Wahrscheinlichkeitstheorie

Fakultät für Mathematik,
Bergische Universität Wuppertal

WS 2016/2017

Martin Friesen*

24. November 2016

*friesen@math.uni-wuppertal.de

Inhaltsverzeichnis

1 Grundbegriffe

1.1 Wahrscheinlichkeitsräume und Zufallsvariablen

Definition 1.1. *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. Ohne Angabe eines Wahrscheinlichkeitsraumes ist ein abstrakter Raum $(\Omega, \mathcal{F}, \mathbb{P})$ gemeint. Wir führen die folgende Notation ein:

- 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.

Beachte

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

und

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

Wir betrachten einige Beispiele für solche Räume.

Beispiel 1.2. *(diskrete Gleichverteilung) Angenommen wir stehen vor einer wichtigen Entscheidung. Die Menge aller möglichen Entscheidungen bezeichnen wir mit* Ω. *Wir nehmen an, dass* Ω *endlich ist, d.h.* $|\Omega| = \sum\limits_{\omega \in \Omega} 1 < \infty$. *Jedes* $\omega \in \Omega$ *entspricht der Wahl einer speziellen Entscheidung. Wir betrachten auf* Ω *die Potenz-*σ*-Algebra* $\mathcal{F} = 2^\Omega$. *Das Wahrscheinlichkeitsmaß sagt wie wahrscheinlich die Wahl* ω *ist. Angenommen jede Auswahl ist gleich wahrscheinlich, z.B. weil wir keine näheren Informationen haben. Dann ist das Wahrscheinlichkeitsmaß bereits eindeutig festgelegt und gegeben durch* $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$ *und*

$$\mathbb{P}(A) = \sum_{x \in \omega} \mathbb{P}(\{\omega\}) = \sum_{x \in \Omega} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}, \quad A \subset \Omega, \tag{1.1}$$

wo $|B| := \sum\limits_{x \in B} 1$ *die Anzahl der Elemente in einer Menge* $B \subset \Omega$ *bezeichnet.*

Wir können auch den Fall betrachten, wo nicht jeder Ausgang $\omega \in \Omega$ gleich wahrscheinlich ist.

Satz 1.3. *Sei Ω eine abzählbare Menge und $\mathcal{F}$ die Potenz-σ-Algebra auf Ω. Dann gilt:*

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

$$\mathbb{P}(A) := \sum_{\omega\in A} p(\omega), \quad A \subset \Omega \tag{1.2}$$

   ein Wahrscheinlichkeitsmaß auf Ω.

2. *Jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf Ω ist von der Form* (1.2) *mit $p(\omega) := P(\{\omega\})$.*

Beweis. 1. $\mathbb{P}$ gegeben durch (1.2) lässt sich umschreiben zu

$$\mathbb{P}(\mathrm{d}\omega) = \sum_{\eta\in\Omega} \delta_\eta(\mathrm{d}\omega)p(\eta),$$

denn es gilt

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega) = \sum_{\eta\in\Omega} p(\eta)\mathbb{1}_A(\eta) = \sum_{\eta\in\Omega} p(\eta)\delta_\eta(A).$$

2. Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf Ω. Sei $A \subset \Omega$, dann gilt

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

und die Behauptung folgt mit $p(\omega) := \mathbb{P}(\{\omega\})$. □

Wir betrachten noch Beispiele für überabzählbare Ω.

Beispiel 1.4. *(uniforme Verteilung) Es sei $\Omega \subset \mathbb{R}^d$ messbar. Wir betrachten einen Dartwurf mit Zielgebiet Ω. Für eine messbare Menge $A \subset \Omega$ ist $\mathbb{P}(A)$ die Wahrscheinlichkeit das Gebiet A zu treffen. Die uniforme Verteilung besagt, dass die Wahrscheinlichkeit überall gleich verteilt ist. Analog zu* (1.1) *definieren wir daher*

$$\mathbb{P}(A) = \frac{\int_A 1\mathrm{d}x}{\int_\Omega 1\mathrm{d}x} =: \frac{vol(A)}{vol(\Omega)},$$

also $\mathbb{P}(\mathrm{d}x) = \frac{1}{vol(\Omega)}\mathrm{d}x$, wo $\mathrm{vol}(\Omega) = m(\Omega)$ das Lebesgue-Maß von Ω bezeichnet.

Haben wir ein Ziel auf welches der Dart geworfen werden soll, so ist davon auszugehen, dass die Verteilung des Darts nicht mehr uniform verteilt ist. Eine bessere Beschreibung auf $\mathbb{R}$ ist unten beschrieben.

Beispiel 1.5. *(Cauchy Verteilung) Es sei $\Omega = \mathbb{R}$ und $\mathbb{P}(\mathrm{d}x) = p(x)\mathrm{d}x$ mit $c > 0$, $a \in \mathbb{R}$ und*

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

Dann ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß und es gilt

$$\int_{\mathbb{R}} |x|^k p(x)\mathrm{d}x = \infty, \quad k \in \mathbb{N}.$$

Eine weitere Wahl auf $\mathbb{R}$ ist die Gaussverteilung bzw. Normalverteilung.

Beispiel 1.6. *(Normalverteilung) Es sei $\Omega = \mathbb{R}$ und $\mathbb{P}(\mathrm{d}x) = p(x)\mathrm{d}x$ mit $\mu \in \mathbb{R}$ sowie $\sigma > 0$ und*

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

Dann ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß mit

$$\int_{\mathbb{R}^d} xp(x)\mathrm{d}x = \mu$$

und

$$\int_{\mathbb{R}} (x-\mu)^2 p(x)\mathrm{d}x = \sigma^2.$$

Definition 1.7. *Seien $\mathbb{P}, \mathbb{Q}$ zwei Wahrscheinlichkeitsmaße auf Ω. $\mathbb{Q}$ heißt absolut stetig bezüglich $\mathbb{P}$, falls*

$$\mathbb{P}(A) = 0 \Longrightarrow \mathbb{Q}(A) = 0.$$

Sei $g : \Omega \longrightarrow \mathbb{R}_+$ integrierbar bezüglich $\mathbb{P}$ und sei

$$\mathbb{Q}(A) := \int_A g(\omega)\mathrm{d}\mathbb{P}(\omega), \quad A \in \mathcal{F}.$$

Dann ist $\mathbb{Q}$ ein Maß und $\mathbb{Q}$ ist absolut stetig bezüglich $\mathbb{P}$. Die Umkehrung ist im nächsten Satz formuliert.

Satz 1.8. *(Satz von Radon-Nikodym) Seien $\mathbb{P}, \mathbb{Q}$ zwei Wahrscheinlichkeitsmaße auf Ω und sei $\mathbb{Q}$ absolut stetig bezüglich $\mathbb{P}$. Dann gibt es eine integrierbare Funktion $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ derart dass*

$$\mathbb{Q}(A) = \int_A \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \mathrm{d}\mathbb{P}, \quad A \in \mathcal{F}. \tag{1.3}$$

Ist $g : \Omega \longrightarrow \mathbb{R}_+$ eine weitere integrierbare Funktion, sodass (1.3) gilt. Dann ist $g = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ fast sicher bezüglich $\mathbb{P}$. Wir schreiben auch $\mathbb{Q} = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\mathbb{P}$. Diese Funktion wird Radon-Nikodym Ableitung von $\mathbb{Q}$ bezüglich $\mathbb{P}$ genannt.

Häufig wird $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ auf als Dichte von $\mathbb{Q}$ bezüglich $\mathbb{P}$ bezeichnet. Der Fall $\Omega = \mathbb{R}^d$ ist von besonderer Bedeutung.

Definition 1.9. *Ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathbb{R}^d$ hat eine Dichte, falls es absolut stetig bezüglich des Lebesgue Maßes $m(\mathrm{d}x) = \mathrm{d}x$ ist. Die Dichte p ist definiert als die Radon-Nikodym Ableitung $p = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}m}$. Folglich gilt für jede messbare Menge $A \subset \mathbb{R}^d$*

$$\mathbb{P}(A) = \int\limits_A p(x)\mathrm{d}x.$$

Betrachten wir eine Folge von möglichen Experimenten, so sind wir gezwungen kompliziertere Räume Ω zu wählen. Die spezielle Wahl von Ω ergibt sich häufig aus Existenzresultaten (siehe letztes Kapitel zu stoch. Prozessen). Im Folgenden soll Ω daher nicht weiter spezifiziert werden.

Ziel: Beschreibung von zeitabhängigen Problemen (Dynamiken).
Genauer, wir wollen ein System beschreiben welches z.B. physikalischer, mathematischer, biologischer, etc. Natur ist. Ein System besteht aus verschiedenen Zuständen, die Menge aller Zustände ist der Zustandsraum E. Wir nehmen im Folgenden an, dass E ein messbarer Raum mit σ-Algebra $\mathcal{E}$ ist. Später werden wir weitere Eigenschaften fordern. Wahrscheinlichkeitsmaße auf E werden mit μ, ν bezeichnet.

Beispiel 1.10.

1. *Diskrete Zustandsräume sind häufig $E \in \{\mathbb{Z}, \mathbb{Z}^d\}$. Solche Räume können durch eine Diskretisierung entstehen und sind in manchen Fällen einfacher zu untersuchen. Auf diese Weise lassen sich z.B. Gitterstrukturen von Atomen, bzw. Spins von Atomen beschreiben.*

2. *Kontinuierliche Zustandsräume $E \in \{\mathbb{R}, \mathbb{R}^d\}$. Im Fall $E = \mathbb{R}_+$ können wir $x \in \mathbb{R}_+$ als Länge, Zeit oder Alter von Objekten vorstellen. Im Fall $E = \mathbb{R}^d$ ist $x \in \mathbb{R}^d$ z.B. die Position eines zu beschreibenden Teilchens.*

Beschreiben wir ein System mit Zustandsraum E, so wollen wir Zustände $X \in E$ als zufällige Variablen modellieren. Insbesondere ist die Position, Alter, Grösse etc. von modellierten Teilchen zufällig.

Definition 1.11. *Eine Zufallsvariable (ZV) ist eine messbare Abbildung $X : \Omega \longrightarrow E$.*

Wir sagen auch

- X ist eine E-wertige ZV.
- X ist eine ZV mit Werten in E.
- X ist eine ZV auf E.

Definition 1.12. *(Erinnerung) Sei μ ein Wahrscheinlichkeitsmaß auf E und $\varphi : E \longrightarrow E'$ wo E' ein messbarer Raum ist. Das Bildmaß $\mu \circ \varphi^{-1}$ ist definiert über*

$$\mu \circ \varphi^{-1}(A) = \mu(\varphi^{-1}(A)).$$

Es ist nicht schwer zu zeigen, dass $\mu \circ \varphi^{-1}$ ein Wahrscheinlichkeitsmaß auf E' ist. Ferner gilt für jede nicht-negative meßbare Funktion $f : E' \longrightarrow \mathbb{R}_+$

$$\int_E f(\varphi(x))\mathrm{d}\mu(x) = \int_{E'} f(y)\mu \circ \varphi^{-1}(\mathrm{d}y). \tag{1.4}$$

Sei $f : E \longrightarrow \mathbb{R}$ messbar, genau dann ist $f \circ \varphi$ integrierbar bezüglich μ, wenn f bezüglich $\mu \circ \varphi^{-1}$ integrierbar ist. In diesem Fall gilt (1.4).

Definition 1.13. *Sei X eine E-wertige Zufallsvariable. Dann ist $\mu_X := \mathbb{P} \circ X^{-1}$ die Verteilung von X.*

Wir betrachten das Beispiel der Normalverteilung auf $\mathbb{R}^d$.

Definition 1.14. *(mehrdimensionale Normalverteilung)*

1. *Die Standardnormalverteilung auf $\mathbb{R}^d$ ist gegeben durch die Dichte*

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

2. *Es sei A eine reelle $d \times d$-Matrix und $b \in \mathbb{R}^d$. Sei $T(x) = Ax + b$ die dazugehörige affine Abbildung und $\mu(\mathrm{d}x) := p(x)\mathrm{d}x$. Dann heißt $\mu \circ T^{-1}$ (allgemeine) Normalverteilung.*

3. *Sei X eine $\mathbb{R}^d$-wertige Zufallsvariable mit Verteilungsfunktion μ_X. Diese heißt normalverteilt bzw. gaußsch, wenn μ_X eine Normalverteilung ist.*

Bemerkung 1.15. *Es gibt gaußsche Zufallsvariablen. Wähle $\Omega = \mathbb{R}^d = E$ mit $X(\omega) = \omega$. Sei $\mathbb{P}$ normalverteilt, dann ist $\mu_X(\mathrm{d}x) = \mathbb{P} \circ X^{-1}$ eine Normalverteilung.*

Satz 1.16. *(a) Sei A invertierbar. Dann hat $\mu \circ T^{-1}$ eine Dichte gegeben durch*

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

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

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

Beweis. (a) Sei A invertierbar. Dann ist auch $\Sigma := AA^*$ invertierbar mit $\Sigma^{-1} = (A^*)^{-1}A$. Es ist $T^{-1}(y) = A^{-1}(y-b) = A^{-1}y - A^{-1}b$ und somit folgt mit der Translationsinvarianz von m und dem Transformationssatz für jedes Recteck R

$$m \circ T^{-1}(R) = m(T^{-1}(R)) = m(A^{-1}(R)) = \det(A^{-1})m(R) = \det(A)^{-1}m(R) \tag{1.5}$$

Rechtecke bilden einen Schnittstabilen Erzeuger der Borel-σ-Algebra, also

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

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

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

Die Behauptung folgt nun aus der folgenden Rechnung

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

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. Ist A nicht invertierbar, so ist $\mathrm{Im}(T) \subset \mathbb{R}^d$ ein höchstens $d-1$-dimensionaler Unterraum und folglich eine Nullmenge bezüglich m. Sei $B := \mathrm{Im}(T)$, dann ist $m(B) = 0$ und es reicht zu zeigen, dass $\mu \circ T^{-1}(B) \neq 0$. Dieses folgt aus $\mathbb{1}_B(T(x)) = 1$ und

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

(b) Verknüpfungen affin linearer Abbildungen sind affin linear. □

Speziell für $\mathbb{R}$-wertige ZV haben wir das nützliche Konzept einer Verteilungsfunktion.

Definition 1.17. *Sei X eine $\mathbb{R}$-wertige ZV. Die Verteilungsfunktion von X ist eine Funktion $F_X : \mathbb{R} \longrightarrow [0,1]$ gegeben durch*

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; X(\omega) \leq t\right\}\right), \quad t \in \mathbb{R}.$$

Der nächste Satz wird ohne Beweis aufgeführt.

Satz 1.18. *(a) Sei X eine $\mathbb{R}$-wertige ZV. Dann ist die Verteilungsfunktion F_X monoton wachsend, rechtsstetig mit*

$$\lim_{t \to -\infty} F_X(t) = 0, \quad \lim_{t \to \infty} F_X(t) = 1. \tag{1.6}$$

(b) Sei $F : \mathbb{R} \longrightarrow [0,1]$ monoton wachsend, rechtsstetig mit (1.6). Dann gibt es genau ein Wahrscheinlichkeitsmaß μ auf $\mathbb{R}$ mit $\mu((-\infty, t]) = F(t)$.

1.2 Erwartungswerte und Varianz

Definition 1.19. *Es sei $\mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathbb{P}) = \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ die Menge aller integrierbaren, reellwertigen Zufallsvariablen. Für jedes $X \in \mathcal{L}^1$ ist der Erwartungswert definiert über*

$$\mathbb{E}(X) = \int\limits_{\Omega} X \mathrm{d}\mathbb{P}.$$

Definition 1.20. *Wir sagen, dass eine Eigenschaft $\mathbb{P}$-fast sicher gilt, falls es ein $A \in \mathcal{F}$ gibt mit $\mathbb{P}(A) = 0$ und für alle $\omega \in A^c$ ist diese Eigenschaft erfüllt.*

Bemerkung 1.21. *Die Bedingung $X \in \mathcal{L}^1$ ist eine Bedingung an die Verteilung μ_X von X. Es gilt nämlich*

$$\mathbb{E}|X| = \int\limits_{\mathbb{R}} |x| \mathrm{d}\mu_X(x).$$

In diesem Fall kann der Erwartungswert ausgedrück werden über die Verteilung μ_X

$$\mathbb{E}(X) = \int\limits_{\mathbb{R}} x \mathrm{d}\mu_X(x).$$

Der Vektorraum $\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathbb{P})$ mit $p \in (0, \infty)$ ist definiert als der Raum der messbaren, reellwertigen Zufallsvariablen mit

$$\int\limits_{\Omega} |X|^p \mathrm{d}\mathbb{P} < \infty.$$

Für Zufallsvariablen mit Werten in E ist der Erwartungswert ohne weiteres nicht definiert. Es gilt jedoch das folgende Lemma.

Lemma 1.22. *Sei X eine E-wertige Zufallsvariable und $\varphi : E \longrightarrow \mathbb{R}$ eine messbare Funktion. Dann ist $\varphi \circ X$ messbar. Sei $\mu_X := \mathbb{P} \circ X^{-1}$ die Verteilung von X auf E. Dann*

$$\varphi \circ X \in \mathcal{L}^1(\Omega, \mathbb{P}) \Longleftrightarrow \varphi \in \mathcal{L}^1(E, \mu_X).$$

Ist das der Fall, so gilt

$$\mathbb{E}(\varphi(X)) = \int\limits_E \varphi(x) \mathrm{d}\mu_X(x).$$

Beweis. Folgt unmittelbar aus der Definition vom Bildmaß. □

Der folgende Spezialfall ist von besonderer Bedeutung. Sei X eine $\mathbb{R}^d$-wertige Zufallsvariable und $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ derart, dass $f(X) = f(X_1, \ldots, X_d)$ integrierbar ist. Dann gilt

$$\mathbb{E}(f(X)) = \int\limits_{\mathbb{R}^d} f(x_1, \ldots, x_d) \mathrm{d}\mu_X(x_1, \ldots, x_d).$$

Insbesondere gilt für $f = \mathbb{1}_A$ mit $A \in \mathcal{B}(\mathbb{R}^d)$

$$\mu_X(A) = \int\limits_{\mathbb{R}^d} \mathbb{1}_A(x) \mathrm{d}\mu_X(x) = \mathbb{E}(\mathbb{1}_A(X)) = \mathbb{P}(X \in A).$$

D.h. $\mathbb{E}(f(X))$ legt die Verteilung von X eindeutig fest.

Beispiel 1.23. *Sei X eine $\mathbb{R}$-wertige Zufallsvariable mit Verteilungsfunktion*

$$\mu_X(\mathrm{d}x) = \sum_{n=1}^{\infty} a_n \delta_{x_n}(\mathrm{d}x),$$

wo $a_n \geq 0$ mit $\sum\limits_{n=1}^{\infty} a_n = 1$ und $(x_n)_{n \in \mathbb{N}} \subset \mathbb{R}$. Dann ist

$$\int\limits_{\mathbb{R}} |x| \mu_X(\mathrm{d}x) = \sum_{n=1}^{\infty} |x_n| a_n.$$

Folglich ist X genau dann integrierbar, wenn $\sum\limits_{n=1}^{\infty} |x_n| a_n < \infty$. In diesem Fall gilt

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} x_n a_n.$$

Allgemeiner sei $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ eine messbare Funktion. Dann gilt

$$\mathbb{E}(|\varphi(X)|) < \infty \Longleftrightarrow \sum_{n=1}^{\infty} |\varphi(x_n)| a_n < \infty.$$

In diesem Fall gilt

$$\mathbb{E}(\varphi(X)) = \sum_{n=1}^{\infty} \varphi(x_n) a_n.$$

Definition 1.24. *Sei X eine $\mathbb{R}$-wertige Zufallsvariable derart, dass*

$$\mathbb{E}(|X|^n) < \infty, \quad n \in \mathbb{N}.$$

Die Momente sind definiert als

$$\mathbb{E}(X^n) = \int_\Omega X^n \mathrm{d}\mathbb{P} = \int_\mathbb{R} x^n \mathrm{d}\mu_X(x), \quad n \in \mathbb{N}_0.$$

Die zentrierten Momente sind definiert als

$$\mathbb{E}((X - \mathbb{E}(X))^n) = \int_\Omega (X - \mathbb{E}(X))^n \mathrm{d}\mathbb{P} = \int_\mathbb{R} (x - \mathbb{E}(X))^n \mathrm{d}\mu_X(x).$$

Lemma 1.25.

1. *Der Erwartungswert ist linear, d.h. sind $X, Y \in \mathcal{L}^1$ und $a, b \in \mathbb{R}$ so gilt*
$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$
2. *Ist $X \geq 0$ und $\mathbb{E}(X) = 0$ so ist $X = 0$ fast sicher.*
3. *Sei $h : \mathbb{R} \longrightarrow \mathbb{R}$ eine konvexe Funktion und sei $X \in \mathcal{L}^1$ eine Zufallsvariable auf $\mathbb{R}$ mit $h(X) \in \mathcal{L}^1$. Dann ist*
$$h(\mathbb{E}(X)) \leq \mathbb{E}(h(X)).$$

Beweis. Die erste Aussage folgt aus der Linearität des Integrals. Die Zweite ist aus der Maßtheorie bekannt. Wir zeigen die letzte Aussage. 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

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

□

Korollar 1.26. *Seien $0 < p' \leq p$, so gilt $\mathcal{L}^p(\Omega) \subset \mathcal{L}^{p'}(\Omega)$ mit*

$$(\mathbb{E}|X|^{p'})^{\frac{1}{p'}} \leq (\mathbb{E}|X|^p)^{\frac{1}{p}}.$$

Beweis. Die Funktion $h(x) := |x|^{\frac{p}{p'}}$ ist für $0 < p' \leq p$ konvex. Es folgt mit $Y := |X|^{p'}$

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

und somit die Behauptung. □

Satz 1.27. *Sei X eine Zufallsvariable mit Werten in $\mathbb{R}$ und $h \geq 0$ monoton steigend. Dann gilt für alle $c \in \mathbb{R}$*

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

Beweis. Es gilt

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

□

Der Fall $h(x) = |x|^p$ ist bekannt als Markovsche Ungleichung.

Korollar 1.28. *(Markovsche Ungleichung) Sei X eine $\mathbb{R}$-wertige Zufallsvariable, $p > 0$ und $h > 0$. Es gilt*

$$\mathbb{P}(|X| \geq h) \leq \frac{1}{h^p}\mathbb{E}(|X|^p).$$

Bemerkung 1.29. *Der Fall $p = 2$ ist als Tschebyscheff Ungleichung bekannt. Insbesondere gilt*

- $\mathbb{E}(|X|) = 0 \iff X = 0$ *fast sicher.*
- $\mathbb{E}(|X|) < \infty \implies |X| < \infty$ *fast sicher.*

Die Umkehrung der zweiten Aussage ist falsch. Denn sei

$$(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), m(\mathrm{d}x))$$

und $X(\omega) = \frac{1}{\omega}$. Dann ist $|X| < \infty$ fast sicher und

$$\mathbb{E}(|X|) = \int\limits_{[0,1]} \frac{1}{\omega} m(\mathrm{d}\omega) = \infty.$$

Satz 1.30. *(Cauchy-Schwarz Ungleichung) Sind $X, Y \in \mathcal{L}^2$ so gilt $X \cdot Y \in \mathcal{L}^1$ und*

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

Satz 1.31. *(Hölder Ungleichung) Seien $1 \leq p, q < \infty$ mit $\frac{1}{p} + \frac{1}{q} = 1$ und $X \in \mathcal{L}^p$, $Y \in \mathcal{L}^q$. Es gilt $X \cdot Y \in \mathcal{L}^1$ und*

$$\mathbb{E}(|X \cdot Y|) \leq (\mathbb{E}(|X|^p))^{\frac{1}{p}}(\mathbb{E}(|Y|^q))^{\frac{1}{q}}.$$

Definition 1.32. *Für eine Zufallsvariable $X \in \mathcal{L}^1$ ist die Varianz definiert als*

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

Für $X, Y \in \mathcal{L}^1$ mit $X \cdot Y \in \mathcal{L}^1$ ist die Konvarianz definiert über

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

X und Y heißen unkorreliert, wenn $\mathrm{cov}(X, Y) = 0$.

Die Varianz einer Zufallsvariable X ist somit das zweite zentrierte Moment der Verteilungsfunktion von X.

Lemma 1.33. *Es gelten die folgenden Rechenregeln.*

1. $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$ *für alle* $a, b \in \mathbb{R}$.
2. $\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
3. $\mathrm{var}(X) < \infty$ *genau dann, wenn* $X \in \mathcal{L}^2$.
4. $\mathrm{var}(X) = 0$ *genau dann, wenn* $X = \mathbb{E}(X)$ *fast sicher.*
5. $\mathbb{P}(|X - \mathbb{E}(X)| \geq h) \leq \frac{1}{h^2}\mathrm{var}(X)$.
6. $\mathbb{E}(X \cdot Y) =: \langle X, Y\rangle$ *definiert eine Abbildung* $\mathcal{L}^2 \times \mathcal{L}^2 \longrightarrow \mathbb{R}$ *mit den folgenden Eigenschaften:*
   - $\mathcal{L}^2 \times \mathcal{L}^2 \ni (X, Y) \longmapsto \langle X, Y\rangle$ *ist linear in jeder komponente (bilinear).*
   - $\langle X, Y\rangle = \langle Y, X\rangle$.
   - $\langle X, X\rangle \geq 0$.
   - *Ist* $\langle X, X\rangle = 0$, *so folgt* $X = 0$ *fast sicher.*
7. *Für* $X, Y \in \mathcal{L}^1$ *mit* $X \cdot Y \in \mathcal{L}^1$ *gilt* $\mathrm{cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y)$.
8. *Für* $X, Y \in \mathcal{L}^2$ *ist*
$$\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y). \tag{1.7}$$

Beweis. Übung. $\square$

Aus der Definition der Kovarianz folgt, dass $\mathrm{cov}(X, Y)$ bilinear ist.

Satz 1.34. *(Satz von Bienayme) Seien* $X_1, \dots, X_n \in \mathcal{L}^2$ *paarweise unkorreliert. Dann gilt*
$$\mathrm{var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{var}(X_k).$$

Beweis. Wir beweisen den Satz über Induktion. Der Fall $n = 2$ folgt aus (1.7). Für $n \longmapsto n + 1$ erhalten wir
$$\begin{aligned}\mathrm{var}\left(\sum_{k=1}^{n+1} X_k\right) &= \mathrm{var}\left(\sum_{k=1}^{n} X_k\right) + \mathrm{var}(X_{n+1}) + \mathrm{cov}\left(\sum_{k=1}^{n} X_k, X_{n+1}\right)\\ &= \sum_{k=1}^{n+1} \mathrm{var}(X_k) + \sum_{k=1}^{n} \mathrm{cov}(X_k, X_{n+1}) = \sum_{k=1}^{n+1} \mathrm{var}(X_k),\end{aligned}$$
wo wir in der zweiten Gleichheit benutzt haben, dass cov linear im ersten Argument ist. $\square$

Beispiel 1.35. *Sei X eine Normalverteilte Zufallsvariable auf $\mathbb{R}$, d.h. mit Verteilung*

$$\mu_X(\mathrm{d}x) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{|x-\mu|^2}{2\sigma^2}} \mathrm{d}x.$$

Dann ist $\mathbb{E}(X) = \mu$ und $\mathrm{var}(X) = \sigma^2$.

Im folgenden betrachten wir $\mathbb{R}^d$-wertige Integrale. Das nächste Lemma liefert ein nützliches Kriterium um die Integrierbarkeit von $\mathbb{R}^d$-wertigen Zufallsvariablen zu prüfen. Hierbei bezeichnet $|a| := \sqrt{\sum_{k=1}^d a_k^2}$ die euklidische Norm von $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$.

Lemma 1.36. *Es sei X eine $\mathbb{R}^d$-wertige Zufallsvariable und $p \in [1, \infty)$. Es sei $X = (X_1, \ldots, X_d)$ die Darstellung in Komponenten. Dann gilt*

$$|X| \in \mathcal{L}^p \iff X_i \in \mathcal{L}^p, \quad \forall i = 1, \ldots, d.$$

Beweis. Es reicht zu zeigen, dass

$$\int_\Omega |X|^p \mathrm{d}\mathbb{P} < \infty \iff \int_\Omega |X_i|^p \mathrm{d}\mathbb{P} < \infty, \quad \forall i = 1, \ldots, d.$$

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

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

Anderseits gilt auch

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

□

Es sei $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) = \mathcal{L}^p(\Omega; \mathbb{R}^d)$ der Raum der $\mathbb{R}^d$-wertigen Zufallsvariablen mit $\mathbb{E}(|X|^p) < \infty$. Insbesondere gilt $X \in \mathcal{L}^1(\Omega; \mathbb{R}^d)$ genau dann, wenn $X_i \in \mathcal{L}^1$. Ferner ist $X \in \mathcal{L}^2(\Omega; \mathbb{R}^d)$ genau dann, wenn $X_i \in \mathcal{L}^2$, d.h. $|X_i|^2 \in \mathcal{L}^1$ für alle $i = 1, \ldots, d$.

Bemerkung 1.37. *Es sei $X \in \mathcal{L}^2(\Omega; \mathbb{R}^d)$. Dann gilt*

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

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

Definition 1.38. *Sei $X = (X_1, \dots, X_d)^T \in \mathcal{L}^1(\Omega; \mathbb{R}^d)$. Der Erwartungswert von X ist komponentenweise definiert, d.h.*

$$\mathbb{E}(X) = \big(\mathbb{E}(X_1) \quad \dots \quad \mathbb{E}(X_d)\big)^T.$$

Ist zusätzlich $X \in \mathcal{L}^2(\Omega; \mathbb{R}^d)$, so ist die Kovarianzmatrix definiert über $\Sigma(X) := (\sigma_{ij}(X))_{ij}$ mit $\sigma_{ij}(X) = \operatorname{cov}(X_i, X_j)$.

Bemerkung 1.39. *$\Sigma(X)$ ist symmetrisch und positiv definit.*

Beweis. Seien $\lambda_1, \dots, \lambda_d \in \mathbb{R}$, dann gilt

$$0 \leq \mathbb{E}\left(\left(\sum_{i=1}^{d} \lambda_i (X_i - \mathbb{E}(X_i))\right)^2\right) = \sum_{i,j=1}^{d} \lambda_i \lambda_j \operatorname{cov}(X_i, X_j) = \langle \lambda, \Sigma(X)\lambda \rangle$$

wo $\lambda = (\lambda_1, \dots, \lambda_d)$. Die Symmetrie von $\Sigma(X)$ folgt aus der von $\sigma_{ij}(X)$, denn

$$\sigma_{ij}(X) = \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))).$$

□

Beispiel 1.40. *(a) Sei X standardnormalverteilt mit Werten in $\mathbb{R}^d$. Dann gilt $\mathbb{E}(X) = 0$ und $\operatorname{cov}(X_i, X_j) = \delta_{ij}$, wo*

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

das Kronecker-Delta Symbol bezeichnet.

(b) Sei X wie oben und $Y := AX + b$ mit $b \in \mathbb{R}^m$ und A eine reelle $m \times d$-Matrix. Es gilt $\mathbb{E}(Y) = b$ und $\Sigma(Y) = AA^T$. Ferner ist Y Normalverteilt mit Verteilung $\mu_X \circ T^{-1}$, wo $T(x) := Ax + b$ eine affin lineare Abbildung ist. Denn es gilt für alle beschränkten messbaren Funktionen $\varphi : \mathbb{R}^m \longrightarrow \mathbb{R}$

$$\mathbb{E}(\varphi(Y)) = \mathbb{E}(\varphi(T(X))) = \int_{\mathbb{R}^d} \varphi(T(x)) \mathrm{d}\mu_X(x) = \int_{\mathbb{R}^m} \varphi(y)(\mu_X \circ T^{-1})(\mathrm{d}y).$$

Betrachte jetzt $\varphi(y) = \mathbb{1}_A(y)$ für $A \in \mathcal{B}(\mathbb{R}^m)$ so folgt

$$\mathbb{P}(Y \in A) = (\mu_X \circ T^{-1})(A),$$

d.h. $\mu_Y = \mu_X \circ T^{-1}$.

1.3 Unabhängigkeit

Definition 1.41. *Sei I eine beliebige Indexmenge.*

(a) Eine Familie von Ereignissen $\{A_i \mid i \in I\} \subset \mathcal{F}$ heißt unabhängig, wenn

$$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i)$$

für alle endlichen Teilmengen $J \subset I$.

(b) Die Familie heißt paarweise unabhängig, wenn

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

Bemerkung 1.42. *Unabhängigkeit impliziert paarweise Unabhängigkeit. Die Umkehrung ist im Allgemeinen falsch.*

Beispiel 1.43. *(a) Sei $\Omega = \{1, 2, 3, 4\}$ mit $\mathcal{F} = 2^\Omega$ und Gleichverteilung $\mathbb{P}$ auf Ω. Seien*

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

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 A, B, C sind paarweise unabhängig. Die Ereignisse sind nicht unabhängig, da

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

(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 1.44. *Sei I eine beliebige Indexmenge und $\mathcal{E}_i \subset \mathcal{F}$. 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 1.45.

(a) Seien $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ mit $i = 1, \dots n$ Wahrscheinlichkeitsräume und setze $\Omega = \Omega_1 \times \Omega_n$ 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

$$\begin{aligned}
&A_1 := B_1 \times \Omega_2 \times \dots \times \Omega_n, \quad B_1 \in \mathcal{F}_1 \\
&\dots \\
&A_n := \Omega_1 \times \dots \times \Omega_{n-1} \times B_n, \quad B_n \in \mathcal{F}_n
\end{aligned}$$

Dann ist die Familie $\{A_1, \dots, A_n\}$ unabhängig unter $\mathbb{P}$. Für jede Auswahl von k-Tupeln mit $k \leq n$ gilt nämlich

$$\mathbb{P}(A_{j_1} \cap \dots \cap A_{j_k}) = \mathbb{P}_{j_1}(B_1) \cdots \mathbb{P}_{j_k}(B_k) = \mathbb{P}(A_{j_1}) \cdots \mathbb{P}(A_{j_k}).$$

(b) Sei $\{\mathcal{E}_i \mid i \in I\}$ 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 1.46. *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 1.45 sei ohne Einschränkung I endlich mit $I = \{1, \dots, n\}$. Für $i_0 \in I$ sei

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

Dann ist $\mathcal{D}_{i_0}$ ein Dynkin-System. Hierfür sei $\{i_1, \dots, i_k\} \subset I \backslash \{i_0\}$ beliebig und wähle $A_{i_\nu} \in \mathcal{E}_{i_\nu}$, wo $\nu = 1, \dots, k$. Es folgt:

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

Korollar 1.47. *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.*

Satz 1.48. *Sei $\{A_i \mid i \in I\}$ unabhängig mit $A_i \in \mathcal{F}$. Für jedes $i \in I$ sein $A_i^* \in \{\varnothing, A_i, A_i^c, \Omega\}$. 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 1.47 ist auch $\{\sigma(\mathcal{E}_i) \mid i \in I\}$ unabhängig. Die Behauptung folgt aus $\sigma(\mathcal{E}_i) = \{\varnothing, A_i, A_i^c, \Omega\}$. □

Lemma 1.49. *(Gruppieren) Sei $\{\mathcal{E}_i \mid i \in I\}$ unabhängig und $\cap$-stabil mit $\mathcal{E}_i \subset \mathcal{F}$. Sei $I = \bigcup_{j \in J} I_j$ eine disjunkte Zerlegung und $\mathcal{A}_j = \sigma\left(\bigcup_{i \in I_j} \mathcal{E}_j\right)$. Dann ist $\{\mathcal{A}_j \mid j \in J\}$ eine unabhängige Familie.*

Beweis. Für $k \in J$ sei

$$\mathcal{D}_k = \left\{ A_{i_1} \cap \dots \cap A_{i_n} \;\middle|\; A_{i_l} \in \mathcal{E}_{i_l}, \quad i_1, \dots, i_n \in I_k, \quad n \in \mathbb{N} \right\}$$

die Menge aller endlichen Schnitte von Mengen aus $\bigcup\limits_{i \in I_k} \mathcal{E}_i$. Dann ist $\mathcal{D}_k$ ist 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 1.47 ist $\{\mathcal{A}_k \mid k \in J\}$ unabhängig. $\square$

Satz 1.50. *(Kolmogoroffsches 0-1-Gesetz) Sei $(\mathcal{F}_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger σ-Algebren mit $\mathcal{F}_n \subset \mathcal{F}$. Definiere*

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

und die terminale-σ-Algebra $\mathcal{T}_\infty := \bigcap\limits_{n \geq 1} \mathcal{A}_n$. 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:** $A \in \mathcal{T}_\infty \Longrightarrow A$ ist unabhängig von A, d.h. $\mathbb{P}(A)^2 = \mathbb{P}(A)$.
Nach Korollar 1.47 und Lemma 1.49 ergibt sich für $n \in \mathbb{N}$

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

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

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

da zu jeder Menge aus $\bigcup\limits_{n \geq 1} \bigvee\limits_{k=1}^{n} \mathcal{F}_k$ es ein $n \in \mathbb{N}$ gibt, sodass diese in $\bigvee\limits_{k=1}^{n} \mathcal{F}_k$ liegt. Korollar 1.47 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 $\mathcal{T}_\infty \subset \mathcal{A}_k \subset \mathcal{A}_1 = \bigvee\limits_{k=1}^{\infty} \mathcal{F}_k$ für alle $k \in \mathbb{N}$ und somit $\mathcal{T}_\infty \perp \mathcal{T}_\infty$. $\square$

Satz 1.51. *Sei $\mathcal{T} \subset \mathcal{F}$ eine σ-Algebra mit $\mathbb{P}(A) \in \{0,1\}$ für alle $A \in \mathcal{T}$. Sei $Z : \Omega \longrightarrow \overline{\mathbb{R}}$ eine $\mathcal{T}$ messbare Zufallsvariable. Dann gibt es ein $c \in \overline{\mathbb{R}}$ mit $\mathbb{P}(Z = c) = 1$.*

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

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

Dann ist

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

Als $\mathbb{P}(Z = t_0) = \lim_{n\to\infty} \mathbb{P}\left(t_0 - \frac{1}{n} \leq Z \leq t_0 + \frac{1}{n}\right) = 1$, d.h. $c = t_0$.

$\square$

Definition 1.52. *Sei I eine beliebige Indexmenge und X_i Zufallsvariablen mit Werten in $(E_i, \mathcal{E}_i)$ für alle $i \in I$. $\{X_i \mid i \in I\}$ heißt unabhängige Familie von Zufallsvariablen, 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 1.53.

(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 auch $\{X_i \mid i \in I\}$ unabhängig.

(b) $\{X_i \mid i \in I\}$ 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 $A_{i_1} \in \mathcal{E}_{i_1}, \ldots, A_{i_n} \in \mathcal{E}_{i_n}$

$$\mathbb{P}(X_{i_1} \in A_{i_1}, \ldots, X_{i_n} \in A_{i_n}) = \prod_{j=1}^{n} \mathbb{P}(X_{i_j} \in A_{i_j}). \tag{1.8}$$

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

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

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

(c) *Seien* $\{X_i \mid i \in I\}$ *unabhängig und* $\varphi_i : E_i \longrightarrow W_i$ *seien* $\mathcal{E}_i$*–*$\mathcal{W}_i$*-messbar. Dann ist* $\{\varphi_i \circ X_i \mid i \in I\}$ *unabhängig.*

Lemma 1.54. *Seien* X_i *Zufallsvariablen mit Werten in* $\mathbb{R}$*. 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* $t_1, \ldots, t_n \in \mathbb{R}$

$$\mathbb{P}(X_{i_1} \leq t_1, \ldots, X_{i_n} \leq t_n) = \prod_{j=1}^{n} \mathbb{P}(X_{i_j} \leq t_j). \tag{1.9}$$

Beweis. Ist $\{X_i \mid i \in I\}$ unabhängig, so folgt (1.9) direkt aus der Definition. Es gelte also (1.9). Sei $n \in \mathbb{N}$ und $i_1, \ldots, i_n \in I$ paarweise verschieden. Es reicht (1.8) zu zeigen. Diese gilt wegen (1.9) 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. $\square$

Lemma 1.55. *Seien* X, Y *unabhängige* $\mathbb{R}$*-wertige Zufallsvariablen. Sei* $X = X^+ - X^-$ *und* $Y = Y^+ - Y^-$ *die Zerlegung in Positiv- und Negativteil. Dann sind* $X^\pm, Y^\pm$ *unabhängig.*

Beweis. **Behauptung:** $X^\pm$ ist messbar bezüglich $\sigma(X)$. Sei $A \in \mathcal{B}(\mathbb{R})$, dann gilt

$$\begin{aligned}(X^+)^{-1}(A) &= \big\{\omega \in \Omega \mid X^+(\omega) \in A\big\} = \big\{\omega \in \Omega \mid X^+(\omega) \in A \cap [0, \infty)\big\} \\ &= \{\omega \in \Omega \mid X(\omega) \in A \cap [0, \infty)\} \in \sigma(X).\end{aligned}$$

Analog erhalten wir

$$\begin{aligned}(X^-)^{-1}(A) &= \big\{\omega \in \Omega \mid X^-(\omega) \in A \cap [0, \infty)\big\} = \{\omega \in \Omega \mid -X(\omega) \in A \cap [0, \infty)\} \\ &= \{\omega \in \Omega \mid X(\omega) \in -(A \cap [0, \infty))\} \in \sigma(X).\end{aligned}$$

Hierbei haben wir benutzt, dass $\mathcal{B}(\mathbb{R}) = \{-A \mid A \in \mathcal{B}(\mathbb{R})\}$, welches aus der Tatsache folgt, dass $x \longmapsto -x$ eine stetige Abbildung mit stetiger Inverser ist. Dieses zeigt die erste Behauptung. Wegen $\sigma(X^\pm) \subset \sigma(X)$ und $\sigma(Y^\pm) \subset \sigma(Y)$ ist das Lemma bewiesen. $\square$

Satz 1.56. *Seien* $X, Y \in \mathcal{L}^1$ *unabhängige,* $\mathbb{R}$*-wertige Zufallsvariablen. Dann gilt.*

(a) $X \cdot Y \in \mathcal{L}^1$ *und* $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

(b) *Seien* $X, Y \in \mathcal{L}^2$*. Dann ist* $\operatorname{cov}(X, Y) = 0$ *und*

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

Beweis. Behauptung (b) folgt unmittelbar aus (a). Wir zeigen daher (a). Sei $X = X^+ - X^-$ und $Y = Y^+ - Y^-$ die Zerlegung in Positiv- und Negativteil. Dann gilt $X \cdot Y = X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+$. Die Behauptung folgt, falls wir

$$\mathbb{E}(X^\pm \cdot Y^\pm) = \mathbb{E}(X^\pm)\mathbb{E}(Y^\pm)$$

gezeigt haben. Da $X^{\pm}, Y^{\pm}$ nicht-negativ und unabhängig sind, können wir ohne Einschränkung der Allgemeinheit $X, Y \geq 0$ annehmen.
Fall 1: $X = \mathbb{1}_A$ und $Y = \mathbb{1}_B$, dann gilt $0 \leq X \cdot Y \leq 1$, also $X \cdot Y \in \mathcal{L}^1$ und

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

Fall 2: $X = \sum\limits_{n=1}^{N} a_n \mathbb{1}_{A_n}$ und $Y = \sum\limits_{m=1}^{M} b_m \mathbb{1}_{B_m}$ mit $a_n, b_m \geq 0$, $N, M \in \mathbb{N}$ und $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$, $(B_n)_{n \in \mathbb{N}} \subset \mathcal{F}$. Dann gilt

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

und folglich $X \cdot Y \in \mathcal{L}^1$. Unabhängigkeit liefert

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

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

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\mathbb{E}(Y) < \infty$$

und somit $X \cdot Y \in \mathcal{L}^1$.

$\square$

Satz 1.57. *Seien X, Y zwei Zufallsvariablen mit Werten in $\mathbb{R}^d$. Dann sind die folgenden Aussagen äquivalent:*

(i) *X, Y sind unabhängig.*

(ii) *Für alle beschränkten messbaren Funktionen $f, g : \mathbb{R}^d \longrightarrow \mathbb{R}$ gilt*

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

(iii) *Für alle stetigen beschränkten Funktionen $f, g : \mathbb{R}^d \longrightarrow \mathbb{R}$ gilt*

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

Beweis. $(i) \Longrightarrow (ii)$: Da $f(X), g(Y)$ unabhängig sind.
$(ii) \Longrightarrow (i)$: Betrachte $f = \mathbb{1}_A$ sowie $g = \mathbb{1}_B$, dann folgt

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

$(ii) \Longrightarrow (iii)$: Jede stetige Funktion ist messbar.
$(iii) \Longrightarrow (i)$: Es seien $A = [a, b], B = [c, d]$ mit $a < b$, $c < d$. Wähle Folgen f_n, g_n stetiger Funktionen mit $0 \leq f_n, g_n \leq 1$ sowie $f_n \nearrow \mathbb{1}_A$ und $g_n \nearrow \mathbb{1}_B$ für $n \to \infty$. Dann gilt

$$\mathbb{E}(f_n(X)g_n(Y)) = \mathbb{E}(f_n(X))\mathbb{E}(g_n(Y)), \quad n \in \mathbb{N}.$$

Mit dem Satz von Lebesgue folgt mit $n \to \infty$

$$\mathbb{P}(X \in [a, b], Y \in [c, d]) = \mathbb{P}(X \in [a, b])\mathbb{P}(Y \in [c, d]).$$

Da die abgeschlossenen Intervalle einen $\cap$-stabilen Erzeuger von $\mathcal{B}(\mathbb{R})$ bilden, folgt

$$\mathbb{P} \circ (X, Y)^{-1} = \mathbb{P} \circ X^{-1} \otimes \mathbb{P} \circ Y^{-1}.$$

$\square$

1.4 Konvergenzbegriffe

Definition 1.58. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von $\mathbb{R}$-wertigen Zufallsvariablen.*

(a) $(X_n)_{n\in\mathbb{N}}$ konvergiert $\mathbb{P}$-fast sicher gegen eine Zufallsvariable X, falls

$$\mathbb{P}(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

(b) $(X_n)_{n\in\mathbb{N}}$ konvergiert im p-ten Mittel bzw. in $\mathcal{L}^p$ gegen eine Zufallsvariable X, falls $X_n, X \in \mathcal{L}^p$ und

$$\mathbb{E}(|X_n - X|^p) \longrightarrow 0, \quad n \to \infty.$$

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

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0.$$

Satz 1.59. *(Satz von Lebesgue, Satz der dominierten Konvergenz) Sei $(X_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1$ eine Folge von Zufallsvariablen und $X \in \mathcal{L}^1$. Angenommen es gibt $Y \in \mathcal{L}^1$ mit*

$$|X_n| \leq Y, \quad n \in \mathbb{N}, \quad \textit{fast sicher}$$

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

Beweis. Siehe Maßtheorie. □

Satz 1.60. *(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* $X_n \longrightarrow X$ *fast sicher, wo* X *eine Zufallsvariable ist. Dann gilt*

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

Beweis. Siehe Maßtheorie. □

Wendet man diesen Satz für $Y_n := \sum_{k=1}^{n} X_k$ an, so ergibt sich das folgende Korollar.

Korollar 1.61. *Sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge nicht-negativer Zufallsvariablen. Dann gilt*

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n). \tag{1.10}$$

Hierbei ist $\infty = \infty$ *zugelassen. In diesem Fall ist die linke Seite endlich genau dann, wenn die rechte Seite endlich ist. Ist* $(X_n)_{n\in\mathbb{N}}$ *eine Folge* $\mathbb{R}$*-wertiger Zufallsvariablen mit* $\sum_{n=1}^{\infty} \mathbb{E}(|X_n|) < \infty$*, so gilt* (1.10) *immernoch und beide Seiten sind endlich.*

Lemma 1.62. *(Lemma von Fatou) Sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge nicht-negativer Zufallsvariablen. Dann gilt*

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

Beweis. Siehe Maßtheorie. □

Satz 1.63. *(a) Angenommen* $X_n \longrightarrow X$ *fast sicher, dann gilt* $X_n \longrightarrow X$ *in Wahrscheinlichkeit.*

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

Beweis. (a) Sei $Y_n = \mathbb{1}_{|X_n - X| \geq \varepsilon}$, $\varepsilon > 0$ fest. Dann ist $Y_n \longrightarrow 0$ fast sicher und $Y_n \leq 1$. Mit 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 für alle $\varepsilon > 0$

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

□

Die so definierten Konvergenzbegriffe wurden ohne Rückgriff auf eine Metrik definiert. Daher ist ohne weiteres nicht klar, ob die so definierten Grenzwerte bezüglich einer der obigen Konvergenzarten eindeutig sind. Dieses soll im nächsten Lemma bewiesen werden.

Lemma 1.64.

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

(b) Angenommen $X_n \longrightarrow X$ fast sicher und $X_n \longrightarrow Y$ fast sicher. Dann gilt $X = Y$ fast sicher.

(c) Angenommen $X_n \longrightarrow X$ in $\mathcal{L}^p$ und $X_n \longrightarrow Y$ in $\mathcal{L}^p$. Dann gilt $X = Y$ fast sicher.

Beweis. Es reicht Behauptung (a) zu zeigen. Behauptungen (b) und (c) sind eine direkte Folgerung aus (a). Für alle $\varepsilon > 0$ gilt

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

□

Beispiel 1.65. *Alle anderen denkbaren Pfeile gelten nicht.*
Sei hierzu $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \mathrm{d}x)$.

(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. Wegen

$$\mathbb{P}(|X_n| \geq \varepsilon) \leq \mathbb{P}(|X_n| > 0) \leq \frac{1}{n}$$

folgt $X_n \longrightarrow 0$ in Wahrscheinlichkeit. Aber $\mathbb{E}(|X_n|^p) = \frac{1}{n}(n^{\frac{1}{p}})^p = 1$ zeigt, dass X_n nicht gegen $X = 0$ in $\mathcal{L}^p$ konvergiert. Nach Lemma 1.64 kann somit die Folge $(X_n)_{n\in\mathbb{N}}$ nicht in $\mathcal{L}^p$ konvergieren.

(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 $X_n \longrightarrow 0$ in Wahrscheinlichkeit und in $\mathcal{L}^p$. Wir zeigen, dass $X_n(\omega)$ für kein $\omega \in [0,1)$ konvergiert. Sei $\omega \in [0,1)$ fest. 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 $X_{k(m,\omega)+2^m}(\omega) = 1$. 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.

Satz 1.66. *(Satz von Lebesgue für $\mathcal{L}^p$) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit $X_n \longrightarrow X$ fast sicher und es gelte $|X_n| \le Y$ für eine Zufallsvariable $Y \in \mathcal{L}^p$. Dann gilt $X_n \longrightarrow X$ in $\mathcal{L}^p$.*

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

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

Somit gilt für $n \ge n_0$ und alle $\omega \in M$ mit $(a+b)^p \le 2^{p-1}(a^p + b^p)$

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

Die Behauptung folgt aus dem Satz von Lebesgue. □

Lemma 1.67. *(1. Borel-Cantelli Lemma) Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge von Ergeignissen $A_n \in \mathcal{F}$ mit $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Dann gilt*

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

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

Beweis. Seien $B_n := \bigcup_{k\ge n} A_k$, dann ist $B_n \searrow \limsup_{n\to\infty} A_n$. Es folgt

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

□

Satz 1.68. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen 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

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

Ohne Einschränkung der Allgemeinheit sei $n_k < n_{k+1}$. Setze $M := \limsup_{k\to\infty} A_k$ und $A_k = \left\{|X_{n_k} - X| \ge \frac{1}{k}\right\}$. Nach Borel-Cantelli folgt $\mathbb{P}(M) = 0$. Für $\omega \in M^c$ gilt $\omega \notin \bigcap_{n\ge 1} \bigcup_{k\ge n} A_k$, d.h. $\omega \in \bigcup_{n\ge 1} \bigcap_{k\ge n} A_k^c$. Für jedes $\omega \in M^c$ gibt es also $n \ge 1$ sodass für alle $k \ge n$, $\omega \in A_k^c = \left\{|X_{n_k} - X| < \frac{1}{k}\right\}$. Wir erhalten somit

$$\lim_{k\to\infty} |X_{n_k}(\omega) - X(\omega)| = 0, \quad \omega \in M^c$$

und wegen $\mathbb{P}(M^c) = 1$ die Behauptung. □

Wir wollen die Vollständigkeit der Konvergenzbegriffe beweisen.

Satz 1.69. *(Vollständigkeit bei fast sicherer Konvergenz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen auf $\mathbb{R}$ mit*

$$|X_n - X_m| \longrightarrow 0, \quad n, m \to \infty \quad \textit{fast sicher.} \tag{1.11}$$

Dann gibt es eine Zufallsvariable X auf $\mathbb{R}$ mit $X_n \longrightarrow X$ fast sicher.

Beweis. Sei $M \in \mathcal{F}$ mit $\mathbb{P}(M) = 1$ derart, dass (1.11) für jedes $\omega \in M$ gilt. Da $\mathbb{R}$ vollständig ist gibt es für jedes $\omega \in M$ einen Grenzwert $X(\omega) := \lim\limits_{n\to\infty} X_n(\omega)$. 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. Es sei

$$X|_M : M \longrightarrow \mathbb{R}, \quad \omega \longmapsto X|_M(\omega) := \lim_{n\to\infty} X_n(\omega).$$

Dann ist $X|_M$ als Grenzwert messbarer Funktionen messbar bezüglich

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

Wegen $M \in \mathcal{F}$ folgt $\mathcal{F} \cap M \subset \mathcal{F}$, also ist $X|_M$ messbar bezüglich $\mathcal{F}$. Sei $a \in \mathbb{R}$, es reicht zu zeigen dass $\{X \leq a\} \in \mathcal{F}$.
Fall 1: Ist $a < 0$, so gilt $\{X \leq a\} = \{X|_M \leq a\} \in \mathcal{F}$.
Fall 2: Ist $a \geq 0$, so schreiben wir

$$\{X \leq a\} = \{X < 0\} \cup \{X = 0\} \cup \{0 < X \leq a\} = \{X|_M < 0\} \cup \{X = 0\} \cup \{0 < X|_M \leq a\}.$$

Die erste und letzte Menge liegt in $\mathcal{F}$. Ferner gilt

$$\{X = 0\} = \{X|_M = 0\} \cup M^c \in \mathcal{F}.$$

□

Satz 1.70. *(Vollständigkeit 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

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

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

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

Damit ist $(X_{n_k}(\omega))_{k\in\mathbb{N}}$ eine Cauchy Folge in $\mathbb{R}$ und konvergiert folglich. Setze $X(\omega) := \lim_{k\to\infty} X_{n_k}(\omega)$. Ferner sei $X|_M := 0$. Dann ist X eine Zufallsvariable. Da $X_{n_k} \longrightarrow X$ fast sicher, gilt die Konvergenz auch in Wahrscheinlichkeit. Es bleibt zu zeigen, dass die ganze Folge $X_n \longrightarrow X$ fast sicher konvergiert. Sei $n \in \mathbb{N}$ und $n_k \geq n$, dann ist

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

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

Für den Fall $\mathcal{L}^p$ betrachten wir $p \in [1, \infty)$. Für $X \in \mathcal{L}^p$ sei

$$\|X\|_{\mathcal{L}^p} := (\mathbb{E}(|X|^p))^{\frac{1}{p}} = \left(\int_\Omega |X|^p \mathrm{d}\mathbb{P}\right)^{\frac{1}{p}}. \tag{1.12}$$

Angenommen es ist $\|X\|_{\mathcal{L}^p} = 0$. Mit der Markov Ungleichung folgt für jedes $\varepsilon > 0$

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}(|X|^p) = 0$$

und folglich $\mathbb{P}(|X| > 0) = 0$. Also ist $\mathbb{P}(X = 0) = 1$, d.h. $X = 0$ fast sicher. Es sei

$$\mathcal{N} := \{X \in \mathcal{L}^p \,||\mathbb{P}(X = 0) = 1\} = \{X \in \mathcal{L}^p \mid X = 0 \text{ fast sicher }\}$$

der Raum der Zufallsvariablen, welche fast sicher verschwinden. Der Quotientenraum

$$L^p := \mathcal{L}^p/\mathcal{N}$$

ist der Quotientenraum der Äquivalenzklassen von Zufallsvariablen X, welche $\mathbb{P}$-fast überall übereinstimmen. Für eine solche Äquivalenzklasse von Zufallsvariablen X ist das Lebesgue Integral bezüglich $\mathbb{P}$ wohldefiniert.

Satz 1.71. *Der Raum $(L^p, \|\cdot\|_{\mathcal{L}^p})$ ist ein Banachraum für $1 \leq p < \infty$. Im Fall $p = 2$ ist es ein Hilbertraum mit Skalarprodukt*

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

Beweis. Siehe Maßtheorie oder [Wer00]. □

Bemerkung 1.72. *Wir haben den Fall von $\mathbb{R}$-wertigen Zufallsvariablen betrachtet. Der Fall von $\mathbb{R}^d$-wertigen Zufallsvariablen geht analog. Wir behalten hierfür alle bisherigen Notationen bei.*

2 Schwache Konvergenz

Sei (E, d) ein metrischer Raum und sei $\mathcal{P}(E)$ der Raum der Wahrscheinlichkeitsmaße auf E. Im ersten Teil untersuchen wir die Konvergenz von Wahrscheinlichkeitsmaßen.
Idee: Mengenweise Konvergenz, d.h.

$$\mu_n(A) \longrightarrow \mu(A), \quad A \in \mathcal{E},\ n \to \infty. \tag{2.1}$$

Problem: Für viele praktische Anwendungen ist diese Art der Konvergenz zu stark. Dazu betrachten wir die Folge $\mu_n(\mathrm{d}x) = n\mathbb{1}_{[0,\frac{1}{n}]}(x)\mathrm{d}x$ und $A = (0, 1]$. Dann gilt $\mu_n((0, 1]) = 1$. Es liegt nahe anzunehmen, dass $\mu_n \longrightarrow \delta_0$ konvergiert. Es gilt jedoch $\delta_0((0, 1]) = 0$. Eine mengenweise Konvergenz liegt in diesem Beispiel nicht vor.
Wir brauchen daher eine abgeschwächte Form von (2.1). In vielen Anwendungen treten häufig Integrale der Form $\int_E f(x)\mathrm{d}\mu(x)$ auf, es liegt daher Nahe (2.1) durch

$$\int_E f(x)\mathrm{d}\mu_n(x) \longrightarrow \int_E f(x)\mathrm{d}\mu(x), \quad f \in M(E)$$

zu ersetzen, wo $M(E)$ eine geeignete Familie von Funktionen auf E bezeichnet. Wir betrachten den Fall $M(E) = C_b(E)$, wo $C_b(E)$ den Raum der stetigen beschränkten Funktionen $f : E \longrightarrow \mathbb{R}$ bezeichnet. Dieser ist ein Banachraum mit der Supremumsnorm

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

Der Zusammenhang zu Zufallsvariablen wird ebenfalls hergestellt. Hierbei führen wir einen einen Konvergenzbegriff für Folgen von Zufallsvariablen X_n ein. D.h. X_n konvergiert in Verteilung gegen X, wenn die Folgen der Verteilungen schwach konvergieren.

2.1 Konvergenz von Wahrscheinlichkeitsmaßen

Definition 2.1. *Eine Folge $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ konvergiert schwach gegen $\mu \in \mathcal{P}(E)$, falls*

$$\int_E f(x)\mathrm{d}\mu_n(x) \longrightarrow \int_E f(x)\mathrm{d}\mu(x), \quad f \in C_b(E),\ n \to \infty.$$

Das nächste Lemma zeigt, dass die Grenzwerte eindeutig sind.

Lemma 2.2. *Sei $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$. Angenommen es gibt $\mu, \nu \in \mathcal{P}(E)$ derart dass μ_n schwach gegen μ und schwach gegen ν konvergiert. Dann gilt $\mu = \nu$.*

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

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

Es gilt

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

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

Beweis. Aus der schwachen Konvergenz folgt

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

Es reicht zu zeigen, dass $\mu(K) = \nu(K)$ für alle abgeschlossenen Mengen $K \subset E$ gilt. 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 $n \in \mathbb{N}$. Ferner gilt $f_n(x) \longrightarrow \mathbb{1}_K(x)$ für $n \to \infty$ für jedes $x \in E$. Damit folgt

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

□

Beispiel 2.3. *1. Sei $(x_n)_{n\in\mathbb{N}} \subset E$ und $x \in E$. Falls $d(x_n,x) \longrightarrow 0$ gilt, so konvergiert δ_{x_n} schwach gegen δ_x.*

2. Sei $\mu_n(\mathrm{d}x) = p_n(x)\mathrm{d}x$ mit

$$p_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{n}{2}x^2}.$$

Dann konvergiert μ_n schwach gegen δ_0.

3. Sei $\mu_n(\mathrm{d}x) := n\mathbb{1}_{[0,\frac{1}{n}]}(x)\mathrm{d}x$, dann konvergiert $\mu_n \longrightarrow \delta_0$ schwach.

Satz 2.4. *(Portmanteau-Theorem) Sei (E,d) ein metrischer Raum mit Borel-σ-Algebra $\mathcal{E}$. 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.

(b) Für alle gleichmässig stetigen Funktionen $f \in C_b(E)$ gilt

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

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

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

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

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

(e) Für alle Mengen $A \in \mathcal{E}$ 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 2.5. *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 $x \in \partial\{y \in E \mid g(y) \leq a\}$. 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 $(x_n)_{n\in\mathbb{N}}$ sowie $(y_n)_{n\in\mathbb{N}}$ mit $d(x,x_n) < \frac{1}{n}$ sowie $d(x,y_n) < \frac{1}{n}$ derart, dass $g(x_n) > a$ und $g(y_n) \leq a$. Weiterhin gilt $x_n, y_n \longrightarrow x$ und somit

$$a \leq \lim_{n\to\infty} g(x_n) = g(x) = \lim_{n\to\infty} g(y_n) \leq a.$$

□

Wir kommen zum Beweis von Satz 2.4

Beweis. $(c) \Leftrightarrow (d)$: klar durch Komplementbildung.
$(a) \Rightarrow (b)$: trivial.
$(b) \Rightarrow (c)$: Sei $F \subset E$ abgeschlossen und setze

$$G_m := \left\{x \in E \mid d(x,F) < \frac{1}{m}\right\},$$

wo $d(x,F) := \inf\{d(x,y) \mid y \in F\}$. Dann ist G_m offen, $G_{m+1} \subset G_m$ und $F = \bigcap_{m\geq 1} G_m$. Folglich gilt auch $\mu(G_m) \longrightarrow \mu(F)$. Sei

$$\varphi(t) := \begin{cases} 1-t, & t \in [0,1] \\ 0, & t \geq 1 \end{cases}$$

und setze $f_m(x) := \varphi(md(x,F))$. Dann gilt $0 \leq f_m \leq 1$, $f_m|_{G_m^c} = 0$, $f_m|_F = 1$ und f_m ist Lipschitz stetig. Insbesondere ist f_m gleichmässig stetig und es gilt $\mathbb{1}_F \leq f_m \leq \mathbb{1}_{G_m}$. Daraus folgt

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \int_E f_m \mathrm{d}\mu_n = \int_E f_m \mathrm{d}\mu \leq \mu(G_m) \longrightarrow \mu(F).$$

$(c) \Rightarrow (e)$: Sei $A \in \mathcal{E}$ mit $\mu(\partial A) = 0$. Dann gilt

$$\begin{aligned} \mu(A) = \mu(\mathring{A}) \leq \liminf_{n\to\infty} \mu_n(\mathring{A}) \leq \liminf_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(A) \\ \leq \limsup_{n\to\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}) = \mu(A). \end{aligned}$$

$(e) \Rightarrow (c)$: Sei $F \subset E$ abgeschlossen. Aus Lemma 2.5 folgt für alle $\delta > 0$

$$\partial\{x \in E \mid d(x,F) \leq \delta\} \subset \{x \in E \mid d(x,F) = \delta\}. \tag{2.2}$$

Da $\{x \in E \mid d(x,F) = \delta\}$ für alle $\delta > 0$ eine abgeschlossene Menge ist, liegt sie in $\mathcal{E}$. Weiterhin gilt

$$\{x \in E \mid d(x,F) = \delta\} \cap \{x \in E \mid d(x,F) = \delta'\} = \varnothing, \quad \delta \neq \delta'. \tag{2.3}$$

Für $n \in \mathbb{N}$ sei

$$D_n = \left\{ \delta > 0 \;\middle|\; \mu\left(\left\{x \in E \mid d(x,F) \geq \frac{1}{n}\right\}\right)\right\}.$$

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 (2.3)

$$1 \geq \mu\left(\bigcup_{k\geq 1}\{x \in E \mid d(x,F) = \delta_k\}\right) = \sum_{k=1}^{\infty} \mu\left(\{x \in E \mid d(x,F) = \delta_k\}\right) \geq \sum_{k=1}^{\infty} \frac{1}{n} = \infty.$$

□

Also ist $D = \bigcup\limits_{n\geq 1} D_n$ abzählbar. Folglich gibt es eine Nullfolge $(\delta_k)_{k\in\mathbb{N}} \subset [0,\infty)\backslash D$. Aus (2.2) folgt

$$\mu(\partial F_k) \leq \mu\left(\{x \in E \mid d(x,F) = \delta_k\}\right) = 0$$

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

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F_k) = \mu(F_k) \longrightarrow \mu(F), \quad k \to \infty.$$

$(c) \Rightarrow (a)$: Sei $f \in C_b(E)$. Es reicht zu zeigen, dass

$$\limsup_{n\to\infty} \int\limits_E f(x)\mathrm{d}\mu_n(x) \leq \int\limits_E f(x)\mathrm{d}\mu(x).$$

Denn, daraus folgt für $(-f)$ anstelle von f

$$-\liminf_{n\to\infty} \int\limits_E f(x)\mathrm{d}\mu_n(x) = \limsup_{n\to\infty} \int\limits_E (-f(x))\mathrm{d}\mu_n(x) \leq \int\limits_E (-f(x))\mathrm{d}\mu(x)$$

und somit

$$\liminf_{n\to\infty} \int\limits_E f(x)\mathrm{d}\mu_n(x) \geq \int\limits_E f(x)\mathrm{d}\mu(x).$$

Für $f = 0$ ist nichts zu zeigen. Sei also $f \neq 0$. Wir können ohne Einschränkung der Allgemeinheit annehmen, dass $0 \leq f \leq 1$. Andernfalls betrachten wir $\frac{f+\|f\|_\infty}{2\|f\|_\infty}$. Sei $k \in \mathbb{N}$ fest und für $i \in \mathbb{N}$ definiere

$$F_i := \left\{ f \geq \frac{i}{k} \right\}. \tag{2.4}$$

Dann ist F_i abgeschlossen, $F_{i+1} \subset F_i$ und es gilt

$$F_i \backslash F_{i+1} = \left\{ \frac{i}{k} \leq f < \frac{i+1}{k} \right\}.$$

Lemma 2.6. *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 $E = \bigcup\limits_{i=0}^{k} F_i\backslash F_{i+1}$ mit $F_0 = E$, $F_{k+1} = \varnothing$, wobei die Vereinigung disjunkt ist. Daraus, und aus (2.4) ergibt sich die mittlere Ungleichung. Für die beiden Identitäten beachte

$$\begin{aligned}\sum_{i=0}^{k}(i+1)\mathbb{1}_{F_i\setminus F_{i+1}} &= \sum_{i=0}^{k}(i+1)\mathbb{1}_{F_i} - \sum_{i=0}^{k-1}(i+1)\mathbb{1}_{F_{i+1}}\\ &= \sum_{i=0}^{k}\mathbb{1}_{F_i} + \sum_{i=0}^{k} i\mathbb{1}_{F_i} - \sum_{i=1}^{k} i\mathbb{1}_{F_i} = 1 + \sum_{i=1}^{k}\mathbb{1}_{F_i}.\end{aligned}$$

Dieses impliziert die rechte Identität. Für die linke Identität beachte $1 = \sum_{i=0}^{k} \mathbb{1}_{F_i\setminus F_{i+1}}$ woraus mit obiger Rechnung $\sum_{i=1}^{k} i\mathbb{1}_{F_i\setminus F_{i+1}} = \sum_{i=1}^{k}\mathbb{1}_{F_i}$ folgt. □

Die Behauptung folgt aus

$$\begin{aligned}\limsup_{n\to\infty}\int_E f\mathrm{d}\mu_n - \frac{1}{k} &\leq \frac{1}{k}\limsup_{n\to\infty}\sum_{i=1}^{k}\mu_n(F_i)\\ &\leq \frac{1}{k}\sum_{i=1}^{k}\limsup_{n\to\infty}\mu_n(F_i) \leq \frac{1}{k}\sum_{i=1}^{k}\mu(F_i) \leq \int_E f\mathrm{d}\mu.\end{aligned}$$

□

Satz 2.7. *Seien (E,d) und (E',d') zwei metrische Räume. Seien $\mu_n,\mu \in \mathcal{P}(E)$ und $h : E \longrightarrow E'$ stetig. Angenommen $\mu_n \longrightarrow \mu$ schwach, dann $\mu_n h^{-1} \longrightarrow \mu h^{-1}$ schwach auf E'.*

Beweis. Sei $f \in C_b(E')$. Dann gilt wegen $f \circ h \in C_b(E)$

$$\int_{E'} f(x)\mathrm{d}\mu_n h^{-1}(x) = \int_E f(h(x))\mathrm{d}\mu_n(x) \longrightarrow \int_E f(h(x))\mathrm{d}\mu(x) = \int_{E'} f(x)\mathrm{d}\mu h^{-1}(x).$$

□

Beispiel 2.8. *Auf die Voraussetzung, dass h stetig ist kann im allgemeinen nicht verzichtet werden. Denn sei $(x_n)_{n\in\mathbb{N}} \subset E$ mit $x_n \longrightarrow x$, $x_n \neq x$. Dann gilt $\delta_{x_n} \longrightarrow \delta_x$ schwach. Wähle $h(y) = \mathbb{1}_{\{x\}}(y)$, dann gilt*

$$\delta_{x_n}h^{-1}(A) = \delta_{x_n}(h^{-1}(A)) = \begin{cases}1, & x_n \in h^{-1}(A)\\ 0, & x_n \notin h^{-1}(A)\end{cases} = \begin{cases}1, & h(x_n) \in A\\ 0, & h(x_n) \notin A\end{cases} = \delta_{h(x_n)}(A) = \delta_0(A).$$

Analog zeigen wir $\delta_x h^{-1} = \delta_{h(x)} = \delta_1$.

Wir müssen daher eine Bedingung an die Stetigkeitsstellen von h stellen.

Satz 2.9. *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 $\mu_n h^{-1} \longrightarrow \mu h^{-1}$ schwach.

Beweis. Wegen

$$D_h = \bigcup_{n\geq 1} \bigcap_{m\geq 1} \left\{ x \in E \mid \exists y, z \in E : \; d(x,y) < \frac{1}{m}, \; d(x,z) < \frac{1}{m}, \; d'(h(y), h(z)) \geq \frac{1}{n} \right\}$$

folgt $D_h \in \mathcal{B}(E)$. Sei als $F \subset E'$ 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 $x_n \longrightarrow x$. Falls h stetig ist, so folgt $h(x_n) \longrightarrow h(x)$ und da F abgeschlossen ist auch $h(x) \in F$, d.h. $x \in h^{-1}(F)$. Damit gilt $\overline{h^{-1}(F)} \subset h^{-1}(F) \cup D_h$. Daraus folgt

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

Portmanteau liefert die Behauptung. □

Satz 2.10. *Seien $(\mu_n)_{n\in\mathbb{N}}, \mu \in \mathcal{P}(\mathbb{R})$, $F_n(t) = \mu_n((-\infty, t])$ und $F(t) = \mu((-\infty, t])$. Dann sind äquivalent:*

1. *$\mu_n \longrightarrow \mu$ schwach für $n \to \infty$.*
2. *$F_n(t) \longrightarrow F(t)$ für $n \to \infty$ und alle Stetigkeitspunkte t von F.*

Definition 2.11. *(a) Eine Teilmenge $\Gamma \subset \mathcal{P}(E)$ heißt relativ kompakt, wenn jede Folge in Γ eine schwach konvergente Teilfolge hat.*

(b) $\Gamma \subset \mathcal{P}(E)$ heißt straff, falls für alle $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subset E$ existiert mit

$$\mu(K_\varepsilon) \geq 1 - \varepsilon, \quad \forall \mu \in \Gamma.$$

Bemerkung 2.12. *(a) Ist E kompakt, so ist $\{\mu\}$ für jedes $\mu \in \mathcal{P}(E)$ straff. Denn sei $A_n \subset E$ eine Folge von abgeschlossenen Mengen mit $A_n \nearrow E$. Dann ist jedes A_n als abgeschlossene Teilmenge einer kompakten Menge, selbst wieder kompakt. Die Behauptung folgt aus $\mu(A_n) \nearrow \mu(E) = 1$ für $n \to \infty$.*

(b) Es gebe eine Folge von kompakten Mengen $(K_n)_{n\in\mathbb{N}} \subset E$ mit $\bigcup_{n\geq 1} K_n = E$. Dann ist für jedes $\mu \in \mathcal{P}(E)$, $\mu(K_n) \nearrow 1$ und folglich ist $\{\mu\}$ straff.

(c) $\Gamma = \mathcal{P}(\mathbb{R}^d)$ *ist nicht straff, denn* $\{\delta_x \mid x \in \mathbb{R}^d\}$ *ist nicht straff.*

Definition 2.13. *Ein Polnischer Raum E ist ein topologischer Raum derart dass eine Metrik auf E existiert sodass*

1. *Die Topologie auf E stimmt mit der durch die Metrik induzierte Topologie überein.*
2. *E ist separabel und vollständig bezüglich dieser Metrik.*

Beispiel 2.14.
1. $E = \mathbb{R}^d$ *oder* $E = \mathbb{R}_+$ *sind Polnische Räume.*
2. $E = C_\infty(\mathbb{R}^d)$ *mit Metrik* $d(f,g) := \|f-g\|_\infty$, *wo*

$$C_\infty(\mathbb{R}^d) = \{f \in C_b(\mathbb{R}^d) \mid \forall \varepsilon > 0\ \exists R > 0 \text{ mit } |f(x)| < \varepsilon\ |x| \geq R\}$$

den Raum der im unendlichen verschwindenden Funktionen bezeichnet.

3. $E = L^2(\mathbb{R}^d, m(\mathrm{d}x))$.
4. *Der Raum der endlichen Punktmaße*

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

Behauptung: *Jedes $\eta \in \ddot{\Gamma}_0$ hat eine (eindeutige) Darstellung der Form*

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

wo $N \in \mathbb{N}_0$, $n_1, \dots, n_N \in \mathbb{N}$ und $x_1, \dots, x_N \in \mathbb{R}^d$ paarweise verschieden sind. Die Punkte $x_1, \dots, x_N$ heißen Atome.

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

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

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

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

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 μ ein Maß mit Werten in $\mathbb{N}_0$ und es gilt $\mu(\{x\}) = 0$ für alle $x \in \mathbb{R}^d$. Für jedes $x \in \mathbb{R}^d$ gilt $\mu(B_\varepsilon(x)) \searrow \mu(\{x\}) = 0$ für $\varepsilon \to 0$. Da $\mu(B_\varepsilon(x)) \in \mathbb{N}_0$ gibt es ein $\varepsilon(x) > 0$ derart, dass $\mu(B_{\varepsilon(x)}(x)) = 0$ gilt. Wegen $B_R \nearrow \mathbb{R}^d$ für $R \to \infty$ mit $B_R := \{x \in \mathbb{R}^d \mid |x| \leq R\}$, folgt

$$\mu(B_R) \nearrow \mu(\mathbb{R}^d), \quad R \to \infty$$

und da $\mu(B_R) \in \mathbb{N}_0$ ist, folgt $\mu(B_R) = \mu(\mathbb{R}^d)$ für ein hinreichend grosses R. Wir zeigen

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

woraus die Behauptung folgt. 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 $U_{\varepsilon(x)}(x) = \{y \in \mathbb{R}^d \mid |x-y| < \varepsilon(x)\}$ 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

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

□

5. *Der Raum der endlichen Teilmengen von* $\mathbb{R}^d$

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

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

6. *Der Raum aller lokal endlichen Teilmengen von* $\mathbb{R}^d$

$$\Gamma = \{\gamma \subset \mathbb{R}^d \mid |\gamma \cap \Lambda| < \infty \quad \textit{für alle } \Lambda \subset \mathbb{R}^d \textit{ kompakt } \}.$$

Jedes $\gamma \in \Gamma$ *kann vermöge* $\sum_{x\in\gamma} \delta_x$ *mit einem lokal endlichen Punktmaß auf* $\mathbb{R}^d$ *identifiziert werden. Es sei* $C_c(\mathbb{R}^d)$ *den Raum der stetigen Funktionen* $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ *mit kompaktem Träger bezeichnet. Eine solche Funktion hat kompakten Träger, falls es ein* $R > 0$ *gibt derart, dass* $f(x) = 0$ *für* $|x| \geq 0$*. Für jedes* $f \in C_c(\mathbb{R}^d)$ *sei*

$$I_f(\gamma) := \sum_{x\in\gamma} f(x).$$

Die Topologie auf Γ *ist definiert als die kleinste Topolgie derart, dass alle Abbildungen* I_f *stetig sind. Es lässt sich zeigen, dass* Γ *bezüglich dieser Topolgie ein Polnischer Raum ist, siehe [KK06].*

7. *Der Raum aller lokal endlichen Punktmaße gegeben durch*

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

Behauptung: *Jedes* $\widehat{\gamma} \in \ddot{\Gamma}$ *hat die Darstellung*

$$\widehat{\gamma} = \sum_{x\in\gamma} n_k \delta_x,$$

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

$$\widehat{\gamma}_N(B) := \widehat{\gamma}((B_{N+1} \backslash B_N) \cap B), \quad N \geq 0$$

Dann ist $\widehat{\gamma} = \sum_{N=0}^{\infty} \widehat{\gamma}_N$ und $\widehat{\gamma}_N \in \ddot{\Gamma}_0$ 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

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

$\square$

Satz 2.15. *(Satz von Prokhorov) [KS07] 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 $\Gamma \subset \mathcal{P}(E)$ genau dann straff, wenn es relativ kompakt ist.*

Im Folgenden führen wir einige weitere Resultate ohne Beweis auf.

Definition 2.16. *Definiere eine Abbildung $\rho : \mathcal{P}(E) \times \mathcal{P}(E) \longrightarrow [0, \infty)$ durch*

$$\rho(\mu, \nu) = \inf \{\varepsilon > 0 \mid \mu(F) \leq \nu(F^\varepsilon) + \varepsilon, \quad F \in \mathcal{C}\}.$$

Hierbei ist $\mathcal{C}$ die Menge aller abgeschlossenen Mengen auf E und

$$F^\varepsilon = \{y \in E \mid \inf_{x \in F} d(x, y) < \varepsilon\}.$$

Die Abbildung ρ wird Prokhorov Metrik genannt.

Lemma 2.17. *[EK86] Die Prokhorov Metrik ρ definiert eine Metrik auf E.*

Satz 2.18. *[EK86] Sei E ein separabler 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.*

(b) $\rho(\mu_n, \mu) \longrightarrow 0$.

Satz 2.19. *[EK86] Die folgenden Aussagen gelten.*

1. *Ist E separabel, so ist auch $\mathcal{P}(E)$ separabel.*
2. *Ist E vollständig, so ist auch $\mathcal{P}(E)$ vollständig.*

Insbesondere, ist E ein Polnischer Raum, so ist auch $\mathcal{P}(E)$ ein Polnischer Raum.

Definition 2.20. *Sei E ein metrischer Raum und μ ein Wahrscheinlichkeitsmaß auf $(E, \mathcal{B}(E))$. μ heißt regulär, falls für jedes $A \in \mathcal{B}(E)$*

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

Äquivalent: *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*

$$\mu(U_\varepsilon \backslash K_\varepsilon) < \varepsilon. \tag{2.5}$$

Satz 2.21. *Sei E ein metrischer Raum. Dann gilt für jedes $\mu \in \mathcal{P}(E)$*

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

Beweis. Es sei

$$\mathcal{K} = \{A \in \mathcal{B}(E) \mid \forall \varepsilon > 0\ \exists U_\varepsilon \text{ offen}, \exists K_\varepsilon \text{ abgeschlossen mit } K_\varepsilon \subset A \subset U_\varepsilon \text{ und } (2.5)\}$$

Die Behauptung folgt, sobald wir gezeigt haben, dass $\mathcal{K} = \mathcal{B}(E)$. Hierfür reicht es die folgenden Eigenschaften zu zeigen:

(a) $\mathcal{K}$ enthält alle abgeschlossenen Mengen.

(b) $\mathcal{K}$ ist ein Dynkin System.

(a) 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 $\lim_{n\to\infty} \mu(U_n) = \mu(A)$, d.h. $A \in \mathcal{K}$.

(b) 1. z.Z.: $E \in \mathcal{K}$. Sei $\varepsilon > 0$, $x_0 \in E$ fest und $U_\varepsilon = E$. Die Mengen der Form

$$K_\alpha = \{x \in E \mid d(x, x_0) \leq \alpha\}$$

sind abgeschlossen und es gilt $K_\alpha \nearrow E$ für $\alpha \to \infty$. Folglich gilt $\mu(K_\alpha) \nearrow \mu(E) = 1$. Also gibt es ein $\alpha(\varepsilon) > 0$ mit $\mu(K_{\alpha(\varepsilon)}) > 1 - \varepsilon$. Dann gilt $K_{\alpha(\varepsilon)} \subset A \subset U_\varepsilon$ und (2.5) für diese Mengen.

2. Sei $A \in \mathcal{K}$ und $\varepsilon > 0$. Sei $K_\varepsilon, U_\varepsilon$ die dazugehörigen Mengen. Dann ist $K' := E \backslash K$ offen, $U' := E \backslash U$ abgeschlossen, es gilt $U' \subset E \backslash A \subset K'$ und

$$\mu(K' \backslash U') = \mu(U \backslash K) < \varepsilon.$$

3. Sei $(A_n)_{n\in\mathbb{N}} \subset \mathcal{K}$ eine Folge disjunkter Mengen. Sei $\varepsilon > 0$. Dann gilt

$$0 \leq \sum_{n=1}^{\infty} \mu(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \mu(E) = 1,$$

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_ε^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 $K_\varepsilon \subset \bigcup_{n=1}^{\infty} A_n \subset U_\varepsilon$ 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 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{K}$. $\square$

Korollar 2.22. *Sei $\mu \in \mathcal{P}(E)$ derart, dass $\{\mu\}$ straff ist. Dann ist μ regulär.*

Lemma 2.23. *Sei (E, d) ein vollständiger metrischer Raum und $K \subset E$ eine abgeschlossene Teilmenge. Genau dann ist K 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*

$$K \subset \bigcup_{k=1}^{m} B_\varepsilon(x_k), \tag{2.6}$$

wo $B_\varepsilon(x_k) = \{x \in E \mid d(x, x_k) \leq \varepsilon\}$ der abgeschlossene Ball um x_k mit Radius ε ist.

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

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

eine offene Überdeckung von K, wo $U_\varepsilon(x) = \{y \in E \mid |d(x, y) < \varepsilon\}$. Es gibt daher endlich viele Punkte $x_1, \ldots, x_m \in K$ mit (2.6).
Umgekehrt gelte (2.6) 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 eine Folgen B_m, N_m und $(n_m)_{m \geq 1}$ sodass

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

unendlich viele Elemente enthält und $n_{m+1} \in N_{m+1}$. Dann ist $(x_{n_m})_{m \geq 1}$ 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. $\square$

Satz 2.24. *Sei E ein Polnischer Raum. Dann ist $\{\mu\}$ straff für jedes $\mu \in \mathcal{P}(E)$. Insbesondere 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 $\delta > 0$

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

wo $B_\delta(a_n) = \{x \in E \mid d(x,a) \leq \delta\}$ den abgeschlossenen Ball mit Radius δ um a_n bezeichnet. 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 abgeschlossenen. 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 2.23 kompakt. Die Behauptung folgt aus

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

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. □

2.2 Konvergenz von Zufallsvariablen

Sei (E,d) ein metrischer Raum. Sei $(X_n)_{n\in\mathbb{N}}$ 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 $d_{E\times E}((x,y),(x',y')) := d(x,y)+d(y,y')$ 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. Das folgende Lemma führen wir ohne Beweis.

Lemma 2.25. *Sei E separabel. 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 sei (E,d) stets ein separabler metrischer Raum.

Definition 2.26. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine E-wertige Zufallsvariable.*

(a) $X_n\longrightarrow X$ in 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 für $\mu_n := \mathbb{P}\circ X_n^{-1}$ und $\mu := \mathbb{P}\circ X^{-1}$ bereits $\mu_n\longrightarrow\mu$ schwach konvergiert.

Konvergenz in Verteilung ist, nach 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 2.27. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von E-wertigen Zufallsvariablen und X eine E-wertige Zufallsvariable. Angenommen $X_n\longrightarrow X$ in Verteilung und X ist $\mathbb{P}$-fast sicher konstant, so 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

$$\begin{aligned}\mathbb{P}\circ X^{-1}(A) &= \mathbb{P}(\{X\in A\}) = \mathbb{P}(\{X\in A\}\cap\{X=\alpha\})+\mathbb{P}(\{X\in A\}\cap\{X\neq\alpha\})\\ &= \mathbb{P}(\{\alpha\in A\}\cap\{X=\alpha\})+\mathbb{P}(\{\alpha\in A\}\cap\{X\neq\alpha\})\\ &= \mathbb{P}(\{\alpha\in A\})=\delta_\alpha(A).\end{aligned}$$

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. Analog zu Lemma 2.5 lässt sich zeigen, dass $\partial\{y\in E\mid d(y,\alpha)\geq\varepsilon\}\subset\{y\in E\mid d(y,\alpha)=\varepsilon\}$. Damit ist

$$\mathbb{P}\circ X^{-1}(\partial\{y\in E\mid d(y,\alpha)\geq\varepsilon\})\leq\mathbb{P}\circ X^{-1}(\{y\in E\mid d(y,\alpha)=\varepsilon\})=0,$$

da $\alpha \notin \{y \in E \mid d(y,\alpha) = \varepsilon\}$. Aus Portmanteau folgt

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(d(X_n, X) \geq \varepsilon) &= \limsup_{n\to\infty} \mathbb{P}(d(X_n,\alpha) \geq \varepsilon)\\ &= \limsup_{n\to\infty} \mathbb{P}\circ X_n^{-1}(\{y \in E \mid d(y,\alpha) \geq \varepsilon\})\\ &= \mathbb{P}\circ X^{-1}(\{y \in E \mid d(y,\alpha) \geq \varepsilon\}) = 0.\end{aligned}$$

□

Satz 2.28. *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. Sei $\mu_n := \mathbb{P}\circ X_n^{-1}$ und $\mu := \mathbb{P}\circ X^{-1}$. Dann gilt

$$\begin{aligned}&\left|\int_E f(x)\mathrm{d}\mu_n(x) - \int_E f(x)\mathrm{d}\mu(x)\right| = |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))|\\ &\leq \int_{d(X_n,X)<\delta} |f(X_n) - f(X)|\mathrm{d}\mathbb{P} + \int_{d(X_n,X)\geq\delta} |f(X_n) - f(X)|\mathrm{d}\mathbb{P}\\ &\leq \varepsilon + 2\|f\|_\infty \mathbb{P}(d(X_n,X) \geq \delta).\end{aligned}$$

□

Satz 2.29. *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

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(Y_n \in F) &\leq \limsup_{n\to\infty} \mathbb{P}(d(Y_n,X_n) \geq \varepsilon) + \limsup_{n\to\infty} \mathbb{P}(X_n \in F^\varepsilon)\\ &= \limsup_{n\to\infty} \mathbb{P}\circ X_n^{-1}(F^\varepsilon) \leq \mu(F^\varepsilon).\end{aligned}$$

Wegen $F^\varepsilon \searrow F$ für $\varepsilon \to 0$ folgt die Behauptung. □

Wir zeigen noch einen alternativen Beweis.

Beweis. Wegen Portmanteau reicht es zu zeigen, dass

$$\int_E f(y)(\mathbb{P} \circ Y_n)^{-1}(\mathrm{d}y) \longrightarrow \int_E f(y)\mathrm{d}\mu(y), \quad n \to \infty$$

für jede gleichmässig stetige Funktion $f : E \longrightarrow \mathbb{R}$ gilt. Sei also f gleichmässig stetig und sei $\varepsilon > 0$ fest. 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

$$\left| \int_E f(y)(\mathbb{P} \circ X_n)^{-1}(\mathrm{d}y) - \int_E f(y)\mathrm{d}\mu(y) \right| < \frac{\varepsilon}{3}$$

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

$$\begin{aligned}
&\left| \int_E f(y)(\mathbb{P} \circ Y_n)^{-1}(\mathrm{d}y) - \int_E f(y)\mathrm{d}\mu(y) \right| \\
&\leq |\mathbb{E}(f(Y_n)) - \mathbb{E}(f(X_n))| + \left| \int_E f(y)(\mathbb{P} \circ X_n)^{-1}(\mathrm{d}y) - \int_E f(y)\mathrm{d}\mu(y) \right| \\
&\leq \int_{d(X_n,Y_n) \geq \delta} |f(X_n) - f(Y_n)|\mathrm{d}\mathbb{P} + \int_{d(X_n,Y_n) < \delta} |f(X_n) - f(Y_n)|\mathrm{d}\mathbb{P} + \frac{\varepsilon}{3} \\
&\leq 2\|f\|_\infty \mathbb{P}(d(X_n, Y_n) \geq \delta) + \frac{2}{3}\varepsilon \leq \varepsilon.
\end{aligned}$$

□

3 Charakteristische Funktionen

Im letzten Kapitel haben wir Wahrscheinlichkeitsmaße und die schwache Konvergenz auf metrischen und insbesondere Polnischen Räumen untersucht. Viele Anwendungen werden jedoch für den Fall $E = \mathbb{R}^d$ modelliert oder lassen sich zumindest auf diesen Fall zurückführen. 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) $\widehat{\mu}(\xi)$ mit gewissen Eigenschaften zuzuordnen. Schwache Konvergenz von Wahrscheinlichkeitsmaßen sowie Existenz von Momenten können mithilfe der charakteristischen Funktion beschrieben werden.

3.1 Fourieransformation für Funktionen

Für $\xi, x \in \mathbb{R}^d$ sei $\xi \cdot x := \sum\limits_{k=1}^{d} \xi_k x_k$.

Definition 3.1. *Sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Die Fouriertransformierte von g ist definiert als*

$$\widehat{g}(\xi) := \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} g(x)\mathrm{d}x := \int\limits_{\mathbb{R}^d} \cos(\xi\cdot x) g(x)\mathrm{d}x + i \int\limits_{\mathbb{R}^d} \sin(\xi\cdot x) g(x)\mathrm{d}x, \quad \xi \in \mathbb{R}^d.$$

Die Definition ist wohldefiniert, da für alle $\xi \in \mathbb{R}^d$: $|e^{i\xi\cdot x}| = 1$.

Definition 3.2. *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 $(g(\xi_j - \xi_i))_{i,j=1,\ldots,n}$ positiv (semi-)definit ist.*
Äquivalent: *Für alle $n \in \mathbb{N}$, alle $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$ und $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$ gilt*

$$\sum_{i,j=1}^{n} \lambda_j \overline{\lambda}_i g(\xi_j - \xi_i) \geq 0.$$

Beweis. (der Äquivalenz)
Es sei $n \in \mathbb{N}$, $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$ und $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$. Setze $A := (g(\xi_j - \xi_i))_{i,j=1,\ldots,n}$, 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. $\overline{A} = A^T$ gilt. Es sei g eine positiv (semi-)definite Funktion.

(i) Für $n = 1$ und $\lambda \in \mathbb{C}$ folgt $|\lambda|^2 g(0) \geq 0$, also $g(0) \geq 0$.

(ii) 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 damit $|g(\xi)| \leq g(0)$.

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 kompakten Träger hat.

Lemma 3.3. *[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*

$$\|f_\varepsilon - g\|_{\mathcal{L}^1} := \int_{\mathbb{R}^d} |f_\varepsilon(x) - g(x)| \mathrm{d}x < \varepsilon.$$

Satz 3.4. *Es sei $g \in \mathcal{L}^1(\mathbb{R}^d)$ Es gelten die folgenden Eigenschaften:*

(a) $|\widehat{g}(\xi)| \leq \|g\|_{\mathcal{L}^1}$.

(b) $\widehat{g}(0) = \int_{\mathbb{R}^d} g(x)\mathrm{d}x$.

(c) g ist gleichmässig stetig.

(d) $\widehat{g}$ verschwindet im Unendlichen, d.h. für alle $\varepsilon > 0$ gibt es eine kompakte Menge $K_\varepsilon \subset \mathbb{R}^d$ derart, dass

$$|\widehat{g}(\xi)| < \varepsilon, \quad \xi \notin K_\varepsilon.$$

(e) Ist g fast überall nicht-negativ, so ist $\widehat{g}$ positiv (semi-)definit.

Beweis. (a) klar.

(b) klar.

(c) Seien $\xi_1, \xi_2 \in \mathbb{R}^d$ und $\varepsilon > 0$. Wähle eine kompakte Menge $K_\varepsilon \subset \mathbb{R}^d$ derart, dass

$$\int_{K_\varepsilon^c} |g(x)|\mathrm{d}x < \frac{\varepsilon}{4}.$$

Sei $\delta > 0$ mit $\delta < \frac{\varepsilon}{\sqrt{2}2(1+\sup_{x\in K_\varepsilon}|x|)(1+\|g\|_{\mathcal{L}^1})}$. Dann gilt für alle $|\xi_1 - \xi_2| < \delta$

$$|\widehat{g}(\xi_1) - \widehat{g}(\xi_2)| \leq \int_{\mathbb{R}^d} |e^{i\xi_1\cdot x} - e^{i\xi_2\cdot x}||g(x)|\mathrm{d}x \leq \int_{K_\varepsilon} |e^{i\xi_1\cdot x} - e^{i\xi_2\cdot x}||g(x)|\mathrm{d}x + \frac{\varepsilon}{2}$$

Es gilt mit $z_1 := \xi_1 \cdot x$ und $z_2 := \xi_2 \cdot x$

$$\begin{aligned}|e^{iz_1} - e^{iz_2}| &= \sqrt{|\cos(z_1) - \cos(z_2)|^2 + |\sin(z_1) - \sin(z_2)|^2} \\ &\leq \sqrt{2}|z_1 - z_2| \leq \sqrt{2}|x||\xi_1 - \xi_2|.\end{aligned}$$

Daraus folgt

$$|\widehat{\mu}(\xi_1) - \widehat{\mu}(\xi_2)| \leq \frac{\varepsilon}{2} + \sqrt{2}\|g\|_{\mathcal{L}^1} \sup_{x\in K_\varepsilon} |x| \cdot |\xi_1 - \xi_2| \leq \varepsilon.$$

(d) 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 $g \in C_c^\infty(\mathbb{R}^d)$ verschwindet, erhalten wir

$$\begin{aligned}\widehat{g}(\xi) &= \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} g(x)\mathrm{d}x = \int\limits_{|x|\leq R'} e^{i\xi\cdot x} g(x)\mathrm{d}x \\ &= \frac{1}{i\xi_j} \int\limits_{|x|\leq R'} \frac{\partial e^{i\xi\cdot x}}{\partial x_j} g(x)\mathrm{d}x = -\frac{1}{i\xi_j} \int\limits_{|x|\leq R'} e^{i\xi\cdot x} \frac{\partial g(x)}{\partial x_j}\mathrm{d}x.\end{aligned}$$

Daraus folgt

$$|\widehat{g}(\xi)| \leq \frac{1}{|\xi_j|} \left\| \frac{\partial g(x)}{\partial x_j} \right\|_{\mathcal{L}^1} \leq \frac{\sqrt{d}}{R'} \left\| \frac{\partial g(x)}{\partial x_j} \right\|_{\mathcal{L}^1}.$$

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

$$|\widehat{g}(\xi)| \leq |\widehat{g}(\xi) - \widehat{f_\varepsilon}(\xi)| + |\widehat{f_\varepsilon}(\xi)| \leq \|g - f_\varepsilon\|_{\mathcal{L}^1} + |\widehat{f_\varepsilon}(\xi)|.$$

(e) Es gilt

$$\int\limits_{\mathbb{R}^d} \left| \sum_{j=1}^{\infty} \lambda_j e^{i\xi_j\cdot x} \right|^2 g(x)\mathrm{d}x = \int\limits_{\mathbb{R}^d} \sum_{i,j=1}^{n} \lambda_i \overline{\lambda}_j e^{ix\cdot(\xi_i - \xi_j)} g(x)\mathrm{d}x = \sum_{i,j=1}^{n} \lambda_i \overline{\lambda}_j \widehat{g}(\xi_i - \xi_j).$$

□

Beispiel 3.5. *Es sei $g(x) := e^{-\alpha|x|^2}$ mit $\alpha > 0$. Dann ist*

$$\widehat{g}(\xi) = \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} e^{-\frac{|\xi|^2}{4\alpha}}. \tag{3.1}$$

Beachte, für $\alpha = \frac{1}{2}$ ist $\widehat{g}(\xi) = (2\pi)^{\frac{d}{2}} g(\xi)$.

Beweis. Es gilt

$$\widehat{g}(\xi) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} e^{-\alpha|x|^2}\mathrm{d}x = \prod_{k=1}^{d} \int\limits_{\mathbb{R}} e^{i\xi_k\cdot x_k} e^{-\alpha x_k^2}\mathrm{d}x_k,$$

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

$$\int_{\mathbb{R}} e^{i\xi\cdot x} e^{-\alpha x^2} \mathrm{d}x = e^{-\frac{\xi^2}{4\alpha}} \int_{\mathbb{R}} e^{-\alpha\left(x - \frac{i\xi}{2\alpha}\right)^2} \mathrm{d}x.$$

Für das Integral erhalten wir mittels Cauchys Integralsatz (siehe [FL80]) und der Substitution $y = \sqrt{\alpha}x$

$$\int_{\mathbb{R}} e^{-\alpha\left(x - \frac{i\xi}{2\alpha}\right)^2} \mathrm{d}x = \int_{\mathbb{R}} e^{-\alpha x^2} \mathrm{d}x = \frac{1}{\sqrt{\alpha}} \int_{\mathbb{R}} e^{-x^2} \mathrm{d}x = \sqrt{\frac{\pi}{\alpha}}.$$

□

Lemma 3.6. *Es seien $g, h \in \mathcal{L}^1(\mathbb{R}^d)$. Dann gilt $\widehat{g} \cdot h, g \cdot \widehat{h} \in \mathcal{L}^1(\mathbb{R}^d)$ und*

$$\int_{\mathbb{R}^d} \widehat{g}(\xi) h(\xi) \mathrm{d}\xi = \int_{\mathbb{R}^d} g(\xi) \widehat{h}(\xi) \mathrm{d}\xi.$$

Beweis. Mittels Fubini erhalten wir

$$\int_{\mathbb{R}^d} \widehat{g}(\xi) h(\xi) \mathrm{d}\xi = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{i\xi\cdot x} g(x) h(\xi) \mathrm{d}x \mathrm{d}\xi = \int_{\mathbb{R}^d} g(\xi) \widehat{h}(\xi) \mathrm{d}\xi.$$

Fubini ist anwendbar, da $(x, \xi) \longmapsto g(x)h(\xi)$ integrierbar bezüglich $\mathrm{d}x\mathrm{d}\xi$ ist. □

Lemma 3.7. *Für jedes $g \in C_b(\mathbb{R}^d)$ und jedes $x \in \mathbb{R}^d$ gilt*

$$\int_{\mathbb{R}^d} g(y) \frac{e^{-\frac{|x+y|^2}{4\alpha}}}{(4\pi\alpha)^{\frac{d}{2}}} \mathrm{d}y \longrightarrow g(-x), \quad \alpha \to 0.$$

Beweis. Übung. □

Lemma 3.8. *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. Es reicht zu zeigen, dass

$$\int_{\mathbb{R}^d} \widehat{g}(\xi) e^{i\xi\cdot x} \mathrm{d}\xi = (2\pi)^d g(-x), \quad x \in \mathbb{R}^d.$$

Sei $\alpha > 0$ und setze $\varphi_x(\xi) := e^{ix\cdot\xi}e^{-\alpha|\xi|^2}$. Für diese Funktion gilt nach (3.1)

$$\begin{aligned}\widehat{\varphi_x}(\xi) &= \int_{\mathbb{R}^d} e^{i\xi\cdot y}\varphi_x(y)\mathrm{d}y = \int_{\mathbb{R}^d} e^{i\xi\cdot y}e^{ix\cdot y}e^{-\alpha|y|^2}\mathrm{d}y \\ &= \int_{\mathbb{R}^d} e^{i(\xi+x)\cdot y}e^{-\alpha|y|^2}\mathrm{d}y = \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} e^{-\frac{|\xi+x|^2}{4\alpha}}.\end{aligned}$$

Damit erhalten wir

$$\begin{aligned}\int_{\mathbb{R}^d} \widehat{g}(\xi)e^{i\xi\cdot x}e^{-\alpha|\xi|^2}\mathrm{d}\xi &= \int_{\mathbb{R}^d} g(\xi)\widehat{\varphi_x}(\xi)\mathrm{d}\xi \\ &= \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}}\int_{\mathbb{R}^d} g(\xi)e^{-\frac{|x-\xi|^2}{4\alpha}}\mathrm{d}\xi = (2\pi)^d\int_{\mathbb{R}^d} g(\xi)\frac{e^{-\frac{|x+\xi|^2}{4\alpha}}}{(4\pi\alpha)^{\frac{d}{2}}}\mathrm{d}\xi.\end{aligned}$$

Betrachten wir den Grenzwert $\alpha \to 0$, so folgt die Behauptung aus Lemma 3.7. □

Satz 3.9. *Sei $g \in \mathcal{L}^1(\mathbb{R}^d)$. Definiere eine Abbildung*

$$\check{g}(x) := \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-ix\cdot\xi}g(\xi)\mathrm{d}\xi, \quad x \in \mathbb{R}^d.$$

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 3.8 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 beweisen. □

3.2 Charakteristische Funktion für Maße

Definition 3.10. *Sei μ ein endliches Maß auf $\mathbb{R}^d$. Die charakteristische Funktion von μ ist die Fouriertransformierte von μ gegeben durch*

$$\widehat{\mu}(\xi) := \int_{\mathbb{R}^d} e^{i\xi\cdot x}\mu(\mathrm{d}x), \quad \xi \in \mathbb{R}^d. \tag{3.2}$$

Beispiel 3.11. *(a) Sei $\mu = \sum_{k=1}^{\infty} a_k\delta_{x_k}$ mit $a_k \geq 0$, $x_k \in \mathbb{R}^d$ derart, dass $\sum_{k=1}^{\infty} a_k = 1$. 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. $\mu(\mathrm{d}x) = p(x)\mathrm{d}x$ für eine Wahrscheinlichkeitsdichte $p(x)$, so ist

$$\widehat{\mu}(\xi) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot x} p(x)\mathrm{d}x = \widehat{p}(x).$$

Satz 3.12. *Es sei μ ein endliches Maß auf $\mathbb{R}^d$. Es gelten die folgenden Eigenschaften.*

(a) $|\widehat{\mu}(\xi)| \leq \mu(\mathbb{R}^d)$ für alle $\xi \in \mathbb{R}^d$.

(b) $\widehat{\mu}(0) = 1$.

(c) $\widehat{\mu}$ ist gleichmässig stetig.

(d) $\widehat{\mu}$ ist positiv (semi-)definit.

Beweis. Übung. □

Beispiel 3.13. *(a) Bernoulli Verteilung: $\mu(\mathrm{d}x) = p\delta_1(\mathrm{d}x) + (1-p)\delta_0(\mathrm{d}x)$. Dann ist*

$$\widehat{\mu}(\xi) = pe^{i\xi} + (1-p) = 1 + p(1 - e^{i\xi}), \quad \xi \in \mathbb{R}.$$

(b) Binomialverteilung: $\mu(\mathrm{d}x) = \sum\limits_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k(\mathrm{d}x)$. Dann ist

$$\widehat{\mu}(\xi) = (1 - p + pe^{i\xi})^n, \quad \xi \in \mathbb{R}.$$

(c) Standardnormalverteilung: Im eindimensionalen Fall sei $\mu(\mathrm{d}x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} \mathrm{d}x$. Dann ist

$$\widehat{\mu}(\xi) = e^{-\frac{\xi^2}{2}}, \quad \xi \in \mathbb{R}.$$

Für den d-dimensionalen Fall sei $\mu(\mathrm{d}x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}} \mathrm{d}x$. 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

$$\nu(\mathrm{d}x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}} \mathrm{d}x$$

die Standardnormalverteilung auf dem $\mathbb{R}^d$ ist. Damit folgt

$$\begin{aligned} \widehat{\mu}(\xi) &= \int\limits_{\mathbb{R}^m} e^{i\xi\cdot x} \mu(\mathrm{d}x) = \int\limits_{\mathbb{R}^d} e^{i\xi\cdot T(x)} \nu(\mathrm{d}x) = e^{i\xi\cdot b} \int\limits_{\mathbb{R}^d} e^{i(A^T\xi)\cdot x} \nu(\mathrm{d}x) \\ &= e^{i\xi\cdot b} e^{-\frac{1}{2}|A^T\xi|^2} = e^{i\xi\cdot b} e^{-\frac{1}{2}\langle \xi, (AA^T)\xi \rangle}. \end{aligned}$$

(e) *Uniforme Verteilung: Sei* $\mu(\mathrm{d}x) = \mathbb{1}_{(-a,a)}(x)\frac{1}{2a}\mathrm{d}x$ *mit* $a > 0$. *Dann ist*

$$\widehat{\mu}(\xi) = \begin{cases} \frac{\sin(a\xi)}{a\xi}, & \xi \neq 0 \\ 1, & \textit{sonst} \end{cases}, \quad \xi \in \mathbb{R}.$$

(f) *Cauchy Verteilung: Sei* $\mu(\mathrm{d}x) = \frac{1}{\pi}\frac{c}{c^2+x^2}\mathrm{d}x$ *mit* $c > 0$. *Dann gilt*

$$\widehat{\mu}(\xi) = e^{-c|\xi|}, \quad \xi \in \mathbb{R}.$$

Satz 3.14. *Die charakteristische Funktion charakterisiert das dazugehörige Maß eindeutig. Genauer, sind* μ, ν *zwei endliche Maße* $\mathbb{R}^d$ *und gilt* $\widehat{\mu}(\xi) = \widehat{\nu}(\xi)$ *für alle* $\xi \in \mathbb{R}^d$, *so folgt* $\mu = \nu$.

Beweis. Es gilt $\mu(\mathbb{R}^d) = \widehat{\mu}(0) = \widehat{\nu}(0) = \nu(\mathbb{R}^d)$. Ohne Einschränkung der Allgemeinheit sei $\mu(\mathbb{R}^d), \nu(\mathbb{R}^d) > 0$. Definiere neue Maße $\mu' := \frac{1}{\mu(\mathbb{R}^d)}\mu$ sowie $\nu' := \frac{1}{\nu(\mathbb{R}^d)}\nu$. Dann gilt $\widehat{\mu'}(\xi) = \widehat{\nu'}(\xi)$ für alle $\xi \in \mathbb{R}^d$. Es reicht daher aus, die Behauptung für Wahrscheinlichkeitsmaße zu zeigen.

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 also $K \subset \mathbb{R}^d$ kompakt und $m \in \mathbb{N}$. Definiere

$$f_m(x) = \begin{cases} 1, & x \in K \\ 0, & d(x,K) \geq \frac{1}{m} \\ 1 - md(x,K), & \text{sonst} \end{cases}.$$

Es gilt dann $0 \leq f_m \leq 1$, f_m ist stetig und $f_m \searrow \mathbb{1}_K$ für $m \to \infty$. Wir zeigen

$$\int\limits_{\mathbb{R}^d} f_m(x)\mathrm{d}\mu(x) = \int\limits_{\mathbb{R}^d} f_m(x)\mathrm{d}\mu(x), \quad m \in \mathbb{N}. \tag{3.3}$$

Dann folgt mit dem Satz von Lebesgue $\nu(K) = \mu(K)$. Allgemeiner zeigen wir (3.3) 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 sodass

$$\overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}} \subset [-N,N]^d =: B_N$$

und $\mu(B_N^c) < \varepsilon$, $\nu(B_N^c) < \varepsilon$ gilt. Nach dem Weierstraßschen Approximationssatz (siehe [Wer00] für den Beweis in einer Dimension) gibt es eine Funktion

$$g(x) = \sum_{j=1}^{m} c_j e^{i\frac{\pi}{N}\langle \xi_j, x\rangle}$$

mit $c_j \in \mathbb{C}$, $\xi_j \in \mathbb{Z}^d$ und

$$\sup_{x\in B_N} |f(x) - g(x)| \leq \varepsilon.$$

Es gilt mit $f|_{B_N^c} = 0$ und $\mu(B_N) \leq 1$

$$\begin{aligned} \left| \int\limits_{\mathbb{R}^d} (f-g)\mathrm{d}\mu \right| &\leq \int\limits_{B_N} |f-g|\mathrm{d}\mu + \int\limits_{B_N^c} |f| + |g|\mathrm{d}\mu \\ &\leq \varepsilon\mu(B_N) + \|g\|_\infty \mu(B_N^c) \leq (1 + \|g\|_\infty)\varepsilon. \end{aligned}$$

Analog lässt sich zeigen, dass

$$\left| \int\limits_{\mathbb{R}^d} (f-g)\mathrm{d}\nu \right| \leq (1 + \|g\|_\infty)\varepsilon.$$

Insgesamt erhalten wir

$$\begin{aligned} \left| \int\limits_{\mathbb{R}^d} f\mathrm{d}\mu - \int\limits_{\mathbb{R}^d} f\mathrm{d}\nu \right| &\leq \left| \int\limits_{\mathbb{R}^d} (f-g)\mathrm{d}\mu \right| + \left| \int\limits_{\mathbb{R}^d} g\mathrm{d}\mu - \int\limits_{\mathbb{R}^d} g\mathrm{d}\nu \right| + \left| \int\limits_{\mathbb{R}^d} (g-f)\mathrm{d}\nu \right| \\ &\leq 2\varepsilon + \left| \int\limits_{\mathbb{R}^d} g\mathrm{d}\mu - \int\limits_{\mathbb{R}^d} g\mathrm{d}\nu \right|. \end{aligned}$$

Die Behauptung folgt aus

$$\int\limits_{\mathbb{R}^d} g\mathrm{d}\mu = \sum_{j=1}^{m} c_j \int\limits_{\mathbb{R}^d} e^{i\frac{\pi}{N}\langle \xi_j, x\rangle}\mathrm{d}\mu(x) = \sum_{j=1}^{m} c_j \int\limits_{\mathbb{R}^d} e^{i\frac{\pi}{N}\langle x_j, x\rangle}\mathrm{d}\nu(x) = \int\limits_{\mathbb{R}^d} g\mathrm{d}\nu$$

$\square$

Satz 3.15. *([Jac01, Theorem 3.5.7], Bochner-Theorem) 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 gilt $g(0) = \mu(\mathbb{R}^d)$. Insbesondere ist μ ein Wahrscheinlichkeitsmaß genau dann, wenn $g(0) = 1$.

Es sei μ ein endliches Maß auf $\mathbb{R}^d$ mit $\widehat{\mu} = g$. Dann haben wir bereits gezeigt, dass g die geforderten Eigenschaften hat. Ferner wissen wir, dass μ eindeutig ist. Der Beweis der Umkehrung erfordert einiges mehr an Vorbereitung und soll daher nicht bewiesen werden.

Lemma 3.16. *Für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ gibt es eine Folge $(\varphi_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ mit $\widehat{\varphi_n} \in \mathcal{L}^1(\mathbb{R}^d)$ derart, dass $\varphi_n(x) \longrightarrow \varphi(x)$ für alle $x \in \mathbb{R}^d$.*

Beweis. Definiere
$$\varphi_\alpha(x) := (4\pi\alpha)^{-\frac{d}{2}} \int\limits_{\mathbb{R}^d} \varphi(x-y) e^{-\frac{|y|^2}{4\alpha}} \mathrm{d}y.$$

Dann gilt
$$|\varphi_\alpha(x)| \leq \|\varphi\|_\infty (4\pi\alpha)^{-\frac{d}{2}} \int\limits_{\mathbb{R}^d} e^{-\frac{|y|^2}{4\alpha}} \mathrm{d}y = \|\varphi\|_\infty.$$

Da $(4\pi\alpha)^{-\frac{d}{2}} e^{-\frac{|y|^2}{4\alpha}} \mathrm{d}y$ schwach gegen δ_0 konvergiert, folgt $\varphi_\alpha(x) \longrightarrow \varphi(x)$ für jedes $x \in \mathbb{R}^d$. Es lässt sich zeigen, dass
$$\widehat{\varphi}_\alpha = \widehat{\varphi} \cdot \widehat{p}_\alpha,$$
wo $p_\alpha(x) = (4\pi\alpha)^{-\frac{d}{2}} e^{-\frac{|y|^2}{4\alpha}}$. Da $\widehat{\varphi}$ beschränkt ist und $\widehat{p}_\alpha \in \mathcal{L}^1(\mathbb{R}^d)$ folgt die Behauptung. $\square$

Satz 3.17. *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 $0 \leq p \in C_\infty(\mathbb{R}^d)$ gegeben durch
$$p(x) := \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x} \widehat{\mu}(\xi) \mathrm{d}\xi = \frac{1}{(2\pi)^d} \widehat{\widehat{\mu}}(-x), \quad x \in \mathbb{R}^d. \tag{3.4}$$

(b) Ist $\widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$ derart, dass für ein $m \geq 1$
$$\int\limits_{\mathbb{R}^d} |x|^m \widehat{\mu}(\xi) \mathrm{d}\xi < \infty.$$

Dann hat p stetige Ableitungen bis zur Ordnung m. In diesem Fall gilt für alle $0 \leq k \leq m$ und alle $1 \leq j_1, \ldots, j_k \leq d$
$$\frac{\partial^k p(x)}{\partial x_{j_1} \ldots \partial x_{j_k}} = \frac{(-i)^k}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\xi\cdot x} \xi_{j_1} \cdots \xi_{j_k} \widehat{\mu}(\xi) \mathrm{d}\xi.$$

(c) Ist

$$\int_{\mathbb{R}^d} |x|^m \mu(\mathrm{d}x) < \infty \tag{3.5}$$

für ein $m \in \mathbb{N}$. Dann hat $\widehat{\mu}$ stetige partielle Ableigungen bis zur Ordnung m, und es gilt für alle $0 \le k \le m$ und $1 \le j_1, \ldots, j_k \le d$

$$\frac{\partial^k \widehat{\mu}(\xi)}{\partial \xi_{j_1} \ldots \partial \xi_{j_k}} = i^k \int_{\mathbb{R}^d} x_{j_1} \cdots x_{j_k} e^{i\xi \cdot x} \mu(\mathrm{d}x).$$

Beweis. **(a)** Für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d)$ gilt $\varphi \cdot \widehat{\mu} \in \mathcal{L}^1(\mathbb{R}^d)$ und folglich

$$\int_{\mathbb{R}^d} \varphi(x)\widehat{\mu}(x)\mathrm{d}x = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(x) e^{i\xi\cdot x}\mu(\mathrm{d}\xi)\mathrm{d}x = \int_{\mathbb{R}^d} \widehat{\varphi}(\xi)\mu(\mathrm{d}\xi).$$

Man beachte, dass alle Integrale existieren. Offensichtlich ist $p \in C_\infty(\mathbb{R}^d)$ mit $|p(x)| \le \|\widehat{\mu}\|_{\mathcal{L}^1}$. Wir erhalten für jedes $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ mit $\widehat{\varphi} \in \mathcal{L}^1(\mathbb{R}^d)$

$$\begin{aligned}(2\pi)^d \int_{\mathbb{R}^d} \varphi(x)p(x)\mathrm{d}x &= \int_{\mathbb{R}^d} \varphi(x)\widehat{\widehat{\mu}}(-x)\mathrm{d}x = \int_{\mathbb{R}^d} \varphi(-x)\widehat{\widehat{\mu}}(x)\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \widehat{\varphi(-\cdot)}(x)\widehat{\mu}(x)\mathrm{d}x = \int_{\mathbb{R}^d} \widehat{\widehat{\varphi(-\cdot)}}(x)\mu(\mathrm{d}x) = (2\pi)^d \int_{\mathbb{R}^d} \varphi(x)\mu(\mathrm{d}x).\end{aligned}$$

Aus Lemma 3.16 und dem Satz von Lebesgue folgt für alle $\varphi \in \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \varphi(x)p(x)\mathrm{d}x = \int_{\mathbb{R}^d} \varphi(x)\mu(\mathrm{d}x).$$

Da $C_c(\mathbb{R}^d) \subset \mathcal{L}^1(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ folgt die Behauptung.
(b) Die Behauptung folgt aus der Darstellung (3.4) und Argumenten wie in (c).
(c) Wir zeigen nur den Fall $m = 1$. Der allgemeine Fall folgt dann über Induktion. Es gelte (3.5). Sei e_j der Basisvektor im $\mathbb{R}^d$ in Richtung j. Dann ist für $h > 0$

$$\frac{\widehat{\mu}(\xi + he_j) - \widehat{\mu}(\xi)}{h} = \int_{\mathbb{R}^d} \frac{e^{i(\xi+he_j)\cdot x} - e^{i\xi\cdot x}}{h}\mu(\mathrm{d}x) = \int_{\mathbb{R}^d} e^{i\xi\cdot x}\frac{e^{ix_jh}-1}{h}\mu(\mathrm{d}x).$$

Es gilt $\frac{e^{ix_jh}-1}{h} \longrightarrow ix_j$ und

$$\left|\frac{e^{ix_jh}-1}{h}\right| \le |x_j| + \varepsilon$$

für hinreichend kleine $h, \varepsilon > 0$. Dieses ist wegen (3.5) eine Majorante. Der Satz von Lebesgue impliziert

$$\frac{\partial \widehat{\mu}(\xi)}{\partial \xi_j} = i \int\limits_{\mathbb{R}^d} x_j e^{i\xi \cdot x} \mu(\mathrm{d}x).$$

Die Ableitung ist stetig, ebenfalls nach dem Satz von Lebesgue. □

Bemerkung 3.18. *Die Momente von μ lassen sich daher wie folgt ausdrücken*

$$\frac{\partial^m \widehat{\mu}(0)}{\partial \xi_{j_1} \ldots \partial \xi_{j_m}} = i^m \int\limits_{\mathbb{R}^d} x_{j_1} \cdots x_{j_m} \mu(\mathrm{d}x).$$

Das nachfolgende Kriterium ist nützlich um die schwache Konvergenz zu prüfen.

Lemma 3.19. *Sei E ein metrischer Raum, $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ sowie $\mu \in \mathcal{P}(E)$. Dann sind die folgenden Aussagen äquivalent:*

(i) $\mu_n \longrightarrow \mu$ *schwach.*

(ii) Für jede Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ gibt es eine weitere Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ mit $\mu_{n_{k_l}} \longrightarrow \mu$ schwach.

Beweis. $(i) \Longrightarrow (ii)$: Jede Teilfolge einer konvergenten Folge konvergiert gegen denselben Grenzwert.
$(ii) \Longrightarrow (i)$: Angenommen μ_n konvergiert nicht schwach gegen μ. Dann gibt es ein $\varepsilon > 0$, ein $f \in C_b(E)$ und eine Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ mit

$$\left| \int\limits_E f(x) \mu_{n_k}(\mathrm{d}x) - \int\limits_E f(x)\mu(\mathrm{d}x) \right| \geq \varepsilon, \quad k \in \mathbb{N}. \tag{3.6}$$

Nach (ii) gibt es eine weitere Teilfolge $(\mu_{n_{k_l}})_{l\in\mathbb{N}}$ von $(\mu_{n_k})_{k\in\mathbb{N}}$ sodass $\mu_{n_{k_l}} \longrightarrow \mu$ schwach. Dieses ist ein Wiederspruch zu (3.6). □

Satz 3.20. *(Satz von Cramer-Wald) Seien $\mu_n, \mu \in \mathcal{P}(\mathbb{R}^d)$ und für $x \in \mathbb{R}^d$ setze $g_x(y) := x \cdot y := \sum\limits_{k=1}^{d} x_k y_k$. Dann gilt*

$$\mu_n \longrightarrow \mu \text{ schwach} \iff \mu_n g_x^{-1} \longrightarrow \mu g_x^{-1} \text{ schwach für alle } x \in \mathbb{R}^d.$$

Beweis. $\Rightarrow$: Folgt aus der Stetigkeit von g_x.
$\Leftarrow$: Sei $(\mu_n g_x^{-1})_{n\in\mathbb{N}}$ schwach konvergent gegen μg_x^{-1} für jedes $x \in \mathbb{R}^d$. Dann ist $(\mu_n g_x^{-1})_{n\in\mathbb{N}}$ relativ kompakt und folglich straff für jedes $x \in \mathbb{R}^d$. Seien $e_1, \ldots, e_d$ die Einheitsvektoren

im $\mathbb{R}^d$. Sei $\varepsilon > 0$ beliebig. Dann gibt es für jedes $i = 1, \ldots, d$ eine kompakte Menge $K_i \subset \mathbb{R}$ mit

$$(\mu_n g_{e_i}^{-1})(K_i) \geq 1 - \frac{\varepsilon}{d}, \quad n \in \mathbb{N}.$$

Setze $K := \bigcap_{i=1}^{d} g_{e_i}^{-1}(K_i)$. Diese Menge sind abgeschlossen und beschränkt und somit kompakt. Es gilt

$$\mu_n(K^c) = \mu_n\left(\bigcup_{i=1}^{d} (g_{e_i}^{-1}(K_i))^c\right) \leq \sum_{i=1}^{d} (\mu_n g_{e_i}^{-1})(K_i^c) \leq d\frac{\varepsilon}{d} = \varepsilon$$

und folglich ist $(\mu_n)_{n\in\mathbb{N}}$ straff. Nach dem Satz von Prokhorov ist die Folge relativ kompakt. Sei $(\mu_n')_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$ eine Teilfolge mit $\mu_n' \longrightarrow \mu'$ schwach für ein $\mu' \in \mathcal{P}(\mathbb{R}^d)$. Dann gilt auch $\mu_n' g_x^{-1} \longrightarrow \mu' g_x^{-1}$ schwach für alle $x \in \mathbb{R}^d$. Nach Voraussetzung ist aber auch $\mu_n g_x^{-1} \longrightarrow \mu g_x^{-1}$ schwach für alle $x \in \mathbb{R}^d$. Da $\mu_n' g_x^{-1}$ eine Teilfolge von $\mu_n g_x^{-1}$ ist, konvergiert $\mu_n' g_x^{-1}$ gegen denselben Grenzwert wie $\mu_n g_x^{-1}$, d.h. $\mu g_x^{-1} = \mu' g_x^{-1}$ für alle $x \in \mathbb{R}^d$. Wegen

$$\widehat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{i\xi\cdot x}\mu(\mathrm{d}x) = \int_{\mathbb{R}} e^{it}\mu g_\xi^{-1}(\mathrm{d}t) = \widehat{\mu'}(\xi), \quad \xi \in \mathbb{R}^d$$

folgt $\mu = \mu'$. Die Behauptung folgt aus Lemma 3.19. $\square$

Satz 3.21. *(Levys Stetigkeitssatz) Seien $(\mu_n)_{n\in\mathbb{N}}$ Wahrscheinlichkeitsmaße auf $\mathbb{R}^d$. Dann gelten die folgenden Aussagen.*

(a) Ist $\mu_n \longrightarrow \mu$ schwach, so gilt auch $\widehat{\mu}_n \longrightarrow \widehat{\mu}$ punktweise auf $\mathbb{R}^d$.

(b) Falls es eine in 0 stetige Funktion $f : \mathbb{R}^d \longrightarrow \mathbb{C}$ gibt derart, dass $\widehat{\mu}_n(\xi) \longrightarrow f(\xi)$ für alle $\xi \in \mathbb{R}^d$. Dann gibt es ein Wahrschrscheinlichkeitsmaß μ auf $\mathbb{R}^d$ mit $\widehat{\mu} = f$ und $\mu_n \longrightarrow \mu$ schwach.

Insbesondere gilt für $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$ und $\mu \in \mathcal{P}(\mathbb{R}^d)$

$$\mu_n \longrightarrow \mu \text{ schwach } \iff \widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi), \quad \forall \xi \in \mathbb{R}^d.$$

Beweis. (a) Da $x \longrightarrow e^{i\xi\cdot x}$ stetig und beschränkt ist, folgt die Behauptung.
(b) Wir zeigen nur den Fall $d = 1$.
Schritt 1. $\{\mu_n \mid n \in \mathbb{N}\}$ ist straff.
Für alle $\alpha > 0$ gilt

$$\begin{aligned}\int_{-\alpha}^{\alpha} \widehat{\mu}_n(t)\mathrm{d}t &= \int_{-\alpha}^{\alpha}\int_{\mathbb{R}} e^{itx}\mu_n(\mathrm{d}x)\mathrm{d}t = \int_{\mathbb{R}}\int_{-\alpha}^{\alpha} e^{itx}\mathrm{d}t\mu_n(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \frac{e^{i\alpha x} - e^{-i\alpha x}}{ix}\mu_n(\mathrm{d}x) = \int_{\mathbb{R}} \frac{2\sin(\alpha x)}{x}\mu_n(\mathrm{d}x).\end{aligned}$$

Es folgt

$$\begin{aligned}\int_{-\alpha}^{\alpha}(1-\widehat{\mu}_n(t))\mathrm{d}t &= 2\alpha - \int_{\mathbb{R}} \frac{2\sin(\alpha x)}{x}\mu_n(\mathrm{d}x) \\ &= 2\alpha\left(1-\int_{\mathbb{R}}\frac{\sin(\alpha x)}{\alpha x}\mu_n(\mathrm{d}x)\right) = 2\alpha\int_{\mathbb{R}}\left(1-\frac{\sin(\alpha x)}{\alpha x}\right)\mu_n(\mathrm{d}x).\end{aligned}$$

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

$$\int_{-\alpha}^{\alpha}(1-\widehat{\mu}_n(t))\mathrm{d}t \geq \alpha\mu_n\left(\left[-\frac{2}{\alpha},\frac{2}{\alpha}\right]^c\right).$$

Es ergibt sich mit $\beta = \frac{2}{\alpha}$

$$\mu_n([-\beta,\beta]^c) \leq \frac{1}{\alpha}\int_{-\alpha}^{\alpha}(1-\widehat{\mu}_n(t))\mathrm{d}t = \frac{\beta}{2}\int_{-\frac{2}{\beta}}^{\frac{2}{\beta}}(1-\widehat{\mu}_n(t))\mathrm{d}t.$$

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

$$\lim_{n\to\infty}\frac{\beta}{2}\int_{-\frac{2}{\beta}}^{\frac{2}{\beta}}(1-\widehat{\mu}_n(t))\mathrm{d}t = \frac{\beta}{2}\int_{-\frac{2}{\beta}}^{\frac{2}{\beta}}(1-f(t))\mathrm{d}t \leq \frac{\varepsilon}{4}\frac{\beta}{2}\frac{4}{\beta} = \frac{\varepsilon}{2}$$

und es existiert ein $n_0(\varepsilon) = n_0 \in \mathbb{N}$ mit

$$\left|\frac{\beta}{2}\int_{-\frac{2}{\beta}}^{\frac{2}{\beta}}(1-\widehat{\mu}_n(t))\mathrm{d}t - \frac{\beta}{2}\int_{-\frac{2}{\beta}}^{\frac{2}{\beta}}(1-f(t))\mathrm{d}t\right| \frac{\varepsilon}{2}, \quad n \geq n_0.$$

Damit ist also $\frac{2}{\beta}\int_{-\frac{2}{\beta}}^{\frac{2}{\beta}}(1-\widehat{\mu}_n(t))\mathrm{d}t < \varepsilon$ 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 u nd 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 ein μ' konvergiert. Damit folgt

$$f(\xi) = \lim_{l\to\infty} \widehat{\mu}_{n_{k_l}}(\xi) = \widehat{\mu'}(\xi), \quad \forall \xi \in \mathbb{R}^d.$$

Folglich gibt es für jede Teilfolge eine weitere Teilfolge welche gegen denselben Grenzwert konvergiert. Die Behauptung folgt aus Lemma 3.19. □

3.3 Für Zufallsvariablen

Definition 3.22. *Für eine Zufallsvariable X mit Werten in $\mathbb{R}^d$ ist die charakteristische Funktion definiert als*

$$\varphi_X(\xi) := \mathbb{E}(e^{i\xi\cdot X}) = \widehat{\mathbb{P}X^{-1}}(\xi) = \widehat{\mu_X}(\xi).$$

Korollar 3.23. *Sei X eine $\mathbb{R}^d$-wertige Zufallsvariable. Dann gelten die folgenden Aussagen:*

(a) Ist $\varphi_X \in \mathcal{L}^1(\mathbb{R}^d)$, so hat X eine Dichte $p \in C_\infty(\mathbb{R}^d)$ welche gegeben ist durch

$$p(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\xi\cdot x} \varphi_X(\xi) \mathrm{d}\xi, \quad x \in \mathbb{R}^d.$$

(b) Ist $\varphi_X \in \mathcal{L}^1(\mathbb{R}^d)$ und gilt für ein $m \geq 1$

$$\int_{\mathbb{R}^d} |\xi|^m |\varphi_X(\xi)| \mathrm{d}\xi < \infty,$$

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

$$\frac{\partial^k p(x)}{\partial x_{j_1} \ldots \partial x_{j_k}} = \frac{(-i)^k}{(2\pi)^d} \int_{\mathbb{R}^d} \xi_{j_1} \cdots \xi_{j_k} e^{-i\xi\cdot x} \varphi_X(\xi) \mathrm{d}\xi.$$

(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 $0 \leq k \leq m$, $1 \leq j_1, \dots, j_k \leq d$

$$\frac{\partial^k}{\partial \xi_{j_1} \cdots \partial \xi_{j_k}} \varphi_X(\xi) = i^k \mathbb{E}(X_{j_1} \cdots X_{j_k} e^{i\xi \cdot X}).$$

Insbesondere gilt für die Momente von X die Darstellung

$$\frac{\partial^k}{\partial \xi_{j_1} \cdots \partial \xi_{j_k}} \varphi_X(0) = i^k \mathbb{E}(X_{j_1} \cdots X_{j_k}).$$

Satz 3.24. *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 3.25. *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*

$$\varphi_X(\xi_1)\varphi_Y(\xi_2) = \varphi_{(X,Y)}(\xi_1, \xi_2), \quad \xi_1 \in \mathbb{R}^d, \ \xi_2 \in \mathbb{R}^m. \tag{3.7}$$

Beweis. Seien X, Y unabhängig. Dann ist $\mathbb{P} \circ (X,Y)^{-1} = \mathbb{P} \circ X^{-1} \otimes \mathbb{P} \circ Y^{-1}$. Wir erhalten

$$\begin{aligned}
\varphi_{(X,Y)}(\xi_1, \xi_2) &= \int\limits_{\mathbb{R}^d \times \mathbb{R}^m} e^{i\xi_1 \cdot x} e^{i\xi_2 \cdot y} \mathbb{P} \circ (X,Y)^{-1}(\mathrm{d}x, \mathrm{d}y) \\
&= \int\limits_{\mathbb{R}^d} e^{i\xi_1 \cdot x} \mathbb{P} \circ X^{-1}(\mathrm{d}x) \int\limits_{\mathbb{R}^m} e^{i\xi_2 \cdot y} \mathbb{P} \circ Y^{-1}(\mathrm{d}y) \\
&= \varphi_X(\xi_1)\varphi_Y(\xi_2).
\end{aligned}$$

Umgekehrt gelte (3.7). 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

$$\varphi_{(X',Y')}(\xi_1, \xi_2) = \varphi_{X'}(\xi_1)\varphi_{Y'}(\xi_2) = \varphi_X(\xi_1)\varphi_Y(\xi_2) = \varphi_{(X,Y)}(\xi_1, \xi_2).$$

Damit folgt $\mathbb{P} \circ (X', Y')^{-1} = \mathbb{P} \circ (X,Y)^{-1}$ und somit die Behauptung. □

Korollar 3.26. *Sei $X = (X_1, \dots, X_d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable. Genau dann sind $X_1, \dots, X_d$ unabhängig wenn*

$$\varphi_X(\xi) = \varphi_{X_1}(\xi_1) \cdots \varphi_{X_d}(\xi_d), \quad \forall \xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d.$$

Bemerkung 3.27. *Sei X eine $\mathbb{R}^d$-wertige Zufallsvariable, A eine $m \times d$ Matrix und $b \in \mathbb{R}^m$. Dann gilt hat $Y := AX + b$ die charakteristische Funktion*

$$\varphi_Y(\xi) = e^{i\xi \cdot b}\varphi_X(A\xi), \quad \xi \in \mathbb{R}^m.$$

Beispiel 3.28. *(a) Cauchy-Verteilung. Seien X, Y unabhängig und Cauchy-Verteilt mit Parameter $c > 0$, d.h. diese haben die Verteilung*

$$\mu(\mathrm{d}x) = \frac{1}{\pi}\frac{c}{c^2 + x^2}\mathrm{d}x.$$

Sei $\lambda \in (0,1)$. Dann gilt

$$\varphi_{\lambda X+(1-\lambda)Y}(\xi) = \varphi_X(\lambda\xi)\varphi((1-\lambda)\xi) = e^{-c\lambda|\xi|}e^{-c(1-\lambda)|\xi|} = e^{-c|\xi|}.$$

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

$$\varphi_{X+Y}(\xi) = \varphi_X(\xi)\varphi_Y(\xi) = e^{i\mu_X\xi - \sigma_X^2\frac{\xi^2}{2}}e^{i\mu_Y\xi - \sigma_Y^2\frac{\xi^2}{2}} = e^{i(\mu_X+\mu_Y)\xi - \frac{\xi^2}{2}(\sigma_X^2+\sigma_Y^2)}.$$

Also ist $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Satz 3.29. *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 3.30. *Sei Σ eine symmetrische, 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 derart, dass

$$S\Sigma S^T = D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d).$$

Hierbei bezeichnen $\lambda_1, \ldots, \lambda_d \geq 0$ die Eigenwerte von Σ. Definiere eine neue Matrix

$$D^{\frac{1}{2}} = \mathrm{diag}(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_d^{\frac{1}{2}}).$$

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 3.29)
Eindeutigkeit: Ist μ eine Normalverteilung mit Erwartungswert b und Kovarianzmatrix Σ, so gilt

$$\widehat{\mu}(\xi) = e^{i\langle \xi, b\rangle} e^{-\frac{1}{2}\langle \xi, \Sigma\xi\rangle}, \quad \xi \in \mathbb{R}^d.$$

Die Eindeutigkeit folgt aus der Eindeutigkeit der charakteristischen Funktion.
Existenz: Zu einem gegebenem Σ gibt es immer mindestens eine Matrix A mit $\Sigma = AA^T$. □

Satz 3.31. *Sei $X = (X_1, \dots, X_d)$ eine normalverteilte Zufallsvariable mit Erwartungswert $b \in \mathbb{R}^d$ und Kovarianzmatrix Σ. Dann gelten die folgenden Aussagen.*

(a) Sei $a = (a_1, \dots, a_d) \in \mathbb{R}^d$ 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 $j = 1, \dots, d$ normalverteilt mit Erwartungswert b_j und Varianz Σ_{jj}.

(c) Sind die X_j paarweise unkorreliert, so sind ist $(X_1, \dots, X_d)$ unabhängig.

Umgekehrt, seien $X_1, \dots, X_d$ unabhängige, normalverteilte, $\mathbb{R}$-wertige Zufallsvariablen mit Erwartungswert b_j und Varianz σ_j^2. Dann ist auch $X = (X_1, \dots, X_d)$ normalverteilt mit Erwartungswert $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ und 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

$$\varphi_X(\xi) = \varphi_{X_1}(\xi_1) \cdots \varphi_{X_d}(\xi_d), \quad \forall \xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d.$$

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 $X_1, \dots, X_d$ unabhängig, so folgt

$$\varphi_X(\xi) = \prod_{k=1}^{d} \varphi_{X_k}(\xi_k) = \prod_{k=1}^{d} e^{i\xi_k b_k} e^{-\frac{\sigma_k^2}{2}\xi_k^2} = e^{i\xi\cdot b} e^{-\frac{1}{2}\langle \xi, \Sigma\xi\rangle}.$$

□

4 Appendix

4.1 Topologische Räume

Die folgende Definition ist eine Verallgemeinerung des Begriffs der offenen Mengen für abstrakte Räume Ω. Details und weitere Resultate lassen sich zum Beispiel in [Jän05] nachlesen.

Definition 4.1. *Ein topologischer Raum ist ein paar Paar $(E, \mathcal{T})$, wo E eine nicht-leere Menge ist und $\mathcal{T} \subset 2^E$ ein Mengensystem mit den folgenden Eigenschaften:*

1. *$E, \varnothing \in \mathcal{T}$.*
2. *Seien $A_1, \ldots, A_n \in \mathcal{T}$ mit $n \in \mathbb{N}$, so ist auch $\bigcap\limits_{k=1}^{n} A_k \in \mathcal{T}$.*
3. *Sei I eine beliebige Indexmenge und $(A_i)_{i \in I} \subset \mathcal{T}$, so ist auch $\bigcup\limits_{i \in I} A_i \in \mathcal{T}$.*

Mengen $A \in \mathcal{T}$ heißen offen. Eine Menge $A \subset E$ heißt abgeschlossen, wenn $A^c \in \mathcal{T}$, d.h. A^c ist offen. Das Mengensystem $\mathcal{T}$ heißt Topologie auf E.

Hier und im Folgenden sei $(E, \mathcal{T})$ ein topologischer Raum. Ist die Topologie $\mathcal{T}$ fest gewählt, so schreiben wir kurz E ist ein topologischer Raum.

Definition 4.2. *Sei $(E, \mathcal{T})$ ein topologischer Raum. Eine Menge $A \subset E$ heißt kompakt, falls für jede offene Überdeckung $(A_i)_{i \in I} \subset \mathcal{T}$ von A eine endliche Teilüberdeckung existiert, d.h. ist $A \subset \bigcup\limits_{i \in I} A_i$ so gibt es $i_1, \ldots, i_n \in I$ mit $n \in \mathbb{N}$ derart, dass $A \subset \bigcup\limits_{k=1}^{n} A_k$.*

Beispiel 4.3. *Sei E eine nicht-leere Menge.*

1. *$\mathcal{T} = \{\varnothing, E\}$ ist eine Topologie auf E.*
2. *$\mathcal{T} = 2^E$ ist eine Topologie auf E.*
3. *Sei (E, d) ein metrischer Raum und sei $\mathcal{T}$ das Mengensystem aller offenen Mengen bezüglich der Metrik d. D.h. $A \in \mathcal{T}$, falls für jedes $x \in A$ ein $\varepsilon > 0$ existiert mit*

   $$B_\varepsilon(x) := \{y \in E \mid d(x, y) < \varepsilon\} \subset A.$$

   Dann ist $(E, \mathcal{T})$ ein topologischer Raum.

4.2 Maß und Integrationstheorie

Wir wollen einige Grundlagen der Maß und Integrationstheorie hier aufgreifen. Hier und im Folgenden bezeichnet Ω stets eine nicht-leere Menge. Die Potenzmenge bezeichnen wir mit 2^Ω. Die hier angegebenen Resultate basieren auf der gängigen Literatur zur Maß und Integrationstheorie, siehe zum Beispiel [Kal02, KS07, Sch11].

Maßräume

Definition 4.4. *Eine σ-Algebra $\mathcal{F} \subset 2^{\Omega}$ ist ein Mengensystem mit den folgenden Eigenschaften.*

1. $\Omega \in \mathcal{F}$.

2. $A \in \mathcal{F}$, *so ist auch* $A^c \in \mathcal{F}$.

3. *Ist* $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$, *so ist auch* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Die Potenzmenge 2^{Ω} ist eine σ-Algebra.

Lemma 4.5. *Sei $\mathcal{F} \subset 2^{\Omega}$ ein Mengensystem. Die kleinste σ-Algebra, welche $\mathcal{F}$ enthält ist gegeben durch*

$$\sigma(\mathcal{F}) := \bigcap_{\mathcal{F}\subset\mathcal{A}} \mathcal{A},$$

wo der Schnitt über alle σ-Algebren läuft, welche $\mathcal{F}$ enthalten.

Definition 4.6. *Sei $\mathcal{F}$ eine σ-Algebra. Ein Mengensystem $\mathcal{A} \subset \mathcal{F}$ heißt Erzeugendensystem von $\mathcal{F}$, falls $\sigma(\mathcal{A}) = \mathcal{F}$.*

We sagen, dass Mengen $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ paarweise disjunkt sind, sofern $A_n \cap A_m = \varnothing$ für $n \neq m$ gilt.

Definition 4.7. *Ein Mengensystem $\mathcal{D}$ heißt Dynkin System, falls die folgenden Eigenschaften gelten:*

1. $\Omega \in \mathcal{D}$

2. $A \in \mathcal{D}$, *so gilt auch* $A^c \in \mathcal{D}$.

3. *Ist* $(A_n)_{n\in\mathbb{N}} \subset \mathcal{D}$ *eine Folge paarweise disjunkter Mengen, so gilt* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$.

Ist $\mathcal{A}$ ein Mengensystem, so schreiben wir $d(\mathcal{A})$ für das von $\mathcal{A}$ erzeugte Dynkin System. Es ist das kleinste Dynkin System, welches $\mathcal{A}$ enthält.

Satz 4.8. *Es sei $\mathcal{A}$ ein $\cap$-stabiles Mengensystem, d.h. $A, B \in \mathcal{A}$ so gilt auch $A \cap B \in \mathcal{A}$. Dann gilt $d(\mathcal{A}) = \sigma(A)$.*

Definition 4.9. *Sei Ω ein topologischer Raum. Die Borel-σ-Algebra über Ω ist definiert als die kleinste σ-Algebra, welche das Mengensystem aller offenen Mengen in Ω enthält. Wir bezeichnen diese σ-Algebra mit $\mathcal{B}(\Omega)$.*

Beispiel 4.10. *Ist $\Omega = \mathbb{R}^d$, so ist die Borel-σ-Algebra $\mathcal{B}(\mathbb{R}^d)$ erzeugt durch die Rechtecke der Form*

$$(a_1, b_1) \times \ldots (a_n, b_n), \quad a_j < b_j, \quad j = 1, \ldots, d.$$

Sofern wir mit einem topologischen Raum arbeiten, betrachten wir stets die Borel-σ-Algebra auf diesem. In diesem Fall werden wir nicht explizit auf die Borel-σ-Algebra hinweisen.

Definition 4.11. *Ein Maß ist eine Abbildung $\mu : \mathcal{F} \longrightarrow [0, \infty]$ mit den folgenden Eigenschaften.*

1. $\mu(\varnothing) = 0$.
2. *Ist $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ eine Folge von Mengen mit $A_n \cap A_m = \varnothing$ für alle $n \neq m$, so gilt*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Beispiel 4.12. *Sei $\mathcal{F}$ eine σ-Algebra und $\omega \in \Omega$. Die Abbildung $\delta_\omega : \mathcal{F} \longrightarrow [0, 1]$ gegeben durch*

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

definiert ein Maß auf Ω. Sei $(a_n)_{n \geq 1}$ eine Folge nicht-negativer Zahlen und $(\omega_n)_{n \geq 1} \subset \Omega$. Dann definiert

$$\mu(A) := \sum_{n \geq 1} a_n \delta_{\omega_n}(A)$$

ein Maß auf $(\Omega, \mathcal{F})$.

Definition 4.13. *Ein Maßraum ist $(\Omega, \mathcal{F}, \mu)$, wo $\mathcal{F}$ eine σ-Algebra auf Ω ist und μ ein Maß. Das Tupel $(\Omega, \mathcal{F})$ heißt messbarer Raum.*

Lebesgue-Integral

Definition 4.14. *Seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ zwei messbare Räume und $f : \Omega \longrightarrow \Omega'$. Die Abbildung f heißt messbar bezüglich $\mathcal{F}$-$\mathcal{F}'$, falls gilt*

$$f^{-1}(A) \in \mathcal{F}, \quad A \in \mathcal{F}'.$$

Es ist nicht schwer zu zeigen, dass

$$f^{-1}(\mathcal{F}') = \{f^{-1}(A) \mid A \in \mathcal{F}'\}$$

eine σ-Algebra ist.

Lemma 4.15. *Sei $f : \Omega \longrightarrow \Omega'$ eine Abbildung zwischen den messbaren Räumen $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$. Die folgenden Aussagen gelten.*

1. *f ist genau dann messbar bezüglich $\mathcal{F}$-$\mathcal{F}'$, wenn $f^{-1}(\mathcal{F}') \subset \mathcal{F}$.*

2. *Sei $h : \Omega' \longrightarrow \Omega''$ eine weitere Abbildung. Angenommen f und h sind messbar. Dann ist auch $h \circ f$ messbar.*

Wir lassen den Zusatz $\mathcal{F}$-$\mathcal{F}'$ weg, sofern aus dem Zusammenhang hervorgeht um welche σ-Algebren es sich handelt.

Lemma 4.16. *Es gelten die folgenden Aussagen.*

1. *Sei $f : \Omega \longrightarrow \mathbb{R}$. Dann ist f genau dann messbar, wenn $\{\omega \in \Omega \mid f(\omega) \leq a\} \in \mathcal{F}$ für alle $a \in \mathbb{R}$.*

2. *Messbarkeit ist abgeschlossen unter abzählbaren Operationen, d.h. $\lim\limits_{n\to\infty} f_n$, $\sup\limits_{n\in\mathbb{N}} f_n$, $\inf\limits_{n\to\infty} f_n$ sind messbar, falls alle f_n messbar sind und die Ausdrücke Sinn machen.*

Satz 4.17. *Sei $(\Omega, \mathcal{F})$ ein messbarer Raum und Γ eine Menge nicht-negativer, messbarer Funktionen für die gilt:*

1. *$f, g \in \Gamma$, $a, b \geq 0$, so ist auch $af + bg \in \Gamma$.*

2. *$f_n \in \Gamma$, $n \in \mathbb{N}$, und ist $f_n(\omega) \nearrow f(\omega)$ für alle $\omega \in \Omega$, so ist auch $f \in \Gamma$.*

3. *$\mathbb{1}_A \in \Gamma$ für alle $A \in \mathcal{F}$.*

Dann ist Γ die Menge aller nicht-negativer, messbarer Funktionen.

Wir konstruieren das Lebesgue-Integral. Sei hierfür $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Eine Elementarfunktion $f : \Omega \longrightarrow \mathbb{R}_+$, ist eine Funktion der Form

$$f(\omega) = \sum_{n=1}^{N} a_n \mathbb{1}_{A_n}(\omega) \tag{4.1}$$

mit $a_n \geq 0$ und $A_n \in \mathcal{F}$. Für solche Funktionen definieren wir das Lebesgue-Integral über

$$\int\limits_{\Omega} f(\omega) \mathrm{d}\mu(\omega) := \int\limits_{\Omega} f(\omega)\mu(\mathrm{d}\omega) := \sum_{n=1}^{N} a_n \mu(A_n).$$

Es lässt sich zeigen, dass obige Definition unabhängig von der Darstellung (4.1) ist.

Lemma 4.18. *Sei $f : \Omega \longrightarrow \mathbb{R}_+$, dann ist f genau dann messbar, wenn es eine Folge von nicht-negativen Elementarfunktionen f_n gibt mit $f_n \leq f_{n+1} \leq f$ und $f_n(\omega) \longrightarrow f(\omega)$ für alle $\omega \in \Omega$.*

Sei $f : \Omega \longrightarrow \mathbb{R}_+$ messbar und $(f_n)_{n\in\mathbb{N}}$ eine Folge von positiven Elementarfunktionen wie oben. Das Lebesgue-Integral von f bezüglich μ ist definiert als

$$\int_\Omega f(\omega)\mathrm{d}\mu(\omega) := \lim_{n\to\infty} \int_\Omega f_n(\omega)\mathrm{d}\mu(\omega).$$

Auch diese Definition ist unabhängig von der Wahl der Folge $(f_n)_n$. Ist $f : \Omega \longrightarrow \mathbb{R}$ messbar, so sei $f = f^+ - f^-$ die Zerlegung in Positiv- und Negativteil, wo $f^\pm \geq 0$. Angenommen

$$\int_\Omega f^\pm(\omega)\mathrm{d}\mu(\omega) < \infty, \tag{4.2}$$

dann ist das Integral definiert über

$$\int_\Omega f(\omega)\mathrm{d}\mu(\omega) := \int_\Omega f^+(\omega)\mathrm{d}\mu(\omega) - \int_\Omega f^-(\omega)\mathrm{d}\mu(\omega).$$

Bedingung (4.2) ist äquivalent zu

$$\int_\Omega |f(\omega)|\mathrm{d}\mu(\omega) < \infty.$$

Eine Funktion f mit dieser Eigenschaft heißt integrierbar bezüglich μ. Für $A \in \mathcal{F}$ und eine integrierbare Funktion f definieren wir

$$\int_A f(\omega)\mathrm{d}\mu(\omega) := \int_\Omega \mathbb{1}_A(\omega)f(\omega)\mathrm{d}\mu(\omega).$$

Satz 4.19. *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Es gelten die folgenden Eigenschaften.*

1. *Seien $a, b \in \mathbb{R}$, f, g integrierbar, so gilt*

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

2. *Seien $f \leq g$ integrierbar, so gilt*

$$\int_\Omega f(\omega)\mathrm{d}\mu(\omega) \leq \int_\Omega g(\omega)\mathrm{d}\mu(\omega).$$

3. *Seien $A, B \in \mathcal{F}$ mit $A \cap B = \varnothing$ und f integrierbar. Dann gilt*

$$\int_{A \cup B} f(\omega) \mathrm{d}\mu(\omega) = \int_A f(\omega) \mathrm{d}\mu(\omega) + \int_B f(\omega) \mathrm{d}\mu(\omega).$$

4. *Sei f integrierbar und $N \in \mathcal{F}$ mit $\mu(N) = 0$. Dann gilt*

$$\int_\Omega f(\omega) \mathrm{d}\mu(\omega) = \int_{\Omega \setminus N} f(\omega) \mathrm{d}\mu(\omega).$$

Bemerkung 4.20. *Die erste Behauptung im obigen Satz ist immernoch richtig, sofern wir $a, b \geq 0$ und $f, g \geq 0$ fordern. In diesem Fall können wir auf die Integrierbarkeit von f, g verzichten. Die Integrale können dann jedoch den Wert ∞ annehmen.*

Definition 4.21. *(Produkträume) Sei I eine Indexmenge, $(\Omega_i, \mathcal{F}_i)_{i \in I}$ eine Familie messbarer Räume und $\Omega = \prod_{i \in I} \Omega_i$. Die Produkt-$\sigma$-Algebra $\mathcal{F} := \bigotimes_{i \in I} \mathcal{F}_i$ ist definiert als die kleinste σ-Algebra derart dass alle Projektionen $\pi_i : \Omega \longrightarrow \Omega_i$, $(\omega_i)_{i \in I} \longmapsto \omega_i$ messbar sind, d.h.*

$$\mathcal{F} = \sigma\left(\pi_i \mid i \in I\right) := \sigma\left(\bigcup_{i \in I} \{\pi_i(A) \mid A \in \mathcal{F}_i\}\right).$$

Definition 4.22. *(σ-endliche Maße) Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Dann heißt μ σ-endlich, wenn es eine Folge $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ gibt mit $\bigcup_{n=1}^{\infty} A_n = \Omega$ und $\mu(A_n) < \infty$. In diesem Fall sagen wir, dass $(\Omega, \mathcal{F}, \mu)$ ein σ-endlicher Maßraum ist.*

Satz 4.23. *(Produktmaß) Seien $(\Omega_1, \mathcal{F}_1, \mu_1)$ und $(\Omega_2, \mathcal{F}_2, \mu_2)$ zwei Maßräume. Dann gibt es ein Maß $\mu = \mu_1 \otimes \mu_2$ auf $\mathcal{F}_1 \otimes \mathcal{F}_2$ mit der Eigenschaft*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \quad A_1 \in \mathcal{F}_1, \; A_2 \in \mathcal{F}_2.$$

Sind zusätzlich μ_1, μ_2 σ-endliche Maße, so ist μ eindeutig.

Satz 4.24. *(Satz von Fubini-Tonelli) Seien $(\Omega_1, \mathcal{F}_1, \mu_1)$, $(\Omega_2, \mathcal{F}_2, \mu_2)$ zwei σ-endliche Maßräume und sei $f : \Omega_1 \times \Omega_2 \longrightarrow \mathbb{R}$ eine messbare Funktion. Angenommen einer der drei Integrale*

1. $\int_{\Omega_1} \int_{\Omega_2} |f(\omega_1, \omega_2)| \mathrm{d}\mu_2(\omega_2) \mathrm{d}\mu_1(\omega_1)$

2. $\int_{\Omega_2} \int_{\Omega_1} |f(\omega_1, \omega_2)| \mathrm{d}\mu_1(\omega_1) \mathrm{d}\mu_2(\omega_2)$

3. $\int\limits_{\Omega_1 \times \Omega_2} |f(\omega_1, \omega_2)| \mathrm{d}\mu(\omega_1, \omega_2)$

ist endlich. Dann sind auch alle anderen Integrale endlich und es gilt

$$\int\limits_{\Omega_1} \int\limits_{\Omega_2} f(\omega_1, \omega_2) \mathrm{d}\mu_2(\omega_2) \mathrm{d}\mu_1(\omega_1) = \int\limits_{\Omega_2} \int\limits_{\Omega_1} f(\omega_1, \omega_2) \mathrm{d}\mu_1(\omega_1) \mathrm{d}\mu_2(\omega_2) = \int\limits_{\Omega_1 \times \Omega_2} f(\omega) \mathrm{d}\mu(\omega). \quad (4.3)$$

Bemerkung 4.25. *Für $f \geq 0$ messbar gilt* (4.3) *ohne die Bedingungen 1–3. In diesem Fall kann in* (4.3) *jedoch $\infty = \infty = \infty$ auftreten. Eines der Integrale ist dann endlich genau dann, wenn die anderen Integrale endlich sind.*

Definition 4.26. *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Eine Menge $N \subset \Omega$ heißt Nullmenge, wenn es ein $A \in \mathcal{F}$ gibt mit $N \subset A$ und $\mu(A) = 0$. Eine Eigenschaft gilt fast überall, wenn es eine Nullmenge N gibt derart, dass diese Eigenschaft für alle $x \in N^c$ gilt.*

Lebesgue-Maß

Satz 4.27. *Es gibt genau ein Maß $m(\mathrm{d}x) = \mathrm{d}x$ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ sodass*

$$m([a_1, b_1) \times \ldots [a_d, b_d)) = \prod_{j=1}^{d} (b_j - a_j)$$

für alle $a_j < b_j$ mit $j = 1, \ldots, d$. Ferner, $m(\mathrm{d}x)$ ist translationsinvariant, d.h.

$$m(A) = m(A + b), \quad A \in \mathcal{B}(\mathbb{R}^d), \ b \in \mathbb{R}^d.$$

Sei $\Omega \subset \mathbb{R}^d$ offen und $\Phi : \Omega \longrightarrow \Phi(\Omega)$. Ein C^1-Diffeomorphismus ist eine bijektive, stetig differenzierbare Abbildung derart, dass die Inverse $\Phi^{-1} : \Phi(\Omega) \longrightarrow \Omega$ ebenfalls stetig differenzierbar ist.

Satz 4.28. *(Transformationssatz) Sei $\Omega \subset \mathbb{R}^d$ offen und $\Phi : \Omega \longrightarrow \Phi(\Omega)$ ein C^1-Diffeomorphismus. Dann ist $f : \Phi(\Omega) \longrightarrow \mathbb{R}$ genau dann integrierbar, wenn $\Omega \longmapsto f(\Phi(x)) |\det(D\Phi(x))|$ integrierbar ist. In diesem Fall gilt*

$$\int\limits_{\Phi(\Omega)} f(y) \mathrm{d}y = \int\limits_{\Omega} f(\Phi(x)) |\det(D\Phi(x)) \mathrm{d}x.$$

Hierbei bezeichnet $D\Phi(x)$ die Jacobi Matrix von Φ an der Stelle $x \in \Omega$.

Korollar 4.29. *Für $f = 1$ folgt*

$$m(\Phi(\Omega)) = \int\limits_{\Omega} |\det(D\Phi(x))| \mathrm{d}x.$$

Ist $\Phi(x) = Ax + b$ für A eine $d \times d$ Matrix und $b \in \mathbb{R}^d$, so gilt

$$m(\Phi(\Omega)) = |\det(A)| m(\Omega).$$

Korollar 4.30. *Sei $A \in \mathcal{B}(\mathbb{R}^d)$ ein höchstens $d-1$-dimensionaler Unterraum. Dann ist A eine Nullmenge.*

Bemerkung 4.31. *Die Annahmen an Φ lassen sich wie folgt abschwächen.*

1. *Φ braucht nicht stetig differenzierbar zu sein. Es reicht, wenn Φ lokal Lipschitz-stetig ist. In diesem Fall ist Φ fast überall differenzierbar und $x \longmapsto \det(D\Phi(x))$ ist lokal beschränkt und lokal integrierbar.*

2. *Φ braucht nicht injektiv zu sein. Es genügt, wenn Φ fast überall injektiv ist. In diesem Fall ist Φ fast überall differenzierbar und die Funktionaldeterminante ist lokal beschränkt und lokal integrierbar.*

3. *$\det(D\Phi(x))$ darf auch den Wert Null annehmen. Nach dem Satz von Sard (siehe [Sar42, Hir94]) ist*
$$\{\Phi(x) \mid x \in \mathbb{R}^d \text{ mit } \det(D\Phi(x)) = 0\}$$
eine Nullmenge.

Literatur

[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.

[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.

[Hir94] Morris W. Hirsch. *Differential topology*, volume 33 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1994. Corrected reprint of the 1976 original.

[Jac01] Nils Jacob. *Pseudo differential operators and Markov processes. Vol. I.* Imperial College Press, London, 2001. Fourier analysis and semigroups.

[Jän05] Klaus Jänich. *Topologie.* Springer-Verlag, Berlin, eighth edition, 2005.

[Kal02] Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[KK06] Yuri Kondratiev and Oleksandr 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.

[Sar42] Arthur Sard. The measure of the critical values of differentiable maps. *Bull. Amer. Math. Soc.*, 48:883–890, 1942.

[Sch11] Klaus D. Schmidt. *Maß und Wahrscheinlichkeit.* Graduate Texts in Mathematics. Springer-Verlag, New York, 2011.

[Wer00] Dirk Werner. *Funktionalanalysis.* Springer-Verlag, Berlin, extended edition, 2000.