \int_\Omega e^{i\frac{\xi}{D_n}X_k}d\mathbb{P} \\
&= \int_{|X_k|\geq\varepsilon D_n}\left(1 + \frac{i\xi}{D_n}X_k + \frac{\Theta_1(X_k)(\xi X_k)^2}{2D_n^2}\right)d\mathbb{P} \\
&\quad + \int_{|X_k|<\varepsilon D_n}\left(1 + \frac{i\xi}{D_n}X_k - \frac{\xi^2}{2D_n^2}X_k^2 + \frac{\Theta_2(X_k)(\xi X_k)^3}{6D_n^3}\right)d\mathbb{P} \\
&= 1 - \frac{\xi^2\sigma_k^2}{2D_n^2} + \frac{\xi^2}{2D_n^2}\int_{|X_k|\geq\varepsilon D_n}(1 + \Theta_1(X_k))X_k^2d\mathbb{P} + \frac{\xi^3}{6D_n^3}\int_{|X_k|<\varepsilon D_n}\Theta_2(X_k)X_k^3d\mathbb{P}.
\end{aligned}$$

Definiere

$$a_k^{(n)} := \frac{\xi^2}{2D_n^2}\int_{|X_k|\geq\varepsilon D_n}(1 + \Theta_1(X_k))X_k^2d\mathbb{P} + \frac{\xi^3}{6D_n^3}\int_{|X_k|<\varepsilon D_n}\Theta_2(X_k)X_k^3d\mathbb{P}.$$

Betrachten wir $\sum\limits_{k=1}^{n}|a_k^{(n)}|$ so erhalten wir für den ersten Term aus der Lindeberg Bedingung

$$\frac{\xi^2}{2D_n^2}\sum_{k=1}^{n}\int_{|X_k|\geq\varepsilon D_n}(1 + |\Theta_1(X_k)|)X_k^2d\mathbb{P} \leq \frac{\xi^2}{D_n^2}\sum_{k=1}^{n}\int_{|X_k|\geq\varepsilon D_n}X_k^2d\mathbb{P} \longrightarrow 0, \quad n \to \infty.$$

Für den zweiten Term erhalten wir

$$\frac{|\xi|^3}{6D_n^3}\sum_{k=1}^{n}\int_{|X_k|<\varepsilon D_n}|\Theta_2(X_k)||X_k|^3d\mathbb{P} \leq \frac{|\xi|^3}{6D_n^3}\sum_{k=1}^{n}\int_{|X_k|<\varepsilon D_n}\varepsilon D_n|X_k|^2d\mathbb{P} \leq \frac{\varepsilon|\xi|^3}{6D_n^2}\sum_{k=1}^{n}\sigma_k^2 = \frac{\varepsilon|\xi|^3}{6}.$$

Da ε beliebig war, kann dieser Term beliebig klein gemacht werden. $\square$

Eine weitere wichtige Ungleichung ist im nächsten Lemma formuliert.

Lemma 6.15. *Es gilt*

$$\lim_{n\to\infty} \max_{1\le k\le n} \frac{\sigma_k^2}{D_n^2} = 0. \tag{6.1}$$

Beweis. Für jedes $\varepsilon > 0$ gilt

$$\max_{1\le k\le n} \frac{\sigma_k^2}{D_n^2} \le \max_{1\le k\le n} \frac{1}{D_n^2} \int\limits_{|X_k|\ge\varepsilon D_n} X_k^2 d\mathbb{P} + \max_{1\le k\le n} \frac{1}{D_n^2} \int\limits_{|X_k|\le\varepsilon D_n} X_k^2 d\mathbb{P}.$$

Der erste Term konvergiert für $n \to \infty$ wegen der Lindeberg Bedingung gegen 0. Für den zweiten Term gilt

$$\max_{1\le k\le n} \frac{1}{D_n^2} \int\limits_{|X_k|\le\varepsilon D_n} X_k^2 d\mathbb{P} \le \varepsilon^2.$$

Da ε beliebig war, folgt die Behauptung. □

Kommen wir zum Beweis vom zentralen Grenzwertsatz.

Beweis. (Satz 6.13) Es reicht zu zeigen, dass

$$\varphi_{\tau_n}(\xi) \longrightarrow e^{-\frac{\xi^2}{2}}, \quad \xi \in \mathbb{R}.$$

Fixiere $\xi \in \mathbb{R}$. Dann gilt

$$\varphi_{\tau_n}(\xi) = \mathbb{E}(e^{i\xi\tau_n}) = \mathbb{E}\left(e^{i\frac{\xi}{D_n}\sum\limits_{k=1}^n X_k}\right) = \prod_{k=1}^n \varphi_k\left(\frac{\xi}{D_n}\right).$$

Daraus folgt

$$\ln\left(\varphi_{\tau_n}(\xi)\right) = \sum_{i=1}^n \ln\left(\varphi_i\left(\frac{\xi}{D_n}\right)\right)$$

und somit reicht es

$$\sum_{i=1}^n \ln\left(\varphi_i\left(\frac{\xi}{D_n}\right)\right) \longrightarrow -\frac{\xi^2}{2}, \quad n\to\infty \tag{6.2}$$

zu zeigen. Für hinreichend große $n \in \mathbb{N}$ ist $|z_{k,n}| < \frac{1}{4}$ für $z_{k,n} = -\frac{\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}$ (siehe (6.1). Daraus folgt

$$\begin{aligned}
\sum_{k=1}^n \ln\left(\varphi_k\left(\frac{\xi}{D_n}\right)\right) &= \sum_{k=1}^n \ln\left(1 - \frac{\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}\right)\\
&= -\frac{\xi^2}{2D_n^2}\sum_{k=1}^n \sigma_k^2 + \sum_{k=1}^n a_k^{(n)} + \sum_{k=1}^n \Theta_{k,n}\left|\frac{-\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}\right|^2\\
&= -\frac{\xi^2}{2} + \sum_{k=1}^n a_k^{(n)} + \sum_{k=1}^n \Theta_{k,n}\left|\frac{-\xi^2\sigma_k^2}{2D_n^2} + a_k^{(n)}\right|^2
\end{aligned}$$

mit $\Theta k, n = \Theta(z_{k,n})$, $|\Theta_{k,n}| \leq 1$. Es reicht daher zu zeigen, dass der letzte Term gegen Null konvergiert. Es gilt

$$\sum_{k=1}^{n} |\Theta_{k,n}| \left| \frac{-\xi^2 \sigma_k^2}{2D_n^2} + a_k^{(n)} \right|^2 \leq \max_{1 \leq k \leq n} \left(\frac{\xi^2 \sigma_k^2}{2D_n^2} + |a_k^{(n)}| \right) \sum_{k=1}^{n} \left(\frac{\xi^2 \sigma_k^2}{2D_n^2} + |a_k^{(n)}| \right).$$

Die Summe ist wegen

$$\sum_{k=1}^{n} \left(\frac{\xi^2 \sigma_k^2}{2D_n^2} + |a_k^{(n)}| \right) = \frac{\xi^2}{2} + \sum_{k=1}^{n} |a_k^{(n)}|$$

offensichtlich beschränkt. Der erste Faktor konvergiert gegen Null, welches die Behauptung zeigt.

□

Bemerkung 6.16. *Sei $K \subset \mathbb{R}$ kompakt. Dann gilt mit denselben Notationen wie im letzten Beweis*

$$\lim_{n \to \infty} \sup_{\xi \in K} \left| \varphi_{\mathcal{T}_n}(\xi) - e^{-\frac{\xi^2}{2}} \right| = 0.$$

Der Beweis ist Übung.

Korollar 6.17. *Sei $X_n \in \mathcal{L}^2$ ein Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $m := \mathbb{E}(X_1)$ und $0 < \sigma^2 := \operatorname{var}(X_1)$. Dann konvergiert $\frac{S_n - nm}{\sqrt{n\sigma^2}}$ schwach gegen die Normalverteilung.*

Beweis. Da alle X_n die gleiche Verteilung haben folgt $\sigma_i^2 = \sigma_1^2$ und somit $D_n = \sqrt{n\sigma^2}$. Ebenso gilt $M_n = mn$ und

$$\int\limits_{|X_k - m_k| \geq \varepsilon D_n} (X_k - m_k)^2 d\mathbb{P} = \int\limits_{|X_1 - m_1|} (X_1 - m_1)^2 d\mathbb{P}.$$

Die Lindeberg Bedingung folgt daher aus

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int\limits_{|X_k - m_k| \geq \varepsilon D_n} (X_k - m_k)^2 d\mathbb{P} = \frac{1}{\sigma^2} \int\limits_{|X_1 - m_1| \geq \varepsilon \sqrt{\sigma^2 n}} (X_1 - m_1)^2 d\mathbb{P}$$

und dem Satz von Lebesgue wegen $X_1 - m_1 \in \mathcal{L}^2$. □

Definition 6.18. *(Lyapunov Bedingung) Eine Folge unabhängiger Zufallsvariablen $X_n \in \mathcal{L}^2$ erfüllt die Lyapunov Bedingung, wenn es ein $\delta > 0$ gibt derart, dass*

$$\lim_{n \to \infty} \frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}(|X_k - m_k|^{2+\delta}) = 0.$$

Korollar 6.19. *(zentraler Grenzwertsatz, Lyapunov Bedingung) Sei $X_n \in \mathcal{L}^2$ eine Folge unabhängiger Zufallsvariablen, welche die Lyapunov Bedingung erfüllen. Dann konvergiert $\frac{S_n - M_n}{D_n}$ schwach gegen die Normalverteilung $\mathcal{N}(0, 1)$.*

Beweis. Sei $\varepsilon > 0$ und $\delta > 0$ gegeben durch die Lyapunov Bedingung. Dann gilt

$$\begin{aligned}\frac{1}{D_n^2}\int\limits_{|X_k-m_k|\geq\varepsilon D_n}(X_k-m_k)^2 d\mathbb{P} &\leq \frac{1}{D_n^2(\varepsilon D_n)^\delta}\int\limits_{|X_k-m_k|\geq\varepsilon D_n}(X_k-m_k)^{2+\delta}d\mathbb{P}\\ &\leq \varepsilon^{-\delta}\frac{\mathbb{E}(|X_k-m_k|^{2+\delta})}{D_n^{2+\delta}}.\end{aligned}$$

Folglich gilt die Lindeberg Bedingung. □

Satz 6.20. *(Lindeberg-Feller, [KS07]) Sei $X_n \in \mathcal{L}^2$ eine Folge unabhängiger Zufallsvariablen mit*

$$\lim_{n\to\infty}\max_{1\leq i\leq n}\frac{\sigma_i^2}{D_n^2} = 0.$$

Dann sind die folgenden Aussagen äquivalent

- *Die Folge X_n erfüllt die Lindeberg Bedingung.*
- *Die Folge $\frac{S_n-M_n}{D_n}$ konvergiert schwach gegen die Normalverteilung $\mathcal{N}(0,1)$.*

6.3 Weitere Grenzwertsätze

Sei $p(x) \geq 0$ eine Wahrscheinlichkeitsdichte. Wir schreiben

$$p(x) \sim q(x), \quad |x| \longrightarrow \infty$$

falls es ein $R > 0$ und eine beschränkte Funktion g gibt mit $g(x) \longrightarrow 0$ für $|x| \longrightarrow \infty$ und

$$p(x) = q(x)(1+g(x)), \quad |x| \geq R.$$

Lemma 6.21. *Sei p eine Wahrscheinlichkeitsdichte mit $p(x) = p(-x)$,*

$$p(x) \sim \frac{c}{|x|^{\alpha+1}}, \quad |x| \longrightarrow \infty$$

und $0 < \alpha < 2$, $c > 0$. Sei φ die charakteristische Funktion von p. Dann gilt

$$\varphi(\xi) = 1 - c_1|\xi|^\alpha + o(|\xi|^\alpha), \quad \xi \to 0. \tag{6.3}$$

Die Behauptung besagt, dass es eine Konstante c_1 gibt derart, dass

$$\frac{\varphi(\xi) - 1 + c_1|\xi|^\alpha}{|\xi|^\alpha} \longrightarrow 0, \quad \xi \to 0.$$

Äquivalent, es gibt eine Funktion h mit $\frac{h(\xi)}{|\xi|^\alpha} \longrightarrow 0$ für $\xi \to 0$ und es gilt

$$\varphi(\xi) = 1 + c_1|\xi|^\alpha + h(\xi)$$

für alle hinreichend kleinen $\xi \in \mathbb{R}$. Wegen

$$\varphi(-\xi) = \int\limits_{\mathbb{R}} p(x)e^{-i\xi x}m(dx) = \int\limits_{\mathbb{R}} p(-x)e^{i\xi x}m(dx) = \varphi(\xi)$$

reicht es die Asymptotik nur für positive ξ zu betrachten. Im folgenden schreiben wir der einfachhalber Lebesgue Integrale wie gewöhnliche Riemann Integrale.

Beweis. Sei $R > 0$ derart, dass

$$p(x) = \frac{c}{|x|^{\alpha+1}}(1 + g(x)), \quad |x| \geq R$$

für eine beschränkte Funktion g mit $g(x) \longrightarrow 0$ für $|x| \to \infty$. Der einfachhalber betrachten wir $\xi > 0$. Für $0 < \xi < \frac{1}{R}$ gilt

$$\begin{aligned}
\varphi(\xi) &= \int_{-\infty}^{-\frac{1}{\xi}} p(x)e^{i\xi x}dx + \int_{-\frac{1}{\xi}}^{-R} p(x)e^{i\xi x}dx + \int_{-R}^{R} p(x)e^{i\xi x}dx + \int_{R}^{\frac{1}{\xi}} p(x)e^{i\xi x}dx + \int_{\frac{1}{\xi}}^{\infty} p(x)e^{i\xi x}dx \\
&= I_1(\xi) + I_2(\xi) + I_3(\xi) + I_4(\xi) + I_5(\xi).
\end{aligned}$$

Für I_3 sehen wir, dass I_3 beliebig oft stetig differenzierbar ist mit $I_3(0) = \int_{-R}^{R} p(x)dx$ und wegen $p(x) = p(-x)$ ist auch

$$I_3'(0) = \int_{-R}^{R} p(x)ixdx = 0.$$

Daraus folgt

$$I_3(\xi) = I_3(0) + I_3'(0)\xi + O(\xi^2) = \int_{-R}^{R} p(x)dx + O(\xi^2), \quad \xi \to 0.$$

Da für jede Funktion $h \in O(\xi^2)$ bereits $h \in o(|\xi|^\alpha)$ gilt, folgt

$$I_3(\xi) = \int_{-R}^{R} p(x)dx + o(|\xi|^\alpha), \quad \xi \to 0.$$

Für I_1 erhalten wir mit $y := \xi x$ und $dy = \xi dx$

$$\begin{aligned}
I_1(\xi) &= \int_{-\infty}^{-\frac{1}{\xi}} p(x)e^{i\xi x}dx = \int_{-\infty}^{-\frac{1}{\xi}} \frac{c(1+g(x))}{|x|^{\alpha+1}}e^{i\xi x}dx = c\xi^\alpha \int_{-\infty}^{-1} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}e^{iy}dy \\
&= c\xi^\alpha \int_{-\infty}^{-1} \frac{e^{iy}}{|y|^{\alpha+1}}dy + c\xi^\alpha \int_{-\infty}^{-1} \frac{e^{iy}g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}dy.
\end{aligned}$$

Da g beschränkt ist und $g\left(\frac{y}{\xi}\right) \longrightarrow 0$ für $\xi \to 0$, folgt mit dem Satz von Lebesgue

$$I_1(\xi) = c\xi^\alpha \int_{-\infty}^{-1} \frac{e^{iy}}{|y|^{\alpha+1}}dy + o(|\xi|^\alpha).$$

Analog erhalten wir

$$I_5(\xi) = \int_{\frac{1}{\xi}}^{\infty} p(x) e^{i\xi x} dx = \int_{\frac{1}{\xi}}^{\infty} \frac{c(1+g(x))}{|x|^{\alpha+1}} e^{i\xi x} dx = c\xi^{\alpha} \int_{1}^{\infty} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}} e^{iy} dy$$

$$= c\xi^{\alpha} \int_{1}^{\infty} \frac{e^{iy}}{|y|^{\alpha+1}} dy + c\xi^{\alpha} \int_{1}^{\infty} \frac{e^{iy} g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}} dy = c\xi^{\alpha} \int_{1}^{\infty} \frac{e^{iy}}{|y|^{\alpha+1}} dy + o(|\xi|^{\alpha}).$$

Für die letzten zwei Terme erhalten wir aus $p(x) = p(-x)$

$$I_2(\xi) + I_4(\xi) = \int_{-\frac{1}{\xi}}^{-R} p(x)(e^{i\xi x} - 1 - i\xi x) dx + \int_{R}^{\frac{1}{\xi}} p(x)(e^{i\xi x} - 1 - i\xi x) dx$$

$$+ \int_{-\frac{1}{\xi}}^{-R} p(x) dx + \int_{R}^{\frac{1}{\xi}} p(x) dx.$$

Der dritte Term liefert

$$\int_{-\frac{1}{\xi}}^{-R} p(x) dx = \int_{-\infty}^{-R} p(x) dx - \int_{-\infty}^{-\frac{1}{\xi}} \frac{c(1+g(x))}{|x|^{\alpha+1}} dx = \int_{-\infty}^{-R} p(x) dx - c\xi^{\alpha} \int_{-\infty}^{-1} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}} dy$$

$$= \int_{-\infty}^{-R} p(x) dx - c\xi^{\alpha} \int_{-\infty}^{-1} \frac{1}{|y|^{\alpha+1}} dy + o(|\xi|^{\alpha}).$$

Für den vierten Term erhalten wir analog

$$\int_{R}^{\frac{1}{\xi}} p(x) dx = \int_{R}^{\infty} p(x) dx - \int_{\frac{1}{\xi}}^{\infty} \frac{c(1+g(x))}{|x|^{\alpha+1}} dx = \int_{R}^{\infty} p(x) dx - c\xi^{\alpha} \int_{1}^{\infty} \frac{1}{|y|^{\alpha+1}} dy + o(|\xi|^{\alpha}).$$

Für den ersten Term benutzen wir $e^{iy} - 1 - iy = \frac{\Theta_1(y)}{2} y^2$ mit $|\Theta_1(y)| \leq 1$. Weiterhin ist $\alpha - 1 \in$

$(-1,1)$ und somit ist $\frac{\Theta_1(y)}{|y|^{\alpha-1}}$ auf $[-1,0]$ integrierbar. Daraus folgt

$$\begin{aligned}
\int_{-\frac{1}{\xi}}^{-R} p(x)(e^{i\xi x}-1-i\xi x)dx &= c\xi^\alpha \int_{-1}^{-\xi R} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}(e^{iy}-1-iy)dy \\
&= \frac{c\xi^\alpha}{2}\int_{-1}^{-\xi R} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha+1}}\Theta_1(y)|y|^2 dy \\
&= \frac{c\xi^\alpha}{2}\int_{-1}^{0} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha-1}}\Theta_1(y)dy - \frac{c\xi^\alpha}{2}\int_{-\xi R}^{0} \frac{1+g\left(\frac{y}{\xi}\right)}{|y|^{\alpha-1}}\Theta_1(y)dy \\
&= \frac{c\xi^\alpha}{2}\int_{-1}^{0} \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy + o(|\xi|^\alpha).
\end{aligned}$$

Analog zeigen wir

$$\int_{R}^{\frac{1}{\xi}} p(x)(e^{i\xi x}-1-i\xi x)dx = \frac{c\xi^2}{2}\int_0^1 \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy + o(|\xi|^\alpha).$$

Setze

$$c_1 := c\left(2\int_1^\infty \frac{\cos(y)-1}{|y|^{\alpha+1}}dy + \frac{1}{2}\int_0^1 \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy + \frac{1}{2}\int_{-1}^0 \frac{\Theta_1(y)}{|y|^{\alpha-1}}dy\right).$$

Dann gilt wegen $\int_{\mathbb{R}} p(x)m(dx) = 1$

$$\varphi(\xi) = 1 + c_1\xi^\alpha + o(|\xi|^\alpha).$$

□

Lemma 6.22. *Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge reeller Zahlen mit $x_n \longrightarrow x$ für $n \to \infty$. Dann gilt*

$$\lim_{n\to\infty}\left(1+\frac{x_n}{n}\right)^n = e^x.$$

Beweis. Sei $A > 0$ mit $|x_n|, |x| \leq A$ für alle $n \in \mathbb{N}$. Wegen

$$\left|\left(1+\frac{x_n}{n}\right)^n - e^x\right| \leq \left|\left(1+\frac{x_n}{n}\right)^n - \left(1+\frac{x}{n}\right)^n\right| + \left|\left(1+\frac{x}{n}\right)^n - e^x\right|$$

reicht es zu zeigen, dass der erste Term auf der rechten Seite gegen Null konvergiert. Das Letztere folgt aus

$$\begin{aligned}
\left|\left(1+\frac{|x_n|}{n}\right)^n - \left(1+\frac{x}{n}\right)^n\right| &\leq \sum_{k=1}^n \left|\left(1+\frac{x_n}{n}\right) - \left(1+\frac{x}{n}\right)\right| \cdot \left(1+\frac{A}{n}\right)^{n-k}\left(1+\frac{A}{n}\right)^{k-1} \\
&= |x-x_n|\left(1+\frac{A}{n}\right)^{n-1}.
\end{aligned}$$

$\square$

Satz 6.23. *Sei X_n eine Folge von unabhängigen Zufallsvariablen mit Verteilung*

$$\mathbb{P} \circ X_n^{-1}(dx) = p(x)dx, \quad n \in \mathbb{N}.$$

Es gelte $p(x) = p(-x)$ und für ein $c > 0$, $0 < \alpha < 2$

$$p(x) \sim \frac{c}{|x|^{\alpha+1}}, \quad |x| \to \infty.$$

Setze $S_n := \frac{1}{n^{\frac{1}{\alpha}}} \sum\limits_{k=1}^{n} X_k$. Dann gilt

$$\mathbb{P} \circ S_n^{-1} \longrightarrow \nu, \quad n \to \infty$$

schwach und die Verteilung ν ist bestimmt durch die charakteristische Funktion $\psi(\xi) = e^{-c_1|\xi|^\alpha}$.

Beweis. Es sei $\varphi(\xi) = \int\limits_{\mathbb{R}} e^{i\xi x} p(x) m(dx)$. Dann gilt wegen Unabhängigkeit

$$\mathbb{E}(e^{i\xi S_n}) = \left(\varphi\left(\frac{\xi}{n^{\frac{1}{\alpha}}}\right)\right)^n.$$

Für jedes $\xi \in \mathbb{R}$ gibt nach es (6.3) eine Folge $a_n(\xi)$ mit $na_n(\xi) \longrightarrow 0$ für $n \to \infty$ derart, dass

$$\varphi\left(\frac{\xi}{n^{\frac{1}{\alpha}}}\right) = 1 - \frac{c_1|\xi|^\alpha}{n} + a_n(\xi).$$

Wir erhalten somit

$$\begin{aligned}\left(\varphi\left(\frac{\xi}{n^{\frac{1}{\alpha}}}\right)\right)^n &= \left(1 - \frac{c_1|\xi|^\alpha}{n} + a_n(\xi)\right)^n \\ &= \left(1 + \frac{na_n(\xi) - c_1|\xi|^\alpha}{n}\right)^n.\end{aligned}$$

Wegen Lemma 6.22 mit $x_n := na_n(\xi) - c_1|\xi|^\alpha \longrightarrow -c_1|\xi|^\alpha =: x$ gilt

$$\mathbb{E}(e^{i\xi S_n}) \longrightarrow e^{-c_1|\xi|^\alpha}, \quad n \to \infty.$$

Die Behauptung folgt somit aus Levys Stetigkeitssatz. $\square$

Bemerkung 6.24. *(a) Sei $X_n \in \mathcal{L}^2$ eine Folge unabhängig, identisch verteilter Zufallsvariablen mit $\mathbb{E}(X_1) = 0$ und $\mathbb{E}(X_1^2) = 1$. Sei $M_n := \max\left\{\frac{|X_k|}{\sqrt{n}}, \dots, \frac{|X_k|}{\sqrt{n}}\right\}$. Dann gilt*

$$\mathbb{P}(M_n \leq t) = \mathbb{P}\left(\frac{|X_k|}{\sqrt{n}} \leq t, \quad k = 1, \dots, n\right) = \prod_{k=1}^{n} \mathbb{P}\left(\frac{|X_k|}{\sqrt{n}} \leq t\right) = \mathbb{P}(|X_1| \leq \sqrt{n}t)^n.$$

Weiterhin erhalten wir

$$\begin{aligned}\mathbb{P}(|X_k| > \sqrt{n}t) &= \int\limits_{\mathbb{R}} \mathbb{1}_{|y|>\sqrt{n}t} \mathbb{P} \circ X_1^{-1}(dy) \\ &\leq \int\limits_{\mathbb{R}} \mathbb{1}_{|y|>\sqrt{n}t} \frac{y^2}{nt^2} \mathbb{P} \circ X_1^{-1}(dy) \\ &= \frac{1}{nt^2} \int\limits_{|y|>\sqrt{n}t} y^2 \mathbb{P} \circ X_1^{-1}(dy) = \frac{\Theta_n(t)}{n}\end{aligned}$$

mit $\Theta_n(t) \longrightarrow 0$ *für alle* $t > 0$. *Daraus folgt*

$$\begin{aligned}\mathbb{P}(M_n \leq t) = \mathbb{P}(|X_1| \leq \sqrt{n}t)^n &= (1 - \mathbb{P}(|X_1| > \sqrt{n}t))^n \\ &\geq \left(1 - \frac{\Theta_n(t)}{n}\right)^n \longrightarrow 1, \quad n \to \infty.\end{aligned}$$

Also gilt $\lim\limits_{n\to\infty} \mathbb{P}(M_n \leq t) = 1$ *für alle* $t > 0$ *und somit*

$$\mathbb{P} \circ M_n^{-1} \longrightarrow \delta_0, \quad n \to \infty$$

schwach.
Konsequenz: *Der wesentliche Beitrag im zentralen Grenzwertsatz kommt von der Summe und nicht von dem maximalen Summanden.*

(b) *Sei* X_n *wie in Satz 6.23. Für* $t > 0$ *gilt dann*

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq k\leq n} |X_k| \geq tn^{\frac{1}{\alpha}}\right) &= 1 - \mathbb{P}\left(\max_{1\leq k\leq n} |X_k| < tn^{\frac{1}{\alpha}}\right) \\ &= 1 - \left(\mathbb{P}\left(|X_1| < tn^{\frac{1}{\alpha}}\right)\right)^n \\ &= 1 - \left(1 - \mathbb{P}\left(|X_1| \geq tn^{\frac{1}{\alpha}}\right)\right)^n.\end{aligned}$$

Ferner gilt für hinreichend große $n \in \mathbb{N}$

$$\mathbb{P}(|X_1| \geq tn^{\frac{1}{\alpha}}) \sim \int\limits_{|x|\geq tn^{\frac{1}{\alpha}}} \frac{c}{|x|^{\alpha+1}} m(dx) = \frac{2c}{\alpha t^\alpha n}.$$

Daher gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{1\leq k\leq n} |X_k| \geq tn^{\frac{1}{\alpha}}\right) = \lim_{n\to\infty} \left(1 - \left(1 - \frac{2c}{\alpha t^\alpha n}\right)^n\right) = 1 - e^{-\frac{2c}{\alpha t^\alpha}} > 0.$$

Konsequenz: *Der meximale Term hat in diesem Fall einen Einfluss auf die Konvergenz in Satz 6.23.*

Definition 6.25. *Ein Wahrscheinlichkeitsmaß $\mu(dx)$ auf $\mathbb{R}$ heißt Poissonverteilt, wenn es ein $\lambda > 0$ gibt mit*

$$\mu(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots,$$

d.h. $\mu = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta_k$. Eine Zufallsvariable X heißt Poissonverteilt, wenn ihre Verteilung $\mathbb{P} \circ X^{-1}$ Poissonverteilt ist.

Sei $n \in \mathbb{N}$. Wir betrachten n unabhängige Münzwürfe, wo 1 für Kopf und 0 für Zahl steht. Ein Ausgang ist daher ein n-Tupel $\omega = (\omega_1, \dots, \omega_n) \in \{0,1\}^n$. Wir wählen

$$\Omega_n = \{\omega = (\omega_1, \dots, \omega_n) \in \{0,1\}^n\}$$

mit $\mathcal{F} = 2^{\Omega_n}$ und

$$\mathbb{P}_n(\{\omega\}) = p_n^{\sum_{k=1}^{n} \omega_k} (1 - p_n)^{\sum_{k=1}^{n} (1-\omega_k)}, \quad \omega \in \Omega_n.$$

Hier ist $p_n \in [0,1]$ die Wahrscheinlichkeit bei einem der Münzwürfe Kopf zu werfen. Demnach ist $1 - p_n$ die Wahrscheinlichkeit Zahl zu werfen.

Satz 6.26. *(Poissonscher Grenzwertsatz) Sei $p_n \in [0,1]$ mt $np_n \longrightarrow \lambda$ für ein $\lambda > 0$. Dann gilt*

$$\lim_{n\to\infty} \mathbb{P}_n \left(\left\{ \omega \in \Omega_n \;\middle|\; \sum_{j=1}^{n} \omega_j = k \right\} \right) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0.$$

Beweis. Sei $k \in \mathbb{N}_0$ fest. Wegen $np_n \longrightarrow \lambda$ folgt $p_n \longrightarrow 0$. Wir können daher annehmen, dass $p_n < 1$ gilt. Die Anzahl der Elemente in

$$A_k := \left\{ \omega \in \Omega_n \;\middle|\; \sum_{j=1}^{n} \omega_j = k \right\}$$

ist gleichbedeutend mit der Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Diese hat demnach $\binom{n}{k}$ Elemente. Für jedes $\omega \in A_k$ gilt

$$\mathbb{P}_n(\{\omega\}) = p_n^k (1 - p_n)^{n-k}$$

und somit

$$\mathbb{P}_n(A_k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} p_n^k \left(1 - \frac{np_n}{n}\right)^{n-k}.$$

Wegen $\left(1 - \frac{np_n}{n}\right)^k \longrightarrow 1$ für $n \to \infty$ folgt

$$\lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^{n-k} = \frac{\lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^n}{\lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^k} = e^{-\lambda}.$$

Weiterhin gilt

$$\lim_{n\to\infty} n(n-1)\cdots(n-k+1) p_n^k = \lim_{n\to\infty} (np_n)^k = \lambda^k.$$

Daraus folgt die Behauptung. $\square$

Korollar 6.27. *Sei X_n eine Folge von unabhängigen Zufallsvariablen mit*

$$\mathbb{P}(X_n = 1) = p_n, \quad \mathbb{P}(X_n = 0) = 1 - p_n, \quad n \in \mathbb{N}$$

für eine Folge $p_n \in [0,1]$ mit $np_n \longrightarrow \lambda > 0$. Sei $S_n = \sum_{j=1}^{n} X_j$. Dann gilt

$$\mathbb{P}(S_n = k) \longrightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}.$$

Die Verteilung von S_n konvergiert also gegen die Poissonverteilung.

Zum Schlußbetrachten wir ein Beispiel aus der Thermodynamik.

Beispiel 6.28. *Es sei $V_R \subset \mathbb{R}^3$ eine Box mit Kantenlänge $R > 0$ und uniformer Verteilung $\frac{1}{R^3} m(dx)$ auf V_R. Dann beschreibt $x \in V_R$ die Position eines einzelnen Atoms. Wir betrachten ein ideales Gas aus $n(R) \in \mathbb{N}$ Atomen welche nicht miteinander wechselwirken, d.h. unabhängig voneinander sind. Die Positionen der Atome seien $x_1, \ldots, x_{n(R)} \in V_R$. Der gemeinsame Zustandsraum ist $\Omega := V_R^{n(R)}$ und wir nehmen an dass alle Atome uniform verteilt sind, d.h.*

$$\mathbb{P} = \frac{1}{(R^3)^{n(R)}} m^{\otimes n(R)}.$$

Ist $U \subset V_R$ eine weitere Box, so bezeichnet

$$X_U(x_1, \ldots, x_{n(R)}) = \sum_{k=1}^{n(R)} \mathbb{1}_U(x_k)$$

die Anzahl der Atome in U. Im thermodynamischen Limes betrachten wir den Grenzfall $R \to \infty$ mit $n(R) \to \infty$ geeignet gewählt. Ziel ist es die Verteilung von X_U in diesem Limes zu bestimmen. Dieses kann als Verteilung von sehr vielen Atomen in einem großen Volumen interpretiert werden.

Die Wahrscheinlichkeit 1 Atom in U zu finden ist gegeben durch $\mathbb{P}(X_U = 1) = \frac{m(U)}{R^3} =: p_n(R)$. Es gilt

$$n(R) p_{n(R)} = \frac{n(R) m(U)}{R^3} = m(U) \frac{n(R)}{R^3}.$$

Wir betrachten die Annahme

$$\frac{n(R)}{R^3} \longrightarrow \rho, \quad R \to \infty.$$

In diesem Fall sei $\lambda = \rho m(U)$. Der Poissonsche Grenzwertsatz zeigt

$$\lim_{R \to \infty} \mathbb{P}(X_U = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0.$$

Der Wert $\rho > 0$ kann als Dichte des Gases interpretiert werden. Damit sit λ die mittlere Anzahl der Atome in dem Gebiet U.

7 Appendix

7.1 Elementare Ungleichungen

Lemma 7.1. *Für jedes $N \in \mathbb{N}_0$ gibt es eine stetige Funktion $\Theta_{N+1}(x) \in \mathbb{C}$ mit $|\Theta_{N+1}(x)| \leq 1$ derart, dass*

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + \frac{(ix)^{N+1}}{(N+1)!}\Theta_{N+1}(x), \quad x \in \mathbb{R}.$$

Beweis. Wir zeigen über Induktion, dass für alle $N \in \mathbb{N}_0$ mit $t_0 = 1$ gilt

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + (ix)^{N+1} \int_0^{t_0} \cdots \int_0^{t_N} e^{ixt_{N+1}} dt_{N+1} \cdots dt_1$$

Dann hat

$$\Theta_{N+1}(x) := (N+1)! \int_0^{t_0} \cdots \int_0^{t_N} e^{ixt_{N+1}} dt_{N+1} \cdots dt_1$$

die gewünschten Eigenschaften. Für $N = 0$ gilt offensichtlich

$$e^{ix} = 1 + ix \int_0^{t_0} e^{ixt} dt.$$

$N \longmapsto N+1$: Dann gilt nach Induktionsvoraussetzung

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + (ix)^{N+1} \int_0^{t_0} \cdots \int_0^{t_N} e^{ixt_{N+1}} dt_{N+1} \cdots dt_1.$$

Für das innere Integral gilt nach $N = 0$

$$\int_0^{t_N} e^{ixt_{N+1}} dt_{N+1} = \int_0^{t_N} 1 dt_N + ix \int_0^{t_N} \int_0^{t_{N+1}} e^{ixt_{N+2}} dt_{N+2} dt_{N+1}.$$

Daraus folgt

$$e^{ix} = \sum_{n=0}^{N} \frac{(ix)^n}{n!} + (ix)^{N+1} \int_0^{t_0} \cdots \int_0^{t_N} 1 dt_{N+1} \cdots dt_1 + (ix)^{N+2} \int_0^{t_0} \cdots \int_0^{t_{N+1}} e^{ixt_{N+2}} dt_{N+2} \cdots dt_1.$$

Die Behauptung folgt, da

$$\int_0^{t_0} \cdots \int_0^{t_N} 1 dt_{N+1} \cdots dt_1 = \frac{1}{(N+1)!}.$$

□

Lemma 7.2. *Es gilt*

$$1 - x \leq e^{-x}, \quad x \geq 0.$$

Beweis. Die Behauptung ist äquivalent zu

$$f(x) := (1-x)e^x - 1 \leq 0, \quad x \geq 0.$$

Für $x = 0$ ist gilt $f(0) = 0$. Ferner gilt für $x \geq 0$ mit

$$f'(x) = -e^x + (1-x)e^x = -xe^x \leq 0.$$

Daraus folgt mit dem Mittelwertsatz der Differentialrechnung für ein $\xi \geq 0$

$$f(x) = f(0) + xf'(\xi) = xf'(\xi) \leq 0,$$

da $x \geq 0$. □

7.2 Riemann Integral und Lebesgue Integral

Hier und im folgenden sei m das Lebesgue Maß auf $\mathbb{R}$. Eine messbare Funktion $f : (a,b) \longrightarrow \mathbb{R}$ heißt Lebesgue integrierbar, falls

$$\int_{(a,b)} |f(x)| m(dx) < \infty,$$

wo $-\infty \leq a < b \leq +\infty$.

Satz 7.3. *Sei $f : [a,b] \longrightarrow \mathbb{R}$ stetig. Dann ist $f|_{(a,b)}$ Lebesgue integrierbar und es gilt*

$$\int_a^b f(x)dx = \int_{(a,b)} f|_{(a,b)}(x) m(dx)$$

wo $\int_a^b f(x)dx$ das Riemann Integral von f bezeichnet.

7.3 Der Raum der stetigen Funktionen

Sei (E,d) ein kompakter metrischer Raum und sei $C(E)$ der Banachraum der stetigen Funktionen auf E versehen mit der Supremumsnorm $\|\cdot\|_\infty$.

Der Satz von Arzela-Ascoli

Die kompakten Mengen in einem solchen Raum werden charakterisiert durch den Satz von Arzela-Ascoli. Eine Menge $K \subset C(E)$ heisst präkompakt, falls ihr Abschluss $\overline{K}$ kompakt ist.

Satz 7.4. *[Wer00, Satz II.3.4] Sei (E,d) ein kompakter metrischer Raum und $K \subset C(E)$. Dann ist K genau dann präkompakt, wenn die folgenden Eigenschaften gelten*

- *K ist beschränkt, d.h.* $\sup_{f \in K} \|f\|_\infty < \infty$.
- *K ist gleichgradig stetig, d.h. für alle $\varepsilon > 0$ gibt es ein $\delta > 0$ sodass*

$$\sup_{f \in K} \sup_{|x-y|<\delta} |f(x) - f(y)| < \varepsilon.$$

Man beachte, dass δ unabhängig von f sein muss. Es kann jedoch sehr wohl von K abhängen.

Der Satz von Stone-Weierstraß

Es ist nicht schwer zu zeigen, dass $C(E)$ befüglich der punktweisen Multiplikation

$$(f \cdot g)(x) := f(x) \cdot g(x), \quad f, g \in C(E)$$

eine Algebra ist. Eine Teilalgebra ist eine Menge $A \subset C(E)$ mit

$$f, g \in A \Longrightarrow f \cdot g \in A.$$

Eine Menge $A \subset C(E)$ heisst punktetrennend, falls es für alle $x \neq y$ Funktionen $f, g \in A$ gibt mit $f(x) \neq g(y)$. Der Satz von Stone-Weierstraß liefert ein nützliches Kriterium wann eine Teilalgebra A dicht in $C(E)$ liegt.

Satz 7.5. *(Stone-Weierstraß) Sei (E, d) ein kompakter metrischer Raum und sei $A \subset C(E)$ eine Teilalgebra mit $1 \in A$. Dann sind äquivalent*

- *A ist dicht in $C(E)$.*
- *A ist punktetrennend.*

Rieszscher Darstellungssatz

Der Dualraum von $C(E)$ ist gegeben durch

$$C(E)^* = \{l : C(E) \longrightarrow \mathbb{R} \mid \exists C > 0 \text{ mit } |l(f)| \leq C\|f\|_\infty, \quad \forall f \in C(E)\}.$$

Es sei $C_+(E)^* \subset C(E)^*$ der Teilraum mit $l(f) \geq 0$ für alle $f \geq 0$. Sei $\mathcal{M}(E)$ der Raum aller regulären Borelmaße auf E.

Satz 7.6. *[Wer00, Theorem II.2.5] Sei (E, d) ein kompakter metrischer Raum. Definiere eine Abbildung*

$$T : \mathcal{M}(E) \longrightarrow C_+(E)^*, \quad (T\mu)(f) = \int_E f(x) d\mu(x).$$

Dann hat T die folgenden Eigenschaften

(a) T ist bijektiv.

(b) $\mu(E) = \sup_{\|f\|_\infty = 1} |(T\mu)(f)| = (T\mu)(1)$.

Literatur

[Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[Ete81] N. Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete*, 55(1):119–122, 1981.

[FL80] Wolfgang Fischer and Ingo Lieb. *Funktionentheorie*, volume 47 of *Vieweg Studium: Aufbaukurs Mathematik [Vieweg Studies: Mathematics Course]*. Friedr. Vieweg & Sohn, Braunschweig, 1980. Aufbaukurs Mathematik.

[Jac01] N. Jacob. *Pseudo differential operators and Markov processes. Vol. I*. Imperial College Press, London, 2001. Fourier analysis and semigroups.

[KK06] Yu. G. Kondratiev and O. V. Kutoviy. On the metrical properties of the configuration space. *Math. Nachr.*, 279(7):774–783, 2006.

[KS07] Leonid B. Koralov and Yakov G. Sinai. *Theory of probability and random processes*. Universitext. Springer, Berlin, second edition, 2007.

[Wer00] Dirk Werner. *Funktionalanalysis*. Springer-Verlag, Berlin, extended edition, 2000.