Wahrscheinlichkeitstheorie

Prof. Dr. Thorsten Schmidt[1]

Albert-Ludwigs-Universität Freiburg

Vorlesung im WS 2019/00

Stand: 22. Januar 2020

[1] www.stochastik.uni-freiburg.de/schmidt

Inhaltsverzeichnis

Vorwort

Dieses Skriptum ist vorläufig ! Es enthält offensichtlich noch viele Fehler – also bitte mit dem entsprechenden Fingerspitzengefühl lesen. Für Rückmeldungen sind wir natürlich äußerst dankbar und würden uns über eine Email sehr freuen.

Freiburg, im Oktober 2019 Thorsten Schmidt

1. Ein kurzer Ausflug in die Maßtheorie

Die Entwicklung eines präzisen Begriffs für Wahrscheinlichkeit hat die Mathematiker sehr lange beschäftigt. Es war die Idee von Andrei N. Kolmogorov (1903–1987) das Hilfsmittel *„Maß“* aus der Analysis hierfür zu verwenden und Zugang zu der mächtigen Maß-Integrationstheorie von Henri L. Lebesgue (1875–1941) zu erlangen.

Aus diesem Grund sind die Begriffe *Maß* und *Integral* für Stochastiker besonders wichtig. Im Unterschied zur Analysis, beziehungsweise zum Lebesgue-Maß auf $\mathbb{R}^n$, sind wir allerdings stets an endlichen Maßen ($\mu(\Omega) = 1$) interessiert, aber mit einem beliebigen Ω. Dieser Abschnitt wiederholt kurz die notwendigen Techniken aus der Analysis III.

1.1 Einführung

Ein zentrales Konzept wird die Erzeugung von σ-Algebren sein, worauf wir einen besonderen Augenmerk richten. Sei Ω eine (beliebige) Menge und $\mathscr{P}(\Omega)$ die Menge aller Teilmengen von Ω.

Definition 1.1. Ein Mengensystem $\mathscr{F} \subset \mathscr{P}(\Omega)$ heißt *σ-Algebra* (auf Ω), falls für alle $A, A_1, A_2, \ldots \in \mathscr{F}$ gilt, dass

(i) $\varnothing \in \mathscr{F}$

(ii) $A^c \in \mathscr{F}$

(iii) $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$

Mit $\mathscr{P}(\Omega)$ bezeichnen wir die Menge aller Teilmengen von Omega. Der Schnitt von (beliebig vielen) σ-Algebren ist wieder eine σ-Algebra ($\rightarrow$ Übung), so dass man für ein $\mathscr{C} \subseteq \mathscr{P}(\Omega)$ die von ***C* erzeugte $\boldsymbol{\sigma}$-Algebra** definiert durch

$$\sigma(\mathscr{C}) = \bigcap \big\{ \mathscr{F} \supseteq \mathscr{C} : \mathscr{F} \text{ ist } \sigma\text{-Algebra} \big\}.$$

Ein *topologischer Raum* $(\Omega, \mathcal{O})$ ist ein Raum Ω mit einer Menge $\mathcal{O} \subset \mathscr{P}(\Omega)$, wobei die Elemente von $\mathcal{O}$ als offen bezeichnet werden, so dass

(i) $\varnothing, \Omega \in \mathscr{O}$

(ii) der endliche Schnitt von offenen Mengen ist wieder offen

(iii) die beliebige Vereinigung von offenen Mengen ist wieder offen.

Definition 1.2. Sei $(\Omega, \mathscr{O})$ ein topologischer Raum. Dann heißt $\mathscr{B}(\Omega) = \sigma(\mathscr{O})$ die *Borel-σ-Algebra* auf $(\Omega, \mathscr{O})$.

Hat $\mathscr{O}$ eine abzählbare Basis[1] B, so ist $\sigma(\mathscr{O}) = \sigma(B)$. $\to$ Übung.

Insbesondere erzeugt $C = \{(-\infty, x] : x \in \mathbb{Q}\}$ die Borel-σ-Algebra $\mathscr{B}(\mathbb{R})$. Das gilt auch für die 2-Punkt-Kompaktifizierung

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}.$$

Oft gibt es geeignete, einfache erzeugende Systeme[2]

Definition 1.3. Ein Mengensystem $\mathscr{D} \subset \mathscr{P}(\Omega)$ heißt *Dynkin-System*, falls für alle $A_1, A_2, A_3, \ldots \in \mathscr{D}$ gilt:

(i) $\Omega \in \mathscr{D}$

(ii) $A_1 \subseteq A_2 \Rightarrow A_2 \setminus A_1 \in \mathscr{D}$

(iii) $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathscr{D}$

Natürlich ist jede σ-Algebra ein Dynkin-System. Ein Mengensystem $\mathscr{C}$ heißt *durchschnittsstabil*, fall mit $A, B \in \mathscr{C}$ auch $A \cap B \in \mathscr{C}$.

Lemma 1.4. *Ist $\mathscr{D}$ ein Dynkin-System und $\mathscr{C} \subseteq \mathscr{D}$ durchschnittsstabil, so gilt*

$$\sigma(\mathscr{C}) \subseteq \mathscr{D}.$$

Insbesondere ist jedes durchschnittsstabile Dynkin-System eine σ-Algebra.

[1] Eine Menge B heißt Basis, falls sich jede offene Menge als Vereinigung beliebig vieler Mengen aus B schreiben lässt.

[2] In der englischen Literatur heißt ein durchschnittliches System π-System; ein Dynkin-System λ-System (s. Billingsley(1995)).

Beweis. Wir definieren

$$\lambda(\mathscr{C}) = \bigcap \{\mathscr{F} \supseteq \mathscr{C} : \mathscr{F} \text{ Dynkin-System}\}$$

Da der Schnitt von Dynkin-Systemen wieder ein Dynkin-System ist, ist $\lambda(\mathscr{C}) \subset \mathscr{D}$. Der Rest des Beweises unterteilt sich in drei Schritte:

1. $\lambda(\mathscr{C})$ ist durchschnittsstabil: Seien $A, B \in \lambda(\mathscr{C})$. Sind $A, B \in \mathscr{C}$, so folgt auch $A \cap B \in \mathscr{C} \subseteq \lambda(\mathscr{C})$. Sei nun lediglich $B \in \mathscr{C}$ und

$$\mathscr{D}_B := \{A \subseteq \Omega : A \cap B \in \lambda(\mathscr{C})\}.$$

Dann ist $\mathscr{D}_B$ ein Dynkin-System, denn für $A_1, A_2, \ldots \in \mathscr{D}_B$ gilt

(i) $\Omega \in \mathscr{D}_B$

(ii) Sind $A_1 \cap B, A_2 \cap B \in \lambda(\mathscr{C})$ mit $A_1 \subseteq A_2$, so folgt

$$(A_2 \cap B) \setminus (A_1 \cap B) = (A_2 \setminus A_1) \cap B \in \lambda(\mathscr{C}).$$

(iii) ebenso $A_i \cap B \subseteq A_{i+1} \cap B \Rightarrow$

$$\Big(\bigcup A_i\Big) \cap B = \bigcup \Big(A_i \cap B\Big) \in \lambda(\mathscr{C}).$$

Da $\mathscr{C} \subset \mathscr{D}_B$ folgt nun $\lambda(\mathscr{C}) \subseteq \mathscr{D}_B$.

Wir erhalten für $A \in \lambda(\mathscr{C}) \subseteq \mathscr{D}_B$, dass

$$A \cap B \in \lambda(\mathscr{C}). \qquad \text{(Def. von } \mathscr{D}_B\text{)}$$

Nun setzen wir für $A \in \lambda(\mathscr{C})$

$$\mathscr{D}_A := \{B \subseteq \Omega : A \cap B \in \lambda(\mathscr{C})\}.$$

Wie vorher zeigt man, dass $\mathscr{D}_A$ ein Dynkin-System ist mit $\mathscr{C} \subset \mathscr{D}_A$. Wieder folgt $\lambda(\mathscr{C}) \subseteq \mathscr{D}_A$, also für $A, B \in \lambda(\mathscr{C})$ gerade

$$A \cap B \in \lambda(\mathscr{C}). \qquad \text{Das war 1)}$$

2. $\lambda(\mathscr{C})$ ist σ-Algebra: Für $A_1, A_2, \ldots \in \lambda(\mathscr{C})$ ist

$$\Big(\bigcup_{i=1}^{n} A_i\Big) = \Big(\bigcap_{i=1}^{n} A_i^c\Big)^c \in \lambda(\mathscr{C}) \qquad (A^c \in \mathscr{D},\ \mathscr{D} \text{ Dynkin})$$

und

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \bigcup_{i=1}^{n} A_i \in \lambda(\mathscr{C}).$$

3. Abschließend erhalten wir $\sigma(\mathscr{C}) \subseteq \sigma(\lambda(\mathscr{C})) = \lambda(\mathscr{C}) \subseteq \mathscr{C}$.

□

Nun kommen wir zu den wichtigen Begriffen **Maß** und **Kapazität**. Die folgenden Begriffe kann man auch für Ringe und Halbringe anstelle von σ-Algebren definieren. Auch kann man signierte Maße mit möglicherweise negativen Werten betrachten. Wir setzen $\bar{\mathbb{R}}_{\geq 0} = \mathbb{R}_{\geq 0} \cup \{\infty\}$.

Definition 1.5. Sei $\mathscr{F}$ eine σ-Algebra und $\mu : \mathscr{F} \to \bar{\mathbb{R}}_{\geq 0}$. Gilt für $(A_i)_{i\geq 1} \subset \mathscr{F}$ paarweise disjunkt, dass

(i) $\mu(\varnothing) = 0$,

(ii) $\mu\Big(\sum_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i)$,

so heißt μ *Maß*. Das Maß μ heißt *endlich*, falls $\mu(\Omega) < \infty$, und es heißt *σ-finit*, falls es $(A_n)_{n\in\mathbb{N}} \subset \mathscr{F}$ gibt, so dass $\sum_{n\in\mathbb{N}} A_n = \Omega$ und $\mu(A_n) < \infty$, für alle $n \in \mathbb{N}$.

Die Eigenschaft (ii) heißt *σ-Additivität*. Interessant ist es, sich Mengenfunktionen anzuschauen, wo in (ii) $\leq$ auftaucht, μ also nur sub-additiv (monoton) ist. Dann heißt μ **Kapazität** oder – was wir später noch kennen lernen – äußeres Maß.

Endliche, additive Mengenfunktionen heißen oft **Inhalt**, σ-additive Mengenfunktionen, die nicht auf σ-Algebren definiert sind, **Prämaß**.

Die Fortsetzung von Maßen ist ein wichtiges Hilfsmittel: Das Lebesgue-Maß etwa definiert man durch

$$\mu([a, b]) := b - a \qquad a, b \in \mathbb{Q}$$

und setzt dann geeignet auf $\sigma(\mathbb{Q})$ fort. Wie das technisch funktioniert, soll nun erläutert werden.

Definition 1.6. Ein Mengensystem $\{\varnothing\} \neq \mathscr{H} \subseteq \mathscr{P}(\Omega)$ heißt *Halbring*, falls

(i) $\varnothing \in \mathscr{H}$

(ii) $A, B \in \mathscr{H} \Rightarrow A \cap B \in \mathscr{H}$ (durchschnittsstabil)

(iii) Für $A, B \in \mathscr{H}$ gibt es p. d. Mengen $C_1, \ldots, C_n \in \mathscr{H}$, s.d.

$$B \setminus A = \sum_{i=1}^{n} C_i.$$

Definition 1.7. Ein nicht-leeres Mengensystem $\mathscr{R} \subseteq \mathscr{P}(\Omega)$ heißt *Ring*, falls

(i) $\varnothing \in \mathscr{R}$

(ii) $A, B \in \mathscr{R} \Rightarrow A \cup B \in \mathscr{R}$

(iii) $A, B \in \mathscr{R} \Rightarrow B \setminus A \in \mathscr{R}$.

Ein Ring $\mathscr{R}$ heißt Algebra, falls $\Omega \in \mathscr{R}$. Eine σ-Algebra erfüllt zusätzlich die σ-Additivität.

Beispiel 1.8. $\mathscr{H} = \{(a,b], a, b \in \mathbb{Q}, a \leq b\}$ ist Halbring und $\sigma(\mathscr{H}) = \mathscr{B}(\mathbb{R})$.

Wir nennen eine Funktion $\mu : \mathscr{H} \to \bar{\mathbb{R}}_{\geq 0}$ *additiv*, falls für alle paarweise disjunkten $A, B \in \mathscr{H}$ mit $A \cup B \in \mathscr{H}$ gilt, dass

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Lemma 1.9. *Sei $\mathscr{H}$ ein Halbring mit $\mu : \mathscr{H} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ additiv. Dann gilt:*

(i) μ *ist monoton und subadditiv*

(ii) μ *ist* σ*-additiv* $\Rightarrow$ μ *ist* σ*-subadditiv*

Wir beginnen mit einer Aussage, die wir noch öfter brauchen: Sei $\mathscr{H}$ ein Halbring und $A_1, \ldots, A_n \in \mathscr{H}$ Dann ist

$$\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} B_i = \sum_{i=1}^{n} \sum_{j=1}^{k_i} D_{i,j} \tag{1.10}$$

mit disjunkten Mengen $B_1, \ldots, B_n \in \mathscr{H}$ bzw. $(D_{i,j}) \in \mathscr{H}$. Den Beweis hierfür führen wir mit Induktion: Für $n = 2$ folgt das aus der Halbring-Eigenschaft (ii). Gilt die Behauptung für ein n, so ist

$$B_{n+1} = A_{n+1} \setminus \bigcup_{i=1}^{n} A_i = \Big(\underbrace{A_{n+1} \setminus \bigcup_{i=2}^{n} A_i}_{=\sum_{i=1}^{k} D_i} \Big) \setminus A_1 = \sum_{i=1}^{k} D_i \setminus A_1.$$

Da $\mathscr{H}$ ein Halbring ist gibt es $(E_{ij})_{i=1,\ldots,k}$ so dass

$$D_i \setminus A_1 = \sum_{j=1}^{k_i} E_{ij}, \text{ also ist}$$

$$B_{n+1} = \sum_{i=1}^{k} \sum_{j=1}^{k_i} E_{ij}.$$

Beweis. Wir zeigen zunächst, dass μ monoton ist. Seien $A, B \in \mathscr{H}$, $A \subseteq B \Rightarrow$

$$B = A + B \setminus A = A + \sum_{i=1}^{k} D_i,$$

mit $D_i \in \mathscr{H}$, also ist $\mu(B) = \mu(A) + \sum_{i=1}^{k} \mu(D_i) \geq \mu(A)$.

Nun zeigen wir, dass μ auch subadditiv ist: Seien $A_1, \dots, A_n, \bigcup_{i=1}^{n} A_i \in \mathscr{H}$. Dann ist

$$\begin{aligned}\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \mu\Big(\sum_{i=1}^{n} B_i\Big) = \mu\Big(\sum_{i=1}^{n}\sum_{j=1}^{k_i} D_{ij}\Big) &= \sum_{i=1}^{n}\sum_{j=1}^{k_i} \mu(D_{ij}) \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{k_i} \mu(D_{ij}) + \sum \dots = \sum_{i=1}^{n} \mu(A_i) \qquad (1.11)\end{aligned}$$

Als Letztes fehlt noch μ σ-additiv $\Rightarrow$ μ σ-subadditiv. Dies folgt direkt wie (1.11). $\square$

Beispiel 1.12. Die Rückrichtung gilt **nicht**: Man wähle zum Beispiel den Halbring $\mathscr{H} = \{n : n \in \mathbb{N}\} \cup \mathbb{N}$. Für zwei unterschiedliche Mengen A und B ist dann $A \cap B$ entweder leer oder enthält nur ein Element. Wir wählen ein σ-subadditives Maß durch

$$\mu(A) := \begin{cases} 1 & A = \mathbb{N} \\ n^{-2} & A = \{n\}. \end{cases}$$

Allerdings ist dann

$$P\Big(\sum_{n \in \mathbb{N}} \{n\}\Big) = \sum \frac{1}{n^2} \neq 1 = P(\mathbb{N})$$

und μ ist demnach nicht σ-additiv.

Ist $\mathscr{H}$ ein Halbring, so nennen wir

$$\mathscr{R}(\mathscr{H}) = \Big\{ \sum_{i=1}^{n} A_i : A_1, \dots, A_n \in \mathscr{H} \text{ p. d.}, n \in \mathbb{N} \Big\}$$

den von $\mathscr{H}$ **erzeugten Ring**.

Lemma 1.13. *Sei $\mathscr{H}$ ein Halbring mit $\mu : \mathscr{H} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ endlich und additiv. Auf $\mathscr{R}(\mathscr{H})$ definieren wir*

$$\widetilde{\mu}\Big(\sum_{i=1}^{n} A_i\Big) := \sum_{i=1}^{n} \mu(A_i)$$

für $A_1, \dots, A_n \in \mathscr{H}$ p. d. Dann ist $\widetilde{\mu}$ die einzige additive Fortsetzung *auf $\mathscr{R}$ die auf $\mathscr{H}$ mit μ übereinstimmt. Außerdem gilt: $\widetilde{\mu}$ σ-additiv $\Leftrightarrow$ μ σ-additiv.*

Beweis. Es ist nur zu zeigen, dass $\widetilde{\mu}$ wohldefiniert ist. Seien dazu $A_1, \dots, A_m, B_1, \dots, B_n$ zwei paarweise disjunkte Darstellung der gleichen Menge, also $\sum_{i=1}^n A_i = \sum_{j=1}^n B_i$. Dann folgt

$$\begin{aligned} A_\mathrm{i} &= \sum_{\mathrm{j}=1}^n A_\mathrm{i} \cap B_\mathrm{j}, \\ B_\mathrm{j} &= \sum_{\mathrm{i}=1}^m B_\mathrm{j} \cap A_\mathrm{i}, \\ \sum_{i=1}^m \mu(A_i) &= \sum_{i=1}^m \sum_{j=1}^n \mu(A_i \cap B_j) = \sum_{j=1}^n \mu(B_j) \end{aligned}$$

und somit die Behauptung. □

Sei $\mathscr{M} \subseteq \mathscr{P}(\Omega)$. Dann heißt $\mu : \mathscr{M} \to \bar{\mathbb{R}}_{\geq 0}$

- **stetig von unten**, falls $\mu\big(\bigcup_{i=1}^\infty A_i\big) = \lim_{i\to\infty} \mu(A_i)$ für $A_1 \subseteq A_2 \subseteq \dots$, $A_i \in \mathscr{M}$ für alle $i \geq 1$ und $\bigcup_{i=1}^\infty A_i \in \mathscr{M}$

- mit $\mu(A_1) < \infty$ **stetig von oben**, falls $\mu(\bigcap_{i=1}^\infty A_i) = \lim \mu(A_i)$ für $A_1 \supseteq A_2 \supseteq \dots$, mit $A_i \in \mathscr{M}$

- mit $\mu(A_1) < \infty$ **stetig von oben in** $\varnothing$, falls die obige Eigenschaft $\forall \bigcap_{i=1}^\infty A_1 = \varnothing$ gilt.

Satz 1.14. *Sei $\mathscr{R}$ ein Ring und $\mu : \mathscr{R} \to \bar{\mathbb{R}}_{\geq 0}$ additiv und $\mu(\varnothing) = 0$,*

(i) μ *ist* σ*-additiv*

(ii) μ *ist* σ*-subadditiv*

(iii) μ *ist stetig von unten*

(iv) μ *ist stetig von oben in* $\varnothing$

(v) μ *ist stetig von oben*

Dann gilt: (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) $\Rightarrow$ (iv) $\Leftrightarrow$ (v) und (iv) $\Rightarrow$ (iii) falls $\mu(A) < \infty$ $\forall A \in \mathscr{R}$.

Beweis:

(i) ⇒ (ii): Lemma 1.9

(ii) ⇒ (i): Seien $A_1, A_2, \ldots \in \mathscr{R}$, p. d. so dass $\bigcup_{i=1}^\infty A_i \in \mathscr{R}$

$$\sum_{i=1}^\infty \mu(A_i) = \lim_{n\to\infty} \sum_{i=1}^n \mu(A_i) = \lim_{n\to\infty} \mu\Big(\sum_{i=1}^n A_i\Big) \overset{\mu \text{ monoton}}{\leq} \mu\Big(\sum_{i=1}^\infty A_i\Big) \overset{(ii)}{\leq} \sum_{i=1}^\infty \mu(A_i)$$

(i) ⇒ (iii): Seien $A_1 \subseteq A_2 \subseteq \cdots \in \mathscr{R}$ und $\bigcup_{i=1}^\infty A_i \in \mathscr{R}$. Dann ist

$$\mu\Big(\sum_{i=1}^\infty A_i\Big) = \sum_{i=1}^\infty \mu(A_i \backslash A_{i-1}) = \lim_{n\to\infty} \sum_{i=1}^n \mu(A_i \backslash A_{i-1}) = \lim_{n\to\infty} \mu(A_n).$$

(iii) ⇒ (i): Seien $A_1, A_2, \ldots \in \mathscr{R}$ p. d. und $\bigcup_{i=1}^\infty A_i \in \mathscr{R}$. Dann ist $\big(\bigcup_{i=1}^n A_i \in \mathscr{R}\big)_{n\geq 1}$ monoton wachsend, und

$$\mu\Big(\sum_{i=1}^\infty A_i\Big) = \lim_{n\to\infty} \mu\Big(\sum_{i=1}^n A_i\Big) = \sum_{i=1}^\infty \mu(A_i).$$

(iii) ⇒ (iv): (Übergang zu Komplementen) Seien $A_1 \supseteq A_2 \supseteq \ldots \in \mathscr{R}$ mit $\bigcap A_i = \varnothing$ und $B_n = A_1 \setminus A_n$. Dann ist $B_1 \subseteq B_2 \subseteq \ldots$ und $\bigcup B_i = A_1$ da $\bigcap A_i = \varnothing$. Außerdem ist

$$\begin{aligned} \mu(A_1) = \lim_{n\to\infty} \mu(B_n) &= \lim_{n\to\infty} \mu(A_1) - \mu(A_n) \\ &= \mu(A_1) - \lim_{n-\infty} \mu(A_n), \end{aligned}$$

so dass $\lim_{n\to\infty} \mu(A_n) = 0$ folgt. Hierbei haben wir $\mu(A_1) < \infty$. genutzt.

(iv) ⇒ (v): Seien $A_1 \supseteq A_2 \supseteq \cdots \supseteq \in \mathscr{R}$ und $A = \bigcap_{n=1}^\infty A_n \in \mathscr{R}$. Setze $B_n = A_n \setminus A$ und nutze (iv).

(v) ⇒ (iv): klar

(iv) ⇒ (iii): (mit μ endlich) Seien $A_1 \subset A_2 \subseteq \cdots \in \mathscr{R}$ und $\bigcup_{i=1}^\infty A_i \in \mathscr{R}$. Setze $B_n = \bigcup_{i=1}^\infty A_i \setminus A_n$, so dass $\bigcap_{n=1}^\infty B_n = \varnothing$, also

$$0 = \lim_{n\to\infty} \mu(B_n) = \mu\Big(\bigcup_{i=1}^\infty A_i\Big) - \lim_{n\to\infty} \mu(A_i). \qquad \square$$

1.2 Die Fortsetzung von Maßen

In diesem Abschnitt wollen wir nun von einem Halbring auf eine σ-Algebra einen Inhalt geeignet zu einem Maß fortsetzen.

Satz 1.15 (Eindeutigkeit). *Sei $\mathscr{C} \subset \mathscr{P}(\Omega)$ durchschnittsstabil, es gebe $C_1 \subseteq C_2 \subseteq \dots$, mit $C_n \in \mathscr{C}$, $n \geq 1$, so dass $\bigcup_{n=1}^\infty C_n = \Omega$ und $\mathscr{F} = \sigma(\mathscr{C})$. Weiterhin seien $\mu, \nu : \mathscr{F} \to \bar{\mathbb{R}}_{\geq 0}$ Maße, so dass $\mu\big|_{\mathscr{C}}$, $\nu\big|_{\mathscr{C}}$ σ-finit sind.*

Dann ist

$$\mu = \nu \Leftrightarrow \mu(C) = \sigma(C) \quad \forall C \in \mathscr{C}.$$

Beweis:

„$\Rightarrow$“ klar

„$\Leftarrow$“ Sei $C \in \mathscr{C}$ mit $\mu(C) = \nu(C) < \infty$ und

$$\mathscr{D}_C := \{A \in \mathscr{F} : \mu(A \cap C) = \sigma(A \cap C\} \supseteq \mathscr{C}.$$

Dann ist $\mathscr{D}_C$ Dynkin-System, denn:

(i) $\Omega \in \mathscr{D}_C$,

(ii) $A, B \in \mathscr{D}_C$,

$$\begin{aligned} A \subseteq B \Rightarrow \mu((B \setminus A) \cap C) &= \mu(B \cap C) - \mu(A \cap C) = \nu(B \cap C) - \nu(A \cap C) \\ &= \nu((B \setminus A) \cap C), \quad \text{also } B \setminus A \in \mathscr{D}_C. \end{aligned}$$

(iii) Ist $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{D}_C$ mit $\bigcup_{i=1}^\infty A_i \in \mathscr{D}_C$, so folgt wegen der σ-Stetigkeit von μ, ν

$$\mu\Big(\bigcup_{i=1}^\infty A_1 \cap C\Big) = \lim_{n\to\infty} \mu(A_n \cap C) = \lim_{n\to\infty} \nu(A_n \cap C) = \nu\Big(\bigcup_{i=1}^\infty A_i\Big)$$

Dann ist $\mathscr{F} = \sigma(C) = \mathscr{D}_C$, was ebenfalls für alle C_n gilt. Nun betrachten wir $A \in \mathscr{F}$. Dann ist $A \cap C_n \in \mathscr{D}_{C_n}$, also $\mu(A \cap C_n) = \nu(A \cap C_n)$. Wir erhalten

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap C_n) = \lim_{n\to\infty} \nu(A \cap C_n) = \nu(A). \qquad \square$$

Wir kommen zu den von C. Carathéodory (1873–1950) eingeführten, zentralen Begriff von einem äußeren Maß. Wie bereits erwähnt, wird lediglich die σ-Additivität durch die σ-Subaddititivtät ersetzt. In der folgenden Definition wird aber auch die Wortgebung äußeres Maß klar, vergleiche Satz 1.18. Alternativ wird für subadditive Mengenfunktionen (auf σ-Algebran) auch der Begriff Kapazität verwendet.

Definition 1.16. Eine Abbildung $\eta : \mathscr{P}(\Omega) \to \bar{\mathbb{R}}_{\geq 0}$ heißt *äußeres Maß*, falls

(i) $\eta(\varnothing) = 0$

(ii) $A \subseteq B \subseteq \Omega \Rightarrow \eta(A) \leq \eta(B)$ (Monotonie)

(iii) Für $A_1, A_2, \dots \in \mathscr{P}(\Omega)$ gilt

$$\eta\Big(\bigcup_{i=1}^{\infty} A_1\Big) \leq \sum_{i=1}^{\infty} \eta(A_i) \qquad (\sigma\text{-Subadditivität})$$

Ein äußeres Maß ist natürlich auch subadditiv. Da es auf allen Teilmengen von Ω definiert ist, gibt es zunächst keinen vernünftigen Begriff der Meßbarkeit.

Definition 1.17. Ist η ein äußeres Maß und $A \subseteq \Omega$, so heißt A *σ-messbar*, falls für alle $B \subseteq \Omega$

$$\eta(B) \geq \eta(B \cap A) + \eta(B \cap A^c).$$

Damit ist eine Menge A genau dann messbar, wenn sie **jede** Menge $B \subseteq \Omega$ zerlegt in die disjunkten Mengen $B \cap A$, $B \cap A^c$, auf denen sich η **additiv** verhält.

Satz 1.18. *Sei $\mathscr{H}$ ein Halbring und μ ein* Inhalt *auf $\mathscr{H}$. Setze für $A \subseteq \Omega$ ($\inf \varnothing = \infty$)*

$$\mu^*(A) := \inf\Big\{\sum_{n=1}^{\infty} \mu(A_n) : A_n \in \mathscr{H}, A \subseteq \bigcup_{n=1}^{\infty} A_n\Big\}. \tag{1.19}$$

Dann ist μ^* äußeres Maß.

$$(1.19) \quad \Leftrightarrow \qquad \mu^*(A) = \inf\Big\{\sum_{i=1}^{\infty} \mu(A_n) : A_n \in \mathscr{H} \text{ p. d.}, A \subseteq \sum_{n=1}^{\infty} A_n\Big\}$$

Beweis. Offensichtlich ist μ^* monoton und $\mu^*(\varnothing) = \mu(\varnothing) = 0$.

Seien $A_1, A_2, \dots \subseteq \Omega$. Ist $\mu(A_n) = \infty$ für ein $n \in \mathbb{N}$, so sind wir fertig.

Sei also $\mu^*(A_n) < \infty \, \forall n$ und $\varepsilon > 0$. Dann gibt es (Infimum!) für jedes $n \in \mathbb{N}$ eine Folge $(B_{nk})_{k\geq 1} \subset \mathscr{H}$, so dass $A_n \subseteq \bigcup_{k=1}^{\infty} B_{nk}$ und

$$\sum_{k=1}^{\infty} \mu(B_{nk}) \leq \mu^*(A_n) + \varepsilon \cdot 2^{-n}.$$

Nun ist $(B_{nk})_{n,k}$ eine abzählbare Familie von Mengen aus $\mathscr{H}$ und

$$\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty}\bigcup_{k=1}^{\infty} B_{nk}, \text{ also}$$

$$\begin{aligned}\mu^*\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\mu(B_{nk}) &\leq \sum_{n=1}^{\infty}(\mu^*(A_n)+\varepsilon\cdot 2^{-n}) \\ &= \sum_{n=1}^{\infty}\mu^*(A_n)+2\varepsilon.\end{aligned}$$

Da ε beliebig war, ist μ^* äußeres Maß. □

Satz 1.20. *Sei μ^* ein äußeres Maß. Dann ist*

$$\mathscr{F}^* := \{A \in \mathscr{P}(\Omega) : A \textit{ ist } \mu^*\textit{-messbar}\}$$

eine σ-Algebra und $\mu^*\big|_{\mathscr{F}^*}$ ein Maß.

Beweis. (i): Wir zeigen zunächst, dass $\mathscr{F}^*$ eine Algebra ist.

$$\Omega \in \mathscr{F}^* \text{ und } A \in \mathscr{F}^* \Rightarrow A^c \in \mathscr{F}^* \qquad \text{(direkt aus Def. von } \mu^*\text{-messbar)}$$

Seien $A, B \in \mathscr{F}^*$ und $C \subset \Omega$. Dann ist

$$\begin{aligned}\mu^*(C) &\geq \mu^*(C\cap A)+\mu^*(C\cap A^c) \qquad (A\in\mathscr{F}*)\\ &\geq \mu^*(C\cap A)+\mu^*(C\cap A^c\cap B)+\mu^*(C\cap A^c\cap B^c) \qquad (B\in\mathscr{F}^*)\\ &\geq \mu^*(C\cap(A\cup B))+\mu^*(C\cap(A\cup B)^c)\end{aligned}$$

und somit sind $A\cup B \in \mathscr{F}^*$. Sind A, B auch noch paarweise disjunkt, so folgt

$$\mu^*(C) \geq \mu^*(C\cap A)+\mu^*(C\cap B)+\mu^*(C\cap(A\cup B)^c). \tag{1.21}$$

(ii) Nun zeigen wir: Sind $(A_n)_{n\geq 1} A_i \in \mathscr{F}^*$ p. d. $\Rightarrow A = \bigcup_{i=1}^{\infty} \in \mathscr{F}^*$ und μ^* ist σ-additiv: Zunächst folgt $\sum_{i=1}^{n} A_i \in \mathscr{F}^*$ aus (i). Für jedes $C \subseteq \Omega$ gilt

$$\begin{aligned}\mu^*(C) &\geq \mu^*\Big(C\cap\Big(\sum_{i=1}^{n}A_i\Big)\Big)+\mu^*\Big(C\cap\Big(\sum_{i=1}^{n}A_i\Big)^c\Big)\\ &\geq \sum_{i=1}^{n}\mu^*(C\cap A_i)+\mu^*(C\cap A^c). \qquad \text{(Monotonie \& (1.21))}\end{aligned}$$

Dies gilt für alle n, also folgt

$$\begin{aligned}\mu^*(C) &\geq \sum_{i=1}^{\infty} \mu^*(C \cap A_i) + \mu^*(C \cap A^c) \\ &\geq \mu^*(C \cap A) + \mu^*(C \cap A^c) \geq \mu^*(C)\end{aligned}$$

wegen der Sub-σ-Additivität. Es folgt $A \in \mathscr{F}^*$ und mit $A = C$ die σ-Additivität von μ^*. □

Theorem 1.22 (Fortsetzungssatz)**.** Sei $\mathscr{H}$ ein Halbring und $\mu : \mathscr{H} \to \bar{\mathbb{R}}_{\geq 0}$ ein Inhalt.

(i) Dann sind alle Mengen aus $\mathscr{H}$ μ^*-messbar.

(ii) Ist μ σ-additiv, so gilt $\mu^*\big|_{\mathscr{H}} = \mu$.

(iii) Ist μ nicht σ-additiv, so gibt es $A \in \mathscr{H}$, so dass $\mu^*(A) < \mu(A)$.

Insbesondere ist $\mu^*\big|_{\mathscr{F}^*}$ Maß und damit auch $\mu^*\big|_{\sigma(\mathscr{H})}$.

Beweis. (i) Sei $A \in \mathscr{H}$ und $C \subseteq \Omega$ mit $\mu^*(C) < \infty$ (für $\mu^*(C) = \infty$ ist nicht zu zeigen) und $(B_n)_{n\geq 1} \subseteq \mathscr{R}(\mathscr{H})$ so, dass $C \subseteq \bigcup_{i=1}^{\infty} B_n$ und $\mu^*(C) + \varepsilon \geq \sum \widetilde{\mu}(B_n)$ (existiert, da $\mu^*(C) < \infty$).

Wir betrachten den auf $\mathscr{R}(\mathscr{H})$ erzeugten Inhalt $\widetilde{\mu}$.

$$\begin{aligned}\mu^*(C) + \varepsilon \geq \sum_{n=1}^{\infty} \widetilde{\mu}(B_n) &= \sum_{n=1}^{\infty} \widetilde{\mu}(B_n \cap A) + \sum_{n=1}^{\infty} \widetilde{\mu}(B_n \cap A^c) \\ &\geq \mu^*(C \cap A) + \mu^*(C \cap A^c).\end{aligned}$$

Da $\varepsilon > 0$ beliebig, folgt $A \in \mathscr{F}^*$.

(ii) Sei $A \in \mathscr{H}$. Für jedes $\varepsilon > 0$ gibt es $(A_n)_{n\geq 1} \subset \mathscr{H}$, so dass $A \subseteq \bigcup_{n\geq 1} A_n$ und (Def. μ^*)

$$\mu^*(A) + \varepsilon \geq \sum_{n\geq 1} \mu(A_n) \underset{\sigma\text{-Sub.-Add.}}{\geq} \mu\Big(\bigcup_{n\geq 1} A_n\Big) \geq \mu(A) \geq \mu^*(A)$$

und es folgt $\mu^*(A) = \mu(A)$.

(iii) Ist μ nicht σ-additiv, so gibt es p. d. $(A_n) \subseteq \mathscr{H}$, $A = \sum_{n\geq 1} A_n \in \mathscr{H}$ mit

$$\mu\Big(\sum_{i=1}^{\infty} A_n\Big) \neq \sum_{i=1}^{\infty} \mu(A_n).$$

Allerdings ist stets (μ Inhalt) $\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\Big(\sum_{n=1}^{\infty} A_n\Big)$, also $\sum_{n=1}^{\infty} \mu(A_n) < \mu(A)$. Nun ist $\mu^*(A_n) = \mu(A)$ und μ^* σ-additiv, also $\sum_{n=1}^{\infty} \mu^*(A_n) = \mu^*(A) < \mu(A)$. □

Man kann noch zeigen, dass

$$\mathscr{F}^* = \{A \setminus N : A \in \sigma(\mathscr{H}), N \in \Omega, \mu^*(N) = 0\}.$$

1.3 Maße auf $\mathbb{R}$

Wir können nun das Lebesgue-Maß λ eindeutig auf $\mathscr{B}(\mathbb{R})$ definieren durch

$$\lambda((a,b]) = b - a, \qquad a, b \in \mathbb{Q} \text{ mit } a \leq b.$$

Wir setzen hierbei $(a,a] = \varnothing$. Ebenso erhalten wir eine Charakterisierung aller Wahrscheinlichkeitsmaße auf $\mathscr{B}(\mathbb{R})$.

Satz 1.23. *Eine Funktion $P : \mathscr{B}(\mathbb{R}) \to [0,1]$ ist ein Wahrscheinlichkeitsmaß genau dann, wenn es eine wachsende rechtsstetige Funktion $F : \mathbb{R} \to [0,1]$ gibt, so dass*

$$P((a,b]) = F(b) - F(a), \qquad a, b \in \mathbb{Q}, a \leq b.$$

Offensichtlich ist P eindeutig durch F bestimmt!

Beweis. „$\Rightarrow$ “: klar

„$\Leftarrow$ “: Wir zeigen, dass P eine σ-additive Mengenfunktion auf dem Erzeugenden-Halbring

$$\mathscr{H} = \{(a,b] : a \leq b, \ \ a, b \in \mathbb{Q}\} \text{ ist.}$$

Seien also $\{(a_i, b_i] : i \geq 1\}$ paarweise disjunkt (p. d.) und $\sum_{i \geq 1}(a_i, b_i] = (a,b]$. Ohne Einschränkung können wir umordnen in $\{(c_i, c_{i-1}]\}$ mit $\sum_{i \geq 1}(c_i, c_{i-1}] = (a,b]$ und $c_1 \geq c_2 \geq \ldots$ Dann gilt $c_i \to a$ und $c_1 = b$. (!)

Da F rechtsstetig ist, folgt

$$\begin{aligned} P((a,b]) &= F(b) - F(a) = F(c_1) - \lim_{n \to \infty} F(c_n) \\ &= \sum_{n=2}^{\infty} F(c_n) - F(c_{n-1}) = \sum_{n=2}^{\infty} P((c_n, c_{n-1}]). \end{aligned}$$

Die Behauptung folgt nun aus Theorem 1.22. $\square$

Für uns wichtig werden Verteilungen, also Bildmaße sein.

Definition 1.24. Ein Tripel $(\Omega, \mathscr{F}, \mu)$ heißt *Maßraum*, falls $\mathscr{F}$ σ-Algebra und μ Maß auf $(\Omega, \mathscr{F})$ ist.

Das Paar $(\Omega, \mathscr{F})$ bezeichnen wir als *messbaren Raum*.

Definition 1.25. Sei $(\Omega, \mathscr{F}, \mu)$ ein Maßraum, $\mathscr{F}'$ eine σ-Algebra auf Ω'. Eine Funktion $f : \Omega \to \Omega'$ heißt *$\mathscr{F}$-$\mathscr{F}'$-messbar*, falls

$$f^{-1}(A) \in \mathscr{F} \text{ für alle } A \in \mathscr{F}'.$$

Für ein solches f heißt $f_*\mu : \mathscr{F}' \to \mathbb{R}_{\geq 0} \cup \{0\}$, definiert durch

$$f_*\mu(A') = \mu(f^{-1}(A')) = \mu(\omega \in \Omega : f(\omega) \in A'), \qquad A' \in \mathscr{F}'$$

Bildmaß von μ unter f.

Man zeigt leicht, dass $f_*\mu$ wieder ein Maß ist. $f_*\mu$ wird oft auch als $f_*(\mu)$, $\mu \circ f^{-1}$ notiert und *push-forward* genannt.

Die **von f erzeugte** σ-Algebra $f^{-1}(\mathscr{F}')$ ist in der Tat eine σ-Algebra. Ebenso gilt für $\sigma(\mathscr{C}') = \mathscr{F}'$, dass

$$\sigma(f^{-1}(\mathscr{C}')) = f^{-1}(\sigma(\mathscr{C}')).$$

Ist $(\Omega', \mathscr{F}') = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, so heißt f **reellwertig** und ist f $\mathscr{F}$-$\mathscr{B}(\mathbb{R})$-messbar, so nennen wir f **Borel-messbar**, oder oft auch einfach messbar.

Messbarkeit ist ein wichtiger Begriff, so dass wir ein paar Rechenregeln wiederholen. Wir schreiben $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Lemma 1.26. *Seien* $(\Omega, \mathscr{F})$, $(\Omega', \mathscr{F}')$, $(\Omega'', \mathscr{F}'')$ *Maßräume auf* $f : \Omega \to \Omega'; g : \Omega' \to \Omega''$.

(i) $\mathscr{F}' = \sigma(\mathscr{C}') \Rightarrow f$ *messbar* $\Leftrightarrow f^{-1}(\mathscr{C}') \subseteq \mathscr{F}$

(ii) f, g *messbar* $\Rightarrow f \circ g$ *messbar*

(iii) stetige Abbildungen sind messbar (auf topologischen Räumen !) bzgl. der Borel-σ*-Algebren.*

(iv) Eine reellwertige Funktion f *ist genau dann Borel-messbar, falls*

$$\{\omega : f(\omega) \leq x\} \in \mathscr{F} \qquad \forall x \in \mathbb{Q}$$

(v) $f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$ *ist messbar* $\Leftrightarrow A_1, \dots, A_n \in \mathscr{F}$.

(vi) $f, g : \Omega \to \overline{\mathbb{R}}$ *messbar* $\Rightarrow f \cdot g, a \cdot f + b \cdot g, \frac{f}{g} \mathbb{1}_{\{g \neq 0\}}$ *messbar* $(a, b \in \mathbb{R})$

(vii) $f_1, f_2, \cdots : \Omega \to \overline{\mathbb{R}}$ *messbar* $\Rightarrow \sup_n f_n, \inf_n f_n, \limsup_n f_n, \liminf_n f_n$ *sind messbar.*

Den Beweis kann man als eine Übungsaufgabe führen. Am interessantesten ist wohl der Beweis für die Messbarkeit von $\sup_n f_n$:

$$\{\omega : \sup f_n(\omega) \leq x\} = \bigcap_{n \geq 1} \{\omega : f_n(\omega) \leq x\} \in \mathscr{F}.$$

Eine wichtige Eigenschaft nicht-negativer, messbarer Funktionen ist, dass sie stets durch einfache Funktionen approximierbar sind: Wir setzen:

$$A_{j,n} = \begin{cases} \{\frac{j}{2^n} \leq f \leq \frac{j+1}{2^n}\}, & j = 0, \dots, n \cdot 2^n - 1 \\ \{f \geq n\}, & j = n \cdot 2^n \end{cases}.$$

Mit

$$f_n = \sum_{j=0}^{n2^n} \frac{j}{2^n} \mathbb{1}_{\{A_{j,n}\}}$$

erhalten wir die gewünschte, monotone Approximation, $f_n \uparrow f \geq 0$. Dies ist der Schlüssel zum Integral! Wir definieren für einfache Funktionen

$$\int f d\mu = \int \sum_{i=1}^{n} c_i \mathbb{1}_{A_i} d\mu = \sum_{i=1}^{n} c_i \int_{A_i} d\mu = \sum_{i=1}^{n} c_i \mu(A_i)$$

und für messbare, nicht-negative Funktionen und $f_n \uparrow f$

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu = \sup\Big\{ \int g d\mu : g \text{ einfach und } 0 \leq g \leq f \Big\}.$$

Dieses Integral können wir fortsetzten, falls mindestens ein Integral $\int f^+ d\mu, \int f^- d\mu$ endlich ist. Dann definieren wir

$$\mathscr{L}^1(\mu) := \Big\{ f : \Omega \to \overline{\mathbb{Q}} : \int |f| d\mu < \infty \Big\} \qquad \text{(und analog } \mathscr{L}^p\text{)}$$

Leider hat dieser Raum schlechte Trennungseigenschaften: Gibt es eine nichtleere Null-Menge, so gibt es immer verschiedene meßbare Funktionen, die den Abstand Null haben. Um das zu beheben, und letzten Endes Banach-Räume zu erhalten, betrachtet man Äquivalenzklassen, siehe auch Abschnitt VI.2 in Elstrodt(2013). Auch im Buch Bauer(1990) gibt es ein kleines Kapitel über L^p-Räume.

Auf den messbaren Funktionen kann man folgende Äquivalenzrelation einführen: $f \sim g$ falls $\mu(f = g) = 1$. Die hierdurch erhaltenen Äquivalenzklassen definieren die L^p-Räume. In der Notation machen wir dies durch die Unterscheidung $\mathscr{L}$ und L kenntlich. Im Folgenden kürzt f. ü. fast überall ab, etwa $f \leq g$ f. ü. ist gleichbedeutend mit $\mu(\omega \in \Omega : f > g) = 0$.

Satz 1.27. *Seien $f, g, f_1, f_2, \dots$ meßbar. Dann gilt:*

(i) $f \leq g$ *f. ü.* $\Rightarrow \int f d\mu \leq \int g d\mu$

(ii) $|\int f d\mu| \leq \int |f| d\mu$ *(Dreiecksungleichung)*

(iii) $\int (af + bg) d\mu = \int f d\mu + b \int g d\mu$ *(Linearität)*

(iv) $f = 0$ *f. ü.* $\Rightarrow \int f d\mu = 0$

(v) $\int f d\mu < \infty \Rightarrow f < \infty$ *f. ü.*

Als Übung beweisen wir den Substitutionssatz

Satz 1.28. $g \in \mathscr{L}^1(f_*\mu) \Rightarrow g \circ f \in \mathscr{L}^1(\mu)$ *und*

$$\int g \circ f d\mu = \int g \, d(f_*\mu).$$

Beweis. Es genügt, die Aussage für einfache nicht-negative g zu zeigen. Der allgemeine Fall folgt durch Approximation.

Sei $g = \sum_{i=1}^n c_i \mathbb{1}_{A_i} \Rightarrow g \circ f = \sum_{i=1}^n c_i \mathbb{1}_{\{f \in A_i\}}$, also

$$\int g \circ f d\mu = \sum_{i=1}^n c_i \mu(f \in A_i) = \sum_{i=12}^n c_i f_* \mu(A_i) = \int g d(f_* \mu). \qquad \square$$

Äußerst wichtig sind die folgenden Konvergenzsätze.

Theorem 1.29 (Monotone Konvergenz)**.** Sei $(\Omega, \mathscr{F}, \mu)$ Maßraum, $f_n \geq 0$ und $f : \Omega \to \overline{\mathbb{R}}$ messbar mit $f_n \uparrow f$ f. ü. $\Rightarrow$

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

Beweis. Folgt direkt aus der monotonen Konvergenz (Teil (vi) in 1.27) mit

$$g_n = (f - f_n) \mathbb{1}_{\{\omega \in \Omega : f_n(\omega) \uparrow f(\omega)\}}. \qquad \square$$

Bemerkung 1.30. Auf die Forderung $f \geq 0$ kann nicht verzichtet werden. Betrachten wir das Lebesgue-Maß und die Funktionenfolge

$$f_n(x) = -\mathbb{1}_{\{x \leq -n\}},$$

so konvergiert $f_n \uparrow f = 0$ monoton, aber $\int f_n d\mu = -\infty$ und $\int f d\mu = 0$.

Theorem 1.31 (Fatou)**.** Sei $(\Omega, \mathscr{F}, \mu)$ Maßraum $f_1, f_2, \ldots : \Omega \to \overline{\mathbb{R}}$ messbar mit $f_n \geq 0$. Dann gilt

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

Beweis. Ist n fest, so gilt für alle $j \geq n$, dass $f_j \geq \inf_{k \geq n} f_k$. Wegen der Monotonie des Integrals erhält man, dass

$$\inf_{j \geq n} \int f_j d\mu \geq \int \inf_{k \geq n} f_k d\mu.$$

Für $n \to \infty$ erhalten wir

$$\liminf_{n \to \infty} \int f_n d\mu \geq \sup_{n \in \mathbb{N}} \int \inf_{k \geq n} f_k d\mu = \int \liminf_{n \to \infty} f_n d\mu$$

wegen monotoner Konvergenz ($\inf\limits_{k \geq n} f_k \uparrow \liminf\limits_{n \to \infty} f_n$). $\square$

Theorem 1.32 (Majorisierte Konvergenz)**.** Sei $(\Omega, \mathscr{F}, \mu)$ Maßraum und $f, g, (f_n)_{n\geq 1} : \Omega \to \overline{\mathbb{R}}$ messbar.

Sei $|f_n| \leq g$ f. ü. mit $g \in \mathscr{L}^1(\mu)$ und $\lim_{n\to\infty} f_n = f$. Dann gilt

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

Beweis. Zunächst sei ohne Einschränkung $|f_n| \leq g \ \ \forall \omega \in \Omega$. Die Idee ist, Fatou auf $g+f$, $g-f$ anzuwenden, wobei $g + f_n \geq 0$ und $g - f_n \geq 0$ aus der Voraussetzung folgt. Es gilt, dass

(1) $\displaystyle\int (g+f)d\mu \leq \liminf_{n\to\infty} \int (g+f_n)d\mu = \int g d\mu + \liminf_{n\to\infty} \int f_n d\mu$

(2) $\displaystyle\int (g-f)d\mu \leq \int g d\mu - \limsup_{n\to\infty} \int f_n d\mu$, indem wir Fatou auf $(g-f_n)$ anwenden.

Aus diesen beiden Gleichungen erhalten wir, dass

$$\int f d\mu \overset{(1)}{\leq} \liminf_{n\to\infty} \int f_n d\mu \leq \limsup_{n\to\infty} \int f_n d\mu \overset{(2)}{\leq} \int f d\mu.$$

□

Wir erinnern kurz an die $L^p(\mu)$-Räume

$$\begin{aligned} \mathscr{L}^p(\mu) &= \{f : \Omega \to \overline{\mathbb{R}} \text{ messbar mit } \|f\|_p < \infty\} \text{ und} \\ \|f\|_p &= \Big(\int |f|^p d\mu\Big)^{1/p}. \\ \|f\|_\infty &= \inf\{K : \mu(|f| > K) = 0\}. \end{aligned}$$

Es gilt für messbare f, g, dass

(i) $0 < p, q, r \leq \infty$ und $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$: $\quad \|fg\|_r \leq \|f\|_p \|g\|_q$ (Hölder-Ungleichung)

(ii) $1 \leq p \leq \infty$: $\quad \|f+g\|_p \leq \|f\|_p + \|g\|_p$ (Minkowski-Ungleichung)

Definition 1.33 (Konvergenz im p-ten Mittel)**.** Eine Folge $(f_n)_{n\geq 1} \subset \mathscr{L}^p(\mu)$ **konvergiert im p-ten Mittel**, falls

$$\|f_n - f\|_p \underset{n\to\infty}{\longrightarrow} 0.$$

Wir schreiben

$$f_n \overset{\mathscr{L}^p(\mu)}{\underset{n\to\infty}{\longrightarrow}} f.$$

Konvergiert eine Folge im p-ten Mittel, so auch im q-ten Mittel, falls $q < p$. (Hölder) Außerdem ist $\mathscr{L}^p(\mu)$, $p \geq 1$ **vollständig** (jede Cauchy-Folge konvergiert); das ist der berühmte Satz von Riesz-Fischer, siehe etwa Satz VI.2.5 in Elstrodt(2013). Allerdings ist $\mathscr{L}^p(\mu)$ kein (!) normierter Raum, denn $N_p = (\int |f|^p d\mu)^{1/p}$ ist lediglich eine Halbnorm. Dies können wir, wie bereits erwähnt, beheben durch Übergang zu Äquivalenzklassen: $L^p(\mu)$ ist ein normierter und vollständer Rau, und damit ein Banach-Raum.

1.4 Hilberträume

Auf dem $\mathscr{L}^2(\mu)$ hat man eine besondere Struktur, denn dieser Raum ist sogar ein Hilbertraum mit Skalarprodukt

$$\langle f, g \rangle = \int f\, g\, d\mu.$$

Dann lassen sich die allgemeinen Resultate anwenden (siehe Werner(2000)).

Lemma 1.34 (Parallelogrammidentität)**.** *Ein Banachraum* $(X, \|\cdot\|)$ *ist genau dann ein Hilbertraum falls* $\forall\, f, g$ *gilt, dass*

$$\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|^2).$$

Beweis. Siehe Werner(2000), Satz V.1.6. □

Satz 1.35. *Sei* M *ein abgeschlossener, linearer Unterraum des Hilbertraums* H *und* $f \in H$*. Dann gibt es genau ein* $g \in M$*,* $h \perp M$*, so dass*

$$f = g + h.$$

Beweis. Siehe Werner(2000), Theorem V.3.4. □

Satz 1.36 (Riesz-Fréchet)**.** *Sei* H *ein Hilbertraum und* $F : H \to \mathbb{R}$*.* F *ist genau dann stetig und linear, falls es ein* $f \in H$ *gibt, so dass*

$$F(h) = \langle f, h \rangle \qquad \forall\, h \in H.$$

Hierbei ist f *eindeutig.*

Beweis. Siehe Werner(2000), V.3.6. □

1.5 Signierte Maße und der Satz von Radon-Nikodým

Wir wenden uns kurz diesem allgemeineren Begriff zu. Vorstellen kann man sich ein signiertes Maß als Ladungsverteilung mit positiver und negativer Ladung. Sei $(\Omega, \mathscr{F})$ ein fester messbarer Raum.

Definition 1.37. Eine Abbildung $\nu : \mathscr{F} \to \mathbb{R}$ heißt *signiertes Maß*, falls

(i) $\nu(\varnothing) = 0$

(ii) $\nu(\mathscr{F}) = \{\nu(F) : F \in \mathscr{F}\}$ ist entweder Teilmenge von $(-\infty, +\infty]$ oder von $[-\infty, \infty)$.

(iii) Sind $(A_n) \in \mathscr{F}$ p. d., so gilt

$$\nu\Big(\sum_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \nu(A_n).$$

Beispiel 1.38. Hat man Maße μ_1, μ_2, wobei eines davon endlich ist, so ist $\nu = \mu_1 - \mu_2$ signiertes Maß. In der Tat hat jedes signierte Maß eine solche Gestalt und μ_1, μ_2 sind bei geeignet „minimaler“ Wahl sogar eindeutig.

Definition 1.39. Ist ν signiertes Maß und $F \in \mathscr{F}$.

(i) F heißt *ν-positiv*, falls $\nu(A) \geq 0$ für alle $\mathscr{F} \ni A \subseteq F$

(ii) F heißt *ν-negativ*, falls $\nu(A) \leq 0$ für alle $\mathscr{F} \ni A \subseteq F$

(iii) F heißt *ν-Nullmenge*, falls $\nu(A) = 0$ für alle $\mathscr{F} \ni A \subseteq F$

Lemma 1.40. *Ist $\nu : \mathscr{F} \to [-\infty, \infty)$ signiertes Maß und $A \in \mathscr{F}$ mit $\nu(A) \neq -\infty \Rightarrow$ es existiert eine positive Menge P mit $\nu(P) \geq \nu(A)$.*

Beweis. (i) Wir zeigen: Zu jedem $\varepsilon > 0$ gibt es $\mathscr{F} \ni A_\varepsilon \subseteq A$ mit $\nu(A_\varepsilon) \geq \nu(A)$ und $\nu(B) \geq -\varepsilon$ für alle $\mathscr{F} \ni B \subseteq A_\varepsilon$.

Durch Widerspruch: Angenommen, es existiert $\varepsilon > 0$, so dass die Behauptung falsch ist. Dann enthält jede messbare Menge $C \subseteq A$ mit $\nu(C) \supseteq \nu(A)$ ein $B \in \mathscr{F}$, so dass $\nu(B) \leq -\varepsilon$. Wir erhalten eine Folge $B_1 \subseteq A$, $B_k \subseteq A \setminus (B_1 \cup \cdots \cup B_{k-1})$ mit $\nu(B_k) \leq -\varepsilon$, $k \geq 1 \Rightarrow \nu(\sum B_k) = -\infty$. ↯

Wir erhalten eine fallende Folge $(A_{1/n}) \subseteq \mathscr{F}$ und $P := \bigcap A_{1/n}$ ist positiv sowie $\nu(A_{1/n}) \geq \nu(A)$, also auch $\lim_{n\to\infty} \nu(A_{1/n}) = \nu(P)$. □

Satz 1.41 (Hahn-Zerlegung)**.** *Zu jedem signierten Maß ν existiert eine disjunkte Zerlegung $\Omega = P + N$, in eine ν-positive Menge $P \in \mathscr{F}$ und eine ν-negative Menge $N \in \mathscr{F}$.*

Beweis. Ohne Beschränkung der Allgemeinheit sei $\nu(\mathscr{F}) \subseteq [-\infty, \infty)$ und $\alpha = \sup\{\nu(A) : A \in \mathscr{F}\}$. Nach Lemma 1.4 gibt es eine Folge positiver Mengen (A_n) mit $\nu(A_n) \to \alpha$. Dann ist $P = \bigcup_{n\geq 1} A_n \in \mathscr{F}$ positiv und $\nu(P) \geq \nu(A_n)$, also $\nu(P) = \alpha$. Außerdem ist nach Voraussetzung $\nu(P) < \infty$ und somit $\alpha \in \mathbb{R}$.

Nun ist $N = P^c$ negativ: Gäbe es $B \subset N$ mit $N(B) > 0$, so wäre $\nu(P \cup B) > \alpha$. □

Man sieht leicht, dass diese Zerlegung **eindeutig** bis auf Nullmengen ist.

Für ein signiertes Maß ν mit Hahn-Zerlegung $\Omega = P + N$ heißen

$$\left.\begin{array}{ll} \nu^+(A) & := \nu(A \cap P) \\ \nu^-(A) & := \nu(A \cap N) \end{array}\right\}, \qquad A \in \mathscr{F}$$

positive (negative) Variation von ν. Da P und N bis auf Nullmengen eindeutig sind, ist $\nu^{+/-}$ wohldefiniert.

Definition 1.42. Zwei signierte Maße μ, ν heißen *singulär* (und wir schreiben $\mu \perp \nu$), falls $\Omega = A + A^c$, $A \in \mathscr{F}$ mit $\mu(A) = 0$, $\nu(A^c) = 0$.

Satz 1.43 (Jordan-Zerlegung)**.** *Jedes signierte Maß ν erfüllt $\nu = \nu^+ + \nu^-$ mit $\nu^+ \perp \nu^-$. Diese Zerlegung ist* minimal*, d.h.: ist $\nu = \varrho - \sigma$ mit Maßen ϱ, σ, von denen mindestens eines endlich ist, so ist $\nu^+ \leq \varrho$, $\nu^- \leq \sigma$.*

Beweis. Es bleibt lediglich die Minimalität:

$$\nu^+(A) = \nu(A \cap P) = \varrho(A \cap P) - \sigma(A \cap P) \leq \varrho(A \cap P) \leq \varrho(A). \qquad \square$$

Beispiel 1.44. Sei $f : \Omega \to \overline{\mathbb{R}}$ quasi-integrierbar. Dann ist durch

$$d\mu = f dx$$

ein signiertes Maß definiert und $P = f^{-1}([0, \infty])$, $N = f^{-1}([-\infty, 0))$.

Definition 1.45. Wieder betrachten wir den festen meßbaren Raum $(\Omega, \mathscr{F})$.

(i) Das (signierte) Maß ν heißt *absolut stetig* bzgl. des Maßes μ, falls

$$\mu(A) = 0 \Rightarrow \nu(A) = 0 \qquad \text{für alle } A \in \mathscr{F}.$$

Dann schreiben wir $\nu \ll \mu$. Gilt zusätzlich $\mu \ll \nu$, so heißen μ und ν *äquivalent* und wir schreiben $\mu \sim \nu$.

(ii) Gilt

$$\nu(A) = \int_A f d\mu,$$

für alle $A \in \mathscr{F}$, so heißt f *Dichte* von ν bzgl. μ und wir schreiben

$$\frac{d\nu}{d\mu} = f \quad \text{oder} \quad d\nu = f d\mu.$$

Der Schlüssel zu dem Satz von Radon-Nikodým ist folgendes Lemma:

Lemma 1.46. *Sind ν, ϱ endliche Maße mit $\nu \leq \varrho$, so gibt es seine messbare Funktion $h : \Omega \to [0,1]$ mit*

$$d\nu = h\, d\varrho. \tag{1.47}$$

Beweis. Die Voraussetzungen ergeben, dass $L^2(\varrho) \subseteq L^2(\nu) \subseteq L^1(\nu)$. Damit ist die Abbildung $f \mapsto \int_\Omega f d\nu$, mit $f \in L^2(\varrho)$ wohldefiniert und stetig (da L^2 vollständig ist). Außerdem ist $L^2(\varrho)$ ein Hilbertraum. Nach Satz 1.36 von Riesz-Fréchet gibt es $h \in L^2(\varrho)$, so dass

$$\int f d\nu = \langle f, h \rangle_\varrho = \int f h\, d\varrho.$$

Dies gilt für alle $f \in L^2(\varrho)$, insbesondere für $f = 1_F$, $F \in \mathscr{F}$. Man sieht direkt, dass $h(\Omega) \in [0,1]$ ϱ-f.s. □

Wir sammeln einige technische Eigenschaften

Lemma 1.48. *Sei μ ein Maß auf $(\Omega, \mathscr{F})$ und $f, g : \Omega \to \bar{\mathbb{R}}$ messbar.*

(i) *Sind $f, g \in \mathscr{L}^1(\mu)$ und gilt*

$$\int_A f d\mu \leq \int_A g d\mu, \qquad A \in \mathscr{F}, \tag{1.49}$$

so folgt $f \leq g$ μ-fast sicher.

(ii) *Ist ν σ-finit und $d\nu = f d\mu = g d\mu$, so folgt*

$$f = g \qquad \mu\text{-f.s.}$$

(iii) *Ist μ σ-finit und sind f, g quasi-integrierbar mit $d\nu = f d\mu = g d\mu$, so folgt*

$$f = g \qquad \mu\text{-f.s.}$$

(iv) *Seien $f : \Omega \to \mathbb{R}_{\geq 0}$ und $g : \Omega \to \mathbb{R}$ messbar. Dann gilt*

$$\int g \cdot (f d\mu) = \int (f \cdot g) d\mu.$$

Beweis. (i) Wir setzen $A_n = \{f > g + \frac{1}{n}\} \in \mathscr{F}$, so dass $A_n \uparrow A = \{f > g\} \in \mathscr{F}$. Wir erhalten, dass

$$\int_{A_n} f d\mu \geq \int_{A_n} \Big(g + \frac{1}{n}\Big) d\mu = \int_{A_n} g d\mu + \frac{1}{n}\mu(A_n) \geq \int_{A_n} f d\mu + \frac{1}{n}\mu(A_n),$$

also $\mu(A_n) = 0$ (da $\int f d\mu < \infty$). Da aber $A_n \uparrow A$, folgt auch $\mu(A) = 0$. Gilt in (1.49) sogar Gleichheit, so erhalten wir damit auch $f = g$ μ-fast sicher.

(ii) Da ν σ-finit ist, gibt es $\Omega_1 \subseteq \Omega_2 \subseteq \cdots \subseteq \mathscr{F}$, so dass $\bigcup_{n \geq 1} \Omega_n = \Omega$ und $\nu(\Omega_n) < \infty$.

Wir setzen $A_n = \Omega_n \cap \{f < g\} \Rightarrow \int_{A_n} (f - g) d\mu = v(A_n) - \nu(A_n) = 0$, also $\mu(A_n) = 0$.

Damit ist

$$\mu(f > g) = \mu\Big(\bigcup_{n \geq 1} A_n\Big) = \lim_{n \to \infty} \mu(A_n) = 0.$$

(iii) Wir zeigen die Behauptung für $\leq$ - das Resultat folgt durch Vertauschung von f und g. Aus Symmetriegründen können wir f^- integrierbar annehmen. Dann ist auch g^- integrierbar. Sei (Ω_n) eine Folge wachsender, messbarer Mengen mit $\mu(\Omega_n) < \infty$ und $\Omega_n \to \Omega$. Wir setzen

$$A_n = \Omega_n \cap \{g \leq n\} \uparrow \{g < \infty\} =: A.$$

Dann ist $\mu(A_n) < \infty$ und $g\mathbb{1}_{A_n} \in \mathscr{L}^1(\mu)$. Für alle $F \in \mathscr{F}$ folgt, dass

$$-\infty < \int_F f\,\mathbb{1}_{A_n} d\mu \leq \int_F g\,\mathbb{1}_{A_n} d\mu < \infty.$$

Mit $F = \Omega$ erhalten wir, dass auch $f\mathbb{1}_{A_n} \in \mathscr{L}^1(\mu)$ ist. Nach Teil (i) folgt, dass $f\mathbb{1}_{A_n} \leq g\mathbb{1}_{A_n}$ μ-fast sicher (und zwar für alle n).

Mit $n \to \infty$ folgt, dass $f\mathbb{1}_{\{g<\infty\}} \leq g\mathbb{1}_{\{g<\infty\}}$ μ-fast sicher. Auf der Menge $A^c\{g = \infty\}$ gilt das natürlich auch und die Behauptung folgt.

(iv) Offensichtlich ist für $g = 1_A$

$$\int g(f \cdot d\mu) = \int_A f d\mu = \int (1_A \cdot f) d\mu.$$

Die übliche Approximation von messbaren g liefert das Resultat.

□

Beispiel 1.50. Natürlich kennen wir bereits Dichten bzgl. des Lebesgue-Maßes (Normalverteilungen etc.). Jede diskrete Verteilung hat allerdings auch eine Dichte bezüglich des Zählmaßes

$$\mu = \sum_{n\geq 1} \delta_{x_n}$$

mit dem Dirac-Maß $\delta_{x_n}(A) := \mathbb{1}_A(x_n)$ und geeignet gewählten (x_n).

Satz 1.51 (Radon-Nikodým)**.** *Es sei μ σ-finites Maß und ν ein signiertes Maß, so dass $\nu \ll \mu$. Dann hat ν eine (μ-f.s.) eindeutige Dichte $f : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ bzgl. μ.*

Beweis. Die Eindeutigkeit folgt aus (f, g quasi-i.b., μ σ-finit) Lemma 1.48. Mit der Jordan-Zerlegung ist $\nu \ll \mu \Leftrightarrow \nu^+ \ll \mu$ und $\nu^- \ll \mu$, so dass wir ohne Beschränkung der Allgemeinheit annehmen können, dass ν ein Maß ist.

(1) μ, ν endlich $\Rightarrow$
Wir betrachten $\tau = \mu + \nu$. Nach Lemma 1.46 existieren h, g so dass $d\mu = g d\tau$ und $d\nu = h d\tau$. Wir setzen $N = \{g = 0\}$, so dass $\mu(N) = 0$, also auch $\nu(N) = 0$, da $\nu \ll \mu$.

Definiere

$$f(x) := \frac{h(x)}{g(x)} \mathbb{1}_{\{x \in N\}}$$

$$\Rightarrow \qquad \begin{aligned} \nu(A) = \nu(A \cap N^c) &= \int_{A\cap N^c} h d\tau \\ &= \int_{A\cap N^c} f d\mu = \int_A f d\mu \end{aligned}$$

(2) μ endlich $\Rightarrow$ Setze

$$\alpha = \sup\{\mu(B) : B \in \mathscr{F}, \nu(B) < \infty\} \qquad (< \infty).$$

Dann gibt es $B_1 \subseteq B_2 \subseteq \cdots \subseteq \mathscr{F}$ mit $\nu(B_n) < \infty$, $\mu(B_n) \to \alpha$, also $\mu(\bigcup_{n\geq 1} B_n) = \alpha$.

Für $A \in \mathscr{F}$, $A \subseteq (\bigcup B_n)^c$, $\nu(A) < \infty$ gilt

$$\alpha + \mu(A) = \lim_{n\to\infty} \mu(B_n \cup A) \leq \alpha,$$

also $\mu(A) = 0$ und somit $\nu(A) = 0$.

Wir erhalten $A \subseteq (\bigcup B_n)^c \Rightarrow \mu(A) = \nu(A) = 0$ oder $\mu(A) > 0$, $\nu(A) = \infty$.

Es ist $E_n := B_n \setminus B_{n-1}$, $E_1 = B_1$, $E := \bigcup E_n$ eine Folge p. d. Mengen mit $\nu(E_n) < \infty \overset{(1)}{\Rightarrow} \exists f_n$ mit $d(\mathbb{1}_{E_n}\nu) = f_n d\mu$

$$\begin{aligned}\nu &= \Big(\sum_{n\geq 1} \mathbb{1}_{E_n} + \mathbb{1}_{E^c}\Big) \cdot \nu \\ &= \sum_{n\geq 1} f_n d\mu + \infty \cdot \mathbb{1}_{E_c} d\mu = \Big(\sum_{n\geq 1} f_n + \infty \cdot \mathbb{1}_{E^c}\Big) d\mu\end{aligned}$$

(3) μ σ-finit
$\Rightarrow$ Es existieren p. d. (Ω_n), $\mu(\Omega_n) < \infty$, so dass $\Omega = \sum_{n\geq 1} \Omega_n$.
Wir definieren $d\mu_n = \mathbb{1}_{\Omega_n} d\mu$, $\nu_n = \mathbb{1}_{\Omega_n} d\nu$
und wenden (2) an $\Rightarrow d\nu_n = f_n d\mu_n$ und $\sum f_n \mathbb{1}_{A_n}$ ist Dichte von ν bzgl. μ. □

Beispiel 1.52. Auf σ-Finitheit kann nicht verzichtet werden: $\Omega \neq \varnothing$, $\mathscr{F} = \{\varnothing, \Omega\}$, $\mu(\varnothing) = 0$, $\mu(\Omega) = \infty$, $\nu(\varnothing) = 0$, $\nu(\Omega) = \mathbb{1}$, so ist $\nu \ll \mu$, aber es gibt keine Dichte $(c \cdot \infty \neq \mathbb{1})$.

Satz 1.53 (Lebesgue-Zerlegung)**.** *Ist μ σ-finites Maß und ν σ-finites, signiertes Maß, so gibt es genau eine Zerlegung*

$$\nu = \varrho + \sigma$$

mit signierten Maßen ϱ, ν, so dass $\varrho \ll \mu$, $\sigma \perp \mu$.

Beweis. Wieder genügt es nach Satz 1.43, Maße zu betrachten.

Wir setzen $\tau = \mu + \nu \underset{\text{Satz 1.51}}{\Rightarrow} \exists\, g$ so dass $d\mu = g d\tau$.

Definiere $N = \{g = 0\}$ und

$$\varrho(A) = \nu(A \cap N^c), \quad \sigma(A) = \nu(A \cap N), \qquad A \in \mathscr{F}.$$ □

Eindeutigkeit: ÜA

1.6 Produkträume

Die grundlegende Idee dieses Kapitels ist es auf dem Raum $\Omega_1 \times \Omega_2$ ein Maß einzuführen - zunächst werden wir zeigen, dass Mengen vom Typ $A_1 \times A_2$ auf diesem Raum eine σ-Algebra erzeugen und sich darauf ein Maß konstruieren lässt, was die Masse $\mu(A_1) \cdot \mu(A_2)$ vergibt.

Für uns interessant werden polnische Räume sein, da auf ihnen die zentralen Grenzwertsätze der Stochastik gelten werden.[3]

Definition 1.54. Ein topologischer Raum $(\Omega, \mathscr{O})$ heißt **polnisch**, falls

(i) es eine abzählbare Menge $\Omega' \subseteq \Omega$ gibt, so dass $\overline{\Omega}' = \Omega$,

(ii) er vollständig metrisierbar ist.[4]

Beispiel 1.55. (i) $(\mathscr{C}[0,1], \sup)$ ist polnisch,

(ii) $(\mathscr{C}[0,\infty), d)$ mit

$$d(f,g) = \sum_{k \geq 1} \frac{1}{2^k} \frac{d_k(f,g)}{1 + d_k(f,g)} \qquad \text{mit } d_k = \sup_{[0,k]} |f - g|$$

ist polnisch (Topologie der gleichmäßigen Konvergenz auf Kompakta) und

(iii) jeder **kompakte** metrische Raum ist polnisch.

Für eine Familie $(\Omega_i)_{i \in I}$ von Mengen heißt

$$\bigtimes_{i \in I} \Omega_i := \{(\omega_i)_{i \in I} : \omega_i \in \Omega_i\}$$

Produktraum der $(\Omega_i)_{i \in J}$. Für $H \subseteq J \subseteq I$ definieren wir die **Projektionen**

$$\pi_H^J : \bigtimes_{j \in J} \Omega_j \to \bigtimes_{h \in H} \Omega_h \qquad \text{durch}$$

$$\pi_H^J((\omega_j)_{j \in I}) = (\omega_h)_{h \in H}.$$

Außerdem schreiben wir $\pi_H^I = \pi_H$ und $\pi_i = \pi_{\{i\}}$, $i \in I$.

[3]Für einen topologischen Raum $(\Omega, \mathscr{O})$ mit $A \subseteq \Omega$ setzen wir

$$A^o := \bigcup \{O \subseteq A : O \in \mathscr{O}\} \qquad \text{(Innere von } A\text{)}$$

$$\overline{A} := \bigcap \{F \supseteq A : F^c \in \mathscr{O}\} \qquad \text{(Abschluss von } A\text{)}$$

Definition 1.56. Ist $(\Omega_i, \mathscr{O}_i)_{i\in I}$ eine Familie von topologischen Räumen, so heißt die von
$$\Big\{ \bigtimes_{j\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega_i : J \subseteq I, J \text{ endlich}\Big\}$$
erzeugte Topologie *Produkttopologie* auf $\Omega = \bigtimes_{i\in I} \Omega_i$.

Bezüglich der Produkttopologie $\mathscr{O}$ sind alle Projektionen π_i, $i \in J$ stetig:
$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{j\in I\setminus\{i\}} \Omega_j \in \Omega, \qquad A_i \in \mathscr{O}_i.$$

Besonders schöne Eigenschaften erhält man unter Abzählbarkeit.

Satz 1.57. *Sei $(\Omega_i, \mathscr{O}_i)_{i\in\mathbb{N}}$ eine Familie polnischer Räume. Dann ist der Produktraum versehen mit der Produkttopologie wieder polnisch.*

Beweis. Da Ω_i separabel ist, gibt es abzählbare viele $\Omega_i' = \{\omega_1, \omega_2, \dots\} \subseteq \Omega_i$, so dass $\overline{\Omega_i'} = \Omega_i$ für alle $i \in \mathbb{N}$. Mit d_i bezeichnen wir die vollständige Metrik, die $\mathscr{O}_i$ erzeugt. Für $\omega, \omega' \in \Omega = \bigtimes_{i\in\mathbb{N}} \Omega_i$ wird durch
$$d(\omega, \omega') = \sum_{i\geq 1} \frac{1}{2^i}(d_i(\omega_i, \omega_i') \wedge 1)$$
eine vollständige Metrik auf Ω definiert, die die Produkt-Topologie $\mathscr{O}$ erzeugt.
Für $\omega' \in \bigtimes_{i\in\mathbb{N}} \Omega_i'$ ist
$$B_{\omega'} := \Big\{\omega \in \bigtimes_{i\in\mathbb{N}} \Omega_i' : \omega_i \neq \omega_i' \text{ für endlich viele } i \in \mathbb{N}\Big\}$$
abzählbar und dicht in Ω und die Behauptung folgt. □

Definition 1.58. Für eine Familie $(\Omega_i, \mathscr{F}_i)_{i\in I}$ von Maßräumen heißt die σ-Algebra
$$\bigotimes_{i\in I} \mathscr{F}_i := \sigma\Big(\Big\{ \bigtimes_{j\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega_i : J \subseteq I, I \text{ endlich}, A_i \in \mathscr{F}_i\Big\}\Big)$$
die Produkt-σ-Algebra auf $\Omega = \bigtimes_{i\in J} \Omega_i$. Ist $\mathscr{F}_i = \mathscr{F}_j$, so schreiben wir auch $\mathscr{F}^I = \bigotimes_{i\in I} \mathscr{F}$.

Wieder sind die Projektionen verträglich mit der Konstruktion: Da

$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{j \in I \setminus \{i\}} \Omega_j \in \bigotimes_{i \in I} \mathscr{F}_i,$$

ist π_i messbar bzgl. $\otimes_{i \in I} \mathscr{F}_i$. Außerdem ist

$$\bigotimes_{i \in I} \mathscr{F}_i = \sigma\Big(\Big\{A_i \times \bigtimes_{j \in J \setminus \{i\}} \Omega_j : A_i \in \mathscr{F}_i, i \in I\Big\}\Big).$$

Unter geeigneten Voraussetzungen ist die Borel-σ-Algebra der Produkttopologie gleich der Produkt-σ-Algebren der einzelnen Borel-σ-Algebren. Abzählbarkeit ist eine solche (vgl. Elstrodt, Kapitel III.5).

Lemma 1.59. *Ist I abzählbar und $(\Omega_i, \mathscr{O}_i)$ polnisch für jedes $i \in I$, so gilt*

$$\mathscr{B}(\Omega) = \bigotimes_{i \in I} \mathscr{B}(\Omega_i).$$

Insbesondere gilt $\mathscr{B}(\mathbb{R}^d) = \bigotimes_{i=1}^d \mathscr{B}(\mathbb{R}^d)$.

Beweis: Wir nutzen folgendes, einfaches topologisches Resultat: Ist $(\Omega, \mathscr{O})$ separabel, so hat $\mathscr{O}$ eine abzählbare Basis $\mathscr{B}$, d.h.

$$\mathscr{O} = \Big\{\bigcup_{i \in I} B_i : B_i \in \mathscr{B}, i \in I, I \text{ beliebig}\Big\}$$

Da alle $(\Omega_i, \mathscr{O}_i), i \in I$, separabel sind, haben sie jeweils abzählbare Basen $\mathscr{B}_i$. Dann ist

$$\mathscr{B} := \Big\{ \bigtimes_{j \in J} A_j \times \bigtimes_{i \in I \setminus J} \Omega_i : J \subseteq I, J \text{ endlich}, A_i \in \mathscr{B}_i\Big\}$$

eine abzählbare Basis von $(\Omega, \mathscr{O})$. Dann ist $\sigma(\mathscr{B}) = \mathscr{B}(\Omega)$ (siehe das folgende Lemma).

Außerdem ist $\mathscr{B} \subseteq \bigotimes_{i \in I} \mathscr{B}(\Omega_i)$, also $\mathscr{B}(\Omega) \subseteq \bigotimes_{i \in I} \mathscr{B}(\Omega_i)$.

Umgekehrt ist für $A_i \in \mathscr{F}_i$

$$\begin{aligned} A_i \times \bigtimes_{j \in I \setminus \{i\}} \Omega_j &\in \sigma\Big(\Big\{A_i \times \bigtimes_{j \in I \setminus \{i\}} \Omega_i : A_i \in \mathscr{O}_i\Big\}\Big) \\ &\subseteq \mathscr{B}(\Omega). \end{aligned}$$

□

Man erhält leicht folgendes Resultat über erzeugende Halbringe:

Lemma 1.60. *(i) Sei I endlich, $\mathscr{H}_i$ Halbringe mit $\sigma(\mathscr{H}_i) = \mathscr{F}_i$. Dann ist*

$$\mathscr{H} := \Big\{ \bigtimes_{i \in I} A_i : A_i \in \mathscr{H}_i, i \in I \Big\}$$

ein Halbring mit $\sigma(\mathscr{H}) = \bigotimes_{i \in I} \mathscr{F}_i$.

(ii) Sei I beliebig, $\mathscr{H}_i$ durchschnittstabil und $\sigma(\mathscr{H}_i) = \mathscr{F}_i$. Dann ist

$$\mathscr{H} := \Big\{ \bigtimes_{j \in J} A_j \times \bigtimes_{i \in I \setminus J} \Omega_i : J \subseteq I, J \text{ endlich}, A_j \in \mathscr{H}_j, j \in J \Big\}$$

durchschnittstabil mit $\sigma(\mathscr{H}) = \bigotimes_{i \in I} \mathscr{F}_i$.

1.7 Der Satz von Fubini

Ein wichtiges Hilfsmittel für Maße auf Produkträumen sind **Übergangskerne**, falls

Definition 1.61. Seien $(\Omega_1, \mathscr{F}_1)$ und $(\Omega_2, \mathscr{F}_2)$ Messräume.

Eine Abbildung $\kappa : \Omega_1 \times \mathscr{F}_2 \to \mathbb{R}_{\geq 0}$ heißt *Übergangskern*, falls

(i) für alle $\omega_1 \in \Omega_1$ ist $\kappa(\omega_1, \cdot)$ ein Maß auf $(\Omega_2, \mathscr{F}_2)$

(ii) für alle $A_2 \in \mathscr{F}_2$ ist $\kappa(\cdot, A_2)$ $\mathscr{F}_1$-messbar.

Ein Übergangskern heißt *σ-finit*, falls es $A_1, A_2, \ldots \uparrow \Omega_2$ gibt mit $\sup_{\Omega_1} \kappa(\omega_1, A_n) < \infty$ für alle $n = 1, 2, \ldots$

Er heißt *stochastischer Kern* oder *Markovscher Kern*, falls

$$\kappa(\omega_1, \Omega_2) = 1 \qquad \text{für alle } \omega_1 \in \Omega_1.$$

Beispiel 1.62 (Markov-Kette)**.**
Sei $\Omega = \{\omega_1, \ldots, \omega_n\}$ mit Wahrscheinlichkeiten $(p_{ij})_{1 \leq i,j \leq n}$, so dass mit $\sum_{j=1}^n p_{ij} = 1 \forall i$. Dann ist

$$\kappa(\omega_i, \cdot) := \sum_{j=1}^n p_{ij} \delta_{\omega_j}$$

ein stochastischer Kern von $(\Omega, \mathscr{P}(\Omega))$ nach $(\Omega, \mathscr{P}(\Omega))$. Die (p_{ij}) bilden die Übergangswahrscheinlichkeiten einer homogenen **Markov-Kette**.

Lemma 1.63 (Messbarkeit integrierbarer Schritte)**.** *Sei κ ein σ-finiter Übergangskern und $f : \Omega_1 \times \Omega_2 \to \mathbb{R}_{\geq 0}$ $\mathscr{F}_1 \otimes \mathscr{F}_2$-messbar. Dann ist*

$$\omega_1 \mapsto \int f(\omega_1, \omega_2) \kappa(\omega_1, d\omega_2)$$

$\mathscr{F}_1$-messbar.

Beweis (Skizze). Zunächst nehmen wir $\kappa(\omega_1, \Omega_2) < \infty$ an. (Dann Erweiterung mittels $A_1, A_2 \uparrow \Omega_2$.) Wir betrachten

$$\mathscr{D} = \left\{ A \in \mathscr{F}_1 \otimes \mathscr{F}_2 : \omega_1 \mapsto \int \mathbb{1}_A(\omega_2) \kappa(\omega_1, d\omega_2) \text{ ist } \mathscr{F}_1\text{-messbar} \right\}$$

Dies ist ein durchschnittstabiles Dynkin-System mit $\mathscr{H} = \{A_1 \times A_2, A_i \in \mathscr{F}_i\} \subseteq \mathscr{D}$. Nun gilt $\sigma(\mathscr{H}) = \mathscr{F}_1 \otimes \mathscr{F}_2$, also $\sigma(\mathscr{H}) = \mathscr{F}_1 \otimes \mathscr{F}_2 \subseteq \mathscr{D} \subseteq \mathscr{F}_1 \otimes \mathscr{F}_2$ und die Aussage folgt für $f = \mathbb{1}_A$, $A \in \mathscr{F}_1 \otimes \mathscr{F}_2$.

Wir erhalten die Aussage für Treppenfunktionen und mit monotoner Konvergenz für alle messbaren $f \geq 0$. □

Theorem 1.64 (Ionescu-Tulcea)**.** Seien $(\Omega_i, \mathscr{F}_i)$, $i = 0, \ldots, n$ Messräume, μ ein σ-finites Maß auf $(\Omega_0, \mathscr{F}_0)$ und κ_i σ-finite Übergangskerne von $(\bigtimes_{j=0}^{i-1} \Omega_j, \bigotimes_{j=0}^{i-1} \mathscr{F}_j)$ nach $(\Omega_i, \mathscr{F}_i)$ für $i = 1, \ldots, n$.

Dann gibt es genau ein σ-finites Maß $\kappa = \mu \bigotimes_{i=1}^{n} \kappa_i$ auf $(\bigtimes_{j=0}^{n} \Omega_j, \bigotimes_{j=0}^{n} \mathscr{F}_j)$, so dass

$$\kappa(A_0 \times \cdots \times A_n) = \int_{A_0} \cdots \int_{A_n} \kappa_n(\omega_0, \ldots, \omega_{n-1}, d\omega_n) \cdots \kappa_1(\omega_0, d\omega_1)\mu(d\omega_0). \tag{1.65}$$

Beweis. Wir beweisen die Aussage für $n = 1$, der allgemeine Fall folgt dann per Induktion. Der Beweis ist eine typische Anwendung des Maßfortsetzungssatzes: Wir starten mit dem durchschnittstabilen Halbring

$$\mathscr{H} = \left\{ \bigtimes_{i=0}^{n} A_i : A_i \in \mathscr{F}_i \right\}$$

und betrachten die durch (1.65) definierte Mengenfunktion κ auf $\mathscr{H}$.

(i) *κ ist σ-finit:* Seien $(\Omega_n^i) \subseteq \mathscr{F}_i$ so dass $(\Omega_n^i) \uparrow \Omega^i$, $i = 0, 1$ und $\mu(\Omega_n^0) < \infty$ und $\sup_{\omega^0 \in \Omega^0} \kappa_1(\omega_0, \Omega_n^1) =: C_n < \infty$.

$$\Rightarrow \qquad \begin{aligned} \kappa(\Omega_n^0 \times \Omega_n^1) &= \int_{\Omega_n^0} \int_{\Omega_n^1} \kappa(\omega^0, d\omega^1)\mu(d\omega^0) \\ &\leq C_n \mu(\Omega_n^1) < \infty. \end{aligned}$$

Da $\Omega_n^0 \times \Omega_n^1 \uparrow \Omega_0 \times \Omega_1$ ist κ σ-finit.

(ii) *κ ist σ-additiv:* Für $A_1, A_2 \in \mathscr{H}$ p. d. mit $A = \sum_{i=1}^{\infty} A_i \in \mathscr{H}$ ist

$$\begin{aligned} \kappa(A) &= \int \int \mathbb{1}_A(\omega_0, \omega_1)\kappa_1(\omega_0, d\omega_1) \\ &= \int \sum_{n \geq 1} \int \mathbb{1}_{A_n}(\omega_0, \omega_1)\kappa_1(\omega_0, d\omega_1) \\ &\overset{(*)}{=} \sum_{n \geq 1} \int \int \mathbb{1}_{A_n}(\omega_0, \omega_1)\kappa_1(\omega_0, d\omega_1)\mu(d\omega_0) \end{aligned}$$

wobei wir für (*) monotone Konvergenz nutzten. Mit dem Maßfortsetzungssatz erhalten wir die Behauptung. □

Bemerkung 1.66. Mit etwas mehr Aufwand kann man die Aussage auch für $n = \infty$ beweisen, siehe Bogachev(1991).

Theorem 1.67 (Satz von Fubini)**.** Seien $(\Omega_i, \mathscr{F}_i)$, μ_0, κ_i und $\kappa = \mu_0 \bigotimes_{i=1}^n \kappa_i$ wie in Theorem 1.64 gegeben und $f : \bigtimes_{i=0}^n \Omega_0 \to \mathbb{R}_{\geq 0}$ messbar bzgl. $\bigotimes_{i=0}^n \mathscr{F}_i$. Dann gilt

$$\int f d\kappa = \int \Big(\cdots \Big(\int f(\omega_0, \ldots, \omega_n) \kappa_n(\omega_0, \ldots, \omega_{n-1} d\omega_n) \Big) \cdots \kappa_1(\omega_0, d\omega_1) \Big) \mu(d\omega_0). \tag{1.68}$$

Die Aussage gilt auch für beliebig messbare $\mathscr{F}$ mit $\int |f| d\kappa < \infty$.

Beweis. Die Messbarkeit der einzelnen Integrale folgt nach Lemma 1.63.

Für alle $A \in \bigotimes_{i=0}^n \mathscr{F}_i$ betrachten wir $A \mapsto \int_A d\kappa =: I(A)$. Zunächst gilt für $A \in \left\{ \bigtimes_{i=0}^n A_i : A_i \in \mathscr{F}_i \right\}$, dass $I(A) = \kappa(A)$ und somit (1.68). Wegen der Linearität des Integrals und dem Satz der monotonen Konvergenz folgt (1.68) für alle nicht-negativen, messbaren Funktionen.

$\square$

Die aus der Analysis bekannte Aussage für Produktmaße $\mu_0 \otimes \cdots \otimes \mu_n$ erhält man als einfachen Spezialfall!

Beispiel 1.69 (Mehrdimensionales Lebesgue-Maß)**.** Mit $\lambda_d = \bigotimes_{i=1}^d \lambda$ bezeichnen wir das Lebesgue-Maß auf dem $\mathbb{R}^d$

$$f(x,y) = \frac{xy}{(x^2+y^2)^2} \Rightarrow \int f(x,y) dy = 0 \quad \forall x$$

$$\Rightarrow \qquad \iint f(x,y) \lambda(dy) \lambda(dx) = \iint f(x,y) \lambda(dx) \lambda(dy) = 0.$$

Allerdings ist $|f|$ nicht bzgl. λ_2 integrierbar!

Aus Gleichheit und Endlichkeit der Mehrfachintegrale kann man also nicht auf die Integrierbarkeit des Integrand schließen.

Betrachten wir zwei unabhängige Zufallsvariable X und Y, so heißt die Verteilung von $X + Y$ die **Faltung**. Die führt zu folgendem Konzept.

Definition 1.70. Seien $\mu_1, \ldots, \mu_n$ σ-finite Maße auf $\mathscr{B}(\mathbb{R})$, $\mu = \bigotimes_{i=1}^n \mu_i$ deren Produktmaß und $S(x) = x_1 + \cdots + x_n$, $x \in \mathbb{R}^n$. Dann heißt

$$S_*\mu := \mu_1 * \cdots * \mu_n$$

die *Faltung* der Maße $\mu_1, \ldots, \mu_n$.

Übungsaufgabe 1.71. Haben die Maße μ_i Dichten f_i bzgl. λ_i, $i = 1, 2$ und setzt man

$$f_{\mu*\nu}(t) := \int f_\mu(s) f_\nu(t-s)\lambda(ds),$$

so gilt

$$\mu * \nu = f_{\mu*\nu} \cdot \lambda. \qquad \text{(Fubini)}$$

Betrachte $\quad P(X + Y \leq x) = P(X \leq x - y)$!

Übungsaufgabe 1.72. Die Faltung zweier Normalverteilungen ist wieder normalverteilt.

1.8 Der Existenzsatz von Kolmogorov

Bisher haben wir endliche bzw. abzählbare Produkträume betrachtet, was für unsere Anwendungen nicht ausreichen wird. Der Satz von Kolmogorov erlaubt die Erweiterungen auf größere Räume, kommt allerdings nicht ohne die Voraussetzung aus, dass diese *polnisch* sind.

Als Beispiele können wir den unendlichen Würfelwurf betrachten und die Ergebnisse als eine Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ auffassen. Oder wir betrachten einen Zählprozess $(N_t)_{t\geq 0}$ (das ist ein Prozess, der in Null startet und dann jeweils an einer zufälligen Zeit um 1 springt).

Beispiel 1.73. (i) $\mathbb{R}^d = \overline{\mathbb{Q}}^d$ ist separabel und vollständig

(ii) Sei $K \subseteq \mathbb{R}$ kompakt $\Rightarrow (C(K), \sup) = (C(K), \|\cdot\|_\infty)$ ist **polnisch** (Weierstraß'scher Approximationssatz – Polynome)

Wir erinnern daran, dass π_H^J die Projektion von J auf H bezeichnete.

Definition 1.74. Sei $(\Omega, \mathscr{F})$ ein Maßraum und I eine beliebige Indexmenge. Eine Familie von Wahrscheinlichkeitsmaße $P = (P_J : J \subseteq I, J \text{ endlich})$ heißt *projektive Familie*, falls P_J Wahrscheinlichkeitsmaß auf $(\Omega^J, \mathscr{F}^J)$ und

$$P_H = \pi^J_{H*} P_J$$

für alle $H \subseteq J \subseteq I$, J endlich.

Beispiel 1.75. Betrachten wir etwa den Poisson-Prozess $(N_t)_{t\geq 0}$: In diesem Fall gilt $P(N_t \in \{0,1,2,\dots\} = 1)$, N is wachsend und

$$P(N_{t_j} = n_j,\ j \leq k) = \prod_{j=1}^{k} e^{-\lambda(t_j - t_{j-1})} \frac{\lambda(t_j - t_{j-1})^{n_j - n_{j-1}}}{(n_j - n_{j-1})!}.$$

Man erhält man eine projektive Familie durch die *endlich-dimensionalen Randverteilungen*

$$\mu_{t_1,\dots,t_k}(H) = P\big((N_{t_1},\dots,N_{t_k}) \in H\big), \quad H \in \mathbb{R}^k.$$

Die bestimmen allerdings die Verteilung des Prozesses noch nicht vollständig! Konstruieren Sie einen Prozess N' mit den gleichen Randverteilungen als Übungsaufgabe.

Definition 1.76. Existiert für eine solche projektive Familie P ein Wahrscheinlichkeitsmaß P_I auf $\mathscr{F}^I = \otimes_{i\in I}\mathscr{F}_i$ mit $P_J = \pi_J * P_I$ für alle $J \subseteq I$, J endlich, so heißt P_I *projektiver Limes* der Familie P. Wir schreiben

$$P_I = \varprojlim_{J \Subset I} P_J$$

mit der Konvention $J \Subset I = J \subseteq J$ und J endlich.

2. Wahrscheinlichkeitstheorie

In diesem Abschnitt beginnen wir nun mit einer systematischen Entwicklung von Zufallsvariablen und Konvergenzen von Zufallsvariablen mit den dazugehörigen Hilfsmitteln. Dabei werden wir natürlich massiv von der im ersten Kapitel entwickelten Maßtheorie profitieren. Zentrale Resultate in diesem Abschnitt sind ein allgemeines starkes Gesetz der großen Zahl und ein allgemeiner zentraler Grenzwertsatz. Zunächst beginnen wir mit Grundlagen, die einige Resultate aus der Stochastik I wiederholen und in etwas allgemeinerer Form formulieren.

2.1 Grundlagen

Wir betrachten einen Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, P)$ und einen Bildraum $(\Omega', \mathscr{F}')$. Eine Abbildung

$$X : \Omega \to \Omega'$$

heißt **Zufallsvariable**, falls sie $\mathscr{F} - \mathscr{F}'$-messbar ist. Wir schreiben

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\},$$
$$\sigma(X) = \{X^{-1}(B) : B \in \mathscr{F}'\}$$

für das *Urbild* von X und die von X erzeugte σ-Algebra. Das Bildmaß

$$X_*P(B) = P(X^{-1}(B)) : \mathscr{F}' \to [0, 1]$$

heißt **Verteilung** von X. Haben X und Y die gleiche Verteilung (equality in law), so schreiben wir

$$X \stackrel{\mathscr{L}}{=} Y.$$

Wir sagen, X hat die **Dichte** f bzgl. des Maßes ν, falls $X_*P = f \cdot \nu$, also

$$P(X \in A) = \int_A f d\nu, \qquad A \in \mathscr{F}'.$$

Für eine messbare Zufallsvariable $X \geq 0$ definieren wir

$$E[X] = \int X dP$$

und im Falle der Quasi-integrierbarkeit den Erwartungswert als Summe der Integrale der Positiv- und Negativ-Teile.

Monotonie und Linearität des Integrals liefern für Zufallsvariable X, Y die Regeln

$$E[aX + bY] = aE[X] + bE[Y],$$
$$0 \leq X \leq Y \Rightarrow E[X] \leq E[Y].$$

Eine Zufallsvariable ist meßbar bezüglich der von X erzeugten Filtration, wenn sie als meßbare Funktion von X geschrieben werden kann. Das die Umkehrung überraschenderweise auch gilt zeigt folgendes Lemma.

Lemma 2.1. *Sei $X : \Omega \to \Omega'$ eine Zufallsvariable. $Z : \Omega \to \overline{\mathbb{R}}$ ist genau dann $\sigma(X)$-messbar, falls*

$$Z = f(X)$$

mit $f : \Omega' \to \overline{\mathbb{R}}$ $\mathscr{F}' - \mathscr{B}(\overline{\mathbb{R}})$-messbar.

Beweis. „$\Leftarrow$“ klar.

„$\Rightarrow$“ Ohne Beschränkung der Allgemeinheit sei $Z \geq 0$. (Ansonsten betrachte $Z = Z^+ - Z^-$.) Der Beweis benutzt die Approximation von Z durch Elementarfunktionen und geht in drei Schritten:

(i) Sei $Z = \mathbb{1}_A$, $A \in \sigma(X)$. Dann existiert $A' \in \mathscr{F}'$ mit $X^{-1}(A') = A$, also

$$Z = \mathbb{1}_A = \mathbb{1}_{X^{-1}(A')} = \mathbb{1}_{A'} \circ X \Rightarrow f = \mathbb{1}_{A'}.$$

(ii) Mit Linearität erhalten wir die Aussage für einfache $Z = \sum c_i \mathbb{1}_{A_i}$.

(iii) Ist $Z \geq 0$ messbar, so gibt es einfache $(Z^n)_{n\geq 1} \uparrow Z$ und $Z_n = f_n \circ X \;\Rightarrow$

$$(\sup_n f_n) \circ X = \sup_n f_n \circ X = \sup Z_n = Z$$

und mit $f = \sup f_n$ sind wir fertig, denn nach Lemma 1.26 ist $\sup f_n$ meßbar. □

Als nächstes Resultat beweisen wir eine Reihe von wichtigen Ungleichungen, die wir im Folgenden ständig benutzen werden. Vorab definieren wir noch $\| \cdot \|_p = E[|X|^p]^{1/p}$. Die Minkowski-Ungleichung wird zeigen, dass dies für $p \geq 1$ auch eine Norm auf dem L^p ist.

Satz 2.2. *Seien X, Y reellwertige Zufallsvariablen. Dann gilt:*

(i) *Für alle monoton wachsenden $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ und $\varepsilon > 0$ so dass $f(\varepsilon) > 0$*

$$P(|X| \geq \varepsilon) \leq \frac{E[f(|X|)]}{f(\varepsilon)}, \qquad \text{Markov-Ungleichung}$$

(ii) *für $E[X^2] < \infty$, dass*

$$P(|X - E[X]| > \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}, \qquad \text{Tschebyscheff-Ungleichung}$$

(iii) *für $0 < p, q, r \leq \infty$ mit $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, dass*

$$\|XY\|_r \leq \|X\|_p \|Y\|_q \qquad \text{Hölder-Ungleichung}$$

für $p = q = 2$ und $r = 1$ erhalten wir die wichtige Cauchy-Schwarz-Ungleichung,

(iv) *für $1 \leq p \leq \infty$, dass*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p, \qquad \text{Minkowski-Ungleichung}$$

(v) *für $0 < p \leq q$, und $X \in \mathscr{L}^q$, dass*

$$\|X\|_p \leq \|X\|_q.$$

Bevor wir diesen Satz beweisen schließen wir noch die wichtige Jensen-Ungleichung an:

Satz 2.3 (Jensensche Ungleichung)**.** *Sei $g : \mathbb{R} \to \mathbb{R}$ konvex und $x \in \mathscr{L}^1$. Dann gilt*

$$E[g(X)] \geq g(E[X]).$$

Gleichheit gilt genau dann, wenn für jede Gerade $a + bx$ tangential zu g an $x = E[X]$ gilt, dass $P(g(X) = a + bX) = 1$.

Beweis zu Satz 2.2:

(i) Da $f(\varepsilon) > 0$ und $f(x) \geq 0$ ist, folgt für alle $x \in \mathbb{R}_{\geq 0}$, dass

$$\begin{aligned} \mathbb{1}_{\{|x|\geq\varepsilon\}} &\leq \mathbb{1}_{\{f(|x|)\geq f(\varepsilon)\}} = \mathbb{1}_{\left\{\frac{f(|x|)}{f(\varepsilon)}\geq 1\right\}} \\ &\leq \frac{f(|x|)}{f(\varepsilon)}\mathbb{1}_{\left\{\frac{f(|x|)}{f(\varepsilon)}\geq \mathbb{1}\right\}} \leq \frac{f(|x|)}{f(\varepsilon)}. \end{aligned}$$

Durch Monotonie des Erwartungswertes folgt die Behauptung.

(ii) Wir verwenden (i): Setzen wir $Y = |X - E[X]|$ und wählen $f(x) = x^2$ auf $\mathbb{R}_{\geq 0}$ (diese Funktion ist monoton), so gibt (i) das Resultat.

(iii) Zunächst ist e^x konvex und damit für $x, y \geq 0$

$$\begin{aligned} (xy)^r = \exp(r \log x + r \log y) &= \exp\left(\frac{r}{p}\log x^p + \frac{r}{q}\log x^q\right) \\ &\leq \frac{r}{p}x^p + \frac{r}{q}x^p. \qquad \left(\text{da } \tfrac{r}{p} + \tfrac{r}{q} = 1\right) \end{aligned}$$

Das wenden wir an auf $X' = \frac{|X|}{\|X\|_p}$, $Y' = \frac{|Y|}{\|Y\|_p}$ und erhalten

$$E[(X'Y')^r] \leq \frac{r}{p}\frac{E[|X|^p]}{\|X\|_p} + \frac{r}{q}\frac{E[|Y|^q]}{\|Y\|_q} = 1,$$

also

$$1 \geq E[(X'Y')^r = \frac{E[|X'Y'|^r]}{\|X\|_p^r\|Y\|_q^r},$$

und die Behauptung folgt.

(iv) Dieser Beweis ist trickreicher. Zunächst ist

$$|x + y|^p = |x + y| \cdot |x + y|^{p-1} \leq |x| \cdot |x + y|^{p-1} + |y| \cdot |x + y|^{p-1}.$$

Mit (iii) folgt, dass

$$E[|x| \cdot |x + y|^{p-1}] \leq \|x\|_p \cdot \||x + y|^{p-1}\|_q \quad \text{mit } \tfrac{1}{p} + \tfrac{1}{q} = 1.$$

Dabei ist

$$\||x + y|^{p-1}\|_q = E\left[|x + y|^{(p-1)\cdot q}\right]^{1/q} = \|x + y\|_p^{p/q} = \|x + y\|_p^{p-1},$$

da $(p - 1) \cdot q = p$ und $\frac{p}{q} = p - 1$.

Das liefert $E[|x + y|^p] = \|x + y\|_p^p \leq \big(\|x\|_p + \|y\|_p\big) \cdot \|x + y\|_p^{p-1}$

(v) Zunächst ist $x \mapsto x^{p/q}$ konkav auf $\mathbb{R}_{\geq 0}$, also folgt mit Satz 2.3

$$E[|X|^p] = E[(|X|^q)^{p/q}] \leq (E[|X|^q])^{p/q}$$

und wir sind fertig. □

Beweis zu Satz 2.3: Für eine konvexe Funktion f gibt es an jedem Punkt x eine Tangente $y \mapsto a + by$, so dass

$$f(y) \geq a + by \quad \forall y.$$

Hierbei kann die Gerade auch als $f(x) + b(y - x)$ parametrisiert werden. Wir betrachten diese unterstützende Gerade am Punkt $E[X]$.

Dann ist

$$f(y) \geq f(E[X]) + b \cdot (y - E[X]).$$

Anwendung des Erwarungswertes liefert

$$E[f(X)] \geq f(E[X]) + b \cdot (E[X] - E[X]) = f(E[X]). \qquad \square$$

Beweis. Für eine konvexe Funktion gibt es zu jedem Punkt x eine Gerade g durch x, so dass $f \geq g$. Wählen wir $x = E[X]$, so bedeutet das, es gibt $a \in \mathbb{R}$ mit

$$f(x') \geq f(E[X]) + a(x' - E[X]).$$

Wir erhalten $E[f(x)] \geq f(E[X]) + 0$ und die Behauptung folgt. $\square$

2.2 Charakteristische Funktionen

Als ein wichtiges Hilfsmittel zur Charakterisierung von Verteilungen erweisen sich Fourier- und Laplace-Transformierte.

Definition 2.4. Ist X eine d-dimensionale Zufallsvariable, so heißt $\varphi : \mathbb{R}^d \to \mathbb{C}$, gegeben durch

$$\varphi(t) = E[e^{i\langle t,x\rangle}]$$

charakteristische Funktion oder *Fourier-Transformierte* von X.

Die Funktion $\mathscr{L} : \mathbb{R}^d \to \mathbb{R}_{\geq 0} \cup \{\infty\}$, gegeben durch

$$\mathscr{L}(t) = E[e^{-\langle t,x\rangle}]$$

heißt (verallgemeinerte) *Laplace-Transformierte* von X.

Oft ist die genaue Bezeichnung Laplace/Fourier-Transformierte in der Literatur unterschiedlich, je nach Anwendungsgebiet. In diesem Skript habe ich mich für die obige Notation entschieden, weil sie in dieser Form für die Stochastik die einfachsten Ausdrücke liefert.

Satz 2.5. *Es gilt*

(i) $|\varphi(t)| \leq 1$ *für alle* $t \in \mathbb{R}^d$ *und* $\varphi(0) = 1$.

(ii) φ *ist gleichmäßig stetig,*

(iii) $\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at) \qquad a \in \mathbb{R}, b \in \mathbb{R}^q, t \in \mathbb{R}^d$

Beweis. (i) und (iii) sind klar.

Für (ii) verwenden wir, dass mit majorisierter Konvergenz folgt:

$$\left|E[e^{itX} \cdot (e^{ihX} - 1)]\right| \overset{(i)}{\leq} \left|E[(e^{ihX} - 1)]\right| \xrightarrow[h\to 0]{} 0,$$

also $\sup_{t\in\mathbb{R}} |\varphi(t+h) - \varphi(t)| \leq |E[(e^{ihX} - 1)]| \xrightarrow[h\to 0]{} 0$, so dass φ gleichmäßig stetig ist. $\square$

Beispiel 2.6. Folgende Beispiele sind wichtige Fourier-Transformierte:

(i) Ist $X \sim B(n,p)$, so erhalten wir

$$E[e^{itx}] = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-h} e^{itk} = \left(1 - p + pe^{it}\right)^n.$$

(ii) Ist $X \sim \text{Poiss}(\lambda)$, so erhalten wir

$$E[e^{itx}] = \sum_{k\geq 0} e^{-\lambda}\frac{\lambda^k}{k!}e^{itk} = e^{-\lambda}e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

(iii) Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so gilt zunächst, dass $E[e^{itx}] = E[e^{it\mu+it\sigma\xi}]$ mit $\xi \sim \mathcal{N}(0,1)$. Es bleibt die charakteristische Funktion für die Standardnormalverteilung zu berechnen:

$$E[e^{it\xi}] = \frac{1}{\sqrt{2\pi}}\int e^{itx}e^{-\frac{x^2}{2}}\,dx = e^{-\frac{t^2}{2}},$$

wobei $\varphi(t) = e^{it\mu-\frac{\sigma^2t^2}{2}}$ die Dichte der Standardnormalverteilung ist.

(iv) Ist schließlich $X \sim \text{Exp}(t)$, so erhalten wir

$$E[e^{-tx}] = \int_0^\infty e^{-tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda+t}.$$

(v) Als Übungsaufgaben könnte man leicht die Laplace-transformierte an Stelle der charakteristischen Funktion berechnen, oder etwa mit der *momentenerzeugenden Funktion* $E[e^{tX}]$ einige der ersten Momente.

Satz 2.7. *Sei X eine reellwertige Zufallsvariable.*

(i) Ist $X \in L^p$, so ist φ_X p-mal stetig differenzierbar und

$$\varphi_X^{(k)}(t) = E[(iX)^k e^{itx}], \qquad k = 0, \ldots, p$$

(ii) Für $X \in L^2$ gilt

$$\varphi(t) = 1 + itE[X] - \frac{t^2}{2}E[X^2] + \varepsilon(t)\cdot t^2$$

mit $\varepsilon(t) \underset{t\to 0}{\longrightarrow} 0$.

Beweis. (i) Da $X \in L^p$ existiert der Erwartungswert für $k \leq p$. Wir nutzen Induktion: $k = 0$ klar. Gelte die Behauptung für ein k. Dann ist

$$\left|\frac{d^{k+1}e^{itx}}{dt^{k+1}}\right| = \lim_{h\to 0}\left|\frac{(ix)^k e^{i(t+h)x} - (ix)^k e^{itx}}{h}\right| \leq \lim_{h\to 0}\left|x^k\frac{e^{ihx}-1}{h}\right|,$$

da $|e^{ix}| \leq 1$. Weiterhin gilt, dass

$$\begin{aligned}\lim_{h\to 0}\frac{e^{ihx}-1}{h} &= \lim_{h\to 0}\left|\frac{ihx + \frac{(ihx)^2}{2} + \cdots}{h}\right| \\ &\leq |x|\cdot\lim_{h\to 0}\left(1 + |hx| + \frac{|hx|^2}{2} + \cdots\right) = |x|,\end{aligned}$$

und somit gibt es für $\frac{d^{k+1}e^{itx}}{dt^{k+1}}$ eine integrierbare Majorante. Damit folgt

$$\varphi^{(kn)}(t) = E\Big[\frac{d}{dt}(iX)^k e^{itX}\Big] = E[(iX)^{k+1}e^{itX}].$$

Ebenso folgert man Stetigkeit der Ableitung mit majorisierter Konvergenz.

(ii) Die Taylor-Entwicklung mit Restglied liefert

$$e^{itX} = 1 + itX - \frac{t^2X^2}{2}(\cos\theta_1 tX + i\sin\theta_2 tX)$$

mit Zufallsvariablen θ_1, θ_2 und $|\theta_i| \leq 1$, $i = 1, 2$. Das ergibt

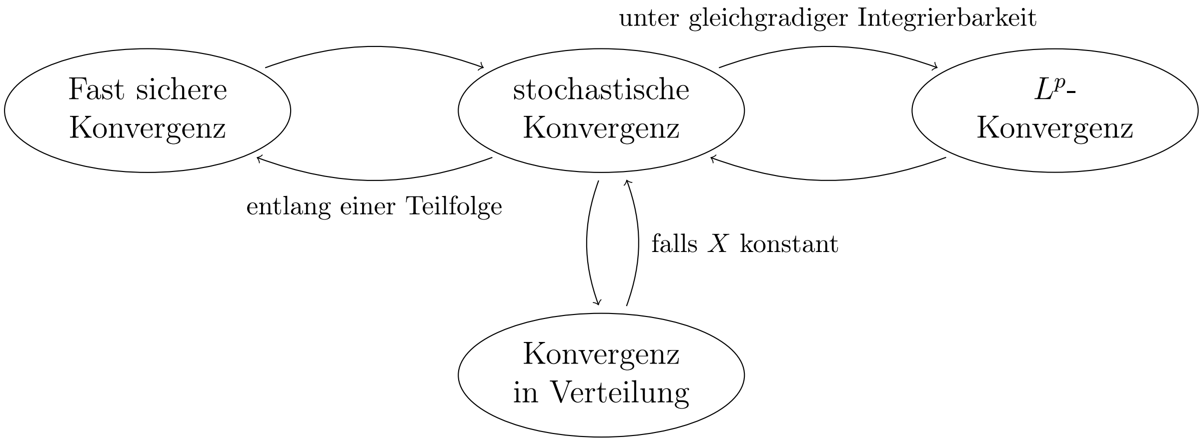
$$\varphi(t) = 1 + itE[X] - \frac{t^2}{2}E[X^2] + \varepsilon(t)\, t^2$$

mit $2\varepsilon(t) = E[X^2(1 - \cos\theta_1 tX - i\sin\theta_2 tX)] \xrightarrow{t\to 0} 0$ wegen majorisierter Konvergenz.

□

2.3 Stochastische Konvergenz

Wir werden nun verschiedene Konvergenzen kennenlernen.



Zusammenfassend ist die Konvergenz in Verteilung die schwächste Konvergenzform, sie wird von fast sicherer, L^p- und stochastischer Konvergenz impliziert. Fast sicherere und L^p-Konvergenz sind zwei recht unterschiedliche Konvergenzarten, für die jeweils unter Zusatzbedingungen Beziehungen aufgezeigt werden können. Eine kennen wir schon aus der Stochastik I: Der Übergang zu Teilfolgen. Ein neues Konzept, was wir in Kürze einführen werden, ersetzt das Kriterium majorisierte Konvergenz durch eine schwächere Bedingung, gleichgradige Integrierbarkeit, welche erlaubt aus fast sicherer Konvergenz L^p-Konvergenz zu schließen.

Definition 2.8. Seien $X, X_1, X_2, \dots$ Zufallsvariable in einem metrischen Raum (E, d).

(i) Gilt
$$\lim_{n\to\infty} d(X_n, X) = 0 \qquad P\text{-f.s.},$$
so konvergiert (X_n) *fast sicher* gegen X
$$(X_n \xrightarrow[n\to\infty]{\text{f.s.}} X).$$

(ii) Gilt für alle $\varepsilon > 0$, dass
$$\lim_{n\to\infty} P(d(X_n, X) > \varepsilon) = 0,$$
so konvergiert (X_n) *stochastisch* gegen X
$$(X_n \xrightarrow[n\to\infty]{P} X).$$

(iii) Sind $X, (X_n)$ reellwertig und gilt für $p > 0$, dass
$$\lim_{n\to\infty} E[|X_n - X|^p] = 0,$$
so *konvergiert* (X_n) *in* L^p (oder *im p-ten Mittel*) gegen X
$$(X_n \xrightarrow[n\to\infty]{L_p} X).$$

Beispiel 2.9 (Gegenbeispiele). (i) Stochastische Konvergenz impliziert nicht fast sichere Konvergenz: Sei
$$A_1 = \left[0, \frac{1}{2}\right], \qquad A_2 = \left[\frac{1}{2}, 1\right],$$
$$A_3 = \left[0, \frac{1}{4}\right], \dots, \quad A_6 = \left[\frac{3}{4}, 1\right], \dots$$
und $X_n = \mathbb{1}_{\{U \in A_n\}}$ mit $U \sim U(0,1)$.

$\Rightarrow$
$$\lim_{n\to\infty} P(|X_n| > \varepsilon) = \lim_{n\to\infty} P(U \in A_n) = 0,$$
aber $X_n \overset{\text{f.s.}}{\not\longrightarrow} 0$, denn für jedes $n \in \mathbb{N}$ gilbt es $n > n_k$ mit $X_n = 1$.

(ii) f.s. impliziert nicht L^1: Sei $U \sim U[0,1]$ und $B_n = [0, \frac{1}{n}]$ und $X_n = n \cdot \mathbb{1}_{\{U \in B_n\}}$. Dann $X_n \xrightarrow{\text{f.s.}} 0$, aber
$$E[X_n] = n \cdot P(U \in B_n) = 1.$$

Lemma 2.10. *Seien $X, Y, X_1, X_2, \ldots$ Zufallsvariable auf dem metrischen Raum (E, d). Gilt $X_n \xrightarrow{P} X, X_n \xrightarrow{P} Y$, so folgt*

$$X = Y \ P\text{-f.s.}$$

Beweis. Aus der stochastischen Konvergenz folgt für alle $\varepsilon > 0$, dass

$$P(d(X,Y) > 2\varepsilon) \leq P\Big(d(X_n, X) > \varepsilon \text{ oder } d(X_n, Y) > \varepsilon\Big) \underset{n\to P}{\to} 0,$$

also ist $P(d(X,Y) > 2\varepsilon) = 0$. Hieraus folgt

$$P(X \neq Y) = P\Big(\bigcup_{k=1}^{\infty} \Big\{d(X,Y) > \frac{1}{k}\Big\}\Big) \leq \sum_{k=1}^{\infty} P\Big(d(X,Y) > \frac{1}{k}\Big) = 0. \qquad \square$$

Lemma 2.11. *Seien $X, X_1, \ldots$ Zufallsvariable auf dem metrischen Raum (E, d). Dann gilt*

$$X_n \xrightarrow[n\to\infty]{P} X \Leftrightarrow E[d(X_n, X) \wedge 1] \xrightarrow[n\to\infty]{} 0.$$

Beweis. „$\Rightarrow$" Aus $X_n \xrightarrow{P} X$ folgt $\forall \varepsilon > 0$, dass

$$\lim_{n\to\infty} E[d(X_n, X) \wedge 1] \leq \lim_{n\to\infty} (\varepsilon + P(d(X_n, X) > \varepsilon)) = \varepsilon.$$

„$\Leftarrow$" Umgekehrt folgt mit $0 < \varepsilon \leq 1$ aus der Markov-Ungleichung, dass

$$P(d(X_n, X) > \varepsilon) \leq \frac{E[d(X_n, X) \wedge 1]}{\varepsilon} \underset{n\to\infty}{\to} 0. \qquad \square$$

Satz 2.12. *Es gelten folgende Aussagen:*

(i) Fast-sichere impliziert stochastische Konvergenz

(ii) Fast-sichere Konvergenz ist äquivalent zu

$$\lim_n P\Big(\bigcup_{k\geq n} \{|Z_k - Z| > \varepsilon\}\Big) = 0.$$

(iii) $\sum_n P(|Z_n - Z| > \varepsilon) < \infty \ \forall \varepsilon > 0 \quad \Rightarrow \quad Z_n \to Z$ *fast sicher.*

Das Kriterium (ii) nennt man auch *Cauchy-Kriterium* für die fast-sichere Konvergenz. Wir erhalten

Korollar 2.13. *Gilt $Z_n \xrightarrow{P} Z$, so existiert eine Teilfolge $Z_{n(j)}$, welche fast sicher gegen Z konvergiert.*

Beweis. Wähle $n(j)$ so dass $P\big(|Z_{n(j)} - Z| \geq \frac{1}{j}\big) \leq \frac{1}{j^2}$. Behauptung folgt aus (ii) des vorigen Satzes. □

Beweis von Satz 2.12. Teil (i)+(ii): Zunächst einmal ist

$$1 = P(\lim_n Z_n = Z) = P\Big(\bigcap_{\varepsilon>0}\bigcup_{n\geq 1}\bigcap_{k\geq n}\{|Z_k - Z| \leq \varepsilon\}\Big). \tag{2.14}$$

Die Mengen $\bigcap_{k\geq n}\{|Z_k - Z| \leq \varepsilon\}$ sind monoton wachsend in n und ε, was wir nun geschickt ausnutzen. Mit der σ-Stetigkeit von P folgt, dass (2.14) äquivalent dazu ist, dass

$$1 = P\Big(\bigcup_{n\geq 1}\bigcap_{k\geq n}\{|Z_k - Z| \leq \varepsilon\}\Big) \tag{2.15}$$

für alle[1] $\varepsilon > 0$. Eine zweite Anwendung der σ-Stetigkeit liefert, dass (2.14) äquivalent dazu ist, dass

$$\lim_n P\Big(\bigcap_{k\geq n}\{|Z_k - Z| \leq \varepsilon\}\Big) = 1 \ \ \forall \varepsilon > 0.$$

Hieraus folgt

$$0 = \lim_n P\Big(\bigcup_{k\geq n}\{|Z_k - Z| > \varepsilon\}\Big), \tag{2.16}$$

was die Behauptung (ii) zeigt. Weiterhin ist

$$\lim_n P\Big(\bigcup_{k\geq n}\{|Z_k - Z| > \varepsilon\}\Big) \geq \lim_n P(|Z_n - Z| > \varepsilon)$$

und somit folgt die gesuchte stochastische Konvergenz.

Teil (iii): Für ein festes $\varepsilon > 0$ folgt mit Borel-Cantelli (Das Resultat werden wir aber erst in Theorem 2.33 beweisen), dass

$$\sum_{n=1}^{\infty} P(|Z_n - Z| > \varepsilon) < 0 \quad \Rightarrow \quad P\Big(\bigcap_n \bigcup_{k\geq n}\{|Z_k - Z| > \varepsilon\}\Big) = 0$$

Das Komplement ist

$$P\Big(\bigcup_n \bigcap_{k\geq n}\{|Z_k - Z| \leq \varepsilon\}\Big) = 1. \tag{2.17}$$

Da ε beliebig gewählt war, liefert die Äquivalenz (2.15) nun die fast sichere Konvergenz. □

[1] In der Tat: Ist $1 = P(\cap_{\varepsilon>0} A_\varepsilon)$ und gilt $A_\varepsilon \supset A_{\varepsilon'}$ für $\varepsilon > \varepsilon'$, so folgt $1 = P(\cap_{n\geq 0} A_{n^{-1}}) = \lim_n P(A_{n^{-1}})$. Da $\lim_n P(A_{n^{-1}}) \leq P(A_{n'^{-1}})$ für alle $n' \geq 1$, folgt sogar, dass $1 = P(A_{n^{-1}})$ für alle $n \geq 1$ und schließlich $1 = P(A_\varepsilon)$ für alle $\varepsilon > 0$.

Für die Folgerung, dass fast sichere Konvergenz L^p-Konvergenz impliziert, benötigen wir eine integrierbare Majorante. Im Folgenden wollen wir diese Forderung abschwächen.

Definition 2.18. Eine Familie von reellen Zufallsvariablen $(X_i)_{i\in I}$ heißt *gleichgradig (uniform) integrierbar*, falls

$$\lim_{k\to\infty} \sup_{i\in I} E\left[|X_i| 1_{\{|X_i|>k\}}\right] = 0.$$

Die englische Übersetzung ist uniformly integrable.

Beispiel 2.19. (i) Integrierbare Majorante: Sei $Y \in L^1$ und $\sup |X_i| < Y$. Dann ist $(X_i)_{i\in I}$ gleichgradig integrierbar, denn nach majorisierter Konvergenz

$$\sup_{i\in I} E\left[|X_i| \mathbb{1}_{\{|X_i|>k\}}\right] \leq E\left[|Y| \mathbb{1}_{\{|Y|>k\}}\right] \underset{k\to\infty}{\longrightarrow} 0.$$

(ii) Jede endliche Familie integrierbarer Zufallsvariablen ist gleichgradig integrierbar, denn $\sup_{1\leq i\leq n} |X_i| \in L^1$ und

$$\sup_{1\leq i\leq n} E\left[|X_i| \mathbb{1}_{\{|X_i|>k\}}\right] \leq E\left[\sup_{1\leq i\leq n} |X_i| \mathbb{1}_{\{\sup|X_i|>k\}}\right] \underset{k\to\infty}{\longrightarrow} 0.$$

(iii) Für $X_n = n\mathbb{1}_{\{U\in[0,\frac{1}{n}]\}}$, $U \sim U[0,1]$ ist für $n > k$

$$E\left[|X_n| \mathbb{1}_{\{|X_n|>k\}}\right] = E[X_n],$$

also ist (X_n) nicht gleichgradig integrierbar.

(iv) Ist $\sup_{i\in I} \|X_i\|_{1+\varepsilon} < \infty$ mit $\varepsilon > 0$, so ist

$$\begin{aligned} \sup_{i\in I} E[|X_i| \mathbb{1}_{\{|X_i|>k\}}] &= \sup_{i\in I} E[|X_i| \mathbb{1}_{\{|X_i|^\varepsilon>k^\varepsilon\}}] \\ &\leq \sup_{i\in I} \frac{E[|X_i|^{1+\varepsilon}]}{k^\varepsilon} \underset{k\to\infty}{\longrightarrow} 0. \end{aligned}$$

Lemma 2.20. *Es sind äquivalent:*

(i) $(X_i)_{i\in I}$ *ist gleichgradig integrierbar*

(ii) $\sup_{i\in I} E[|X_i|] < \infty$ *und* $\lim_{\varepsilon\to 0} \sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i| \mathbb{1}_A] = 0$

(iii) $\lim_{k\to\infty} \sup_{i\in I} E[(|X_i| - k)^+] = 0.$

(iv) *Es gibt* $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *mit* $\frac{f(x)}{x} \underset{x\to\infty}{\longrightarrow} \infty$ *und* $\sup_{i\in I} E[f(|X_i|)] < \infty.$

Beweis:

(i) $\Rightarrow$ *(ii)*: Für $\delta > 0$ gibt es $k = k(\delta)$, sodass $\sup_{i\in I} E[|X_i|\mathbb{1}_{\{|X_i|>k\}}] < \delta$. Dann ist für alle $A \in \mathscr{F}$

$$\begin{aligned} E[|X_i|\mathbb{1}_A] &= E[|X_i|\mathbb{1}_{A\cap\{|X_i|>k\}}] + E[|X_i|\mathbb{1}_{A\cap\{|X_i|\leq k\}}] \\ &\leq \delta + k \cdot P(A). \end{aligned}$$

Mit $A = \Omega$ folgt

$$\sup_{i\in I} E[|X_i|] \leq \delta + k < \infty.$$

Außerdem ist

$$\sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i|] \leq \delta + k \cdot \varepsilon \underset{\varepsilon\to 0}{\longrightarrow} \delta$$

Da δ beliebig war folgt (ii).

(ii) $\Rightarrow$ *(iii)*: Zunächst ist $(|X_i| - k)^+ \leq |X_i|\mathbb{1}_{\{|X_i|\geq k\}}$. Sei $\varepsilon > 0$ und $k = k(\varepsilon)$, so dass

$$\frac{\sup_{i\in I} E[|X_i|]}{k} < \varepsilon.$$

Für $\varepsilon \to 0$ folgt, dass $k(\varepsilon) \to \infty$ und wir erhalten

$$\begin{aligned} \lim_{k\to\infty} \sup_{i\in I} E[(|X_i| - k)^+] &= \lim_{\varepsilon\to 0} \sup_{i\in I} E[(|X_i| - k(\varepsilon))^+] \\ &\leq \lim_{\varepsilon\to 0} \sup_{i\in I} E[|X_i|\mathbb{1}_{\{|X_i|>k(\varepsilon)\}}] \\ &\leq \lim_{\varepsilon\to 0} \sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i|\mathbb{1}_A] = 0. \end{aligned}$$

(iii) $\Rightarrow$ *(iv)*: Wir wählen $(k_n) \uparrow \infty$, so dass $\sup_{i\in I} E[(|X_i| - k_n)^+] \leq 2^{-n}$. Wir setzen

$$f(x) = \sum_{n\geq 1} (x - k_n)^+.$$

Dann ist f monoton wachsend und konvex. Für $x \geq 2k_n$ ist $\frac{f(x)}{x} \geq \sum_{k=1}^n \left(1 - \frac{k_n}{x}\right) \geq \frac{n}{2}$, also $\frac{f(x)}{x} \underset{x\to\infty}{\longrightarrow} \infty$. Außerdem gilt wegen monotoner Konvergenz, dass

$$E[f(|X_i|)] = \sum_{n\geq 1} E[(|X_i| - k_n)^+] \leq \sum_{n\geq 1} 2^{-k} = 1.$$

(iv) $\Rightarrow$ *(i)*: Setze $a(k) := \inf_{x\geq k} \frac{f(x)}{x}$, so dass $\lim_{k\to\infty} a(k) = \infty$. Dann gilt

$$\begin{aligned} \sup_{i\in I} E[|X_i|\mathbb{1}_{\{|X_i|\geq k\}}] &\leq \frac{1}{a(k)} \sup_{i\in I} E[f(|X_i|)\mathbb{1}_{\{|X_i|\geq k\}}] \\ &\leq \frac{1}{a(k)} \sup_{i\in I} E[f(|X_i|)] \underset{k\to\infty}{\longrightarrow} 0. \end{aligned}$$ □

Beispiel 2.21. Ist $X \in L^1$ und ist $(X_i)_{i\in J}$ gleichgradig integrierbar, so ist $(X_i - X)_{i\in I}$ gleichgradig integrierbar: Zunächst ist X gleichgradig integrierbar nach Beispiel 2.19. Außerdem ist

$$\sup_{i\in I} E[|X_i - X|] \leq E[|X|] + \sup_{i\in I} E[|X_i|] < \infty$$

und

$$\sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i - X|\mathbb{1}_A] \leq \sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i|\mathbb{1}_A] + \sup_{A:P(A)<\varepsilon} E[|X|\mathbb{1}_A] \xrightarrow[\varepsilon\to 0]{} 0$$

und nach Lemma 2.20 ist $(X_i - X)_{i\in I}$ gleichgradig integrierbar. Die Umkehrung folgt analog.

Der folgende Satz ist der *Konvergenzsatz von Vitali.*

Theorem 2.22. Sei $0 < p < \infty$ und $X_1, X_2, \ldots \in L^p$. Dann sind äquivalent

(i) Es gibt $X \in L^p$ und $X_n \overset{L^p}{\to} X$

(ii) $(|X_n|^p)$ ist gleichgradig integrierbar und es gibt eine Zufallsvariable X mit $X_n \xrightarrow{P} X$.

Gilt (i) oder (ii), so stimmen die Limiten überein.

Beweis. *(i)* $\Rightarrow$ *(ii)*: Mit der Markov-Ungleichung gilt für alle $\varepsilon > 0$, dass

$$P(|X_n - X| > \varepsilon) \leq \frac{E[|X_n - X|^p]}{\varepsilon^p} = \frac{\|X_n - X\|_p^p}{\varepsilon^p} \longrightarrow 0,$$

so dass $X_n \xrightarrow{P} X$.

Nun verwenden wir Lemma 2.20 (ii): Sei $\varepsilon > 0$ und $N = N(\varepsilon)$ so, dass $\|X_n - X\|_p < \varepsilon\ \forall n \geq N$. Dann ist

$$\begin{aligned}\sup E[|X_n|^p]^{1\wedge\frac{1}{p}} &= \sup \|X_n\|_p^{p\wedge 1} \\ &\leq \sup_{n<N} \|X_n\|_p^{p\wedge 1} + \underbrace{\sup_{n\geq N} \|X_n - X\|_p^{p\wedge 1}}_{\leq \varepsilon^{p\wedge 1}} + \|X\|_p^{p\wedge 1} < \infty\end{aligned}$$

unter Verwendung der Minkowski-Ungleichung[2].

[2] Das kann als Übungsaufgabe nachgerechnet werden. Sie können einfach $0 < p < 1$ und $p \geq 1$ getrennt betrachten.

Für $A \in \mathscr{F}$ erhalten wir ebenso

$$\begin{aligned} \sup \big(E[|X_n|^p \mathbb{1}_A]\big)^{1 \wedge \frac{1}{p}} &= \sup \|X_n \mathbb{1}_A\|_p^{p \wedge 1} \\ &\leq \sup_{n<N} \|X_n \mathbb{1}_A\|_p^{p \wedge 1} + \underbrace{\sup_{n \geq N} \|(X_n - X)\mathbb{1}_A\|_p^{p \wedge 1}}_{\leq \varepsilon^{p \wedge 1}} + \|X \mathbb{1}_A\|_p^{p \wedge 1}. \end{aligned}$$

Da $\|X \mathbb{1}_A\|_p \to 0$ für $P(A) \to 0$, folgt

$$\lim_{\delta \to 0} \sup_{A : P(A) < \delta} \sup_{n \in \mathbb{N}} E[|X_n|^p \mathbb{1}_A] \leq \varepsilon^{2p \wedge 1}$$

Die Behauptung folgt, weil $\varepsilon > 0$ beliebig war.

(ii) $\Rightarrow$ *(i)* Aus $X_n \xrightarrow{P} X$ folgt die Existenz einer Teilfolge (n_k) so dass $X_{n_k} \xrightarrow{\text{f.s.}} X$. Außerdem ist $\sup E[|X_n|^p] < \infty$ nach Lemma 2.20(ii). Nach dem Lemma von Fatou, Theorem 2.56, ist

$$E[|X|^p] = E\big[\liminf_{k \to \infty} |X_{n_k}|^p\big] \leq \sup_{n \in \mathbb{N}} E[|X_n|^p] < \infty$$

und somit $X \in L^p$. Weiterhin ist $|X_n - X|^p$ gleichgradig integrierbar (siehe Beispiel 2.21).

Da nach Voraussetzung

$$P(|X_n - X| > \delta) \underset{n \to \infty}{\longrightarrow} 0,$$

folgt

$$\begin{aligned} \lim_{n \to \infty} E[|X_n - X|^p] &= \lim_{n \to \infty} E[|X_n - X|^p \mathbb{1}_{\{|X_n - X| > \delta\}}] + \lim_{n \to \infty} \underbrace{E[|X_n - X|^p \mathbb{1}_{\{|X_n - X| < \delta\}}]}_{\leq \delta^p} \\ &\leq \delta^p \end{aligned}$$

und die Behauptung folgt, weil δ beliebig war. ☐

Korollar 2.23. *Seien* $0 < p < \infty$, $X_1, X_2, \ldots \in L^p$ *und* $X_n \xrightarrow{P} X$. *Dann sind äquivalent*

(i) $X_n \xrightarrow{L^p} X$

(ii) $\|X_n\|_p \to \|X\|_p$

(iii) $(|X_n|^p)_{n \geq 1}$ *ist gleichgradig integrierbar*

Wir schließen noch einige Bemerkungen über Konvergenz in Verteilung an, welche wir aber später noch ausführlicher behandeln. Wir betrachten den metrischen Raum (E, d). Wir bezeichnen

$$\begin{aligned} \mathscr{M}(E) &= \{\mu : \mu \text{ Wahrscheinlichkeitsmaß auf } \mathscr{B}(E)\} \\ \mathscr{P}_{\leq 1}(E) &= \{\mu : \mu \text{ Maß auf } \mathscr{B}(E) \text{ mit } \mu(E) \leq 1\} \end{aligned}$$

Aus dem Kontext heraus sollte keine Verwechslungsgefahr von $\mathscr{M}(E)$ mit der Potenzmenge von E bestehen. Mit $C_b(E)$ bezeichnen wir die Menge der stetigen, beschränkten, reellwertigen Funktionen auf $\mathbb{Z}$.

Definition 2.24. Die Folge $P_1, P_2, \ldots \in \mathscr{M}(E)$ konvergiert *schwach* gegen $P \in \mathscr{M}(E)$, falls

$$\int f dP_n \underset{n\to\infty}{\longrightarrow} \int f dP$$

für alle $f \in C_b(E)$. Wir schreiben dann $P_n \Rightarrow P$.

Lemma 2.25. *Gilt $\mu_n \Rightarrow \mu$ und $\mu_n \Rightarrow \nu \Rightarrow \mu = \nu$.*

Beweis. Nach Satz 1.15 reicht es, $\mu(A) = \nu(A)$ für alle abgeschlossenen $A \subseteq E$ zu zeigen, da die Menge der abgeschlossenen Mengen ein durchschnittsstabiler Erzeuger von $\mathscr{B}(E)$ ist. Setze

$$d(x, A) = \inf_{y\in A} d(x, y)$$
$$f_m(x) := \big(1 - m \cdot d(x, A)\big)^+.$$

Dann gilt $f_m \Rightarrow \mathbb{1}_A$, also

$$\mu(A) = \lim_{m\to\infty} \int f_m d\mu = \lim_{m\to\infty} \lim_{n\to\infty} \int f_m d\mu_n = \lim \int f_m d\nu - \nu(A)$$

wegen majorisierter Konvergenz. □

Definition 2.26. Sind $X, X_1, X_2, \ldots$ Zufallsvariable auf $(\Omega, \mathcal{A}, P)$, $(\Omega_1, \mathcal{A}_1, P_1)$, $(\Omega_2, \mathcal{A}_2, P_2)$, ... mit Werten in $\mathbb{Z}$. Wir sagen $X_1, X_2, \ldots$ *konvergieren in Verteilung* gegen X, falls $(X_n)_* P_n \Rightarrow X_* P$. Wir schreiben

$$X_n \Rightarrow X.$$

Satz 2.27. $X_n \xrightarrow{P} X \Rightarrow X_n \Rightarrow X$. *Ist X konstant, so gilt auch die Umkehrung.*

Beweis. Sei also $X_n \xrightarrow{P} X$. Wir zeigen die Behauptung durch Widerspruch. Angenommen für $f \in C_b(Z)$ gilt $\lim_{n\to\infty} E[f(X_n)] \neq E[f(X)]$. Dann gibt es eine Teilfolge n_k und ein $\varepsilon > 0$, so dass

$$\lim_{k\to\infty} |E[f(X_{n_k})] - E[f(x)]| > \varepsilon.$$

Da $X_n \xrightarrow{P} X$ gibt es eine Teilfolge (n_{k_e}),$e = 1, 2, \ldots$, so dass $X_{n_{k_e}} \xrightarrow{f.\,s.} X$. Mit majorisierter Konvergenz folgt

$$\lim_{e\to\infty} E[f(X_{n_{k_e}})] = E[f(X)],$$

ein Widerspruch.

Ist umgekehrt $X = c \in Z$. Dann ist $x \to d(x, c) \wedge \mathbb{1} \in C_b(E)$ und somit

$$\mathbb{E}[d(X_n, c) \wedge \mathbb{1}] \to E[d(X, c) \wedge \mathbb{1}] = 0,$$

also $X_n \xrightarrow{P} X$ wegen der Definition von stochastischer Konvergenz. □

Sind $X, X_1, X_2, \ldots$ identisch verteilt $\Rightarrow X_n \Rightarrow X$, allerdings gilt typischerweise weder f. s. noch stochastische Konvergenz (außer wenn X konstant ist, wie ein folgender Satz zeigt).

2.4 Das starke Gesetz der großen Zahlen

Ein zentrales Konzept dieses Grenzwertsatzes ist die Unabhängigkeit. Wir werden in diesem Zusammenhang auch die wichtigen Hilfsmittel von Borel-Cantelli und das Kolmogorov'sche 0-1-Gesetz kennenlernen.

Definition 2.28. (i) Eine Familie messbarer (d. h. $A_i \in \mathscr{F}$) Mengen $(A_i)_{i\in I}$ heißt *unabhängig*, falls

$$P\Big(\bigcap_{j\in I} A_j\Big) = \prod_{j\in I} P(A_j) \tag{2.29}$$

für alle $I \subset \mathscr{F}$ endlich.

(ii) Eine Familie von Mengensystemen $(C_i)_{i\in I}$, $C_i \subseteq \mathscr{F}$ $\forall i \in I$ heißt *unabhängig*, falls (2.29) für alle $I \subseteq \mathscr{F}$, $\mathscr{F}$ endlich und $A_j \in C_j$, $j \in I$, gilt.

(iii) Eine Familie von Zufallsvarialben $(X_i)_{i\in I}$ heißt *unabhängig*, falls $(\sigma(X_i))_{i\in I}$ unabhängig sind.

Man erhält unmittelbar, dass $(X_i)_{i\in I}$ unabhängig sind, falls sich ihre gemeinsamen Verteilungen als Produkt von den Marginalien ergibt, d.h.

$$\left.\begin{aligned} & P(X_i \in A_i, i \in I) = \prod_{i\in I} P(X_i \in A_i) \\ \Leftrightarrow \quad & ((X_i)_{i\in I})_* = \bigotimes_{i\in I} (X_i)_* P \end{aligned}\right\} \text{für alle } I \subset \mathscr{F},\ I \text{ endlich.}$$

Als wichtiges Hilfsmittel erweist sich das folgende „Blockungslem“:

Lemma 2.30. *Seien $(\Omega_i', \mathscr{F}_i'')$, $(\Omega_i'', \mathscr{F}_i'')$, $i \in J$, messbare Räume und $(X_i)_{i\in J}$ unabhängige Zufallsvariable mit $X_i : \Omega \to \Omega_i'$. Sind $\varphi_i : \Omega_i' \to \Omega_i''$ messbar, $i \in J$, so ist*

$$(\varphi_i(X_i))_{i\in J} \qquad \textit{unabhängig.}$$

Beweis. Zunächst ist $\sigma(\varphi_i(X_i)) \subset \sigma(X_i)$, da φ_i messbar. Da $\sigma(X_i)_{i\in J}$ bereits unabhängig sind, folgt die Behauptung. □

Zwei Zufallsvariablen heißen **unkorreliert**, falls $\mathrm{Cov}(X, Y) = 0$. Wir erhalten leicht, dass Unabhängigkeit Unkorreliertheit impliziert.

Satz 2.31. *Sind $X, Y \in L^1$ unabhängig, so ist $X \cdot Y \in L^1$ und*

$$E[XY] = E[X] \cdot E[Y].$$

Beweis. Die Aussage gilt zivilerweise für $X = \mathbb{1}_A$ und $Y = \mathbb{1}_B$. Mit Linearität folgt die Aussage für einfache X, Y mit monotoner Konvergenz für Positiv-/Negativteile und somit die Behauptung. □

Beispiel 2.32. Ist $X \sim \mathcal{N}(0, 1)$, so ist X, X^2 unkorreliert aber nicht unabhängig.

Sind $A_1, A_2, \dots$ Ereignisse, so definiert der folgende Limes Superior das Ereignis „unendlich viele der (A_n) treten ein“.

Definition 2.33. Für $A_1, A_2, \dots \in \mathscr{F}$ definieren wir

$$\limsup_{n\to\infty} A_n := \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m.$$

Theorem 2.34 (Borel-Cantelli).

(i) Seien $A_1, A_2, \dots \in \mathscr{F} \Rightarrow$

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

(ii) Sind die $(A_n)_{n\geq 1}$ unabhängig, so gilt

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

Beweis:

(i) Stetigkeit von oben von P impliziert

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

(ii) Für $x \in [0,1]$ gilt: $\log(1-x) \leq -x$ also folgt mit Stetigkeit von unten, dass

$$\begin{aligned} P((\limsup A_n)^c) &= P\Big(\bigcup_{n\geq 1}\bigcap_{m\geq n} A_m^c\Big) \\ &= \lim_{n\to\infty} \exp\Big(\sum_{m\geq n} \log\big(1-P(A_m)\big)\Big) \\ &\leq \lim_{n\to\infty} \exp\Big(-\sum_{m\geq n} P(A_m)\Big) = 0. \end{aligned}$$

□

Beispiel 2.35 (Unendlicher Münzwurf). $X_1, X_2, \dots$ iid, $X_i \in \{0,1\}$ und $P(X_n = 1) = \frac{1}{2} \Rightarrow \sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow$ fast sicher erscheint unendlich mal Kopf.

Auf die Unabhängigkeit kann nicht verzichtet werden: Ist

$$B_1 = B_2 = B_n = \{X_1 = \text{ Kopf}\},$$

so gilt $\sum P(B_n) = \infty$, aber $P\big(\limsup_{n\to\infty} B_n\big) = P(X_1 = \text{ Kopf}) = \frac{1}{2}$.

Die Aussagen „unendlich oft tritt Kopf ein“ bilden eine spezielle Klasse von Ereignissen, die wir nun genau untersuchen.

Satz 2.36. *Sei $(C_i)_{i\in J}$ eine Familie unabhängiger Mengensysteme. Ist jedes C_i, $i \in J$, durchschnittstabil, so sind auch $(\sigma(C_i))_{i\in J}$ unabhängig.*

Beweis. Sei $J = \{j_1, \dots, j_n\} \subset I$ und ohne Einschränkung $|J| > 1$. Wir definieren

$$\mathscr{D} := \Big\{A \in J : P(A \cap A_{j_2} \cap \cdots \cap A_{j_n}) = P(A) \cdot \prod_{i=2}^{n} P(A_{j_i})\Big\}.$$

Dann ist $\mathscr{D}$ Dynkin-System, denn

1) $A \subseteq B, A, B \in \mathscr{D} \Rightarrow B \setminus A \in \mathscr{D}$: Mit $C = \bigcap_{i=2}^{n} A_{j_i}$ ist

$$\begin{aligned} P\Big(B \setminus A \cap \bigcap_{i=2}^{n} A_{j_i}\Big) &= P(B \cap C) - P(A \cap C) = (P(B) - P(A)) \prod_{i=2}^{n} P(A_{j_i}) \\ &= P(B \setminus A) \cdot \prod_{i=2}^{n} P(A_{j_i}) \end{aligned}$$

2) Für $A_1, A_2, \dots \in \mathscr{D}$ mit $A_1 \subseteq A_2 \subseteq \cdots$ ist $\quad$ (σ-Stetigkeit von unten)

$$\begin{aligned} P\Big(\bigcup_{n=1}^{\infty} A_n \cap C\Big) &\overset{(2.29)}{=} \lim_{m\to\infty} P(A_m \cap C) \quad = \quad \lim_{n\to\infty} P(A_m) \cdot \prod_{i=3}^{n} P(A_{j_i}) \\ &\overset{(2.29)}{=} P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \cdot \prod_{i=2}^{n} P(A_{j_i}). \end{aligned}$$

Da C_{i_1} durchschnittstabil ist, ist $\sigma(C_{i_1}) \subseteq \mathscr{D}$, also folgt

$$P(A \cap A_{j_2} \cap \cdots \cap A_{j_n}) = P(A) \cdot \prod_{i=2}^{n} P(A_{j_i})$$

für alle $A \in \sigma(C_{i_1})$. Da die Wahl von i_1 beliebig war, erhält man die Aussage. ☐

Bemerkung 2.37. (i) Als leichte Folgerung erhält man, dass $(A_i)_{i\in J}$ genau dann unabhängig sind, falls $(\mathbb{1}_{A_i})_{i\in J}$ unabhängig sind.

(ii) Ist $(J_i)_{i\in J}$ eine Familie unabhängiger σ-Algebren und $\mathcal{I} = \{I_k : K \in \mathcal{K}\}$ eine Partition von J, so sind auch $(\sigma(\mathscr{F}_i : i + I_k) : k + K)$ unabhängig.

2.4.1 Der Beweis des starken Gesetzes

Dieser Abschnitt widmet sich dem Beweis des starken Gesetzes der großen Zahl. Wir werden noch einige Hilfsmittel entwickeln, bevor wir den Beweis führen können.

Ein erstes, zentrales Hilfsmittel ist die Maximalungleichung von Kolmogorov. Sie wird oft mit Hilfe von Stoppzeiten[3] bewiesen, was in dem Setting hier aber nicht nötig ist und wir

[3]Siehe etwa Rüschendorf(2016), Satz 2.6.3.

präsentieren einen elementaren Beweis.

Satz 2.38 (Maximal-Ungleichung). *Seien $X_1, X_2, \dots$ stochastisch unabhängige, reellwertige Zufallsvariablen mit $E[X_i] = 0,\ \mathrm{Var}(X_i) = \sigma_i^2 < \infty$ und $S_n = \sum_{i=1}^n X_i$. Dann gilt für alle $\varepsilon > 0$, das*

$$P\Big(\max_{1\le k\le n} |S_k| \ge \varepsilon\Big) \le \frac{1}{\varepsilon^2}\sum_{k=1}^n \sigma_k^2. \tag{2.39}$$

Beweis. Zentral in dem Beweis ist die disjunkte Zerlegung von $\{\max_{1\le k\le n} |S_k| \ge \varepsilon\}$ in $\sum_{k=1}^n A_k$ mit $A_k = \{|S_k| \ge \varepsilon, |S_j| < \varepsilon \text{ für } j < k\}$.

Dann ist

$$P(A_k) = E\Big[\mathbb{1}_{\{|S_k|\varepsilon^{-1}\ge 1, |S_j|<\varepsilon,\ j<k\}}\Big] \le E\Big[\frac{S_k^2}{\varepsilon^2}\mathbb{1}_{A_k}\Big].$$

Als nächsten Schritt möchten wir S_k^2 durch S_n^2 abschätzen, wofür wir die Unabhängigkeit und Zentriertheit der (X_i) ausnutzen. Denn es gilt, dass

$$\begin{aligned} E[S_k(S_n - S_k)\mathbb{1}_{A_k}] &= 0, \\ E\Big[(S_n - S_k)^2\mathbb{1}_{A_k}\Big] &\ge 0 \end{aligned}$$

und somit

$$P(A_k) \le \varepsilon^{-2} E\Big[\big(S_k + (S_n - S_k)\big)^2 \mathbb{1}_{A_k}\Big] = \varepsilon^{-2}E\Big[S_n^2\mathbb{1}_{A_k}\Big] \le \frac{1}{\varepsilon^2}\sum_{k=1}^n \sigma_k^2$$

und die Behauptung ist bewiesen.

□

Kombinieren wir die Maximal-Ungleichung mit unseren Konvergenzkriterien für fastsichere Konvergenz erhalten wir folgenden Konvergenzsatz. Bemerkenswert ist hieran, dass nicht i.i.d. vorausgesetzt wird, sondern lediglich Unabhängigkeit (und eben Summierbarkeit der Varianzen).

Satz 2.40. *Sind $X_1, X_2, \dots$ unabhängig und gilt $E[X_n] = 0$ für alle $n \ge 1$ sowie*

$$\sum_{n=1}^\infty \mathrm{Var}(X_n) < \infty,$$

so konvergiert $\sum_{n\ge 1} X_n$ P-fast sicher.

Beweis. Zunächst müssen wir nach dem Cauchy-Kriterium für fast-sichere Konvergenz, Satz 2.12 (ii) zeigen, dass

$$P\Big(\sup_{n\geq m}|S_n - S_m| \geq \varepsilon\Big) \longrightarrow 0$$

für $m \to \infty$. Die Maximal-Ungleichung liefert aber bereits, dass

$$P\Big(\max_{m\leq n\leq k}|S_n - S_m| \geq \varepsilon\Big) \leq \frac{1}{\varepsilon^2}\sum_{n=m+1}^{k}\mathrm{Var}(X_n).$$

Mit Hilfe der σ-Stetigkeit folgt, dass

$$P\Big(\sup_{n\geq m}|S_n - S_m| \geq \varepsilon\Big) = \lim_{k\to\infty} P\Big(\max_{m\leq n\leq k}|S_k| \geq \varepsilon\Big) \leq \frac{1}{\varepsilon^2}\sum_{n=m+1}^{\infty}\mathrm{Var}(X_k).$$

Für $m \to \infty$ konvergiert der letzte Ausdruck gegen Null und die Behauptung ist bewiesen.

□

Oft wollen wir eine Folge von Zufallsvariablen nur ein bisschen verändern, ohne dass sich das Konvergenzverhalten ändert.

Definition 2.41. Zwei Folgen von Zufallsvariablen (X_n) und (Y_n) heißen *stochastisch äquivalent*, falls

$$\sum_{n\geq 1} P(X_n \neq Y_n) < \infty. \tag{2.42}$$

Mit dem Satz von Borell-Cantelli folgt sofort, dass beide Folgen fast sicher konvergieren, falls sie stochastisch äquivalent sind: Denn (2.42) impliziert, dass

$$P(\limsup\{X_n \neq Y_n\}) = 0.$$

D.h. es existiert eine Menge A mit Wahrscheinlichkeit eins und eine Zufallsvariable $n_0 \in \mathbb{N}$, so dass für alle $\omega \in A$ und $n \geq n_0(\omega)$ gilt, dass $X_n(\omega) = Y_n(\omega)$. Dann folgt, dass $\sum_{n\geq 1}(X_n - Y_n)$ fast sicher konvergiert.

Zentrales Konvergenzkriterium wird der folgende *Drei-Reihensatz* sein:

Satz 2.43. *Seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen. Dann konvergiert $S_n = \sum_{k=1}^n X_k$ genau dann P-fast sicher, falls ein $c \in (0,\infty)$ existiert, so dass die folgenden drei Reihen konvergieren:*

$$(i)\ \sum_{n\geq 1} P(|X_n| \geq c), \qquad (ii)\ \sum_{n\geq 1}\mathrm{Var}(X_n \mathbb{1}_{\{|X_n|\geq c\}}) \qquad (iii)\ \sum_{n\geq 1} E\big[X_n \mathbb{1}_{\{|X_n|\geq c\}}\big].$$

Beweis. Betrachten wir $Y_n = X_n \mathbb{1}_{\{|X_n| \geq c\}} - E[X_n \mathbb{1}_{\{|X_n| \geq c\}}]$. Dann impliziert Bedingung (ii) mit Satz 2.40, dass $\sum_{k=1}^n Y_k$ fast sicher konvergiert. Mit (iii) erhalten wir, dass auch

$$\sum_{k=1}^n X_k \mathbb{1}_{\{|X_k| \geq c\}} = \sum_{k=1}^n Y_k + \sum_{k=1}^n E[X_k \mathbb{1}_{\{|X_k| \geq c\}}]$$

konvergiert.

Bedingung (i) impliziert, dass (X_n) und $(X_n \mathbb{1}_{\{|X_n| \geq c\}})$ asymptotisch äquivalent sind, denn

$$\sum_{n \geq 1} P\big(X_n \neq X_n \mathbb{1}_{\{|X_n| \geq c\}}\big) = \sum_{n \geq 1} P(|X_n| \geq c).$$

Nach obiger Bemerkung konvergiert also auch $\sum_{k=1}^n X_n$. □

Als letzter Schritt benötigen wir das Kronecker Lemma.

Lemma 2.44. *Seien $x_1, x_2, \dots$ reelle Zahlen, $a_1, a_2, \dots$ positive Zahlen mit $a_n \uparrow \infty$ so, dass $\sum \frac{x_n}{a_n}$ konvergiert. Dann gilt, dass*

$$\frac{1}{a_n} \sum_{k=1}^n x_k \longrightarrow 0.$$

Beweis folgt.

Wir erhalten das starke Gesetz der großen Zahl von Kolmogorov unter der Annahme über existierende zweite Momente. Im Folgenden soll diese Annahme noch so weit wie möglich abgeschwächt werden.

Theorem 2.45 (Starkes Gesetz großer Zahlen)**.** Seien $X_1, X_2, \dots \subseteq L^2$ unabhängige und reellwertige Zufallsvariablen. Gibt es eine Folge positiver Zahlen $a_n \uparrow \infty$, so dass

$$\sum_{n \geq 1} \frac{\operatorname{Var}(X_n)}{a_n^2} < \infty,$$

so folgt, dass

$$\frac{1}{a_n} \sum_{k=1}^n (X_k - E[X_k]) \longrightarrow 0$$

P-fast sicher.

Beweis. Wir setzen $Y_n = (a_n)^{-1}(X_n - E[X_n])$. Dann sind $Y_1, Y_2, \ldots$ unabhängig, zentriert und $\sum \operatorname{Var}(Y_n) < \infty$ nach Voraussetzung. Nach Satz 2.40 konvergiert $\sum_{k=1}^n Y_k$ P-fast sicher.

Nach dem Kronecker Lemma folgt, dass

$$\frac{1}{a_n}\sum_{k=1}^n a_k Y_k = \frac{1}{a_n}\sum_{k=1}^n (X_k - E[X_k]) \longrightarrow 0$$

P-fast sicher. □

Theorem 2.46 (Starkes Gesetz großer Zahlen unter L^1-Voraussetzung)**.** Seien $X_1, X_2, \ldots$ reellwertige Zufallsvariablen und i.i.d. Gilt $E[|X_1|] < \infty$, so folgt, dass

$$\frac{1}{n}\sum_{k=1}^n X_k \longrightarrow E[X_1],$$

P-fast sicher.

Für den Fall, dass $E[|X_1|] = \infty$ folgt Divergenz, siehe etwa Rüschendorf(2016), Satz 2.6.13.

Beweis. Die Idee für den Beweis ist, die Zufallsvariablen geschickt abzuschneiden. Hierzu setzen wir $Y_n = X_n \mathbb{1}_{\{|X_n| \le n\}}$, $n \ge 1$. Wir zeigen, dass (X_n) und (Y_n) asymptotisch äquivalent sind, denn

$$\begin{aligned}
\sum_{n\ge 1} P(X_n \ne Y_n) &= \sum_{n\ge 1} P(|X_n| > n) = \sum_{n\ge 1} P(|X_1| > n) \\
&= \sum_{n\ge 1}\sum_{j\ge n} P(j < |X_1| \le j+1) \\
&= \sum_{j\ge 1}\sum_{n=1}^{j} P(j < |X_1| \le j+1) = \sum_{j\ge 1} jP(j < |X_1| \le j+1) \\
&\le E[|X_1|] < \infty
\end{aligned}$$

Nun kann man Theorem 2.45 anwenden, und zwar auf (Y_n). In der Tat, es gilt

$$\begin{aligned}\sum_{n\geq 1}\frac{Var(Y_n)}{n^2} &\leq \sum_{n\geq 1}\frac{E[Y_n^2]}{n^2} = \sum_{n\geq 1}\frac{1}{n^2}\int_{\{|X_1|\leq n\}}(X_1)^2 dP \\ &= \sum_{n\geq 1}\frac{1}{n^2}\sum_{j=1}^{n}\int_{\{j-1<|X_1|\leq j\}}(X_1)^2 dP \\ &= \sum_{j\geq 1}\int_{\{j-1<|X_1|\leq j\}}(X_1)^2 dP\cdot\sum_{n=j}^{\infty}\frac{1}{n^2} \\ &\leq \sum_{j\geq 1}\int_{\{j-1<|X_1|\leq j\}} j\cdot |X_1| dP\cdot\sum_{n=j}^{\infty}\frac{1}{n^2}.\end{aligned}$$

Die letzte Reihe können wir direkt auswerten: Zunächst ist für $j=1$, $\sum_{n\geq 1}\frac{1}{n^2}=\frac{\pi^2}{6}$. Für $j>1$ ist

$$\sum_{n=j}^{\infty}\frac{1}{n^2}\leq\int_{(j-1,\infty)}\frac{1}{x^2}dx = -\frac{1}{x}\Big|_{j-1}^{\infty} = \frac{1}{j-1}.$$

Multiplizieren wir diese Reihe mit j, so erhalten wir zwei als obere Schranke:

$$j\sum_{n=j}^{\infty}\frac{1}{n^2}\leq\begin{cases}\frac{\pi^2}{6}\leq 2 & j=1\\ \frac{j}{j-1}\leq 2 & j>1.\end{cases}$$

Es folgt, dass

$$\sum_{n\geq 1}\frac{Var(Y_n)}{n^2}\leq 2E[|X_1|]<\infty.$$

Theorem 2.45 impliziert nun, dass

$$\frac{1}{n}\sum_{j=1}^{n}(Y_k - E[Y_k])\longrightarrow 0$$

P-fast sicher, wobei $E[Y_k]\to E[X_1]$ mit Hilfe von majorisierter Konvergenz. Nutzt man nun das Lemma von Césaro, so folgt die Behauptung. □

2.4.2 Empirische Verteilungen und der Satz von Glivenko-Cantelli

Als wichtige Anwendung des starken Gesetzes der großen Zahl betrachten wir empirische Verteilungen. Diese sind das zentrale Hilfsmittel der nichtparametrischen Statistik, da sie unter sehr geringen Annahmen in der Lage sind, statistische Aussagen zu treffen.

Definition 2.47. Seien $X_1, X_2, \dots$ Zufallsvariablen. Dann heißt das Maß μ_n, definiert durch

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

die *empirische Verteilung* von $X_1, \dots, X_n$. Sind die Zufallsvarialben reellwertig, so heißt

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}}, \qquad x \in \mathbb{R}$$

die *empirische Verteilungsfunktion*.

Theorem 2.48 (Glivenko-Cantelli)**.** Sind $X_1, X_2, \dots$ reellwertig und iid (unabhängig und identisch verteilt), mit $X_1 \sim F$, so gilt

$$\sup_{x \in \mathbb{R}} \left| F_n(x) - F(x) \right| \xrightarrow[n \to \infty]{f.s.} 0.$$

Beweis. Wir setzen $Y_n = \mathbb{1}_{\{X_n \le x\}}$ und $Z_n = \mathbb{1}_{\{X_n < x\}}$. Nach dem starken Gesetz der großen Zahlen gilt

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow{f.s.} E[Y_1] = F(x)$$
$$F_n(x-) = \frac{1}{n} \sum_{i=1}^{n} Z_i \xrightarrow{f.s.} F(x-).$$

Es bleibt zu zeigen, dass die Konvergenz auch gleichmäßig ist. Dazu fixieren wir ein N und setzen

$$x_j^N := \inf\left\{ x \in \mathbb{R} : F(x) \ge \frac{j}{N} \right\}, \quad j = 0, \dots, N$$

und

$$R^N := \max_{j=1,\dots,N-1} \Big(\left| F_n(x_j^N) - F(x_j^N) \right| + \left| F_n(x_j^N-) - F(x_j^N-) \right| \Big).$$

Nach dem starken Gesetz der großen Zahlen gilt also

$$R_n^N \xrightarrow{f.s.} 0.$$

Außerdem ist für $x \in \left(x_{n-1}^N, x_j^N\right)$

$$\begin{aligned} F_n(x) &\leq F_n(x_j^N) \leq F_n(x_j^N) + R_n^N \leq F(x) + R_n^N + \frac{1}{N} \qquad \left(F(x)+\tfrac{1}{N} \geq F_n(x_{j-1}^N)\right) \\ F_n(x) &\geq F_n(x_{j-1}^N) \geq F_n(x_{j-1}^N) - R_n^N \geq F(x) - R_n^N - \frac{1}{N}. \end{aligned}$$

Damit ist $\sup\left|F_n(X) - F(X)\right| \leq \dfrac{1}{N} + R_n^N \xrightarrow{f.s.} \dfrac{1}{N}$. Da die linke Seite nicht von N abhängt, folgt die Aussage durch $N \to \infty$. □

2.5 Schwache Konvergenz

Wir betrachten den metrischen Raum (E, d). Wir bezeichnen

$$\begin{aligned} \mathscr{M}(E) &= \{\mu : \mu \text{ Wahrscheinlichkeitsmaß auf } \mathscr{B}(E)\} \\ \mathscr{M}_{\leq 1}(E) &= \{\mu : \mu \text{ Maß auf } \mathscr{B}(E) \text{ mit } \mu(E) \leq 1\}, \end{aligned}$$

wobei man die Schranke 1 auch durch eine beliebige andere Konstante ersetzen kann (man erhält dann den Raum der beschränkten Maße). Uns interessiert hier ganz besonders der Zusammenhang mit Wahrscheinlichkeitsmaßen, weswegen die Konstante 1 von besonderem Interesse ist. Allgemeinere Betrachtungen findet man z.B. bei Bauer(1990). Weiterhin bezeichnen wir mit $C_b(E)$ die Menge der stetigen, beschränkten, reellwertigen Funktionen auf $\mathbb{Z}$, sowie mit $C_c(E)$ die Menge der stetigen, reellwertigen Funktionen mit kompaktem Träger.

Definition 2.49. Die Folge $P_1, P_2, \ldots \in \mathscr{M}(E)$ konvergiert *schwach* gegen $P \in \mathscr{M}(E)$, falls

$$\int f dP_n \underset{n\to\infty}{\longrightarrow} \int f dP$$

für alle $f \in C_b(E)$. Wir schreiben dann $P_n \Rightarrow P$.

Bei dieser Definition fällt auf, dass die Gesamtmasse $P(\Omega) = 1$ in allen Fällen gleich bleibt. Dies ist eine besondere Eigenschaft der *schwachen* Konvergenz. Im allgemeineren Rahmen fällt dies als Zusatzbedingung auf - weswegen wir ein noch schwächeres Konzept der Konvergenz einführen, die vage Konvergenz. Eine schöne Darstellung zur vagen Konvergenz findet sich in Bauer(1990). Zunächst einmal bemerken wir, dass für ein endliches Maß P und eine Funktion $f \in C_c(E)$ die Integrale $\int f dP := Pf$ stets definiert sind. Die Konvergenz der reellen Zahlen Pf_n impliziert dass der Grenzwert wie die einzelnen Folgenglieder eine Linearform auf $C_c(E)$ bildet. Nach dem Darstellungssatz von Riesz lässt sich zu diesem Grenzwert wieder ein Maß zuordnen, was folgenden Begriff motiviert.

Definition 2.50. Seien $\mu_1, \mu_2, \ldots \subseteq \mathscr{M}_{\leq 1}(E)$ und μ ein Maß auf $\mathscr{B}(E)$. Gilt

$$\int f d\mu_n \to \int f d\mu$$

für alle $f \in C_c(E)$, so konvergiert die Folge (μ_n) *vage* gegen μ, und wir schreiben

$$\mu_n \Rightarrow_{\mathrm{v}} \mu.$$

Vage Konvergenz kann man auch für allgemeinere Maße als Wahrscheinlichkeitsmaße definieren (Radon-Maße, das sind lokal endliche und von innen reguläre Maße), was wir hier aber nicht benötigen.

Im Gegensatz zu der schwachen Konvergenz kann man bei der vagen Konvergenz Masse verlieren, wie folgendes Resultat zeigt.

Lemma 2.51. *Sind $P_1, P_2, \dots$ Wahrscheinlichkeitsmaße auf $\mathscr{B}(\mathbb{R})$ und μ Maß auf $\mathscr{B}(\mathbb{R})$ mit $P_n \Rightarrow_v \mu$, so gilt $\mu(\mathbb{R}) \leq 1$.*

Beweis: Wir wählen $(f_n) \subset C_c(\mathbb{R})$ mit $f_n \uparrow 1$. Mit monotoner Konvergenz folgt

$$\mu(\mathbb{R}) = \sup_{n \geq 1} \int f_n d\mu = \sup_{n \geq 1} \lim_{k \to \infty} \int f_n dP_k \leq 1. \qquad \square$$

Wir erinnern an die Konvergenz in Verteilung:

Definition 2.52. Sind $X, X_1, X_2, \dots$ Zufallsvariable auf $(\Omega, \mathcal{A}, P)$, $(\Omega_1, \mathcal{A}_1, P_1)$, $(\Omega_2, \mathcal{A}_2, P_2)$, $\dots$ mit Werten in E. Wir sagen $X_1, X_2, \dots$ *konvergieren in Verteilung* gegen X, falls $(X_n)_* P_1 \Rightarrow X_* P$ und schreiben

$$X_n \Rightarrow X.$$

Bemerkung 2.53.
- Damit gilt $X_n \Rightarrow X \Leftrightarrow P_n(f(X_n) \in \cdot\,) \to P(f(X) \in \cdot\,)$ $\forall f \in C_b(E)$.
- Da $f \in C_b(E)$ ist der schwache Limes von Wahrscheinlichkeitsmaßen immer ein Wahrscheinlichkeitsmaß.

  Allerdings ist $1 \notin \mathbb{1}_C(E)$, so dass dies für die vage Konvergenz nicht gelten muss.
- Die schwache Topologie ist die schwächste Topologie bzgl. der $P \to \int f dP$ für alle $f \in C_b(E)$ stetig ist. Ebenso ist die vage Topologie die schwächste Tobologie, bezüglich der dies für alle $f \in C_c(E)$ gilt.

Beispiel 2.54. (i) Seien $x, x_1, x_2, \dots \in \mathbb{R}$ mit $x_n \to x$. Setze $P = \delta_x$, $P_n = \delta x_n$. Dann gilt $P_n \Rightarrow P$, da

$$\int f dP_n = f(x_n) \to f(x) = \int f dP$$

für alle $f \in C_b(R)$.

(ii) Gilt $x_n = n$, so folgt $P_n \Rightarrow_v 0$, allerdings nicht $P_n \Rightarrow 0$, da 0 kein Wahrscheinlichkeitsmaß ist.

(iii) Sind $X, X_1, X_2, \dots$ identisch verteilt so gilt $X_n \Rightarrow X$, allerdings gilt typischerweise weder fast sichere noch stochastische Konvergenz.

(iv) Der zentrale Grenzwertsatz.

Als wichtiges Resultat erhalten wir den Satz von Skorokhod, der einen Zusammenhang zwischen schwacher und fast sicherer Konvergenz bildet.

Theorem 2.55 (Skorokhod)**.** Sei (E, r) vollständig[4] und separabel, $X, X_1, \dots$ Zufallsvariable mit Werten in E. Dann sind äquivalent

(i) $X_n \Rightarrow X$

(ii) Es gibt ein $(\Omega', \mathcal{A}', P')$ mit Zufallsvariablen Y, Y_1, Y_2 und $Y_n \xrightarrow{f.\,s.} Y$ sowie $Y \overset{\mathscr{L}}{=} X$, $Y_i \overset{\mathscr{L}}{=} X_i$, $i = 1, 2, \dots$

Beweis. „$\Leftarrow$“ klar.

„$\Rightarrow$“ Wir setzen hier noch die Existenz eines Wahrscheinlichkeitsraumes mit abzählbar vielen unabhängigen Zufallsvariablen voraus. Dies werden wir in Kürze noch beweisen.

Zunächst nehmen wir $E = \{1, \dots, m\}$ an. Setze $p_k = P(X = k)$ und $p_k^* = P(X_n > k)$, $k \in E$. Nimmt man $U \sim U[0,1]$ unabhängig von X, so kann man leicht Zufallsvariablen $\xi_n \overset{\mathscr{L}}{=} X_n$ konstruieren, so dass $\xi_n = k$, falls $X = k$ und $U \leq \frac{P_k^n}{p_k}$. Da $p_k^n \to p_k$ folgt $\xi_n \Rightarrow X$ aus $X_n \Rightarrow X$, und da es auch nur endlichviele Werte sind sogar f. s.

Sei nun E allgemein. Wir fixieren $p \in \mathbb{N}$ und teilen E in Borel-Mengen $B_1, B_2, \dots$ mit Durchmesser kleiner 2^{-p} mit $P(X \in \partial B_n) = 0$.

Nun wählen wir m groß genug so, dass

$$P\Big(X = \bigcup_{k \leq m} B_k\Big) < 2^{-p}$$

und setzen $B_0 := E \setminus \bigcup_{k \leq m} B_k$.

Für $k = 0, \dots, m$ setzen wir $\xi = k$ falls $x \in B_k$ und $\xi_n = k$ falls $X_n \in B_k, n \in \mathbb{N}$. Dann gilt $\xi_n \Rightarrow \xi$ und wegen des Resultats über endliches E finden wir $\xi_n^N \overset{\mathscr{L}}{=} \xi_n$, so dass $\xi_n^n \Rightarrow \xi$ *f. s.* □

[4] Hinweis: Kallenberg verwendet das Theorem ohne die Vollständigkeit, damit ist die Existenz von iid Variablen etwas schwerer!

Satz 2.56 (Helly-Bray)**.** *Seien $P_1, P_2, \ldots \in \mathscr{M}(\mathbb{R})$. Dann gibt es eine Teilfolge $(P_{n_k})_{k\geq 1}$ und $\mu \in \mathscr{M}_{\leq 1}(\mathbb{R})$, so dass*

$$P_n \Rightarrow_v \mu.$$

Anders formuliert, sagt dieser Satz, dass der Raum aller Verteilungsfunktionen von Maßen $\mu \in \mathscr{M}_{\leq 1}(\mathbb{R})$ folgenkompakt ist. Man erhält, dass für eine Folge (P_n) mit einem einzigen Häufungspunkt μ bereits Verteilungskonvergenz folgt.

Beweis. Die Werte der Verteilungsfunktion $F_1, F_2, \ldots$ der (P_n) leben auf dem kompakten Intervall $[0,1]$. Wir nutzen eine Abzählung (x_n) von $\mathbb{Q}$ und finden wegen der Kompaktheit von $[0,1]$ eine Teilfolge (F_{n_k}) und eine Funktion G, so dass

$$F_{n_k}(x_i) \underset{k\to\infty}{\longrightarrow} G(x_i) \qquad i \geq 1$$

mit $G(x_i) \in [0,1]$ für alle $i \geq 1$. Nun definieren wir

$$F(x) = \inf\{G(r) : r \in 0 : r > x\},$$

und mit (F_{n_k}) ist auch G und somit F monoton sowie rechtsstetig. Damit existiert ein zugehöriges Maß μ, so dass $\mu((x,y]) = F(y) - F(x)$, $x < y \in \mathbb{R}$.

Wir zeigen noch $P_n \Rightarrow_{\mathrm{v}} \mu$. Dazu sei $f \in C_c(\mathbb{R})$. Ohne Einschränkung sei $f \geq 0$. Nach Konstruktion gilt $F_n(x) \to F(x)$ an allen Stetigkeitsstellen von F. Da F wachsend und rechtsstetig ist, gibt es nur abzählbar viele Unstetigkeitsstellen $D \subset \mathbb{R}$. Es folgt für jedes $[x,y]$ mit $x, y \in D$

$$P_n(U) \to \mu(U)$$

ebenso wie für endliche Vereinigungen, welche wir mit $\mathcal{V}$ bezeichnen.

Wir wählen $B \subseteq \mathbb{R}$ offen und (U_k), $(V_k) \leq \mathcal{V}$, so dass $U_k \uparrow B$, $V_k \downarrow \overline{B} \Rightarrow$
(Ziel: $\liminf = \limsup$)

$$\begin{aligned}\int_B d\mu &= \lim_{n\to\infty} \mu(U_k) = \lim_{k\to\infty} \liminf_{n\to\infty} P_n(U_k) \leq \liminf_{n\to\infty} P_n(B) \\ &\leq \limsup_{n\to\infty} P_n(B) \underset{\substack{\uparrow \\ \lim\limits_{k\to\infty}}}{\leq} \limsup_{n\to\infty} P_n(V_k) = \lim_{k\to\infty} \mu(V_n) = \int_{\overline{B}} d\mu \end{aligned} \tag{2.57}$$

Da $\mu(f > t) = 0$ für Lebesgue-fast sicher folgt mit majorisierter Konvergenz

$(P_n(f > t) \leq 1)$

$$
\begin{aligned}
\int f d\mu = \int_0^\infty \mu(f > t) dt &\overset{(2.57)}{\leq} \int_0^\infty \liminf P_n(f > t) dt \\
&\leq \liminf \int_0^\infty P_n(f > t) dt = \liminf \int f dP_n \\
&\leq \limsup \int f dP_n = \limsup \int_0^\infty P_n(f > t) dt \\
&\leq \int \limsup P_n(f > t) dt \leq \int_0^\infty \mu(f \geq t) dt = \int f d\mu,
\end{aligned}
$$

also $\lim \int f dP_n = \int f d\mu$. □

Wir benötigen folgendes nützliche Resultat was man mit dem Fubini-Theorem beweist (!): Für reellwertiges $X \geq 0$ ist

$$E[X] = \int_0^\infty P(X > t) dt$$

Theorem 2.58 (Portmanteau)**.** Seien $X, X_1, X_2, \ldots$ Zufallsvarialbe mit werten in E. Dann sind äuqivalent

(i) $X_n \Rightarrow X$

(ii) $E[f(X_n)] \Rightarrow E[f(X)]$ für alle Lipschitz-stetigen und beschränkte f

(iii) $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$ für alle offenen $G \subseteq E$

(iv) $\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$ für alle abgeschlossenen $F \subseteq E$

(v) $\lim_{n\to\infty} P(X_n \in B) = P(X \in B)$ für alle $B \in \mathscr{B}(E)$ mit $P(X \in \partial B) = 0$.

Die $B \in \mathscr{B}(E)$ mit $P(X \in \partial B)$ heißen auch $P_* X$-randlos. $\partial B = \overline{B} \setminus B^0$ ($\overline{B}$ = Abschluss, B^0 = Innere von B).

Beweis:

$(i) \Rightarrow (ii)$: klar

$(ii) \Rightarrow (iv)$: Sei $F \subseteq E$ abgeschlossen und f Lipschitz-ststig, so dass $f_k \downarrow \mathbb{1}_F$ (etwa $f_m(x) = (1 - mr(x, F))^+$)

$\Rightarrow \limsup P(X_n + F) \leq \inf_{k\geq 1} \limsup_{n\to\infty} E[f_k(X_n)]] \overset{(ii)}{=} \inf_{k\geq 1} E[f_k(X)] = P(x \in F)$

$(iv) \Leftrightarrow (iii)$: Setze $G = E \setminus F$ bzw. $F = E \setminus G$.

$(iii) \Rightarrow (i)$: Sei $f \geq 0$ stetig. Dann ist

$$E[f(X)] = \int_0^\infty \underbrace{P(f(X) > t)}_{\substack{P(X \in G) \text{ mit} \\ G = \{x : f(x) > t\} \text{ offen} \\ = f^{-1}((t,\infty))}} dt \overset{(iii)}{\leq} \int_0^\infty \liminf P(X_n \in G) dt \tag{2.59}$$

$$\overset{\text{Fatou}}{\leq} \liminf \int_0^\infty P(f(X_n) > t) dt = \liminf E[f(X_n)]$$

Ist f beschränkt, etwa $|f| \leq c$, so folgt $-f + c \geq 0$, also

$$\begin{aligned} \limsup E[f(X_n)] &= c - \liminf E[-f(X_n) + c] \overset{(2.59)}{=} c - E[-f(X) + c] \\ &= E[f(X)] = E[f(X) + c] - c \\ &\overset{(2.59)}{=} \liminf E[f(X_n) + c] - c = \liminf E[f(X_n)] \end{aligned}$$

und sonst (i).

$(iv) \Rightarrow (v)$: Für $B \in \mathscr{B}(E)$:

$$P(X \in B^0) \overset{(iii)}{\leq} \liminf \underbrace{P(X_n \in B^0)}_{\leq P(X_n \in \overline{B})} \leq \limsup P(X_n \in \overline{B}) \leq P(X \in \overline{B}).$$

Ist $P(X \in \partial B) = 0$, so folgt $P(X \in B^0) = P(X \in \overline{B})$, also $P(X_n \in B) \to P(X \in B)$.

$(v) \Rightarrow (iv)$: Sei $F \subseteq E$ abgeschlossen. Wir setzen $F^\varepsilon := \{x \in E : r(X \leq F) \leq \varepsilon\}$ für $\varepsilon > 0$. Die Mengen $\partial F^\varepsilon \subseteq \{x : r(X, F) = \varepsilon\}$ sind disjunkt, also gilt

$$P(X \in \partial F^\varepsilon) = 0$$

für Lebesgue-fast alle ε. Wir wählen eine Folge $(\varepsilon_k) \downarrow 0$, so dass $\varepsilon \Rightarrow$

$$\begin{aligned} \limsup P(X_n \in F) &\leq \inf_{k \geq 1} \limsup P(X_n \in F^{\varepsilon_k}) \\ &= \inf_{k \geq 1} P(X \in F^{\varepsilon_k}) \underset{\substack{\uparrow \\ \text{Stetigkeit von oben} \\ F \text{ abg.}}}{=} P(X \in F). \end{aligned}$$

$\square$

Korollar 2.60. *Seien $X, X_1, X_2, \ldots$ reellwertige Zufallsvariable mit Verteilungsfunktion $F, P_1, F_2, \ldots$ Dann*

$$X_n \Rightarrow X \Leftrightarrow F_n(x) \to F(x) \text{ für alle Stetigkeitsstellen } x \text{ von } F.$$

Beweis. „$\Rightarrow$“ Ist X Stetigkeitsstelle von F, so ist $P(X \in \partial(-\infty, x]) = P(X = x) = 0$

$$\overset{(v)}{\Rightarrow} \qquad F_n(x) = P(X_n \in (-\infty, x]) \to F(x).$$

„$\Leftarrow$“ Wir verwenden (ii) des vorigen Theorems: Ohne Einschränkung sei $|f| \leq 1$, f stetig mit Lipschitz-Konstante 1. Für $\varepsilon > 0$ wählen wir $N \in \mathbb{N}$ und Stetigkeitspunkte $y_1, \ldots, y_N$ von F so, dass

$$F(y_0) < \varepsilon, \qquad F(y_N) > 1 - \varepsilon$$

und $y_i - y_{i-1} < \varepsilon$, $i = 1, \ldots, N$. Dann gilt nach Voraussetzung $F_n(y_i) \to F(y_i)$ und

$$f \leq \mathbb{1}_{(-\infty, y_0)} + \mathbb{1}_{(y_N, \infty)} + \sum_{i=1}^{N-1} (f(y_i) + \varepsilon) \mathbb{1}_{(y_i, y_{i+1}]}.$$

Es folgt, dass

$$\begin{aligned}\limsup_{n\to\infty} E[f(X_n)] &\leq \limsup \Big[F_n(y_0) + 1 - F_n(y_N) + \sum_{i=1}^{N-1} (f(y_i) + c)\big(F_n(y_{i+1} - F_n(y_i))\big)\Big] \\ &\leq 3\varepsilon + \sum_{i=1}^{N-1} f(y_i)\big(F(y_{i+1}) - F(y_i)\big) \\ &\leq 4\varepsilon + E[f(x)].\end{aligned}$$

Verwenden wir $1 - f$, so erhalten wir eine analoge Aussage für den $\liminf$ und die Behauptung folgt mit $\varepsilon \to 0$. □

Korollar 2.61 (Slutsky)**.**

$$X_n \Rightarrow X, \qquad d(X_n, Y_n) \xrightarrow{P} 0 \Rightarrow Y_n \Rightarrow X.$$

Beweis. Übungsaufgabe □

Theorem 2.62 (Continuous mapping Theorem)**.** Seien (E, d) und (E', d') metrische Räume, $f : E \to E'$ messbar und $\mu_f = \{x \in E : f \text{ nicht stetig an } X\}$ messbar.

(i) Gilt $P \Rightarrow P$ und $P(U_f) = 0$, so folgt $f_* P_n \Rightarrow f_* P$

(ii) Gilt $X_n \Rightarrow X$ und $P(X \in U_f) = 0$, so folgt $f(X_n) \Rightarrow f(X)$

Beweis. (ii) folgt direkt aus (i) mit $P_n = X_{n*} P$

Für Teil (i) betrachte $G \subseteq E'$ offen und $x \in f^{-1}(G) \cap U_f^c$.

Da f stetig in x ist, ist $f^{-1}(G)$ offen und so existiert $\delta > 0$ dass $f(y) \in G$ für alle $y \in E$ mit $d(x,y) < \delta$.

Dann ist $\varphi^{-1}(G) \cap U_f^c \subseteq (\varphi^{-1}(G))^o$ und mit dem Portmanteau-Theorem 2.58 ($i \Rightarrow iii$) folgt

$$\begin{aligned} f_*P(G) = P(f^{-1}(G) &\overset{P(\mu_f)=0}{=} P(f^{-1}(G) \cap U_f^c) \leq P((f^{-1}(G))^o) \\ &\leq \liminf_{n\to\infty} P_n((f^{-1}(G))^o) \leq \liminf_{n\to\infty} P_n((f^{-1}(G))) \\ &= \liminf f_*P_n(G) \end{aligned}$$

und wieder mit dem Portmanteau-Theorem 2.58 ($iii \Rightarrow i$) folgt die Behauptung. □

2.5.1 Straffheit und relative Kompaktheit

Nun betrachten wir wieder einen allgemeinen metrischen Raum (E,r). Mit $\mathcal{K}$ bezeichnen wir die kompakten Mengen in E.

Definition 2.63. Eine Familie $(P_i)_{i\in I} \subseteq \mathscr{M}(E)$ heißt *straff*, falls

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} P_i(K) = 1.$$

Eine Familie $(x_i)_{i\in I}$ von Zufallsvarialben heißt *straff*, falls $(X_{i*}P)_{i\in I}$ straff sind, also

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} P(X_i \in K) = 1.$$

Bemerkung 2.64. Äquivalente Formulierungen:

(i) Für alle $\varepsilon > 0$ gibt es $K_E \in \mathcal{K}$, so dass $\inf_{i\in I} P_i(K) \geq 1 - \varepsilon$.

(ii) Ist $E = \mathbb{R}^d$, so ist $(P_i)_{i\in I}$ straff $\Leftrightarrow$

$$\sup_{r>0} \inf_{i\in I} P_i(B_r(0)) = \mathbb{1} \qquad (B_r(0) = \{x : \|x\| \leq r\})$$

(iii) Wir haben gezeigt, dass P straff ist, falls (E,r) vollständig und separabel ist. Damit folgt dies für jede endliche Familie von Wahrscheinlichkeitsmaßen.

Lemma 2.65. *Eine abzählbare Familie $(P_n)_{n\geq 1} \subseteq \mathscr{M}(E)$ auf einem polnischen Raum E ist genau dann straff, wenn*

$$\sup_{K\in\mathcal{K}} \liminf_{n\geq 1} P_n(K) = 1. \tag{2.66}$$

Beweis. **„⇒“:** klar, da $\liminf P_n(K) \geq \inf P_n(K)$

„⇐“: Sei $\varepsilon > 0$ und K so, dass $\liminf P_n(K) \geq 1 - \frac{\varepsilon}{2}$.

Wir wählen n_0 so, dass $\inf_{n\geq n_0} P_n(K) \geq 1 - \varepsilon$ und $K_1, \ldots, K_{n_0} \in \mathcal{K}$ so, dass $P_n(K_n) \geq 1 - \varepsilon$, $n = 1, \ldots, n_0$. Da $K' = K_1 \cup \cdots \cup K_{n_0} \in \mathcal{K}$ wieder kompakt ist und $\inf P_n(K') \geq 1 - \varepsilon$ folgt die Straffheit. (Zu jedem K finde ich $K' \geq K$, so dass $\inf P_n(K') \geq 1 - \varepsilon$.) □

Beispiel 2.67. (i) Ist E kompakt, so ist jede Familie von Wahrscheinlichkeitsmaßen kompakt.

(ii) Sind $(X_i)_{i\in I}$ reellwertig mit

$$\sup_{i\in I} E[|X_i|] < \infty,$$

so ist (X_i) straff: Denn es gilt

$$\inf_{r>0} \sup_{i\in I} P(|X_i| \geq r) \leq \inf_{r>0} \sup_{i\in I} \frac{E|X_i|]}{r} = 0$$

(iii) $(\delta_n)_{n\geq 1}$ ist nicht straff.

(iv) Verteilungskonvergente Folgen sind straff.

(v) Die Menge der Normalverteilungen $\{\mathcal{N}(\mu, \sigma^2).(\mu, \sigma) \in \Theta\}$ ist genau dann straff wenn Θ beschränkt ist.

Lemma 2.68. *Seien $P_1, P_2, \ldots \in \mathscr{P}(\mathbb{R})$ und $\mu < \mathscr{P}_{\leq 1}(\mathbb{R})$ mit*

$$P_n \Rightarrow_v \mu.$$

Dann ist $\mu(\mathbb{R}) = 1$ genau dann, wenn (P_n) straff ist. In diesem Fall folgt $P_n \Rightarrow \mu$.

Korollar 2.69. *Seien $P, P_1, P_2, \ldots \in \mathscr{P}(\mathbb{R})$. Gilt $P_n \Rightarrow P$, so ist (P_n) straff.*

Definition 2.70. Auf einem topologischen Raum $(\Omega, \mathcal{O})$ heißt $K \subseteq \Omega$

(i) *kompakt*, falls jede offene Überdeckung von K eine endliche Teilüberdeckung besitzt,

(ii) *relativ kompakt*, falls $\overline{K}$ kompakt ist,

(iii) *relativ folgenkompakt*, falls es für jede Folge $\omega_1, \omega_2, \ldots \in K$ eine konvergente Teilfolge gibt.

(iv) Ist d eine Metrik, die $\mathcal{O}$ erzeugt, so heißt $K \subset \Omega$ *total beschränkt*, falls es für jedes $\varepsilon > 0$ ein $N \in \mathbb{N}$ und $\omega_1, \ldots, \omega_N \in K$ gibt, so dass $K \subseteq \bigcup_{n=1}^{N} \mathscr{B}_\varepsilon(\omega_n)$.

Die Menge K heißt folgenkompakt, falls jede Folge $\omega_1, \omega_2, \ldots \in K$ eine konvergente Teilfolge mit Grenzwert $k \in K$ hat - im Unterschied dazu haben in einer *relativ* folgenkompakten Menge die Teilfogen Grenzwerte in $\overline{K}$.

Satz 2.71. *Sei (E, d) ein metrischer Raum und $\mathcal{O}$ die von d erzeugte Topologie.*

(i) K ist relativ kompakt

(ii) Sind $F_i \subseteq \overline{K}$ abgeschlossen, $i \in I$, $\bigcap_{i \in I} F_i = \varnothing$, dann gibt es endliches $I \leq \mathcal{I}$ mit $\bigcap_{i \in J} F_i = \varnothing$.

(iii) K ist relativ folgenkompakt.

(iv) K ist total beschränkt.

Dann gilt (iv) $\Leftarrow$ (i) $\Leftrightarrow$ (ii) $\Rightarrow$ (iii).

Ist (E, d) separabel, so folgt (iii) $\Rightarrow$ (ii).

Ist (E, d) vollständig, so folgt (iv) $\Rightarrow$ (iii).

Ist der Raum (E, d) also polnisch, so sind die Eigenschaften äquivalent.

Beweis. (i) $\Rightarrow$ (iv): Sei $\overline{K}$ kompakt und $\varepsilon > 0$.

Wir wählen eine offene Überdeckung $\bigcup_{\omega \in K} B_E(\omega)$ von $\overline{K}$, also gibt es eine endliche Teilüberdeckung, d.h. es gibt $\omega_1, \ldots, \omega_N$, so dass $\overline{K} \subseteq \bigcup_{n=1}^{N} B_E(\omega_n)$. Da ε beliebig war, folgt, dass K total beschränkt ist.

(i) $\Rightarrow$ (ii): Seien $F_i \subseteq \overline{K}$ abgeschlossen und $\bigcap_{i \in J} F_i = \varnothing$. Dann ist $\bigcup F_i^c = (\bigcap F_i)^c = \Omega$. Dies ist natürlich eine offene Überdeckung von $\overline{K}$, also gibt es endliches $I \subseteq J$, so dass $\overline{K} \subseteq \bigcup_{j \in J} F_j^c$.

Außerdem ist dann $\left(\bigcup_{j\in J} F_j^c\right)^c = \bigcap_{j\in J} F_j \subseteq \overline{K}^c$, aber $F_i \subseteq \overline{K}$, also muss $\bigcap_{j\in J} F_j = \varnothing$ sein.

(ii) ⇒ (i): Seien O_i offen und $\overline{K} \subseteq \bigcup_{i\in J} O_i$. Wir setzen $F_i = O_i^c \cap \overline{K}$, so dass $F_i^c \in \mathcal{O}$ und $\bigcap_{i\in J} F_i = \overline{K} \cap (\bigcup O_i)^c = \varnothing$. Also gibt es nach Voraussetzung (ii) endliches $I \subseteq J$ mit $\bigcap_{j\in J} F_j = \varnothing$. Dann ist

$$\overline{K}^c \cup \bigcup_{j\in J} O_i = \bigcup_{j\in J} F_j^c = \Omega,$$

also $\overline{K} \subseteq \bigcup_{j\in J} O_j$ und somit $\overline{K}$ kompakt.

(ii) ⇒ (iii): Seien $\omega_1, \omega_2, \ldots \in K$. Wir setzen $F_n = \overline{\{\omega_n, \omega_{n-1}, \ldots\}} \subseteq \overline{K}$. Angenommen es gibt keine konvergente Teilfolge von (ω_n). Dann ist $\bigcap_{n=1}^\infty F_n = \varnothing$.

Mit (ii) folgt dann, dass es ein $N \in \mathbb{N}$ gibt mit $\varnothing = \bigcap_{n=1}^N F_n = F_N$, da $F_N \neq \varnothing$.

(iii) ⇒ (i): Diesen Teil unterteilen wir in zwei Schritte: Erstens, falls (E,d) separabel ist, erhalten wir: Sei E' abzählbar mit $\overline{E'} = E$ und $\mathscr{B} = \{B_{\frac{1}{n}}(\omega) : \omega \in E', n \in \mathbb{N}\}$. Dann ist $\mathscr{B}$ abzählbare Basis von $\mathcal{O}$, und wir schreiben $\mathscr{B} = \{C_1, C_2, \ldots\}$.

Zweitens: Angenommen $\overline{K}$ ist nicht kompakt: Dann gibt es $A_i \in \mathcal{O}$, $i \in J$ mit $\overline{K} \subseteq \bigcup_{i\in J} A_i$ und es gibt keine endliche Teilüberdeckung. Wir setzen für $i \in J$

$$J_i = \{j \in \mathbb{N} : C_j \subseteq A_i\} \subseteq \mathbb{N}$$

sowie $J := \bigcup_{i\in I} \partial_i \subseteq \mathbb{N}$. Dann ist $A_i = \bigcup_{j\in J_i} C_j$, also

$$K \subseteq \bigcup_{i\in J} A_i = \bigcup_{i\in J} \bigcup_{j\in J_i} c_j = \bigcup_{j\in J} C_j.$$

Damit sind die (C_j) eine offene Überdeckung. Da es für (A_i) keine Teilüberdeckung gibt, gibt es ebenso keine für die (C_i). Wir setzen $\omega_n \in \overline{K} \setminus \bigcup_{j\in J, j\leq n} C_j$. Da K relativ folgenkompakt ist, hat (ω_n) einen Häufungspunkt in $\omega \in \overline{K}$. Da $\overline{K} \underset{j\leq J}{\subseteq} C_j$, gibt es $k \in J$ mit $\omega \in C_k$. Da C_k offen liegen unendlich viele der w_n in C_k, andererseits ist $\omega_i \notin c_k$, $i \geq k$. ↯

(iii) ⇒ (iv): Seien $\omega_1, \omega_2, \ldots \in K$. Wir konstruieren eine Teilfolge, die eine Cauchyfolge ist. Da E vollständig ist, ist die Folge auch konvergent und somit K relativ folgenkompakt.

Wir wählen $E_n \downarrow 0$. Da K total beschränkt ist, gibt es $N \in \mathbb{N}$ ε_1-Bälle, die K überdecken. Mindestens einer dieser Bälle enthält unendlich viele der (ω_1). Diese haben dann höchstens den Abstand $2\varepsilon_1$. Wir wählen ein ω_{k_1} davon. Dieser ε_1-Ball wird durch endlich viele ε_2-Bälle überdeckt. Mindestens einer dieser Bälle enthält unendlich viele der (ω_i) und wir wählen $\omega_{k_2} \neq \omega_{k_1}$. Iterieren liefert (ω_{k_n}) mit $d(\omega_{k_n}, \omega_{k_m}) \leq 2\varepsilon_{m\wedge n}$. □

2.6 Der Satz von Prochorow

Theorem 2.72 (Prochorow)**.** Sei (E,d) polnisch. Dann ist $(P_i)_{i\in I}$ genau dann straff, wenn $(P_i)_{i\in I}$ relativ kompakt bzgl. der schwachen Topologie ist.

Hierbei heißt relativ kompakt bezüglich der schwachen Topologie (da wir auf einem polnischen Raum arbeiten) , dass jede Folge eine schwach konvergente Teilfolge hat. Relative Kompaktheit ist ein wichtiger Begriff und der Satz von Prochorow schlägt die Brücke zur Analysis (über den Satz von Arzela–Ascoli).

Der folgende Beweis enthält auch die nützliche Tatsache, dass aus relativer Kompaktheit Gleichung (2.73) folgt.

Beweis. Wir zeigen zunächst: Aus relativer Kompaktheit folgt $\forall \varepsilon > 0\ \exists N \in \mathbb{N}$ und $x_1, \dots, x_N \in E$ s. d. (B_ε offener Ball)

$$\inf_{i\in I} P_i\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_k)\Big) \geq 1-\varepsilon. \tag{2.73}$$

Angenommen, das gilt nicht. Dann existiert $\varepsilon > 0$ so dass für alle $x_1, \dots, x_N$, $N \in \mathbb{N}$ ein P_{i_N} existiert mit

$$P_{i_N}\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_k)\Big) \geq 1-\varepsilon.$$

Da nach Voraussetzung (P_i) relativ kompakt ist, also auch relativ folgenkompakt (nach Satz 2.71) ist, also gibt es eine Teilfolge (P_{i_N}) die schwach gegen ein $P \in \mathscr{M}(E)$ konvergiert. Wegen der Separabilität von E und des Portmanteau-Theorems gilt

$$1 = P(E) = \sup_{N\in\mathbb{N}} P\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_i)\Big) \leq \sup_{N\in\mathbb{N}} \liminf_{M\to\infty} P_{i_{N_M}}\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_i)\Big) \leq 1-\varepsilon,$$

ein Widerspruch.

Nun zeigen wir: aus (2.73) folgt Straffheit. Sei dazu $\varepsilon > 0$. Für $j \in \mathbb{N}$ wählen wir $x_1^j, \dots, x_{n_j}^j$ so, dass

$$\inf_{i\in I} P_i\Big(\bigcup_{k=1}^{n_j} B_{\frac{\varepsilon}{2^j}}(x_k^j)\Big) > 1-\frac{\varepsilon}{2^j}.$$

Wir setzen

$$K := \bigcap_{j=1}^{\infty} \bigcup_{k=1}^{n_j} B_{\frac{\varepsilon}{2^j}}(x_k^j).$$

Dann ist K total beschränkt. Da E vollständig und separabel ist, ist K relativ kompakt nach Satz 2.71, also $\overline{K}$ kompakt. Außerdem ist

$$\sup_{i\in I} P_i(\overline{K}^c) \leq \sup_{i\in I} \sum_{j=1}^{\infty} P_i\Big(\bigcap_{k=1}^{n_j} (B_{\frac{\varepsilon}{2^j}}(x_k^j))^c\Big) \leq \varepsilon \sum_{j=1} \frac{1}{2^j} = \varepsilon,$$

also ist $(P_i)_{i\in I}$ straff.

Im Folgenden betrachten wir die Umkehrung. Ziel ist es, zu zeigen, dass $(P_i)_{i\in I}$ relativ folgenkompakt ist. Sei also $(Q_n)_{n\geq 1}$ eine Teilfolge von $(P_i)_{i\in I}$. Da (P_i) straff ist, können wir kompakte Mengen $K_1 \subseteq K_2 \subseteq \cdots$ so wählen dass für jedes $j \geq 1$

$$\inf_{n\geq 1} Q_n(K_j) \geq 1 - \frac{1}{j}.$$

Nun möchten wir messbare Mengen geeignet durch diese kompakten Mengen approximieren. Da der Raum (E, d) separabel ist, gibt es abzählbar viele $x_1, x_2, \dots$ und $\varepsilon_1, \varepsilon_2, \dots$, so dass die Menge $\{(B_{\varepsilon_k}(x_k))_{k\geq 1} : k \in \mathbb{N}\}$ die σ-Algebra $\mathscr{B}(E)$ erzeugt. Wir setzen[5]

$$\mathscr{H} := \Big\{ \bigcup_{k=1}^{N} K_{j_k} \bigcap \overline{B_{\varepsilon_k}(x_k)} : N, \in \mathbb{N}, (j_k)_{k\geq 1} \subseteq \mathbb{N} \Big\}.$$

Dann ist $\mathscr{H}$ ein Halbring und ein durchschnittstabiler Erzeuger von $\mathscr{B}(E)$. Da $\mathscr{H}$ abzählbar ist, finden wir eine Teilfolge (Q_{n_k}), so dass $(Q_{n_k}(A))$ für alle $A \in \mathscr{H}$ konvergiert (denn $[0, 1]$ ist kompakt).

Wir setzen $\mu(A) = \lim_{k\to\infty} Q_{n_k}(A)$, $A \in \mathscr{H}$. Diesen Inhalt kann man eindeutig zu einem Maß μ auf $\mathscr{B}(E)$ fortsetzen. Dann gilt, dass

$$1 \geq \mu(E) \geq \sup_{j\geq 1} \mu(K_j) = \sup_{j\geq 1} \lim_{k\to\infty} Q_{n_k}(K_j) \geq \sup_{j\geq 1} \Big(1 - \frac{1}{j}\Big) = 1.$$

Für jedes offene G gilt, dass

$$\mu(G) = \sup_{H\in\mathscr{H}: H\subset G} \mu(H) = \sup_{H\in\mathscr{H}: H\subset G} \lim_{k\to\infty} Q_{n_k}(H) \leq \liminf Q_{n_k}(G).$$

Damit erhalten wir $Q_{n_k} \Rightarrow P$ aus dem Portmanteau-Theorem 2.58 und somit die Behauptung. □

Auf einem polnischen Raum kann man die volle Wahrscheinlichkeit durch kompakte Mengen ausschöpfen, was sich als sehr nützlich erweisen wird.

[5] $(\overline{B}_\varepsilon(x) = \{\|x' - x\| \leq \varepsilon\})$.

Lemma 2.74. *Ist* $(\Omega, \mathcal{O})$ *polnisch und* μ *ein endliches Maß auf* $\mathscr{B}(\mathcal{O})$, *so existiert zu jedem* $\varepsilon > 0$ *eine kompakte Menge* K, *so dass* $\mu(\Omega \setminus K) < \varepsilon$.

Beweis: Sei $\varepsilon > 0$. Aus der Separabilität folgt

$$\Omega = \bigcup_{k=1}^{\infty} B_{\frac{1}{n}}(\omega_k^n) \quad \text{für alle } n \geq 1 \text{ und geeignete } (\omega_k^n)_{k \geq 1}.$$

Mit Stetigkeit von oben erhält man

$$0 = \mu\Big(\Omega \setminus \bigcup_{k=1}^{\infty} B_{\frac{1}{n}}(\omega_k^n)\Big) = \lim_{N \to \infty} \mu\Big(\Omega \setminus \bigcup_{k=1}^{N} B_{\frac{1}{n}}(\omega_k^n)\Big).$$

Demnach gibt es $N_n \in \mathbb{N}$ so dass $\mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega_k^n)\Big) \leq \frac{\varepsilon}{2^n}$.
Die Menge $A := \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega_k^n)$ ist total beschränkt, also ist $\overline{A}$ kompakt nach Satz 2.71[6]

$$\Rightarrow \qquad \begin{aligned} \mu(\Omega \setminus \overline{A}) \leq \mu((\Omega \setminus A) &= \mu\Big(\bigcup_{k=1}^{\infty} \Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega - k^n)\Big)\Big) \\ &\leq \sum_{n=1}^{\infty} \mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega_k^n)\Big) \leq \varepsilon. \end{aligned}$$

□

Als Folgerung erhalten wir, dass auf einem polnischen Raum jede ein-elementige Familie (P) straff ist.

[6] Hierbei nutzen wir, dass

$$\Omega \cap \Big(\bigcup \Big(\bigcup B_{\frac{1}{n}}\Big)^c\Big) = \bigcup \Omega \setminus \Big(\bigcup B_{\frac{1}{n}}\Big)$$

3. Der Zentrale Grenzwertsatz

In diesem Abschnitt beweisen wir den zentralen Grenzwertsatz von Lindeberg und Feller. Das wesentliche Hilfsmittel werden die charakteristischen Funktionen sein, die wir zunächst einführen. Anschließend beweisen wir den Satz von LÉvy, welcher schwache Konvergenz mit Hilfe von charakteristischen Funktionen (oder allgemeiner: separierenden Funktionenklassen) verknüpft.

3.0.1 Charakteristische Funktionen

Definition 3.1. Eine Menge $\mathcal{M} \subseteq C(E)$ von stetigen Funktionen auf E heißt *punktetrennend*, falls für alle $x \neq y \in E$ ein $f \in \mathcal{M}$ existiert, so dass $f(x) \neq f(y)$.

$\mathcal{M}$ heißt *separierend*, falls aus $P, Q \in \mathscr{M}(E)$ und $E_P[f] = E_Q[f]$ für alle $f \in \mathcal{M}$ folgt, dass $P = Q$.

Beispiel 3.2. $C_b(E)$ ist punktetrennend und separierend: Wähle $f(x) = d(x, y) \wedge 1$. Ist $P \neq Q$, so gibt es einen offenen Ball A mit $P(A) \neq Q(A)$. Wir wählen $f_n \uparrow \mathbb{1}_A$ und erhalten $P(A) = \lim E_P[f_n]$ und $Q(A) = \lim E_Q[f_n]$. Wäre $C_b(E)$ nicht separierend, so erhielten wir $Q(A) = P(A)$, ein Widerspruch.

Wir nennen ein Mengensystem $\mathcal{M} \subset \mathcal{C}(E)$ **Algebra**, falls

(i) $1 \in \mathcal{M}$

(ii) $\alpha f + \beta g \in \mathcal{M}$ für $\alpha, \beta \in \mathbb{R}$ und $f, g \in \mathcal{M}$

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

Theorem 3.3 (Stone-Weierstraß)**.** Sei (E, d) kompakt und $\mathcal{M} \subseteq C_b(E)$ eine punktetrennende Algebra. Dann liegt $\mathcal{M}$ dicht in $C_b(Z)$ bzgl. der Supremumsnorm.

Theorem 3.4. Sei (E, d) vollständig und separabel. Ist $\mathcal{M} \subseteq C_b(E)$ punktetrennend und folgt aus $f, g \in \mathcal{M}$ dass $fg \in \mathcal{M}$, dann ist M separierend.

Wir erhalten direkt

Satz 3.5. *Ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$ wird eindeutig durch die Fourier-Transformierte bestimmt, denn e^{iux}, $u \in \mathbb{R}^d$ ist eine separierende Familie auf dem $\mathbb{R}^d$.*

Beweis zu Theorem 3.4: Ohne Einschränkung ist $\mathcal{M}$ eine Algebra, denn aus $E_Q[f] = E_P[f]$ und ebenso für G folgt $\alpha E_Q[f] + \beta E_Q[f] = \alpha E_P[f] + \beta E_P[f]$, sowie $E_P[1] = E_Q[1]$.

Nach Lemma 2.74 können wir K kompakt wählen, so dass $P(K) > 1-\varepsilon$ und $Q(K) > 1-\varepsilon$. Nach dem Stone-Weierstraß-Theorem gibt es für $g \in C_b(E)\quad (g_n) \subseteq \mathcal{M}$ so dass

$$\sup_{x\in K} |g_n(x) - g(x)| \underset{n\to\infty}{\rightarrow} 0. \tag{3.6}$$

Wir betrachten $ge^{-\varepsilon g^2}$, so dass außerhalb von K diese Funktionenklasse schnell verschwindet. Durch Ableiten und Null setzen errechnet man das folgende Maximum

$$\sup_{x\in\mathbb{R}} xe^{-\varepsilon x^2} = e^{-1/2}\sqrt{\frac{1}{2\varepsilon}}.$$

Wir erhalten, dass

$$|E[ge^{-\varepsilon g^2}] - E[ge^{-\varepsilon g^2} 1_K]| \leq C\varepsilon^{-1/2} \cdot P(K^c) \leq C \cdot \sqrt{\varepsilon} \underset{\varepsilon\to 0}{\rightarrow} 0. \tag{3.7}$$

Wir schätzen ab:

$$
\begin{aligned}
& \left|E_P[ge^{-\varepsilon g^2}] - E_Q[ge^{-\varepsilon g^2}]\right| \\
(1)\quad & \leq \left|E_P[ge^{-\varepsilon g^2}] - E_P[ge^{-\varepsilon g^2}\mathbb{1}_K]\right| \\
(2)\quad & \quad + \left|E_P[ge^{-\varepsilon g^2}\mathbb{1}_K] - E_P[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K]\right| \\
(3)\quad & \quad + \left|E_P[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K] - E_P[g_n e^{-\varepsilon g_n^2}]\right| \\
(4)\quad & \quad + \left|E_P[g_n e^{-\varepsilon g_n^2}] - E_Q[g_n e^{-\varepsilon g_n^2}]\right| \\
(5)\quad & \quad + \left|E_Q[g_n e^{-\varepsilon g_n^2}] - E_Q[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K]\right| \\
(6)\quad & \quad + \left|E_Q[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K] - E_Q[ge^{-\varepsilon g^2}\mathbb{1}_K]\right| \\
(7)\quad & \quad + \left|E_Q[g_n e^{-\varepsilon g^2}\mathbb{1}_K] - E_Q[ge^{-\varepsilon g^2}]\right|
\end{aligned}
$$

Dann gilt (1), (3), (5), (7) $\rightarrow 0$ wegen (3.7) und (2) und (6) $\rightarrow 0$ wegen (3.6). Da $\mathcal{M}$ Algebra ist, kann $g_n e^{-\varepsilon g_n^2}$ durch Funktionen f_ε in $\mathcal{M}$ approximiert werden (Stone-Weierstraß). Es folgt

$$|E_P[g] - E_Q[g]| \leq \lim_{\varepsilon\to 0} f_\varepsilon = 0.$$

Da $g \in \mathcal{C}_b(E)$ beliebig und $C_b(E)$ separierend folgt $P = Q$. $\square$

Bemerkung 3.8. Wir erhalten unmittelbar

(i) Auf dem $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ bestimmen charakteristische Funktionen eindeutig das Wahrscheinlichkeitsmaß.

(ii) Eine Familie reellwertiger Zufallsvariablen $(X_i)_{i\in J}$ ist genau dann unabhängig, wenn für alle endlichen $J \subseteq I$

$$E\Big[\prod_{j\in J} e^{iu_jX_j}\Big] = \prod_{j\in J} E[e^{iu_jX_j}], \quad (u_j)_{j\in J} \subseteq \mathbb{R}^0.$$

3.1 Der Satz von Lévy

Als nächsten großen Schritt beweisen wir den Satz von Lévy. Der Satz von Lévy wird schwache Konvergenz vollständig mit Hilfe von charakteristischen Funktionen klassifizieren.

Satz 3.9. *Sei (E,d) vollständig und separabel und $P, P_1, P_2, \ldots \in \mathscr{M}(E)$.*

Dann sind äquivalent:

(i) $P_n \Rightarrow P$

(ii) $(P_n)_{n\geq 1}$ *ist straff und es gibt eine separierende Familie* $\mathcal{M} \subseteq \mathcal{C}_b(Z)$, *s. d.*

$$P_n(f) \to P(f) \quad \forall f \in \mathcal{M}.$$

Beweis. „⇒“ klar

„⇐“ Angenommen (P_n) ist straff, aber $P_n \not\Rightarrow P$.
Dann gibt es $\varepsilon > 0$, $f \in \mathcal{C}_b(Z)$ und eine Teilfolge, so dass

$$\big|E_{P_{n_k}}[f] - E_P[f]\big| > \varepsilon \quad \text{für alle } k. \tag{3.10}$$

$\Rightarrow$ Nach dem Satz von Prohorov gibt es eine Teilfolge $(P_{n_{k_l}})$ der straffen Familie $(P_{n_k})_{k\geq 1}$, so dass $P_{n_{k_l}} \overset{l\to\infty}{\Rightarrow} Q \in \mathscr{M}(E)$.

Dann ist

$$|E_P[f] - E_Q[f]| \geq \liminf\Big(E_P[f] - E_{P_{n_{k_l}}}[f]\Big) + \liminf\Big(E_{P_{n_{k_l}}}[f] - E_{@}[f]\Big) > \varepsilon$$

nach (3.10), also $P \neq Q$. Andererseits gilt nach Voraussetzung $\forall g \in \mathcal{M}$, dass

$$E_P[g] = \lim_{l\to\infty} E_{P_{n_{k_l}}}[g] = E_Q[g],$$

ein Widerspruch daszu, dss $\mathcal{M}$ separierend ist. $\square$

Im unendlichdimensionalen Banachräumen sind abgeschlossene Kugeln nicht mehr kompakt (dies folgt aus dem Kopaktheitssatz von Riesz: ein normierter Vektorraum ist endlichdimensional genau dann, wenn $\overline{B}_1$ kompakt ist).

Der Satz von Arzela-Ascoli verknüpft Kompaktheit elegant mit folgendem Stetigkeitsbegriff:

Definition 3.11. Eine Familie von reellwertigen Funktionen $\mathcal{M}$ auf dem $\mathbb{R}^d$ heißt *gleichgradig stetig in* $x \in \mathbb{R}^d$, falls

$$\sup_{f\in\mathcal{M}} |f(x) - f(y)| \xrightarrow{y\to x} 0.$$

Dies kann man natürlich auch allgemein für Funktionen auf metrischen Räumen definieren. (**Gleichmäßige** Stetigkeit bezieht sich allerdings nur auf **eine** Funktion!)

Bemerkung 3.12. Handelt es sich um eine Folge (f_n) von Funktionen, so ist gleichgradige Stetigkeit äquivalent zu

$$\limsup_{n\to\infty} |f_n(x) - f_n(y)| \xrightarrow{y\to x} 0.$$

Wir erhalten folgendes Lemma:

Lemma 3.13. *Sei* $(f_n) \subseteq \mathcal{C}(\mathbb{R}^d)$ *und gelte* $f_n(t) \to f(t)$ $\forall t \in \mathbb{R}^d$.

Dann ist (f_n) *genau dann gleichgradig stetig in* 0*, wenn* f *in* 0 *stetig ist.*

Beweis. **„⇒“:** $|f(x) - f(0) = |\underbrace{\lim f_n(t) - \lim f_n(0)}_{=\lim(f_n(t)-f_n(0))}| \leq \limsup |f_n(t) - f_n(0)| \to 0,$

da (f_n) gleichgradig stetig ist in 0.

„⇐“: $\limsup |f_n(t) - f_n(0)| \leq \limsup_{n\to\infty} \big(|f_n(t) - f(t)| + |f(t) - f(0)| + |f(0) - f_n(0)|\big)$

$$\overset{f_n(t)\to f(t)}{=} |f(t) - f(0)| \overset{t\to 0}{\to} 0. \qquad \square$$

Wir benötigen noch eine handliche Abschätzung für die Fourier-Transformierte.

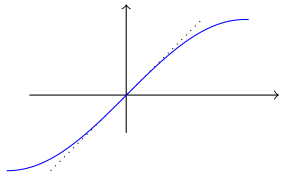
Lemma 3.14. *Für alle* $r > 0$ *und* $P \in \mathscr{P}(\mathbb{R})$

$$P\Big((-\infty, -r] \cup [r, \infty)\Big) \leq \frac{r}{2} \int_{-2/r}^{2/r} (1 - \varphi_P(t))dt.$$

Beweis. Wir betrachten X mit Verteilung P. Dann ist

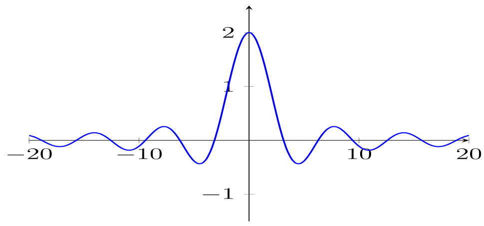
$$\int_{-c}^{c}(1-\varphi_P(t)dt = E\Big[\int_{-c}^{c}(1-e^{itX})dt\Big] = 2c - E\Big[\frac{1}{iX}e^{itX}\,\Big|_{-c}^{c}\Big], \tag{3.15}$$

da $e^{ix} = \cos x + i\sin x$ ist, und cos symmetrisch ist um Null, folgt $\cos t\,|_{-\xi cX}^{\xi cX} = 0$, $\sin t\,|_{-\xi cX}^{\xi cX} = 2\sin cX$, also ist



$$\text{Gleichung (3.15)} \;= 2c\Big(E\Big[1-\frac{\sin cX}{cX}\Big]\Big).$$

Nun ist $\frac{\sin x}{x}\begin{cases} > 0 & \text{für } |x|\le 2,\\ \le \frac{1}{2} & \text{für } |x|\ge 2,\end{cases}$
also $1-\frac{\sin x}{x}\ge\frac{1}{2}$ für $|x|\ge 2 \Rightarrow$



$$\begin{aligned}\text{Gl. (3.15)} \quad &\ge 2cE\Big[\Big(1-\frac{\sin cX}{cX}\Big)\mathbb{1}_{\{|cX|\ge 2\}}\Big]\\ &\ge c\cdot P(|cX|\ge 2).\end{aligned}$$

Die Behauptung folgt nun mit $c=\frac{2}{r}$. □

Satz 3.16 (Arzelà-Ascoli)**.** *Sei $(P_i)_{i\in I}\subseteq\mathscr{M}(\mathbb{R}^d)$. (P_i) ist gleichgradig stetig in 0 genau dann, wenn $(P_i)_{i\in I}$ straff ist.*

Beweis: Wir beweisen lediglich die eine Richtung: gleichgradige Stetigkeit an 0 impliziert Straffheit.

Es genügt zu zeigen, dass $(\prod_k, P_i)_{i\in I}$ für alle Projektionen $\prod_1,\dots,\prod_d$ straff ist, und somit können wir ohne Einschränkung $d=1$ annehmen.

Zunächst ist $\varphi_{P_i}(0)=1$ für alle $i\in J$. Mit der Voraussetzung folgt, dass

$$\sup_{i\in I}|1-\varphi_{P_i}(t)| \overset{t\to 0}{\to} 0.$$

Wir erhalten

$$\begin{aligned}\sup_{r>0}\inf_{i\in J} P_i([-r,r]) &\ge 1-\inf_{r>0}\sup_{i\in J}\frac{r}{2}\int_{-2/r}^{2/r}(1-\varphi_{P_i}(t))dt\\ &\ge 1-\inf_{r>0}\frac{r}{2}\int_{-2/r}^{2/r}\sup_{i\in J}(2-\varphi_{P_i}(t))dt\\ &\ge 1-2\inf_{r>0}\sup_{[0,2/5]}\sup_{i\in J}(1-\varphi_{p_i}(t)) = 1.\end{aligned}$$ □

Theorem 3.17 (Lévy)**.** Seien $(P_n) \subseteq \mathscr{P}(\mathbb{R}^d)$ und $\varphi_{P_n}(t) \to \varphi(t)\ \forall t \in \mathbb{R}^d$ mit $\varphi : \mathbb{R}^d \to \mathbb{C}$. Ist φ stetig in 0, so folgt $P_n \to P$ für ein $P \in \mathscr{P}(\mathbb{R}^d)$ und $\varphi = \varphi_P$.

Beweis. Da φ_{P_n} punktweise konvergieren und φ in 0 stetig ist, folgt mit Lemma 3.13, dass (φ_{P_n}) in 0 gleichgradig stetig ist. Mit Satz 3.16 folgt also, dass (P_n) straff ist, also relativ folgenkompakt. Wir wählen eine teilfolge $P_{n_k} \Rightarrow P \in \mathscr{P}(\mathbb{R}^d)$.

Da die Abbildung $x \mapsto e^{itx}$ stetig und beschränkt ist, folgt

$$\varphi_{P_{n_k}}(t) \to \varphi_P(t) \quad \text{für alle } t \in \mathbb{R}^d.$$

Nach Voraussetzung gilt auf $\varphi_{P_n}(t) \to \varphi_P(t)$, also $\varphi = \varphi_P$. □

Beispiel 3.18 (Satz von de Moivre-Laplace)**.** Ist S_n binomialverteilt, $S_n \sim B(n,p)$, so konvergiert die standardisierte Zufallsvariable S_n^* gegen eine Standard-Normalverteilung, also $S_n^* := \frac{S_n - np}{\sqrt{np(1-p)}} \Rightarrow \mathcal{N}(0,1)$:

Zunächst ist $S_n = \sum_{i=1}^n X_i$ mit $X_i \sim \text{Bin}(1,p)$, also

$$\begin{aligned} \varphi_{S_n}(t) &= \left(pe^{it} + (1-p)\right)^n \\ \text{und} \quad \varphi_{S_n/\sqrt{np(1-p)}}(t) &= \left((1-p) + pe^{it/\sqrt{np(1-p)}}\right)^n, \end{aligned}$$

also

$$\begin{aligned} \varphi_{S_n^*} &> e^{-it\sqrt{\frac{np}{1-p}}}\left((1-p) + pe^{\frac{it}{\sqrt{np(1+p)}}}\right)^n \\ &= \left((1-p)e^{it\sqrt{\frac{p}{n(1-p)}}} + pe^{\frac{it}{\sqrt{np(1-p)}} - it\sqrt{\frac{p}{n(1-p)}}}\right) \\ &\overset{\text{Taylor}}{=} \Bigg((1-p)\Bigg(1 - it\sqrt{\frac{p}{n(1-p)}} - \frac{1}{2}\frac{t^2}{n(1-p)} - \underbrace{\frac{1}{b}it^3\Big(\frac{p}{n(1-p)}\Big)^{3/2}}_{\to O\left(\frac{1}{n^{3/2}}\right)}\Bigg) \\ &\qquad + p\Bigg(1 + \frac{it\sqrt{1-p}}{\sqrt{np}} - \frac{1}{2}\frac{t^2(1-p)}{np} + \underbrace{\frac{1}{b}it^3\Big(\frac{p}{n(1-p)}\Big)^{3/2}}_{\to O\left(\frac{1}{n^{3/2}}\right)}\Bigg)\Bigg)^n \\ &= \Bigg(1 - \underbrace{\frac{t^2}{2}\Big(\frac{p}{n} + \frac{1-p}{n}\Big)}_{= -\frac{t^2}{2n}} + C_n \cdot \frac{1}{n^{3/2}}\Bigg)^n \to e^{-\frac{t^2}{2}} = \varphi_{N(0,1)}(t). \end{aligned}$$

Für die Laplace-Transformierte erhält man eine analoge Aussage.

Beispiel 3.19. Gilt $X_n \sim$Geom(p_n) und $n \cdot p_n \to \lambda$, so folgt $\frac{X_n}{n} \Rightarrow$Exp$(\lambda)$:
Wir erhalten, dass

$$
\begin{aligned}
E\Big[e^{it\frac{X_n}{n}}\Big] &= \sum_{k=1}^{\infty}(1-p_n)^{k-1} \cdot p_n e^{\frac{itk}{n}} \\
&= p_n\Big(e \sum_{k=0}^{n}\Big((1-p_n) \cdot e^{\frac{it}{n}}\Big) \cdot \frac{1}{1-p_n} - 1\Big) \\
&\quad \frac{1}{1-p_n}\Big(\frac{1}{1-(1-p_n)e^{\frac{it}{n}}} - \frac{1-(1-p_n)^{\frac{it}{n}}}{1-(1-p_n)e^{\frac{it}{n}}}\Big) \\
&= \frac{\cancel{(1-p_n)}e^{\frac{it}{n}}}{1-(1-p_n)e^{\frac{it}{n}}} \cdot \frac{1}{\cancel{1-p_n}} \\
&= \frac{np_n e^{\frac{it}{n}}}{(1-(1-p_n)e^{\frac{it}{n}}} \to \frac{\lambda}{\lambda - it} = \varphi_{\mathrm{Exp}(\lambda)}
\end{aligned}
$$

$$
n\big(e^{-\frac{it}{n}} - (1-p_n)\big) = \Big(1 - \frac{it}{n} + O\Big(\frac{1}{n^2}\Big) - 1 + p_n\Big) = -it + n \cdot p_n
$$

3.1.1 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz von Lindeberg (1922) und Feller (1935, Theorem 3.20, verallgemeinert den Satz von de Moivre und Laplace sowie den zentralen Grenzwertsatz von Lindeberg und Lévy. Hierbei wird die Annahmen von unabhängig und identischer Verteilung durch die Annahme von Unabhängigkeit in Kombination mit der *Lindeberg-Bedingung* ersetzt (Bedingung (ii)). Lindeberg zeigte hierbei zunächst, aus (ii) folgt (i), während Feller die Rückrichtung bewies.

Theorem 3.20 (Zentraler Grenzwertsatz von Lindeberg-Feller)**.**
Sei $(X_{nj})_{n=1,2,\ldots,j=1,\ldots,m_n}$ eine Familie von Zufallsvariablen, so dass für $n = 1, 2, \ldots$ die Zufallsvariablen $X_{n1}, \ldots, X_{nm_n}$ unabhängig sind. Sei außerdem

$$
\sum_{j=1}^{m_n} E[X_{nj}] \xrightarrow{n\to\infty} \mu, \qquad \sum_{j=1}^{m_n} \mathrm{Var}(X_{nj}) \xrightarrow{n\to\infty} \sigma^2
$$

und $X \sim \mathscr{N}(\mu, \sigma^2)$. Dann sind folgende Aussagen äquivalent:

(i) $\displaystyle\sum_{j=1}^{m_n} X_{nj} \overset{n\to\infty}{\Longrightarrow} X$ und $\displaystyle\sup_{j=1,\ldots,m_n} \mathrm{Var}[X_{nj}] \xrightarrow{n\to\infty} 0$,

(ii) $\displaystyle\sum_{j=1}^{m_n} E\Big[(X_{nj} - E[X_{nj}])^2 \mathbb{1}_{\{|X_{nj}-E[X_{nj}]|>\varepsilon\}}\Big] \xrightarrow{n\to\infty} 0$ für alle $\varepsilon > 0$.

Als einfache Folgerund erhalten wir den zentralen Grenzwertsatz für i.i.d. Zufallsvariablen, den wir bereits in der Stochastik I kennengelernt hatten.

Korollar 3.21. *Seien $X_1, X_2, \dots$ i.i.d. mit $E[X_1] = \mu, \mathrm{Var}(X_1) = \sigma^2 \in (0, \infty)$. Dann gilt*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \xRightarrow{n\to\infty} \mathscr{N}(0,1).$$

Beweis. Sei $m_n = n$ und $X_{nj} = \frac{X_n - \mu}{\sqrt{n\sigma^2}}$. Dann erfüllt die Familie $(X_{nj})_{n=1,2,\dots,j=1,\dots,n}$ die Voraussetzungen von Theorem 3.20 mit $\mu = 0, \sigma^2 = 1$. Außerdem gilt

$$\sum_{j=1}^{n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] = \frac{1}{\sigma^2} E[(X_1 - \mu)^2 \mathbb{1}_{\{|X_1 - \mu| > \varepsilon\sqrt{n\sigma^2}\}}] \xrightarrow{n\to\infty} 0$$

mit majorisierter Konvergenz. □

Oftmals ist die Lindeberg-Bedingung nicht einfach nachzuprüfen. Einfacher ist oft die stärkere[1] *Lyapunov-Bedingung.*

Bemerkung 3.22. Die *Lyapunov-Bedingung* gilt im Kontext von Theorem 3.20, falls für ein $\delta > 0$

$$\sum_{j=1}^{m_n} E\big[|X_{nj} - E[X_{nj}]|^{2+\delta}\big] \xrightarrow{n\to\infty} 0.$$

Die Lyapunov-Bedingung impliziert die Lindeberg-Bedingung: Ohne Einschränkung sei $E[X_{nj}] = 0$. Es gilt für $\varepsilon > 0$

$$x^2 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta} 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta}.$$

Gilt nun die Lyapunoff-Bedingung, so folgt die Lindeberg-Bedingung aus

$$\sum_{j=1}^{m_n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \leq \frac{1}{\varepsilon^\delta} \sum_{j=1}^{m_n} E[|X_{nj}|^{2+\delta}] \xrightarrow{n\to\infty} 0.$$ ◇

Der Beweis von Theorem 3.20 basiert auf der geschickten Verwendung der charakteristischen Funktionen der Zufallsvariable X_{nj} und Taylor-Approximationen. Wir bereiten den Beweis des Theorems mit zwei Lemmata vor.

[1] Aktuelle deutsche Schreibweise nach Wikipedia ist Lyapunow, im Englischen of Lyapunov.

Lemma 3.23. *Für komplexe Zahlen $z_1, \dots, z_n, z_1', \dots, z_n'$ mit $|z_i| \le 1, |z_i'| \le 1$ für $i = 1, \dots, n$ gilt*

$$\Big| \prod_{k=1}^{n} z_k - \prod_{k=1}^{n} z_k' \Big| \le \sum_{k=1}^{n} |z_k - z_k'|. \tag{3.24}$$

Beweis. Für $n = 1$ ist die Gleichung offensichtlich richtig. Gilt (3.24) für ein n, so ist

$$\begin{aligned} \Big| \prod_{k=1}^{n+1} z_k - \prod_{k=1}^{n+1} z_k' \Big| &\le \Big| z_{n+1} \Big(\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} z_k' \Big) \Big| + \Big| (z_{n+1} - z_{n+1}') \prod_{k=1}^{n} z_k' \Big| \\ &\le \sum_{k=1}^{n} |z_k - z_k'| + |z_{n+1} - z_{n+1}'|. \end{aligned}$$

Daraus folgt die Behauptung. □

Wir verwenden weiterhin folgende Taylor-Approximation der Exponentialfunktion.

Lemma 3.25. *Sei $t \in \mathbb{C}$ und $n \in \mathbb{Z}_+$. Dann gilt*

$$\Big| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \Big| \le \frac{2|t|^n}{n!} \wedge \frac{|t|^{n+1}}{(n+1)!}. \tag{3.26}$$

Beweis. Bezeichne $h_n(t)$ die Differenz auf der linken Seite. Für $n = 0$ folgt (3.26) aus

$$|h_0(t)| = \Big| \int_0^t e^{is} ds \Big| \le \int_0^t |e^{is}| ds = |t|$$

und

$$|h_0(t)| \le |e^{it}| + 1 = 2.$$

Allgemein gilt für $t \in \mathbb{R}, n \in \mathbb{N}$

$$\Big| \int_0^t h_n(s) ds \Big| = \Big| -i(e^{it} - 1) + i \sum_{k=0}^{n} \frac{(it)^{k+1}}{(k+1)!} \Big| = \Big| ie^{it} - i \sum_{k=0}^{n+1} \frac{(it)^k}{k!} \Big| = |h_{n+1}(t)|,$$

woraus (3.26) mittels Induktion folgt. □

Im folgenden Beweis werden wir für Funktionen a und b genau dann $a \lesssim b$ schreiben, falls es eine Konstante C gibt mit $a \le Cb$.

Beweis von Theorem 3.20. O.E. sei $E[X_{nj}] = \mu = 0$ und $\sigma^2 = 1$; ansonsten ersetzen wir X_{nj} durch $\frac{X_{nj} - E[X_{nj}]}{\sqrt{\sigma^2}}$. Sei $\sigma_{nj}^2 := \mathrm{Var}[X_{nj}]$ sowie $\sigma_n^2 := \sum_{j=1}^{m_n} \sigma_{nj}^2 \xrightarrow{n \to \infty} 1$. Bezeichne außerdem ψ_{nj} die charakteristische Funktion von X_{nj}.

2. ⇒ 1. Da für jedes $\varepsilon > 0$

$$\sup_{j=1,\dots,m_n} \sigma_{nj}^2 \le \varepsilon^2 + \sup_{j=1,\dots,m_n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \le \varepsilon^2 + \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \xrightarrow{n\to\infty} \varepsilon^2, \quad (3.27)$$

ist der zweite Teil von 1. bereits gezeigt.

Seien $(Z_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ unabhängige Zufallsvariablen mit $Z_{nj} \sim N(0, \sigma_{nj}^2)$. Damit ist $Z_n := \sum_{j=1}^{m_n} Z_{nj} \sim N(0, \sigma_n^2)$. Insbesondere gilt also $Z_n \overset{n\to\infty}{\Longrightarrow} X$, was man etwa direkt aus der Form der charakteristischen Funktionen der Normalverteilung abliest. Sei $\widetilde{\psi}_{nj}$ die charakteristische Funktion von Z_{nj}. Dann genügt es zu zeigen, siehe Theorem 3.17, dass

$$\prod_{j=1}^{n_j} \psi_{nj}(t) - \prod_{j=1}^{m_n} \widetilde{\psi}_{nj}(t) \xrightarrow{n\to\infty} 0 \quad (3.28)$$

für alle t. Mittels Lemma 3.23 und Lemma 3.25 schreiben wir

$$\begin{aligned}
\Big| \prod_{j=1}^{m_n} \psi_{nj}(t) - \prod_{j=1}^{m_n} \widetilde{\psi}_{nj}(t) \Big| &\le \sum_{j=1}^{m_n} |\psi_{nj}(t) - \widetilde{\psi}_{nj}(t)| \\
&\le \sum_{j=1}^{m_n} |\psi_{nj}(t) - 1 + \tfrac{1}{2} t^2 \sigma_{nj}^2| + \sum_{j=1}^{m_n} |\widetilde{\psi}_{nj}(t) - 1 + \tfrac{1}{2} t^2 \sigma_{nj}^2| \\
&\lesssim 2 \sum_{j=1}^{m_n} E[X_{nj}^2 (1 \wedge |X_{nj}|) + \sum_{j=1}^{m_n} |e^{-\frac{1}{2}\sigma_{nj}^2 t^2} - 1 + \tfrac{1}{2} t^2 \sigma_{nj}^2|.
\end{aligned}$$

Weiter ist

$$\sum_{j=1}^{m_n} E[X_{nj}^2 (1 \wedge |X_{nj}|)] \le \varepsilon \sum_{j=1}^{m_n} \sigma_{nj}^2 + \sum_{j=1}^{m_n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \xrightarrow{n\to\infty} \varepsilon$$

und

$$\sum_{j=1}^{m_n} |e^{-\frac{1}{2}\sigma_{nj}^2 t^2} - 1 + \tfrac{1}{2} t^2 \sigma_{nj}^2| \lesssim \sum_{j=1}^{m_n} \sigma_{nj}^4 \le \sigma_n^2 \sup_{j=1,\dots,m_n} \sigma_{nj}^2 \xrightarrow{n\to\infty} 0$$

wegen (3.27). Damit ist (3.28) bereits bewiesen.

1. ⇒ 2. Nach dem zweiten Teil von 1. ist für jedes $\varepsilon > 0$ mit der Chebyshev-Ungleichung

$$\sup_{j=1,\dots,m_n} P[|X_{nj}| > \varepsilon] \le \sup_{j=1,\dots,m_n} \frac{\sigma_{nj}^2}{\varepsilon^2} \xrightarrow{n\to\infty} 0. \quad (3.29)$$

Mit Lemma 3.25 gilt

$$\sup_{j=1,\dots,m_n} |\psi_{nj}(t) - 1| \le \sup_{j=1,\dots,m_n} E[2 \wedge |t \cdot X_{nj}|] \le 2 \sup_{j=1,\dots,m_n} P[|X_{nj}| > \varepsilon] + \varepsilon |t| \xrightarrow{n\to\infty} \varepsilon |t|.$$

Insbesondere ist $\sum_{j=1}^{m_n} \log \psi_{nj}(t)$ für jedes t definiert, falls n groß genug ist. Aus der Gültigkeit von 1. folgt

$$\sum_{j=1}^{m_n} \log \psi_{nj}(t) \xrightarrow{n\to\infty} -\frac{t^2}{2}. \tag{3.30}$$

Außerdem gilt wegen $\psi'_{nj}(0) = iE[X_{nj}] = 0, \psi''_{nj}(0) = -\mathbf{V}[X_{nj}] = -\sigma_{nj}^2$ mit Hilfe einer Taylorentwicklung von ψ_{nj} um 0

$$|\psi_{nj}(t) - 1| \lesssim \sigma_{nj}^2 |t|^2$$

und

$$\begin{aligned}\Big|\sum_{j=1}^{m_n} \log \psi_{nj}(t) - \sum_{j=1}^{m_n} (\psi_{nj}(t) - 1)\Big| &\lesssim \sum_{j=1}^{m_n} |\psi_{nj}(t) - 1|^2 \\ &\lesssim \sum_{j=1}^{m_n} (\sigma_{nj}^2)^2 |t|^4 \lesssim |t|^4 \sup_{j=1,\dots,m_n} \sigma_{nj}^2 \xrightarrow{n\to\infty} 0.\end{aligned} \tag{3.31}$$

Da aus der Konvergenz einer imaginären Reihe die Konvergenz ihres Realteils folgt, folgern wir aus (3.30) und (3.31) wegen $\mathrm{Re}(\psi_{nj}(t)) = E[\cos(tX_{nj})]$

$$\sum_{j=1}^{m_n} E[\cos(tX_{nj}) - 1] \xrightarrow{n\to\infty} -\frac{t^2}{2}$$

Für $\varepsilon > 0$ ist nun wegen $0 \le 1 - \cos(\theta) \le \frac{\theta^2}{2}$

$$\begin{aligned} 0 \le \limsup_{n\to\infty} \sum_{j=1}^{m_n} E[X_{nj}^2; |X_{nj}| > \varepsilon] &= \limsup_{n\to\infty}\ 1 - \sum_{j=1}^{m_n} E[X_{nj}^2; |X_{nj}| \le \varepsilon] \\ &\le \limsup_{n\to\infty}\ 1 - \frac{2}{t^2} \sum_{j=1}^{m_n} E[1 - \cos(tX_{nj}); |X_{nj}| \le \varepsilon] \\ &= \limsup_{n\to\infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{E}[1 - \cos(tX_{nj}); |X_{nj}| > \varepsilon] \\ &\le \limsup_{n\to\infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{P}[|X_{nj}| > \varepsilon] \\ &\le \frac{2}{\varepsilon^2 t^2} \limsup_{n\to\infty} \sum_{j=1}^{m_n} \sigma_{nj} = \frac{2}{\varepsilon^2 t^2}. \end{aligned} \tag{3.32}$$

Da $t, \varepsilon > 0$ willkürlich waren, ist 2. gezeigt, wenn man in der letzten Ungleichungskette $t\infty$ betrachtet. □

3.1.2 Mehrdimensionale Grenzwertsätze

Bisher haben wir schwache Grenzwertsätze nur für den Fall $\mathbb{R}$–wertiger Zufallsvariablen betrachtet. Wir verallgemeinern dies nun zu $\mathbb{R}^d$–wertigen Zufallsgrößen. Insbesondere geben wir eine Variante des mehrdimensionalen zentralen Grenzwertsatzes an. Einen Vektor im $\mathbb{R}^d$ fassen wir als Spaltenvektor auf[2]. Das Skalarprodukt schreibt sich dann als $\langle a, b\rangle = a^\top b$.

Definition 3.33. Seien $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d\times d}$ eine strikt positiv definite symmetrische Matrix.[3,4] Die *d-dimensionale Normalverteilung* mit Erwartungswert μ und Covarianzmatrix Σ hat die Dichte

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big(- \tfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\Big).$$

Satz 3.34. *Seien $\mu \in \mathbb{R}^d$, $\Sigma = AA^\top \in \mathbb{R}^{d\times d}$ strikt positiv definit und symmetrisch. Dann sind äquivalent:*

(i) $X \sim \mathscr{N}(\mu, \Sigma)$

(ii) $\langle t, X\rangle \sim \mathscr{N}(\langle t, \mu \rangle, t^\top \Sigma t)$ *für jedes* $t \in \mathbb{R}^d$

(iii) $\varphi_X(t) = \exp\Big(i\langle t, \mu\rangle - \frac{1}{2} t^\top \Sigma t\Big)$ *für jedes* $t \in \mathbb{R}^d$.

In jedem dieser Fälle gilt

(iv) $X \overset{\mathscr{L}}{=} AY + \mu$ *für* $Y \sim \mathscr{N}(0, Id_d)$

(v) $E[X_i] = \mu_i$ *für* $i = 1, \dots, d$

(vi) $\mathrm{Cov}(X_i, X_j) = \Sigma_{ij}$ *für* $i, j = 1, \dots, d$

Theorem 3.35. Seien $X_1, X_2, \dots$ unabhängige, identische verteilte Zufallsvariablen mit Werten in $\mathbb{R}^d$ mit $E[X_n] = \mu \in \mathbb{R}^d$ und $\mathrm{Cov}[X_{n,i}, X_{n,j}] = \Sigma_{ij}$ für $i, j = 1, \dots, d$. Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xRightarrow{n\to\infty} \mathscr{N}_{0,\Sigma}.$$

Beweis. Wir wenden den eindimensionalen zentralen Grenzwertsatz, Korollar 3.21, auf die unabhängigen, identisch verteilten Zufallsvariablen $tX_1, tX_2, \dots$ an. Dieser liefert

$$t^\top \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xRightarrow{n\to\infty} t^\top X.$$

[2]Im Skriptum von Prof. Pfaffelhuber sind dies gerade Zeilenvektoren.

Da t beliebig war, folgt die Aussage mit Hilfe des Cramer-Wold Device. $\square$

Literatur

H. Bauer. *Maß- und Integrationstheorie*. Walter de Gruyter, Berlin, 1990.

P. Billingsley. *Probability and Measure*. Wiley, 3rd edition, 1995.

V. I. Bogachev. *Gaussian Measures*. American Mathematical Society, 1991.

J. Elstrodt. *Maß- und Integrationstheorie*. Springer Berlin Heidelberg, 2013.

L. Rüschendorf. *Wahrscheinlichkeitstheorie*. Springer, 2016.

D. Werner. *Funktionalanalysis*. Springer, 2000.

Index