Gerhard Keller

Wahrscheinlichkeitstheorie

Vorlesungen im WS 2007/2008 und im SS 2008

Mathematisches Institut, Universität Erlangen-Nürnberg

Letzte Änderung: 14. November 2007

Inhaltsverzeichnis

Literaturverzeichnis

[Bauer-MT] H. Bauer, *Maß- und Integrationstheorie*, de Gruyter (1990).

[Bauer-WT] H. Bauer, *Wahrscheinlichkeitstheorie*, de Gruyter (1991).

[Beltrami] E. Beltrami, *What Is Random?*, Springer (1999).

[Billingsley] P. Billingsley, *Probability and Measure*, Wiley (1986).

[Breiman] L. Breiman, *Probability*, Addison-Wesley (1968).

[Denker-Woyczyński] M. Denker, W.A. Woyczyński, *Introductory Statistics and Random Phenomena - Uncertainty, Complexity and Chaotic Behavior in Engineering and Science*, Birkhäuser (1998).

[Durrett] R. Durrett, *Probability: Theory and Examples*, Second Edition, Duxbury Press (1996).

[Gänssler-Stute] P. Gänssler, W. Stute, *Wahrscheinlichkeitstheorie*, Springer (1977).

[Georgii] H.-O. Georgii, *Stochastik*, de Gruyter (2002).

[Kallenberg] O. Kallenberg, *Foundations of Modern Probability*, Second Edition, Springer (2002).

[Klenke] A. Klenke, *Wahrscheinlichkeitstheorie*, Springer (2006).

[Resnick] S. Resnick, *A Probability Path*, Birkhäuser (1999).

[Revuz-Yor] D. Revuz, M. Yor, *Continuous Martingales and Brownian Motion*, Springer (1991).

[Širjaev] A.N. Širjaev, *Wahrscheinlichkeit*, VEB Verlag der Wissenschaften (1988). (Davon gibt es auch eine erweiterte englische Ausgabe.)

Eine Reihe dieser Bücher waren mir bei der Erstellung dieses Vorlesungsskripts sehr hilfreich. Insbesondere habe ich mich auf die Darstellungen in [Billingsley] und [Revuz-Yor] gestützt. Das Buch [Georgii] schlägt die Brücke von der elementaren Stochastik zur maßtheoretisch fundierten Wahrscheinlichkeitstheorie, um die es in dieser Vorlesung geht. Das Buch von [Kallenberg] stellt die zur Zeit wohl vollständigste Darstellung der Wahrscheinlichkeitstheorie von ihren maßtheoretischen Grundlagen bis zur stochaastischen Analysis dar.

Kapitel 1

Mengensysteme

In diesem Kapitel werden die mengentheoretischen Grundlagen der Wahrscheinlichkeitstheorie gelegt. Semiringe, Algebren, σ-Algebren und Dynkin-Systeme von Teilmengen werden eingeführt, erste Beispiele dafür angegeben, und es werden die Beziehungen zwischen diesen Strukturen geklärt.

Sei Ω eine Menge.

1.1 Definition (Semiring, $\cap$-stabil) *$\mathcal{S} \subseteq \mathfrak{P}(\Omega)$ ist ein Semiring, falls*

SR1) $\emptyset \in \mathcal{S}$,

SR2) $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$ *($\cap$-Stabilität),*

SR3) $A, B \in \mathcal{S}$ *und* $A \subseteq B \implies \exists A_1, \dots, A_k \in \mathcal{S}$ *paarweise disjunkt mit* $B \setminus A = A_1 \cup \dots \cup A_k$.

1.2 Beispiele (Semiringe)

a) $\mathcal{S} = \{\emptyset\}$ und $\mathcal{S} = \mathfrak{P}(\Omega)$ (triviale Extrembeispiele).

b) Ω beliebig, $At_\Omega := \{\{\omega\} : \omega \in \Omega\}$ (*Atome*). Dann ist $At_\Omega \cup \{\emptyset\}$ ein Semiring.

c) $\Omega = \mathbb{R}$, $\mathcal{S} = \mathcal{I} := \{(a, b] : -\infty < a \le b < +\infty\}$. $(a, b] \setminus (c, d]$ ist leer, ein Intervall oder disjunkte Vereinigung von zwei Intervallen.

d) $\Omega = \mathbb{R}^k$, $\mathcal{S} = \mathcal{I}^k := \{(\underline{a}, \underline{b}] : -\underline{\infty} < \underline{a} \le \underline{b} < +\underline{\infty}\}$ mit $(\underline{a}, \underline{b}] := \{\underline{x} \in \mathbb{R}^k : \underline{a} < \underline{x} \le \underline{b}\}$. Dabei sei $\underline{x} \le \underline{y}$ für $\underline{x}, \underline{y} \in \mathbb{R}^k$, falls $x_i \le y_i$ für alle $i = 1, \dots, k$ und es sei $\underline{x} < \underline{y}$, falls $\underline{x} \le \underline{y}$ und $\underline{x} \ne \underline{y}$. Die weiteren Ordnungssymbole werden analog benutzt. $(\underline{a}, \underline{b}] \setminus (\underline{c}, \underline{d}]$ kann als disjunkte Vereinigung von höchstens $2k$ k-dimensionalen Intervallen geschrieben werden (Übung).

e) Σ endliche Menge, $\Omega = \Sigma^{\mathbb{N}} = \{\omega = (\omega_1, \omega_2, \omega_3, \dots) : \omega_n \in \Sigma\ \forall n \in \mathbb{N}\}$. Für $a_1, \dots, a_N \in \Sigma$ sei

$$[a_1, \dots, a_N] := \{\omega \in \Omega : \omega_n = a_n (n = 1, \dots, N)\}$$

der von $a_1, \dots, a_N$ bestimmte *Zylinder* . Sei

$$\mathcal{Z}_N := \{[a_1, \dots, a_N] : a_1, \dots, a_N \in \Sigma\}$$
$$\mathcal{Z} := \bigcup_{N=1}^{\infty} \mathcal{Z}_N \cup \{\emptyset\} .$$

Dann ist $\mathcal{Z}$ ein Semiring:

SR1) $\emptyset \in \mathcal{S}$ ✔

SR2) $A = [a_1, \dots, a_M]$, $B = [b_1, \dots, b_N]$, o.B.d.A. $0 < M \leq N$.

$$A \cap B = \begin{cases} B, & \text{falls } a_j = b_j \ \forall j = 1, \dots, M \\ \emptyset & \text{sonst} \end{cases}$$

Außerdem: $A \cap \emptyset = \emptyset \ \forall A \in \mathcal{Z}$.

SR3) Seien $A, B \in \mathcal{Z}$. Ist $A = \emptyset$, so ist $B \setminus A = B \in \mathcal{Z}$, ist $B = \emptyset$, so ist $B \setminus A = \emptyset \in \mathcal{Z}$. Seien nun $A = [a_1, \dots, a_M], B = [b_1, \dots, b_N] \neq \emptyset$ und $A \subseteq B$. Dann ist $M \geq N$ und $a_j = b_j$ $\forall j = 1, \dots, N$.

- ▷ Ist $M = N$, so ist $B = A$, also $B \setminus A = \emptyset$.
- ▷ Ist $M > N$, so ist

$$\begin{aligned} B \setminus A &= \{\omega \in \Omega : \omega_j = a_j \ \forall j = 1, \dots, N \text{ aber} \\ &\qquad \exists j \in \{N+1, \dots, M\} : \omega_j \neq a_j\} \\ &= \bigcup_{\substack{(c_{N+1}, \dots, c_M) \in \Sigma^{M-N} \\ \exists j \in \{N+1, \dots, M\} : c_j \neq a_j}} [a_1, \dots, a_N, c_{N+1}, \dots, c_M] \end{aligned}$$

1.3 Bemerkung Der Raum $\Omega = \Sigma^{\mathbb{N}}$ des letzten Beispiels wird mit folgender Metrik zu einem kompakten metrischen Raum:

$$d(\omega, \omega') := \sum_{n=1}^{\infty} \delta_{\omega_n \neq \omega'_n} \cdot 2^{-n}$$

Die Eigenschaften einer Metrik weist man leicht nach. Ebenfalls leicht ist der Beweis, dass alle Zylinder zugleich offen und abgeschlossen sind. Statt der Kompaktheit zeigen wir hier eine zunächst etwas schwächere Eigenschaft, aus der die Kompaktheit aber leicht gefolgert werden kann (Übung!). Diese Eigenschaft, auf die wir später noch einmal zurückgreifen werden, lautet:

> Sind $A, A_k \in \mathcal{Z}$ und ist $A \subseteq \bigcup_{k=1}^{\infty} A_k$, so gibt es ein $m \in \mathbb{N}$ mit $A \subseteq \bigcup_{k=1}^{m} A_k$.

Sei dazu $\mathcal{E} := \{B \in \mathcal{Z} : \exists m \in \mathbb{N} \text{ s.d. } B \subseteq \bigcup_{k=1}^{m} A_k\}$. Zu zeigen ist also, dass $A \in \mathcal{E}$ ist. Da Σ endlich ist, existiert für jeden Zylinder $[b_1, \dots, b_n] \in \mathcal{Z} \setminus \mathcal{E}$ ein $b_{n+1} \in \Sigma$, so dass $[b_1, \dots, b_{n+1}] \in \mathcal{Z} \setminus \mathcal{E}$. Angenommen, der Zylinder $A = [a_1, \dots, a_N] \in \mathcal{Z} \setminus \mathcal{E}$. Dann folgt induktiv die Existenz von $a_{N+1}, a_{N+2}, \dots \in \Sigma$ derart, dass $[a_1, \dots, a_j] \in \mathcal{Z} \setminus \mathcal{E}$ für alle $j > N$. Andererseits ist $(a_1, a_2, a_3, \dots) \in A = \bigcup_{k=1}^{\infty} (A_k \cap A)$. Also gibt es ein $m \in \mathbb{N}$, für das $(a_1, a_2, a_3, \dots) \in A_m \cap A$. Aber $A_m \cap A$ ist ein Zylinder der Form $[a_1, \dots, a_N, a_{N+1}, \dots, a_{N+\ell}]$, und es folgt $[a_1, \dots, a_{N+\ell}] \subseteq A_m \subseteq \bigcup_{k=1}^{m} A_k$, im Widerspruch zu $[a_1, \dots, a_{N+\ell}] \in \mathcal{Z} \setminus \mathcal{E}$.

1.4 Definition ((σ)-Algebra) *$\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ ist eine Algebra, falls*

A1) $\Omega \in \mathcal{A}$,

A2) $A \in \mathcal{A} \implies A^c := \Omega \setminus A \in \mathcal{A}$,

A3) $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$.

$\mathcal{A}$ heißt σ-Algebra, falls statt A3) gilt

A3σ) $A_n \in \mathcal{A}$ $(n \in \mathbb{N}) \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.
Die Mengen $A \in \mathcal{A}$ heißen $\mathcal{A}$-messbar (oder einfach messbar).
Das Paar $(\Omega, \mathcal{A})$ heißt ein messbarer Raum.

1.5 Bemerkungen

a) Falls A2) gilt, so ist A1) äquivalent zu

A1') $\emptyset \in \mathcal{A}$

und A3) ist äquivalent zu

A3') $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$.

b) Aus A3),A3') folgt sofort: $A_1, \ldots, A_n \in \mathcal{A} \Longrightarrow A_1 \cup \cdots \cup A_n \in \mathcal{A}$, $A_1 \cap \cdots \cap A_n \in \mathcal{A}$.

c) Aus A2) und A3') folgt: $B \setminus A \in \mathcal{A}$ für alle $A, B \in \mathcal{A}$.

d) Insbesondere gilt: $\mathcal{A}$ Algebra $\Longrightarrow \mathcal{A}$ Semiring.

1.6 Lemma *Ist $(\mathcal{A}_i)_{i\in I}$ eine Familie von (σ)-Algebren, I eine beliebige Indexmenge, so ist $\bigcap_{i\in I} \mathcal{A}_i$ eine (σ)-Algebra.*

Beweis:

A1) $\forall i \in I : \Omega \in \mathcal{A}_i \implies \Omega \in \bigcap_{i\in I} \mathcal{A}_i$

A3σ)

$$\forall n \in \mathbb{N} : A_n \in \bigcap_{i\in I} \mathcal{A}_i \Longrightarrow \forall n \forall i : A_n \in \mathcal{A}_i \overset{A3\sigma}{\Longrightarrow} \forall i : \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_i$$
$$\Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \bigcap_{i\in I} \mathcal{A}_i$$

A2), A3) ähnlich.

□

1.7 Definition (Erzeugte σ-Algebra) *Sei $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$.*

$$\sigma(\mathcal{C}) := \bigcap_{\substack{\mathcal{A}\supseteq\mathcal{C} \\ \mathcal{A}\ \sigma\text{-Algebra}}} \mathcal{A}$$

heißt die von $\mathcal{C}$ erzeugte σ-Algebra. (Beachte, dass $\mathfrak{P}(\Omega)$ eine σ-Algebra ist, so dass der Durchschnitt immer über eine nicht leere Familie von σ-Algebren genommen wird.)

1.8 Bemerkungen

a) $\mathcal{C} \subseteq \sigma(\mathcal{C})$,

b) $\mathcal{C}_1 \subseteq \mathcal{C}_2 \implies \sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$,

c) $\mathcal{A}$ σ-Algebra $\Longrightarrow \sigma(\mathcal{A}) = \mathcal{A}$.

1.9 Beispiele (σ-Algebren) Die meisten interessanten σ-Algebren können nicht explizit angegeben werden, sondern sind nur durch $\mathcal{A} = \sigma(\mathcal{S})$ bestimmt, wo $\mathcal{S}$ ein Semiring oder ein anderes einfaches Mengensystem ist. Tatsächlich ist jede σ-Algebra entweder endlich oder überabzählbar.

a) Triviale Extrem-Beispiele sind $\mathcal{A} = \{\emptyset, \Omega\} = \sigma(\{\emptyset\})$ und $\mathcal{A} = \mathfrak{P}(\Omega) = \sigma(\mathfrak{P}(\Omega))$.

b) $\sigma(At_\Omega) = \{A \subseteq \Omega : A$ oder A^c höchstens abzählbar$\}$, denn da die rechte Seite eine in $\sigma(At_\Omega)$ enthaltene σ-Algebra ist, ist sie gleich $\sigma(At_\Omega)$.

c) $\mathcal{B} := \sigma(\mathcal{I})$ heißt die σ-Algebra der Borelschen Mengen in $\mathbb{R}$. Analog: $\mathcal{B}^k := \sigma(\mathcal{I}^k)$.

d) Sei $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$, $\bar{\mathcal{B}} := \{A, A \cup \{-\infty\}, A \cup \{+\infty\}, A \cup \{-\infty, +\infty\} : A \in \mathcal{B}\}$. Dann ist $\bar{\mathcal{B}}$ eine σ-Algebra.

e) $\mathcal{F} := \sigma(\mathcal{Z})$ ist die von den Zylindermengen erzeugte σ-Algebra auf $\Omega = \Sigma^{\mathbb{N}}$.

1.10 Bemerkung Für $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$ bezeichne $\mathcal{C}^*$ die Menge aller $A \subseteq \Omega$, für die $A \in \mathcal{C}$, $A^c \in \mathcal{C}$ oder A höchstens abzählbare Vereinigung von Mengen aus $\mathcal{C}$ ist. Definiere dann induktiv

$$\begin{aligned} \mathcal{I}_0 &= \mathcal{I}, & \mathcal{I}_n &= \mathcal{I}_{n-1}^* \quad (n \geq 1) \\ \mathcal{Z}_0 &= \mathcal{Z} \cup \{\emptyset\}, & \mathcal{Z}_n &= \mathcal{Z}_{n-1}^* \quad (n \geq 1) \end{aligned}$$

Keine der Mengen $\mathcal{I}_n, \mathcal{Z}_n, \bigcup_{n\geq 0} \mathcal{I}_n, \bigcup_{n \geq 0} \mathcal{Z}_n$ ist eine σ-Algebra. (Für $\mathcal{I}_n^*$ steht das am Ende von Sektion 2 in [Billingsley], und für $\mathcal{Z}_n$ kann man ganz analog vorgehen.) Andererseits ist $\mathcal{B} \neq \mathfrak{P}(\mathbb{R})$, $\mathcal{F} \neq \mathfrak{P}(\Sigma^{\mathbb{N}})$ und $\sigma(At_\Omega) \neq \mathfrak{P}(\Omega)$ falls Ω überabzählbar ist (ohne Beweis, siehe z.B. [Bauer-MT, Satz 8.6]).

Für einen flexiblen Umgang mit σ-Algebren benötigen wir noch eine weitere Struktur:

1.11 Definition (Dynkin-System) *$\mathcal{D} \subseteq \mathfrak{P}(\Omega)$ heißt Dynkin-System, falls*

D1) $\Omega \in \mathcal{D}$

D2) $A, B \in \mathcal{D}, A \subseteq B \implies B \setminus A \in \mathcal{D}$

D3) $A_n \in \mathcal{D}$ $(n \in \mathbb{N})$ *paarweise disjunkt* $\Longrightarrow \biguplus_{n=1}^{\infty} A_n \in \mathcal{D}$
($\biguplus$ bezeichnet eine disjunkte Vereinigung.)

1.12 Bemerkungen
a) D1),D2) $\Longrightarrow \emptyset \in \mathcal{D}$

b) Jede σ-Algebra ist ein Dynkin-System.

c) Der Durchschnitt beliebig vieler Dynkin-Systeme ist ein Dynkin-System (analog Lemma 1.6).

1.13 Definition (Erzeugtes Dynkin-System) *Für $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$ sei*

$$\mathcal{D}(\mathcal{C}) := \bigcap_{\substack{\mathcal{D} \supseteq \mathcal{C} \\ \mathcal{D} \text{ Dynkin-System}}} \mathcal{D}$$

das von $\mathcal{C}$ erzeugte Dynkin-System. (Es gelten die Bemerkungen 1.8 entsprechend.)

1.14 Lemma *Ein $\cap$-stabiles Dynkin-System $\mathcal{D}$ ist eine σ-Algebra.*

Beweis:

A1) D1 ✔

A2) $\Omega \in \mathcal{D} \Longrightarrow A^c = \Omega \setminus A \in \mathcal{D}$ für $A \in \mathcal{D}$.

A3) $A, B \in \mathcal{D} \Longrightarrow A \cup B = A \uplus (B \setminus (A \cap B)) \in \mathcal{D}$.

A3σ) $A_n \in \mathcal{D}$ $(n \geq 1) \overset{\text{A3}}{\Longrightarrow} B_n := \bigcup_{i=1}^{n} A_i \in \mathcal{D}$ $(n \geq 0) \Longrightarrow \bigcup_{n=1}^{\infty} A_n = \biguplus_{n=1}^{\infty} (B_n \setminus B_{n-1}) \in \mathcal{D}$ wegen D3.

□

1.15 Satz (Dynkin-System / σ-Algebra)
Sei $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$ $\cap$-stabil. Dann ist $\sigma(\mathcal{C}) = \mathcal{D}(\mathcal{C})$.

Beweis:
"$\supseteq$": $\sigma(\mathcal{C})$ ist ein Dynkin-System und $\mathcal{C} \subseteq \sigma(\mathcal{C})$. Also $\mathcal{D}(\mathcal{C}) \subseteq \mathcal{D}(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$.
"$\subseteq$": Da $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{D}(\mathcal{C}))$, reicht es zu zeigen, dass $\mathcal{D}(\mathcal{C})$ eine σ-Algebra ist, denn dann ist $\sigma(\mathcal{D}(\mathcal{C})) = \mathcal{D}(\mathcal{C})$. Also ist zu zeigen, dass $\mathcal{D}(\mathcal{C})$ $\cap$-stabil ist (Lemma 1.14): Für $C \in \mathcal{D}(\mathcal{C})$ sei

$$\mathcal{D}_C := \{A \in \mathcal{D}(\mathcal{C}) : A \cap C \in \mathcal{D}(\mathcal{C})\} \quad \text{(Beachte: } A \in \mathcal{D}_C \Longleftrightarrow C \in \mathcal{D}_A\text{)}$$

Wir zeigen zunächst, dass $\mathcal{D}_C$ ein Dynkin-System ist:

D1) $\Omega \cap C = C \in \mathcal{D}(\mathcal{C})$, also $\Omega \in \mathcal{D}_C$.

D2) $A, B \in \mathcal{D}_C, A \subseteq B \Longrightarrow (B \setminus A) \cap C = (B \cap C) \setminus (A \cap C) \in \mathcal{D}(\mathcal{C})$, also $B \setminus A \in \mathcal{D}_C$.

D3) $A_n \in \mathcal{D}_C$ $(n \geq 1)$ paarweise disjunkt $\Longrightarrow (\biguplus_{n=1}^{\infty} A_n) \cap C = \biguplus_{n=1}^{\infty} (A_n \cap C) \in \mathcal{D}(\mathcal{C})$, also $\biguplus_{n=1}^{\infty} \in \mathcal{D}_C$.

Die $\cap$-Stabilität folgt nun aus folgenden drei Überlegungen:

- Ist $G \in \mathcal{C}$, so ist $C \in \mathcal{D}_G$ für alle $C \in \mathcal{C}$, also $\mathcal{C} \subseteq \mathcal{D}_G$ und damit $\mathcal{D}(\mathcal{C}) \subseteq \mathcal{D}(\mathcal{D}_G) = \mathcal{D}_G$.
- Sei $B \in \mathcal{D}(\mathcal{C})$ und $G \in \mathcal{C}$. Dann ist $B \in \mathcal{D}_G$ und damit $G \in \mathcal{D}_B$ für alle $G \in \mathcal{C}$. Es folgt: $\mathcal{C} \subseteq \mathcal{D}_B$, also $\mathcal{D}(\mathcal{C}) \subseteq \mathcal{D}_B$.
- Sind nun $A, B \in \mathcal{D}(\mathcal{C})$, so folgt $A \in \mathcal{D}_B$, d.h. $A \cap B \in \mathcal{D}(\mathcal{C})$.

$\square$

Kapitel 2

Mengenfunktionen

In diesem Kapitel werden Maße eingeführt. Das sind Mengenfunktionen, die jeder Menge einer σ-Algebra eine Zahl zwischen 0 und $+\infty$ zuordnen und gewisse Additivitätseigenschaften besitzen. Werden die Werte im Intervall $[0,1]$ angenommen, so spricht man von *Wahrscheinlichkeitsmaßen*.

2.1 Definition (Prämaße, Maße) *Sei* $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$, $\mu : \mathcal{C} \to [0,+\infty]$ *eine* Mengenfunktion.

a) μ *heißt* *endlich additiv auf* $\mathcal{C}$, *falls für alle paarweise disjunkten* $A_1,\dots,A_n \in \mathcal{C}$ *gilt:*

$$\bigcup_{i=1}^{n} A_i \in \mathcal{C} \implies \mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i)\ .$$

b) μ *heißt* *σ-additiv auf* $\mathcal{C}$, *falls für alle paarweise disjunkten* $A_n \in \mathcal{C}$ $(n \in \mathbb{N})$ *gilt:*

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

c) μ *heißt* *subadditiv auf* $\mathcal{C}$ *(auch: endlich subadditiv), falls* für alle $A, A_1,\dots,A_n \in \mathcal{C}$ $(n \in \mathbb{N})$ *gilt:*

$$A \subseteq \bigcup_{i=1}^{n} A_i \implies \mu\,(A) \le \sum_{i=1}^{n} \mu(A_i)\ .$$

d) μ *heißt* *σ-subadditiv auf* $\mathcal{C}$, *falls* für alle $A, A_n \in \mathcal{C}$ $(n \in \mathbb{N})$ *gilt:*

$$A \subseteq \bigcup_{n=1}^{\infty} A_n \implies \mu\,(A) \le \sum_{n=1}^{\infty} \mu(A_n)\ .$$

e) Ist $\mathcal{A}$ *eine Algebra (σ-Algebra) und ist* $\mu : \mathcal{A} \to [0,+\infty]$ *mit* $\mu(\emptyset) = 0$ *σ-additiv auf* $\mathcal{A}$, *so heißt* μ *ein Prämaß (Maß).*

f) Ist μ *ein Maß auf der σ-Algebra* $\mathcal{A}$, *so heißt das Tripel* $(\Omega,\mathcal{A},\mu)$ *ein Maßraum. Ist außerdem* $\mu(\Omega) = 1$, *so heißt* $(\Omega,\mathcal{A},\mu)$ *ein Wahrscheinlichkeitsraum und* μ *ein Wahrscheinlichkeitsmaß.*

Achtung: *Billingsley nennt* μ *auch dann ein Maß, wenn* $\mathcal{A}$ *nur eine Algebra ist.*

2.2 Lemma *Sei* $\mathcal{S}$ *ein Semiring,* $\mu : \mathcal{S} \to [0,+\infty]$ *additiv. Dann gilt:*

i) $A, B \in \mathcal{S}, A \subseteq B \Longrightarrow \mu(A) \le \mu(B)$ *(Monotonie)*

ii) μ *ist endlich subadditiv.*

Beachte, dass die Subadditivität keine triviale Konsequenz der Additivität ist, da bei der Subadditivität nicht die paarweise Disjunktheit der B_i angenommen wird!

Beweis:

i) $B \setminus A = \bigcup_{i=1}^n A_i$ für paarweise disjunkte $A_i \in \mathcal{S}$. Daher:

$$\mu(B) = \mu(A \uplus A_1 \uplus \cdots \uplus A_n) = \mu(A) + \mu(A_1) + \cdots + \mu(A_n) \geq \mu(A)$$

ii) Betrachte zunächst den Fall $A = \bigcup_{i=1}^n A_i$. Für jedes A_i gibt es $C_i^1, \ldots, C_i^{k_i} \in \mathcal{S}$, so dass $A \setminus A_i = C_i^1 \uplus \cdots \uplus C_i^{k_i}$. Setze $C_i^0 := A_i$. Dann ist $A = \biguplus_{j=0}^{k_i} C_i^j \quad (i = 1, \ldots, n)$. Setze $\mathcal{C} := \{C_1^{j_1} \cap \cdots \cap C_n^{j_n} : 0 \leq j_i \leq k_i \ (i = 1, \ldots, n)\}$ und $\mathcal{C}_i := \{S \in \mathcal{C} : S \subseteq A_i\}$. Dann ist

$$A = \biguplus_{S \in \mathcal{C}} S\ , \quad A_i = \biguplus_{S \in \mathcal{C}_i} S \ (i = 1, \ldots, n)$$

und

$$\mathcal{C} = \bigcup_{i=1}^n \mathcal{C}_i \subseteq \mathcal{S}\ .$$

Da außerdem $\mu(S) \geq 0$ für jedes $S \in \mathcal{S}$, folgt aus der endlichen Additivität von μ auf $\mathcal{S}$

$$\mu(A) = \sum_{S \in \mathcal{C}} \mu(S) \leq \sum_{i=1}^n \sum_{S \in \mathcal{C}_i} \mu(S) = \sum_{i=1}^n \mu(A_i)\ .$$

Ist nun A eine echte Teilmenge von $\bigcup_{i=1}^n A_i$, so setze $\tilde{A}_i = A \cap A_i$. Dann sind die $\tilde{A}_i \in \mathcal{S}$ und $A = \bigcup_{i=1}^n \tilde{A}_i$. Also: $\mu(A) \leq \sum_{i=1}^n \mu(\tilde{A}_i)$, und aus Teil i) folgt $\mu(\tilde{A}_i) \leq \mu(A_i)$ □

2.3 Beispiele (Maße, einfache)

a) $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$, $x \in \Omega$, $\mu(A) = \delta_x(A) := 1_A(x)$ (*Punktmasse, Dirac-Maß*)

b) $\mathcal{A} = \{\emptyset, \Omega\}$, $\mu(\emptyset) = 0$, $\mu(\Omega) = 1$

c) $\mathcal{A} = \sigma(At_\Omega)$, $\mu(A) = \mathrm{card}(A)$. (Dieses Beispiel ist dafür verantwortlich, dass wir später Sätze über die Summation von Reihen mit beliebiger Indexmenge als einfache Spezialfälle von Sätzen der Integrationstheorie auffassen können.)

d) $\Omega = \mathbb{N}$, $(p_n)_{n \geq 1}$ mit $p_n \geq 0$, $\sum_{i=1}^\infty p_n = 1$, d.h. $(p_n)_{n \geq 1}$ ist ein Wahrscheinlichkeitsvektor. Setze $\mu(A) := \sum_{n \in A} p_n$ für $A \in \mathcal{A} = \mathfrak{P}(\mathbb{N})$.

2.4 Beispiel (Lebesgue-Stieltjes-Maß auf $\mathbb{R}$) Sei $\mathcal{B} = \sigma(\mathcal{I})$ die Borel-σ-Algebra auf $\mathbb{R}$ und sei $F : \mathbb{R} \to \mathbb{R}$ monoton wachsend und rechtsseitig stetig (d.h. $\lim_{t \downarrow x} F(t) = F(x)$ für alle x). Ziel: Ein Maß λ_F auf $\mathcal{B}$ mit $\lambda_F((a, b]) = F(b) - F(a)$, $\lambda_F(\emptyset) = 0$. Hier zeigen wir:

Das so festgelegte λ_F ist additiv und σ-subadditiv auf $\mathcal{I}$.

Später folgt daraus die Fortsetzbarkeit zu einem Maß auf $\mathcal{B}$. Ist $F(x) = x$, so erhält man gerade das *Lebesgue-Maß* auf $\mathbb{R}$, während jedes F mit $\lim_{x \to -\infty} F(x) = 0$ und $\lim_{x \to \infty} F(x) = 1$ die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes auf $\mathbb{R}$ beschreibt.

Additivität: Seien $A_i = (a_i, b_i] \in \mathcal{I}$ $(i = 1, \ldots, n)$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$. O.B.d.A. sei $a_1 \leq b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_n \leq b_n$. Da $\biguplus_{i=1}^n A_i \in \mathcal{I}$, folgt $b_1 = a_2, \ldots, b_{n-1} = a_n$, also $\bigcup_{i=1}^n A_i = (a_1, b_n]$. Daher

$$\lambda_F\left(\bigcup_{i=1}^n A_i\right) = F(b_n) - F(a_1) = \sum_{i=1}^n (F(b_i) - F(a_i)) = \sum_{i=1}^n \lambda_F(A_i)\ .$$

Die endliche Subadditivität folgt nun aus Lemma 2.2.

<u>σ-Subadditivität:</u> Seien nun $(a,b],(a_n,b_n] \in \mathcal{I}$ $(n \geq 1)$ und $(a,b] \subseteq \bigcup_{n=1}^{\infty}(a_n,b_n]$. O.B.d.A. $a < b$. Sei $\epsilon > 0$. Da F rechtsseitig stetig ist, gibt es $\delta, \delta_n > 0$ derart, dass $F(a+\delta) < F(a)+\epsilon$ und $F(b_n+\delta_n) \leq F(b_n)+\epsilon 2^{-n}$ für alle $n \geq 1$. Es folgt

$$
\begin{aligned}
&[a+\delta,b] \subseteq \bigcup_{n=1}^{\infty}(a_n, b_n+\delta_n) \\
\Longrightarrow &\exists\, m > 0 : [a+\delta,b] \subseteq \bigcup_{n=1}^{m}(a_n, b_n+\delta_n) \quad \text{(Kompaktheit)} \\
\Longrightarrow &F(b)-F(a+\delta) = \lambda_F((a+\delta,b]) \leq \sum_{n=1}^{m}\lambda_F((a_n,b_n+\delta_n]) = \sum_{n=1}^{m}(F(b_n+\delta_n)-F(a_n)) \\
&\qquad\qquad\qquad\text{(endliche Subadditivität)} \\
\Longrightarrow &F(b)-F(a)-\epsilon \leq \sum_{n=1}^{\infty}(F(b_n)-F(a_n))+\epsilon
\end{aligned}
$$

Da $\epsilon > 0$ beliebig war, folgt $\lambda_F((a,b]) \leq \sum_{n=1}^{\infty}\lambda_F((a_n,b_n])$.

2.5 Beispiel (Bernoulli-Maß auf $\Sigma^{\mathbb{N}}$) Sei $\mathcal{F} = \sigma(\mathcal{Z})$, siehe Beispiel 1.9e), und sei $(p_i)_{i\in\Sigma}$ ein Wahrscheinlichkeitsvektor. Wir wollen ein Wahrscheinlichkeitsmaß P auf $\mathcal{F}$ definieren, für das

$$P([a_1,\dots,a_N]) = \prod_{i=1}^{N} p_{a_i} \ (a_1,\dots,a_N \in \Sigma),$$

also insbesondere $P([a]) = p_a$ für alle $a \in \Sigma$. Wie im letzten Beispiel zeigen wir zunächst nur, dass durch diese Festsetzung eine auf $\mathcal{Z}$ additive und σ-subadditive Mengenfunktion gegeben ist.

<u>Additivität:</u> Übung!

<u>σ-Subadditivität:</u> P additiv auf $\mathcal{Z} \overset{\text{Lemma 2.2}}{\Longrightarrow} P$ subadditiv auf $\mathcal{Z}$. Sei nun $A \subseteq \bigcup_{k=1}^{\infty} A_k$, $A, A_k \in \mathcal{Z}$. Da alle Zylindermengen offen und abgeschlossen sind und da $\Sigma^{\mathbb{N}}$ kompakt ist (siehe Bemerkung 1.3), gibt es ein $m > 0$ derart, dass $A \subseteq \bigcup_{k=1}^{m} A_k$. Aus der Subadditivität und der Positivität von P folgt nun: $P(A) \leq \sum_{k=1}^{n} P(A_k) \leq \sum_{k=1}^{\infty} P(A_k)$.

2.6 Bemerkung Sei μ ein Maß auf $\mathcal{A}$, $A \in \mathcal{A}$. Bezeichne

$$\mathcal{A}_{|A} := \{B \subseteq A : B \in \mathcal{A}\} = \{C \cap A : C \in \mathcal{A}\}\ .$$

- ▷ $\mathcal{A}_{|A}$ ist eine σ-Algebra. Übung!
- ▷ Die durch $B \mapsto \mu(B)$ definierte Mengenfunktion auf $\mathcal{A}_{|A}$ ist ein Maß – die <u>Einschränkung von μ auf $\mathcal{A}_{|A}$</u>. (Übung!) Wir bezeichnen sie mit $\mu_{|A}$ oder auch einfach wieder mit μ. Ein wichtiges Beispiel ist die Einschränkung des Lebesgue-Maßes auf $[0,1]$.

Notation: Seien $A, A_1, A_2, A_3, \dots \subseteq \Omega$. Wir schreiben

$$A_n \nearrow A, \text{ falls } A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \text{ und } \bigcup_{n=1}^{\infty} A_n = A$$

$$A_n \searrow A, \text{ falls } A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots \text{ und } \bigcap_{n=1}^{\infty} A_n = A$$

2.7 Satz (Monotone Stetigkeit von (Prä)maßen)
Sei μ ein Prämaß auf der Algebra $\mathcal{A}$ und seien $A, A_n \in \mathcal{A}$.

i) $A_n \nearrow A \Longrightarrow \mu(A_n) \nearrow \mu(A)$ *(Stetigkeit von unten)*

ii) $A_n \searrow A$ *und* $\mu(A_1) < \infty \Longrightarrow \mu(A_n) \searrow \mu(A)$ *(Stetigkeit von oben)*

iii) $A \subseteq \bigcup_{n=1}^{\infty} A_n \Longrightarrow \mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)$

Beweis:

i) Sei $A_0 := \emptyset$. Wegen Lemma 2.2 ist $\mu(A_0) \leq \mu(A_1) \leq \mu(A_2) \leq \dots$. Außerdem ist

$$A = \bigcup_{n=1}^{\infty} A_n = \biguplus_{n=1}^{\infty} (A_n \setminus A_{n-1}), \quad A_n \setminus A_{n-1} \in \mathcal{A}\,,$$

so dass

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n-1}) = \sup_{k>0} \sum_{n=1}^{k} \mu(A_n \setminus A_{n-1}) = \sup_{k>0} \mu(A_k)$$

ii) $A_n \searrow A \Longrightarrow (A_1 \setminus A_n) \nearrow (A_1 \setminus A)$. Unter Beachtung von $\mu(A_1) < \infty$ folgt nun aus Teil i):

$$\mu(A_1) - \mu(A_n) = \mu(A_1 \setminus A_n) \nearrow \mu(A_1 \setminus A) = \mu(A_1) - \mu(A)\,,$$

also $\mu(A_n) \searrow \mu(A)$.

iii) Setze $B_n := A \cap \bigcup_{k=1}^{n} A_k$. Dann ist $B_n \in \mathcal{A}$ und $B_n \nearrow A$. Aus Teil i) und Lemma 2.2 folgt:

$$\mu(A) = \sup_{n>0} \mu(B_n) \leq \sup_{n>0} \mu\left(\bigcup_{k=1}^{n} A_k\right) \leq \sup_{n>0} \sum_{k=1}^{n} \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k)\,.$$

□

2.8 Bemerkungen

a) Auf die Voraussetzung $\mu(A_1) < \infty$ in Satz 2.7ii) kann i.A. nicht verzichtet werden. (Beispiel?)

b) Die Aussage i) von Satz 2.7 ist sogar äquivalent zur σ-Additivität, und ist $\mu(\Omega) < \infty$, so ist auch Aussage ii) äquivalent dazu.

2.9 Beispiel $\lambda([a,b]) = \lambda\left(\bigcap_{n=1}^{\infty} (a - \frac{1}{n}, b]\right) = \lim_{n\to\infty} b - (a - \frac{1}{n}) = b - a$. Insbesondere $\lambda(\{b\}) = 0$ und $\lambda((a,b)) = \lambda((a,b] \setminus \{b\}) = \lambda((a,b]) - 0 = b - a$.

2.10 Definition (endliche und σ-endliche Maße) *Sei $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$ und $\mu : \mathcal{C} \to [0,\infty]$ monoton.*

a) μ ist <u>endlich</u>, falls $\Omega \in \mathcal{C}$ und $\mu(\Omega) < \infty$.

b) μ ist <u>σ-endlich</u>, falls es $C_1, C_2, \dots \in \mathcal{C}$ gibt mit $C_n \nearrow \Omega$ und $\mu(C_n) < \infty$ für alle n.

2.11 Beispiele a) Sei $F : \mathbb{R} \to \mathbb{R}$, λ_F wie in Beispiel 2.4. Da $F((-n,n]) = F(n) - F(-n) < \infty$ für $n \in \mathbb{N}$ und $(-n,n] \nearrow \mathbb{R}$, ist λ_F auf $\mathcal{I}$ σ-endlich. Das gilt insbesondere für das Lebesgue-Maß λ.

b) Sei $\mu : \mathfrak{P}(\mathbb{R}) \to [0,+\infty]$, $\mu(A) = \mathrm{card}(A)$. μ ist *nicht* σ-endlich, da $\mu(C_n) < \infty \Longrightarrow C_n$ endlich $\Longrightarrow \bigcup_{n=1}^{\infty} C_n$ abzählbar $\Longrightarrow \mathbb{R} \neq \bigcup_{n=1}^{\infty} C_n$.

Bevor wir im nächsten Kapitel mit nicht ganz geringem Aufwand zeigen, dass sich jede additive und σ-subadditive Mengenfunktion auf einem Semiring zu einem Maß auf der von dem Semiring erzeugten σ-Algebra fortsetzen lässt, zeigen wir hier schon einmal, dass es in den meisten Fällen höchstens eine solche Fortsetzung geben kann:

2.12 Satz (Eindeutigkeit der Maßfortsetzung)
Sei $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$ $\cap$-stabil und seien μ_1, μ_2 Maße auf $\sigma(\mathcal{C})$. Sind die Einschränkungen $\mu_1|_\mathcal{C}$ und $\mu_2|_\mathcal{C}$ σ-endlich und stimmen sie überein, so ist $\mu_1 = \mu_2$.

Beweis: Für $C \in \mathcal{C}$ mit $\mu_1(C) = \mu_2(C) < \infty$ setze

$$\mathcal{D}_C := \{A \in \sigma(\mathcal{C}) : \mu_1(A \cap C) = \mu_2(A \cap C)\} \ .$$

Dann ist $\mathcal{C} \subseteq \mathcal{D}_C$ und $\mathcal{D}_C$ ist ein Dynkin-System, denn:

D1) $\mu_1(\Omega \cap C) = \mu_1(C) = \mu_2(C) = \mu_2(\Omega \cap C) \Longrightarrow \Omega \in \mathcal{D}_C$.

D2) Seien $A, B \in \mathcal{D}_C$, $A \subseteq B$. Dann ist für $i = 1, 2$

$$\mu_i((B \setminus A) \cap C) = \mu_i((B \cap C) \setminus (A \cap C)) \overset{\mu_i(C)<\infty}{=} \mu_i(B \cap C) - \mu_i(A \cap C) \ .$$

Da $A, B \in \mathcal{D}_C$, folgt: $\mu_1((B \setminus A) \cap C) = \mu_2((B \setminus A) \cap C)$.

D3) Seien $A_n \in \mathcal{D}_C$ paarweise disjunkt. Dann ist für $i = 1, 2$

$$\mu_i\left(\left(\biguplus_{n=1}^{\infty} A_n\right) \cap C\right) = \mu_i\left(\biguplus_{n=1}^{\infty} (A_n \cap C)\right) = \sum_{n=1}^{\infty} \underbrace{\mu_i(A_n \cap C)}_{\text{gleich für } i=1,2}$$

Also stimmt der erste dieser Ausdrücke auch für $i = 1$ und $i = 2$ überein, so dass $\biguplus_{n=1}^{\infty} A_n \in \mathcal{D}_C$.

Aus Satz 1.15 folgt nun: $\sigma(\mathcal{C}) = \mathcal{D}(\mathcal{C}) \subseteq \mathcal{D}(\mathcal{D}_C) = \mathcal{D}_C \subseteq \sigma(\mathcal{C})$, d.h.

$$\mu_1(A \cap C) = \mu_2(A \cap C) \quad \text{für alle } A \in \sigma(\mathcal{C}) \ ,$$

und das gilt für jedes $C \in \mathcal{C}$. Seien nun $C_k \in \mathcal{C}$, $C_k \nearrow \Omega$, $\mu_i(C_k) < \infty$. Für jedes $A \in \sigma(\mathcal{C})$ folgt dann aus Satz 2.7

$$\mu_1(A) = \sup_k \mu_1(A \cap C_k) = \sup_k \mu_2(A \cap C_k) = \mu_2(A) \ .$$

□

Kapitel 3

Äußere Maße und Maße

Hauptziel dieses Abschnitts ist es zu zeigen, dass sich jedes Maß auf einer Algebra zu einem Maß auf der erzeugten σ-Algebra fortsetzen lässt.

3.1 Definition (Äußeres Maß) *Die Mengenfunktion $\mu^* : \mathfrak{P}(\Omega) \to [0, +\infty]$ ist ein <u>äußeres Maß</u>, falls μ^* σ-subadditiv und $\mu^*(\emptyset) = 0$ ist.*

3.2 Bemerkung Ein äußeres Maß ist *monoton*, denn ist $A \subseteq B$, also $A \subseteq B \cup \emptyset \cup \emptyset \cup \ldots$, so folgt $\mu^*(A) \leq \mu^*(B)$.

3.3 Satz (Konstruktion äußerer Maße)
Sei $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$, $\emptyset \in \mathcal{C}$, und sei $\rho : \mathcal{C} \to [0, +\infty]$ eine Mengenfunktion mit $\rho(\emptyset) = 0$. (Z.B. kann $\mathcal{C}$ ein Semiring sein.) Für $A \subseteq \Omega$ setze

$$\rho^*(A) := \inf \left\{ \sum_{n=1}^{\infty} \rho(C_n) : A \subseteq \bigcup_{n=1}^{\infty} C_n, C_n \in \mathcal{C}\, \forall n \right\} .$$

(Falls keine abzählbare $\mathcal{C}$-Überdeckung von A existiert, heißt das $\rho^*(A) = +\infty$.)
Dann ist ρ^* ein äußeres Maß.

Beispiel: $\Omega = \mathbb{R}^d$, $\mathcal{C} = \{B_r(x) : x \in \mathbb{R}^d, r > 0\}$ die Familie aller offenen Kugeln im $\mathbb{R}^d$, $\rho(B_r(x)) = r^\alpha$ für ein $\alpha > 0$. Wir werden sehen, dass $\rho^*|_{\mathcal{M}(\rho^*)}$ ein Maß ist, das sogenannte *Hausdorff-Maß* der Dimension α. Für $\alpha = d$ stimmt es bis auf einen Vorfaktor mit dem Lebesgue-Maß λ^d überein.

Beweis: $\rho^*(\emptyset) = 0$ ist klar. Wir zeigen die σ-Subadditivität von ρ^*: Sei $A \subseteq \bigcup_{n=1}^\infty A_n$. Ist $\rho^*(A_n) = \infty$ für ein n, so ist $\rho^*(A) \leq \infty = \sum_{n=1}^\infty \rho^*(A_n)$. Also können wir annehmen, dass $\rho^*(A_n) < \infty$ für alle n.

Sei $\epsilon > 0$. Zu jedem A_n gibt es eine Überdeckung $A_n \subseteq \bigcup_{k=1}^\infty C_n^k$ mit $C_n^k \in \mathcal{C}$ und

$$\sum_{k=1}^{\infty} \rho(C_n^k) \leq \rho^*(A_n) + \epsilon 2^{-n} .$$

Es folgt:

$$\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{\infty} C_n^k \quad \text{und}$$

$$\sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \rho(C_n^k) \leq \sum_{n=1}^{\infty} \rho^*(A_n) + \epsilon ,$$

d.h. $\rho^*\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \rho^*(A_n) + \epsilon$. Lasse nun $\epsilon \to 0$ gehen. $\square$

3.4 Definition (μ^*-Messbarkeit) *Sei μ^* ein äußeres Maß auf Ω. $A \subseteq \Omega$ heißt μ^*-messbar, falls*

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(E) \quad \text{für alle } E \subseteq \Omega\,.$$

(Wegen der Subadditivität von μ^* ist das äquivalent zu

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E) \quad \text{für alle } E \subseteq \Omega\,.)$$

Bezeichnung:

$$\mathcal{M}(\mu^*) := \{A \subseteq \Omega : A\ \mu^*\text{-messbar}\}$$

3.5 Satz
Ist μ^* ein äußeres Maß auf Ω, so ist $\mathcal{M}(\mu^*)$ eine σ-Algebra, und die Einschränkung von μ^* auf $\mathcal{M}(\mu^*)$ ist ein Maß.

Der Satz ist eine sofortige Konsequenz der folgenden drei Lemmata:

3.6 Lemma *Ist μ^* ein äußeres Maß, so ist $\mathcal{M}(\mu^*)$ eine Algebra.*

Beweis:

A1) $\Omega \in \mathcal{M}(\mu^*)$, denn $\mu^*(E) + \mu^*(\emptyset) = \mu^*(E)$ für alle $E \subseteq \Omega$.

A2) $A \in \mathcal{M}(\mu^*) \Longleftrightarrow A^c \in \mathcal{M}(\mu^*)$ ✔

A3') Wir zeigen, dass mit $A, B \in \mathcal{M}(\mu^*)$ auch $A \cap B \in \mathcal{M}(\mu^*)$: Sei $E \subseteq \Omega$. Dann ist

$$\begin{aligned}
&\mu^*((A \cap B) \cap E) + \mu^*((A \cap B)^c \cap E) \\
=&\mu^*(A \cap B \cap E) + \mu^*((A^c \cap B \cap E) \cup (A^c \cap B^c \cap E) \cup (A \cap B^c \cap E)) \\
\overset{\text{Subadditivität}}{\leq}&\mu^*(A \cap (B \cap E)) + \mu^*(A^c \cap (B \cap E)) + \mu^*(A^c \cap (B^c \cap E)) + \mu^*(A \cap (B^c \cap E)) \\
\overset{A \in \mathcal{M}(\mu^*)}{=}&\mu^*(B \cap E) + \mu^*(B^c \cap E) \\
\overset{B \in \mathcal{M}(\mu^*)}{=}&\mu^*(E)
\end{aligned}$$

□

3.7 Lemma *Ein äußeres Maß μ^* ist σ-additiv auf $\mathcal{M}(\mu^*)$.*

Beweis: Einfache Additivität: Seien $A, B \in \mathcal{M}(\mu^*)$, $A \cap B = \emptyset$. Dann ist

$$\begin{aligned}
\mu^*(A \cup B) &= \mu^*(A \cap (A \cup B)) + \mu^*(A^c \cap (A \cup B)) \quad \text{da } A \in \mathcal{M}(\mu^*) \\
&= \mu^*(A) + \mu^*(B)
\end{aligned}$$

Die endliche Additivität folgt daraus durch Induktion, da $\mathcal{M}(\mu^*)$ eine Algebra ist (Lemma 3.6).
σ-Additivität: Seien $A, A_n \in \mathcal{M}(\mu^*)$, $A = \biguplus_{n=1}^{\infty} A_n$. Für $m > 0$ ist

$$\sum_{n=1}^{m} \mu^*(A_n) \overset{\text{Additivität}}{=} \mu^*\left(\bigcup_{n=1}^{m} A_n\right) \overset{\text{Bem. 3.2}}{\leq} \mu^*(A)$$

Es folgt $\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A_n)$, und die umgekehrte Ungleichung folgt aus der σ-Subadditivität des äußeren Maßes μ^*. □

3.8 Lemma *Ist μ^* ein äußeres Maß, so ist $\mathcal{M}(\mu^*)$ eine σ-Algebra.*

Beweis: Da $\mathcal{M}(\mu^*)$ eine Algebra und damit $\cap$-stabil ist (Lemma 3.6), folgt aus Satz 1.15 dass $\mathcal{D}(\mathcal{M}(\mu^*)) = \sigma(\mathcal{M}(\mu^*))$. Es reicht daher zu zeigen, dass $\mathcal{M}(\mu^*)$ ein Dynkin-System ist, denn dann folgt $\mathcal{M}(\mu^*) = \mathcal{D}(\mathcal{M}(\mu^*))$:
D1) und D2) folgen direkt aus der Algebra-Eigenschaft von $\mathcal{M}(\mu^*)$.
Für D3) seien $A_n \in \mathcal{M}(\mu^*)$ paarweise disjunkt, $A = \bigcup_{n=1}^{\infty} A_n$. Zu zeigen ist:

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E) \quad \text{für alle } E \subseteq \Omega\ . \tag{*}$$

Unter Ausnutzung von Lemma 3.7 zeigen wir zunächst durch Induktion nach n:

$$\mu^*\left(E \cap \biguplus_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mu^*(E \cap A_k)$$

$\underline{n = 1}$: ✔
$\underline{n \Rightarrow n+1}$:

$$\begin{aligned}
\mu^*\left(E \cap \biguplus_{k=1}^{n+1} A_k\right) &= \mu^*\left(\left(E \cap \biguplus_{k=1}^{n+1} A_k\right) \cap \biguplus_{k=1}^{n} A_k\right) + \mu^*\left(\left(E \cap \biguplus_{k=1}^{n+1} A_k\right) \cap \left(\biguplus_{k=1}^{n} A_k\right)^c\right) \\
&= \mu^*\left(E \cap \biguplus_{k=1}^{n} A_k\right) + \mu^*(E \cap A_{n+1}) \\
&= \sum_{k=1}^{n+1} \mu^*(E \cap A_k)
\end{aligned}$$

Daraus und aus der Monotonie von μ^* (Bemerkung 3.2) folgt für alle n

$$\mu^*(E) = \mu^*\left(E \cap \biguplus_{k=1}^{n} A_k\right) + \mu^*\left(E \setminus \biguplus_{k=1}^{n} A_k\right) \geq \sum_{k=1}^{n} \mu^*(E \cap A_k) + \mu^*(E \cap A^c)$$

Für $n \to \infty$ folgt mit der σ-Subadditivität von μ^*:

$$\mu^*(E) \geq \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \setminus A) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)\ ,$$

also (*) . □

Satz 3.5 folgt nun direkt aus den drei letzten Lemmata. Er ist der Schlüssel zum Hauptergebnis dieses Kapitels:

3.9 Satz (Fortsetzungssatz für Maße, klassische Form)
Jedes Prämaß μ auf einer Algebra $\mathcal{A}$ lässt sich zu einem Maß auf $\sigma(\mathcal{A})$ fortsetzen. Ist μ σ-endlich auf $\mathcal{A}$, so ist diese Fortsetzung eindeutig.

Da jede Algebra insbesondere ein Semiring ist, ist dieser Satz enthalten in der folgenden, etwas technischeren, aber flexibler anwendbaren Aussage:

3.10 Satz (Fortsetzungssatz für Maße, flexible Form)
Sei $\mathcal{S}$ ein Semiring. $\mu : \mathcal{S} \to [0, +\infty]$ sei additiv, σ-subadditiv und erfülle $\mu(\emptyset) = 0$. Dann lässt sich μ zu einem Maß auf $\sigma(\mathcal{S})$ fortsetzen. Ist μ σ-endlich auf $\mathcal{S}$, so ist diese Fortsetzung eindeutig.

Beweis: Die Eindeutigkeit wurde schon in Satz 2.12 bewiesen. Zum Beweis der Existenz wird ein äußeres Maß μ^* wie in Satz 3.3 definiert:

$$\mu^*(A) = \inf\left\{\sum_{n=1}^{\infty}\mu(S_n) : A \subseteq \bigcup_{n=1}^{\infty} S_n, S_n \in \mathcal{S}\ \forall n\right\}$$

Ist $A \in \mathcal{S}$, so folgt aus der σ-Subadditivität von μ auf $\mathcal{S}$: $\mu(A) \le \sum_{n=1}^{\infty}\mu(S_n)$ wann immer $A \subseteq \bigcup_{n=1}^{\infty} S_n$ mit $S_n \in \mathcal{S}$. Daher ist $\mu(A) \le \mu^*(A)$, und da $A \subseteq A \cup \emptyset \cup \emptyset \cup \dots$, ist auch $\mu^*(A) \le \mu(A)$. Also: $\mu^*(A) = \mu(A)$ für $A \in \mathcal{S}$.

Bleibt zu zeigen, dass $\mathcal{S} \subseteq \mathcal{M}(\mu^*)$. Denn dann ist auch $\sigma(\mathcal{S}) \subseteq \mathcal{M}(\mu^*)$, und nach Satz 3.5 ist μ^* ein Maß auf $\mathcal{M}(\mu^*)$.

Seien also $A \in \mathcal{S}$ und $E \subseteq \Omega$ beliebig. Zu zeigen:

$$\mu^*(E \cap A) + \mu^*(E \cap A^c) \le \mu^*(E)\ . \tag{**}$$

Sei o.B.d.A. $\mu^*(E) < \infty$. Dann gibt es zu $\epsilon > 0$ Mengen $E_n \in \mathcal{S}$ mit

$$E \subseteq \bigcup_{n=1}^{\infty} E_n \quad \text{und} \quad \sum_{n=1}^{\infty}\mu(E_n) < \mu^*(E) + \epsilon\ .$$

Da $\mathcal{S}$ ein Semiring ist, gilt für alle $n \in \mathbb{N}$

$$\begin{aligned} B_n &:= E_n \cap A \in \mathcal{S}\ , \\ E_n \setminus A = E_n \setminus B_n &= \biguplus_{k=1}^{m_n} C_n^k \quad \text{für geeignete } C_n^k \in \mathcal{S}. \end{aligned}$$

Also:

$$E \cap A \subseteq \bigcup_{n=1}^{\infty} B_n\ , \quad E \setminus A \subseteq \bigcup_{n=1}^{\infty}\biguplus_{k=1}^{m_n} C_n^k$$
$$E_n = B_n \uplus \biguplus_{k=1}^{m_n} C_n^k\ ,$$

so dass aus der Definition von μ^* und aus der endlichen Additivität von μ auf $\mathcal{S}$ folgt:

$$\begin{aligned} \mu^*(E \cap A) + \mu^*(E \cap A^c) &\le \sum_{n=1}^{\infty}\mu(B_n) + \sum_{n=1}^{\infty}\sum_{k=1}^{m_n}\mu(C_n^k) \\ &= \sum_{n=1}^{\infty}\mu(E_n) < \mu^*(E) + \epsilon\ . \end{aligned}$$

Mit $\epsilon \to 0$ folgt daraus (**). □

3.11 Bemerkung In Zukunft werden die Maße $\mu^*_{|\mathcal{M}(\mu^*)}$ und $\mu^*_{|\sigma(\mathcal{S})}$ wieder mit μ bezeichnet.

3.12 Beispiele (Fortsetzung von 2.4 und 2.5) $\lambda_F : \mathcal{I} \to [0, +\infty]$ und $P : \mathcal{Z} \to [0, +\infty]$ erfüllen die Voraussetzungen von Satz 3.10. Damit ist auch der letzte Schritt zur Konstruktion von *Lebesgue-Stieltjes-Maßen* auf $\mathbb{R}$ und von *Bernoulli-Maßen* auf $\Sigma^{\mathbb{N}}$ getan.

3.13 Satz (Approximationssatz für Maße)
Seien $\mathcal{S}$ und $\mu : \mathcal{S} \to [0, \infty]$ wie in Satz 3.10, μ σ-endlich.

a) *Zu $A \in \mathcal{M}(\mu^*)$ und $\epsilon > 0$ gibt es eine Folge $S_1, S_2, \dots \in \mathcal{S}$ paarweise disjunkter Mengen mit $A \subseteq \bigcup_{n=1}^{\infty} S_n$ und $\mu\left(\bigcup_{n=1}^{\infty} S_n \setminus A\right) < \epsilon$.*

b) *Zu $A \in \mathcal{M}(\mu^*)$ mit $\mu(A) < \infty$ und $\epsilon > 0$ gibt es eine endliche Folge $S_1, \dots, S_n \in \mathcal{S}$ paarweise disjunkter Mengen mit $\mu\left(A \triangle \bigcup_{k=1}^{n} S_k\right) < \epsilon$.*

c) *Zu $A \in \mathcal{M}(\mu^*)$ gibt es $A_-, A_+ \in \sigma(\mathcal{S})$ mit $A_- \subseteq A \subseteq A_+$ und $\mu(A_+ \setminus A_-) = 0$.*

Beweis: Sei $A \in \mathcal{M}(\mu^*)$.

a) Sei $\epsilon > 0$. Da μ σ-endlich auf $\mathcal{S}$ ist, gibt es Mengen $C_n \in \mathcal{S}$ derart, dass $C_n \nearrow \Omega$ und $\mu(C_n) = \mu^*(C_n) < \infty$ für alle n. Nach Konstruktion von μ^* gibt es zu jedem n Mengen $B_n^k \in \mathcal{S}$ mit

$$C_n \cap A \subseteq \bigcup_{k=1}^{\infty} B_n^k \quad \text{und} \quad \sum_{k=1}^{\infty} \mu\left(B_n^k\right) < \mu^*(C_n \cap A) + \epsilon 2^{-n} = \mu(C_n \cap A) + \epsilon 2^{-n}. \qquad (*)$$

Wir nummerieren die Mengen B_n^k zu einer Folge $(B_\ell)_\ell$ um und konstruieren daraus eine spezielle Überdeckung von A durch *paarweise disjunkte* Mengen aus $\mathcal{S}$: Sei $R_1 = B_1$, $R_2 = B_2 \setminus B_1 = B_2 \setminus (R_1 \cap B_2)$, und für allgemeine ℓ sei

$$R_{\ell+1} = B_{\ell+1} \setminus \left(\biguplus_{k=1}^{n} R_k \cap B_{\ell+1} \right) .$$

Die R_ℓ sind paarweise disjunkt, und man überlegt sich leicht (Induktion, Übung!), dass jedes R_ℓ endliche disjunkte Vereinigung von Mengen aus $\mathcal{S}$ ist. Daher gibt es $S_1, S_2, \dots \in \mathcal{S}$ derart, dass

$$A = \bigcup_{n=1}^{\infty} (C_n \cap A) \subseteq \bigcup_{\ell=1}^{\infty} B_\ell = \biguplus_{\ell=1}^{\infty} R_\ell = \biguplus_{j=1}^{\infty} S_j$$

und

$$\begin{aligned}
\mu\left(\biguplus_{j=1}^{\infty} S_j \setminus A \right) &= \mu\left(\bigcup_{\ell=1}^{\infty} B_\ell \setminus A \right) \leq \mu\left(\bigcup_{n=1}^{\infty} \left(\bigcup_{k=1}^{\infty} B_n^k \setminus (C_n \cap A) \right) \right) \\
&\leq \sum_{n=1}^{\infty} \mu\left(\bigcup_{k=1}^{\infty} B_n^k \setminus (C_n \cap A) \right) = \sum_{n=1}^{\infty} \left(\mu\left(\bigcup_{k=1}^{\infty} B_n^k \right) - \mu(C_n \cap A) \right) \\
&\leq \sum_{n=1}^{\infty} \left(\sum_{k=1}^{\infty} \mu(B_n^k) - \mu(C_n \cap A) \right) \\
&\leq \sum_{n=1}^{\infty} 2^{-n} \epsilon \\
&= \epsilon ,
\end{aligned}$$

wobei (*) für die letzte Ungleichung benutzt wurde.

b) Da $\mu(A) < \infty$, ist auch $\mu\left(\biguplus_{j=1}^{\infty} S_j\right) < \mu(A) + \epsilon < \infty$, und es gibt ein $m > 0$ derart, dass $\sum_{j=m+1}^{\infty} \mu(S_j) < \epsilon$. Also:

$$\mu\left(A \triangle \biguplus_{j=1}^{m} S_j \right) \leq \mu\left(\biguplus_{j=m+1}^{\infty} S_j \right) + \mu\left(\biguplus_{j=1}^{\infty} S_j \setminus A \right) < 2\epsilon ,$$

d.h. Aussage b).

c) Wegen a) gibt es zu $A \in \mathcal{M}(\mu^*)$ eine Folge von Mengen $A_n \in \sigma(\mathcal{S})$ mit $A \subseteq A_n$ und $\mu(A_n \setminus A) < \frac{1}{n}$. Für $A_+ := \bigcap_{n=1}^{\infty} A_n$ ist dann $A \subseteq A_+$ und $\mu(A_+ \setminus A) \leq \inf_n \mu^*(A_n \setminus A) = 0$. Das gleiche Argument angewandt auf A^c liefert $A_- := ((A^c)_+)^c$, denn $((A^c)_+)^c \subseteq A$ und $\mu(A \setminus ((A^c)_+)^c) = \mu(A \cap (A^c)_+) = \mu((A^c)_+ \setminus A^c) = 0$.

□

3.14 Bemerkung Sei $\mathcal{N}(\mu^*) := \{A \subseteq \Omega : \mu^*(A) = 0\}$. Aus der Monotonie des äußeren Maßes μ^* folgt leicht
– $\mathcal{N}(\mu^*) \subseteq \mathcal{M}(\mu^*)$ und
– ist $A \in \mathcal{N}(\mu^*)$ and $B \subseteq A$ beliebig, so ist $B \in \mathcal{N}(\mu^*)$.
Damit lässt sich leicht zeigen (Übung!)

$$\begin{aligned}\mathcal{M}(\mu^*) &= \{B \cup N : B \in \sigma(\mathcal{S}), N \in \mathcal{N}(\mu^*)\}\\ &= \{A \subseteq \Omega : \exists\, B \in \sigma(\mathcal{S}) \text{ mit } \mu^*(A \triangle B) = 0\}\ .\end{aligned}$$

3.15 Bemerkung (Vollständigkeit, Vervollständigung) Ein Maß ν auf einer σ-Algebra $\mathcal{A}$ heißt vollständig, falls für $A \in \mathcal{A}$ mit $\nu(A) = 0$ gilt, dass $B \in \mathcal{A}$ für alle Teilmengen $B \subseteq A$.

Bezeichnet man in der Situation von Satz 3.10 die Fortsetzung von μ auf $\sigma(\mathcal{S})$ wieder mit μ, so ist das Maß $\mu^*_{|\mathcal{M}(\mu^*)}$ die Vervollständigung von μ. Man überlegt sich leicht, dass die Vervollständigung eindeutig bestimmt ist. (Übung!)

Oft wird erst die *Vervollständigung* von λ das Lebesgue-Maß genannt. $\mathcal{B}_0 := \mathcal{M}(\lambda^*)$ nennt man die Lebesgue-Mengen.

3.16 Bemerkung (Regularität) Sei Ω ein topologischer Raum, $\mathcal{O} \subseteq \mathfrak{P}(\Omega)$ das System der offenen Mengen, $\mathcal{B} := \sigma(\mathcal{O})$ die Borel'sche σ-Algebra. Ein Maß μ auf $(\Omega, \mathcal{B})$ heißt

i) von außen regulär, falls zu jedem $A \in \mathcal{B}$ und $\epsilon > 0$ eine offene Menge $G \supseteq A$ mit $\mu(G \setminus A) < \epsilon$ existiert;

ii) von innen regulär, falls zu jedem $A \in \mathcal{B}$ mit $\mu(A) < \infty$ und $\epsilon > 0$ eine kompakte Menge $K \subseteq A$ mit $\mu(A \setminus K) < \epsilon$ existiert.

Aus Satz 3.13 folgert man leicht die *Regularität* aller Maße μ auf $(\mathbb{R}^k, \mathcal{B}^k)$, die auf allen Intervallen $(\underline{a}, \underline{b}] \in \mathcal{I}^k$ endlich sind (Beweis: Übung!). Insbesondere ist auch das Lebesgue-Stieltjes-Maß λ auf $\mathbb{R}$ regulär, ebenso wie das später zu konstruierende Lebesgue-Maß auf $\mathbb{R}^k$.

Die Regularität des Bernoulli-Maßes P (bzw. seiner Vervollständigung, die ebenfalls mit P bezeichnet wird) ist in Satz 3.13 enthalten: Da alle $S \in \mathcal{S} = \mathcal{Z}$ offen sind, bedeutet Aussage 3.13a) die Regularität von außen. Da $\Omega = \Sigma^{\mathbb{N}}$ kompakt und da P endlich ist, sind Regularität von innen und Regularität von außen äquivalent.

Kapitel 4

Messbare Funktionen und Abbildungen

In ähnlicher Weise wie man auf topologischen Räumen stetige Abbildungen definiert, führen wir in diesem Kapitel den Begriff der *messbaren Abbildung* zwischen messbaren Räumen ein. Im Rahmen der Wahrscheinlichkeitstheorie werden messbare Funktionen später als Zufallsvariablen interpretiert.

4.1 Definition (Messbare Abbildung) *Seien* $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ *messbare Räume.*

a) *Eine Abbildung* $T : \Omega \to \Omega'$ *heißt* <u>$\mathcal{A}$-$\mathcal{A}'$-*messbar*</u> *(oder einfach messbar, wenn keine Verwechslungsgefahr besteht), falls* $T^{-1}A' \in \mathcal{A}$ *für alle* $A' \in \mathcal{A}'$ *(kurz:* $T^{-1}\mathcal{A}' \subseteq \mathcal{A}$*).* Erinnerung: $T^{-1}A' = \{x \in \Omega : Tx \in A'\}$

b) *Eine Funktion* $f : \Omega \to \mathbb{R}(\bar{\mathbb{R}})$ *heißt* <u>$\mathcal{A}$-*messbar*</u> *(oder einfach messbar), falls sie* $\mathcal{A}$-$\mathcal{B}(\bar{\mathcal{B}})$*-messbar ist.*

4.2 Bemerkung $f : \Omega \to \bar{\mathbb{R}}$ ist messbar genau dann, wenn $f^{-1}\{-\infty\}, f^{-1}\{+\infty\} \in \mathcal{A}$ und $f^{-1}M \in \mathcal{A}$ für alle $M \in \mathcal{B}$.

4.3 Lemma *Seien* $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ *messbare Räume und sei* $T : \Omega \to \Omega'$*. Dann sind* $T^{-1}\mathcal{A}'$ *und* $\mathcal{A}'_0 := \{A' \in \mathcal{A}' : T^{-1}A' \in \mathcal{A}\}$ σ*-Algebren.*

Beweis: Leichte Übung, die nur die Vertauschbarkeit der üblichen Mengenoperationen mit der Urbildoperation T^{-1} benutzt. □

Folgender Satz hilft, die Messbarkeit einer Abbildung in konkreten Fällen zu beweisen:

4.4 Satz (Messbarkeit auf Erzeugern)

a) *Seien* $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ *messbare Räume,* $\mathcal{M}' \subseteq \mathcal{A}'$*, und sei* $T : \Omega \to \Omega'$*. Dann ist*
$$\sigma(T^{-1}\mathcal{M}') = T^{-1}\sigma(\mathcal{M}').$$

b) *Ist außerdem* $\sigma(\mathcal{M}') = \mathcal{A}'$ *und ist* $T^{-1}A' \in \mathcal{A}$ *für alle* $A' \in \mathcal{M}'$*, so ist* T *messbar.*

Beweis:

a) Da $T^{-1}\mathcal{M}' \subseteq T^{-1}\sigma(\mathcal{M}')$ und da $T^{-1}\sigma(\mathcal{M}')$ eine σ-Algebra ist (Lemma 4.3), ist auch $\sigma(T^{-1}\mathcal{M}') \subseteq T^{-1}\sigma(\mathcal{M}')$. Für die Umkehrung betrachte $\mathcal{A}'_0 := \{A' \in \sigma(\mathcal{M}') : T^{-1}A' \in \sigma(T^{-1}\mathcal{M}')\}$. $\mathcal{A}'_0$ ist eine σ-Algebra (Lemma 4.3) und $\mathcal{M}' \subseteq \mathcal{A}'_0$. Also ist $\sigma(\mathcal{M}') \subseteq \mathcal{A}'_0$ und daher $T^{-1}\sigma(\mathcal{M}') \subseteq T^{-1}\mathcal{A}'_0 = \sigma(T^{-1}\mathcal{M}')$.

b) Ist $\sigma(\mathcal{M}') = \mathcal{A}'$, so folgt aus Teil a):
$$T^{-1}\mathcal{A}' = T^{-1}\sigma(\mathcal{M}') = \sigma(T^{-1}\mathcal{M}') \subseteq \sigma(\mathcal{A}) = \mathcal{A}\,.$$
□

4.5 Beispiele a) Betrachte $f : \Omega \to \mathbb{R}$ und $\mathcal{M} = \{(-\infty, x] : x \in \mathbb{R}\}$. Da $\mathcal{I} \subseteq \sigma(\mathcal{M})$ und $\mathcal{M} \subseteq \mathcal{B}$, ist $\mathcal{B} = \sigma(\mathcal{I}) \subseteq \sigma(\mathcal{M}) \subseteq \mathcal{B}$, also $\mathcal{B} = \sigma(\mathcal{M})$. Daher ist f messbar, falls für alle $x \in \mathbb{R}$

$$\{f \leq x\} := \{\omega \in \Omega : f(\omega) \leq x\} = f^{-1}((-\infty, x]) \quad \text{messbar ist.}$$

Analoges gilt für die Mengen $\{f \geq x\}$, $\{f < x\}$ und $\{f > x\}$ (nicht aber für $\{f = x\}$!).

b) Alle $\pi_i : \mathbb{R}^k \to \mathbb{R}, \underline{x} \mapsto x_i$ $(i = 1, \ldots, k)$ sind messbar, da

$$\pi_i^{-1}((-\infty, x]) = \mathbb{R}^{i-1} \times (-\infty, x] \times \mathbb{R}^{k-i} = \bigcup_{n=1}^{\infty} \left((-n, n]^{i-1} \times (-n, x] \times (-n, n]^{k-i}\right) \in \mathcal{B}^k$$

c) Betrachte nun $f : \Omega \to \mathbb{R}^k$ und $\mathcal{M} = \{(-\underline{\infty}, \underline{x}] : \underline{x} \in \mathbb{R}^k\}$. Beachte, dass sich jedes $(\underline{a}, \underline{b}] \in \mathcal{I}^k$ darstellen lässt als

$$(-\underline{\infty}, \underline{b}] \setminus \left(\bigcup_{i=1}^{k} (-\underline{\infty}, (b_1, \ldots, b_{i-1}, a_i, b_{i+1}, \ldots, b_k)]\right) .$$

Also ist $\mathcal{I}^k \subseteq \sigma(\mathcal{M})$ und daher $\mathcal{B}^k = \sigma(\mathcal{I}^k) = \sigma(\mathcal{M})$. Schreibe f als $(f_1, \ldots, f_k)$, jedes $f_i : \Omega \to \mathbb{R}$. Für $\underline{x} \in \mathbb{R}^k$ ist dann

$$\{\omega \in \Omega : f(\omega) \leq \underline{x}\} = \bigcap_{i=1}^{k} \{\omega \in \Omega : f_i(\omega) \leq x_i\} .$$

Daher ist f messbar, falls alle f_i messbar sind. (Die Umkehrung folgt aus Teil b und dem nächsten Satz, da $f_i = \pi_i \circ f$.)

d) Ähnlich zeigt man, dass $f : \Omega \to \bar{\mathbb{R}}^k$ genau dann messbar ist, falls alle $f_i : \Omega \to \bar{\mathbb{R}}$ messbar sind.

4.6 Satz
Seien $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}'), (\Omega'', \mathcal{A}'')$ messbare Räume, $T_1 : \Omega \to \Omega'$, $T_2 : \Omega' \to \Omega''$ messbar. Dann ist $T_2 \circ T_1 : \Omega \to \Omega''$ messbar.

Beweis: $(T_2 \circ T_1)^{-1}(\mathcal{A}'') = T_1^{-1}(T_2^{-1}\mathcal{A}'') \subseteq T_1^{-1}\mathcal{A}' \subseteq \mathcal{A}$. □

4.7 Satz
Jedes stetige $f : \mathbb{R}^k \to \mathbb{R}^n$ ist messbar.

Beweis: Wegen Beispiel 4.5c) reicht es, den Fall $n = 1$ zu betrachten. Dann gilt:

$$f \text{ stetig} \Longrightarrow \{f < x\} \text{ offen für alle } x \in \mathbb{R} \Longrightarrow \{f < x\} \in \mathcal{B}^k \text{ für alle } x \in \mathbb{R},$$

und die Messbarkeit von F folgt aus Satz 4.4. □

Im Folgenden wollen wir auch mit $\bar{\mathbb{R}}$-wertigen Funktionen rechnen. Daher vereinbaren wir folgende Regeln für das Rechnen mit $\pm\infty$:

$$0 \cdot (\pm\infty) = 0, \quad x \cdot (+\infty) = +\infty \text{ für } x > 0, \quad +\infty + \infty = +\infty, \quad (+\infty) \cdot (+\infty) = +\infty, \quad \text{u.s.w.}$$

Verboten wird natürlich $+\infty - \infty$!

4.8 Satz
Sei $(\Omega, \mathcal{A})$ ein messbarer Raum. Summen, Produkte, Maxima und Minima endlich vieler messbarer Funktionen von Ω nach $\mathbb{R}$ oder $\bar{\mathbb{R}}$ sind messbar.

Beweis: Seien $f_1, \ldots, f_n : \Omega \to \mathbb{R}$ messbar. Dann ist $F := (f_1, \ldots, f_n) : \Omega \to \mathbb{R}^n$ messbar, und wegen Satz 4.7 und 4.6 ist $g \circ F$ für jedes stetige $g : \mathbb{R}^n \to \mathbb{R}$ messbar. Wähle nun für g:

$$g(\underline{x}) = \sum_{i=1}^{n} x_i,\ \prod_{i=1}^{n} x_i,\ \max_{i=1}^{n} x_i,\ \min_{i=1}^{n} x_i\ .$$

Der Beweis für $\bar{\mathbb{R}}$-wertige f_i verläuft ähnlich. □

4.9 Satz
Sei $f_1, f_2, \ldots$ eine Folge messbarer $\bar{\mathbb{R}}$-wertiger Funktionen auf $(\Omega, \mathcal{A})$.

a) Die Funktionen $\sup_n f_n$, $\inf_n f_n$, $\limsup_{n\to\infty} f_n$ und $\liminf_{n\to\infty} f_n$ sind messbar.

b) $E := \{\omega : \lim_{n\to\infty} f_n(\omega)$ existiert in $\bar{\mathbb{R}}\} \in \mathcal{A}$ und $1_E \cdot \lim_{n\to\infty} f_n$ ist messbar.

Beweis:

a)

$$\left\{\sup_n f_n \le x\right\} = \bigcap_n \{f_n \le x\} \in \mathcal{A} \quad \text{für alle } x \in \bar{\mathbb{R}}\ ,$$
$$\left\{\inf_n f_n \ge x\right\} = \bigcap_n \{f_n \ge x\} \in \mathcal{A} \quad \text{für alle } x \in \bar{\mathbb{R}}\ ,$$
$$\limsup_{n\to\infty} f_n = \inf_n \sup_{k\ge n} f_k, \quad \liminf_{n\to\infty} f_n = \sup_n \inf_{k\ge n} f_k$$

b) $\left\{\lim_{n\to\infty} f_n \text{ existiert in } \bar{\mathbb{R}}\right\} \in \mathcal{A}$, da die Menge folgendermaßen geschrieben werden kann:

$$\left\{\liminf_{n\to\infty} f_n = +\infty\right\} \cup \left\{\limsup_{n\to\infty} f_n = -\infty\right\}$$
$$\cup \left(\left\{\liminf_{n\to\infty} f_n > -\infty\right\} \cap \left\{\limsup_{n\to\infty} f_n < +\infty\right\} \cap \left\{\limsup_{n\to\infty} f_n - \liminf_{n\to\infty} f_n \le 0\right\}\right)\ .$$

Dann ist $1_E \cdot \lim_{n\to\infty} f_n = 1_E \cdot \limsup_{n\to\infty} f_n$ messbar.

□

Messbare Funktionen lassen sich durch *Elementarfunktionen* monoton approximieren:

4.10 Definition *Sei $(\Omega, \mathcal{A})$ ein messbarer Raum. $f : \Omega \to \mathbb{R}$ heißt $\mathcal{A}$-Elementarfunktion, falls $f = \sum_{i=1}^n \alpha_i 1_{A_i}$ für gewisse $\alpha_i \ge 0$ und paarweise disjunkte $A_i \in \mathcal{A}$.*

4.11 Satz
Sei $(\Omega, \mathcal{A})$ ein messbarer Raum und $f : \Omega \to [0, +\infty]$ messbar.

a) Es gibt eine Folge von $\mathcal{A}$-Elementarfunktionen f_n derart, dass $f_n \nearrow f$.

b) Es gibt eine Folge messbarer Mengen A_n und $\alpha_n \ge 0$ derart, dass $f = \sum_{n=1}^\infty \alpha_n 1_{A_n}$.

Beweis:

a) Wähle

$$f_n(\omega) = \begin{cases} k2^{-n} & \text{falls } k2^{-n} \leq f(\omega) < (k+1)2^{-n},\ 0 \leq k < n2^n, \\ n & \text{falls } n \leq f(\omega) \leq +\infty \end{cases}$$
$$= \sum_{k=0}^{n2^n-1} k \cdot 2^{-n} \cdot 1_{\{k2^{-n} \leq f < (k+1)2^{-n}\}} + n \cdot 1_{\{f \geq n\}}\ .$$

b) Für jedes n ist $f_{n+1} - f_n$ eine endliche Linearkombination von Indikatorfunktionen. Da $f_{n+1} - f_n \geq 0$, zeigt man leicht, dass alle Koeffizienten nichtnegativ sind. Also ist $f = f_1 + \sum_{n=1}^{\infty}(f_{n+1} - f_n)$ eine Darstellung der gesuchten Form.

□

Maße lassen sich durch messbare Abbildungen transportieren:

4.12 Satz
Seien $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ messbare Räume, $T : \Omega \to \Omega'$ messbar, und sei μ ein Maß auf $(\Omega, \mathcal{A})$. Definiere

$$\mu \circ T^{-1} : \mathcal{A}' \to [0, +\infty] \quad \textit{durch} \quad \mu \circ T^{-1}(A') = \mu(T^{-1}A')\ .$$

Dann ist $\mu \circ T^{-1}$ ein Maß auf $\mathcal{A}'$ (das <u>Bildmaß</u> von μ unter T). Es ist ein Wahrscheinlichkeitsmaß genau dann, wenn μ eines ist.

Beweis: Elementar, da $T^{-1}\emptyset = \emptyset$ und $T^{-1}(\biguplus_{n=}^{\infty} A_n) = \biguplus_{n=1}^{\infty} T^{-1}A_n$. □

Kapitel 5

Zufallsvariablen, Unabhängigkeit

In diesem Abschnitt sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Die Mengen $A \in \mathcal{A}$ heißen auch Ereignisse. Oft kann man nicht alle Ereignisse beobachten, sondern nur solche, die sich durch gegebene "Messinstrumente" erfassen lassen. Solche Situationen werden mit Hilfe messbarer Funktionen modelliert, die wir dann *Zufallsvariablen* nennen werden.

5.1 Definition *a) Sei $(M, \mathcal{M})$ ein messbarer Raum. Eine $\mathcal{A}$-$\mathcal{M}$-messbare Abbildung $X : \Omega \to M$ heißt eine M-wertige Zufallsvariable. Ist $(M, \mathcal{M}) = (\mathbb{R}, \mathcal{B})$ oder $= (\bar{\mathbb{R}}, \bar{\mathcal{B}})$, so sprechen wir einfach von einer Zufallsvariablen.*

b) Das Wahrscheinlichkeitsmaß $P_X := P \circ X^{-1}$ auf $(M, \mathcal{M})$ heißt die Verteilung von X.

c) Ist μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$, so heißt $F : \mathbb{R} \to [0, 1], x \mapsto \mu((-\infty, x])$ die Verteilungsfunktion von μ. Ist $\mu = P_X$ die Verteilung einer reellwertigen Zufallsvariablen, so spricht man auch von der Verteilungsfunktion von X, also $F(x) = P_X((-\infty, x]) = P(X \leq x)$.

d) Eine Familie $(X_i)_{i \in I}$ heißt identisch verteilt, falls P_{X_i} für alle i die gleiche Verteilung ist.

5.2 Bemerkung Sind F_1 und F_2 Verteilungsfunktionen von μ_1 und μ_2 und ist $F_1 = F_2$, so ist $\mu_1 = \mu_2$. Das folgt sofort aus dem Eindeutigkeitssatz für die Maßfortsetzung (Satz 2.12), denn bezeichnet $\mathcal{C}$ den $\cap$-stabilen Erzeuger $\{(-\infty, x] : x \in \mathbb{R}\}$ von $\mathcal{B}$, so bedeutet $F_1 = F_2$ gerade, dass $\mu_1|_{\mathcal{C}} = \mu_2|_{\mathcal{C}}$.

5.3 Lemma *Ist F die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes μ, so gilt:*

1) $x \leq y \Longrightarrow F(x) \leq F(y)$.

2) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$.

3) F ist rechtsseitig stetig, d.h. $\lim_{y \downarrow x} F(y) = F(x)$.

Beweis: Übung! Benutze die Monotonie und die monotone Stetigkeit von μ. □

5.4 Satz
Jede Funktion $F : \mathbb{R} \to [0, 1]$, die die Eigenschaften 1)-3) des vorhergehenden Lemmas erfüllt, ist Verteilungsfunktion einer auf $([0, 1], \mathcal{B}, P = \lambda)$ definierten Zufallsvariablen X (und damit auch eines Wahrscheinlichkeitsmaßes P_X auf $\mathbb{R}$).

Beweis: Betrachte die Quantilfunktion $\phi : (0, 1) \to \mathbb{R}$ zu F:

$$\phi(u) := \inf \{x \in \mathbb{R} : u \leq F(x)\} \ .$$

Es ist $\phi \nearrow$ (da $F \nearrow$) und $u \leq F(x) \Longrightarrow \phi(u) \leq x$ nach Definition. Umgekehrt gilt:

$$\phi(u) \leq x \implies \exists x_n \downarrow x : u \leq F(x_n) \implies u \leq F(x), \text{ da } F \text{ stetig von rechts.}$$

Also:

$$\phi(u) \le x \quad \Longleftrightarrow \quad u \le F(x) \ .$$

Definiere nun $X(\omega) = \phi(\omega)$ für $\omega \in (0,1)$ und $X(0) = X(1) = 0$ (diese Werte können beliebig gewählt werden). X ist also eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $([0,1], \mathcal{B}, \lambda)$. Dann ist für jedes $a \in \mathbb{R}$

$$P_X((-\infty, a]) = P(X \le a) = \lambda\{u \in (0,1) : \phi(u) \le a\} = \lambda\{u \in (0,1) : u \le F(a)\} = F(a)$$

Beachte, dass mit λ_F wie in Beispiel 2.4 gilt

$$P_X((a,b]) = P_X((-\infty, b]) - P_X((-\infty, a]) = F(b) - F(a) = \lambda_F((a,b]) \ .$$

Aus dem Eindeutigkeitssatz 2.12 folgt dann $P_X = \lambda_F$.

□

5.5 Beispiel Sei $\Omega = \Sigma^{\mathbb{N}}$, P das Bernoulli-Maß zum Wahrscheinlichkeitsvektor $(p_i)_{i\in\Sigma}$ auf Σ. Betrachte Zufallsvariablen $X_n : \Omega \to \mathbb{R}$, $X_n(\omega) := \omega_n$ (Messbarkeit?). Dann ist P_{X_n} für alle n das durch den Wahrscheinlichkeitsvektor $(p_i)_{i\in\Sigma}$ auf Σ gegebene Wahrscheinlichkeitsmaß.

Sei nun $\Sigma = \{0,1\}$, $p_0 = p, p_1 = 1 - p$. Setze $S_n := \sum_{k=1}^n X_k$. Aus der elementaren Stochastik ist bekannt, dass P_{S_n} die *Binomialverteilung* mit Parameter (n,p) ist,

$$P(S_n = k) = P_{S_n}(\{k\}) = \binom{n}{k}(1-p)^k p^{n-k} \ .$$

Sei wieder $\Omega = \Sigma^{\mathbb{N}}$ und wähle ein $s \in \Sigma$ mit $p_s > 0$. Setze $Z(\omega) := \min\{n > 0 : \omega_n = s\}$, $Z(\omega) = +\infty$ falls kein solches ω_n existiert. Dann ist P_Z die *geometrische Verteilung* mit Parameter $1 - p_s$: $P(Z = 1) = p_s$, $P(Z = 2) = (1-p_s)p_s$, $P(Z = 3) = (1-p_s)^2 p_s, \ldots$.

5.6 Bemerkung Viele weitere Beispiele klassischer Verteilungen findet man in dem Textbuch von Georgii [Georgii].

5.7 Definition (Unabhängige Familien von Mengensystemen) *Eine Familie $(\mathcal{E}_i)_{i\in I}$, $\mathcal{E}_i \subseteq \mathcal{A}$ $(i \in I)$ mit beliebiger Indexmenge I heißt unabhängig, falls für jede endliche Teilmenge $I_0 \subseteq I$ und jede Wahl von $E_i \in \mathcal{E}_i$ $(i \in I_0)$ gilt*

$$P\left(\bigcap_{i\in I_0} E_i\right) = \prod_{i\in I_0} P(E_i) \ . \qquad (*)$$

5.8 Satz (Unabhängige Familien von Mengensystemen)

a) $(\mathcal{E}_i)_{i\in I}$ ist unabhängig genau dann, wenn $(\mathcal{E}_i)_{i\in I_0}$ für jede endliche Teilmenge $I_0 \subseteq I$ unabhängig ist.

b) Ist I endlich und ist $\Omega \in \mathcal{E}_i$ für alle $i \in I$, so ist $(\mathcal{E}_i)_{i\in I}$ unabhängig genau dann, wenn (*) für $I_0 = I$ gilt.

c) Ist $(\mathcal{E}_i)_{i\in I}$ unabhängig und sind alle $\mathcal{E}_i \cup \{\emptyset\}$ $\cap$-stabil, so ist auch $(\sigma(\mathcal{E}_i)_{i\in I})$ unabhängig.

Beweis:

a) Klar, da jedes endliche I_0 Teilmenge von sich selbst ist.

b) $\Longrightarrow$: Klar.
$\Longleftarrow$: Sei $I_1 \subseteq I$, $E_i \in \mathcal{E}_i$ $(i \in I_1)$, $E_i := \Omega$ $(i \notin I_1)$. Dann ist

$$P\left(\bigcap_{i\in I_1} E_i\right) = P\left(\bigcap_{i\in I} E_i\right) = \prod_{i\in I} P(E_i) = \prod_{i\in I_1} P(E_i) \ .$$

c) O.B.d.A. I endlich, siehe a). Zu zeigen ist: (*) gilt für jede Wahl von $E_i \in \sigma(\mathcal{E}_i)$, $i \in I$. Sei dazu $i_0 \in I$, $I_1 = I \setminus \{i_0\}$, o.B.d.A. $I_1 \neq \emptyset$. Für jede Wahl der $E_i \in \mathcal{E}_i$, $i \in I_1$, sind

$$A \mapsto P\left(A \cap \bigcap_{i \in I_1} E_i\right) \quad \text{und} \quad A \mapsto P(A) \cdot \prod_{i \in I_1} P(E_i)$$

endliche Maße auf $(\Omega, \mathcal{A})$, die auf $\mathcal{E}_{i_0} \cup \{\emptyset\}$ und damit auch auf $\sigma(\mathcal{E}_{i_0})$ übereinstimmen (Satz 2.12). Also sind $\sigma(\mathcal{E}_{i_0}), \mathcal{E}_i (i \in I)$ unabhängig. Durch Induktion über $i_0 \in I$ folgt die Behauptung.

□

5.9 Definition (von Zufallsvariablen erzeugte σ-Algebra) *Ist $(X_i)_{i \in I}$ eine Familie von Zufallsvariablen, so heißt*

$$\sigma(X_i : i \in I) := \bigcap_{\substack{\mathcal{F} \subseteq \mathcal{A}\ \sigma\text{-Algebra} \\ X_i\ \mathcal{F}\text{-messbar}\ \forall\, i \in I}} \mathcal{F} = \sigma\left(\bigcup_{i \in I} X_i^{-1} \bar{\mathcal{B}}\right)$$

die von X_i $(i \in I)$ erzeugte σ-Algebra. Insbesondere: $\sigma(X) = X^{-1}\bar{\mathcal{B}}$ für jede Zufallsvariable X.

5.10 Definition (Unabhängigkeit von Zufallsvariablen)
Die Zufallsvariablen $(X_i)_{i \in I}$ sind unabhängig, falls die σ-Algebren $(\sigma(X_i))_{i \in I}$ unabhängig sind.

5.11 Übung Die Ereignisse $E_i \in \mathcal{A}$ $(i \in I)$ heißen unabhängig, wenn für jede endliche Teilmenge $I_0 \subseteq I$ gilt $P\left(\bigcap_{i \in I_0} E_i\right) = \prod_{i \in I_0} P(E_i)$. Zeigen Sie, dass die folgenden Aussagen äquivalent sind:

i) $(E_i)_{i \in I}$ ist unabhängig

ii) $(\sigma(\{E_i\}))_{i \in I}$ ist unabhängig

iii) $(1_{E_i})_{i \in I}$ ist unabhängig

5.12 Bemerkung Sind $(X_i)_{i \in I}$ unabhängig, $f_i : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ messbar, so sind $(f_i(X_i))_{i \in I}$ unabhängig, denn $\sigma(f_i \circ X_i) = X_i^{-1}(f_i^{-1}\bar{\mathcal{B}}) \subseteq X_i^{-1}\bar{\mathcal{B}} = \sigma(X_i)$.

5.13 Beispiel (Unabhängigkeit des Bernoulli-Prozesses) Sei $\Omega = \Sigma^{\mathbb{N}}$, P ein Bernoulli-Maß auf Ω und $X_n(\omega) = \omega_n$ für alle $n \in \mathbb{N}$. Dann sind die $(X_n)_{n \in \mathbb{N}}$ unabhängig. (Übung!)

5.14 Bemerkung (Paarweise Unabhängigkeit impliziert nicht Unabhängigkeit) Sind in einer Familie $(X_i)_{i \in I}$ alle Paare (X_i, X_j) mit $i \neq j$ unabhängig, so muss $(X_i)_{i \in I}$ nicht unabhängig sein. Das gilt schon für eine Familie (X_1, X_2, X_3). *Beispiel:* Das wohl einfachste Beispiel erhält man, wenn man die Gleichverteilung P auf $\Omega = \{0, 1, 2, 3\}$ und die Zufallsvariablen X_1, X_2, X_3 mit $X_i(0) = X_i(i) = 1$ und $X_i(j) = 0$ sonst betrachtet (Übung).

5.15 Satz (Unabhängigkeit nach "Klumpung")
Seien die Zufallsvariablen $(X_i)_{i \in I}$ unabhängig und seien I_k $(k \in K)$ paarweise disjunkte Teilmengen von I. Dann sind die σ-Algebren $(\sigma(X_i : i \in I_k))_{k \in K}$ unabhängig.

Beweis: Für $k \in K$ sei

$$\mathcal{E}_k = \left\{\bigcap_{i \in I_k} A_i : A_i \in \sigma(X_i)\ (i \in I_k),\ A_i \neq \Omega \text{ für höchstens endlich viele } i\right\}.$$

Da die $\mathcal{E}_k$ $\cap$-stabil sind und da $\sigma(X_i : i \in I_k) = \sigma(\mathcal{E}_k)$, reicht es wegen Satz 5.8c) zu zeigen, dass $(\mathcal{E}_k)_{k \in K}$ unabhängig ist, und wegen Satz 5.8a) können wir annehmen, dass K endlich ist.

Seien $E_k = \bigcap_{i \in I_k} A_i^k \in \mathcal{E}_k$. Da nur endlich viele A_i^k von Ω verschieden sind, können wir in der folgenden Rechnung annehmen, dass auch die I_k endlich sind.

$$P\left(\bigcap_{k\in K} E_k\right) = P\left(\bigcap_{k\in K}\bigcap_{i\in I_k} A_i^k\right) = \prod_{k\in K}\prod_{i\in I_k} P(A_i^k) = \prod_{k\in K} P\left(\bigcap_{i\in I_k} A_i^k\right) = \prod_{k\in K} P(E_k)$$

□

Kapitel 6

Produkträume, Stochastische Prozesse

In diesem Abschnitt zeigen wir, wie man mehrere (auch unendlich viele) zufällige Einzelereignisse (-messungen, -beobachtungen) auf einem einzigen großen Wahrscheinlichkeitsraum modelliert. Unter dem Blickwinkel der Modellbildung tun wir den ersten Schritt, um aus (einfachen) Modellen für einzelne zufällige Phänomene ein (komplizierteres) Modell für die gleichzeitige Beschreibung vieler solcher Phänomene zu konstruieren.

Sei $(\Omega_t)_{t\in T}$ eine Familie von Mengen, wobei T eine beliebige Indexmenge ist. (In vielen Anwendungen ist $T = \mathbb{N}, \mathbb{Z}, \mathbb{Z}^d, \mathbb{R}$ oder $\mathbb{R}^d$ oder eine Teilmenge dieser Mengen.) Sei

$$\bigtimes_{t\in T} \Omega_t = \{(\omega_t)_{t\in T} : \omega_t \in \Omega_t \ \forall t \in T\} \ .$$

Für $\emptyset \neq S \subseteq T$ sei die kanonische Projektion

$$\pi_S^T : \bigtimes_{t\in T} \Omega_t \to \bigtimes_{t\in S} \Omega_t, \quad (\omega_t)_{t\in T} \mapsto (\omega_t)_{t\in S}$$

Ist aus dem Zusammenhang klar, um welche Menge T es geht, so schreiben wir auch einfach π_S statt π_S^T. Ist $S = \{j\}$, so schreiben wir π_j statt $\pi_{\{j\}}$.

6.1 Definition *Sind $(\Omega_t, \mathcal{A}_t)$ $(t \in T)$ messbare Räume, so heißt*

$$\left(\bigtimes_{t\in T} \Omega_t, \bigtimes_{t\in T} \mathcal{A}_t\right) \quad \textit{das } \underline{\textit{Produkt der }} (\Omega_t, \mathcal{A}_t) \ ,$$

wobei

$$\bigtimes_{t\in T} \mathcal{A}_t := \sigma\left(\bigcup_{t\in T} \pi_t^{-1} \mathcal{A}_t\right) \quad \textit{die } \underline{\textit{Produkt } \sigma\textit{-Algebra}} \textit{ ist.}$$

($\bigtimes_{t\in T} \mathcal{A}_t$ ist also die kleinste σ-Algebra auf $\bigtimes_{t\in T} \Omega_t$, für die alle Projektionen π_t messbar sind.)

6.2 Bemerkungen

a) Ist z.B. $T = \{1, 2, 3\}$ und $\Omega_t = \mathbb{R}$, $\mathcal{A}_t = \mathcal{B}$ für alle $t \in T$, so ist $\pi_2^{-1}\mathcal{A}_2 = \{\mathbb{R} \times A \times \mathbb{R} : A \in \mathcal{B}\}$.

b) Definitionsgemäß ist jedes $\pi_t^T : \Omega_T \to \Omega_t$ messbar.

c) *Beachte:* $\bigtimes_{t\in T} \mathcal{A}_t \neq \{\bigtimes_{t\in T} A_t : A_t \in \mathcal{A}_t \ \forall t \in T\}$. Ganz allgemein gilt in keiner der beiden Richtungen Inklusion, aber für abzählbare T gilt "$\supseteq$".

6.3 Satz
Seien $(\Omega, \mathcal{A})$ und $(\Omega_t, \mathcal{A}_t)$ $(t \in T)$ messbare Räume. Dann gilt: $f : \Omega \to \bigtimes_{t\in T} \Omega_t$ ist messbar genau dann, wenn $\pi_t \circ f : \Omega \to \Omega_t$ für alle $t \in T$ messbar ist.

Beweis: "$\Longrightarrow$": π_t ist $(\times_{s\in T}\mathcal{A}_s)$-$\mathcal{A}_t$-messbar, also $\pi_t \circ f$ $\mathcal{A}$-$\mathcal{A}_t$-messbar.
"$\Longleftarrow$": $f^{-1}(\pi_t^{-1}\mathcal{A}_t) = (\pi_t \circ f)^{-1}\mathcal{A}_t \subseteq \mathcal{A}$ für alle $t \in T$. Also ist

$$f^{-1}\left(\underset{t\in T}{\times}\mathcal{A}_t\right) = f^{-1}\left(\sigma\left(\bigcup_{t\in T}\pi_t^{-1}\mathcal{A}_t\right)\right) \overset{\text{Satz 4.4}}{=} \sigma\left(f^{-1}\left(\bigcup_{t\in T}\pi_t^{-1}\mathcal{A}_t\right)\right)$$
$$= \sigma\left(\bigcup_{t\in T} f^{-1}\left(\pi_t^{-1}\mathcal{A}_t\right)\right) \subseteq \sigma(\mathcal{A}) = \mathcal{A}$$

□

6.4 Korollar $\emptyset \neq S \subseteq T \Longrightarrow \pi_S^T$ *ist* $(\times_{t\in T}\mathcal{A}_t)$-$(\times_{t\in S}\mathcal{A}_t)$*-messbar.*

Beweis: Für alle $t \in S$ ist $\pi_t^S \circ \pi_S^T = \pi_t^T$ messbar. □

6.5 Satz
Seien $(\Omega_t, \mathcal{A}_t)$, $t \in T$, *messbare Räume,* $\mathcal{A}_t = \sigma(\mathcal{C}_t)$ *für alle* $t \in T$. *Dann ist*

$$\underset{t\in T}{\times}\mathcal{A}_t = \sigma\left(\bigcup_{t\in T}\pi_t^{-1}\mathcal{C}_t\right)$$

Beweis: "$\supseteq$": klar ✔
"$\subseteq$": Für jedes $j \in T$ ist $\pi_j^{-1}\mathcal{C}_j \subseteq \sigma\left(\bigcup_{t\in T}\pi_t^{-1}\mathcal{C}_t\right)$, also nach Satz 4.4

$$\pi_j^{-1}\mathcal{A}_j = \pi_j^{-1}\sigma(\mathcal{C}_j) = \sigma(\pi_j^{-1}\mathcal{C}_j) \subseteq \sigma\left(\bigcup_{t\in T}\pi_t^{-1}\mathcal{C}_t\right) .$$

Es folgt

$$\underset{j\in T}{\times}\mathcal{A}_j = \sigma\left(\bigcup_{j\in T}\pi_j^{-1}\mathcal{A}_j\right) \subseteq \sigma\left(\bigcup_{t\in T}\pi_t^{-1}\mathcal{C}_t\right) .$$

□

6.6 Definition (Zylindermengen) *Seien* $(\Omega_t, \mathcal{A}_t)$, $t \in T$, *messbare Räume. Wir bezeichnen mit*

$$\mathcal{Z} = \mathcal{Z}(\mathcal{A}_t, t \in T) := \bigcup_{\substack{S\subseteq T\\ S \text{ endlich}}}\left\{\underset{t\in T}{\times}A_t : A_t \in \mathcal{A}_t\ (t \in S), A_t = \Omega_t\ (t \notin S)\right\}$$

die Zylindermengen (oder messbaren Rechtecke) zu $(\mathcal{A}_t, t \in T)$.

6.7 Satz
$\mathcal{Z} = \mathcal{Z}(\mathcal{A}_t, t \in T)$ *ist ein Semiring und* $\sigma(\mathcal{Z}) = \times_{t\in T}\mathcal{A}_t$.

Beweis:

SR1) Setze $S = \{t_0\}$, $A_{t_0} = \emptyset$. Dann ist $\times_{t\in T}A_t = \emptyset$.

SR2) Seien $(\times_{t\in T}A_t), (\times_{t\in T}B_t) \in \mathcal{Z}$. Dann ist

$$\underset{t\in T}{\times}A_t \cap \underset{t\in T}{\times}B_t = \underset{t\in T}{\times}(A_t \cap B_t) \quad \text{mit } A_t \cap B_t \in \mathcal{A}_t\ (t \in T)$$

und $A_t \cap B_t = \Omega_t$ falls $A_t = B_t = \Omega_t$, also für alle t bis auf endlich viele.

SR3) Seien $(\times_{t\in T} A_t), (\times_{t\in T} B_t) \in \mathcal{Z}$, $(\times_{t\in T} A_t) \subseteq (\times_{t\in T} B_t)$. Dann ist $A_t \subseteq B_t$ für alle $t \in T$. Ist $S \subseteq T$ endlich und $A_t = \Omega_t \;\forall\, t \in S$, so ist auch $B_t = \Omega_t \;\forall\, t \in S$. Daher ist

$$\times_{t\in T} B_t = \biguplus_{j\in\{0,1\}^S} \times_{t\in T} C_t^j$$

mit $C_t^j = \Omega_t$ $(t \notin S)$, $C_t^j = A_t$ $(t \in S, j_t = 0)$, $C_t^j = B_t \setminus A_t$ $(t \in S, j_t = 1)$. Es folgt, dass

$$\times_{t\in T} B_t \setminus \times_{t\in T} A_t = \times_{t\in T} B_t \setminus \times_{t\in T} C_t^{(0,\dots,0)}$$

endliche disjunkte Vereinigung von Mengen in $\mathcal{Z}$ ist.

Schließlich ist

$$\times_{t\in T} A_t = \bigcap_{t\in T} \pi_t^{-1} A_t = \bigcap_{t\in S} \pi_t^{-1} A_t \in \times_{t\in T} \mathcal{A}_t \quad \text{da } S \text{ endlich ist,}$$

also $\sigma(\mathcal{Z}) \subseteq \times_{t\in T} \mathcal{A}_t$, und die Umkehrung gilt da $\pi_t^{-1}\mathcal{A}_t \subseteq \mathcal{Z}$ für alle $t \in T$. □

6.8 Korollar *Sind T_1, T_2 disjunkte Indexmengen, $(\Omega_t, \mathcal{A}_t)_{t\in T_1}$ und $(\Omega_t, \mathcal{A}_t)_{t\in T_2}$ Familien messbarer Räume, so ist*

$$\left(\times_{t\in T_1} \mathcal{A}_t\right) \times \left(\times_{t\in T_2} \mathcal{A}_t\right) = \left(\times_{t\in T_1\cup T_2} \mathcal{A}_t\right)$$

Beweis: Wir haben die folgenden beiden Inklusionen:

$$\mathcal{Z}(\mathcal{A}_t, t \in T_1 \cup T_2) \subseteq \mathcal{Z}\left(\times_{t\in T_1} \mathcal{A}_t, \times_{t\in T_2} \mathcal{A}_t\right) \subseteq \times_{t\in T_1\cup T_2} \mathcal{A}_t\,.$$

Die erste ist klar, da jede Menge $\times_{t\in T_1\cup T_2} A_t$ mit $A_t \in \mathcal{A}_t$ als $(\times_{t\in T_1} A_t)\times(\times_{t\in T_2} A_t)$, d.h. als Element von $\mathcal{Z}\,(\times_{t\in T_1} \mathcal{A}_t, \times_{t\in T_2} \mathcal{A}_t)$ aufgefasst werden kann. Für die zweite Inklusion seien $T := T_1 \cup T_2$, $\Omega^{(i)} := \times_{t\in T_i} \Omega_t$ und $B^{(i)} \in \times_{t\in T_i} \mathcal{A}_t$ beliebig $(i = 1, 2)$. Dann ist

$$B^{(1)} \times B^{(2)} = (B^{(1)} \times \Omega^{(2)}) \cap (\Omega^{(1)} \times B^{(2)}) \in \sigma\left(\pi_{T_1}^{-1}\left(\times_{t\in T_1} \mathcal{A}_t\right) \cup \pi_{T_2}^{-1}\left(\times_{t\in T_2} \mathcal{A}_t\right)\right)\ ,$$

und das ist gerade die zweite Inklusion. Die Behauptung folgt nun durch Übergang zu den erzeugten σ-Algebren. □

6.9 Definition (Stochastischer Prozess) *Sei $(X_t)_{t\in T}$ eine Familie von Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Wir nennen $(X_t)_{t\in T}$ auch einen stochastischen Prozess (oder einfach* Prozess*), insbesondere, wenn $T = \mathbb{N}, \mathbb{Z}^d, \mathbb{Z}_+^d, \mathbb{R}^d$ oder $\mathbb{R}_+^d$ ist.*

Bezeichne $\mathbb{R}^T := \times_{t\in T} \mathbb{R}$ und $\mathcal{B}^T := \times_{t\in T} \mathcal{B}$, und sei $\mathcal{X} : \Omega \to \mathbb{R}^T$ definiert durch

$$(\mathcal{X}(\omega))_t = X_t(\omega) \quad (\textit{so dass } \pi_t \circ \mathcal{X} = X_t).$$

$\mathcal{X}$ ist $\mathcal{A}$-$\mathcal{B}^T$-messbar wegen Satz 6.3, und $P_{\mathcal{X}} = P \circ \mathcal{X}^{-1}$ ist die Verteilung des Prozesses $(X_t)_{t\in T}$ auf $\mathbb{R}^T$.

6.10 Definition (Stationarität) *Sei $(T, +)$ eine kommutative Halbgruppe.*

a) Für $t_0 \in T$ bezeichne

$$S_{t_0} : \mathbb{R}^T \to \mathbb{R}^T, \quad (S_{t_0}x)_t = x_{t+t_0}$$

die Verschiebung (shift) um t_0. (Beachte: Da $\pi_t \circ S_{t_0} = \pi_{t+t_0}$ für alle $t \in T$, ist S_{t_0} nach Satz 6.3 messbar.)

b) Der Prozess $(X_t)_{t\in T}$ heißt stationär, falls für seine Verteilung $P_{\mathcal{X}}$ auf $\mathbb{R}^T$ gilt: $P_{\mathcal{X}} \circ S_{t_0}^{-1} = P_{\mathcal{X}}$ für alle $t_0 \in T$.

6.11 Satz (Charakterisierung der Stationarität)
Sei T wie vorher. Der Prozess $(X_t)_{t\in T}$ ist stationär genau dann, wenn für jede endliche Menge $R \subseteq T$, jedes $t_0 \in T$ und jede Familie $(B_t)_{t\in R}$ aus $\mathcal{B}$ gilt:

$$P(X_t \in B_t \;\forall\, t \in R) = P(X_{t+t_0} \in B_t \;\forall\, t \in R) \tag{*}$$

Beweis: Seien $(B_t)_{t\in R}$ gegeben. Setze $B_t = \mathbb{R}$ für $t \in T \setminus R$ und betrachte den Zylinder $B = \times_{t\in T} B_t$. Dann ist (*) äquivalent zu $P_{\mathcal{X}}(B) = P_{\mathcal{X}}(S_{t_0}^{-1}B)$, d.h. (*) gilt für alle solchen Mengen genau dann, wenn die Wahrscheinlichkeitsmaße $P_{\mathcal{X}} \circ S_{t_0}^{-1}$ und $P_{\mathcal{X}}$ auf $\mathcal{Z}$ übereinstimmen. Da $\mathcal{Z}$ ein Semiring und $\sigma(\mathcal{Z}) = \mathcal{B}^T$ ist (Satz 6.7), ist das gleichbedeutend mit $P_{\mathcal{X}} \circ S_{t_0}^{-1} = P_{\mathcal{X}}$ aus Satz 2.12. □

6.12 Korollar *Ist $(X_t)_{t\in T}$ unabhängig identisch verteilt, so ist $(X_t)_{t\in T}$ stationär.*

Beweis:

$$\begin{aligned} P(X_t \in B_t \;\forall\, t \in R) &= \prod_{t\in R} P_{X_t}(B_t) = \prod_{t\in R} P_{X_{t+t_0}}(B_t) \\ &= P(X_{t+t_0} \in B_t \;\forall\, t \in R) \end{aligned}$$

□

6.13 Bemerkung (Kanonisches Modell eines Prozesses) Sei wieder T eine beliebige Indexmenge, $\Omega = \mathbb{R}^T$. Betrachte die messbaren Abbildungen $\pi_t : \Omega \to \mathbb{R}$. Ist μ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B}^T)$, so ist $(\pi_t)_{t\in T}$ ein stochastischer Prozess auf $(\Omega, \mathcal{B}^T, \mu)$ mit Verteilung μ auf $\mathbb{R}^T$. Ist $(T, +)$ eine kommutative Halbgruppe, so ist dieser Prozess stationär genau dann, wenn $\mu \circ S_t^{-1} = \mu$ für alle $t \in T$.

Ist nun $(X_t)_{t\in T}$ ein stochastischer Prozess auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, so ist $(\pi_t)_{t\in T}$ ein Prozess auf $(\mathbb{R}^T, \mathcal{B}^T, P_{\mathcal{X}})$, der die gleiche Verteilung wie $(X_t)_{t\in T}$ hat, denn $X_t = \pi_t \circ \mathcal{X}$ ($\mathcal{X}$ aus Definition 6.9). Er heißt das kanonische Modell für $(X_t)_{t\in T}$.

6.14 Bemerkung Im folgenden Diagramm ist das Zusammenspiel der verschiedenen Wahrscheinlichkeitsräume bei der Definition der Stationarität eines Prozesses schematisch dargestellt:

$$\begin{tikzcd}
(\Omega,\mathcal{A},P) \arrow[r, "\mathcal{X}"] \arrow[dr, "X_t"'] & (\mathbb{R}^T,\mathcal{B}^T,P_\phi) \arrow[r, "S_t"] \arrow[d, "\pi_t"] & (\mathbb{R}^T,\mathcal{B}^T,P_\phi) \arrow[dl, "\pi_0"] \\
& (\mathbb{R},\mathcal{B},P_{X_t}) &
\end{tikzcd}$$

6.15 Definition (Ergodischer Prozess) *Sei $(T, +)$ eine kommutative Halbgruppe. Ein stationärer Prozess $(X_t)_{t\in T}$ mit Verteilung $P_{\mathcal{X}}$ heißt ergodisch, falls für alle $A \in \mathcal{B}^T$ gilt*

$$S_t^{-1}(A) = A \text{ für alle } t \in T \implies P_{\mathcal{X}}(A) = 0 \text{ oder } 1$$

6.16 Satz (u.i.v. Prozesse sind ergodisch)
Sei $(T, +)$ eine kommutative Halbgruppe mit der Eigenschaft, dass für jedes endliche $R \subseteq T$ ein $t \in T$ existiert, für das $(R + t) \cap R = \emptyset$. Ist dann $(X_t)_{t\in T}$ unabhängig identisch verteilt, so ist $(X_t)_{t\in T}$ ergodisch.

Beweis: Sei $A \in \mathcal{B}^T$, $S_t^{-1}(A) = A$ für alle $t \in T$. Sei $\epsilon > 0$ beliebig. Wegen des Approximationssatzes 3.13 (in Verbindung mit Satz 6.7) gibt es eine endliche Vereinigung A' von Zylindern aus $\mathcal{Z}$ derart, dass $P_\mathcal{X}(A \triangle A') < \epsilon$. Da $(X_t)_t$ stationär ist (Korollar 6.12), folgt für alle $t \in T$

$$P_\mathcal{X}\left(S_t^{-1}A \triangle S_t^{-1}A'\right) = P_\mathcal{X}\left(S_t^{-1}(A \triangle A')\right) = P_\mathcal{X}(A \triangle A') < \epsilon$$
$$P_\mathcal{X}\left((A \cap S_t^{-1}A) \triangle (A' \cap S_t^{-1}A')\right) \le P_\mathcal{X}\left(A \triangle A'\right) + P_\mathcal{X}\left(S_t^{-1}(A \triangle A')\right) \le 2\epsilon$$

Da A' eine endliche Vereinigung von Zylindern ist, gibt es eine endliche Teilmenge $R \subseteq T$ und ein $B' \in \mathcal{B}^R$ derart, dass $A' = (\pi_R^T)^{-1}B'$. Sei nun $t_0 \in T$ so, dass $(R + t_0) \cap R = \emptyset$. Dann ist

$$\begin{aligned}
P_\mathcal{X}\left(A' \cap S_{t_0}^{-1}A'\right) &= P_\mathcal{X}\left((\pi_R^T)^{-1}B' \cap (\pi_{R+t_0}^T)^{-1}B'\right) \\
&= P\left((X_t)_{t\in R} \in B' \text{ und } (X_t)_{t \in R+t_0} \in B'\right) \\
&= P\left((X_t)_{t\in R} \in B'\right) \cdot P\left((X_t)_{t \in R+t_0} \in B'\right) \\
&= P_\mathcal{X}\left(A'\right) \cdot P_\mathcal{X}\left(S_{t_0}^{-1}A'\right)
\end{aligned}$$

Zusammen folgt

$$\begin{aligned}
P_\mathcal{X}\left(A\right) - P_\mathcal{X}(A)^2 &= P_\mathcal{X}\left(A \cap S_{t_0}^{-1}A\right) - P_\mathcal{X}\left(A\right) \cdot P_\mathcal{X}\left(S_{t_0}^{-1}A\right) \\
&\le P_\mathcal{X}\left(A' \cap S_{t_0}^{-1}A'\right) - P_\mathcal{X}\left(A'\right) \cdot P_\mathcal{X}\left(S_{t_0}^{-1}A'\right) + 4\epsilon \\
&= 4\epsilon
\end{aligned}$$

Da $\epsilon > 0$ beliebig war, folgt $P_\mathcal{X}\left(A\right) - P_\mathcal{X}(A)^2 = 0$, d.h. $P_\mathcal{X}(A) = 0$ oder 1. □

6.17 Beispiel Die mit der gewöhnlichen Addition versehenen Indexmengen $T = \mathbb{Z}, \mathbb{Z}_+, \mathbb{Z}^d, \mathbb{R}, \mathbb{R}_+, \mathbb{R}^d$ erfüllen die Voraussetzung des Satzes.

6.18 Bemerkung Die Überlegungen in 6.9-6.16 bleiben - mit unveränderten Beweisen - auch für Zufallsvariablen mit Werten in anderen messbaren Räumen als $(\mathbb{R}, \mathcal{B})$ richtig, insbesondere für $\bar{\mathbb{R}}$-wertige X_t. Im nächsten Kapitel werden wir ein Beispiel kennen lernen, das auf einem Prozess $\{0,1\}^2$–wertiger, unabhängiger, identisch verteilter Zufallsvariablen fußt. Die explizite Konstruktion solcher Prozesse liefern wir in einem späteren Kapitel nach.

Beispiel: Perkolation

Das Beispiel, das wir in diesem Abschnitt näher betrachten wollen, gehört zur großen Klasse der *Perkolationsprobleme*. Eine einfache (idealisierte) physikalische Interpretation für unser Modell ist die folgende: In einem unendlich großen Netzwerk von Rohrleitungen, die an Knotenpunkten miteinander verknüpft sind, Ist jedes Teilstück mit einer gewissen Wahrscheinlichkeit p unabhängig von allen anderen Teilstücken geöffnet. Wie groß ist die Wahrscheinlichkeit, dass es ein unendlich großes Teilnetzwerk gibt, in dem jegliche zwei beliebigen Knoten durch eine komplett offene Leitung miteinander verbunden sind? Die Antwort darauf hängt natürlich von der Struktur des Netzes ab, deshalb machen wir einige spezielle Annahmen.

Wir betrachten den ungerichteten Graphen $\mathcal{G} = (\mathbb{Z}^2, E)$ mit Kantenmenge

$$E = \left\{ \langle i, i+v \rangle : \, i \in \mathbb{Z}^2, \; v = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ oder } v = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \, .$$

Anschaulich: $\mathcal{G}$ ist $\mathbb{Z}^2$ mit den waagerechten und senkrechten "Gitterlinien" als Kanten. "Ungerichtet" bedeutet hier, dass wir $\langle i, j \rangle$ mit $\langle j, i \rangle$ identifizieren. Wir werden einige elementare graphentheoretische Begriffsbildungen verwenden, ohne an dieser Stelle immer ganz strenge Definitionen zu geben.

- ▷ Eine Teilmenge $\tilde{E} \subseteq E$ definiert den Teilgraphen $\mathcal{G}(\tilde{E}) := (\mathbb{Z}^2, \tilde{E})$ von $\mathcal{G}$.
- ▷ Eine Folge $(i_0, i_1, \dots, i_n)$ aus $\mathbb{Z}^2$ ist ein *Weg* in $\mathcal{G}(\tilde{E})$, falls $\langle i_{k-1}, i_k \rangle \in \tilde{E}$ für alle $k = 1, \dots, n$.
- ▷ Eine Teilmenge $J \subseteq \mathbb{Z}^2$ heißt $\underline{\mathcal{G}(\tilde{E})\text{-zusammenhängend}}$, falls für beliebige $i, j \in J$ ein Weg in $\mathcal{G}(\tilde{E})$ von i nach j existiert.

Sei $(X_e)_{e \in E}$ unabhängig identisch verteilt auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P_p)$ mit

$$P_p(X_e = 1) = p \text{ und } P_p(X_e = 0) = 1 - p \text{ für ein } p \in [0, 1].$$

Mit $E(\omega)$ bezeichnen wir die *zufällige Menge*

$$E(\omega) := \{e \in E : X_e(\omega) = 1\}$$

und mit $J_0(\omega)$ die $\mathcal{G}(E(\omega))$-Zusammenhangskomponente von $\mathbb{Z}^2$ die $0 \in \mathbb{Z}^2$ enthält. ($J_0(\omega)$ kann also die Menge sein, die nur den Nullpunkt enthält.) Die folgenden Teilmengen von Ω sind messbar:

$$\begin{aligned} U_0 &:= \{\omega \in \Omega : J_0(\omega) \text{ ist unendlich}\} \\ U &:= \{\omega \in \Omega : \text{Es gibt eine unendliche } \mathcal{G}(E(\omega))\text{-Zusammenhangskomponente}\} \end{aligned}$$

Für U_0 folgt das aus

$$U_0 = \bigcap_{n \in \mathbb{N}} \; \bigcup_{i \in \mathbb{Z}^2, |i| \geq n} \; \bigcup_{(i_0 = 0, \dots, i_k = i) \text{ Weg in } \mathcal{G}} \{\omega : X_{\langle i_{j-1}, i_j \rangle}(\omega) = 1 (j = 1, \dots, k)\} \, ,$$

für U zeigt man das ähnlich (Übung!).

6.19 Satz (Kritischer Perkolationsparameter)
a) Die Abbildungen $p \mapsto P_p(U_0)$ *und* $p \mapsto P_p(U)$ *sind monoton wachsend*

b) Es gibt ein $p_c \in [\frac{1}{3}, \frac{4}{5})$ *(*kritischer Parameter*) derart, dass*

(i) $\forall\, p \in [0, p_c) : P_p(U_0) = P_p(U) = 0$

(ii) $\forall\, p \in (p_c, 1] : P_p(U_0) > 0$, $P_p(U) = 1$

6.20 Bemerkung (Verhalten am kritischen Parameter) Der genaue Wert des kritischen Parameters in diesem Modell ist bekannt: $p_c = \frac{1}{2}$, siehe [G. Grimmett, *Percolation*, Theorem 9.11]. Ein Teil des Beweises - das dortige Lemma 9.12 - liefert darüber hinaus, dass $P_{\frac{1}{2}}(U_0) = P_{\frac{1}{2}}(U) = 0$ und dass $p \mapsto P_p(U_0)$ auf ganz $[0, 1]$ stetig ist.

Beweis von Satz 6.19:

a) Um die Monotonie in p zu zeigen, verschafft man sich *einen* Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, auf dem die Prozesse $(X_e)_{e\in E}$ für alle $p \in [0, 1]$ in spezieller Weise darstellbar sind. Dieser Raum ist so zu wählen, dass es auf ihm einen unabhängigen identisch verteilten Prozess $(Z_e)_{e\in E}$ gibt, wobei P_{Z_e} für alle e das Lebesgue-Maß auf $[0, 1]$ ist. (Die explizite Konstruktion eines solchen Prozesses erfolgt in einem späteren Kapitel.) Ist dieser Prozess gegeben, so definieren wir für jedes p den Prozess $(X_e^p)_{e\in E}$ durch

$$X_e^p(\omega) = 1 \text{ wenn } Z_e(\omega) \le p \text{ und } X_e^p(\omega) = 0 \text{ sonst.}$$

$(X_e^p)_{e\in E}$ ist unabhängig, da $(Z_e)_{e\in E}$ unabhängig ist, und

$$P(X_e^p = 1) = P(Z_e \le p) = p, \quad P(X_e^p = 0) = P(Z_e > p) = 1 - p\,.$$

Der Prozess $(X_e^p)_{e\in E}$ hat unter P also die gleiche Verteilung wie der ursprüngliche Prozess $(X_e)_{e\in E}$ unter P_p, und wir können mit dieser speziellen Version arbeiten. (Man sagt auch: Die Prozesse mit den verschiedenen Parametern p wurden *gekoppelt*.) Die mit dem Prozess $(X_e^p)_{e\in E}$ assoziierten Mengen $E(\omega)$, $J_0(\omega)$, U_0 und U bezeichnen wir mit $E^p(\omega)$, $J_0^p(\omega)$, U_0^p und U^p. Dann folgt sofort:

$$p \le p' \implies \forall e\, \forall \omega : X_e^p(\omega) \le X_e^{p'}(\omega) \implies E^p(\omega) \subseteq E^{p'}(\omega) \implies U_{(0)}^p \subseteq U_{(0)}^{p'}\,,$$

und daher

$$p \le p' \implies P_p(U_{(0)}) = P(U_{(0)}^p) \le P(U_{(0)}^{p'}) = P_{p'}(U_{(0)})$$

b) Wir können jetzt wieder mit dem $\{0, 1\}$–wertigen Prozess $(X_e)_{e\in E}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P_p)$ arbeiten. Da $E = \mathbb{Z}^2 \times \left\{\binom{1}{0}, \binom{0}{1}\right\}$, können wir genau so gut den unabhängigen $\{0, 1\}^2$–wertigen Prozess $\Big((X_{\langle i, i+\binom{1}{0}\rangle}, X_{\langle i, i+\binom{0}{1}\rangle})\Big)_{i\in\mathbb{Z}^2}$ betrachten. Deshalb folgt aus Satz 6.16 die Ergodizität des betrachteten Prozesses. Außerdem können wir annehmen, dass der Prozess in kanonischer Version vorliegt (Bemerkung 6.13), also $\Omega = \{0, 1\}^E$, $X_e = \pi_e$, und P_p ist gleichzeitig die Verteilung des Prozesses.

Die Mengen U und U_0 sind nun Mengen von Konfigurationen $\omega \in \{0, 1\}^E$, und aus der Definition von U folgt sofort, dass U unter allen Translationen S_i des Gitters invariant ist. Also ist $P_p(U) = 0$ oder 1 für alle p. Trivialerweise ist $P_0(U) = 0$ und $P_1(U) = 1$. Sei

$$p_c := \inf\,\{p \in [0, 1] : P_p(U) = 1\}\,.$$

Da $p \mapsto P_p(U)$ monoton wächst, gilt

$$P_p(U) = 0 \text{ für } p < p_c, \quad P_p(U) = 1 \text{ für } p > p_c\,.$$

Da $U = \bigcup_{i\in\mathbb{Z}^2} S_i(U_0)$ ist, folgt $P_p(U_0) = 0$ für $p < p_c$ und $P_p(U_0) > 0$ für $p > p_c$. Es bleibt zu zeigen, dass $\frac{1}{3} \le p_c < \frac{23}{30}$.

"$p_c \ge \frac{1}{3}$": Sei $p \in [0, \frac{1}{3})$. Zu zeigen ist $P_p(U_0) = 0$.
Ist $\omega \in U_0$, dann ist $J_0(\omega)$ unendlich, so dass es für alle $n > 0$ einen irreduziblen Weg $(0 = i_0, i_1, \dots, i_n)$ in $J_0(\omega)$ gibt. [1] Insbesondere sind alle seine n Kanten $\langle i_{k-1}, i_k\rangle$ verschieden. (Zunächst gibt es einen Weg, dessen Endpunkt von 0 mindestens den euklidischen Abstand n hat. Reduziert man diesen so weit wie möglich, bleibt immer noch ein Weg von 0 zum selben Endpunkt, und der hat mindestens Länge n.)

Wir betrachten nun zunächst ein beliebiges festes n. Dann ist

$$U_0 \subseteq \bigcup_{\substack{(0=i_0,i_1,\dots,i_n)\\ \text{irreduzibler Weg in } \mathcal{G}}} \bigcap_{k=1}^{n} \{X_{\langle i_{k-1},i_k\rangle} = 1\} \ ,$$

so dass

$$P_p(U_0) \le \sum_{\substack{(0=i_0,i_1,\dots,i_n)\\ \text{irreduzibler Weg in } \mathcal{G}}} p^n \le \frac{4}{3}(3p)^n \ , \tag{6.1}$$

da es höchstens $4\cdot 3^{n-1}$ von 0 ausgehende irreduzible Wege der Länge n in $\mathcal{G}$ gibt. Im Limes $n \to \infty$ folgt $P_p(U_0) = 0$ falls $p < \frac{1}{3}$.

"$p_c < \frac{4}{5}$": Zu zeigen ist $P_p(\Omega \setminus U_0) < 1$ für $p = \frac{4}{5}$.
Sei $\omega \in \Omega \setminus U_0$. Dann ist $J_0(\omega)$ endlich, und es gibt einen geschlossenen irreduziblen Weg im dualen Graphen zu $\mathcal{G}(E \setminus E(\omega))$ der $J_0(\omega)$ "umrundet".[2] Daher

$$\Omega \setminus U_0 \subseteq \bigcup_{n\in\mathbb{N}} \bigcup_{\substack{(i_0,i_1,\dots,i_n=i_0)\\ \text{geschl. irred. Weg in } \mathcal{G}\text{, der 0 "umrundet"}}} \bigcap_{k=1}^{n} \{X_{\langle i_{k-1},i_k\rangle} = 0\}$$

Solche Wege haben Länge $n \ge 4$. Da jeder dieser Wege die "Halbachse" $\{(k+\frac{1}{2}, 0) : k \in \mathbb{N}\}$ in einem Punkt $(k+\frac{1}{2}, 0)$ mit $0 < k \le \frac{n}{2}$ berührt, gibt es höchstens $\frac{n}{2}\cdot 3^n$ von ihnen. Daher ist

$$P_p(\Omega \setminus U_0) \le \sum_{n=4}^{\infty} \frac{n}{2}\cdot 3^n(1-p)^n = \frac{1}{2}\cdot\frac{3(1-p)}{(1-3(1-p))^2} - \sum_{n=0}^{3}\frac{n}{2}\cdot 3^n(1-p)^n$$

Dieser Ausdruck strebt gegen 0 für $p \to 1$, und durch numerische Auswertung stellt man fest, dass er gleich 0.891, also kleiner als 1 ist für $p = \frac{4}{5}$.

□

6.21 Bemerkung (Kopplung) Die im Teil a) des Beweises benutzte Technik, Prozesse, die ursprünglich auf verschiedenen Wahrscheinlichkeitsräumen definiert sind, auf einem gemeinsamen Wahrscheinlichkeitsraum zu redifinieren, so dass

1. die originalen und die redifinierten Prozesse die gleichen Verteilungen haben und
2. die redifinierten Prozesse für "möglichst viele" ω den gleichen Wert (oder zumindest nahe beieinander liegende Werte) annehmen

nennt man *Kopplung*. Diese Technik findet vielfache Verwendung.

[1] Ein Weg $(i_0, \dots, i_n)$ heißt reduzibel, falls es möglich ist, eine endliche Anzahl seiner Kanten so zu entfernen dass ein Weg von i_0 nach i_n übrig bleibt.

[2] Das ist der Graph, dessen Knoten die Punkte $(m+\frac{1}{2}, n+\frac{1}{2})$, $m, n \in \mathbb{Z}$, und dessen Kanten gerade diejenigen nächste-Nachbar-Verbindungen zwischen diesen Punkten sind, die eine Kante aus $E \setminus E(\omega)$ senkrecht "kreuzen".

Kapitel 7

Integral und Erwartungswert

Aufbauend auf dem Maß- bzw. Wahrscheinlichkeitsbegriff führen wir jetzt die Begriffe *Integral* bzw. *Erwartungswert* ein und stellen einige elementare, aber für die Wahrscheinlichkeitstheorie zentrale Gleichungen und Ungleichungen mit Erwartungswerten bereit. Während des ganzen Kapitels sei $(\Omega, \mathcal{A}, \mu)$ ein *Maßraum*.

7.1 Definition (Integral) *Sei $f : \Omega \to \bar{\mathbb{R}}$ messbar. Ist $f \geq 0$, so ist das Integral von f (bzgl. μ) definiert als*

$$\int f \, d\mu = \sup \left\{ \sum_i \left(\inf_{\omega \in A_i} f(\omega) \right) \mu(A_i) \right\} ,$$

wobei sich das Supremum über alle endlichen messbaren Partitionen $(A_i)_i$ von Ω (d.h. Partitionen in $\mathcal{A}$-messbare Mengen) erstreckt.

Für allgemeine f betrachte

$$f^+(\omega) := \max\{f(\omega), 0\} \quad \textit{und} \quad f^-(\omega) := \max\{-f(\omega), 0\} .$$

f^+ und f^- sind messbar (Satz 4.8), nichtnegativ und $f = f^+ - f^-$. Man setzt

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu ,$$

vorausgesetzt, dass nicht $\int f^+ \, d\mu = \int f^- \, d\mu = +\infty$. In diesem Fall ist $\int f \, d\mu$ nicht definiert. Ist $f \geq 0$, so ist $f^+ = f$ und $f^- = 0$, und diese Festlegung ist mit der vorherigen für nichtnegative f konsistent.

Wenn es erforderlich ist, schreiben wir auch $\int f(\omega) \, d\mu(\omega)$ anstatt $\int f \, d\mu$.

7.2 Lemma *Seien f, g, f_n nichtnegative, messbare $\bar{\mathbb{R}}$-wertige Funktionen.*

a) Ist $f = \sum_{i=1}^n \alpha_i 1_{A_i}$, $\alpha_i \in [0, +\infty]$, $A_i \in \mathcal{A}$, so ist $\int f \, d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$.
Insbesondere ist $\int 1_A \, d\mu = \mu(A)$.

b) Aus $0 \leq f \leq g$ folgt $0 \leq \int f \, d\mu \leq \int g \, d\mu$.

c) Aus $0 \leq f_n \nearrow f$ folgt $0 \leq \int f_n \, d\mu \nearrow \int f \, d\mu$.

d) Sind $\alpha, \beta \in \bar{\mathbb{R}}$, $\alpha, \beta \geq 0$, so gilt

$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu .$$

Beweis:

a) Wir nehmen zunächst an, dass $(A_i)_i$ eine endliche *Partition* von Ω ist. Sei $(B_j)_j$ eine beliebige endliche messbare Partition von Ω, $\beta_j := \inf_{\omega\in B_j} f(\omega)$. Ist $A_i \cap B_j \neq \emptyset$, so ist $\beta_j \leq \alpha_i$, also

$$\sum_j \beta_j \mu(B_j) = \sum_{i,j} \beta_j \mu(A_i \cap B_j) \leq \sum_{i,j} \alpha_i \mu(A_i \cap B_j) = \sum_i \alpha_i \mu(A_i) \ ,$$

d.h. $\int f\, d\mu \leq \sum_i \alpha_i \mu(A_i)$. Da $(A_i)_i$ selbst eine endliche messbare Partition von Ω ist, folgt auch die Umkehrung. Den Beweis für allgemeine endliche Familien $(A_i)_i$ stellen wir zurück.

b) folgt sofort aus der Definition des Integrals.

c) Wegen b) gilt

$$0 \leq \int f_n\, d\mu \nearrow \sup_n \int f_n\, d\mu \leq \int f\, d\mu \ .$$

Es bleibt zu zeigen: $\int f\, d\mu \leq \sup_n \int f_n\, d\mu$. Sei also $(A_i)_i$ eine endliche messbare Partition von Ω, $v_i := \inf_{\omega\in A_i} f(\omega)$. Zu zeigen ist

$$\sum_i v_i \mu(A_i) \leq \sup_n \int f_n\, d\mu \ . \qquad (*)$$

Sei $0 < \epsilon < 1$. Setze $u_i := (1-\epsilon)v_i$ falls $v_i < \infty$ und $u_i := \epsilon^{-1}$ falls $v_i = \infty$. Betrachte $\omega \in A_i$.

▷ Ist $v_i = 0$, so ist $u_i = 0 \leq f_n(\omega)$ für alle n.

▷ Ist $v_i > 0$, so ist $u_i < v_i \leq f(\omega) = \sup_n f_n(\omega)$, also $u_i \leq f_n(\omega)$ für hinreichend große $n \geq n_0(\omega)$.

Setze nun

$$A_{i,n} = \{\omega \in A_i : f_n(\omega) \geq u_i\} \ .$$

Dann strebt $A_{i,n} \nearrow A_i$ mit $n \to \infty$, so dass $\mu(A_{i,n}) \nearrow \mu(A_i)$ (Satz 2.7). Beachte nun, dass $f_n \geq \sum_i u_i \cdot 1_{A_{i,n}} + \sum_i 0 \cdot 1_{A_i \setminus A_{i,n}}$. Wegen b) und dem schon bewiesenen Spezialfall von a) folgt:

$$\int f_n\, d\mu \geq \sum_i u_i \mu(A_{i,n}) \nearrow \sum_i u_i \mu(A_i) \quad \text{für } n \to \infty \ ,$$

d.h.

$$\sup_n \int f_n\, d\mu \geq \sum_i u_i \mu(A_i) = (1-\epsilon) \sum_{\substack{i \\ v_i<\infty}} v_i \mu(A_i) + \epsilon^{-1} \sum_{\substack{i \\ v_i=\infty}} \mu(A_i) \ .$$

Die Behauptung (*) folgt nun im Limes $\epsilon \to 0$.

d) Seien zunächst $f = \sum_i \alpha_i 1_{A_i}$ und $g = \sum_j \beta_j 1_{B_j}$ Elementarfunktionen. Wir können annehmen, dass $(A_i)_i$ und $(B_j)_j$ endliche Partitionen sind, indem wir beide Familien gegebenenfalls um eine Menge erweitern, auf der die jeweilige Funktion den Wert 0 annimmt. Dann ist

$$\alpha f + \beta g = \sum_{i,j} (\alpha\alpha_i + \beta\beta_j) 1_{A_i \cap B_j}$$

auch Elementarfunktion, und aus dem schon bewiesenen Spezialfall von a) folgt

$$\begin{aligned}
\int (\alpha f + \beta g)\, d\mu &= \sum_{i,j} (\alpha\alpha_i + \beta\beta_j) \mu(A_i \cap B_j) \\
&= \alpha \sum_i \left(\alpha_i \sum_j \mu(A_i \cap B_j) \right) + \beta \sum_j \left(\beta_j \sum_i \mu(A_i \cap B_j) \right) \\
&= \alpha \sum_i \alpha_i \mu(A_i) + \beta \sum_j \beta_j \mu(B_j) \\
&= \alpha \int f\, d\mu + \beta \int g\, d\mu \ .
\end{aligned}$$

Beachte, dass diese Rechnung richtig bleibt, falls wir für α, β den Wert $+\infty$ zulassen!

Für beliebige nichtnegative, messbare f, g gibt es nach Satz 4.11 Folgen $(f_n)_{n>0}$ und $(g_n)_{n>0}$ $\mathcal{A}$-messbarer Elementarfunktionen mit $0 \le f_n \nearrow f$ und $0 \le g_n \nearrow g$. Daraus folgt $0 \le \alpha f_n + \beta g_n \nearrow \alpha f + \beta g$, so dass wegen c)

$$\begin{aligned}\int (\alpha f + \beta g)\, d\mu &= \sup_n \int (\alpha f_n + \beta g_n)\, d\mu = \sup_n \left(\alpha \int f_n \, d\mu + \beta \int g_n \, d\mu \right) \\ &= \alpha \int f \, d\mu + \beta \int g \, d\mu \, .\end{aligned}$$

a') Schließlich folgt a) für beliebige messbare Mengen $(A_i)_i$ aus d):

$$\int \left(\sum_i \alpha_i 1_{A_i} + 0 \cdot 1_{\Omega \setminus A_i} \right) \, d\mu = \sum_i \alpha_i \int 1_{A_i} \, d\mu = \sum_i \alpha_i \mu(A_i) \, .$$

□

7.3 Definition (Nullmenge, fast sicher)

a) $N \in \mathcal{A}$ heißt Nullmenge, falls $\mu(N) = 0$.

b) Eine Eigenschaft $E(\omega)$ die Punkten in Ω zukommen kann, gilt fast sicher (f.s.), falls es eine Nullmenge N gibt, so dass $E(\omega)$ für alle $\omega \in \Omega \setminus N$.

7.4 Satz
Seien $f, g : \Omega \to \bar{\mathbb{R}}$ messbar, $f \ge 0$.

a) $f(\omega) = 0$ f.s. $\Longleftrightarrow \int f \, d\mu = 0$

b) $\int f \, d\mu < \infty \Longrightarrow f(\omega) < \infty$ f.s.

c) $f(\omega) \le g(\omega)$ f.s. $\Longrightarrow \int f \, d\mu \le \int g \, d\mu$

Beweis:

a) "$\Longrightarrow$" Sei $N = \{f > 0\}$, Also $\mu(N) = 0$. Setze $N_n := \{0 < f \le n\} \subseteq N$. Dann gilt $0 \le f \cdot 1_{N_n} + n \cdot 1_{N \setminus N_n} \nearrow f$, so dass

$$\begin{aligned} 0 &\overset{\text{Lemma 7.2b)}}{\le} \int f \, d\mu \overset{\text{Lemma 7.2c)}}{=} \sup_n \int (f \, 1_{N_n} + n \, 1_{N \setminus N_n}) \, d\mu \overset{\text{Lemma 7.2b)}}{\le} \sup_n \int n \, 1_N \, d\mu \\ &\overset{\text{Lemma 7.2a)}}{\le} \sup_n n \cdot \mu(N) = 0 \end{aligned}$$

"$\Longleftarrow$" Für $n \in \mathbb{N}$ sei $A_n := \{f > \frac{1}{n}\}$. Dann ist

$$0 = \int f \, d\mu \overset{\text{Lemma 7.2b)}}{\ge} \int \frac{1}{n} \cdot 1_{A_n} \, d\mu \overset{\text{Lemma 7.2a)}}{=} \frac{1}{n} \mu(A_n) \, ,$$

also $\mu(A_n) = 0$. Da $A_n \nearrow \{f > 0\}$, ist $\mu\{f > 0\} = \sup_n \mu(A_n) = 0$ (Satz 2.7), also $f(\omega) = 0$ f.s.

b) Es ist für alle $n \in \mathbb{N}$

$$n \cdot \mu\{f \ge n\} \overset{\text{Lemma 7.2a)}}{=} \int n \cdot 1_{\{f \ge n\}} \, d\mu \overset{\text{Lemma 7.2b)}}{\le} \int f \, d\mu < \infty \, ,$$

also $\mu\{f = \infty\} \le \mu\{f \ge n\} \to 0$ mit $n \to \infty$, d.h. $f < \infty$ f.s.

c) Sei $N = \{f > g\}$, also $\mu(N) = 0$. Es folgt

$$\int f\,d\mu \overset{\text{Lemma 7.2d)}}{=} \int \underbrace{f \cdot 1_N}_{=0 \text{ f.s.}}\,d\mu + \int f \cdot 1_{N^c}\,d\mu \overset{\text{Teil a) und Lemma 7.2b)}}{\leq} \int g \cdot 1_{N^c}\,d\mu \overset{\text{Lemma 7.2b)}}{\leq} \int g\,d\mu$$

□

7.5 Definition (Integrierbarkeit) *Eine messbare Funktion $f : \Omega \to \bar{\mathbb{R}}$ heißt integrierbar (oder genauer μ-integrierbar), falls*

$$\int f^+\,d\mu < \infty \quad \textit{und} \quad \int f^-\,d\mu < \infty\,.$$

Äquivalent dazu ist die Forderung $\int |f|\,d\mu < \infty$, da $|f| = f^+ + f^-$. Sei $\mathcal{L}^1_\mu := \{f : \int |f|\,d\mu < \infty\}$ der Raum der μ-integrierbaren Funktionen.

7.6 Lemma *a) $|f| \leq |g|$ und $g \in \mathcal{L}^1_\mu \Longrightarrow f \in \mathcal{L}^1_\mu$*

b) $\mu(\Omega) < \infty$ und f beschränkt $\Longrightarrow f \in \mathcal{L}^1_\mu$

Beweis:

a) $\int |f|\,d\mu \leq \int |g|\,d\mu < \infty$ wegen Lemma 7.2b).

b) Sei $|f| \leq M$. Dann ist

$$\int |f|\,d\mu \overset{\text{Lemma 7.2b)}}{\leq} \int M\,d\mu \overset{\text{Lemma 7.2a)}}{=} M \cdot \mu(\Omega) < \infty$$

□

7.7 Satz (Monotonie Linearität und Positivität des Integrals)
Seien $f, g \in \mathcal{L}^1_\mu$.

a) $f \leq g$ f.s. $\Longrightarrow \int f\,d\mu \leq \int g\,d\mu$

b) Für alle $\alpha, \beta \in \mathbb{R}$ ist $\alpha f + \beta g \in \mathcal{L}^1_\mu$ und

$$\int (\alpha f + \beta g)\,d\mu = \alpha \int f\,d\mu + \beta \int g\,d\mu \tag{*}$$

c) $\left|\int f\,d\mu\right| \leq \int |f|\,d\mu$ (Insbesondere: $f \geq 0 \Longrightarrow \int f\,d\mu \geq 0$)

Beweis:

a) Für $f, g \geq 0$ benutze Satz 7.4c). Für beliebige f, g beachte, dass aus $f \leq g$ f.s. folgt: $f^+ \leq g^+$ f.s. und $g^- \leq f^-$ f.s., also

$$\int f^+\,d\mu \leq \int g^+\,d\mu \quad \text{und} \quad \int g^-\,d\mu \leq \int f^-\,d\mu\,,$$

und daher $\int f\,d\mu \leq \int g\,d\mu$.

b) Da $|\alpha f + \beta g| \leq |\alpha| \cdot |f| + |\beta| \cdot |g|$, folgt aus Lemma 7.2b) und d), dass $(\alpha f + \beta g) \in \mathcal{L}^1_\mu$. Um (*) zu zeigen, beweisen wir folgende drei Aussagen:

i) $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$
ii) $\int \alpha f\,d\mu = \alpha \int f\,d\mu$ für $\alpha \geq 0$

iii) $\int (-f)\, d\mu = -\int f\, d\mu$.

i) Es gilt

$$(f+g)^+ - (f+g)^- = f + g = f^+ - f^- + g^+ - g^- \ ,$$

also

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

Diese Funktionen sind ≥ 0, so dass nach Lemma 7.2d)

$$\int (f+g)^+\, d\mu + \int f^-\, d\mu + \int g^-\, d\mu = \int (f+g)^-\, d\mu + \int f^+\, d\mu + \int g^+\, d\mu \ ,$$

also

$$\begin{aligned}\int (f+g)\, d\mu &= \int (f+g)^+\, d\mu - \int (f+g)^-\, d\mu \\ &= \int f^+\, d\mu - \int f^-\, d\mu + \int g^+\, d\mu - \int g^-\, d\mu \\ &= \int f\, d\mu + \int g\, d\mu \ .\end{aligned}$$

ii) Sei $\alpha \geq 0$. Da $(\alpha f)^\pm = \alpha f^\pm$, ist

$$\int \alpha f\, d\mu = \int \alpha f^+\, d\mu - \int \alpha f^-\, d\mu = \alpha \int f^+\, d\mu - \alpha \int f^-\, d\mu = \alpha \int f\, d\mu$$

iii)

$$\int (-f)\, d\mu = \int (-f)^+\, d\mu - \int (-f)^-\, d\mu = \int f^-\, d\mu - \int f^+\, d\mu = -\int f\, d\mu$$

c) Da $f \leq |f|$ und $-f \leq |f|$, ist $\int f\, d\mu \leq \int |f|\, d\mu$ und $-\int f\, d\mu = \int (-f)\, d\mu \leq \int |f|\, d\mu$ nach Teil a) und b).

□

7.8 Bemerkung Das Integral einer $\mathbb{C}$-wertigen Funktion wird folgendermaßen definiert: $f : \Omega \to \mathbb{C}$ ist messbar, falls $\Re f$ und $\Im f$ messbar sind. f ist integrierbar, falls $\Re f$ und $\Im f$ integrierbar sind. Für ein solches integrierbares f sei

$$\int f\, d\mu := \int \Re f\, d\mu + i \int \Im f\, d\mu \ .$$

Dann gelten Satz 7.7b) und c) analog, wobei auch die Koeffizienten α und β komplex sein können.

7.9 Definition *Für* $1 \leq p < \infty$ *sei*

$$\mathcal{L}^p_\mu := \left\{ f : \Omega \to \bar{\mathbb{R}} \middle| f \textit{ messbar, } \int |f|^p\, d\mu < \infty \right\}$$

7.10 Bemerkung Ist $\mu(\Omega) < \infty$, so ist $\mathcal{L}^p_\mu \subseteq \mathcal{L}^{p'}_\mu$ falls $p' \leq p$. Zum Beweis vergleiche Bemerkung 10.2.

7.11 Bemerkung Folgende Beobachtung ist oft nützlich: Ist $0 \leq f : \Omega \to \mathbb{R}$ messbar, so ist

$$\sum_{k=1}^\infty \mu\{f \geq k\} \leq \int f\, d\mu \leq \sum_{k=0}^\infty \mu\{f > k\} \ ,$$

denn für $n < f(\omega) < n+1$ gilt

$$\sum_{k=1}^\infty 1_{\{f \geq k\}}(\omega) = n \leq f(\omega) \leq n + 1 = \sum_{k=0}^\infty 1_{\{f>k\}}(\omega) \ ,$$

und falls $f(\omega) = n + 1$ kann man das erste n durch $n + 1$ ersetzen.

7.12 Definition *Ist $f : \Omega \to \mathbb{R}$ integrierbar oder ≥ 0 und ist $A \in \mathcal{A}$, so ist*

$$\int_A f \, d\mu := \int 1_A \cdot f \, d\mu \, .$$

7.13 Bemerkung (Riemann- und Lebesgue-Integral) Ist $f : [a, b] \to \mathbb{R}$ Riemann-integrierbar, so ist f auch integrierbar bzgl. der Einschränkung des Lebesgue-Maßes auf $[a, b]$, und die beiden Integrale stimmen überein, d.h.

$$\text{R-}\int_a^b f(x) \, dx = \int_{[a,b]} f \, d\lambda \, .$$

Der Beweis ist nicht schwierig. Da in dieser Vorlesung das Riemann-Integral aber nur eine untergeordnete Rolle spielt, warten wir einfach, bis er als direktes Korollar aus Satz 8.8 (Konvergenzsatz von Lebesgue) folgt.

Nun wenden wir den Integralbegriff auf Zufallsvariablen an. Sei daher $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

7.14 Definition (Erwartungswert, Varianz, Momente) *Sei X eine Zufallsvariable.*

a) $E[X] := \int X \, dP$ heißt der <u>Erwartungswert</u> von X. Wir schreiben auch EX.

b) Ist $X \in \mathcal{L}_P^r$ oder $X^r \geq 0$, so ist das <u>r-te Moment</u> $E[X^r]$ von X wohldefiniert $(r = 1, 2, 3, \dots)$.

c) Sei $X \in \mathcal{L}_P^1$. $V(X) := E[(X - EX)^2]$ heißt die <u>Varianz</u> von X, $\sigma(X) := \sqrt{V(X)}$ die <u>Streuung</u>.

7.15 Lemma *Für $X \in \mathcal{L}_P^1$ gilt:*

a) $V(X) = E[X^2] - (EX)^2$,

b) $V(X) = 0 \iff X = EX$ f.s. $\iff X = \text{const}$ f.s.,

c) $V(X) < \infty \iff X \in \mathcal{L}_P^2$

d) Sei $X \in \mathcal{L}_P^2$. Dann ist

$$V(X) \leq E[(X - a)^2] \quad \text{für alle } a \in \mathbb{R}$$

mit Gleichheit genau dann, wenn $a = EX$.

Beweis:

a)

$$V(X) + \underbrace{(2E|X|)^2}_{<\infty} = E[(X - EX)^2 + 2|X| \cdot E|X|] = E[X^2 \underbrace{-2X \cdot EX + 2|X| \cdot E|X|}_{\geq 0} + (EX)^2]$$
$$= E[X^2] - 2\,EX\,EX + 2E|X|\,E|X| + (EX)^2 = E[X^2] - (EX)^2 + 2(E|X|)^2.$$

b) Folgt aus Satz 7.4.

c) Folgt aus Teil a).

d)
$$f(a) := E[(X - a)^2] = E[X^2] - 2a\,EX + a^2 = (a - EX)^2 + V(X) \, .$$

f hat ein eindeutiges Minimum bei $a = EX$ und $f(EX) = V(X)$.

□

7.16 Satz (Markov- und Chebyshev-Ungleichung)
Sei X eine Zufallsvariable, $\epsilon > 0$. Dann ist $P(|X| \geq \epsilon) \leq \epsilon^{-1} E|X|$ (Markov-Ungleichung) und daher auch (Chebyshev-Ungleichung)

$$P(|X| \geq \epsilon) \leq \epsilon^{-2} E[X^2]$$

Beweis:

$$P(|X| \geq \epsilon) \leq \int_{\{|X| \geq \epsilon\}} \epsilon^{-1}|X|\, d\mu \leq \int \epsilon^{-1}|X|\, d\mu = \epsilon^{-1} E|X|$$

□

7.17 Definition *Für $X, Y \in \mathcal{L}_P^2$ heißt $Cov(X,Y) := E[(X - EX)(Y - EY)]$ die Kovarianz von X und Y. X und Y heißen unkorreliert, falls $Cov(X,Y) = 0$.*

7.18 Satz
Seien $X_i, Y_j \in \mathcal{L}_P^2$ $(i = 1, \ldots, m;\ j = 1, \ldots, n)$ und $a_i, b_j, c, d \in \mathbb{R}$. Dann ist

$$Cov\left(\sum_{i=1}^m a_i X_i + c, \sum_{j=1}^n b_j Y_j + d\right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j\, Cov(X_i, Y_j)$$

Beweis:

$$\begin{aligned} Cov\left(\sum_{i=1}^m a_i X_i + c, \sum_{j=1}^n b_j Y_j + d\right) &= E\left[\sum_{i=1}^m a_i(X_i - EX_i) \cdot \sum_{j=1}^n b_j(Y_j - EY_j)\right] \\ &= \sum_{i=1}^m \sum_{j=1}^n E\left[a_i b_j (X_i - EX_i)(Y_j - EY_j)\right] \\ &= \sum_{i=1}^m \sum_{j=1}^n a_i b_j\, Cov(X_i, Y_j) \end{aligned}$$

□

7.19 Korollar *Seien $X, X_i, Y \in \mathcal{L}_P^2$. Dann gilt*

a)

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^n Cov(X_i, X_j)$$

Insbesondere: Sind die X_i paarweise unkorreliert, so ist $V(\sum_{i=1}^n X_i) = \sum_{i=1}^n V(X_i)$.

b)

$$Cov(X - EX, Y - EY) = Cov(X,Y) \quad \text{und daher} \quad Cov(X,Y) = E[XY] - EX \cdot EY.$$

Insbesondere: Sind X, Y unkorreliert, so ist $E[XY] = EX \cdot EY$.

7.20 Satz
Seien $X_1, \dots, X_n \in \mathcal{L}^1_P$ unabhängig. Dann ist $\prod_{i=1}^n X_i \in \mathcal{L}^1_P$ und

$$E\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n EX_i \qquad (*)$$

Beweis: Ist $X_i = 1_{A_i}$ $(i = 1, \dots, n)$, so gilt (*) wegen der Unabhängigkeit. Da beide Seiten von (*) multilinear in $X_1, \dots, X_n$ sind, gilt (*) auch für endliche Linearkombinationen von Indikatorfunktionen, und aus Satz 4.11 zusammen mit Lemma 7.2c) folgt, dass (*) für nichtnegative $X_i \in \mathcal{L}^1_P$ gilt. Aufgrund der Zerlegung $X_i = X_i^+ - X_i^-$ folgt (*) dann durch nochmalige Ausnutzung der Multilinearität auch für beliebige $X_i \in \mathcal{L}^1_P$. (Dieses Beweisschema wird bei [Gänssler-Stute] auch "algebraische Induktion" genannt.) Einen davon unabhängigen Beweis erhalten wir später als Korollar zum Satz von Fubini. □

7.21 Korollar *Sind $X_i \in \mathcal{L}^2_\mu$ unabhängig $(i \in I)$, so sind sie paarweise unkorreliert.*

7.22 Satz (Integration mit transformierten Maßen)
Seien $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ messbare Räume, $T : \Omega \to \Omega'$ messbar, und sei μ ein Maß auf $(\Omega, \mathcal{A})$. Ist $f : \Omega' \to \mathbb{R}$ messbar und ist f nichtnegativ oder $f \circ T \in \mathcal{L}^1_\mu$, so gilt

$$\int_\Omega f \circ T \, d\mu = \int_{\Omega'} f \, d(\mu \circ T^{-1}) \,. \qquad (**)$$

(Ist $\mu = P$ ein Wahrscheinlichkeitsmaß und $T = X$ eine Ω'-wertige Zufallsvarable, so schreibt sich das als $E_P[f(X)] = \int f \, dP_X$.)
Insbesondere gilt im Fall $\Omega = \Omega'$: Ist $\mu \circ T^{-1} = \mu$, so ist $\int f \circ T \, d\mu = \int f \, d\mu$.

Beweis: (**) gilt nach Definition von $\mu \circ T^{-1}$ falls $f = 1_A$. Wie oben folgt die Behauptung durch "algebraische Induktion". □

7.23 Bemerkung *Achtung:* Satz 7.22 stellt keine Integraltransformationsformel bereit, wie man sie aus der Analysis kennt. Dazu müsste auf der rechten Seite das transportierte Maß $\mu \circ T^{-1}$ auf Ω' durch ein dort vorgegebenes Maß μ' mal einer "Jacobischen Determinante" ersetzt werden. Das wäre ein viel tiefer liegendes Ergebnis.

Kapitel 8

Konvergenzsätze

In diesem Kapitel sei $(\Omega, \mathcal{A}, \mu)$ wieder ein Maßraum. Im Zentrum unserer Überlegungen steht die Frage, unter welchen Voraussetzungen Grenzwertbildung von Folgen messbarer Funktionen und Integration in der Reihenfolge vertauscht werden können. Dabei werden verschiedene Konvergenzbegriffe untersucht, und mit dem Borel-Cantelli Lemma und dem schwachen Gesetz der großen Zahl zwei einfache, aber für die Wahrscheinlichkeitstheorie wichtige Aussagen bewiesen.

8.1 Satz (Monotone Konvergenz)
Seien $f, f_n : \Omega \to \bar{\mathbb{R}}$ messbar.

a) *Aus $0 \leq f_n \nearrow f$ f.s. folgt $\int f_n \, d\mu \nearrow \int f \, d\mu$.*

b) *Sind die $f_n \geq 0$, so gilt $\int \sum_{n=1}^{\infty} f_n \, d\mu = \sum_{n=1}^{\infty} \int f_n \, d\mu$.*

c) *Sind die $f_n \in \mathcal{L}^1_\mu$, gilt $f_n \nearrow f$ f.s. und ist $\sup_n \int f_n \, d\mu < \infty$, so ist $f \in \mathcal{L}^1_\mu$ und $\int f_n \, d\mu \nearrow \int f \, d\mu$. (Diese Aussage heißt auch* Satz von Beppo Levi.*)*

Beweis:

a) Das ist gerade Lemma 7.2 c).

b) Das folgt aus a), da $0 \leq \sum_{n=1}^{m} \int f_n \, d\mu = \int \sum_{n=1}^{m} f_n \, d\mu \nearrow \int \sum_{n=1}^{\infty} f_n \, d\mu$.

c) $f_1 \leq f_n \Longrightarrow 0 \leq f_n - f_1 \nearrow f - f_1$, und aus Teil a) folgt

$$\int (f - f_1) \, d\mu = \sup_n \int (f_n - f_1) \, d\mu = \sup_n \int f_n \, d\mu - \int f_1 \, d\mu < \infty \qquad (*)$$

da das Supremum endlich ist und $f_1 \in \mathcal{L}^1_\mu$. Daher $f - f_1 \in \mathcal{L}^1_\mu$, und da $f_1 \in \mathcal{L}^1_\mu$, ist auch $f \in \mathcal{L}^1_\mu$. Es folgt $\int (f - f_1) \, d\mu = \int f \, d\mu - \int f_1 \, d\mu$ und schließlich mit (*) auch $\sup_n \int f_n \, d\mu = \int f \, d\mu$.

$\square$

8.2 Bemerkung In Satz 8.1c) sowie in vielen anderen Konvergenzausasgen dieses Kapitels kann man auf die Voraussetzung, dass f messbar ist, verzichten, falls $(\Omega, \mathcal{A}, \mu)$ vollständig ist, denn dann folgt die Messbarkeit von f aus der fast sicheren Konvergenz von f_n gegen f.

8.3 Korollar (Maße mit Dichten) *Ist $0 \leq f$ messbar, so wird durch $\nu(A) := \int_A f \, d\mu$ ein Maß auf $(\Omega, \mathcal{A})$ definiert. ν ist endlich genau dann, wenn $f \in \mathcal{L}^1_\mu$, und ν ist ein Wahrscheinlichkeitsmaß genau dann, wenn $\int f \, d\mu = 1$. Schreibweise: $\nu = f \cdot \mu$ oder $\nu = f\mu$.*

Beweis: Wir zeigen die σ-Additivität, der Rest ist trivial: Seien $A_1, A_2, \dots \in \mathcal{A}$ paarweise disjunkt. Mit $f_n := 1_{A_n} f$ folgt aus Satz 8.1b)

$$\nu\left(\biguplus_{n=1}^{\infty} A_n\right) = \int \sum_{n=1}^{\infty} f_n \, d\mu = \sum_{n=1}^{\infty} \nu(A_n)$$

□

Wir führen folgende Bezeichnung ein:

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega : \omega \in A_k \text{ für unendlich viele } k\}$$

8.4 Korollar (Borel-Cantelli Lemma) *Seien $A_n \in \mathcal{A}$, $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. Dann ist*

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=n}^{\infty} A_k\right) = 0 \, ,$$

d.h. $\{n \in \mathbb{N} : \omega \in A_n\}$ ist für μ-fast alle ω endlich.

Beweis: Aus Satz 8.1b) folgt für $f := \sum_{n=1}^{\infty} 1_{A_n}$, dass $\int f \, d\mu = \sum_{n=1}^{\infty} \mu(A_n) < \infty$, so dass $f < \infty$ f.s. nach Satz 7.4b). Das ist gleichbedeutend mit $\operatorname{card}\{n \in \mathbb{N} : \omega \in A_n\} < \infty$ f.s. □

8.5 Bemerkung Sind die Ereignise A_n unabhängig, so gilt auch folgende "Umkehrung": Ist $\sum_{n=1}^{\infty} \mu(A_n) = \infty$, so ist $\mu\left(\limsup_{n\to\infty} A_n\right) = 1$, d.h. $\{n \in \mathbb{N} : \omega \in A_n\}$ ist für μ-fast alle ω unendlich.

8.6 Beispiel Sei $(X_n)_n$ eine Folge integrierbarer, identisch verteilter Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Dann ist $\lim_{n\to\infty} n^{-1} X_n = 0$ f.s., denn

$$\left\{n^{-1} X_n \not\longrightarrow 0\right\} = \bigcup_{j=1}^{\infty} \limsup_{n\to\infty} \left\{n^{-1}|X_n| > \frac{1}{j}\right\} = \bigcup_{j=1}^{\infty} \limsup_{n\to\infty} \{j|X_n| > n\}$$

und für jedes j ist $P(\limsup_{n\to\infty} \{j|X_n| > n\}) = 0$, da nach Bemerkung 7.11

$$\sum_{n=1}^{\infty} P(j|X_n| > n) = \sum_{n=1}^{\infty} P(j|X_1| > n) \le j \int |X_1| \, dP < \infty \, .$$

8.7 Satz (Lemma von Fatou)
Seien $0 \le f_n : \Omega \to \bar{\mathbb{R}}$ messbar. Dann ist

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

Beweis: Sei $g_k := \inf_{n \ge k} f_n$ (punktweise!), so dass $0 \le g_k \nearrow \liminf f_n$. Wegen der Monotonie des Integrals ist $\int g_k \, d\mu \le \int f_n \, d\mu$ für alle $n \ge k$, also $\int g_k \, d\mu \le \inf_{n \ge k} \int f_n \, d\mu$. Aus dem Satz von der monotonen Konvergenz folgt nun

$$\int \liminf_{n\to\infty} f_n \, d\mu = \sup_k \int g_k \, d\mu \le \sup_k \inf_{n \ge k} \int f_n \, d\mu = \liminf_{n\to\infty} \int f_n \, d\mu$$

□

8.8 Satz (Majorisierte Konvergenz, auch Konvergenzsatz von Lebesgue)
Seien $f, f_n, g : \Omega \to \bar{\mathbb{R}}$ messbar, $|f_n| \leq g$ f.s. und $g \in \mathcal{L}^1_\mu$. Existiert $f = \lim_{n\to\infty} f_n$ f.s., so ist $f \in \mathcal{L}^1_\mu$ und

$$\lim_{n\to\infty} \int |f - f_n|\, d\mu = 0, \quad \textit{insbesondere} \quad \int f\, d\mu = \lim_{n\to\infty} \int f_n\, d\mu\ .$$

Beweis: Da $|f_n| \leq g$ f.s., ist auch $|f| = \lim_{n\to\infty} |f_n| \leq g$ f.s., also $f \in \mathcal{L}^1_\mu$ wegen der Monotonie des Integrals. Sei $h_n := |f - f_n|$. Dann gilt

$$0 \leq h_n \leq |f| + |f_n| \leq 2g\ , \quad \text{also} \quad 0 \leq 2g - h_n \leq 2g \quad \text{für alle } n$$

und $h_n \to 0$ f.s., so dass aus dem Lemma von Fatou folgt

$$\begin{aligned}\int 2g\, d\mu &= \int \liminf_{n\to\infty} (2g - h_n)\, d\mu \leq \liminf_{n\to\infty} \int (2g - h_n)\, d\mu \\ &= \int 2g\, d\mu - \limsup_{n\to\infty} \int h_n\, d\mu \leq \int 2g\, d\mu\ ,\end{aligned}$$

also $\limsup_{n\to\infty} \int h_n\, d\mu = 0$. Daher

$$\left| \int f\, d\mu - \int f_n\, d\mu \right| \overset{\text{Satz 7.7}}{\leq} \int |f - f_n|\, d\mu = \int h_n\, d\mu \to 0\ .$$

□

8.9 Korollar *Ist $\mu(\Omega) < \infty$ und ist $(f_n)_n$ eine Folge gleichmäßig beschränkter messbarer Funktionen, die fast sicher gegen eine messbare Funktion f konvergiert, so ist $f \in \mathcal{L}^1_\mu$ und $\lim_{n\to\infty} \int |f - f_n|\, d\mu = 0$.*

Beweis: Wähle $g = M := \sup_n \sup_\omega f_n(\omega)$ im Satz von der majorisierten Konvergenz. □

8.10 Korollar *Seien $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subseteq \Omega$ messbar, $A = \bigcup_{n=1}^\infty A_n$ und $f \in \mathcal{L}^1_\mu$. Dann ist*

$$\lim_{n\to\infty} \int_{A_n} f\, d\mu = \int_A f\, d\mu\ .$$

Beweis: Wähle $f_n := f 1_{A_n}$ und $g := |f|$, und berücksichtige, dass $\int_{A_n} f\, d\mu = \int f 1_{A_n}\, d\mu$. □

8.11 Korollar (Riemann- und Lebesgue-Integral, siehe auch Bem. 7.13) *Ist $f : [a,b] \to \mathbb{R}$ Riemann-integrierbar, so ist f auch integrierbar bzgl. der Einschränkung des Lebesgue-Maßes auf $[a,b]$, und die beiden Integrale stimmen überein, d.h.*

$$\textit{R-}\int_a^b f(x)\, dx = \int_{[a,b]} f\, d\lambda\ .$$

Beweis: Zu Riemann-integrierbarem f gibt es Treppenfunktionen (Unter- und Oberfunktionen) $u_1 \leq u_2 \leq \cdots \leq f \leq \cdots \leq v_2 \leq v_1$ auf $[a,b]$, so dass

$$\begin{aligned}\int_{[a,b]} u_n\, d\lambda &= \int_a^b u_n(x)\, dx \nearrow \text{R-}\int_a^b f(x)\, dx \\ \int_{[a,b]} v_n\, d\lambda &= \int_a^b v_n(x)\, dx \searrow \text{R-}\int_a^b f(x)\, dx\ .\end{aligned}$$

Sei $u_\infty = \sup_n u_n$, $v_\infty = \inf_n v_n$. Dann ist $u_\infty \leq f \leq v_\infty$ und da $|u_n|, |v_n| \leq \max\{|u_1|, |v_1|\} < \infty$, folgt aus Korollar 8.9, dass $\int_{[a,b]} u_\infty \, d\lambda = \int_{[a,b]} v_\infty \, d\lambda =$ R-$\int_a^b f(x)\, dx$. Da $u_\infty \leq v_\infty$, folgt $u_\infty = v_\infty$ λ-f.s., also auch $f = u_\infty$ λ-f.s. f stimmt also λ-f.s. mit einer beschränkten messbaren Funktion überein, und ist damit selbst Lebesgue-integrierbar (Vollstädnigkeit des Lebesgue-Maßes!). Insbesondere ist dann auch $\int_{[a,b]} f \, d\lambda = \int_{[a,b]} u_\infty \, d\lambda =$ R-$\int_a^b f(x)\, dx$. $\square$

8.12 Satz (Differentiation parameterabhängiger Integrale)
Sei $G \subseteq \mathbb{R}$ offen, $f : G \times \Omega \to \mathbb{R}$ sei so, dass $\omega \mapsto f(t,\omega)$ für jedes $t \in G$ μ-integrierbar und $t \mapsto f(t,\omega)$ für jedes $\omega \in \Omega$ differenzierbar ist. Gibt es ein $\Psi \in \mathcal{L}^1_\mu$ derart, dass $|\frac{d}{dt} f(t,\omega)| \leq \Psi(\omega)$ für alle $(t,\omega) \in G \times \Omega$, so existiert $\frac{d}{dt} \int f(t,\omega)\, d\mu(\omega)$ für jedes $t \in G$ und es ist

$$\frac{d}{dt} \int f(t,\omega)\, d\mu(\omega) = \int \frac{d}{dt} f(t,\omega)\, d\mu(\omega) \; .$$

(Entsprechendes gilt auch für $\mathbb{C}$-wertige f.)

Beweis: Sei $(h_n)_n$ eine Nullfolge reeller Zahlen, $h_n \neq 0$ für alle n. Bezeichne

$$g(t,\omega) := \frac{d}{dt} f(t,\omega), \quad g_n(t,\omega) := \frac{1}{h_n} \left(f(t+h_n,\omega) - f(t,\omega) \right) \; .$$

Zu zeigen ist

$$\lim_{n\to\infty} \frac{1}{h_n} \left(\int f(t+h_n,\omega)\, d\mu(\omega) - \int f(t,\omega)\, d\mu(\omega) \right) = \lim_{n\to\infty} \int g_n(t,\omega)\, d\mu(\omega) \overset{!}{=} \int g(t,\omega)\, d\mu(\omega)$$

für alle t. Nach Voraussetzung ist $\lim_{n\to\infty} g_n(t,.) = g(t,.)$ für alle t und $|g_n| \leq \Psi$ (Mittelwertsatz der Differentialrechnung). Daher folgt die Behauptung aus Satz 8.8. $\square$

Im Rest dieses Kapitels beschränken wir uns auf den Fall endlicher Maße (im Hinblick auf unsere Hauptanwendung, die Wahrscheinlichkeitstheorie). Ziel ist es, die Voraussetzungen von Satz 8.8 und Korollar 8.9 abzuschwächen. Analoge Ergebnisse für allgemeine Maße findet man bei [Bauer-MT, §21].

8.13 Definition *Sei μ ein endliches Maß auf $(\Omega, \mathcal{A})$. Eine Familie $\mathcal{F}$ messbarer $\bar{\mathbb{R}}$-wertiger Funktionen auf Ω heißt gleichgradig integrierbar, falls es zu jedem $\epsilon > 0$ ein $a > 0$ gibt, für das*

$$\sup_{f\in\mathcal{F}} \int (|f| - a)^+ \, d\mu < \epsilon \; .$$

8.14 Bemerkung Sei $\mathcal{F} = \{f\}$ mit $f \in \mathcal{L}^1_\mu$. Dann ist $\mathcal{F}$ gleichgradig integrierbar, denn $0 \leq (|f| - n)^+ \leq |f|$ und $\lim_{n\to\infty} (|f| - n)^+ = 0$ auf $\{|f| < \infty\}$, d.h. f.s., so dass aus dem Satz von der majorisierten Konvergenz folgt $\lim_{n\to\infty} \int (|f| - n)^+ \, d\mu = 0$.

8.15 Satz
Sei μ ein endliches Maß auf $(\Omega, \mathcal{A})$ und $\mathcal{F}$ eine Familie messbarer $\bar{\mathbb{R}}$-wertiger Funktionen auf Ω. Äquivalent sind die folgenden Aussagen:

i) $\mathcal{F}$ ist gleichgradig integrierbar.

ii) Für alle $\epsilon > 0$ gibt es ein $a > 0$ derart, dass

$$\sup_{f \in \mathcal{F}} \int_{\{|f|>a\}} |f| \, d\mu < \epsilon \, .$$

iii) $\sup_{f \in \mathcal{F}} \int |f| \, d\mu =: C < \infty$ und

$$\forall \epsilon > 0 \; \exists \delta > 0 \; \forall A \in \mathcal{A}: \; \mu(A) \leq \delta \implies \sup_{f \in \mathcal{F}} \int_A |f| \, d\mu < \epsilon \, .$$

Beweis: Schema: iii) $\Longrightarrow$ ii) $\Longrightarrow$ i) $\Longrightarrow$ iii).
iii) $\Longrightarrow$ ii) Zu $\epsilon > 0$ wähle $\delta > 0$ wie in iii). Dann ist für jedes $a > \frac{C}{\delta}$ und für jedes $f \in \mathcal{F}$ wegen der Markov-Ungleichung

$$\mu\{|f| > a\} \leq \frac{1}{a} \int |f| \, d\mu \leq \frac{C}{a} < \delta \, ,$$

also $\sup_{f \in \mathcal{F}} \int_{\{|f|>a\}} |f| \, d\mu < \epsilon$.
ii) $\Longrightarrow$ i) Beachte dass $(|f| - a)^+ \leq |f| \cdot 1_{\{|f|>a\}}$.
i) $\Longrightarrow$ iii) Sei $\epsilon > 0$. Wähle $a = a(\frac{\epsilon}{2})$ wie in i), $\delta := \frac{\epsilon}{2a}$. Dann ist für $f \in \mathcal{F}$ und $A \in \mathcal{A}$

$$\int_A |f| \, d\mu \leq \int_A (|f| - a)^+ \, d\mu + \int_A a \, d\mu \leq \frac{\epsilon}{2} + a \cdot \mu(A) \, .$$

Für $A = \Omega$ folgt daraus

$$\sup_{f \in \mathcal{F}} \int |f| \, d\mu \leq \frac{\epsilon}{2} + a \cdot \mu(\Omega) < \infty \, ,$$

und für $\mu(A) < \delta$ erhält man

$$\sup_{f \in \mathcal{F}} \int_A |f| \, d\mu \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \, .$$

□

8.16 Lemma *Sei μ ein endliches Maß auf $(\Omega, \mathcal{A})$.*

a) $\mathcal{F}$ gleichgradig integrierbar $\Longleftrightarrow$ $\{|f| : f \in \mathcal{F}\}$ gleichgradig integrierbar

b) $\mathcal{F}$ gleichgradig integrierbar, $S > 0 \Longrightarrow \left\{\sum_{i=1}^N \alpha_i f_i : N \in \mathbb{N}, f_i \in \mathcal{F}, \sum_{i=1}^N |\alpha_i| \leq S\right\}$ gleichgradig integrierbar

c) $\mathcal{F}_1, \mathcal{F}_2$ gleichgradig integrierbar $\Longrightarrow$ $\mathcal{F}_1 \cup \mathcal{F}_2$ gleichgradig integrierbar

d) $\mathcal{F} \subseteq \mathcal{L}^1_\mu$ endlich $\Longrightarrow$ $\mathcal{F}$ gleichgradig integrierbar

Beweis:

a) trivial.

b) Sei $\epsilon > 0$ und seien C und $\delta = \delta(\frac{\epsilon}{S})$ wie in Satz 8.15iii) zu $\mathcal{F}$ gewählt. Dann ist für $A \in \mathcal{A}$ mit $\mu(A) < \delta$, für $f_i \in \mathcal{F}$ und α_i mit $\sum_{i=1}^N |\alpha_i| \leq S$

$$\int |\sum_i \alpha_i f_i| \, d\mu \leq \sum_i |\alpha_i| \cdot \int |f_i| \, d\mu \leq C \cdot S < \infty \, ,$$

$$\int_A |\sum_i \alpha_i f_i| \, d\mu \leq \sum_i |\alpha_i| \cdot \int_A |f_i| \, d\mu < \sum_i |\alpha_i| \, \frac{\epsilon}{S} \leq \epsilon \, .$$

c) Zu $\epsilon > 0$ seien δ_1 (für $\mathcal{F}_1$) und δ_2 (für $\mathcal{F}_2$) wie in Satz 8.15iii) gewählt. Benutze nun $\delta := \min\{\delta_1, \delta_2\} > 0$ für $\mathcal{F}_1 \cup \mathcal{F}_2$.

d) folgt aus Bemerkung 8.14 und Aussage c).

□

Nun beweisen wir die angekündigte Verschärfung von Korollar 8.9.

8.17 Satz (Konvergenzsatz für gleichgradig integrierbare Funktionen)
Sei μ ein endliches Maß auf $(\Omega, \mathcal{A})$, $(f_n)_{n\geq 1}$ eine gleichgradig integrierbare Folge messbarer $\bar{\mathbb{R}}$-wertiger Funktionen auf Ω. Ist $f : \Omega \to \bar{\mathbb{R}}$ messbar und $f = \lim_{n\to\infty} f_n$ f.s., so ist $f \in \mathcal{L}^1_\mu$ und

$$\lim_{n\to\infty} \left| \int f_n \, d\mu - \int f \, d\mu \right| = \lim_{n\to\infty} \int |f_n - f| \, d\mu = 0 \, .$$

Beweis: Es ist $f \in \mathcal{L}^1_\mu$, da

$$\int |f| \, d\mu = \int \liminf_{n\to\infty} |f_n| \, d\mu \overset{\text{Lemma von Fatou}}{\leq} \liminf_{n\to\infty} \int |f_n| \, d\mu \leq \sup_n \int |f_n| \, d\mu \overset{\text{Satz 8.15iii)}}{<} \infty \, .$$

Setze $h_n = |f - f_n|$. Wegen Lemma 8.16 ist $(h_n)_n$ gleichgradig integrierbar, so dass zu $\epsilon > 0$ ein $a > 0$ existiert mit

$$\sup_n \int_{\{h_n > a\}} h_n \, d\mu < \epsilon \, .$$

Daher:

$$\int |f - f_n| \, d\mu = \int h_n \, d\mu = \int_{\{h_n > a\}} h_n \, d\mu + \int h_n \cdot 1_{\{h_n \leq a\}} \, d\mu < \epsilon + \int h_n \cdot 1_{\{h_n \leq a\}} \, d\mu \, .$$

Da $|h_n \cdot 1_{\{h_n \leq a\}}| \leq a$ und $|h_n \cdot 1_{\{h_n \leq a\}}| \leq h_n = |f - f_n| \to 0$ f.s., folgt aus Korollar 8.9

$$\lim_{n\to\infty} \int h_n \cdot 1_{\{h_n \leq a\}} \, d\mu = 0 \, ,$$

also

$$\lim_{n\to\infty} \int |f - f_n| \, d\mu \leq \epsilon \quad \text{für beliebiges } \epsilon > 0 \, .$$

□

Wir diskutieren nun drei Konvergenzbegriffe für messbare Funktionen im Fall endlicher Maße. Für den allgemeinen Fall siehe [Bauer-MT, §20].

Seien f_n, f messbare $\mathbb{R}$-wertige Funktionen auf $(\Omega, \mathcal{A})$, μ ein endliches Maß auf $\mathcal{A}$.

▷ **Fast sichere Konvergenz** $f = \lim_{n\to\infty} f_n$ μ-f.s. genau dann, wenn

$$\begin{aligned} \forall \alpha > 0 : \; \mu\left(\{\limsup_{k\to\infty} |f - f_k| > \alpha\}\right) &= \mu\left(\limsup_{k\to\infty} \{|f - f_k| > \alpha\}\right) \\ &= \lim_{n\to\infty} \mu\left(\bigcup_{k\geq n} \{|f - f_k| > \alpha\}\right) = 0 \end{aligned}$$

▷ **Stochastische Konvergenz** $f_n \underset{\mu}{\to} f \ (n \to \infty)$ genau dann, wenn

$$\forall \alpha > 0: \lim_{n\to\infty} \mu\left(\{|f - f_n| > \alpha\}\right) = 0$$

▷ L^1_μ**-Konvergenz** $\lim_{n\to\infty} \int |f - f_n|\, d\mu = 0$

8.18 Lemma *Seien f_n, f messbare $\mathbb{R}$-wertige Funktionen auf $(\Omega, \mathcal{A})$, μ ein endliches Maß auf $\mathcal{A}$.*

a) $f_n \to f$ *μ-f.s.* $\Longrightarrow f_n \underset{\mu}{\to} f$

b) $\int |f_n - f|\, d\mu \to 0 \Longrightarrow f_n \underset{\mu}{\to} f$

c) Für jeden der drei Konvergenzbegriffe gilt: Geht $f_n \to f$, so folgt

$$f_n \to g \quad \Longleftrightarrow \quad f = g \ \ \mu\text{-f.s.}$$

Beweis:

a) Folgt direkt aus obigen Charakterisierungen der Konvergenzbegriffe.

b) Folgt aus der Markov-Ungleichung $\mu\{|f_n - f| > \alpha\} \le \alpha^{-1} \cdot \int |f_n - f|\, d\mu$.

c) "$\Longrightarrow$" Wegen Teil a) und b) ist das nur für die stochastische Konvergenz zu zeigen: Für $\alpha > 0$ und $n \in \mathbb{N}$ ist

$$\mu\{|f - g| > \alpha\} \le \mu\{|f - f_n| > \frac{\alpha}{2}\} + \mu\{|f_n - g| > \frac{\alpha}{2}\} \overset{n\to\infty}{\longrightarrow} 0\,,$$

also $\mu\{|f - g| > \alpha\} = 0$ und daher

$$\mu\{f \neq g\} = \mu\left(\bigcup_{i=1}^{\infty} \{|f - g| > \frac{1}{i}\}\right) \le \sum_{i=1}^{\infty} \mu\{|f - g| > \frac{1}{i}\} = 0$$

"$\Longleftarrow$" Ist $\mu\{f \neq g\} = 0$, so werden die definierenden Ausdrücke für die drei Konvergenzbegriffe beim Übergang von f zu g oder umgekehrt nur auf μ-Nullmengen geändert, d.h. ihre Werte ändern sich nicht.

□

8.19 Satz
Seien f_n, f wie vorher. Gilt $f_n \underset{\mu}{\to} f$, so gibt es eine Teilfolge $(f_{n_k})_k$ von $(f_n)_n$, die μ-f.s. gegen f konvergiert.

Beweis: Für alle $k > 0$ gibt es ein $n_k \in \mathbb{N}$ mit $\mu\{|f_{n_k} - f| > \frac{1}{k}\} \le 2^{-k}$. Die n_k können so gewählt werden, dass $n_1 < n_2 < n_3 < \ldots$. Aus dem Borel-Cantelli Lemma 8.4 folgt $\operatorname{card}\{k \in \mathbb{N} : |f_{n_k}(\omega) - f(\omega)| > \frac{1}{k}\} < \infty$ für μ-fast alle ω und daher $\lim_{k\to\infty} f_{n_k} = f$ μ-f.s. □

8.20 Satz
Seien f_n, f messbare $\mathbb{R}$-wertige Funktionen auf $(\Omega, \mathcal{A})$, μ ein endliches Maß auf $\mathcal{A}$. Folgende Aussagen sind äquivalent:

i) $\lim_{n\to\infty} \int |f_n - f|\, d\mu = 0$ *und* $f \in \mathcal{L}^1_\mu$

ii) $f_n \underset{\mu}{\to} f$ *und* $(f_n)_n$ *ist gleichgradig integrierbar.*

Beweis: "i) $\Rightarrow$ ii)": $f_n \underset{\mu}{\to} f$ folgt aus Lemma 8.18b). Sei nun $\epsilon > 0$. Wähle $a_1 > 0$ so, dass $\int(|f| - a_1)^+ \, d\mu < \frac{\epsilon}{3}$ (vergl. Bemerkung 8.14), und $n_0 = n_0(\epsilon) \in \mathbb{N}$ derart, dass

$$\int ||f_n| - |f|| \, d\mu \le \int |f_n - f| \, d\mu < \frac{\epsilon}{3} \quad \text{für alle } n \ge n_0.$$

Es folgt

$$\int (|f_n| - a_1)^+ \, d\mu \le \int (|f| - a_1)^+ \, d\mu + \int ||f_n| - |f|| \, d\mu < \frac{2}{3}\epsilon \quad \text{für alle } n \ge n_0.$$

Da die endliche Familie $\{f_1, \dots, f_{n_0}\}$ gleichgradig integrierbar ist (siehe Lemma 8.16), gibt es ein $a_2 > 0$ derart, dass $\int(|f_i| - a_2)^+ \, d\mu < \epsilon$ für $i = 1, \dots, n_0$. Für $a = \max\{a_1, a_2\}$ folgt daher $\sup_{n>0} \int (|f_n| - a)^+ \, d\mu < \epsilon$.

"ii) $\Rightarrow$ i)": Sei $(f_{n_k})_{k>0}$ eine beliebige Teilfolge von $(f_n)_{n>0}$. Wegen Satz 8.19 gibt es eine Teilfolge $(f_{n_{k(i)}})_{i>0}$ von $(f_{n_k})_{k>0}$ mit $\lim_{i\to\infty} f_{n_{k(i)}} = f$ μ-f.s. Aus Satz 8.17 folgt: $f \in \mathcal{L}^1_\mu$ und $\lim_{i\to\infty} \int |f_{n_{k(i)}} - f| \, d\mu = 0$. Also hat jede Teilfolge der Folge $(\int |f_n - f| \, d\mu)_{n>0}$ eine Teilfolge, die gegen 0 konvergiert, so dass die Folge selbst auch gegen 0 konvergiert. $\square$

Wir fassen die Beziehungen zwischen den verschiedenen Konvergenzbegriffen bei endlichem Maß μ zusammen:

$$\begin{tikzcd}
f_n \to f \ \mu\text{-f.s.} \arrow[rr, Rightarrow, "\text{falls } (f_n)_n \text{ gleichgradig int'bar}"'] \arrow[dr, Rightarrow] & & \int |f_n - f| \, d\mu \to 0 \arrow[ll, dotted, bend right, "\text{für Teilfolgen}"'] \arrow[dl, Rightarrow] \\
& f_n \underset{\mu}{\to} f \arrow[ul, dotted, bend left, "\text{für Teilfolgen}"] \arrow[ur, bend right, "\text{falls } (f_n)_n \text{ gleichgradig int'bar}"'] &
\end{tikzcd}$$

8.21 Bemerkung (Schwaches Gesetz der großen Zahl) Seien $X_1, X_2, \dots \in \mathcal{L}^2_P$ unkorreliert und $M := \sup_n V(X_n) < \infty$. Dann konvergiert

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - EX_i) \underset{P}{\to} 0 \, ,$$

denn aus der Chebyshev-Ungleichung 7.16 folgt:

$$P\left(\left| \frac{1}{n} \sum_{i=1}^{n} (X_i - EX_i) \right| > \alpha \right) \le \alpha^{-2} \, V\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - EX_i) \right) = \alpha^{-2} \, \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) \le \alpha^{-2} M n^{-1} \, .$$

Kapitel 9

Der Ergodensatz - ein starkes Gesetz der großen Zahl für stationäre Prozesse

Sei $(X_t)_{t\in T}$ ein stationärer Prozess auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, wobei wir uns in diesem Abschnitt auf $T = \mathbb{Z}^d$ oder $T = \mathbb{Z}_+^d$ einschränken. Wie in Definition 6.9 sei $\mathcal{X} : \Omega \to \mathbb{R}^T$, $(\mathcal{X}\omega)_t = X_t(\omega)$, und es bezeichne $P_{\mathcal{X}}$ die Verteilung des Prozesses auf $\mathbb{R}^T$. Zur Erinnerung: $X_t = \pi_t \circ \mathcal{X}$.

In dieser Situation gilt

9.1 Satz (Birkhoffscher Ergodensatz für stationäre Prozesse)
Sei $\Lambda_n := \{t \in T : |t_i| < n \ (i = 1, \dots, d)\}$. *Ist* $X_t \in \mathcal{L}_P^1$, *so existiert*

$$\bar{X} := \lim_{n\to\infty} \frac{1}{|\Lambda_n|} \sum_{t\in\Lambda_n} X_t$$

P*-f.s. und in* L_P^1, *und für die Zufallsvariable* $\bar{X}$ *gilt:*

a) $E|\bar{X}| \leq E|X_1| < \infty$

b) Ist $(X_t)_{t\in T}$ *ergodisch, so ist* $\bar{X} = EX$ *konstant* P*-f.s. Insbesondere gilt das für unabhängige identisch verteilte* X_t.

Im Spezialfall $T = \mathbb{Z}_+$ erhalten wir (nach Verschiebung des Index um 1)

9.2 Korollar (Starkes Gesetz der großen Zahl für u.i.v. Prozesse) *Sind* $X_1, X_2, \dots \in \mathcal{L}_P^1$ *unabhängig identisch verteilt, so ist* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = EX_1$ P*-f.s. und in* L_P^1.

Wir werden Satz 9.1 nur für den Fall $T = \mathbb{Z}_+$ beweisen. Dann folgt er aus (und ist in der Tat äquivalent zu)

9.3 Satz (Birkhoffscher Ergodensatz für maßerhaltende Transformationen)
Sei $(X, \mathcal{F}, \mu)$ ein Wahrscheinlichkeitsraum, $T : X \to X$ messbar und maßerhaltend (d.h. $\mu \circ T^{-1} = \mu$). Sei $f \in \mathcal{L}^1_\mu$. Dann existiert

$$\bar{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(T^k x)$$

μ-f.s. und in $\mathcal{L}^1_\mu$, und es gilt:

a) $\int |\bar{f}|\, d\mu \le \int |f|\, d\mu < \infty$, also $\bar{f} \in \mathcal{L}^1_\mu$,

b) Ist (T, μ) ergodisch (d.h. ist $\mu(A) = 0$ oder 1 falls $T^{-1}(A) = A$), so ist $\bar{f} = \int f\, d\mu$ μ-f.s.

Satz 9.1 für $T = \mathbb{Z}_+$ folgt daraus wenn man $(X, \mathcal{F}, \mu) = (\mathbb{R}^T, \mathcal{B}^T, P_\mathcal{X})$, $T = S_1$ und $f = \pi_0$ betrachtet, da $X_i = \pi_i \circ \mathcal{X} = (\pi_0 \circ T^i) \circ \mathcal{X}$.

Die Identifikation der Grenzwerte $\bar{X}$ bzw. $\bar{f}$ im nicht-ergodischen Fall wird am Ende von Kapitel 19 nachgeliefert.

Beweis: Wir beweisen den Satz hier nur für ergodische (T, μ). Der allgemeine Fall wird erst in Kapitel 19 behandelt.

Seien $c := \int f\, d\mu$ und $S_n f := \sum_{k=0}^{n-1} f \circ T^k$ $(n \ge 0)$. Wir werden zeigen, dass

$$F := \limsup_{n\to\infty} \frac{1}{n} S_n f \le c \quad \mu\text{-f.s.}$$

Durch Anwendung des gleichen Arguments auf die Funktion $-f$ folgt dann $\limsup_{n\to\infty} \frac{1}{n} S_n(-f) \le -c$ μ-f.s., d.h. $\liminf_{n\to\infty} \frac{1}{n} S_n f \ge c$ μ-f.s. Beide Ungleichungen zusammen ergeben schließlich die Behauptung $\lim_{n\to\infty} \frac{1}{n} S_n f = c$ μ-f.s.

Für $\epsilon > 0$ seien $g_\epsilon := f - c - \epsilon$ und $G_\epsilon := \limsup_{n\to\infty} \frac{1}{n} S_n g_\epsilon$. Dann ist $F = G_\epsilon + c + \epsilon$, und es reicht zu zeigen, dass $\mu\{G_\epsilon > 0\} = 0$ für jedes $\epsilon > 0$, denn daraus folgt $\mu\{F > c + \epsilon\} = 0$ für jedes $\epsilon > 0$, so dass

$$\mu\{F > c\} = \mu\left(\bigcup_{k=1}^{\infty} \{F > c + \frac{1}{k}\} \right) \le \sum_{k=1}^{\infty} \mu\{F > c + \frac{1}{k}\} = 0 ,$$

also $F \le c$ μ-f.s.

Nun zum Beweis von $\mu\{G_\epsilon > 0\} = 0$: Beachte zunächst, dass $G_\epsilon \circ T = G_\epsilon$:

$$G_\epsilon \circ T = \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g_\epsilon \circ T^k = \limsup_{n\to\infty} \frac{n+1}{n} \left(\frac{1}{n+1} \sum_{k=0}^{n} g_\epsilon \circ T^k - \frac{1}{n+1} g_\epsilon \right) = G_\epsilon .$$

Also ist $\{G_\epsilon > 0\}$ eine T-invariante Menge, und aus der Annahme der Ergodizität von (T, μ) folgt $\mu\{G_\epsilon > 0\} = 0$ oder 1.

Annahme: Es gibt ein $\epsilon > 0$ mit $\mu\{G_\epsilon > 0\} = 1$.

Sei $g := g_\epsilon$, und für $n > 0$ sei $M_n := \max\{0, S_1 g, \dots, S_n g\}$. Dann ist $0 \le M_1 \le M_2 \le M_3 \le \dots$, und da $S_k g(Tx) = S_{k+1} g(x) - g(x)$ für alle $k \ge 0$, folgt für $x \in \{M_n > 0\}$ *Hopfs Maximalungleichung*

$$\begin{aligned} M_n(x) - M_n(Tx) &= \max\{0, S_1 g(x), \dots, S_n g(x)\} - \max\{0, S_1 g(Tx), \dots, S_n g(Tx)\} \\ &= \max\{S_1 g(x), \dots, S_n g(x)\} - \max\{S_1 g(x), S_2 g(x), \dots, S_{n+1} g(x)\} + g(x) \\ &\le g(x) . \end{aligned}$$

Da $M_n \ge 0$ für alle n, folgt daraus

$$\int g 1_{\{M_n > 0\}}\, d\mu \ge \int_{\{M_n > 0\}} M_n\, d\mu - \int_{\{M_n > 0\}} M_n \circ T\, d\mu \ge \int M_n\, d\mu - \int M_n \circ T\, d\mu = 0 .$$

Nun ist aber $\{M_n > 0\} \nearrow \bigcup_{n=1}^{\infty}\{S_n g > 0\} \supseteq \{G_\epsilon > 0\}$, so dass aus Korollar 8.10 zum Satz von der majorisierten Konvergenz und aus der Annahme $\mu\{G_\epsilon > 0\} = 1$ folgt

$$\int g\,d\mu = \int_{\{G_\epsilon > 0\}} g\,d\mu = \lim_{n\to\infty} \int_{\{M_n>0\}} g\,d\mu \geq 0$$

im Widerspruch zu $\int g\,d\mu = \int f\,d\mu - c - \epsilon = -\epsilon < 0$. Also ist die Anahme falsch, und wir haben $\mu\{G_\epsilon > 0\} = 0$ für jedes $\epsilon > 0$.

$\square$

9.4 Beispiel (Größe subkritischer Perkolationskomponenten) Sei $(X_e)_{e\in E}$ unabhängig und Bernoulli-verteilt mit $P(X_e = 1) = p$, $P(X_e = 0) = 1 - p$, wobei $E = \mathbb{Z}^2 \times \left\{\binom{1}{0}, \binom{0}{1}\right\}$. Wir betrachten das kanonische Modell von $(X_e)_{e\in E}$, d.h. wir betrachten $X_e = \pi_e : \{0,1\}^E \to \{0,1\}$. Es sei $p < p_c = \frac{1}{2}$ ein subkritischer Perkolationsparameter, siehe Satz 6.19. Für $t \in \mathbb{Z}^2$ betrachten wir die Zufallsvariablen

$$Y_t(\omega) := \text{Mächtigkeit der Zusammenhangskomponente } J_t(\omega) \text{ von } \mathcal{G}(E(\omega)), \text{ die } t \text{ enthält.}$$

Zur Erinnerung: $E(\omega) = \{e \in E : X_e(\omega) = \omega_e = 1\}$. Dann ist $Y_t = Y_0 \circ S_t$, so dass $(Y_t)_{t\in\mathbb{Z}^2}$ ein stationärer, ergodischer Prozess ist (Übung!). Im Beweis zu Satz 6.19 hatten wir in Abschätzung (6.1) gesehen, dass

$$P(\text{Radius von } J_0(\omega) \text{ um } 0 \text{ größer als } n) \leq \frac{4}{3}(3p)^n\,.$$

Es folgt: $P(Y_0 \geq k) \leq (3p)^{\sqrt{k/\pi}}$ und daher $EY_0 \leq \sum_{k=0}^{\infty} P(Y_0 > k) < \infty$ (Bemerkung 7.11). Wir können daher den Ergodensatz anwenden und z.B. schließen

$$\lim_{n\to\infty} \frac{1}{|\Lambda_n|} \sum_{t\in\Lambda_n} Y_t = EY_0 \quad \text{f.s.}$$

Beachte dabei, dass die Y_t *nicht* unabhängig sind: Wählt man zwei beliebige Gitterpunkte t und t' und weiß man, dass $Y_t(\omega)$ sehr groß ist (was sehr unwahrscheinlich ist), so ist die darauf bedingte Wahrscheinlichkeit, dass t' zu $J_t(\omega)$ gehört recht groß, und da dann $J_{t'}(\omega) = J_t(\omega)$, erwartet man nach Kenntnis von $Y_t(\omega)$ ein untypisch großes $Y_{t'}(\omega)$.

Kapitel 10

Minkowski, Hölder und Jensen

In diesem Kapitel stellen wir eine Reihe der wichtigsten Ungleichungen der Analysis zur Verfügung. Insbesondere zeigen wir, dass die sogenannten L^p–Räume Banach-Räume sind.

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$. Für messbare $f : \Omega \to \bar{\mathbb{R}}$ sei

$$\|f\|_p := \left(\int |f|^p \, d\mu \right)^{1/p}$$

Zur Erinnerung: Dann ist $\mathcal{L}^p_\mu = \{f : \Omega \to \bar{\mathbb{R}} \text{ mit } \|f\|_p < \infty\}$.

10.1 Satz (Minkowski- und Hölder-Ungleichung)

a) $\mathcal{L}^p_\mu$ *ist ein Vektorraum.*

b) $\|.\|_p$ *ist eine Semi-Norm auf* $\mathcal{L}^p_\mu$*, d.h. für* $f, g \in \mathcal{L}^p_\mu$ *und* $\alpha \in \mathbb{R}$ *gilt*

1) $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ (Minkowski-Ungleichung)

2) $\|\alpha f\|_p = |\alpha| \cdot \|f\|_p$

3) $\|f\|_p = 0 \Longleftrightarrow f = 0$ μ*-f.s.*

c) $f \in \mathcal{L}^p_\mu, g \in \mathcal{L}^q_\mu \Longrightarrow (fg) \in \mathcal{L}^1_\mu$, $\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q$ (Hölder-Ungleichung)

Beweis:

b2)

$$\|\alpha f\|_p = \left(\int |\alpha f|^p \, d\mu \right)^{1/p} = |\alpha| \cdot \left(\int |f|^p \, d\mu \right)^{1/p} = |\alpha| \cdot \|f\|_p$$

b3)

$$\|f\|_p = 0 \quad \Longleftrightarrow \quad \int |f|^p \, d\mu = 0 \quad \Longleftrightarrow \quad |f|^p = 0 \text{ f.s.} \quad \Longleftrightarrow \quad f = 0 \text{ f.s.}$$

c) Ist $\|f\|_p = 0$ oder $\|g\|_q = 0$, so ist $f = 0$ f.s. oder $g = 0$ f.s., also $fg = 0$ f.s. und daher $\|fg\|_1 = 0$. Seien nun $\|f\|_p > 0$ und $\|g\|_q > 0$. Für beliebige $a, b > 0$ gilt

$$ab = e^{\frac{1}{p}(p \log a) + \frac{1}{q}(q \log b)} \leq \frac{1}{p} e^{p \log a} + \frac{1}{q} e^{q \log b} = \frac{a^p}{p} + \frac{b^q}{q} \, ,$$

da $x \mapsto e^x$ konvex ist. Ist $a = 0$ oder $b = 0$, so gilt diese Ungleichung ebenfalls. Also

$$\frac{|f(\omega)|}{\|f\|_p} \cdot \frac{|g(\omega)|}{\|g\|_q} \leq \frac{1}{p} \frac{|f(\omega)|^p}{\int |f|^p \, d\mu} + \frac{1}{q} \frac{|g(\omega)|^q}{\int |g|^q \, d\mu}$$

Integriert man beide Seiten dieser Ungleichung mit $d\mu$, so folgt

$$\frac{\int |fg|\, d\mu}{\|f\|_p \cdot \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1 \,,$$

also die Behauptung.

a) i) $0 \in \mathcal{L}^p_\mu$ ✔

ii) Da $\|\alpha f\|_p = |\alpha| \cdot \|f\|_p$, ist mit f auch αf in $\mathcal{L}^p_\mu$.

iii) Seien $f, g \in \mathcal{L}^p_\mu$. Da

$$|f+g|^p \leq (|f|+|g|)^p \leq (2\, \max\{|f|,|g|\})^p = 2^p\, \max\{|f|^p, |g|^p\} \leq 2^p\, (|f|^p + |g|^p) \,, \tag{10.1}$$

ist

$$\int |f+g|^p \leq 2^p \left(\int |f|^p\, d\mu + \int |g|^p\, d\mu \right) < \infty \,,$$

also $f + g \in \mathcal{L}^p_\mu$.

b1) Seien $f, g \in \mathcal{L}^p_\mu$. Es gilt

$$\int |f+g|^p\, d\mu \leq \int (|f|+|g|)^p\, d\mu \,, \tag{*}$$

und die Ungleichung folgt sofort für $p = 1$. Sei nun $p > 1$. Dann ist

$$\begin{aligned}\int (|f|+|g|)^p\, d\mu &= \int |f|\, (|f|+|g|)^{p-1}\, d\mu + \int |g|\, (|f|+|g|)^{p-1}\, d\mu \\ &\overset{\text{Hölder-Ungl.}}{\leq} \|f\|_p \left(\int (|f|+|g|)^{(p-1)q}\, d\mu \right)^{1/q} + \|g\|_p \left(\int (|f|+|g|)^{(p-1)q}\, d\mu \right)^{1/q} .\end{aligned}$$

Da aus $\frac{1}{p} + \frac{1}{q} = 1$ folgt, dass $p + q = pq$, ist $(p-1)q = p$, und wir erhalten

$$\int (|f|+|g|)^p\, d\mu \leq (\|f\|_p + \|g\|_p) \left(\int (|f|+|g|)^p\, d\mu \right)^{1/q} ,$$

woraus wegen $|f| + |g| \in \mathcal{L}^p_\mu$ folgt

$$\left(\int (|f|+|g|)^p\, d\mu \right)^{1/p} \leq \|f\|_p + \|g\|_p \,.$$

Mit (*) folgt daraus die behauptete Ungleichung.

□

10.2 Bemerkungen

a) Für $p = q = 2$ erhält man $\|fg\|_1 \leq \|f\|_2 \|g\|_2$ (*Cauchy/Schwarz-Ungleichung*)

b) Ist $\mu(\Omega) = 1$ und $1 \leq p < r$, so gilt für $f \in \mathcal{L}^r_\mu$

$$\|f\|_p = \left(\int |f|^p \cdot 1\, d\mu \right)^{1/p} \overset{\text{Hölder-Ungl.}}{\leq} \left(\||f|^p\|_{\frac{r}{p}} \cdot \|1\|_{\frac{r}{r-p}} \right)^{1/p} = \left(\int |f|^r\, d\mu \right)^{\frac{p}{r} \cdot \frac{1}{p}} = \|f\|_r$$

10.3 Bemerkung Sei

$$\mathcal{L}^\infty_\mu := \{ f : \Omega \to \mathbb{R} \mid \exists\, M > 0 \text{ so dass } |f| \leq M\ \mu\text{-f.s.} \} \,.$$

Für $f \in \mathcal{L}^\infty_\mu$ sei

$$\|f\|_\infty := \inf \{ M > 0 : |f| \leq M\ \mu\text{-f.s.} \} \,.$$

Dann gelten für $p = \infty$ und $q = 1$ die gleichen Aussagen wie in Satz 10.1 (Übung!).

10.4 Bemerkung Sei $\mathcal{F} \subseteq \mathcal{L}^1_\mu$ und $\sup_{f\in\mathcal{F}} \|f\|_r =: C < \infty$ für ein $r > 1$. Dann ist $\mathcal{F}$ *gleichgradig integrierbar*, denn für $a > 0$ und $f \in \mathcal{F}$ gilt

$$\int_{\{|f|>a\}} |f|\,d\mu \le a^{-(r-1)} \int_{\{|f|>a\}} |f|^r\,d\mu \le a^{-(r-1)} C^r \to 0 \quad \text{für } a \to \infty.$$

10.5 Satz (Vollständigkeit von $\mathcal{L}^p_\mu$)
Sei $1 \le p \le \infty$, $f_n \in \mathcal{L}^p_\mu$, und sei $\lim_{n\to\infty} \sup_{m\ge n} \|f_n - f_m\|_p = 0$. Dann existiert ein $f \in \mathcal{L}^p_\mu$ derart, dass $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Beweis: Der Beweis für $p = \infty$ bleibt zur Übung überlassen. Hier sei $1 \le p < \infty$. Wähle eine Teilfolge $(f_{n_k})_k$ von $(f_n)_n$ mit $\|f_{n_{k+1}} - f_{n_k}\|_p \le 2^{-k}$. Wir wollen zeigen, dass $f := \lim_{k\to\infty} f_{n_k}$ fast sicher existiert. Sei dazu $g_m := \sum_{k=1}^m |f_{n_{k+1}} - f_{n_k}|$. Dann folgt aus der Minkowski-Ungleichung, dass

$$\|g_m\|_p \le \sum_{k=1}^m 2^{-k} \le 1\ .$$

Außerdem strebt $g_m \nearrow g := \sum_{k=1}^\infty |f_{n_{k+1}} - f_{n_k}|$, also auch $g_m^p \nearrow g^p$ und daher

$$\int g^p\,d\mu = \sup_m \int g_m^p\,d\mu = \sup_m \|g_m\|_p^p \le 1\ ,$$

also $g^p \in \mathcal{L}^1_\mu$, so dass $g < \infty$ μ-f.s. Insbesondere existiert

$$f := f_{n_1} + \sum_{k=1}^\infty (f_{n_{k+1}} - f_{n_k}) \quad \text{f.s., und} \quad |f| \le |f_{n_1}| + g\ ,$$

so dass

$$\|f\|_p \le \||f_{n_1}| + g\|_p \le \|f_{n_1}\|_p + \|g\|_p < \infty\ ,$$

also $f \in \mathcal{L}^p_\mu$. Aus dem Lemma von Fatou folgt schließlich

$$\begin{aligned}\|f_n - f\|_p^p = \int |f_n - f|^p\,d\mu &= \int \liminf_{m\to\infty} |f_n - f_{n_m}|^p\,d\mu \\ &\le \liminf_{m\to\infty} \int |f_n - f_{n_m}|^p\,d\mu \le \sup_{m\ge n} \|f_n - f_m\|_p^p =: \epsilon_n\end{aligned}$$

und $\epsilon_n \to 0$ nach Voraussetzung. □

10.6 Bemerkung (L^p-Räume) Sei L^p_μ die Menge der $\|.\|_p$-Äquivalenzklassen von $\mathcal{L}^p_\mu$ ($f \sim g$, falls $\|f - g\|_p = 0$, d.h. falls $f = g$ μ-f.s.). Aus dem vorhergehenden Satz folgt, dass L^p_μ ein *Banach-Raum* ist. Im Fall $p = 2$ ist es sogar ein *Hilbert-Raum* mit Skalarprodukt $\langle f, g\rangle = \int fg\,d\mu$. Das Skalarprodukt ist wohldefiniert, da $\int |fg|\,d\mu \le \|f\|_2\|g\|_2 < \infty$, und offensichtlich ist $\|f\|_2 = \langle f, f\rangle^{\frac{1}{2}}$.

Schon im Beweis der Hölder-Ungleichung spielte die Konvexität von $x \mapsto e^x$ eine entscheidende Rolle. Ist μ ein Wahrscheinlichkeitsmaß, so lässt sich eine viel allgemeinere Ungleichung für Integrale konvexer Funktionen beweisen.

10.7 Definition (Konvexität) *Sei $I \subseteq \mathbb{R}$ ein Intervall. $\varphi : I \to \mathbb{R}$ ist konvex, falls*

$$\varphi(ta + (1-t)b) \le t\varphi(a) + (1-t)\varphi(b) \quad \textit{für alle } a, b \in I \textit{ und } t \in [0,1].$$

φ ist strikt konvex, falls diese Ungleichung für $a \ne b$ und $t \in (0,1)$ strikt ist.

10.8 Bemerkung $\varphi : I \to \mathbb{R}$ ist konvex genau dann, wenn

$$\frac{\varphi(v)-\varphi(u)}{v-u} \le \frac{\varphi(w)-\varphi(u)}{w-u} \le \frac{\varphi(w)-\varphi(v)}{w-v} \quad \text{für alle } u<v<w \text{ in } I. \tag{*}$$

(Setze $u=a$, $w=b$, $v=ta+(1-t)b$; dann einfache Übung.) Insbesondere ist jedes konvexe φ stetig im Inneren von I, also auch messbar, und für alle $x \in \mathring{I}$ existiert

$$\varphi^*(x) := \lim_{v \downarrow x} \frac{\varphi(v)-\varphi(x)}{v-x} = \lim_{n\to\infty} \frac{\varphi(x+\frac{1}{n})-\varphi(x)}{\frac{1}{n}} \quad \text{(Monotonie!).}$$

Mit φ ist auch φ^* messbar, und wegen (*) ist

$$\frac{\varphi(x)-\varphi(u)}{x-u} \le \varphi^*(x) \le \frac{\varphi(v)-\varphi(x)}{v-x} \quad \text{für alle } u<x<v \text{ in } I. \tag{**}$$

Ist φ strikt konvex, so sind alle Ungleichungen in (*) und (**) strikt.

10.9 Satz (Jensen'sche Ungleichung)
Sei $I \subseteq \mathbb{R}$ ein Intervall (nicht notwendig endlich!), $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, X eine integrierbare *Zufallsvariable mit Werten in I f.s., und sei $\varphi : I \to \mathbb{R}$ konvex. Dann ist*

$$\varphi(EX) \le E[\varphi(X)] \ .$$

Ist φ strikt konvex, so gilt Gleichheit genau dann, wenn $X = EX$ f.s.

Beweis: Wegen (**) gilt für $x, y \in I$

$$\varphi^*(x)\cdot(y-x)+\varphi(x) \le \varphi(y)$$

mit strikter Ungleichung falls φ strikt konvex und $x \ne y$ ist. Da $X(\omega) \in I$ f.s., ist auch $EX \in I$. Ist EX ein Endpunkt von I, z.B. $EX = \min I$, so ist $X \ge EX$ f.s. und daher $X = EX$ f.s., insbesondere $E[\varphi(X)] = \varphi(EX)$. Andernfalls ist $EX \in \mathring{I}$, und es folgt mit $x = EX$ und $y = X$

$$\varphi^*(EX)\cdot(X-EX)+\varphi(EX) \le \varphi(X) \quad \text{f.s.} \tag{***}$$

Nimmt man auf beiden Seiten dieser Ungleichung den Erwartungswert, so folgt wegen $E[X-EX]=0$ die Behauptung. Ist φ strikt konvex und X nicht fast sicher konstant, so herrscht mit positiver Wahrscheinlichkeit strikte Ungleichung in (***). □

10.10 Bemerkung (Jensen impliziert Hölder) Die Hölder-Ungleichung kann direkt aus der Jensen-Ungleichung gefolgert werden: Seien $f \in \mathcal{L}^p_\mu$, $g \in \mathcal{L}^q_\mu$. Setze

$$\tilde{f} := \left(\frac{|f|}{\|f\|_p}\right)^p, \quad \tilde{g} := \left(\frac{|g|}{\|g\|_q}\right)^q .$$

Dann ist $\int \tilde{f}\,d\mu = \int \tilde{g}\,d\mu = 1$. Insbesondere ist $P := \tilde{f}\,\mu$ ein Wahrscheinlichkeitsmaß (Korollar 8.3). Es folgt

$$\int \frac{|fg|}{\|f\|_p\|g\|_q}\,d\mu = \int \tilde{f}^{\frac{1}{p}}\tilde{g}^{\frac{1}{q}}\,d\mu = \int \tilde{f}\left(\frac{\tilde{g}}{\tilde{f}}\right)^{\frac{1}{q}} d\mu = E_P\left[\left(\frac{\tilde{g}}{\tilde{f}}\right)^{\frac{1}{q}}\right]$$

da $\frac{1}{p}+\frac{1}{q}=1$. (Beachte auch, dass der Quotient P-f.s. wohldefiniert ist!) Da $x \mapsto x^q$ konvex auf $[0,\infty)$ ist, folgt aus der Jensen'schen Ungleichung

$$\left(\int \frac{|fg|}{\|f\|_p\|g\|_q}\,d\mu\right)^q = \left(E_P\left[\left(\frac{\tilde{g}}{\tilde{f}}\right)^{\frac{1}{q}}\right]\right)^q \le E_P\left[\frac{\tilde{g}}{\tilde{f}}\right] = \int \tilde{g}\,d\mu = 1 \ ,$$

also die Hölder-Ungleichung.

Kapitel 11

Übergangskerne und Produktmaße – zwei Faktoren

Das Produkt zweier Maße ist aus der Analysis bekannt. So gilt z.B. für das m- und das n-dimensionale Lebesguemaß $\lambda^m \times \lambda^n = \lambda^{m+n}$. Mit solchen Produkten lassen sich in der Wahrscheinlichkeitstheorie gemeinsame Verteilungen unabhängiger Zufallsvariablen konstruieren. Zur Beschreibung der gemeinsamen Verteilung abhängiger Zufallsvariablen benötigt man aber eine flexiblere Konstruktion, eine Art Schiefprodukt von Maßen. Die dazu nötige Theorie wird in diesem Kapitel bereit gestellt.

11.1 Beispiele Wir bereiten die allgemeine Begriffsbildung mit zwei Beispielen vor.

a) Sei Σ eine endliche oder abzählbare Menge, interpretiert als Menge möglicher Zustände eines "Systems". Übgergänge zwischen den Zuständen sind mit gewissen Wahrscheinlichkeiten möglich: Befindet sich das System im Zustand s, so gehe es mit Wahrscheinlichkeit $p_{s,s'}$ in den Zustand s' über. Das heißt, für jedes $s \in \Sigma$ ist $(p_{ss'})_{s' \in \Sigma}$ ein Wahrscheinlichkeitsvektor. Die $\Sigma \times \Sigma$-Matrix $(p_{ss'})_{ss'}$, in der also jede Zeile ein Wahrscheinlichkeitsvektor ist, ist eine sogenannte *Markov-Matrix*.

   Sei $(\pi_s)_{s\in\Sigma}$ ein weiterer Wahrscheinlichkeitsvektor. Wird das System zum Zeitpunkt $t = 0$ mit Wahrscheinlichkeit π_s in den Zustand $s_0 \in \Sigma$ versetzt und wird dann einmal ein Folgezustand $s_1 \in \Sigma$ gemäß obigem Mechanismus gewählt, so ergibt sich für das Paar (s_0, s_1) die Wahrscheinlichkeit $q_{(s_0,s_1)} := \pi_{s_0} p_{s_0 s_1}$. Dadurch wird ein Wahrscheinlichkeitsmaß Q auf Σ^2 als "Produkt" der *Startverteilung* π und der Markov-Matrix P definiert.

   Vom Zustand s_1 kann man nach der gleichen Regel in einen witeren Zustand s_2 übergehen. Die Wahrscheinlichkeit, dabei die Abfolge (s_0, s_1, s_2) zu beobachten, ist dann $\pi_{s_0} p_{s_0 s_1} p_{s_1 s_2}$. Man modelliert damit eine Situation, in der man sich bei der Entscheidung für s_2 nur vom momentanen Zustand s_1 leiten lässt, nicht aber vom vergangenen Zustand s_0.

b) Hier ist eine Variante des vorherigen Beispiels mit kontinuierlichem Zustandsraum: Sei $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ das "Einheitsintervall modulo 1" und $f : \mathbb{T} \to \mathbb{R}$ eine differenzierbare Funktion (hier als '"Potential" auf I interpretiert). Sei X_0 eine Zufallsvariable mit Werten in $\mathbb{T}$. Ausgehend vom Zustand X_0 zur Zeit $t = 0$ geht das System nun zufällig in einen Zustand X_1 zur Zeit $t = 1$ über, dann mit der gleichen Regel nach X_2 u.s.w. Dabei soll die Art des Zufalls vom Potential abhängen

   $$X_{n+1} = X_n - h \cdot f'(X_n) + \xi_{n+1} \qquad (n = 0, 1, 2, \dots).$$

   Dabei ist $h > 0$ eine fest gewählte Konstante (typischerweise sollte h klein sein), und die $\xi_1, \xi_2, \dots$ sind unabhängige Zufallsvariablen. Dann gilt für die "bedingte Verteilung von X_{n+1} gegeben $X_n = x$":

   $$P(X_{n+1} \in A | X_n = x) = P_{\xi_{n+1}}(A - x + h \cdot f'(x)) \ .$$

   Wieder beeinflussen die vergangenen Zustände $X_0, \dots, X_{n-1}$ nicht die zufällige Wahl von X_{n+1}, solange $X_n = x$ bekannt ist. Die gemeinsame Verteilung von X_0 und X_1 sollte sich dann wie im vorherigen Beispiel als "Produkt" der Verteilung P_{X_0} auf $\mathbb{T}$ mit dem "Markovkern" $K(x, A) := P_{\xi_1}(A - x - h \cdot f'(x))$ schreiben lassen.

Ziel dieses Kapitels ist es, diese Idee in sehr allgemeiner Form zu präzisieren.

11.2 Definition (Übergangskern, stochastischer Kern, Markov-Kern) *Seien $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$, messbare Räume. Eine Abbildung $K : \Omega_1 \times \mathcal{A}_2 \to [0, +\infty)$ heißt Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, falls gilt:*

i) $A \mapsto K(\omega_1, A)$ ist ein endliches Maß auf $(\Omega_2, \mathcal{A}_2)$ für jedes $\omega_1 \in \Omega_1$, und

ii) $\omega_1 \mapsto K(\omega_1, A)$ ist $\mathcal{A}_1$-$\mathcal{B}$-messbar für jedes $A \in \mathcal{A}_2$.

Ist jedes $K(\omega_1, \cdot)$ ein Wahrscheinlichkeitsmaß, so ist K ein stochastischer Kern oder auch Markov-Kern. Integration bzgl. eines Maßes $K(\omega_1, \cdot)$ wird als $\int f(\omega_2)\, K(\omega_1, d\omega_2)$ geschrieben.

11.3 Beispiele a) **(Markov-Matrizen mit höchstens abzählbarem Zustandsraum)** In Beispiel 11.1a ist

$$K(\omega_1, A) := \sum_{j \in A} q_{\omega_1, j}$$

ein Markov-Kern von $(\Sigma, \mathcal{A})$ nach $(\Sigma, \mathcal{A})$, $\mathcal{A} = \mathfrak{P}(\Sigma)$.

b) In Beispiel 11.1b ist K ein Markov-Kern von $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ nach $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$.

c) μ ist ein endliches Maß auf $(\Omega_2, \mathcal{A}_2)$ und $K(\omega_1, A) := \mu(A)$ für alle $\omega_1 \in \Omega_1$. Ist μ ein Wahrscheinlichkeitsmaß, so interpretiert man das folgendermaßen: Die Verteilung von ω_2 ("das Wissen über ω_2") wird durch Informationen über ω_1 nicht beeinflusst.

d) Sei $T : \Omega_1 \to \Omega_2$ messbar, $K(\omega_1, A) := \delta_{T\omega_1}(A) = 1_{T^{-1}A}(\omega_1)$. In diesem Fall hängt ω_2 *deterministisch* von ω_1 ab, denn nur $\omega_2 = T\omega_1$ hat positive Wahrscheinlichkeit.

11.4 Bemerkung Es reicht, ii) der Definition für alle Mengen A aus einem $\cap$-stabilen Erzeuger $\mathcal{C}$ von $\mathcal{A}_2$ mit $\Omega_2 \in \mathcal{C}$ zu fordern, denn

$$\mathcal{D} := \{A \in \mathcal{A}_2 : \omega_1 \mapsto K(\omega_1, A) \text{ ist } \mathcal{A}_1\text{-}\mathcal{B}\text{-messbar}\}$$

ist ein Dynkin-System (Übung!) und $\mathcal{C} \subseteq \mathcal{D}$, also $\mathcal{D} = \mathcal{A}_2$ nach Satz 1.15.

11.5 Lemma *Sei $A \in \mathcal{A}_1 \times \mathcal{A}_2$, $f : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ $\mathcal{A}_1 \times \mathcal{A}_2$-messbar. Dann sind für alle $\tilde{\omega}_1 \in \Omega_1$, $\tilde{\omega}_2 \in \Omega_2$*

$$\begin{aligned}
&A_{\tilde{\omega}_1} := \{\omega_2 \in \Omega_2 : (\tilde{\omega}_1, \omega_2) \in A\} \in \mathcal{A}_2, \\
&A_{\tilde{\omega}_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \tilde{\omega}_2,) \in A\} \in \mathcal{A}_1, \\
&f_{\tilde{\omega}_1} : \Omega_2 \to \bar{\mathbb{R}},\ \omega_2 \mapsto f(\tilde{\omega}_1, \omega_2) \quad \mathcal{A}_2\text{-messbar, und} \\
&f_{\tilde{\omega}_2} : \Omega_1 \to \bar{\mathbb{R}},\ \omega_1 \mapsto f(\omega_1, \tilde{\omega}_2) \quad \mathcal{A}_1\text{-messbar.}
\end{aligned}$$

Beweis: Für $\tilde{\omega}_1$ definiere $i : \Omega_2 \to \Omega$ durch $i(\omega_2) = (\tilde{\omega}_1, \omega_2)$. Da $\pi_1 \circ i = \tilde{\omega}_1 = \mathrm{const}$ und $\pi_2 \circ i = Id_{\Omega_2}$ messbar sind, ist i nach Satz 6.3 $\mathcal{A}_2$-$\mathcal{A}_1 \times \mathcal{A}_2$-messbar, so dass $A_{\tilde{\omega}_1} = i^{-1}(A) \in \mathcal{A}_2$ und $f_{\tilde{\omega}_1} = f \circ i$ $\mathcal{A}_2$-$\bar{\mathcal{B}}$-messbar ist. Analog für $A_{\tilde{\omega}_2}$ und $f_{\tilde{\omega}_2}$. □

11.6 Lemma *Sei K ein Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$ und sei $f : \Omega_1 \times \Omega_2 \to [0, +\infty]$ $\mathcal{A}_1 \times \mathcal{A}_2$-messbar. Dann ist $g : \Omega_1 \to [0, +\infty]$,*

$$g(\omega_1) := \int f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)$$

wohldefiniert und $\mathcal{A}_1$-messbar.

Beweis: Da $g(\omega_1) = \int f_{\omega_1}(\omega_2)\, K(\omega_1, d\omega_2)$, ist $g(\omega_1)$ für jedes $\omega_1 \in \Omega_1$ wohldefiniert (siehe auch Lemma 11.5. Wir zeigen die Messbarkeit von g.

Sei $\mathcal{D} := \{A \in \mathcal{A}_1 \times \mathcal{A}_2 :\ g$ ist messbar für $f = 1_A\}$. Beachte, dass $g(\omega_1) = K(\omega_1, (A)_{\omega_1})$ falls $f = 1_A$. Für $A_i \in \mathcal{A}_i$ ist $A_1 \times A_2 \in \mathcal{D}$, denn dann ist $g(\omega_1) = 1_{A_1}(\omega_1) \cdot K(\omega_1, A_2)$ messbar. Also enthält $\mathcal{D}$ einen $\cap$-stabilen Erzeuger von $\mathcal{A}_1 \times \mathcal{A}_2$, siehe Satz 6.7. Außerdem ist $\mathcal{D}$ ein Dynkin-System:

D1) $f = 1_{\Omega_1 \times \Omega_2} \Longrightarrow g(\omega_1) = K(\omega_1, \Omega_2)$ ist messbar.

D2) $A, B \in \mathcal{D}$, $A \subseteq B \Longrightarrow g(\omega_1) = K(\omega_1, (B \setminus A)_{\omega_1}) = K(\omega_1, (B)_{\omega_1} \setminus (A)_{\omega_1}) = K(\omega_1, (B)_{\omega_1}) - K(\omega_1, (A)_{\omega_1})$ ist messbar. (Hier wird die Endlichkeit der Maße $K(\omega_1, \boldsymbol{.})$ benötigt!)

D3) $A_n \in \mathcal{D}$ $(n \geq 1)$ paarweise disjunkt $\Longrightarrow g(\omega_1) = K(\omega_1, (\biguplus_n A_n)_{\omega_1}) = K(\omega_1, \biguplus_n (A_n)_{\omega_1}) = \sum_n K(\omega_1, (A_n)_{\omega_1})$ ist messbar.

Also ist $\mathcal{D} = \mathcal{A}_1 \times \mathcal{A}_2$ (Satz 1.15), so dass g für alle $f = 1_A$ mit $A \in \mathcal{A}_1 \times \mathcal{A}_2$ messbar ist.

Da die Abbildung $f \mapsto g$ linear ist, folgt, dass g für Elementarfunktionen f messbar ist. Sind schließlich f_n Elementarfunktionen mit $0 \leq f_n \nearrow f$, so folgt aus dem Satz von der monotonen Konvergenz, dass $g_n(\omega_1) \nearrow g(\omega_1)$ für jedes feste $\omega_1 \in \Omega_1$, woraus wieder die Messbarkeit von g folgt. □

11.7 Satz (Existenz von Produktmaßen mit Übergangskernen)
Sei μ ein endliches Maß auf $(\Omega_1, \mathcal{A}_1)$ und sei K ein Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$. Dann wird durch

$$(\mu \times K)(A) := \int_{\Omega_1} \left(\int_{\Omega_2} 1_A(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \right) d\mu(\omega_1)$$

ein Maß auf $\mathcal{A}_1 \times \mathcal{A}_2$ definiert. $\mu \times K$ ist durch die Festlegung

$$(\mu \times K)(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2)\, d\mu(\omega_1)$$

für $A_i \in \mathcal{A}_i$ eindeutig bestimmt. Ist μ ein Wahrscheinlichkeitsmaß und ist K stochastisch, so ist $\mu \times K$ ein Wahrscheinlichkeitsmaß.

Beweis: Wegen Lemma 11.6 angewandt auf $f = 1_A$ ist $\mu \times K : \mathcal{A}_1 \times \mathcal{A}_2 \to [0, +\infty]$ wohldefiniert und $\mu \times K(\emptyset) = 0$. Seien nun $A_n \in \mathcal{A}_1 \times \mathcal{A}_2$ paarweise disjunkt $(n \geq 1)$, so dass für festes $\omega_1 \in \Omega_1$ auch die Mengen $(A_n)_{\omega_1}$, $n \geq 1$, paarweise disjunkt sind. Wegen Lemma 11.5 sind die $(A_n)_{\omega_1} \in \mathcal{A}_2$, und wegen der Maßeigenschaft des Übergangskerns ist

$$K\left(\omega_1, \left(\biguplus_n A_n\right)_{\omega_1}\right) = K\left(\omega_1, \biguplus_n (A_n)_{\omega_1}\right) = \sum_{n \geq 1} K(\omega_1, (A_n)_{\omega_1}) \,.$$

Daher folgt aus dem Satz von der monotonen Konvergenz

$$\begin{aligned}(\mu \times K)\left(\biguplus_{n \geq 1} A_n\right) &= \int_{\Omega_1} K\left(\omega_1, \left(\biguplus_n A_n\right)_{\omega_1}\right) d\mu(\omega_1) \\ &= \sum_{n \geq 1} \int_{\Omega_1} K\left(\omega_1, (A_n)_{\omega_1}\right)\, d\mu(\omega_1) \\ &= \sum_{n \geq 1} (\mu \times K)(A_n) \,.\end{aligned}$$

Aus dem Eindeutigkeitssatz 2.12 und aus Satz 6.7 folgt, dass $(\mu \times K)$ durch seine Werte auf $\mathcal{Z}(\mathcal{A}_1, \mathcal{A}_2)$ eindeutig bestimmt ist, da $\mu \times K$ auf $\mathcal{Z}(\mathcal{A}_1, \mathcal{A}_2)$ σ-endlich ist: Sei $A_n := \{\omega_1 \in \Omega_1 : K(\omega_1, \Omega_2) \leq n\}$. Dann sind die $A_n \in \mathcal{A}_1$, also $A_n \times \Omega_2 \in \mathcal{Z}(\mathcal{A}_1, \mathcal{A}_2)$, $\bigcup_n A_n \times \Omega_2 = \Omega_1 \times \Omega_2$, da alle

$K(\omega_1, \boldsymbol{.})$ endliche Maße sind, und es ist $(\mu \times K)(A_n \times \Omega_2) = \int_{A_n} K(\omega_1, \Omega_2)\, d\mu(\omega_1) \le n \cdot \mu(A_n) < \infty$ für alle n.

Ist μ ein Wahrscheinlichkeitsmaß und ist K stochastisch, so ist

$$(\mu \times K)(\Omega_1 \times \Omega_2) = \int_{\Omega_1} K(\omega_1, \Omega_2)\, d\mu(\omega_1) = 1\ ,$$

also ist $\mu \times K$ ein Wahrscheinlichkeitsmaß. □

11.8 Satz (Satz von Fubini für Übergangskerne)
Seien μ, K und $\mu \times K$ wie in Satz 11.7. Sei $f : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ $\mathcal{A}_1 \times \mathcal{A}_2$-messbar. Ist $f \ge 0$ oder ist $f \in \mathcal{L}^1_{\mu \times K}$, so gilt

$$\int_{\Omega_1 \times \Omega_2} f\, d(\mu \times K) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \right) d\mu(\omega_1)\ . \qquad (*)$$

Beweis: Für $f = 1_A$ mit $A \in \mathcal{A}_1 \times \mathcal{A}_2$ ist das die Definition von $\mu \times K$. Für allgemeine $f \ge 0$ wendet man algebraische Induktion an. Ist $f \in \mathcal{L}^1_{\mu \times K}$, $f = f^+ - f^-$, so gilt (*) für f^+ und f^-. Da $f^+, f^- \in \mathcal{L}^1_{\mu \times K}$, ist die linke Seite von (*) für $f^\pm$ endlich, so dass auch $\int_{\Omega_2} f^\pm_{\omega_1}(\omega_2)\, K(\omega_1, d\omega_2)$ für μ-f.a. $\omega_1 \in \Omega_1$ endlich ist (Satz 7.4b). Daher ist $\int_{\Omega_2} f_{\omega_1}(\omega_2)\, K(\omega_1, d\omega_2)$ für μ-f.a. $\omega_1 \in \Omega_1$ wohldefiniert und endlich, und (*) für f folgt aus der Linearität des Integrals. □

11.9 Korollar *Ist $A \in \mathcal{A}_1 \times \mathcal{A}_2$ eine $\mu \times K$-Nullmenge, so ist $K(\omega_1, (A)_{\omega_1}) = 0$ für μ-f.a. $\omega_1 \in \Omega_1$.*

Beweis: (*) für $f = 1_A$ und Satz 7.4a. □

Im Fall von Produktmaßen (d.h. $K(\omega_1, \boldsymbol{.}) = \nu(\boldsymbol{.})$) gelten die Sätze 11.7 und 11.8 auch ohne die Endlichkeitsannahme an die Maße μ und $K(\omega_1, \boldsymbol{.})$. Zum Beweis benötigen wir die folgende Charakterisierung σ-endlicher Maße:

11.10 Lemma (Charakterisierung σ-endlicher Maße) *Ein Maß μ auf dem messbaren Raum $(\Omega, \mathcal{A})$ ist σ-endlich genau dann, wenn es eine strikt positive Funktion $f \in \mathcal{L}^1_\mu$ gibt. Ein solches f kann beschränkt gewählt werden.*

Beweis: Sei μ σ-endlich. Es gibt also $A_n \in \mathcal{A}$ mit $A_n \nearrow \Omega$ und $0 < \mu(A_n) < \infty$. Setze $f := \sum_{n=1}^\infty (2^n \cdot \mu(A_n))^{-1}\, 1_{A_n}$. Dann ist $f(\omega) > 0$ für alle $\omega \in \Omega$ und $\int f\, d\mu = \sum_{n=1}^\infty 2^{-n} = 1 < \infty$, also $f \in \mathcal{L}^1_\mu$ und $f \le \max_k \mu(A_k)^{-1} \sum_{n=1}^\infty 2^{-n} = \mu(A_1)^{-1}$.

Ist umgekehrt $f \in \mathcal{L}^1_\mu$ strikt positiv, so setze $A_n := \{f \ge \frac{1}{n}\}$. Da $f(\omega) > 0$ für alle $\omega \in \Omega$, gilt $A_n \nearrow \Omega$, und es ist $\mu(A_n) \le \int_{A_n} n \cdot f\, d\mu \le n \cdot \int f\, d\mu < \infty$ für alle $n > 0$. □

11.11 Satz (Existenz von Produktmaßen)
Seien μ und ν σ-endliche Maße auf $(\Omega_1, \mathcal{A}_1)$ bzw. $(\Omega_2, \mathcal{A}_2)$. Dann wird durch

$$\mu \times \nu(A) := \int_{\Omega_1} \int_{\Omega_2} 1_A(\omega_1, \omega_2)\, d\nu(\omega_2)\, d\mu(\omega_1)$$

ein σ-endliches Maß auf $\mathcal{A}_1 \times \mathcal{A}_2$ definiert, das durch die Festlegung

$$\mu \times \nu(A_1 \times A_2) = \mu(A_1) \cdot \nu(A_2) \quad (A_i \in \mathcal{A}_i)$$

eindeutig bestimmt wird. Sind μ und ν endlich (Wahrscheinlichkeitsmaße), so ist auch $\mu \times \nu$ endlich (Wahrscheinlichkeitsmaß).

Beweis: Für endliche μ und ν ist das ein Spezialfall von Satz 11.7. Sonst gibt es $0 < f \in \mathcal{L}^1_\mu$ und $0 < g \in \mathcal{L}^1_\nu$ (Lemma 11.10). Setze $\tilde{\mu} := f \cdot \mu$, $\tilde{\nu} := g \cdot \nu$. Dann sind $\tilde{\mu}$ und $\tilde{\nu}$ endliche Maße (Korollar

8.3) und $h(\omega_1, \omega_2) := f(\omega_1) \cdot g(\omega_2)$ ist $\mathcal{A}_1 \times \mathcal{A}_2$-messbar. Also ist $\tilde{\mu} \times \tilde{\nu}$ ein endliches Maß (Satz 11.8) und für $A \in \mathcal{A}_1 \times \mathcal{A}_2$ ist

$$\begin{aligned} \int_A \frac{1}{h}\, d(\tilde{\mu} \times \tilde{\nu}) &= \int_{\Omega_1} \left(\int_{\Omega_2} 1_A(\omega_1, \omega_2) \frac{1}{g(\omega_2)}\, d\tilde{\nu}(\omega_2) \right) \frac{1}{f(\omega_1)}\, d\tilde{\mu}(\omega_1) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} 1_A(\omega_1, \omega_2)\, d\nu(\omega_2) \right) d\mu(\omega_1) \\ &= \mu \times \nu(A) \,. \end{aligned}$$

Nach Korollar 8.3 ist daher $\mu \times \nu = \frac{1}{h} \cdot (\tilde{\mu} \times \tilde{\nu})$ ein Maß, und da $h > 0$ und $\int h \cdot \frac{1}{h} \cdot (\tilde{\mu} \times \tilde{\nu}) = \tilde{\mu} \times \tilde{\nu}(\Omega_1 \times \Omega_2) < \infty$, ist dieses Maß σ-endlich.

Sei $\mathcal{A}_1' := \{A \in \mathcal{A}_1 : \mu(A) < \infty\}$, $\mathcal{A}_2' := \{A \in \mathcal{A}_2 : \nu(A) < \infty\}$. Da μ und ν σ-endlich sind, sind die $\mathcal{A}_i'$ $\cap$-stabile Erzeuger der $\mathcal{A}_i$, und es folgt aus Satz 6.5, dass $\mathcal{A}' := \{A_1 \times A_2 : A_i \in \mathcal{A}_i'\}$ ein $\cap$-stabiler Erzeuger von $\mathcal{A}_1 \times \mathcal{A}_2$ ist, auf dem das Maß $\mu \times \nu$ σ-endlich ist. Nach Satz 2.12 ist es daher durch seine Werte auf $\mathcal{A}'$ eindeutig bestimmt. □

11.12 Satz (Satz von Fubini für Produktmaße)
Seien μ und ν σ-endliche Maße auf $(\Omega_1, \mathcal{A}_1)$ bzw. $(\Omega_2, \mathcal{A}_2)$, und sei $f : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ $\mathcal{A}_1 \times \mathcal{A}_2$-messbar. Ist $f \geq 0$ oder ist $f \in \mathcal{L}^1_{\mu \times \nu}$, so gilt

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f\, d(\mu \times \nu) &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\, d\nu(\omega_2) \right) d\mu(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu(\omega_1) \right) d\nu(\omega_2) \end{aligned} \qquad (*)$$

Beweis: Die erste Gleichheit wird wie im Beweis zu Satz 11.8 gezeigt. (Die im Fall von Übergangskernen vorausgesetzte Endlichkeit von μ und $K(\omega_1, \centerdot)$ wurde dort nicht benutzt!) Für die zweite Gleichheit beachte, dass durch $A \mapsto \int_{\Omega_2} \int_{\Omega_1} 1_A(\omega_1, \omega_2)\, d\mu(\omega_1)\, d\nu(\omega_2)$ ebenfalls ein σ-endliches Maß auf $\mathcal{A}_1 \times \mathcal{A}_2$ definiert wird, das für $A_1 \times A_2 \in \mathcal{Z}(\mathcal{A}_1, \mathcal{A}_2)$ mit $\mu \times \nu$ übereinstimmt und daher mit $\mu \times \nu$ identisch ist (Satz 11.11). □

11.13 Korollar *Seien μ, ν wie in Satz 11.12 und $A \in \mathcal{A}_1 \times \mathcal{A}_2$. Ist $\mu \times \nu(A) = 0$, so ist $\mu((A)_{\omega_2}) = 0$ für ν-f.a. $\omega_2 \in \Omega_2$ und $\nu((A)_{\omega_1}) = 0$ für μ-f.a. $\omega_1 \in \Omega_1$.*

Beweis: (*) für $f = 1_A$ und Satz 7.4a. □

11.14 Beispiel (Lebesgue-Maß auf $\mathbb{R}^2$) $(\Omega_i, \mathcal{A}_i) = (\mathbb{R}, \mathcal{B})$, $\mu = \nu = \lambda$ (Lebesgue-Maß auf $\mathbb{R}$). $\lambda^2 := \lambda \times \lambda$ ist das Lebesgue-Maß auf $\mathbb{R}^2$. Es ist eindeutig bestimmt durch $\lambda^2((a,b] \times (c,d]) = (b-a) \cdot (d-c)$.

11.15 Korollar (Integral = "Fläche unter dem Funktionsgraphen") *Sei $(\Omega, \mathcal{A})$ messbarer Raum, μ σ-endlich auf $\mathcal{A}$, $f : \Omega \to [0, +\infty]$ $\mathcal{A}$-messbar. Sei $A_f := \{(\omega, x) \in \Omega \times \mathbb{R} : 0 < x \leq f(\omega)\}$. Dann ist*

$$\int_\Omega f\, d\mu = (\mu \times \lambda)(A_f) = \int_0^\infty \mu\{f \geq x\}\, d\lambda(x) \,.$$

Beweis: Betrachte zunächst $f = \sum_{i=1}^n \alpha_i \cdot 1_{A_i}$ mit paarweise disjunkten $A_i \in \mathcal{A}$. Dann ist

$$\int f\, d\mu = \sum_{i=1}^n \alpha_i\, \mu(A_i) = \sum_{i=1}^n \lambda((0, \alpha_i]) \cdot \mu(A_i) = \sum_{i=1}^n (\mu \times \lambda)(A_i \times (0, \alpha_i]) = (\mu \times \lambda)(A_f) \,.$$

Zu beliebigem nicht-negativem messbarem f gibt es Elementarfunktionen $f_n \nearrow f$. Dann ist auch $A_{f_n} \nearrow A_f$, und es folgt

$$\begin{aligned}\int f\,d\mu &= \sup_n \int f_n\,d\mu = \sup_n (\mu\times\lambda)(A_{f_n}) = (\mu\times\lambda)(A_f)\\ &\overset{\text{Fubini}}{=} \int_0^\infty \mu((A_f)_x)\,d\lambda(x) = \int_0^\infty \mu\{f\geq x\}\,d\lambda(x)\end{aligned}$$

□

11.16 Bemerkungen

a) Sind $\mu_1,\ldots,\mu_n$ σ-endliche Maße auf $(\Omega_1,\mathcal{A}_1),\ldots,(\Omega_n,\mathcal{A}_n)$, so wird durch

$$\mu_1\times\cdots\times\mu_n := (\ldots((\mu_1\times\mu_2)\times\mu_3)\cdots\times\mu_n)$$

ein σ-endliches Maß auf $\mathcal{A}_1\times\cdots\times\mathcal{A}_n$ definiert (beachte Korollar 6.8). Wegen der Festlegung von Produktmaßen durch ihre Werte auf Zylindermengen ist die Reihenfolge der Klammerung unwesentlich. Der Satz von Fubini überträgt sich analog. **Beispiel**: $\lambda^n := \lambda\times\cdots\times\lambda$ (*Lebesgue-Maß auf dem* $\mathbb{R}^n$).

b) Sind in der Situation von Teil a) h_i Wahrscheinlichkeitsdichten auf $(\Omega_i,\mathcal{A}_i,\mu_i)$ $(i=1,\ldots,n)$ und ist $h:\Omega_1\times\cdots\times\Omega_n\to[0,+\infty]$ definiert durch $h(\omega_1,\ldots,\omega_n)=h_1(\omega_1)\cdot\cdots\cdot h_n(\omega_n)$, so ist $(h_1\mu_1)\times\cdots\times(h_n\mu_n) = h\cdot(\mu_1\times\cdots\times\mu_n)$, denn beide Maße stimmen auf Zylindermengen $A_1\times\cdots\times A_n$ überein (Beweis für $n=2$):

$$\begin{aligned}(h_1\mu_1)\times(h_2\mu_2)(A_1\times A_2) &= (h_1\mu_1)(A_1)\cdot(h_2\mu_2)(A_2)\\ &= \int_{\Omega_1} 1_{A_1}h_1\,d\mu_1\cdot\int_{\Omega_2} 1_{A_2}h_2\,d\mu_2\\ &= \int_{\Omega_2}\left(\int_{\Omega_1} 1_{A_1}(\omega_1)1_{A_2}(\omega_2)h_1(\omega_1)h_2(\omega_2)\,d\mu_1(\omega_1)\right)d\mu_2(\omega_2)\\ &= \int_{\Omega_1\times\Omega_2} 1_{A_1\times A_2}(\omega_1,\omega_2)h(\omega_1,\omega_2)\,d(\mu_1\times\mu_2)(\omega_1,\omega_2)\\ &= (h\cdot\mu_1\times\mu_2)(A_1\times A_2)\end{aligned}$$

Die Verteilungen von Familien unabhängiger Zufallsvariablen lassen sich als Produktmaße charakterisieren:

11.17 Satz (Verteilungen von Familien unabhängiger Zufallsvariablen sind Produktmaße)
Seien $X_1,\ldots,X_n$ Zufallsvariablen auf $(\Omega,\mathcal{A},P)$. Bezeichne $\mathcal{X}:\Omega\to\bar{\mathbb{R}}^n$, $\omega\mapsto(X_1(\omega),\ldots,X_n(\omega))$. Dann gilt

$$X_1,\ldots,X_n \text{ unabhängig} \iff P_{\mathcal{X}} = P_{X_1}\times\cdots\times P_{X_n}\,,$$

d.h. die gemeinsame Verteilung der X_i ist das Produkt der einzelnen Verteilungen.
In diesem Fall gilt: Ist $f:\bar{\mathbb{R}}^n\to\bar{\mathbb{R}}$ nicht-negativ oder in $\mathcal{L}^1_{P_{\mathcal{X}}}$, so ist

$$E\left[f(X_1,\ldots,X_n)\right] = \int_{\bar{\mathbb{R}}^n} f\,dP_{\mathcal{X}} = \int_{\bar{\mathbb{R}}^n} f\,d(P_{X_1}\times\cdots\times P_{X_n})\,.$$

Beweis: $X_1,\ldots,X_n$ sind unabhängig genau dann, wenn für alle $A=A_1\times\cdots\times A_n\in\mathcal{Z}(\mathcal{A}_1,\ldots,\mathcal{A}_n)$ gilt

$$P_{\mathcal{X}}(A) = P\left(\bigcap_{i=1}^n\{X_i\in A_i\}\right) \overset{!}{=} \prod_{i=1}^n P\{X_i\in A_i\} = \prod_{i=1}^n P_{X_i}(A_i) = (P_{X_1}\times\cdots\times P_{X_n})(A)$$

Das wiederum ist äquivalent zu $P_{\mathcal{X}} = P_{X_1}\times\cdots\times P_{X_n}$. □

11.18 Definition (Faltung von Wahrscheinlichkeitsmaßen auf $\mathbb{R}$) *Bezeichne $\mathcal{W}$ die Menge aller Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$ und sei $s : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x + y$. Für $\mu, \nu \in \mathcal{W}$ sei*

$$\mu * \nu := (\mu \times \nu) \circ s^{-1} \in \mathcal{W}$$

die Faltung (oder das Faltungsprodukt) von μ und ν.

11.19 Satz (Eigenschaften der Faltung)

a) $\mu * \nu(A) = \int_{\mathbb{R}} \mu(A - x)\, d\nu(x) = \int_{\mathbb{R}} \nu(A - x)\, d\mu(x) = \nu * \mu(A)$.

b) Ist $\mu = f\lambda$, so ist $\mu * \nu = h\lambda$ mit $h(x) := \int_{\mathbb{R}} f(x - y)\, d\nu(y)$. Ist darüberhinaus $\nu = g\lambda$, so ist $h(x) = f * g(x) := \int_{\mathbb{R}} f(x - y) g(y)\, dy$.

c) Sind X und Y unabhängige Zufallsvariablen, so hat $X + Y$ die Verteilung $P_{X+Y} = P_X * P_Y$.

d) $*$ ist eine assoziative und kommutative Verknüpfung auf $\mathcal{W}$ mit neutralem Element δ_0.

Beweis:

a) Für $A \in \mathcal{B}$ folgt aus dem Satz von Fubini

$$\begin{aligned}\mu * \nu(A) &= (\mu \times \nu)(s^{-1}A) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x + y)\, d\mu(x)\, d\nu(y) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(y + x)\, d\nu(y)\, d\mu(x) \\ &= (\nu \times \mu)(s^{-1}A) = \nu * \mu(A)\end{aligned}$$

und $\int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x + y)\, d\mu(x)\, d\nu(y) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{A-x}(y)\, d\mu(y)\, d\nu(x) = \int_{\mathbb{R}} \mu(A - x)\, d\nu(x)$.

b) Für jedes $y \in \mathbb{R}$ ist

$$\int_{\mathbb{R}} 1_A(x + y)\, d\mu(x) = \int_{\mathbb{R}} 1_A(x + y) f(x)\, dx = \int_{\mathbb{R}} 1_A(x) f(x - y)\, dx\,,$$

so dass mit dem Satz von Fubini folgt

$$\mu * \nu(A) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x) f(x - y)\, dx\, d\nu(y) = \int_{\mathbb{R}} 1_A(x) \left(\int_{\mathbb{R}} f(x - y)\, d\nu(y) \right) dx\,.$$

c) Da X und Y unabhängig sind, hat $X + Y = s(X, Y)$ die Verteilung $(P_X \times P_Y) \circ s^{-1} = P_X * P_Y$, siehe Satz 11.17.

d) Ist $\nu = \delta_0$, so folgt $\mu * \nu(A) = \int_{\mathbb{R}} 1_A(x + 0)\, d\mu(x) = \mu(A)$. Die Assoziativität und Kommutativität folgen aus c), da jedes Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ Verteilung einer Zufallsvariablen auf $([0, 1], \mathcal{B}, \lambda)$ ist (Satz 5.4) und damit k solcher Zuzfallsvariablen auf $([0, 1]^k, \mathcal{B}^k, \lambda^k)$ unabhängig realisiert werden können (Satz 11.17).

□

Kapitel 12

Übergangskerne und Produktmaße – unendlich viele Faktoren

In diesem Abschnitt zeigen wir zunächst, dass Wahrscheinlichkeitsmaße auf unendlichen Produkträumen durch Festlegung ihrer Werte auf Zylindermengen definiert werden können. Das versetzt uns in die Lage, die Ergebnisse aus Kapitel 11 auf einfache Weise auf den Fall unendlich vieler Faktoren zu übertragen.

12.1 Definition (Verträgliches System von Verteilungen) *Bezeichne $\mathcal{E}(T)$ die Familie aller endlichen Teilmengen der Indexmenge T. Ein System $(\mu_I)_{I\in\mathcal{E}(T)}$ von* Wahrscheinlichkeitsmaßen *auf $(\mathbb{R}^I, \mathcal{B}^I)$ heißt verträglich (oder projektiv), falls*

$$\mu_J = \mu_I \circ \left(\pi_J^I\right)^{-1} \quad \textit{für alle } J \subseteq I \in \mathcal{E}(T).$$

12.2 Bemerkung (Marginalverteilungen) Sei ν ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^T, \mathcal{B}^T)$. Setze $\nu_I := \nu \circ (\pi_I^T)^{-1}$ für $I \in \mathcal{E}(T)$. Ist $J \subseteq I$, so folgt aus $\pi_J^T = \pi_J^I \circ \pi_I^T$, dass $(\nu_I)_{I\in\mathcal{E}(T)}$ verträglich ist.

Die ν_I mit $I \in \mathcal{E}(T)$ heißen endlich-dimensionale Rand- (oder Marginal-)verteilungen von ν.

Der folgende Satz ist die Umkehrung dieser Bemerkung:

12.3 Satz (Existenzsatz von Kolmogorov)
Sei $(\mu_I)_{I\in\mathcal{E}(T)}$ ein verträgliches System von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^I, \mathcal{B}^I)$, $(I \in \mathcal{E}(T))$. Dann gibt es genau ein Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}^T, \mathcal{B}^T)$ zu dem die μ_I die Randverteilungen sind, d.h. für das gilt

$$\mu_I = \mu \circ \left(\pi_I^T\right)^{-1} \quad \textit{für alle } I \in \mathcal{E}(T). \qquad (*)$$

Beweis: Wir schreiben $\mathcal{E} := \mathcal{E}(T)$ und betrachten den Semiring $\mathcal{Z}$ der Zylindermengen in $(\mathbb{R}^T, \mathcal{B}^T)$. Jedes Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}^T, \mathcal{B}^T)$ ist wegen Satz 2.12 durch die Bedingung (*) eindeutig bestimmt, denn (*) legt die Werte von μ auf Zylindermengen fest. Also kann es höchstens ein solches Maß geben. Wir wenden uns der Existenz zu: Für $B = \pi_I^{-1}(A)$ setze

$$\mu(B) := \mu_I(A)\ .$$

Wir zeigen, dass μ auf $\mathcal{Z}$ dadurch eindeutig definiert ist: Hat B außerdem die Darstellung $B = \pi_{I'}^{-1}(A')$, so setze $J = I \cup I'$ und beachte

$$\pi_J^{-1}\left(\left(\pi_I^J\right)^{-1}(A)\right) = B = \pi_J^{-1}\left(\left(\pi_{I'}^J\right)^{-1}(A')\right)\ .$$

Da π_J surjektiv ist, folgt daraus $\left(\pi_I^J\right)^{-1}(A) = \left(\pi_{I'}^J\right)^{-1}(A')$, so dass

$$\mu_I(A) \overset{\text{Verträglkt.}}{=} \mu_J\left(\left(\pi_I^J\right)^{-1}(A)\right) = \mu_J\left(\left(\pi_{I'}^J\right)^{-1}(A')\right) = \mu_{I'}(A')\ ,$$

d.h. $\mu(B)$ ist eindeutig definiert.

Seien nun $B = \pi_I^{-1}(A)$, $B' = \pi_{I'}^{-1}(A')$, $B \cap B' = \emptyset$ und $B \cup B' \in \mathcal{Z}$. Wegen der vorhergehenden Überlegung kann man o.B.d.A. $I = I'$ annehmen, so dass insbesondere $A \cap A' = \emptyset$. Daher ist

$$\mu(B \cup B') = \mu\left(\pi_I^{-1}(A \cup A')\right) = \mu_I(A \cup A') = \mu_I(A) + \mu_I(A') = \mu(B) + \mu(B')\ ,$$

d.h. μ ist *additiv* auf $\mathcal{Z}$.

Um den Fortsetzungssatz 3.10 anwenden zu können, der die Fortsetzbarkeit von μ zu einem Wahrscheinlichkeitsmaß auf $\mathcal{B}^T$ garantiert, müssen wir noch die σ-Subadditivität von μ auf $\mathcal{Z}$ nachweisen. Seien also $B = \pi_I^{-1}(A) \in \mathcal{Z}$ und $B_n = \pi_{I_n}^{-1}(A_n) \in \mathcal{Z}$, $B \subseteq \bigcup_{n=1}^{\infty} B_n$. Setze $J_N := I \cup \bigcup_{n=1}^{N} I_n$ für $N \geq 1$ und $J := I \cup \bigcup_{n=1}^{\infty} I_n$. Dann ist

$$\pi_J^{-1}\left((\pi_I^J)^{-1}A\right) = \pi_I^{-1}(A) = B \subseteq \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} \pi_{I_n}^{-1}(A_n) = \pi_J^{-1}\left(\bigcup_{n=1}^{\infty} (\pi_{I_n}^J)^{-1}(A_n)\right)\ ,$$

und da $\pi_J : \mathbb{R}^T \to \mathbb{R}^J$ surjektiv ist, folgt

$$(\pi_I^J)^{-1}A \subseteq \bigcup_{n=1}^{\infty} (\pi_{I_n}^J)^{-1}(A_n)\ .$$

Sei $\epsilon > 0$. Aus der Regularität der Wahrscheinlichkeitsmaße μ_{I_n} auf $\mathbb{R}^{I_n}$ und μ_{J_n} auf $\mathbb{R}^{J_n}$ (Bemerkung 3.16) folgt, dass es kompakte Teilmengen $K_n \subseteq (\pi_I^{J_n})^{-1}(A)$ und offene Mengen $G_n \supseteq A_n$ derart gibt, dass

$$\mu_{J_n}((\pi_I^{J_n})^{-1}A \setminus K_n) < 2^{-n}\epsilon \quad \text{und} \quad \mu_{I_n}(G_n \setminus A_n) < 2^{-n}\epsilon\ .$$

Für $m > 0$ sei $K^m := \bigcap_{n=1}^{m} (\pi_{J_n}^{J_m})^{-1} K_n \subseteq K_m \subseteq (\pi_I^{J_m})^{-1}(A)$. Dann ist K^m kompakt und

$$(\pi_{J_m}^J)^{-1} K^m = \bigcap_{n=1}^{m} (\pi_{J_n}^J)^{-1} K_n \subseteq (\pi_I^J)^{-1} A \subseteq \bigcup_{n=1}^{\infty} (\pi_{I_n}^J)^{-1}(A_n) \subseteq \bigcup_{n=1}^{\infty} (\pi_{I_n}^J)^{-1}(G_n) \qquad (*)$$

und

$$\begin{aligned}
\mu_{J_m}\left((\pi_I^{J_m})^{-1}A \setminus K^m\right) &= \mu_{J_m}\left(\bigcup_{n=1}^{m} (\pi_I^{J_m})^{-1}A \setminus (\pi_{J_n}^{J_m})^{-1}K_n\right) \\
&\leq \sum_{n=1}^{m} \mu_{J_m}\left((\pi_I^{J_m})^{-1}A \setminus (\pi_{J_n}^{J_m})^{-1}K_n\right) \\
&= \sum_{n=1}^{m} \mu_{J_m}\left((\pi_{J_n}^{J_m})^{-1}((\pi_I^{J_n})^{-1}A \setminus K_n)\right) \\
&= \sum_{n=1}^{m} \mu_{J_n}\left((\pi_I^{J_n})^{-1}A \setminus K_n\right) < \epsilon\ .
\end{aligned}$$

Am Ende des Beweises werden wir zeigen, dass die K^m eine gewisse Kompaktheitseigenschaft haben: Aus (*) folgt

$$\exists\, M > 0 :\ (\pi_{J_M}^J)^{-1} K^M \subseteq \bigcup_{n=1}^{M} (\pi_{I_n}^J)^{-1}(G_n) = (\pi_{J_M}^J)^{-1}\left(\bigcup_{n=1}^{M} (\pi_{I_n}^{J_M})^{-1} G_n\right) \qquad (**)$$

Da $\pi_{J_M}^J$ surjektiv ist, folgt daraus

$$K^M \subseteq \bigcup_{n=1}^{M} (\pi_{I_n}^{J_M})^{-1} G_n\ .$$

Da μ_{J_M} ein Maß auf $\mathcal{B}^{J_M}$ ist, folgt

$$\begin{aligned}\mu(B) &= \mu_{J_M}\left((\pi_I^{J_M})^{-1}A\right) \le \mu_{J_M}\left(K^M\right) + \epsilon \le \sum_{n=1}^{M} \mu_{J_M}\left((\pi_{I_n}^{J_M})^{-1}G_n\right) + \epsilon \\ &= \sum_{n=1}^{M} \mu_{I_n}(G_n) + \epsilon \le \sum_{n=1}^{M}\left(\mu_{I_n}(A_n) + 2^{-n}\epsilon\right) + \epsilon \\ &\le \sum_{n=1}^{\infty} \mu(B_n) + 2\epsilon\,.\end{aligned}$$

Im Limes $\epsilon \to 0$ erhält man die σ-Subadditivität von μ auf $\mathcal{Z}$.

Es bleibt zu zeigen, dass (**) aus (*) folgt. Angenommen (**) ist falsch. Dann gibt es zu jedem $m > 0$ einen Punkt

$$x_m \in (\pi_{J_m}^J)^{-1}K^m \setminus \bigcup_{n=1}^{m} (\pi_{I_n}^J)^{-1}(G_n) \subseteq \mathbb{R}^J\,. \tag{***}$$

Insbesondere ist für $m \ge M > 0$

$$\pi_{J_M}^J x_m = \pi_{J_M}^{J_m}(\pi_{J_m}^J x_m) \in \pi_{J_M}^{J_m} K^m \subseteq \bigcap_{n=1}^{M} \pi_{J_M}^{J_m}\left((\pi_{J_M}^{J_m})^{-1}(\pi_{J_n}^{J_M})^{-1}K_n\right) = K^M\,.$$

Da K^M kompakt ist, hat $(\pi_{J_M}^J x_m)_{m\ge M}$ eine in K^M konvergente Teilfolge, und durch Diagonalisierung erhält man eine Teilfolge $(x_{m_j})_{j>0}$ derart, dass

$$y_M := \lim_{j\to\infty} \pi_{J_M}^J(x_{m_j}) \in K^M \subset \mathbb{R}^{J_M} \quad \text{für jedes } M > 0 \text{ existiert.}$$

Sei $M' > M$. Dann ist

$$\pi_{J_M}^{J_{M'}}(y_{M'}) = \lim_{j\to\infty} \pi_{J_M}^{J_{M'}} \pi_{J_{M'}}^J(x_{m_j}) = \lim_{j\to\infty} \pi_{J_M}^J(x_{m_j}) = y_M\,,$$

und da $J = \bigcup_{M=1}^{\infty} J_M$, gibt es einen Punkt $y \in \mathbb{R}^J$ derart, dass $\pi_{J_M}^J(y) = y_M \in K^M$ für alle $M > 0$. Es folgt aus (*), dass

$$\exists\, N > 0:\ \pi_{I_N}^{J_N}(y_N) = \pi_{I_N}^{J_N}(\pi_{J_N}^J y) = \pi_{I_N}^J(y) \in G_N\,.$$

Da G_N offen ist und da $\pi_{I_N}^{J_N}(y_N) = \lim_{j\to\infty} \pi_{I_N}^{J_N}(\pi_{J_N}^J x_{m_j}) = \lim_{j\to\infty} \pi_{I_N}^J x_{m_j}$, ist $\pi_{I_N}^J(x_{m_j}) \in G_N$ für alle hinreichend großen j. Für $m_j \ge N$ steht das aber im Widerspruch zur Wahl von x_{m_j} in (***).
$\square$

12.4 Bemerkung (Satz von Kolmogorov für polnische Räume) Der Satz von Kolmogorov bleibt richtig, wenn $(\mathbb{R}, \mathcal{B})$ durch einen beliebigen *polnischen Raum* X mit Borel-σ-Algebra $\mathcal{B}$ ersetzt wird. (X heißt polnisch, wenn X ein topologischer Raum mit abzählbarer Basis ist, dessen Topologie durch eine vollständige Metrik definiert wird.) Insbesondere ist auch $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$ polnisch. Als Metrik kann $d(x,y) = |\arctan(x) - \arctan(y)|$ dienen, wobei $\arctan(\pm\infty) = \pm\frac{\pi}{2}$.

Nun suchen wir, wie eingangs angekündigt, den Anschluss an Kapitel 11. Zunächst betrachten wir eine Familie $(\Omega_t, \mathcal{A}_t, \mu_t)$ von Wahrscheinlichkeitsräumen. Zu $I \in \mathcal{E}(T)$ sei $\mu_I := \mathsf{X}_{t\in T}\,\mu_t$ ein Produktmaß wie in Bemerkung 11.16a. Die Verträglichkeit des Systems $(\mu_I)_{I\in\mathcal{E}(T)}$ folgt sofort aus dem Satz von Fubini. Daher erhalten wir als Korollar zum Satz von Kolmogorov:

12.5 Korollar (Existenz und Eindeutigkeit von Produktmaßen – unendlich viele Faktoren)
Sei $(\Omega, \mathcal{A}) = (\mathsf{X}_{t\in T}\,\Omega_t, \mathsf{X}_{t\in T}\,\mathcal{A}_t)$. *Dann gibt es genau ein Wahrscheinlichkeitsmaß* $\mu = \mathsf{X}_{t\in T}\,\mu_t$ *auf* $(\Omega, \mathcal{A})$ *derart, dass*

$$\mu\left(\mathop{\mathsf{X}}_{t\in T} A_t\right) = \prod_{t\in T} \mu_t(A_t)$$

wenn $A_t \in \mathcal{A}_t$ $(t \in T)$ *und* $A_t \ne \Omega_t$ *für höchstens endlich viele* t.

12.6 Korollar (Existenz von unabhängigen Prozessen) *Ist $P = \times_{t\in T}\mu_t$ eine Produktwahrscheinlichkeit auf $\Omega = \times_{t\in T}\Omega_t$ und sind $f_t : \Omega_t \to \mathbb{R}$ $\mathcal{A}_t$-messbar, so sind die Zufallsvariablen $X_t := f_t\circ\pi_t : \Omega\to\mathbb{R}$ unabhängig und $P_{X_t} = \mu_t\circ f_t^{-1}$ $(t\in T)$.*

Beweis: Es ist $P_{X_t}(A) = P(\pi_t^{-1}(f_t^{-1}A)) = \mu_t(f_t^{-1}A)$. Zum Beweis der Unabhängigkeit kann man o.B.d.A. annehmen, dass T endlich ist (Satz 5.8a). Aber dann ist

$$P\left(\bigcap_{t\in T}\{X_t\in A_t\}\right) = P\left(\bigcap_{t\in T}\pi_t^{-1}(f_t^{-1}A_t)\right) = \prod_{t\in T}P(\pi_t^{-1}(f_t^{-1}A_t)) = \prod_{t\in T}P\{X_t\in A_t\}\ .$$

□

12.7 Satz (Produktverteilungen und Unabhängigkeit)
Seien $(X_t)_{t\in T}$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$. Der Prozess $(X_t)_{t\in T}$ ist unabhängig genau dann, wenn

$$P_{\mathcal{X}} = \mathop{\times}_{t\in T} P_{X_t}$$

wo, wie üblich, $\mathcal{X}:\Omega\to\bar{\mathbb{R}}^T$, $\mathcal{X}(\omega) = (X_t(\omega))_{t\in T}$.

Beweis: $(X_t)_{t\in T}$ ist unabhängig genau dann, wenn $(X_t)_{t\in I}$ für alle $I\in\mathcal{E}(T)$ unabhängig ist (Satz 5.8a und Definition 5.10). Das ist äquivalent zu $P_{\pi_I\circ\mathcal{X}} = \times_{t\in I}P_{X_t}$ für $I\in\mathcal{E}(T)$ (Satz 11.17), und da $P_{\pi_I\circ\mathcal{X}} = P_{\mathcal{X}}\circ\pi_I^{-1}$, ist das äquivalent zu $P_{\mathcal{X}} = \times_{t\in T}P_{X_t}$. □

Wir betrachten abschließend eine *Folge* messbarer Räume $(\Omega_0,\mathcal{A}_0),(\Omega_1,\mathcal{A}_1),\dots$, ein Wahrscheinlichkeitsmaß P_0 auf $\mathcal{A}_0$ und stochastische Kerne

$$K^j : (\Omega_0\times\dots\times\Omega_{j-1})\times\mathcal{A}_j\to[0,1]\quad (j=0,1,2,\dots).$$

Mit Satz 11.7 definiert man induktiv Wahrscheinlichkeitsmaße

$$Q_n := P_0\times K^1\times\dots\times K^n := \big(\dots\big((P_0\times K^1)\times K^2\big)\times\dots\times K^n\big)$$

auf $\mathcal{A}_0\times\dots\times\mathcal{A}_n$, die durch ihre Werte auf Zylindermengen in $\mathcal{Z}(\mathcal{A}_0,\dots,\mathcal{A}_n)$ eindeutig bestimmt sind. Ebenfalls durch Induktion folgt aus Satz 11.8 (Fubini):

12.8 Satz
Ist $f:\Omega_0\times\dots\times\Omega_n\to\bar{\mathbb{R}}$ $\mathcal{A}_0\times\dots\times\mathcal{A}_n$-messbar und ist $f\geq 0$ oder $f\in\mathcal{L}^1_{Q_n}$, so ist

$$\int f\,dQ_n = \int_{\Omega_0}\int_{\Omega_1}\dots\int_{\Omega_n} f(\omega_0,\dots,\omega_n)\,K^n(\omega_0,\dots,\omega_{n-1},d\omega_n)\ \dots\ K^1(\omega_0,d\omega_1)\,dP_0(\omega_0)\ .$$

Daraus folgt sofort, dass für Funktionen f, die nur von den Variablen $\omega_0,\dots,\omega_{n-1}$ abhängen,

$$\int f\,dQ_n = \int f\,dQ_{n-1}\ . \tag{*}$$

Nach diesen Vorbereitungen können wir ein Wahrscheinlichkeitsmaß $P_0\times\times_{i=1}^\infty K^i$ auf $\times_{i=0}^\infty\mathcal{A}_i$ konstruieren.

12.9 Satz (Ionescu Tulcea)
Sei $(\Omega,\mathcal{A}) = (\times_{i=0}^\infty\Omega_i, \times_{i=0}^\infty\mathcal{A}_i)$ und seien P_0 und die Kerne K^j wie oben. Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $\mathcal{A}$ derart, dass

$$P\left(\mathop{\times}_{i=0}^{\infty}A_i\right) = \int_{A_0}\int_{A_1}\dots\int_{A_n} K^n(\omega_0,\dots,\omega_{n-1},d\omega_n)\ \dots\ K^1(\omega_0,d\omega_1)\,dP_0(\omega_0)$$

für $A_i\in\mathcal{A}_i$ $(i\in\mathbb{N}_0)$ und $A_i=\Omega_i$ $(i>n)$.

Beweis: Sei Q_n auf $\mathcal{A}_0 \times \cdots \times \mathcal{A}_n$ definiert wie in Satz 12.8. Für $I \in \mathcal{E}(\mathbb{N}_0)$ wähle $n \in \mathbb{N}_0$ so groß, dass $I \subseteq \{0, \ldots, n\}$, und setze

$$\mu_I := Q_n \circ (\pi_I^{\{0,\ldots,n\}})^{-1}, \quad \text{insbesondere } \mu_{\{0,\ldots,n\}} = Q_n\ .$$

Wegen (*) aus Satz 12.8 ist diese Festlegung von der Wahl von n unabhängig, und daher gilt

$$\mu_I = \mu_J \circ (\pi_I^J)^{-1} \quad \text{für } I \subseteq J \in \mathcal{E}(\mathbb{N}_0)\ .$$

Ist $(\Omega, \mathcal{A})$ polnisch (siehe Bemerkung 12.4), so folgt die Existenz von μ aus Satz 12.3 (Kolmogorov). Andernfalls muss ein auf C. Ionescu Tulcea zurückgehender, recht aufwendiger Beweis geführt werden (siehe z.B. [Gänssler-Stute, Satz 1.9.3]), den wir hier aber nicht präsentieren. □

12.10 Bemerkung
Hängt $K^n(\omega_0, \ldots, \omega_{n-1}, \boldsymbol{.}) = P_n(\boldsymbol{.})$ nicht von $\omega_0, \ldots, \omega_{n-1}$ ab, so stimmt das Maß P aus dem vorigen Satz mit dem Maß $\bigtimes_{n=0}^{\infty} P_n$ im Sinne von Korollar 12.5 überein.

Kapitel 13

Gibbs Verteilungen

Bevor wir uns Fragen der Verteilungskonvergenz und dem Zentralen Grenzwertsatz zuwenden, wollen wir uns in diesem Kapitel Gedanken darüber machen, was es eigentlich heißt, dass ein Ereignis "zufällig" ist. Unsere Überlegungen beruhen auf Vorstellungen der klassischen Stochastik – "möglichst zufällig" wird quantifiziert durch den Begriff der *Entropie*.

Sei X eine Zufallsvariable mit Werten im messbaren Raum $(M, \mathcal{M})$ und Verteilung μ. Wir möchten, dass X die Realisierung des "totalen Zufalls" in M beschreibt, d.h. μ repräsentiere die "totale Unwissenheit" über zufällige Realisierungen in M. Wie ein solches μ in konkreten Modellen auszusehen hat, ist eine Frage, die keine eindeutige mathematische Antwort hat. Ist M endlich, so wird man in der Regel die Gleichverteilung auf M für μ wählen, ist $M \subset \mathbb{R}^d$ mit endlichem d-dimensionalen Volumen, so wird μ oft das normalisierte Lebesgue-Maß auf M sein. Ist aber z.B. $M = \mathbb{R}$, so gibt es keine allgemein akzeptierte Wahl von μ.

Haben wir uns aber darauf geeinigt, dass μ die totale Unwissenheit über M repräsentiert, so wird durch jede Wahrscheinlichkeitsdichte f auf $(M, \mathcal{M}, \mu)$ ein Vorwissen über zu erwartende Realisierungen von nach $f\mu$ verteilten Zufallsvariablen repräsentiert. Dieses Wissen wird quantifiziert durch die *Kullback-Leibler Information* (auch *relative Entropie*) von f bzgl. μ,

$$I_\mu(f) := \int f \cdot \log f \, d\mu = \int \varphi \circ f \, d\mu \,, \quad \text{wo } \varphi : [0, \infty) \to \mathbb{R}, \varphi(x) = x \log x \,.$$

Da $\varphi'(x) = 1 + \log x$, $\varphi''(x) = \frac{1}{x} > 0$, ist φ strikt konvex. Man überzeugt sich leicht, dass $\varphi(0) = \lim_{x \downarrow 0} \varphi(x) = 0$, $\varphi(1) = 0$ und dass $\varphi(\frac{1}{e}) = -\frac{1}{e}$ das Minimum von φ ist. Aus der Jensen'schen Ungleichung folgt

13.1 Satz
$I_\mu(f) \geq \varphi\left(\int f \, d\mu\right) = \varphi(1) = 0$ *mit Gleichheit genau dann, wenn* $f = \int f \, d\mu = 1$ *f.s.*

Die Dichte 1, d.h. μ selbst, hat also Informationswert 0 bzgl. μ, während jede andere Dichte echt positiven Informationswert bzgl. μ besitzt.

Der nächste Satz besagt, dass *der Informationswert zweier unabhängiger Beobachtungen gleich der Summe der Informationswerte der einzelnen Beobachtungen ist.* Dieses ist die wichtigste Plausibilitätsforderung an einen quantitativen Informationsbegriff, und $I_\mu(\boldsymbol{.})$ ist dadurch im Wesentlichen eindeutig bis auf Normierung festgelegt.

13.2 Satz
Seien X und Y zwei Zufallsvariablen mit Werten in $(M_1, \mathcal{M}_1)$ bzw. $(M_2, \mathcal{M}_2)$ und Verteilungen $P_X = f\mu_1$, $P_Y = g\mu_2$, und sei $P_{(X,Y)} = h \cdot (\mu_1 \times \mu_2)$. Dann ist

$$I_{\mu_1\times\mu_2}(h) \geq I_{\mu_1}(f) + I_{\mu_2}(g)$$

mit Gleichheit genau dann, wenn X und Y unabhängig sind.

Beweis: Da

$$\int_A f(x)\,d\mu_1(x) = P\{X \in A\} = P\{(X,Y) \in A \times M_2\} = \int_A \left(\int_{M_2} h(x,y)\,d\mu_2(y)\right)\,d\mu_1(x)$$

für alle $A \in \mathcal{M}_1$, ist $f(x) = \int_{M_2} h(x,y)\,d\mu_2(y)$ für μ_1-f.a. $x \in M_1$. Ist $f(x) = 0$ für ein solches $x \in M_1$, so folgt $h(x,y) = 0$ für dieses x und μ_2-f.a. $y \in M_2$. Daher ist

$$\int_{\{f=0\}} \left(\int_{M_2} \varphi(h(x,y))\,d\mu_2(y)\right)\,d\mu_1(x) = 0 = \int_{\{f=0\}} \varphi(f(x))\,d\mu_1(x)\ ,$$

und wir können o.B.d.A. $f > 0$ annehmen. Mit ähnlicher Begründung kann man auch $g > 0$ annehmen.

Für $x \in M_1$ definiere

$$g_x : M_2 \to [0,+\infty], \quad g_x(y) := \frac{h(x,y)}{f(x)} \qquad \text{(“bedingte Dichte gegeben } x\text{”)}$$

Dann ist jedes g_x eine Wahrscheinlichkeitsdichte auf $(M_2, \mathcal{M}_2, \mu_2)$, und es gilt (Satz 11.12)

$$\begin{aligned}
I_{\mu_1\times\mu_2}(h) &= \int\int f(x)g_x(y)\log\left(f(x)g_x(y)\right)\,d\mu_1(x)\,d\mu_2(y) \\
&= \int \underbrace{\left(\int g_x(y)\,d\mu_2(y)\right)}_{=1} f(x)\log f(x)\,d\mu_1(x) \\
&\qquad + \int\int f(x)g_x(y)\left(\log g(y) + \log\frac{g_x(y)}{g(y)}\right)\,d\mu_1(x)\,d\mu_2(y) \\
&= I_{\mu_1}(f) + \int \underbrace{\left(\int h(x,y)\,d\mu_1(x)\right)}_{=g(y)} \log g(y)\,d\mu_2(y) \\
&\qquad + \int\int \varphi\left(\frac{g_x(y)}{g(y)}\right)\,d(f\mu_1)(x)\,d(g\mu_2)(y) \\
&\overset{\text{Jensen}}{\geq} I_{\mu_1}(f) + I_{\mu_2}(g) + \underbrace{\varphi\left(\int\int f(x)g(y)\frac{g_x(y)}{g(y)}\,d\mu_1(x)\,d\mu_2(y)\right)}_{=\varphi(\int\int h(x,y)\,d\mu_1(x)\,d\mu_2(y))=\varphi(1)=0}
\end{aligned}$$

mit Gleichheit genau dann, wenn $(f\mu_1 \times g\mu_2)$-f.s. $\frac{g_x(y)}{g(y)}$ konstant ist. Da g und alle g_x Wahrscheinlichkeitsdichten sind, ist das genau dann der Fall, wenn $h(x,y) = f(x)g_x(y) = f(x)g(y)$ f.s., d.h. genau dann, wenn X und Y unabhängig sind. □

Seien nun $\phi_1, \ldots, \phi_n : M \to \bar{\mathbb{R}}$ *Observablen*, d.h. $\mathcal{M}$-messbare Funktionen mit denen Zufallsrealisierungen in M “ausgewertet” werden können. Wir setzen voraus, dass alle ϕ_i μ-f.s. nur endliche Werte annehmen. Seien $m_i \in \mathbb{R}$ denkbare Erwartungswerte der ϕ_i $(i = 1, \ldots, n)$, $m := (m_1, \ldots, m_n)$, und bezeichne

$$\Phi_m := \left\{f : f \text{ Wahrscheinlichkeitsdichte bzgl. } \mu,\ \phi_i \in \mathcal{L}^1_{f\mu},\ E_{f\mu}[\phi_i] = m_i\ (i = 1, \ldots, n)\right\}\ .$$

Gesucht ist eine (die?) Wahrscheinlichkeitsdichte f, die nur das Wissen $E_{f\mu}[\phi_i] = m_i$ $(i = 1, \ldots, n)$ repräsentiert, ansonsten aber keine Information trägt, also das angemessene Modell für zufällige Realisierungen unter den Nebenbedingungen $E_{f\mu}[\phi_i] = m_i$ ist. Formal ist gesucht:

$$f \in \Phi_m \text{ mit } I_\mu(f) = \inf\{I_\mu(g) : g \in \Phi_m\}$$

Im Folgenden werden wir die gesuchten f unter zusätzlichen Integrabilitätsannahmen an die ϕ_i charakterisieren – sogar in der allgemeineren Situation, wo μ ein beliebiges σ-endliches Maß sein darf, z.B. das Lebesgue-Maß auf $\mathbb{R}$. Es wird sich herausstellen, dass sie von der Form

$$f_\vartheta = \exp(-\gamma + \sum_{i=1}^{n} \vartheta_i \phi_i) \quad \text{für ein } \vartheta = (\vartheta_1, \ldots, \vartheta_n) \in \mathbb{R}^n$$

sind, wobei die Konstante $\gamma = \gamma(\vartheta)$ durch die Normierung $\int f_\vartheta \, d\mu = 1$ bestimmt wird, also $\gamma(\vartheta) = \log \int e^{\sum_{i=1}^n \vartheta_i \phi_i} \, d\mu$. Verteilungen, die durch solche f_ϑ gegeben sind, heißen *Gibbs Verteilungen*. Wir werden Dichten der Form f_ϑ näher untersuchen. Sie spielen auch in der mathematischen Statistik eine wichtige Rolle. Dort heißen sie *exponentielle Familien*.

13.3 Beispiele a) Die Bernoulli-Maße mit Parameter $p \in (0, 1)$ auf $M = \{0, 1\}^n$ schreibt man folgendermaßen als exponentielle Familie: Sei μ das Zählmaß auf Ω, $\phi(\omega) = \sum_{k=1}^n \omega_k$, $\vartheta = \log \frac{p}{1-p}$. Dann ist

$$f_\vartheta(\omega) = e^{-\gamma} \, e^{\log \frac{p}{1-p} \cdot \phi(\omega)} = e^{-\gamma} \, p^{\phi(\omega)} \, (1-p)^{-\phi(\omega)} = e^{-\gamma}(1-p)^{-n} \cdot p^{\phi(\omega)}(1-p)^{n-\phi(\omega)} \ ;$$

das ist die Dichte des Bernoulli-Maßes zum Zählmaß, wenn man γ so wählt, dass $e^{-\gamma}(1-p)^n = 1$.

b) Die Normalverteilungsdichten mit Erwartungswert t und Varianz σ^2 schreibt man folgendermaßen als exponentielle Familie: Sei μ das Lebesgue-Maß auf $\mathbb{R}$, $\phi_1(x) = x$, $\phi_2(x) = x^2$, $\vartheta_1 = \frac{t}{\sigma^2}$, $\vartheta_2 = -\frac{1}{2\sigma^2}$. Dann ist

$$f_\vartheta(x) = e^{-\gamma} \, \exp\left(\frac{t}{\sigma^2}x - \frac{1}{2\sigma^2}x^2\right) = \exp\left(-\gamma + \frac{t^2}{2\sigma^2}\right) \, \exp\left(-\frac{(x-t)^2}{2\sigma^2}\right) \ ,$$

wobei γ so zu wählen ist, dass $\exp(\gamma - \frac{t^2}{2\sigma^2}) = \sqrt{2\pi\sigma^2}$.

13.4 Satz (Momente von Gibbs Verteilungen bzw. exponentiellen Familien)
Seien $\phi_1, \dots, \phi_n$ Observablen, μ ein σ-endliches Maß auf $(M, \mathcal{M})$. Bezeichne

$$\begin{aligned}\Theta &:= \left\{ \vartheta \in \mathbb{R}^n : 0 < \int \exp\left(\sum_{i=1}^n \vartheta_i \phi_i\right) d\mu < \infty \right\} \\ &= \{\vartheta \in \mathbb{R}^n : \ f_\vartheta \text{ existiert als Wahrscheinlichkeitsdichte}\}\end{aligned}$$

Für $\vartheta \in \Theta$ mögen E_ϑ und Cov_ϑ Erwartungswert bzw. Kovarianz bzgl. des Wahrscheinlichkeitsmaßes $f_\vartheta \mu$ bezeichnen. Dann ist Θ konvex, und es gilt

a) Für $j = 1, \dots, n$ und $\vartheta \in \mathring{\Theta}$ ist

$$\frac{\partial}{\partial \vartheta_j} \gamma(\vartheta) = \int_M \phi_j f_\vartheta \, d\mu = E_\vartheta[\phi_j].$$

b) Für $i, j = 1, \dots, n$ und $\vartheta \in \mathring{\Theta}$ ist

$$\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \gamma(\vartheta) = -E_\vartheta[\phi_i] E_\vartheta[\phi_j] + E_\vartheta[\phi_i \phi_j] = Cov_\vartheta(\phi_i, \phi_j) \ .$$

$(Cov_\vartheta(\phi_i, \phi_j))_{ij}$ ist eine positiv semidefinite Matrix. Sie ist positiv definit genau dann, wenn die Familie $\{1, \phi_1, \dots, \phi_n\}$ im Raum der μ-Äquivalenzklassen messbarer Funktionen linear unabhängig ist. (Aus dieser Bedingung sieht man, dass die Matrix entweder für alle ϑ oder für keines positiv definit ist.)

Die Funktion $\vartheta \mapsto \gamma(\vartheta)$ ist sogar unendlich oft differenzierbar in $\mathring{\Theta}$.

Beweis: Konvexität von Θ: Seien $\vartheta, \tilde{\vartheta} \in \Theta$, $0 < \alpha < 1$. Dann ist, wegen der Konvexität der e-Funktion und der Linearität des Integrals,

$$\int \exp\left(\sum_{i=1}^n (\alpha\vartheta_i + (1-\alpha)\tilde{\vartheta}_i)\phi_i\right) d\mu \le \alpha \int \exp\left(\sum_{i=1}^n \vartheta_i \phi_i\right) d\mu + (1-\alpha) \int \exp\left(\sum_{i=1}^n \tilde{\vartheta}_i \phi_i\right) d\mu < \infty \ .$$

Wäre dieses Integral $= 0$, dann wäre $\sum_{i=1}^n (\alpha\vartheta_i + (1-\alpha)\tilde{\vartheta}_i)\phi_i = -\infty$ μ-f.s. im Widerspruch zu der Voraussetzung, dass die ϕ_i μ-f.s. $\mathbb{R}$-wertig sind.

a) Da
$$\gamma(\vartheta) = \log \int_M \exp\left(\sum_{i=1}^n \vartheta_i \phi_i\right) d\mu,$$
folgt aus Satz 8.12, dass
$$\begin{aligned}\frac{\partial}{\partial \vartheta_j} \gamma(\vartheta) &= e^{-\gamma(\vartheta)} \cdot \frac{\partial}{\partial \vartheta_j} \int_M \exp\left(\sum_{i=1}^n \vartheta_i \phi_i\right) d\mu = e^{-\gamma(\vartheta)} \cdot \int_M \phi_j \exp\left(\sum_{i=1}^n \vartheta_i \phi_i\right) d\mu \\ &= \int_M \phi_j f_\vartheta \, d\mu = E_\vartheta[\phi_j]\end{aligned}$$
sobald wir die Integrabilitätsbedingung von Satz 8.12 nachgewiesen haben: Sei dazu $\vartheta \in \Theta$. Es gibt $\eta > 0$ derart, dass $\vartheta + te_j \in \Theta$ für alle $t \in \mathbb{R}$ mit $|t| \le \eta$. Sei $\vartheta^t := \vartheta + te_j$. Dann ist
$$\phi_j \, e^{\sum_i \vartheta_i^t \phi_i} = \frac{1}{t - (\pm\eta)} \cdot e^{\sum_i \vartheta_i^{\pm\eta} \phi_i} \cdot (t - (\pm\eta))\phi_j e^{(t-(\pm\eta))\phi_j} \ .$$

Betrachte nun zunächst den Fall $\phi_j(\omega) \geq 0$. Dann wenden wir diese Zerlegung für $+\eta$ an und erhalten

$$\left|\frac{\partial}{\partial\vartheta_j} e^{\sum_i \vartheta_i\phi_i(\omega)}\right|_{|\vartheta=\vartheta^t} = \left|\phi_j(\omega)\, e^{\sum_i \vartheta_i^t\phi_i(\omega)}\right| \leq \frac{2}{\eta}\cdot e^{\sum_i \vartheta_i^\eta\phi_i(\omega)} \cdot \underbrace{\left|(t-\eta)\phi_j(\omega)\, e^{(t-\eta)\phi_j(\omega)}\right|}_{\leq e^{-1}}$$

für alle t mit $|t| < \frac{\eta}{2}$, denn $\max_{x\leq 0}|xe^x| = \max_{x\geq 0} xe^{-x} = e^{-1}$. Ist $\phi_j(\omega) \leq 0$, so wenden wir die gleiche Überlegung für $-\eta$ an und erhalten schließlich für alle $|t| < \frac{\eta}{2}$

$$\left|\frac{\partial}{\partial\vartheta_j} e^{\sum_i \vartheta_i\phi_i}\right|_{|\vartheta=\vartheta^t} \leq \frac{2}{e\eta}\max\left\{e^{\sum_i \vartheta_i^{+\eta}\phi_i(\omega)}, e^{\sum_i \vartheta_i^{-\eta}\phi_i(\omega)}\right\} \in L^1_\mu\,.$$

b) Unter Berücksichtigung der Formel für $\frac{\partial}{\partial\vartheta_j}\gamma(\vartheta)$ folgt ähnlich wie im vorherigen Teil (aber mit etwas mehr Aufwand bei der Überprüfung der Integrabilitätsvoraussetzung von Satz 8.12):

$$\begin{aligned}\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\gamma(\vartheta) &= \frac{\partial}{\partial\vartheta_i}e^{-\gamma(\vartheta)}\cdot e^{\gamma(\vartheta)}E_\vartheta[\phi_j] + e^{-\gamma(\vartheta)}\cdot\frac{\partial}{\partial\vartheta_i}\underbrace{\left(e^{\gamma(\vartheta)}E_\vartheta[\phi_j]\right)}_{=\int \phi_j \exp(\sum_k \vartheta_k\phi_k)\,d\mu} \\ &= -E_\vartheta[\phi_i]E_\vartheta[\phi_j] + E_\vartheta[\phi_i\phi_j] = Cov_\vartheta(\phi_i,\phi_j).\end{aligned}$$

Diese Matrix der 2. Ableitungen ist positiv semidefinit, denn für $\lambda \in \mathbb{R}^n$ ist

$$\sum_{i=1}^n\sum_{j=1}^n \lambda_i\, Cov_\vartheta(\phi_i,\phi_j)\lambda_j = Cov_\vartheta\left(\sum_{i=1}^n\lambda_i\phi_i, \sum_{j=1}^n\lambda_j\phi_j\right) = V_\vartheta\left(\sum_{i=1}^n\lambda_i\phi_i\right) \geq 0$$

mit Gleichheit genau dann, wenn $\sum_{i=1}^n \lambda_i\phi_i$ $f_\vartheta\mu$-f.s. konstant ist. Da $f_\vartheta > 0$ ist das genau dann der Fall, wenn die Familie $\{1,\phi_1,\ldots,\phi_n\}$ im Raum der μ-Äquivalenzklassen messbarer Funktionen linear abhängig ist.

Ähnlich zeigt man die Existenz der höheren Ableitungen von γ. □

13.5 Korollar *Sei $n = 1$ und μ ein Wahrscheinlichkeitsmaß in der Situation des vorherigen Satzes und $0 \in \mathring{\Theta}$.*

a) Im Limes $\vartheta \to 0$ gilt

$$\gamma(\vartheta) = E_0[\phi_1]\vartheta + \frac{1}{2}V_0(\phi_1)\vartheta^2 + o(\vartheta^2) \quad \textit{und} \quad E_\vartheta[\phi_1] = E_0[\phi_1] + V_0(\phi_1)\vartheta + o(\vartheta)\,.$$

b) Ist ϕ_1 nicht μ-f.s. konstant, so ist $\mathring{\Theta}$ ein nicht leeres, offenes Intervall, und es gibt $\alpha^-,\alpha^+ > 0$ derart, dass die Abbildung

$$E_\cdot[\phi_1] : \mathring{\Theta} \to (E_0[\phi_1]-\alpha^-, E_0[\phi_1]+\alpha^+), \quad \vartheta \mapsto E_\vartheta[\phi_1]$$

ein Diffeomorphismus mit Ableitung $\frac{d}{d\vartheta}E_\vartheta[\phi_1] = V_\vartheta(\phi_1) > 0$ ist.

Beweis:

a) Da $f_0 = \exp(-\gamma(0))$ eine Wahrscheinlichkeitsdichte ist, ist $\gamma(0) = 0$. Der Rest folgt aus dem vorherigen Satz.

b) Ist ϕ_1 nicht μ-f.s. konstant, so ist $V_\vartheta(\phi_1) > 0$ für alle $\vartheta \in \Theta$ (Lemma 7.15b). Aus dem vorherigen Satz folgt, dass γ in Θ zweimal differenzierbar und damit $E_.[\phi_1] = \gamma'$ einmal differenzierbar ist mit Ableitung $\frac{\partial}{\partial\vartheta} E_\vartheta[\phi_1] = \gamma''(\vartheta) = V_\vartheta(\phi_1) > 0$. Daraus folgt die Behauptung mit dem aus der Analysis bekannten Satz über die Differenzierbarkeit von Umkehrfunktionen.

□

13.6 Bemerkung Ein multivariates ($n > 1$) Analogon zu Korollar 13.5 gilt ebenfalls.

Die Aussagen aus 13.4 – 13.6 legen nahe, dass das Gleichungssystem $E_\vartheta[\phi_i] = m_i$ $(i = 1, \ldots, n)$ in vielen Fällen eine eindeutige Lösung $\vartheta(m)$ hat. Genauer: Sei

$$\Lambda := \left\{ (E_\vartheta[\phi_1], \ldots, E_\vartheta[\phi_n]) : \vartheta \in \mathring{\Theta} \right\} .$$

Ist die Familie $\{1, \phi_1, \ldots, \phi_n\}$ im Raum der μ-Äquivalenzklassen messbarer Funktionen linear unabhängig, so ist $\Lambda \subseteq \mathbb{R}^n$ offen, denn die Abbildung $D\gamma : \mathring{\Theta} \to \Lambda$, $D\gamma(\vartheta) = (E_\vartheta[\phi_1], \ldots, E_\vartheta[\phi_n])$ ist nach Satz 13.4b) ein lokaler Diffeomorphismus.

13.7 Satz (Informationsminimierung unter Nebenbedingungen, Gibbs Verteilungen)
Sei μ ein σ-endliches Maß auf $(M, \mathcal{M})$. Ist $m = D\gamma(\vartheta) \in \Lambda$, so ist $f_\vartheta \in \Phi_m$,

$$I_\mu(f_\vartheta) = \inf\{I_\mu(g) : g \in \Phi_m\}$$

und $I_\mu(f_\vartheta) = I_\mu(g)$ genau dann, wenn $g = f_\vartheta$ μ-f.s.

Beweis: Sei $g \in \Phi_m$. Da $m = D\gamma(\vartheta) = (E_\vartheta[\phi_1], \ldots, E_\vartheta[\phi_n])$, ist nach Satz 13.4a) auch $f_\vartheta \in \Phi_m$. Unter Berücksichtigung von $\log f_\vartheta = -\gamma + \sum_{i=1}^n \vartheta_i \phi_i \in \mathcal{L}^1_{g\mu}$ folgt

$$\begin{aligned} I_\mu(g) - I_\mu(f_\vartheta) &= \int g \log g \, d\mu - \int f_\vartheta \log f_\vartheta \, d\mu = \int (g - f_\vartheta) \log f_\vartheta \, d\mu + \int g \log \frac{g}{f_\vartheta} \, d\mu \\ &= \int (g - f_\vartheta) \cdot \left(-\gamma + \sum_{i=1}^n \vartheta_i \phi_i \right) d\mu + \int g \cdot \left(-\log \frac{f_\vartheta}{g} \right) d\mu \\ &= 0 + E_{g\mu}\left[-\log \frac{f_\vartheta}{g} \right] \geq -\log E_{g\mu}\left[\frac{f_\vartheta}{g} \right] = -\log \int f_\vartheta \, d\mu = 0 \ , \end{aligned}$$

wobei in der letzten Zeile die Jensensche Ungleichung für das Wahrscheinlichkeitsmaß $g\mu$ benutzt wurde. Dabei tritt Gleichheit auf genau dann, wenn $\frac{f_\vartheta}{g}$ $g\mu$-f.s. konstant ist, also wenn $g = f_\vartheta$ μ-f.s. □

13.8 Beispiel (Bernoulliverteilung) Sei $M = \{0,1\}^N$, μ das Zählmaß auf M und $\phi(\omega) = \sum_{k=1}^N \omega_k$, siehe Beispiel 13.3. Sei $m = pN$ für ein $p \in (0,1)$. Dann minimiert die Dichte der Bernoulliverteilung zum Parameter p, d.h. $f(\omega) = p^{\phi(\omega)}(1-p)^{n-\phi(\omega)}$, das Funktional $I_\mu(f)$ unter der Nebenbedingung $\int \phi f \, d\mu = m$.

Das folgt so:

$$\gamma(\vartheta) = \log \int e^{\vartheta\phi} \, d\mu = \log \sum_{k=0}^N \binom{N}{k} e^{\vartheta k} = \log(1 + e^\vartheta)^N = N \log(1 + e^\vartheta) \ ,$$

also

$$\Theta = \mathbb{R} \ , \quad D\gamma(\vartheta) = N \frac{e^\vartheta}{1 + e^\vartheta} \quad \text{und} \quad \Lambda = D\gamma(\mathring{\Theta}) = (0, N) \ .$$

Daher ist

$$D\gamma(\vartheta) = m = pn \Longleftrightarrow \frac{e^\vartheta}{1+e^\vartheta} = p \Longleftrightarrow \frac{1}{1+e^\vartheta} = 1-p \Longleftrightarrow e^\vartheta = \frac{p}{1-p} \Longleftrightarrow \vartheta = \log\frac{p}{1-p} \ ,$$

so dass wir als minimierende Dichte erhalten:

$$f(\omega) = f_\vartheta(\omega) = e^{-\gamma(\vartheta)+\vartheta\phi(\omega)} = \left(\frac{1}{1+e^\vartheta}\right)^N e^{\vartheta\phi(\omega)} = p^{\phi(\omega)}(1-p)^{N-\phi(\omega)}$$

13.9 Beispiel (Normalverteilung) Zur Erinnerung: Die Dichte der Normalverteilung $\mathcal{N}(t,\sigma^2)$ relativ zum Lebesgue-Maß λ auf $\mathbb{R}$ ist

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

Um diese Dichte im Sinne von Satz 13.7 zu charakterisieren, bestimmen wir zunächst $I_\lambda(h_{t,\sigma^2})$:

$$\begin{aligned} I_\lambda(h_{t,\sigma^2}) &= \int \log(h_{t,\sigma^2}(x))\, h_{t,\sigma^2}(x)\, dx = -\frac{1}{2}\log(2\pi\sigma^2) - \int \frac{(x-t)^2}{2\sigma^2} h_{t,\sigma^2}(x)\, dx \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2} \ . \end{aligned}$$

Da diese Größe 1) nicht vom Mittelwert t abhängt und 2) im Limes $\sigma^2 \to \infty$ gegen $-\infty$ strebt, wird man zumindest das erste und zweite Moment der Verteilungen vorgeben müssen, um eine Charakterisierung im Sinne von Satz 13.7 zu erhalten. Seien also

$$\phi_1(x) = x, \quad \phi_2(x) = x^2 \ .$$

Da

$$\exp(\vartheta_1\phi_1(x) + \vartheta_2\phi_2(x)) = \exp\left(\vartheta_2(x + \frac{\vartheta_1}{2\vartheta_2})^2 - \frac{\vartheta_1^2}{4\vartheta_2}\right) = \exp\left(-\frac{\vartheta_1^2}{4\vartheta_2}\right)\exp\left(-\frac{(x-(-\frac{\vartheta_1}{2\vartheta_2}))^2}{2\frac{-1}{2\vartheta_2}}\right)$$

bis auf die Normierung die Dichte einer Normalverteilung mit Mittelwert $t = -\frac{\vartheta_1}{2\vartheta_2}$ und Varianz $\sigma^2 = -\frac{1}{2\vartheta_2}$ ist, folgt sofort dass $0 < \int \exp(\vartheta_1\phi_1 + \vartheta_2\phi_2)\, d\lambda < \infty$ und $f_\vartheta = h_{t,\sigma^2}$, falls $\vartheta = (\vartheta_1,\vartheta_2) \in \Theta := \mathbb{R}\times(-\infty,0)$. Nach Satz 13.4a) ist $D\gamma(\vartheta) = (E_\vartheta[\phi_1], E_\vartheta[\phi_2]) = (t, t^2+\sigma^2)$, und es folgt $\Lambda = D\gamma(\Theta) = \{(m_1,m_2)\in\mathbb{R}^2 : m_2 > m_1^2\}$.

Ist nun $g \in \Phi_m$, so ist wegen der Jensenschen Ungleichung immer

$$m_2 = \int x^2 g(x)\, dx > \left(\int x g(x)\, dx\right)^2 = m_1^2 \ ,$$

also $m \in \Lambda$. Mit $(t,\sigma^2) := (m_1, m_2 - m_1^2) \in \mathbb{R}\times(0,\infty)$ folgt daher aus Satz 13.7

$$I_\lambda(g) \geq I_\lambda(f_\vartheta) = I_\lambda(h_{t,\sigma^2}) \quad \text{für alle } g \in \Phi_m$$

mit Gleichheit genau dann wenn $g = h_{t,\sigma^2}$.

13.10 Übung Charakterisieren Sie auf ähnliche Weise die *Exponentialverteilungen* auf $[0,\infty)$, die die Dichten $f_\alpha(x) = \alpha e^{-\alpha x}$ haben, $\alpha > 0$. Dazu muss man nur $m_1 = \int_0^\infty f_\vartheta(x)\, dx > 0$ vorgeben.

Dichten f_ϑ des Typs, wie sie in Satz 13.7 gefundenen wurden, spielen in der statistischen Mechanik und auch in der klassischen Statistik (dort unter dem Namen *exponentielle Familien*) eine wichtige Rolle. Man benutzt sie, um die größtmögliche Unsicherheit bei der Realisierung von Zufallsvariablen X zu modellieren, wenn Erwartungswerte $E[\phi_i(X)]$ für gewisse Observablen ϕ_i bekannt sind.

Kapitel 14

Große Abweichungen

In diesem Kapitel zeigen wir, wie die *Theorie der großen Abweichungen* für Partialsummen unabhängiger identisch verteilter Zufallsvariablen mit den Überlegungen zu Gibbs-Verteilungen aus dem letzten Kapitel zusammenhängt. Sei X eine reelle Zufallsvariable mit Verteilung μ. Die Abbildung

$$\gamma_\mu : \mathbb{R} \to (-\infty, +\infty], \quad \gamma_\mu(\vartheta) := \log E[e^{\vartheta X}] = \log \int e^{\vartheta x}\, d\mu(x)$$

heißt die *logarithmische Laplace-Transformierte* von X. (Sie hängt nur von der Verteilung μ von X ab!) Die Menge $\Theta_\mu := \{\vartheta \in \mathbb{R} : \gamma_\mu(\vartheta) < \infty\}$ ist ein Intervall (Satz 13.4). Für $\vartheta \in \Theta_\mu$ sei $f_\vartheta(x) = e^{-\gamma_\mu(\vartheta)+\vartheta x}$. Dann haben wir die Situation von Satz 13.7 mit $n = 1$ und $\phi_1(x) = x$. Das dortige $\gamma(\vartheta)$ stimmt mit $\gamma_\mu(\vartheta)$ überein. Zur Abkürzung schreiben wir E_ϑ statt $E_\vartheta[\phi_1]$ und V_ϑ statt $V_\vartheta(\phi_1)$.

14.1 Satz (Große Abweichungen für u.i.v. Zufallsvariablen)
Seien $X_1, X_2, \ldots$ unabhängig identisch verteilt mit Verteilung μ und sei $0 \in \mathring{\Theta}_\mu$. Seien $\alpha^+, \alpha^-, > 0$ wie in Korollar 13.5b. Sei $c \in (E_0, E_0 + \alpha^+)$, also $c = E_\vartheta$ für ein $\vartheta \in \Theta_\mu \cap (0, \infty)$. Dann ist für dieses ϑ

$$\lim_{n\to\infty} \frac{1}{n} \log P\left\{\frac{1}{n}\sum_{i=1}^n X_i > c\right\} = \gamma_\mu(\vartheta) - \vartheta c = -I_\mu(f_\vartheta)\ . \qquad (*)$$

Entsprechend gilt für $c = E_\vartheta \in (E_0 - \alpha^-, E_0)$: $\vartheta \in \Theta_\mu \cap (-\delta, \infty)$ und

$$\lim_{n\to\infty} \frac{1}{n} \log P\left\{\frac{1}{n}\sum_{i=1}^n X_i < c\right\} = \gamma_\mu(\vartheta) - \vartheta c = -I_\mu(f_\vartheta)\ .$$

Außerdem ist $I_\mu(f_\vartheta) = \frac{\vartheta^2}{2}V_0 + o(\vartheta^2) = \frac{(c-E_0)^2}{2V_0} + o\left((c-E_0)^2\right)$ im Limes $\vartheta \to 0$, d.h. $c \to E_0$.

Beweis: Beachte zunächst, dass

$$I_\mu(f_\vartheta) = \int f_\vartheta(x) \cdot (-\gamma_\mu(\vartheta) + \vartheta x)\, d\mu(x) = -\gamma_\mu(\vartheta) + \vartheta E_\vartheta = -\gamma_\mu(\vartheta) + \vartheta c\ .$$

Da $c = E_\vartheta = E_0 + V_0\vartheta + o(\vartheta)$, folgt aus Korollar 13.5 außerdem

$$\begin{aligned} I_\mu(f_\vartheta) &= -E_0\vartheta - \frac{1}{2}V_0\vartheta^2 + o(\vartheta^2) + \vartheta E_0 + \vartheta V_0 \vartheta + o(\vartheta^2) \\ &= \frac{\vartheta^2}{2}V_0 + o(\vartheta^2) = \frac{(c-E_0)^2}{2V_0} + o\left((c-E_0)^2\right)\ . \end{aligned}$$

Wir betrachten nun den Fall $c \in (E_0, E_0 + \alpha^+)$, d.h. $\vartheta > 0$. Der Beweis im anderen Fall verläuft analog.

$$\begin{aligned}\frac{1}{n}\log P\left\{\frac{1}{n}\sum_{i=1}^n X_i > c\right\} &= \frac{1}{n}\log P\left\{\exp\left(\vartheta\sum_{i=1}^n X_i - n\vartheta c\right) > 1\right\} \\ &\le \frac{1}{n}\log E_P\left[\exp\left(\vartheta\sum_{i=1}^n X_i - n\vartheta c\right)\right] \\ &= -\vartheta c + \frac{1}{n}\log E_P\left[\prod_{i=1}^n e^{\vartheta X_i}\right] = -\vartheta c + \frac{1}{n}\log\prod_{i=1}^n \underbrace{E_P\left[e^{\vartheta X_i}\right]}_{=e^{\gamma_\mu(\vartheta)}} \\ &= -\vartheta c + \gamma_\mu(\vartheta)\ .\end{aligned} \tag{**}$$

Das beweist die "$\le$"-Ungleichung in (*). Beachte, dass diese Ungleichung für jedes $c \in (E_0, E_0 + \alpha^+)$, jedes $\vartheta > 0$ und jedes $n \in \mathbb{N}$ gilt.

Wir beweisen die Umkehrung. Sei $\epsilon > 0$ so klein, dass $\tilde{c} := c + \epsilon \in (E_0, E_0 + \alpha^+)$. Dann gibt es ein $\tilde{\vartheta} \in \mathring{\Theta}_\mu \cap (0, \infty)$ derart, dass $\tilde{c} = E_{\tilde{\vartheta}}$, und es gilt:

$$\begin{aligned}&P\left\{\frac{1}{n}\sum_{i=1}^n X_i > c\right\} \ge P\left\{\tilde{c} + \epsilon > \frac{1}{n}\sum_{i=1}^n X_i > \tilde{c} - \epsilon\right\} \\ &= \int 1_{\left\{\tilde{c}+\epsilon > \frac{1}{n}\sum_{i=1}^n x_i > \tilde{c}-\epsilon\right\}} d\mu(x_1)\ \ldots\ d\mu(x_n) \\ &= \int 1_{\left\{\tilde{c}+\epsilon > \frac{1}{n}\sum_{i=1}^n x_i > \tilde{c}-\epsilon\right\}} f_{\tilde{\vartheta}}^{-1}(x_1)\cdot\ldots\cdot f_{\tilde{\vartheta}}^{-1}(x_n)\, d(f_{\tilde{\vartheta}}\mu)(x_1)\ \ldots\ d(f_{\tilde{\vartheta}}\mu)(x_n) \\ &= \int \exp\left(n\gamma_\mu(\tilde{\vartheta}) - \tilde{\vartheta}\sum_{i=1}^n x_i\right)\cdot 1_{\left\{\tilde{c}+\epsilon > \frac{1}{n}\sum_{i=1}^n x_i > \tilde{c}-\epsilon\right\}}\, d(f_{\tilde{\vartheta}}\mu)(x_1)\ \ldots\ d(f_{\tilde{\vartheta}}\mu)(x_n)\ .\end{aligned}$$

Seien $X_1', X_2', \ldots$ unabhängig identisch verteilte Zufallsvariablen mit Verteilung $f_{\tilde{\vartheta}}\mu$. Dann folgt

$$P\left\{\frac{1}{n}\sum_{i=1}^n X_i > c\right\} \ge e^{n(\gamma_\mu(\tilde{\vartheta}) - \tilde{\vartheta}(\tilde{c}+\epsilon))}\cdot P\left\{\tilde{c} + \epsilon > \frac{1}{n}\sum_{i=1}^n X_i' > \tilde{c} - \epsilon\right\}\ ,$$

und da $\tilde{c} = E_{\tilde{\vartheta}} = \int x f_{\tilde{\vartheta}}(x)\, d\mu(x) = E[X_i']$, konvergiert der zweite Faktor nach dem schwachen Gesetz der großen Zahl (Bemerkung 8.21) gegen 1 mit $n \to \infty$. Es folgt

$$\liminf_{n\to\infty} \frac{1}{n}\log P\left\{\frac{1}{n}\sum_{i=1}^n X_i > c\right\} \ge \gamma_\mu(\tilde{\vartheta}) - \tilde{\vartheta}(\tilde{c} + \epsilon)\ .$$

Da $c + \epsilon = \tilde{c} = E_{\tilde{\vartheta}}$ und da $\vartheta \mapsto E_\vartheta$ ein Homöomorphismus ist, ist $\lim_{\epsilon\to 0}\tilde{\vartheta} = \vartheta$, und da $\vartheta \mapsto \gamma_\mu(\vartheta)$ stetig ist (Korollar 13.5), folgt daraus im Limes $\epsilon \to 0$ die "$\ge$"-Ungleichung in (*). □

14.2 Bemerkung (Moderate Abweichungen) Seien $t > 0$, $0 < r < 1$ und $c = c(n) = E_0 + tn^{-r}$. Dann folgt aus (**), falls wieder $c = E_\vartheta$ für ein geignetes ϑ:

$$\frac{1}{n}\log P\left\{\frac{1}{n^{1-r}}\sum_{i=1}^n (X_i - E_0) > t\right\} = \frac{1}{n}\log P\left\{\frac{1}{n}\sum_{i=1}^n X_i > c(n)\right\} \le -\vartheta c(n) + \gamma_\mu(\vartheta) = -I_\mu(f_\vartheta).$$

Da $c(n) - E_0 = E_\vartheta - E_0 = V_0\vartheta + o(\vartheta)$ nach Korollar 13.5, ist $\vartheta = V_0^{-1}(c(n) - E_0) + o(|c(n) - E_0|)$, und es folgt unter Beachtung von $I_\mu(f_\vartheta) = \frac{\vartheta^2}{2}V_0 + o(\vartheta^2)$

$$\begin{aligned}\log P\left\{\frac{1}{n^{1-r}}\sum_{i=1}^n (X_i - E_0) > t\right\} &\le -\frac{n}{2V_0}(c - E_0)^2 + n\cdot o((c - E_0)^2) \\ &= -\frac{n^{1-2r}}{2V_0}t^2 + o(n^{1-2r})\end{aligned} \tag{***}$$

Für $r < \frac{1}{2}$ ist daher

$$\sum_{n=1}^{\infty} P\left\{\frac{1}{n^{1-r}}\sum_{i=1}^{n}(X_i - E_0) > t\right\} < \infty$$

und aus dem Borel-Cantelli-Lemma folgt, dass $\limsup_{n\to\infty} \frac{1}{n^{1-r}}\sum_{i=1}^{n}(X_i - E_0) \leq 0$ f.s. Ähnlich zeigt man, dass $\liminf_{n\to\infty} \frac{1}{n^{1-r}}\sum_{i=1}^{n}(X_i - E_0) \geq 0$ f.s., so dass $\lim_{n\to\infty} \frac{1}{n^{1-r}}\sum_{i=1}^{n}(X_i - E_0) = 0$ f.s.. Für $r = \frac{1}{2}$ ist das nicht mehr richtig. In diesem Fall wird das asymptotische Verhalten durch einen *Zentralen Grenzwertsatz* beschrieben, siehe Satz 16.16

Analog zur Herleitung von (***) zeigt man auch

$$\lim_{n\to\infty} \frac{1}{n^{1-2r}} \log P\left\{\frac{1}{n^{1-r}}\sum_{i=1}^{n}(X_i - E_0) > t\right\} = -\frac{t^2}{2V_0} ,$$

eine Aussage, die auch als Satz über *moderate Abweichungen* bezeichnet wird.

14.3 Übung Sei μ die Bernoulli-Verteilung zum Parameter $0 < p < 1$ auf $\{0,1\}$, $f_\vartheta(x) = e^{-\gamma_\mu(\vartheta)+\vartheta x}$. Zeigen Sie, dass für beliebige $\vartheta \in \mathbb{R}$

$$I_\mu(f_\vartheta) = q \log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p} \quad \text{mit } q = \frac{pe^\vartheta}{pe^\vartheta + 1 - p} \in (0,1).$$

14.4 Übung Sei μ die Normalverteilung mit Mittelwert t und Varianz σ^2, $f_\vartheta(x) = e^{-\gamma_\mu(\vartheta)+\vartheta x}$. . Zeigen Sie, dass für beliebige $\vartheta \in \mathbb{R}$ gilt

$$I_\mu(f_\vartheta) = t\vartheta - \frac{\sigma^2}{2}\vartheta^2 .$$

Kapitel 15

Verteilungskonvergenz

In diesem Kapitel führen wir ein Konvergenzkonzept ein, das es uns gestattet, das asymptotische Verhalten von normierten Summen $n^{-1/2}\sum_{i=1}^n X_i$ für unabhängig identisch verteilte X_i zu untersuchen, vergleiche auch die Diskussion in Bemerkung 14.2.

15.1 Definition (Schwache Konvergenz von Verteilungsfunktionen) *Seien F, F_n Verteilungsfunktionen auf $\mathbb{R}$, siehe Definition 5.1. Wir sagen (F_n) konvergiert schwach gegen F (symbolisch: $F_n \Longrightarrow F$), falls $\lim_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitsstellen x von F. (Beachte, dass F als monotone Funktion höchstens abzählbar viele Unstetigkeitsstellen hat.)*

15.2 Beispiel (Satz von de Moivre-Laplace) Seien $X_1, X_2, \dots$ unabhängig identisch verteilte Bernoulli-Zufallsvariablen zum Parameter p,

$$F_n(x) := P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - p) \le x\right)\ , \quad F(x) := \frac{1}{\sqrt{2\pi p(1-p)}}\int_{-\infty}^x e^{\frac{-t^2}{2p(1-p)}}\, dt\ .$$

Dann beschreibt “$F_n \Longrightarrow F$” gerade die Aussage des Satzes von de Moivre-Laplace.

Auch das schwache Gesetz der großen Zahl lässt sich so formulieren. Setze dazu

$$F_n(x) := P\left(\frac{1}{n}\sum_{i=1}^n X_i \le x\right)\ , \quad F(x) := \begin{cases} 1 & \text{falls } x \ge p \\ 0 & \text{falls } x < p\ . \end{cases}$$

Dann besagt das schwache GGZ gerade $\lim_{n\to\infty} F_n(x) = F(x)$ für alle $x \neq p$, d.h. $F_n \Longrightarrow F$.

Da Verteilungsfunktionen eineindeutig Wahrscheinlichkeitsmaßen auf $\mathbb{R}$ entsprechen, soll versucht werden, die schwache Konvergenz durch Maße zu beschreiben. Dazu verlassen wir den messbaren Raum $(\mathbb{R}, \mathcal{B})$ und betrachten eine etwas allgemeinere Situation:

Generalvoraussetzung: (E, d) ist ein metrischer Raum mit Borel-σ-Algebra $\mathcal{E}$.

15.3 Definition (Schwache Konvergenz von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{E})$)
Eine Folge $\mu_1, \mu_2, \dots$ von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{E})$ konvergiert schwach gegen das Wahrscheinlichkeitsmaß μ (symbolisch $\mu_n \Longrightarrow \mu$), falls

$$\lim_{n\to\infty}\int_E f\, d\mu_n = \int_E f\, d\mu \qquad \forall f \in C_b(E;\mathbb{R})\ ,$$

d.h. für alle stetigen beschränkten $f : E \to \mathbb{R}$.

15.4 Bemerkung Schwache Limiten sind eindeutig bestimmt, denn gilt sowohl $\mu_n \Longrightarrow \mu$ als auch $\mu_n \Longrightarrow \nu$, so ist $\int_E f\, d\mu = \int_E f\, d\nu$ für alle $f \in C_b(E;\mathbb{R})$. Um $\mu = \nu$ zu zeigen reicht es, $\mu(A) = \nu(A)$

für alle abgeschlossenen $A \subseteq E$ zu zeigen ($\cap$-stabiler Erzeuger der Borel-σ-Algebra). Für $A \subseteq E$ sei dazu

$$f_{A,\epsilon} : E \to \mathbb{R}, \quad f_{A,\epsilon}(x) = \begin{cases} 1 - \frac{d(x,A)}{\epsilon} & \text{falls } d(x,A) \leq \epsilon \\ 0 & \text{falls } d(x,A) > \epsilon\ . \end{cases}$$

Es ist $1_{\bar{A}} \leq f_{A,\epsilon} \leq 1$, $\lim_{\epsilon \to 0} f_{A,\epsilon}(x) = 1_{\bar{A}}(x)$ für alle x und $f_{A,\epsilon}$ ist Lipschitz-stetig (mit Lipschitz-Konstante ϵ^{-1}). Also folgt für abgeschlossenes A aus dem Satz von der majorisierten Konvergenz

$$\mu(A) = \lim_{k\to\infty} \int f_{A,\frac{1}{k}}\, d\mu = \lim_{k\to\infty} \int f_{A,\frac{1}{k}}\, d\nu = \nu(A)\ .$$

Der folgende Satz wird auch als *Portemanteau-Theorem* bezeichnet.

15.5 Satz (Charakterisierungen der schwachen Konvergenz von Wahrscheinlichkeitsmaßen)
Seien μ_n, μ Wahrscheinlichkeitsmaße auf $(E, \mathcal{E})$. Folgende Aussagen sind äquivalent:

(i) $\mu_n \Longrightarrow \mu$.

(ii) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *für alle* $A \in \mathcal{E}$ *mit* $\mu(\partial A) = 0$.

(iii) $\lim_{n\to\infty} \int_E f\, d\mu_n = \int_E f\, d\mu$ *für alle beschränkten Lipschitz-stetigen* $f : E \to \mathbb{R}$.

(iv) $\lim_{n\to\infty} \int_E f\, d\mu_n = \int_E f\, d\mu$ *für alle beschränkten messbaren* $f : E \to \mathbb{R}$ *mit* $\mu(U_f) = 0$.

Dabei ist $\partial A = \bar{A} \setminus \overset{\circ}{A}$ der Rand von A, und U_f ist die Menge der Unstetigkeitsstellen von f.[1]

Beweis: (i)$\Rightarrow$(iii) trivial.

(iii)$\Rightarrow$(ii) Sei $A \in \mathcal{E}$ mit $\mu(\partial A) = 0$. Mit $f_{A,\epsilon}$ wie in der vorangehenden Bemerkung gilt:

$$\begin{aligned}\limsup_{n\to\infty} \mu_n(A) &\leq \liminf_{k\to\infty} \limsup_{n\to\infty} \int f_{A,\frac{1}{k}}\, d\mu_n = \liminf_{k\to\infty} \int f_{A,\frac{1}{k}}\, d\mu = \mu(\bar{A}) \\ &\leq \mu(A) + \mu(\partial A) = \mu(A)\ .\end{aligned}$$

(Dabei wurde für die erste Ungleichung nur die Monotonie des Integrals benutzt.) Angewandt auf A^c liefert diese Ungleichung

$$\liminf_{n\to\infty} \mu_n(A) = 1 - \limsup_{n\to\infty} \mu_n(A^c) \geq 1 - \mu(A^c) = \mu(A)\ ,$$

und zusammen folgt daraus (ii).

(ii)$\Rightarrow$(iv) Sei $f : E \to \mathbb{R}$ messbar und beschränkt mit $\mu(U_f) = 0$. Da $f^{-1}\{t\} \cap f^{-1}\{t'\} = \emptyset$ falls $t \neq t'$, ist $T_0 := \{t \in \mathbb{R} : \mu(f^{-1}\{t\}) > 0\}$ höchstens abzählbar und zu $\epsilon > 0$ gibt es $t_0 \leq -\|f\|_\infty < t_1 < \cdots < t_{k-1} < \|f\|_\infty \leq t_k$ derart, dass

$$t_i \notin T_0 \quad \text{und} \quad |t_{i+1} - t_i| \leq \epsilon \quad \text{für alle } i.$$

Da $t_0 \leq f(x) \leq t_k$ für alle $x \in E$, ist

$$E = \biguplus_{i=1}^{k} \underbrace{\{x : t_{i-1} \leq f(x) < t_i\}}_{=:E_i}\ ,$$

[1] Die Messbarkeit von U_f zeigt man folgendermaßen: Sei $osc_f(x) := \lim_{\delta\to 0} \sup\{d_2(f(y), f(x)) : d_1(y,x) < \delta\}$. Offensichtlich ist $U_f = \{x : osc_f(x) > 0\} = \bigcup_{k=1}^{\infty} \{x : osc_f(x) \geq \frac{1}{k}\}$, und man überzeugt sich leicht, dass $\{x : osc_f(x) < \frac{1}{k}\}$ für jedes $k > 0$ offen ist. U_f ist also eine sogenannte F_σ-Menge, d.h. eine abzählbare Vereinigung abgeschlossener Mengen.

wobei $\mu(\partial E_i) \leq \mu(f^{-1}\{t_{i-1}\}) + \mu(f^{-1}\{t_i\}) + \mu(U_f) = 0$. Daher ist

$$\begin{aligned}
\int f\,d\mu_n &\leq \sum_{i=1}^{k} \mu_n\{t_{i-1} \leq f < t_i\} \cdot t_i \\
&\leq \sum_{i=1}^{k} \mu_n(E_i) \cdot t_{i-1} + \sum_{i=1}^{k} \mu_n(E_i) \cdot \epsilon \\
&\leq \sum_{i=1}^{k} \mu(E_i) \cdot t_{i-1} + \sum_{i=1}^{k} (\mu_n(E_i) - \mu(E_i)) \cdot t_{i-1} \ + \ \epsilon \\
&\leq \int f\,d\mu + \epsilon + \sum_{i=1}^{k} \underbrace{(\mu_n(E_i) - \mu(E_i))}_{\to 0 \text{ mit } n \to \infty \text{ für jedes } i} \cdot t_{i-1} \ ,
\end{aligned}$$

also

$$\limsup_{n\to\infty} \int f\,d\mu_n \leq \int f\,d\mu$$

und durch Betrachtung von $(-f)$ erhält man

$$\liminf_{n\to\infty} \int f\,d\mu_n \geq \int f\,d\mu$$

(iv)⇒(i) trivial. □

15.6 Beispiele Sei $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$.

a) Sei $\mu_n = h_n\lambda$, $\mu = h\lambda$ und gelte $\lim_{n\to\infty} \|h_n - h\|_1 = 0$. Dann gilt für jedes $A \in \mathcal{B}$:
$|\mu_n(A) - \mu(A)| \leq \int_A |h_n - h|\,d\lambda \leq \|h_n - h\|_1 \to 0$ mit $n \to \infty$.

b) Seien $\mu_n = \delta_{c_n}$, $\mu = \delta_c$, wo $c, c_n \in \mathbb{R}$ und $\lim_{n\to\infty} c_n = c$. Dann gilt für stetiges f:
$|\int f\,d\mu_n - \int f\,d\mu| = |f(c_n) - f(c)| \to 0$ mit $n \to \infty$.

c) Seien $\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{k/n}$, $\mu = 1_{[0,1]}\lambda$. Dann gilt für Lipschitz-stetiges f:

$$\left|\int f\,d\mu_n - \int f\,d\mu\right| \leq \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} |f(k/n) - f(x)|\,dx \leq \frac{1}{n}\mathrm{Lip}(f) \to 0 \quad \text{mit } n \to \infty.$$

15.7 Definition (Verteilungskonvergenz) *Seien X, X_n Zufallsvariablen. Die Folge $(X_n)_{n>0}$ konvergiert in Verteilung (oder auch schwach) gegen X, kurz $X_n \Longrightarrow X$, falls $P_{X_n} \Longrightarrow P_X$.*

15.8 Satz (Charakterisierungen der Verteilungskonvergenz)
Seien X, X_n Zufallsvariablen mit Verteilungsfunktionen F bzw. F_n. Äquivalent sind

(i) $X_n \Longrightarrow X$

(ii) $F_n \Longrightarrow F$

(iii) $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ *für alle* $f \in C_b(\mathbb{R};\mathbb{R})$.

Beweis: (i)⇒(ii) Sei F bei x stetig, d.h. $P_X\{x\} = 0$. Aus $X_n \Longrightarrow X$ und Satz 15.5 folgt

$$F_n(x) = P_{X_n}(-\infty, x] \to P_X(-\infty, x] = F(x) \ .$$

(ii)⇒(i) Zu zeigen ist $\lim_{n\to\infty}\int f\,dP_{X_n} = \int f\,dP$ für alle Lipschitz-stetigen $f:\mathbb{R}\to\mathbb{R}$, siehe Satz 15.5. Sei $\epsilon>0$. Wir können $t_0<t_1<\cdots<t_k$ in $\mathbb{R}$ so wählen, dass

$$F(t_0)=P(-\infty,t_0]<\epsilon,\quad 1-F(t_k)=P(t_k,+\infty)<\epsilon,$$
$$t_i-t_{i-1}<\epsilon \quad \forall\, i=1,\dots,k, \text{ und } F \text{ an allen Stellen } t_0,\dots,t_k \text{ stetig ist.}$$

Insbesondere ist $\lim_{n\to\infty}F_n(t_i)=F(t_i)\ \forall\, i=0,\dots,k$, so dass

$$\exists\, n_0\in\mathbb{N}\ \forall\, n\geq n_0\ \forall\, i=0,\dots,k:\ |F_n(t_i)-F(t_i)|<\frac{\epsilon}{k+1}\ .$$

Setze

$$\overline{f_i}:=\max_{t_{i-1}\leq x\leq t_i} f(x),\qquad \underline{f_i}:=\min_{t_{i-1}\leq x\leq t_i} f(x).$$

Dann ist für $n\geq n_0$:

$$\begin{aligned}
&\int f\,dP_{X_n}\\
&\leq \|f\|_\infty\cdot(F_n(t_0)+1-F_n(t_k))+\sum_{i=1}^k \overline{f_i}\cdot(F_n(t_i)-F_n(t_{i-1}))\\
&\leq \|f\|_\infty\cdot\left(F(t_0)+1-F(t_k)+2\frac{\epsilon}{k+1}\right)+\sum_{i=1}^k\overline{f_i}\cdot(F(t_i)-F(t_{i-1}))+\sum_{i=1}^k 2\frac{\epsilon}{k+1}\,\|f\|_\infty\\
&\leq \|f\|_\infty\cdot 2\epsilon+\sum_{i=1}^k\left(\underline{f_i}+\epsilon\cdot\mathrm{Lip}(f)\right)(F(t_i)-F(t_{i-1}))\\
&\leq 2\epsilon\|f\|_\infty+\int_{t_0}^{t_k} f\,dP_X+\epsilon\,\mathrm{Lip}(f)\\
&\leq \int f\,dP_X+\epsilon\cdot(2\|f\|_\infty+2\|f\|_\infty+\mathrm{Lip}(f))
\end{aligned}$$

und es folgt $\limsup_{n\to\infty}\int f\,dP_{X_n}\leq\int f\,dP_X$. Durch Betrachtung von $(-f)$ an Stelle von f folgt $\liminf_{n\to\infty}\int f\,dP_{X_n}\geq\int f\,dP_X$ und damit die Behauptung.

(i)⇔(iii) Ist nur eine Frage der Schreibweise: $E[f(X)]=\int f\,dP_X$. □

15.9 Satz
Sei $h:(E_1,d_1)\to(E_2,d_2)$ Borel-messbar.

a) *Seien μ,μ_n Wahrscheinlichkeitsmaße auf $(E_1,\mathcal{E}_1)$ mit $\mu_n\Longrightarrow\mu$ und $\mu(U_h)=0$. Dann konvergiert auch $\mu_n\circ h^{-1}\Longrightarrow\mu\circ h^{-1}$.*

b) *Sind $E_1=E_2=\mathbb{R}$ und sind X,X_n Zufallsvariablen mit $X_n\Longrightarrow X$ und $P_X(U_h)=0$, so folgt $h(X_n)\Longrightarrow h(X)$.*

Beweis:

a) Zunächst zeigen wir, dass U_h (und entsprechend $U_{f\circ h}$) messbar ist: Für $x\in E_1$ sei

$$osc_h(x):=\lim_{k\to\infty}\sup\left\{d_2(h(y),h(x)) : y\in E_1, d_1(y,x)<\frac{1}{k}\right\}\ .$$

Offensichtlich ist

$$U_h=\{x\in E_1 : osc_h(x)>0\}=\bigcup_{k=1}^\infty\{x\in E_1 : osc_h(x)\geq\frac{1}{k}\}\ ,$$

und man sieht leicht, dass die Komplemente dieser Mengen, nämlich die Mengen $\{x \in E_1 : osc_h(x) < \frac{1}{k}\}$ offen und damit messbar sind.

Sei nun $f \in C_b(E_2; \mathbb{R})$. Dann ist $f \circ h : E_1 \to \mathbb{R}$ beschränkt und $U_{f \circ h} \subseteq U_h$, also $\mu(U_{f\circ h}) = 0$. Es folgt aus Satz 15.5:

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

also $\mu_n \circ h^{-1} \Longrightarrow \mu \circ h^{-1}$.

b) Folgt aus a), da $P_{h(X)} = P_X \circ h^{-1}$.

□

15.10 Übung Seien X, X_n Zufallsvariablen, $X_n \Longrightarrow X$. Dann gilt:

a) $E|X| \le \liminf_{n\to\infty} E|X_n|$.

b) Sind die X_n gleichgradig integrierbar, so ist $EX = \lim_{n\to\infty} EX_n$.

15.11 Beispiel Aussage b) dieser Übung bleibt ohne die Annahme der gleichgradigen Integrierbarkeit nicht richtig: Betrachte dazu $X_1, X_2, \dots$ mit $P(X_n = 0) = 1 - \frac{1}{n^2}$, $P(X_n = n^2) = \frac{1}{n^2}$, und $X = 0$. Dann ist für $f \in C_b(E; \mathbb{R})$

$$|E[f(X_n)] - E[f(X)]| = \frac{1}{n^2} \left| f(n^2) - f(0) \right| \le \frac{2}{n^2} \, \|f\|_\infty \to 0 \, ,$$

aber $EX_n = 1 \not\to EX = 0$. Aus dem Borel-Cantelli Lemma folgt sogar: $\lim_{n\to\infty} X_n = 0$ f.s.

Zum Begriff der schwachen Konvergenz gehört auch ein Kompaktheitskonzept. Das werden wir im Rest dieses Kapitels untersuchen.

15.12 Definition (Folgenkompaktheit von Wahrscheinlichkeitsmaßen) *Eine Familie $\mathcal{M}$ von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{E})$ heißt relativ folgenkompakt, falls jede Folge in $\mathcal{M}$ eine schwach konvergente Teilfolge hat.*

15.13 Definition (Straffheit) *Eine Familie $\mathcal{M}$ von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{E})$ heißt straff, falls für jedes $\epsilon > 0$ eine kompakte Menge $K \subseteq E$ existiert, so dass $\sup_{\mu\in\mathcal{M}} \mu(E \setminus K) \le \epsilon$.*

Beide Definitionen übertragen sich auf Verteilungsfunktionen, wenn wir sie mit den zugehörigen Wahrscheinlichkeitsmaßen identifizieren.

15.14 Satz (Prohorov: relativ folgenkompakt $\Longleftrightarrow$ straff)
Sei (E, d) ein vollständiger, separabler metrischer Raum. Eine Familie $\mathcal{M}$ von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{E})$ ist genau dann relativ folgenkompakt, wenn sie straff ist.

Beweis: Wir beweisen diesen Satz nur im Fall $E = \mathbb{R}$.
“$\Longrightarrow$” Angenommen, $\mathcal{M}$ ist nicht straff. Dann gibt es ein $\epsilon > 0$ derart, dass

$$\forall\, n \in \mathbb{N}\ \exists\, \mu_n \in \mathcal{M} : \ \mu_n(M_n) > \epsilon, \quad \text{wo } M_n := \mathbb{R} \setminus [-n, n] \, .$$

Wähle eine Teilfolge $\mu_{n_j} \Longrightarrow \mu$ mit $j \to \infty$. Zu jedem n gibt es ein $\eta_n \in [0, 1)$ mit $\mu\{-(n + \eta_n), n + \eta_n\} = 0$. Also folgt für alle $k \in \mathbb{N}$ aus Satz 15.5

$$\begin{aligned} \mu(M_k) &\ge \mu\left(\mathbb{R} \setminus [-(k + \eta_k), k + \eta_k]\right) = \lim_{j\to\infty} \mu_{n_j}\left(\mathbb{R} \setminus [-(k + \eta_k), k + \eta_k]\right) \\ &\ge \limsup_{j\to\infty} \mu_{n_j}(M_{n_j}) \ge \epsilon \, , \end{aligned}$$

und wir erhalten den Widerspruch $0 < \epsilon \leq \liminf_{k\to\infty} \mu(M_k) = 0$.

"$\Longleftarrow$" Zum Beweis der Gegenrichtung benötigen wir den folgenden Satz:

15.15 Satz (Helly)
Ist $(F_n)_{n>0}$ eine Folge von Verteilungsfunktionen, so gibt es eine Teilfolge $(F_{n_i})_{i>0}$ und eine nichtfallende, rechtsseitig stetige Funktion F derart, dass $\lim_{i\to\infty} F_{n_i}(x) = F(x)$ an allen Stetigkeitsstellen x von F.

Es folgt sofort, dass $0 \leq F \leq 1$ und dass $F \nearrow$, aber es ist nicht notwendig

$$\lim_{x\to-\infty} F(x) = 0 \text{ und } \lim_{x\to+\infty} F(x) = 1. \tag{*}$$

Wir kehren zum Beweis von Satz 15.14 zurück. Sei $(\mu_n)_{n>0}$ eine Folge in $\mathcal{M}$, $(F_n)_{n>0}$ die Folge der zugehörigen Verteilungsfunktionen. Seien F_{n_i} und F wie im Satz von Helly. Ist F eine Verteilungsfunktion, so folgt $F_{n_i} \Longrightarrow F$, also $\mu_{n_i} \Longrightarrow \mu$ für das durch F bestimmte Wahrscheinlichkeitsmaß μ auf $E = \mathbb{R}$. Zu zeigen ist also, dass F unter der Straffheitsannahme tatsächlich eine Verteilungsfunktion ist, d.h. (*).

Sei $\epsilon > 0$. Da $\mathcal{M}$ straff ist, folgt

$$\begin{aligned}
&\exists\, K \text{ kompakt} \,\forall\, i > 0: \ \mu_{n_i}(\mathbb{R} \setminus K) \leq \epsilon \\
\Rightarrow &\exists\, M \in \mathbb{N} \,\forall\, i > 0: \ \mu_{n_i}(\mathbb{R} \setminus [-M, M]) \leq \epsilon \\
\Rightarrow &\exists\, M \in \mathbb{N} \,\forall\, i > 0: \ \forall x \leq -M: \ F_{n_i}(x) \leq \epsilon \ \text{ und } \ \forall y \geq M: \ 1 - F_{n_i}(y) \leq \epsilon
\end{aligned}$$

Da $\epsilon > 0$ beliebig ist, folgt (*). □

Beweis von Satz 15.15: Da $0 \leq F_n(x) \leq 1$ für alle $n > 0$ und $x \in \mathbb{R}$, gibt es zu jedem x und jeder Teilfolge von $(F_n(x))$ eine Teilfolge dieser Teilfolge, die konvergiert. Durch einen Diagonalschluss erhält man:

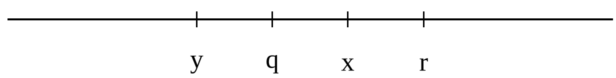
Es gibt eine Teilfolge $(F_{n_i})_{i>0}$ derart, dass $G(q) := \lim_{i\to\infty} F_{n_i}(q)$ existiert $\forall\, q \in \mathbb{Q}$.

Setze $F(x) := \inf\{G(q) : q > x\}$. Dann gilt:

i) F_{n_i} nichtfallend $\Rightarrow G$ nichtfallend $\Rightarrow F$ nichtfallend.

ii) F ist rechtsseitig stetig, denn: Zu $x \in \mathbb{R}$ und $\epsilon > 0$ existiert ein $q > x$ in $\mathbb{Q}$ mit $G(q) < F(x) + \epsilon$. Ist dann $x \leq y < q$, so folgt $F(y) \leq G(q) < F(x) + \epsilon$.

iii) $\lim_{i\to\infty} F_{n_i}(x) = F(x)$ an allen Stetigkeitsstellen x von F. Denn: Sei F stetig bei x, $\epsilon > 0$. Wähle $y < x$ derart, dass $F(x) < F(y) + \epsilon$ und wähle $q, r \in \mathbb{Q}$ mit $y < q < x < r$ und $G(r) < F(x) + \epsilon$. Dann ist
$$F(x) - \epsilon \leq F(y) \leq G(q) \leq G(r) < F(x) + \epsilon\,.$$
Da
$$G(q) = \lim_{i\to\infty} F_{n_i}(q) \leq \liminf_{i\to\infty} F_{n_i}(x) \leq \limsup_{i\to\infty} F_{n_i}(x) \leq \lim_{i\to\infty} F_{n_i}(r) = G(r)\,,$$
folgt
$$F(x) - \epsilon \leq \liminf_{i\to\infty} F_{n_i}(x) \leq \limsup_{i\to\infty} F_{n_i}(x) \leq F(x) + \epsilon\,,$$
also die behauptete Stetigkeit von F bei x.



$$F(x) - \epsilon < F(y) \leq G(y), \quad G(r) < F(x) + \epsilon$$

□

Kapitel 16

Charakteristische Funktionen und der Zentrale Grenzwertsatz

Ein wichtiges Hilfsmittel zur Untersuchung der Verteilungskonvergenz reellwertiger Zufallsvariablen sind die *charakteristischen Funktionen*. Die Grundidee ist, dass man jedem Wahrscheinlichkeitsmaß eine auf ganz $\mathbb{R}$ definierte $\mathbb{C}$-wertige charakteristische Funktion derart eineindeutig zuordnet, dass eine Folge von Wahrscheinlichkeitsmaßen genau dann schwach konvergiert, wenn die Folge ihrer charakteristischen Funktionen punktweise konvergiert.

16.1 Definition (Charakteristische Funktion) *a) Sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$. Die <u>Fourier-Transformierte</u> $\varphi_\mu : \mathbb{R} \to \mathbb{C}$ von μ,*

$$\varphi_\mu(t) := \int e^{itx}\, d\mu(x) \left[= \int \cos(tx)\, d\mu(x) + i \int \sin(tx)\, d\mu(x) \right]$$

wird auch als <u>charakteristische Funktion</u> von μ bezeichnet.

b) Ist X eine Zufallsvariable mit $P\{|X| < \infty\} = 1$, so heißt $\varphi_X := \varphi_{P_X}$ die <u>charakteristische Funktion</u> von X. Es ist

$$\varphi_X(t) = \int e^{itx}\, dP_X(x) = E\left[e^{itX}\right] \;.$$

16.2 Lemma *Sei φ eine charakteristische Funktion.*

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

b) $\varphi : \mathbb{R} \to \mathbb{C}$ ist gleichmäßig stetig.

c) Ist $P\{|X| < \infty\} = 1$, so gilt $\varphi_{aX+b}(t) = e^{itb} \cdot \varphi_X(at)$ für alle $a, b \in \mathbb{R}$.

Beweis:

a) trivial.

b) Für alle $t, h \in \mathbb{R}$ ist

$$|\varphi(t+h) - \varphi(t)| = \left| \int e^{itx}(e^{ihx} - 1)\, d\mu(x) \right| \leq \int \left| e^{ihx} - 1 \right|\, d\mu(x)$$

unabhängig von t, und da $|e^{ihx}| = 1$ für alle h und $\lim_{h\to 0} e^{ihx} = 1$, folgt aus dem Satz von der majorisierten Konvergenz $\lim_{h\to 0} \int |e^{ihx} - 1|\, d\mu(x) = 0$.

c) $E[e^{it(aX+b)}] = e^{itb} \cdot E[e^{iatX}]$.

□

16.3 Lemma *a) Sind $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit $P\{|X_i| < \infty\} = 1$, so ist*

$$\varphi_{X_1+\dots+X_n} = \varphi_{X_1} \cdot \dots \cdot \varphi_{X_n} .$$

b) Sind μ_1, μ_2 Wahrscheinlichkeitsmaße auf $\mathbb{R}$, so ist

$$\varphi_{\mu_1 * \mu_2} = \varphi_{\mu_1} \cdot \varphi_{\mu_2} .$$

Beweis:

a) Wir zeigen das Lemma für $n = 2$. Der Rest folgt durch Induktion.

$$E\left[e^{it(X_1+X_2)}\right] = E\left[e^{itX_1} \cdot e^{itX_2}\right] \overset{\text{Unabhängigkeit}}{=} E\left[e^{itX_1}\right] \cdot E\left[e^{itX_2}\right]$$

Dabei haben wir benutzt, dass ganz allgemein für unabhängige $\mathbb{C}$-wertige beschränkte Zufallsvariablen X, Y gilt $E[XY] = EX \cdot EY$, denn

$$\begin{aligned} E[XY] &= E\left[\Re X \cdot \Re Y - \Im X \cdot \Im Y + i\left(\Re X \cdot \Im Y + \Im X \cdot \Re Y\right)\right] \\ &= E[\Re X] \cdot E[\Re Y] + E[\Im X] \cdot E[\Im Y] + i\left(E[\Re X] \cdot E[\Im Y] + E[\Im X] \cdot E[\Re Y]\right) \\ &= \left(E[\Re X] + iE[\Im X]\right) \cdot \left(E[\Re Y] + iE[\Im Y]\right) \\ &= EX \cdot EY . \end{aligned}$$

b) Betrachte unabhängige Zufallsvariablen X_1, X_2 mit Verteilungen μ_1 bzw. μ_2. Da $P_{X_1+X_2} = \mu_1 * \mu_2$ (siehe Satz 11.19), folgt die Behauptung sofort aus Teil a).

□

Zum Rechnen mit charakteristischen Funktionen benötigen wir ein paar analytische Abschätzungen.

16.4 Lemma *Für $n \geq 0$ und $x \in \mathbb{R}$ ist*

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \underbrace{\frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is}\, ds}_{=:I_n} \quad \textit{mit} \quad |I_n| \leq \min\left\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right\} , \qquad (*)$$

insbesondere

$$\left|e^{ix} - 1\right| \leq \min\{|x|, 2\} \quad \textit{und} \quad \left|e^{ix} - (1 + ix - \frac{x^2}{2})\right| \leq \min\left\{\frac{|x|^3}{6}, x^2\right\} .$$

Beweis: Durch partielle Integration erhält man für $n \geq 1$:

$$\int_0^x (x-s)^{n-1} e^{is}\, ds = \frac{x^n}{n} + \frac{i}{n} \int_0^x (x-s)^n e^{is}\, ds ,$$

also

$$I_{n-1} = \frac{i^n}{(n-1)!} \int_0^x (x-s)^{n-1} e^{is}\, ds = \frac{(ix)^n}{n!} + I_n . \qquad (**)$$

Da außerdem $e^{ix} = 1 + (e^{ix} - 1) = 1 + i \int_0^x e^{is}\, ds = 1 + I_0$, folgt (*) aus (**) durch vollständige Induktion.

Außerdem ist

$$|I_n| \leq \frac{1}{n!} \int_0^{|x|} s^n\, ds = \frac{|x|^{n+1}}{(n+1)!} ,$$

woraus zusammen mit (**) schließlich folgt

$$|I_n| \leq |I_{n-1}| + \frac{|x|^n}{n!} \leq 2\frac{|x|^n}{n!} .$$

□

16.5 Satz (Charakteristische Funktionen und Momente)
Sei $\varphi(t)$ die charakteristische Funktion der Zufallsvariablen X.

a) Ist $E[|X|^n] < \infty$, so ist

$$\left|\varphi(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!} E[X^k]\right| \le E\left[\min\left\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right\}\right]$$

und φ ist n-mal stetig differenzierbar mit Ableitungen

$$\varphi^{(k)}(t) = E\left[(iX)^k e^{itX}\right] \quad (k = 0, \ldots, n) .$$

b) Für jedes $t \in \mathbb{R}$, für das $\lim_{n\to\infty} \frac{|t|^n \, E[|X|^n]}{n!} = 0$, gilt

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} E[X^k] .$$

Das gilt insbesondere, falls $E[e^{|tX|}] < \infty$.

c) Ist $E\left[X^2\right] < \infty$, so ist $\varphi(t) = 1 + itE[X] - \frac{t^2}{2} E\left[X^2\right] + o(t^2)$ im Limes $t \to 0$.

Beweis:

1. Aus Lemma 16.4 folgt

$$\left|\varphi(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!} E[X^k]\right| \le E\left|e^{itx} - \sum_{k=0}^{n} \frac{(itX)^k}{k!}\right| \le E\left[\min\left\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right\}\right] .$$

Aus der zweiten dieser beiden Ungleichungen folgt außerdem

$$\begin{aligned}\left|\varphi(u+h) - \sum_{k=0}^{n} \frac{h^k}{k!} E\left[(iX)^k e^{iuX}\right]\right| &= \left|E\left[e^{iuX}\left(e^{ihX} - \sum_{k=0}^{n} \frac{(ihX)^k}{k!}\right)\right]\right| \\ &\le E\left|e^{ihX} - \sum_{k=0}^{n} \frac{(ihX)^k}{k!}\right| \\ &\le \frac{|h|^n}{n!} E\Bigg[\underbrace{|X|^n \cdot \min\left\{\frac{|h| \cdot |X|}{n+1}, 2\right\}}_{=:Z_h}\Bigg] ,\end{aligned}$$

und es bleibt zu zeigen, dass $\lim_{h\to 0} E[Z_h] = 0$: Da $0 \le Z_h \le 2|X|^n$ und da $\lim_{h\to 0} Z_h = 0$ falls $|X| < \infty$, das heißt fast sicher, folgt die Behauptung aus dem Satz von der majorisierten Konvergenz.

2. Die Konvergenz der Potenzreihendarstellung folgt direkt aus der Abschätzung in Teil a). Außerdem ist

$$\sum_{k=0}^{\infty} \frac{|t|^k}{k!} E\left[|X|^k\right] = E\left[\sum_{k=0}^{\infty} \frac{|t|^k |X|^k}{k!}\right] = E\left[e^{|tX|}\right] ,$$

so dass $\lim_{k\to\infty} \frac{|t|^k E[|X|^k]}{k!} = 0$ falls $E\left[e^{|tX|}\right] < \infty$.

3. Aus Teil a) folgt:

$$\begin{aligned}\left|\varphi(t) - \left(1 + itE[X] - \frac{t^2}{2}E\left[X^2\right]\right)\right| &\le \frac{t^3}{6}\,E\left[|X|^3 \cdot 1_{\{|X|\le t^{-\frac{1}{4}}\}}\right] + t^2\,E\left[X^2 \cdot 1_{\{|X|\ge t^{-\frac{1}{4}}\}}\right] \\ &\le \frac{1}{6}t^{-\frac{9}{4}} + t^2 \cdot o(1) \qquad \text{im Limes } t \to 0.\end{aligned}$$

(Für den letzten Schritt siehe Bemerkung 8.14.)

□

Satz 16.5 zeigt eine Beziehung zwischen *globalen* Eigenschaften von P_X, nämlich den Momenten, und *lokalen* (nämlich Differenzierbarkeits-) Eigenschaften von φ_X auf. Der folgende Satz setzt umgekehrt eine lokale Eigenschaft von P_X voraus (P_X hat Dichte zum Lebesgue-Maß auf $\mathbb{R}$) und folgert daraus eine globale Eigenschaft von φ_X ($\varphi_X(t) \to 0$ für $|t| \to \infty$).

16.6 Satz (Riemann-Lebesgue)
Hat die Verteilung μ Dichte bzgl. des Lebesgue-Maßes λ, so ist $\lim_{|t|\to\infty} \varphi_\mu(t) = 0$.

Beweis: Sei $\mu = f\lambda$, also $f \in \mathcal{L}^1_\lambda$. Sei $\epsilon > 0$ beliebig. Wegen Satz 4.11 gibt es eine Elementarfunktion $g = \sum_{k=1}^n \alpha_k 1_{A_k}$ mit $A_k \in \mathcal{B}$ und $\int |f-g|\,d\lambda < \frac{\epsilon}{2}$. Wegen Satz 3.13b gibt es endliche Vereinigungen B_k von Intervallen derart, dass für $h = \sum_{k=1}^n \alpha_k 1_{B_k}$ gilt:

$$\int |f-h|\,d\lambda \le \frac{\epsilon}{2} + \int |g-h|\,d\lambda \le \frac{\epsilon}{2} + \sum_{k=1}^n |\alpha_k| \cdot \lambda(A_k \triangle B_k) < \epsilon\,.$$

Durch Umnummerierung können wir annehmen, die B_k seien selbst Intervalle, $B_k = (u_k, v_k]$. Dann ist

$$\left|\varphi_\mu(t) - \int h(x)e^{itx}\,dx\right| = \left|\int f(x)e^{itx}\,dx - \int h(x)e^{itx}\,dx\right| \le \int |f-h|\,d\lambda < \epsilon$$

und

$$\begin{aligned}\left|\int h(x)e^{itx}\,dx\right| &\le \sum_{k=1}^n |\alpha_k| \left|\int_{u_k}^{v_k} e^{itx}\,dx\right| \\ &= \sum_{k=1}^n |\alpha_k| \frac{|e^{itv_k} - e^{itu_k}|}{|t|} \\ &\le 2 \cdot \sum_{k=1}^n \frac{|\alpha_k|}{|t|} \\ &\to 0 \qquad \text{im Limes } |t| \to \infty.\end{aligned}$$

Da $\epsilon > 0$ beliebig war, folgt die Behauptung des Satzes. □

Nun wollen wir sehen, wie charakteristische Funktionen uns bei der Untersuchung von zentrierten und standardisierten Summen unabhängiger Zufallsvariablen helfen können.

16.7 Satz
Seien $X_1, X_2, \dots$ unabhängige identisch verteilte Zufallsvariablen mit $m := E[X_i]$ und $\sigma^2 := V(X_i) < \infty$. Sei

$$S_n^* := \frac{1}{\sqrt{n\sigma^2}} \sum_{k=1}^n (X_k - m)$$

die zentrierte und standardisierte Partialsumme *von $X_1, \dots, X_n$. Dann ist*

$$\lim_{n\to\infty} \varphi_{S_n^*}(t) = e^{-\frac{t^2}{2}}\,.$$

Beweis: Sei $\varphi(t)$ die charakteristische Funktion der Zufallsvariablen $X_k - m$, also nach Satz 16.5

$$\varphi(t) = 1 - \frac{\sigma^2}{2}t^2 + t^2\,\beta(t) \qquad \text{wo } \lim_{t\to 0}\beta(t) = 0.$$

Wegen Lemma 16.3 und Lemma 16.2c) ist dann

$$\varphi_{S_n^*}(t) = \left(\varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right)^n = \left(1 - \frac{t^2}{2n}\right)^n - \left[\left(1 - \frac{t^2}{2n}\right)^n - \left(\varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right)^n\right]$$

und im Limes $n \to \infty$ gilt für alle $t \in \mathbb{R}$

$$\left(1 - \frac{t^2}{2n}\right)^n \to e^{-\frac{t^2}{2}}$$

$$\left|\left(1 - \frac{t^2}{2n}\right)^n - \left(\varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right)^n\right| \le n \cdot \left|1 - \frac{t^2}{2n} - \varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right| \le n\,\frac{t^2}{n\sigma^2}\beta\left(\frac{t}{\sqrt{n\sigma^2}}\right) \to 0\,,$$

da $|u^n - v^n| \le |u - v| \cdot n \cdot \max\{|u|, |v|\}^{n-1}$ für beliebige $u, v \in \mathbb{C}$. □

16.8 Bemerkung Seien X, Y unabhängig und normalverteilt, $P_X = \mathcal{N}(0, \sigma^2)$, $P_Y = \mathcal{N}(0, \rho^2)$. Eine einfache Rechnung zeigt, dass dann $P_{X+Y} = P_X * P_Y = \mathcal{N}(0, \sigma^2 + \rho^2)$: Sei $\tau^2 = \frac{\sigma^2\rho^2}{\sigma^2+\rho^2}$. Dann ist

$$\frac{1}{\tau^2} = \frac{1}{\sigma^2} + \frac{1}{\rho^2} \quad \text{und} \quad \frac{1}{\sigma^2} - \frac{\tau^2}{\sigma^4} = \frac{1}{\sigma^2}\left(1 - \frac{\rho^2}{\sigma^2 + \rho^2}\right) = \frac{1}{\sigma^2 + \rho^2}\,,$$

und nach Satz 11.19 hat P_{X+Y} die Dichte

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{\sqrt{2\pi\rho^2}}\int \exp\left(-\frac{(x-y)^2}{2\sigma^2} - \frac{y^2}{2\rho^2}\right)\,dy \\
=&\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{\sqrt{2\pi\rho^2}}\int \exp\left(-\frac{(y - x\frac{\tau^2}{\sigma^2})^2}{2\tau^2} + \frac{x^2\tau^2}{2\sigma^4} - \frac{x^2}{2\sigma^2}\right)\,dy \\
=&\frac{1}{\sqrt{2\pi\frac{\sigma^2\rho^2}{\tau^2}}}\exp\left(-\frac{x^2}{2}\left(\frac{1}{\sigma^2} - \frac{\tau^2}{\sigma^4}\right)\right)\frac{1}{\sqrt{2\pi\tau^2}}\int \exp\left(-\frac{y^2}{2\tau^2}\right)\,dy \\
=&\frac{1}{\sqrt{2\pi(\sigma^2 + \rho^2)}}\exp\left(-\frac{x^2}{2(\sigma^2 + \rho^2)}\right)
\end{aligned}$$

Sind daher die X_k im vorigen Satz nach $\mathcal{N}(0, 1)$ verteilt, so ist $X_1 + \cdots + X_n$ nach $\mathcal{N}(0, n)$ verteilt, und daher ist S_n^* nach $\mathcal{N}(0, 1)$ verteilt. Es folgt sofort, dass $\varphi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}}$.

16.9 Korollar *In der Situation von Satz 16.7 ist für alle* $t \in \mathbb{R}$

$$\lim_{n\to\infty}\varphi_{S_n^*}(t) = \varphi_{\mathcal{N}(0,1)}(t)\,.$$

Im Rest dieses Kapitels werden wir hauptsächlich zeigen, dass daraus folgt

$$P_{S_n^*} \Longrightarrow \mathcal{N}(0, 1) \quad (n \to \infty)\,,$$

d.h. wir werden den *Zentralen Grenzwertsatz* beweisen. Vorher geben wir noch einige charakteristische Funktionen an.

16.10 Beispiel Hier sind einige Beispiele charakteristischer Funktionen:

Für die Normalverteilung wurde das in Bemerkung 16.8 gezeigt, für die übrigen Verteilungen ist die Bestimmung der charakteristischen Funktionen eine einfache *Übung*.

Verteilung	Formel	Charakteristische Funktion
Normalverteilung	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}\,dx \quad (-\infty < x < \infty)$	$e^{-\frac{\sigma^2 t^2}{2}}$
Gleichverteilung	$1_{[0,1]}(x)\,dx$	$\frac{e^{it}-1}{it}$
Exponentialverteilung	$\lambda e^{-\lambda x}\,dx \quad (0 \le x < \infty)$	$\frac{\lambda}{\lambda - it}$
Binomialverteilung	$\binom{n}{k}p^k(1-p)^{n-k}\,\delta_k \quad (k = 0, 1, \ldots, n)$	$(1 + p(e^{it}-1))^n$
Poisson-Verteilung	$e^{-\alpha}\frac{\alpha^k}{k!}\,\delta_k \quad (k = 0, 1, 2, \ldots)$	$e^{\alpha(e^{it}-1)}$

Der folgende Satz zeigt, wie man aus einer charakteristischen Funktion auf eindeutige Weise die zu Grunde liegende Verteilung zurückgewinnen kann.

16.11 Satz (Umkehrformel und Eindeutigkeitssatz)
Seien μ und ν Verteilungen auf $\mathbb{R}$.

a) Ist $\mu(\{a\}) = \mu(\{b\}) = 0$, so ist

$$\mu([a,b]) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\, \varphi_\mu(t)\,dt\ .$$

b) Ist $\varphi_\mu = \varphi_\nu$, so ist auch $\mu = \nu$.

c) Ist $\int_{-\infty}^{\infty} |\varphi_\mu(t)|\,dt < \infty$, so hat μ die Dichte $f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_\mu(t)\,dt$ zum Lebesgue-Maß.

Bevor wir diesen Satz beweisen, halten wir zunächst zwei einfache Korollare fest:

16.12 Korollar *Ist $\varphi_\mu(t) = e^{-\frac{t^2}{2}}$, so ist $\mu = \mathcal{N}(0,1)$.*

16.13 Korollar *μ symmetrisch $\Longleftrightarrow \forall\, t \in \mathbb{R}:\ \varphi_\mu(t) \in \mathbb{R}$.*

Beweis: Sei $\tilde\mu$ die durch $\tilde\mu(A) = \mu(-A)$ definierte Verteilung auf $\mathbb{R}$. Es ist $\varphi_{\tilde\mu}(t) = \int e^{-itx}\,d\mu(x) = \int \overline{e^{itx}}\,d\mu(x) = \overline{\int e^{itx}\,d\mu(x)} = \overline{\varphi_\mu(t)}$, so dass

$$\mu \text{ symmetrisch} \quad\Longleftrightarrow\quad \mu = \tilde\mu \quad\overset{\text{Satz}16.11}{\Longleftrightarrow}\quad \varphi_\mu = \varphi_{\tilde\mu} = \overline{\varphi_\mu} \quad\Longleftrightarrow\quad \varphi_\mu \text{ reellwertig}$$

□

Den Beweis von Satz 16.11 bereiten wir durch das folgende Lemma vor.

16.14 Lemma *Für $T > 0$ sei $S(T) := \int_0^T \frac{\sin t}{t}\,dt$. Dann existiert $S(\infty) := \lim_{T\to\infty} S(T)$, es ist $0 < S(\infty) \le \|S\|_\infty < \infty$, und für $\alpha \in \mathbb{R}$ ist*

$$\int_0^T \frac{\sin(\alpha t)}{t}\,dt = \operatorname{sign}(\alpha)\cdot S(|\alpha|\,T)\ . \tag{*}$$

*(*Achtung: *Das heißt nicht, dass $\frac{\sin(\alpha t)}{t}$ über $(0,\infty)$ integrierbar ist! Vergleiche die alternierende harmonische Reihe!)*

Beweis: Beachte, dass für $k = 1, 2, \ldots$

$$\frac{2}{(2k+1)\pi} \leq \int_{2k\pi}^{(2k+1)\pi} \frac{\sin t}{t}\, dt \leq \frac{2}{2k\pi}$$
$$\frac{-2}{(2k+1)\pi} \leq \int_{(2k+1)\pi}^{2(k+1)\pi} \frac{\sin t}{t}\, dt \leq \frac{-2}{2(k+1)\pi}\,.$$

Addiert man diese beiden Zeilen, so folgt

$$0 \leq \int_{2k\pi}^{2(k+1)\pi} \frac{\sin t}{t}\, dt \leq \frac{1}{k(k+1)\pi}$$

so dass $S(T)$ aus dem gleichen Grund wie die alternierende harmonische Reihe konvergiert. Ebenfalls folgt sofort, dass $\|S\|_\infty < \infty$, und (*) folgt durch Substitution, indem man der Reihe nach die Fälle $\alpha = 0$, $\alpha > 0$ und $\alpha = -1$ betrachtet. □

Beweis von Satz 16.11:

a) Beachte zunächst, dass wegen Lemma 16.4

$$\left|\frac{e^{-ita} - e^{-itb}}{it}\right| = \left|\frac{e^{-itb}(e^{it(b-a)} - 1)}{t}\right| \leq \left|\frac{t(b-a)}{t}\right| = |b-a|\,. \qquad (**)$$

Also ist der Integrand im folgenden Integral beschränkt, und man kann den Satz von Fubini anwenden:

$$\begin{aligned}
I(T) &:= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\, \varphi_\mu(t)\, dt \\
&= \frac{1}{2\pi} \int_{-T}^{T} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it}\, e^{itx}\, d\mu(x)\, dt \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \left[\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\, dt\right] d\mu(x)
\end{aligned}$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \Bigg[\frac{1}{i} \underbrace{\int_{-T}^{T} \frac{\cos(t(x-a))}{t} - \frac{\cos(t(x-b))}{t}\, dt}_{=0\text{, da der Integrand ungerade in } t \text{ ist}}$$
$$+ \underbrace{\int_{-T}^{T} \frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t}\, dt}_{\text{Integrand gerade}}\Bigg]\, d\mu(x)$$

$$\begin{aligned}
&= \frac{1}{2\pi} \cdot 2 \cdot \int_{\mathbb{R}} \left[\int_0^T \frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t}\right] d\mu(x) \\
&= \frac{1}{\pi} \int_{\mathbb{R}} \Big[\operatorname{sign}(x-a) \cdot S(|x-a|\,T) - \operatorname{sign}(x-b) \cdot S(|x-b|\,T)\Big]\, d\mu(x) \\
&\overset{T\to\infty}{\to} \frac{1}{\pi}\, S(\infty) \cdot \int_{\mathbb{R}} (\operatorname{sign}(x-a) - \operatorname{sign}(x-b))\, d\mu(x)
\end{aligned}$$

nach dem Satz von der majorisierten Konvergenz, da $S(|x-a|\,T), S(|x-b|\,T) \to S(\infty)$ μ-f.s. im Limes $T \to \infty$ und $\|S\|_\infty < \infty$. Da $\operatorname{sign}(x-a) - \operatorname{sign}(x-b) = 1_{\{a\}}(x) + 2 \cdot 1_{(a,b)}(x) + 1_{\{b\}}(x)$ und da $\mu(\{a\}) = \mu(\{b\}) = 0$, folgt

$$\lim_{T\to\infty} I(T) = \frac{S(\infty)}{\pi}\, (\mu(\{a\}) + 2\,\mu((a,b)) + \mu(\{b\})) = \frac{2}{\pi} S(\infty) \cdot \mu([a,b])$$

Um den Beweis von Teil a) zu beenden, bleibt zu zeigen, dass $S(\infty) = \frac{\pi}{2}$. Vorher wenden wir uns Aussage c) zu.

c) Sei $\int_{-\infty}^{\infty} |\varphi_\mu(t)|\, dt < \infty$. Dann ist wegen (**)

$$\left| \frac{e^{-ita} - e^{-itb}}{it} \varphi_\mu(t) \right| \leq |b-a| \cdot |\varphi_\mu(t)| \in \mathcal{L}^1_{\text{Leb}}$$

und aus dem Satz von der majorisierten Konvergenz und dem Satz von Fubini folgt

$$\begin{aligned}
\mu([a,b]) &= \frac{\pi}{2S(\infty)} \cdot \lim_{T\to\infty} I(T) = \frac{1}{4S(\infty)} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\mu(t)\, dt \\
&= \frac{1}{4S(\infty)} \int_{-\infty}^{\infty} \left[\int_a^b e^{-itx}\, dx \right] \varphi_\mu(t)\, dt \\
&= \frac{1}{4S(\infty)} \int_a^b \left[\int_{-\infty}^{\infty} e^{-itx} \varphi_\mu(t)\, dt \right] dx
\end{aligned}$$

d.h. μ hat die behauptete Dichte, falls $S(\infty) = \frac{\pi}{2}$.

a') Wir zeigen nun, dass $S(\infty) = \frac{\pi}{2}$. Sei dazu $\mu = \mathcal{N}(0,1)$, also $\varphi_\mu(t) = e^{-\frac{t^2}{2}} \in \mathcal{L}^1_{\text{Leb}}$. Daher ist nach dem soeben gezeigten

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} &= \frac{1}{4S(\infty)} \cdot \int_{-\infty}^{\infty} e^{-itx} e^{-\frac{t^2}{2}}\, dt \\
&= \frac{\pi}{2S(\infty)} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{itx} e^{-\frac{t^2}{2}}\, dt \\
&= \frac{\pi}{2S(\infty)} \cdot \frac{1}{\sqrt{2\pi}} \cdot \varphi_\mu(x) \\
&= \frac{\pi}{2S(\infty)} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} ,
\end{aligned}$$

also $S(\infty) = \frac{\pi}{2}$.

b) Seien μ, ν Verteilungen auf $\mathbb{R}$ mit $\varphi_\mu = \varphi_\nu$. Sei $N := \{x \in \mathbb{R} : \mu\{x\} > 0 \text{ oder } \nu\{x\} > 0\}$. N ist höchstens abzählbar, und für jedes Intervall (a,b) gibt es $a_n, b_n \in \mathbb{R} \setminus N$ mit $[a_n, b_n] \nearrow (a,b)$. Daher ist

$$\mu((a,b)) = \sup_n \mu([a_n, b_n]) \overset{\text{Teil a)}}{=} \sup_n \nu([a_n, b_n]) = \nu((a,b)) \ ,$$

also $\mu = \nu$.

□

Um aus $\lim_{n\to\infty} \varphi_{S_n^*} = \varphi_{\mathcal{N}(0,1)}$ die Konvergenz $P_{S_n^*} \Longrightarrow \mathcal{N}(0,1)$, d.h. den *Zentralen Grenzwertsatz*, schließen zu können, benötigen wir noch den folgenden Satz.

16.15 Satz (Stetigkeitssatz)
Seien μ, μ_n Wahrscheinlichkeitsmaße auf $\mathbb{R}$. Dann gilt:

$$\mu_n \Longrightarrow \mu \quad \textit{genau dann, wenn} \quad \lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t) \textit{ für alle } t \in \mathbb{R}.$$

Daraus folgt nun sofort unter Berücksichtigung von Korollar 16.9 der Zentrale Grenzwertsatz.

16.16 Satz (Zentraler Grenzwertsatz)
Seien $X_1, X_2, \dots$ unabhängige identisch verteilte Zufallsvariablen mit $m := E[X_i]$ und $\sigma^2 := V(X_i) < \infty$. Sei

$$S_n^* := \frac{1}{\sqrt{n\sigma^2}} \sum_{k=1}^{n} (X_k - m)$$

die zentrierte und standardisierte Partialsumme *von $X_1, \dots, X_n$. Dann konvergiert*

$$P_{S_n^*} \Longrightarrow \mathcal{N}(0,1) \, .$$

Beweis von Satz 16.15: Konvergiere $\mu_n \Longrightarrow \mu$. Da $\Re e^{itx} = \cos(tx)$ und $\Im e^{itx} = \sin(tx)$ für jedes $t \in \mathbb{R}$ beschränkte stetige Funktionen in $x \in \mathbb{R}$ sind, folgt $\lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t)$ aus der Definition der schwachen Konvergenz.

Der Beweis der Umkehrung beruht auf den beiden folgenden Lemmata.

16.17 Lemma *Existiert $g(t) := \lim_{n\to\infty} \varphi_{\mu_n}(t)$ für alle $t \in \mathbb{R}$ und ist g stetig bei 0, so ist $(\mu_n)_n$ straff.*

16.18 Lemma *Existiert $g(t) := \lim_{n\to\infty} \varphi_{\mu_n}(t)$ für alle $t \in \mathbb{R}$ und ist $(\mu_n)_n$ straff, so ist $g(t)$ charakteristische Funktion eines Wahrscheinlichkeitsmaßes ν auf $\mathbb{R}$ und $\mu_n \Longrightarrow \nu$.*

Gelte nun $\lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t)$ für alle $t \in \mathbb{R}$ in Satz 16.15. Da φ_μ stetig ist (Lemma 16.2b), folgt aus Lemma 16.17, dass $(\mu_n)_n$ straff ist, und aus Lemma 16.18, dass es ein Wahrscheinlichkeitsmaß ν gibt mit $\mu_n \Longrightarrow \nu$ und $\varphi_\nu(t) = \lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t)$ für alle $t \in \mathbb{R}$. Aus dem Eindeutigkeitssatz 16.11 folgt schließlich $\nu = \mu$ und daher $\mu_n \Longrightarrow \mu$. □

Beweis von Lemma 16.17: Aus dem Satz von Fubini folgt für $u > 0$

$$\begin{aligned}
\frac{1}{u}\int_{-u}^{u} (1 - \varphi_{\mu_n}(t))\, dt &= \int_{-\infty}^{\infty} \left[\frac{1}{u} \int_{-u}^{u} \left(1 - e^{itx}\right) dt \right] d\mu_n(x) \\
&= \int_{-\infty}^{\infty} \left[\frac{1}{u} \int_{-u}^{u} (1 - \cos(tx))\, dt \right] d\mu_n(x) - i \int_{-\infty}^{\infty} \underbrace{\left[\frac{1}{u} \int_{-u}^{u} \sin(tx)\, dt \right]}_{=0} d\mu_n(x) \\
&= \int_{-\infty}^{\infty} \left[\frac{1}{u} \left(t - \frac{1}{x}\sin(tx) \right) \right]_{t=-u}^{u} d\mu_n(x) = \int_{-\infty}^{\infty} \left(2 - 2\frac{\sin(ux)}{ux} \right) d\mu_n(x) \qquad (*) \\
&\geq 2 \cdot \int_{\{|x| \geq \frac{2}{u}\}} \underbrace{\left(1 - \frac{\sin(ux)}{ux} \right)}_{\geq \frac{1}{2}} d\mu_n(x) \geq \mu_n \left\{ x \in \mathbb{R} : |x| \geq \frac{2}{u} \right\} .
\end{aligned}$$

Sei nun $\epsilon > 0$ beliebig. Da $g(0) = \lim_{n\to\infty} \varphi_{\mu_n}(0) = 1$ und da $g(t)$ bei $t = 0$ stetig ist, gibt es zu $\epsilon > 0$ ein $u > 0$ mit

$$\left| \frac{1}{u} \int_{-u}^{u} (1 - g(t))\, dt \right| < \epsilon \, .$$

Daher folgt aus (*) und dem Satz von der majorisierten Konvergenz, dass

$$\limsup_{n\to\infty} \mu_n \left\{ x \in \mathbb{R} : |x| \geq \frac{2}{u} \right\} \leq \limsup_{n\to\infty} \frac{1}{u} \int_{-u}^{u} (1 - \varphi_{\mu_n}(t))\, dt \leq \left| \frac{1}{u} \int_{-u}^{u} (1 - g(t))\, dt \right| < \epsilon \, ,$$

d.h. $(\mu_n)_n$ ist straff. □

Beweis von Lemma 16.18: Da $(\mu_n)_n$ straff ist, hat jede Teilfolge $(\mu_{n_i})_i$ von $(\mu_n)_n$ eine Teilfolge $(\mu_{n_{i(j)}})_j$, die schwach gegen ein ν konvergiert. Aus der schon bewiesenen "⇒"-Richtung von Satz 16.15 folgt für jedes solche Limes-Maß ν

$$\varphi_\nu(t) = \lim_{j\to\infty} \varphi_{\mu_{n_{i(j)}}}(t) = \lim_{n\to\infty} \varphi_{\mu_n}(t) = g(t) \, .$$

Aus dem Eindeutigkeitssatz 16.11 folgt dann, dass alle Teilfolgen $(\mu_{n_{i(j)}})_j$ gegen das gleiche Maß ν konvergieren, also $\mu_n \Longrightarrow \nu$. □

16.19 Beispiel Sei λ_n die Gleichverteilung auf $[-n, n]$, also

$$\varphi_{\lambda_n}(t) = \frac{1}{2n}\int_{-n}^{n} e^{itx}\,dx = \frac{1}{2nt}\big(\sin(tn) - \sin(-tn)\big) = \frac{\sin(tn)}{tn} \quad (t \neq 0)\ .$$

Sei $g(t) := \lim_{n\to\infty}\varphi_{\lambda_n}(t)$. Dann ist $g(t) = 0$ für $t \neq 0$, während $g(0) = 1$, so dass $g(t)$ bei $t = 0$ unstetig und daher keine charakteristische Funktion ist. Daher ist $(\lambda_n)_n$ nicht schwach konvergent, ja nicht einmal straff.

16.20 Beispiel Sei $\alpha > 0$ und seien $X_1, \ldots, X_n$ unabhängig identisch verteilt mit Bernoulli-Verteilung zum Parameter $p = \frac{\alpha}{n}$, $S_n := X_1 + \cdots + X_n$. Dann ist

$$\varphi_{X_i}(t) = pe^{it} + (1-p) = 1 + \frac{\alpha}{n}\left(e^{it} - 1\right)\ ,$$

also

$$\varphi_{S_n}(t) = (\varphi_{X_i}(t))^n \to e^{\alpha(e^{it}-1)} \qquad \text{mit } n \to \infty.$$

Dieser Limes ist die charakteristische Funktion der Poisson-Verteilung $P(\alpha)$ zum Parameter α, siehe Beispiel 16.10. Also konvergiert $P_{S_n} \Longrightarrow P(\alpha)$.

Es gibt eine große Zahl von Verallgemeinerungen und Verschärfungen des Zentralen Grenzwertsatzes. Einige davon sollen im Folgenden zwar diskutiert, aber nicht bewiesen werden. [1] Seien also $X_1, X_2, \ldots$ Zufallsvariablen mit $m_k := EX_k$ und $\sigma_k^2 := V(X_k) < \infty$. Sei $V_n := V(X_1 + \cdots + X_n)$. Mit

$$S_n^* := \frac{1}{\sqrt{V_n}}\sum_{k=1}^{n}(X_k - m_k)$$

bezeichnen wir wieder die zentrierte und normalisierte Partialsumme von $X_1, \ldots, X_n$. Um Verteilungskonvergenz $P_{S_n^*} \Longrightarrow \mathcal{N}(0,1)$ nachzuweisen reicht es wegen des Stetigkeitssatzes 16.15 zu zeigen, dass $\lim_{n\to\infty}\varphi_{S_n^*}(t) = \varphi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}}$. Als Muster dafür (im einfachsten Fall) dient der Beweis von Satz 16.7.

Sind die X_i nicht unabhängig, so lässt sich schon der erste Schritt,

$$\varphi_{S_n^*}(t) = \prod_{k=1}^{n}\varphi_{X_k - m_k}\left(\frac{t}{\sqrt{V_n}}\right) \tag{*}$$

nicht übertragen. Trotzdem kann man die Unabhängigkeitsannahme abschwächen ("mischende stationäre Prozesse" lautet das relevante Stichwort) und einen Zentralen Grenzwertsatz beweisen.

Im Folgenden nehmen wir an, dass die X_i unabhängig sind und dass $\sigma_k^2 > 0$ für alle k. Dann ist $V_n = \sigma_1^2 + \cdots + \sigma_n^2$, und es gilt die Identität (*). Da die zweiten Momente der X_k endlich sind, gibt es nach Satz 16.5 Funktionen $\beta_k : \mathbb{R} \to \mathbb{R}$ derart, dass $\lim_{t\to 0}\beta_k(t) = 0$ und

$$\varphi_{X_k - m_k}(t) = 1 - \frac{t^2}{2}\left(\sigma_k^2 + \beta_k(t)\right) \quad \text{im Limes } t \to 0\ . \tag{**}$$

Beachte nun, dass $-u - u^2 \leq \log(1-u) \leq -u$ für $|u| \leq 0.4$.[2] Geht dann $V_n \to \infty$ für $n \to \infty$, so folgt aus (*) und (**) für hinreichend große n

$$\log\varphi_{S_n^*}(t) = \sum_{k=1}^{n}\log\left(1 - \frac{t^2}{2V_n}\left(\sigma_k^2 + \beta_k\Big(\frac{t}{\sqrt{V_n}}\Big)\right)\right) \leq -\frac{t^2}{2} - \frac{t^2}{2V_n}\sum_{k=1}^{n}\beta_k\Big(\frac{t}{\sqrt{V_n}}\Big)$$

[1] Details findet man z.B. in Kapitel 27 von [Billingsley].

[2] Beachte, dass $\log(1-u) = -u - \frac{u^2}{2} - \sum_{k=3}^{\infty}\frac{u^k}{k}$.

und

$$\log \varphi_{S_n^*}(t) \geq -\frac{t^2}{2} - \frac{t^2}{2V_n}\sum_{k=1}^n \beta_k\Big(\frac{t}{\sqrt{V_n}}\Big) - \frac{t^4}{4V_n^2}\sum_{k=1}^n \Big(\sigma_k^2 + \beta_k\Big(\frac{t}{\sqrt{V_n}}\Big)\Big)^2$$

Ist nun $\lim_{t\to 0}\beta_k(t) = 0$ *gleichmäßig in* k (das gilt, falls die Familie $(X_k^2)_{k>0}$ gleichgradig integrierbar ist) und sind die Summanden σ_k^2 von V_n nicht zu verschieden (es reicht, dass $\min_k \sigma_k^2 > 0$; dann geht automatisch $V_n \to \infty$ mit $n \to \infty$), so zeigt man leicht, dass

$$\lim_{n\to\infty} \log \varphi_{S_n^*}(t) = -\frac{t^2}{2} \quad \text{für alle } t \in \mathbb{R}.$$

Durch ein ähnliches Argument erhält man z.B. mit der obigen Notation:

16.21 Satz (Zentraler Grenzwertsatz unter der Lyapunov-Bedingung)
Seien $X_1, X_2, \ldots$ unabhängig mit $0 < \sigma_k^2 < \infty$ für alle k. Ist die Lyapunov-Bedingung *erfüllt, d.h. gibt es ein $r > 2$ mit*

$$\lim_{n\to\infty} \frac{1}{V_n^{r/2}} \sum_{k=1}^n E\left[|X_k - EX_k|^r\right] = 0\,,$$

so gilt $P_{S_n^*} \Longrightarrow \mathcal{N}(0,1)$.

Die schärfste Aussage dieser Art ist

16.22 Satz (Satz von Lindeberg)
Seien $X_1, X_2, \ldots$ unabhängig mit $0 < \sigma_k^2 < \infty$ für alle k. Ist die Lindeberg-Bedingung *erfüllt, d.h. gilt für jedes $\delta > 0$*

$$\lim_{n\to\infty} \frac{1}{V_n} \sum_{k=1}^n E\left[(X_k - EX_k)^2 \cdot 1_{\{(X_k - EX_k)^2 \geq \delta V_n\}}\right] = 0\,,$$

so gilt $P_{S_n^*} \Longrightarrow \mathcal{N}(0,1)$.

In der Tat ist die Lindeberg-Bedingung fast schon äquivalent zur Gültigkeit des Zentralen Grenzwertsatzes. Sie folgt aus der Konvergenz $P_{S_n^*} \Longrightarrow \mathcal{N}(0,1)$ und der zusätzlichen Bedingung

$$\forall\, \epsilon > 0: \quad \lim_{n\to\infty} \max_{k=1,\ldots,n} P\left\{\frac{X_k^2}{V_n} \geq \epsilon\right\} = 0\,.$$

Eine Verschärfung des Zentralen Grenzwertsatzes liefert der Satz von Berry-Esséen, der die Konvergenzgeschwindigkeit im Zentralen Grenzwertsatz abschätzt. [3]

16.23 Satz (Berry-Esséen)
Seien $X_1, X_2, \ldots$ unabhängig mit $0 < \sigma_k^2 < \infty$ für alle k. Bezeichne F_n die Verteilungsfunktion von S_n^* und Φ die von $\mathcal{N}(0,1)$. Dann gilt für alle $n \in \mathbb{N}$:

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{6}{V_n^{3/2}} \cdot \sum_{k=1}^n E\left[|X_k - EX_k|^3\right]\,.$$

Im Fall unabhängig identisch verteilter X_k wird die rechte Seite zu $\mathrm{const}\cdot n^{-\frac{1}{2}}$.

[3] Siehe z.B. Satz 4.2.10 in [Gänssler-Stute].

Der Beweis basiert auf folgender Abschätzung, die für $T^{-1} = \frac{9}{8V_n^3} \sum_{k=1}^n E\left[|X_k - EX_k|^3\right]$ den Anschluss an obige Überlegungen liefert:

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{1}{\pi} \int_{-T}^{T} \left| \varphi_{S_n^*}\left(\frac{t}{\sqrt{V_n}}\right) - e^{-\frac{t^2}{2}} \right| \frac{dt}{|t|} + \frac{24}{T\pi\sqrt{2\pi}} \,,$$

Kapitel 17

Normalverteilung und zentraler Grenzwertsatz auf dem $\mathbb{R}^d$

Die Idee der Normalverteilung lässt sich auf $\mathbb{R}^d$-wertige Zufallsvariablen verallgemeinern. Der einfachste Fall liegt vor, wenn man einen Zufallsvektor aus d unabhängigen $\mathbb{R}$-wertigen normalverteilten Zufallsvariablen betrachtet. Mit Verallgemeinerungen dieser Situation befassen wir uns in diesem Kapitel.

Sei V eine reelle, symmetrische, positiv definite $d \times d$-Matrix, insbesondere $\det(V) \neq 0$. V lässt sich diagonalisieren als

$$V = U^T \Lambda U, \quad U \text{ Orthogonalmatrix, d.h.} U^T = U^{-1}, \quad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_d),$$

wobei $\lambda_1, \ldots, \lambda_d > 0$ die Eigenwerte von V sind.

Definiere $f_V : \mathbb{R}^d \to \mathbb{R}$,

$$f_V(x) := \frac{1}{\sqrt{(2\pi)^d \, |\det(V)|}} \cdot \exp\left(-\frac{1}{2} x^T V^{-1} x\right) .$$

Für $d = 1$ und $V = (\sigma^2)$ ist f_V gerade die Dichte zu $\mathcal{N}(0, \sigma^2)$. Allgemein gilt:

- ▷ $f_V \geq 0$,
- ▷ $f_V(M^{-1}x) = |\det(M)| \cdot f_{MVM^T}(x)$ für invertierbare reelle $d \times d$-Matrizen M, da $\det(MVM^T)/\det(V) = \det(M)^2$, so dass $f_V(U^{-1}x) = f_\Lambda(x)$, also
- ▷ $\int_{\mathbb{R}^d} f_V(x)\,dx = \int_{\mathbb{R}^d} f_V(U^{-1}x)\,dx = \int_{\mathbb{R}^d} f_\Lambda(x)\,dx = \prod_{i=1}^d \frac{1}{\sqrt{2\pi\lambda_i}} \int_{\mathbb{R}} e^{-\frac{x_i^2}{2\lambda_i}}\,dx_i = 1$, wobei wir den Integraltransformationssatz und die Tatsache $\det(V) = \det(\Lambda) = \prod_{i=1}^d \lambda_i$ benutzt haben.

Daher ist f_V eine Wahrscheinlichkeitsdichte (zum Lebesgue-Maß) auf $\mathbb{R}^d$.

17.1 Definition (Normalverteilung auf $\mathbb{R}^d$) *Die Verteilung mit Dichte f_V auf $\mathbb{R}^d$ heißt d-dimensionale Normalverteilung mit Erwartung 0 und Kovarianzmatrix V, kurz $\mathcal{N}(0, V)$.*

Diese Sprechweise wird durch folgenden Satz gerechtfertigt.

17.2 Satz (Eigenschaften der Normalverteilung auf $\mathbb{R}^d$)
Sei $X = (X_1, \ldots, X_d)^T$ eine $\mathbb{R}^d$-wertige Zufallsvariable, deren Verteilung Dichte f_V hat.

a) Ist $a \in \mathbb{R}^d$, so ist $P_{a^T X} = \mathcal{N}(0, a^T V a)$.

b) $Cov(X_i, X_j) = v_{ij}$ $(i, j = 1, \ldots, d)$.

c) $X_1, \ldots, X_d$ sind unabhängig genau dann, wenn $V = \Lambda$.

Beweis:

a) Wir bestimmen die charakteristische Funktion von $a^T X$:

$$\begin{aligned}
E\left[e^{ita^T X}\right] &= \int_{\mathbb{R}^d} e^{ita^T x} f_V(x)\,dx = \int_{\mathbb{R}^d} e^{ita^T U^T x} f_V(U^{-1}x)\,dx \quad (U^T = U^{-1}!)\\
&= \int_{\mathbb{R}^d} e^{it(Ua)^T x} f_\Lambda(x)\,dx = \prod_{j=1}^d \int_{\mathbb{R}} e^{it(Ua)_j^T x_j} \frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{x_j^2}{2\lambda_j}}\,dx_j\\
&= \prod_{j=1}^d \exp\left(-\frac{1}{2}t^2\left((Ua)_j^T\right)^2 \lambda_j\right) \qquad \text{(Beispiel 16.10)}\\
&= \exp\left(-\frac{1}{2}t^2\sum_{j=1}^d\left((Ua)_j^T\right)^2 \lambda_j\right) = \exp\left(-\frac{1}{2}t^2(Ua)^T\Lambda(Ua)\right)\\
&= \exp\left(-\frac{1}{2}t^2 a^T V a\right)\\
&= \varphi_{\mathcal{N}(0,a^T V a)}(t)
\end{aligned}$$

b) Im Folgenden steht $E[Z]$, wenn Z eine Matrix mit zufälligen Koeffizienten ist, für die Matrix der koeffizientenweisen Erwartungswerte.

$$\begin{aligned}
(Cov(X_i,X_j))_{ij} = E[XX^T] &= \int_{\mathbb{R}^d} xx^T f_V(x)\,dx = \int_{\mathbb{R}^d} U^T x(U^T x)^T f_V(U^{-1}x)\,dx\\
&= \int_{\mathbb{R}^d} U^T xx^T U f_\Lambda(x)\,dx = U^T\cdot\int_{\mathbb{R}^d} xx^T f_\Lambda(x)\,dx\cdot U = U^T\Lambda U = V\ ,
\end{aligned}$$

denn

$$\begin{aligned}
\int_{\mathbb{R}^d} x_i x_j f_\Lambda(x)\,dx &= \int_{\mathbb{R}}\dots\int_{\mathbb{R}} x_i x_j \prod_{k=1}^d \frac{1}{\sqrt{2\pi\lambda_k}} e^{-\frac{x_k^2}{2\lambda_k}}\,dx_1\dots dx_d\\
&= \begin{cases} \int_{\mathbb{R}} x_i \frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{x_i^2}{2\lambda_i}}\,dx_i \int_{\mathbb{R}} x_j\cdot\frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{x_j^2}{2\lambda_j}}\,dx_j = 0 & (i\neq j)\\ \int_{\mathbb{R}} x_i^2 \frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{x_i^2}{2\lambda_i}}\,dx_i = \lambda_i & (i=j)\end{cases}
\end{aligned}$$

c)

$$X_1,\dots,X_d \text{ unabhängig} \iff f_V(x) = \prod_{j=1}^d \frac{1}{\sqrt{2\pi v_{jj}}} e^{-\frac{x_j^2}{2v_{jj}}} \iff f_V = f_\Lambda \overset{V\text{ symm.}}{\iff} V = \Lambda$$

□

Das Transformationsverhalten d-dimensionaler Normalverteilungen beschreibt der folgende Satz.

17.3 Satz (Lineare Transformationen mehrdimensionaler Normalverteilungen)
Sei $d' \leq d$ und M eine $d'\times d$-Matrix mit vollem Rang d'. Ist dann X nach $\mathcal{N}(0,V)$ verteilt, so ist MX eine nach $\mathcal{N}(0,MVM^T)$ verteilte $\mathbb{R}^{d'}$-wertige Zufallsvariable.

Beweis: Wir beweisen diesen Satz in drei Schritten.
1. Schritt: $d' = d$. Dann ist M invertierbar und für messbares $B\subseteq\mathbb{R}^d$ ist

$$\begin{aligned}
P(MX\in B) &= \int 1_B(Mx) f_V(x)\,dx = \int 1_B(y) f_V(M^{-1}y)\,|\det(M^{-1})|\,dy\\
&= \int 1_B(y) f_{MVM^T}(y)\,dy
\end{aligned}$$

2. Schritt: $M = (\tilde{M}, 0)$ für eine invertierbare $d' \times d'$-Matrix $\tilde{M}$. Weitere Notation:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},\ X_1\ \mathbb{R}^{d'}\text{-wertig},\quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},\ V_{11} \text{ eine } d' \times d'\text{-Matrix.}$$

Dann ist $MX = \tilde{M}X_1$, also $P_{MX} = P_{\tilde{M}X_1} = \mathcal{N}(0, \tilde{M}V_{11}\tilde{M}^T) = \mathcal{N}(0, MVM^T)$ nach Schritt 1.

3. Schritt: M sei nun beliebig. Dann gibt es eine orthogonale $d \times d$-Matrix R derart, dass $MR^{-1} =: N = (\tilde{N}, 0)$. [1] Nach Schritt 1 ist $P_{RX} = \mathcal{N}(0, RVR^T)$, nach Schritt 2 daher

$$P_{MX} = P_{NRX} = \mathcal{N}(0, NRVR^TN^T) = \mathcal{N}(0, MVM^T)\ .$$

□

Nun kommt der folgende Zentrale Grenzwertsatz für $\mathbb{R}^d$-wertige Zufallsvariablen nicht überraschend.

17.4 Satz (Zentraler Grenzwertsatz im $\mathbb{R}^d$)
Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte $\mathbb{R}^d$-wertige Zufallsvariablen mit Erwartungswertvektor 0 und Kovarianzmatrix V. Dann konvergiert

$$P_{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i} \Longrightarrow \mathcal{N}(0, V)\ .$$

Der Beweis benutzt die folgende Charakterisierung von Wahrscheinlichkeitsmaßen durch ihre eindimensionalen linearen Projektionen.

17.5 Satz (Cramér-Wold-device)
Seien X und Y Zufallsvariablen mit Werten im $\mathbb{R}^d$. Dann gilt $P_X = P_Y$ genau dann, wenn $P_{a^TX} = P_{a^TY}$ für alle $a \in \mathbb{R}^d$.

Die "⇒"-Richtung ist trivial. Der Beweis der Umkehrung beruht auf dem Eindeutigkeitssatz für charakteristische Funktionen $\mathbb{R}^d$-wertiger Zufallsvariablen [2], denn aus $P_{a^TX} = P_{a^TY}$ für alle $a \in \mathbb{R}^d$ folgt $E[e^{\imath a^T X}] = E[e^{\imath a^T Y}]$ für alle $a \in \mathbb{R}^d$.

Beweis von Satz 17.4: Da die X_i identisch verteilt sind, ist die Konstante $c := \sum_{j=1}^d E[X_{i,j}^2]$ für alle i dieselbe. Sei $Z_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$. Für die euklidische Länge $|Z_n|_2$ des zufälligen Vektors Z_n gilt:

$$E[|Z_n|_2^2] = \sum_{j=1}^d E[Z_{n,j}^2] = \sum_{j=1}^d \frac{1}{n}\sum_{i=1}^n E[X_{i,j}^2] = \frac{1}{n}\sum_{i=1}^n c = c$$

Daher folgt aus der Chebyshev-Ungleichung für jedes $r > 0$

$$P\left(|Z|_2 > r\right) \leq \frac{c}{r^2}$$

und die Folge der Verteilungen P_{Z_n} ist straff, d.h. relativ folgenkompakt bzgl. der schwachen Konvergenz (siehe Satz 15.14). Zu zeigen bleibt daher, dass jeder schwache Häufungspunkt von P_{Z_n} mit der Normalverteilung $\mathcal{N}(0, V)$ übereinstimmt. Gelte also $Z_{n_i} \Rightarrow Z$. Dann folgt $a^T Z_{n_i} \Rightarrow a^T Z$ für alle $a \in \mathbb{R}^d$. Andererseits folgt aus dem eindimensionalen Zentralen Grenzwertsatz

$$a^T Z_{n_i} = \frac{1}{\sqrt{n}}\sum_{i=1}^n a^T X_i \Rightarrow \mathcal{N}(0, a^T V a)\ ,$$

denn $V(a^T X_i) = a^t V a$ nach Satz 7.18. Also ist $P_{a^TZ} = \mathcal{N}(0, a^T V a)$, d.h. $P_{a^TZ} = P_{a^TY}$ falls $P_Y = \mathcal{N}(0, V)$, für alle $a \in \mathbb{R}^d$. Mit dem Cramér-Wold-device schließt man, dass $P_Z = P_Y = \mathcal{N}(0, V)$. □

[1] R transformiert die Zerlegung $\mathbb{R}^d = \ker(M)^\perp + \ker(M)$ nach $\mathbb{R}^{d'} \oplus \mathbb{R}^{d-d'}$.
[2] Siehe z.B. [Gänssler-Stute, Satz 8.7.1]

Wir beschließen dieses Kapitel mit einer Anwendung von Satz 17.3 auf die Statistik normalverteilter Beobachtungen. Seien zunächst $X_1, \ldots, X_n$ u.i.v. Wir bezeichnen das **empirische Mittel** der $\mathcal{X}_i$ mit

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

und ihre **empirische Varianz** mit

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Beachte, dass $E[\bar{X}] = E[X_1]$ und

$$E[S^2] = \frac{1}{n-1} \left(\sum_{i=1}^{n} E[X_i^2] - \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] \right) = V[X_1].$$

Sind die X_i normalverteilt, so gilt

17.6 Satz (Unabhängigkeit von $\bar{X}$ und S^2 bei normalverteilten Stichproben)
Seien $X_1, \ldots, X_n$ u.i.v. nach $\mathcal{N}(\mu, \sigma^2)$. Dann sind $\bar{X}$ und S^2 unabhängig, $\bar{X}$ ist $\mathcal{N}(\mu, \frac{\sigma^2}{n})$-verteilt, und $\frac{n-1}{\sigma^2} S^2$ ist χ^2_{n-1}-verteilt. (χ^2_{n-1} ist die Verteilung von $Z_1^2 + \cdots + Z_n^2$, wenn die Z_i u.i.v. nach $\mathcal{N}(0,1)$ sind.)

Beweis: Sei $a^{(1)}, \ldots, a^{(n)}$ das folgende Orthonormalsystem in $\mathbb{R}^n$:

$$\begin{aligned} a^{(1)} &= \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}} \right) \\ a^{(k)} &= \left(-\frac{1}{\sqrt{k(k-1)}}, \ldots, -\frac{1}{\sqrt{k(k-1)}}, \sqrt{\frac{k-1}{k}}, 0, \ldots, 0 \right) \end{aligned}$$

für $k = 2, \ldots, n$, wobei $a^{(k)}_{k+1}, \ldots, a^{(k)}_n = 0$. Bezeichne A die Matrix mit den Zeilen $a^{(1)}, \ldots, a^{(n)}$ und $\mathbf{X} = (X_1, \ldots, X_n)^T$. Dann ist $\mathbf{Y} := \mathbf{AX} = \mathbf{A}(\mathbf{X} - \mu\mathbf{1}) + \mu\mathbf{A1}$ nach $\mathcal{N}(\mu\mathbf{A1}, \mathbf{A}\sigma^2 Id \mathbf{A}^T) = \mathcal{N}(\mu\mathbf{A1}, \sigma^2 Id)$ verteilt (Satz 17.3), so dass die Y_i unabhängig normalverteilt sind. Da $Y_1 = (a^{(1)})\mathbf{X} = \sqrt{n}\,\bar{X}$ und

$$\sum_{i=2}^{n} Y_i^2 = \mathbf{Y}^T\mathbf{Y} - Y_1^2 = \mathbf{X}^T\mathbf{X} - n\bar{X}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 = (n-1)S^2,$$

sind $\bar{X}$ und S^2 unabhängig.

Aus $\mathbf{A1} = (\sqrt{n}, 0, \ldots, 0)^T$ folgt dann $\bar{X} = \frac{1}{\sqrt{n}} Y_1 \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, und zusammen mit $\frac{n-1}{\sigma^2} S^2 = \sum_{i=1}^{n-1} (\frac{Y_i}{\sigma})^2$ auch $\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$. □

Ein analoges Ergebnis für d-dimensional normalverteilte ZV'en gilt ebenfalls.

Eine Konsequenz dieses Satzes und der Formeln für die Momente der Normalverteilung ist, dass

$$E[\bar{X}] = \mu,\ V[\bar{X}] = \frac{\sigma^2}{n} \to 0 \quad \text{und} \quad E[S^2] = \sigma^2,\ V[S^2] = \frac{2\sigma^4}{n-1} \to 0 \quad \text{mit } n \to \infty,$$

so dass $\bar{X}$ und S^2 erwartungstreue und konsistente Schätzer für μ bzw. σ^2 sind.

Kapitel 18

Zerlegungssätze für Maße

Ziel dieses Kapitels ist es, zu zwei gegebenen σ-endlichen Maßen μ und ν auf einem messbaren Raum $(\Omega, \mathcal{A})$ eine messbare Menge $A \in \mathcal{A}$ mit folgenden Eigenschaften zu finden: 1) $\mu(A^c) = 0$ und 2) $\nu_{|A}$ hat Dichte bzgl. μ.

Zunächst wiederholen wir ein paar einfache Tatsachen aus vorangegangenen Kapiteln. Sei $(\Omega, \mathcal{A})$ ein messbarer Raum.

- ▷ Ist $h : \Omega \to [0, +\infty]$ $\mathcal{A}$-messbar, so wird durch $\nu(A) := \int_A h\, d\mu$ $(A \in \mathcal{A})$ ein Maß auf $\mathcal{A}$ definiert. ν ist endlich genau dann, wenn $h \in \mathcal{L}^1_\mu$. (Korollar 8.3) Man sagt: *ν hat Dichte h bzgl. μ*, kurz:

$$\nu = h\mu \quad \text{und} \quad h = \frac{d\nu}{d\mu}\,.$$

- ▷ Ist $\nu = h\mu$, so gilt für messbare $f : \Omega \to [0, +\infty]$

$$\int f\, d\nu = \int f h\, d\mu \tag{*}$$

  Für beliebige messbare $f : \Omega \to \mathbb{R}$ gilt daher: $f \in \mathcal{L}^1_\nu \iff fh \in \mathcal{L}^1_\mu$, und für jedes $f \in \mathcal{L}^1_\nu$ gilt (*) ebenfalls. (Diese Aussagen wurden nicht explizit als Satz bewiesen, folgen aber leicht durch "algebraische Induktion", vergl. den Beweis zu Satz 7.20.

- ▷ Beispiele sind u.a.:

  a) Normalverteilung auf $\mathbb{R}$: μ Lebesgue-Maß auf $\mathbb{R}$, $h(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$, $m \in \mathbb{R}$, $\sigma^2 > 0$, siehe Beispiel 13.9.

  b) Exponentialverteilung auf $[0, \infty)$: μ Lebesgue-Maß auf $[0, \infty)$, $h(x) = \alpha e^{-\alpha x}$, $\alpha > 0$, siehe Übung 13.10.

In den uns interessierenden Fällen bestimmt ein Maß seine Dichte eindeutig (so es überhaupt eine hat).

18.1 Satz (Eindeutigkeit der Dichte)
Seien μ, ν σ-endlich. Hat ν sowohl Dichte h_1 als auch Dichte h_2 bzgl. μ, so ist $h_1 = h_2$ μ-f.s. Also ist die messbare Funktion $\frac{d\nu}{d\mu}$ μ-f.s. eindeutig definiert.

Beweis: Da μ und ν (und damit auch $\mu + \nu$) σ-endlich sind, gibt es $A_n \in \mathcal{A}$, $A_n \nearrow \Omega$, $\mu(A_n) + \nu(A_n) < \infty$. Es reicht zu zeigen, dass $h_1 = h_2$ μ-f.s. sicher auf jedem A_n. Sei daher $B_n := A_n \cap \{h_1 \geq h_2\}$. Dann ist $\int_{B_n} (h_1 - h_2)\, d\mu = \nu(B_n) - \nu(B_n) = 0$ und da $1_{B_n}(h_1 - h_2) \geq 0$, folgt $h_1 - h_2 = 0$ μ-f.s. auf B_n, siehe Satz 7.4a. Analog zeigt man $h_1 - h_2 = 0$ μ-f.s. auf $A_n \setminus B_n$, zusammen also $h_1 = h_2$ μ-f.s. auf A_n für jedes n. $\square$

18.2 Bemerkung Ist $\nu = h\mu$, so gilt

$$A \in \mathcal{A}, \mu(A) = 0 \quad \Longrightarrow \quad \nu(A) = 0 \tag{**}$$

Dadurch motiviert definiert man

18.3 Definition (Absolut stetig) *Sind μ, ν zwei Maße auf $(\Omega, \mathcal{A})$ und gilt (**), so heißt ν absolut stetig bzgl. μ, kurz: $\nu \ll \mu$. Man sagt μ und ν sind äquivalent, kurz $\mu \approx \nu$, falls $\mu \ll \nu$ und $\nu \ll \mu$.*

Aus dieser Definition folgt sofort: Hat ν Dichte bzgl. μ, so ist $\nu \ll \mu$. Wichtigstes Ergebnis dieses Kapitels wird sein, dass für σ-endliche Maße auch die Umkehrung davon gilt (Satz von Radon-Nikodym).

In vielen Fällen lässt sich Absolutstetigkeit durch ein ϵ-δ-Kriterium charakterisieren.

18.4 Satz (ϵ-δ-Kriterium)
Seien μ, ν Maße auf $(\Omega, \mathcal{A})$. Gilt

$$\forall \epsilon > 0 \, \exists \delta > 0 \, \forall A \in \mathcal{A} : \ \mu(A) < \delta \ \Rightarrow \ \nu(A) < \epsilon \, , \tag{***}$$

so ist $\nu \ll \mu$. Ist ν ein endliches Maß, so gilt auch die Umkehrung.

Beweis:
"$\Rightarrow$" Sei $A \in \mathcal{A}$ mit $\mu(A) = 0$. Wegen (***) ist $\nu(A) < \epsilon$ für jedes $\epsilon > 0$ und daher $\nu(A) = 0$.

"$\Leftarrow$" Sei ν endlich. Angenommen (***) ist falsch. Dann gibt es ein $\epsilon > 0$ und eine Folge von $A_n \in \mathcal{A}$ $(n \in \mathbb{N})$ mit $\mu(A_n) < 2^{-n}$ aber $\nu(A_n) \geq \epsilon$ für alle n. Betrachte $A := \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$. Dann ist

$$\mu(A) \leq \lim_{n\to\infty} \mu\left(\bigcup_{k\geq n} A_k\right) \leq \lim_{n\to\infty} \sum_{k\geq n} \mu(A_k) \leq \lim_{n\to\infty} \sum_{k\geq n} 2^{-k} = 0$$

aber

$$\nu(A) \overset{\nu \text{ endlich}}{=} \lim_{n\to\infty} \nu\left(\bigcup_{k\geq n} A_k\right) \geq \inf_n \nu(A_n) \geq \epsilon > 0$$

im Widerspruch zu $\nu \ll \mu$.

□

18.5 Beispiel Die Äquivalenz im vorhergehenden Satz gilt ohne die Endlichkeitsannahme an ν nicht. Beispiel: μ =Lebesgue-Maß auf $[0,1]$, $\nu(A) := \int_A \frac{1}{x}\, dx$. Dann ist ν σ-endlich aber nicht endlich, $\mu \approx \nu$, aber für die Mengen $A_n = [2^{-n}, 2^{-(n-1)}]$ gilt: $\mu(A_n) = 2^{-n} \to 0$, $\nu(A_n) = \log 2^{-(n-1)} - \log 2^{-n} = \log 2$.

18.6 Definition (σ-additive Mengenfunktion, signiertes Maß) *Sei $(\Omega, \mathcal{A})$ ein messbarer Raum. Eine Funktion $\varphi : \mathcal{A} \to \mathbb{R}$ heißt σ-additive Mengenfunktion oder signiertes Maß, falls*

i) $\varphi(\emptyset) = 0$,

ii) $\varphi(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \varphi(A_n)$ *für paarweise disjunkte* $A_n \in \mathcal{A}$.

18.7 Bemerkungen
a) Da $\varphi(A) \in \mathbb{R}$, also $|\varphi(A)| < \infty$ für alle $A \in \mathcal{A}$, ist $\lim_{n\to\infty} \sum_{k\geq n} \varphi(A_k) = 0$ für paarweise disjunkte A_n. In diesem Sinn ist auch die Konvergenz in Punkt ii) der obigen Definition zu verstehen.Aus Korollar 18.11 folgt dann allerdings später die absolute Konvergenz.

b) Eine σ-additive Mengenfunktion ist in der Regel nicht σ-subadditiv.

18.8 Beispiel Sind μ, ν endliche Maße auf $(\Omega, \mathcal{A})$, so ist $\varphi = \mu - \nu$ ein signiertes Maß. Wir werden zeigen, dass jedes signierte Maß so darstellbar ist.

18.9 Satz (Zerlegungssatz von Hahn)
Sei $\varphi : \mathcal{A} \to \mathbb{R}$ ein signiertes Maß. Dann gibt es disjunkte Mengen $\Omega^+, \Omega^- \in \mathcal{A}$, $\Omega^+ \cup \Omega^- = \Omega$, sodass $\varphi(E) \geq 0$ für alle $E \subseteq \Omega^+$, $E \in \mathcal{A}$, und $\varphi(E) \leq 0$ für alle $E \subseteq \Omega^-$, $E \in \mathcal{A}$.

Beweis: Sei $\alpha := \sup\{\varphi(A) : A \in \mathcal{A}\}$. Wir werden zeigen, dass es ein $\Omega^+ \in \mathcal{A}$ mit $\varphi(\Omega^+) = \alpha$ gibt. Dann ist $\alpha \in \mathbb{R}$ und für $E \subseteq \Omega^+$, $E \in \mathcal{A}$, gilt

$$\alpha \geq \varphi(\Omega^+ \setminus E) = \varphi(\Omega^+) - \varphi(E) = \alpha - \varphi(E) ,$$

also $\varphi(E) \geq 0$. Für $E \subseteq \Omega^- := \Omega \setminus \Omega^+$, $E \in \mathcal{A}$, gilt

$$\alpha \geq \varphi(\Omega^+ \cup E) = \varphi(\Omega^+) + \varphi(E) = \alpha + \varphi(E) ,$$

also $\varphi(E) \leq 0$.

Wir zeigen nun die Existenz von $\Omega^+ \in \mathcal{A}$ mit $\varphi(\Omega^+) = \alpha$. Wähle $A_n \in \mathcal{A}$ mit $\varphi(A_n) \to \alpha$ und setze $A = \bigcup_{n=1}^\infty A_n$. Da φ auch negative Werte annehmen kann, besteht das Problem darin, dass die A_n auch für große n noch "Inseln negativer Masse" enthalten können. So kann man nicht, wie im Fall eines Maßes schließen, dass $\varphi(A) \geq \lim_{n\to\infty} \varphi(A_n) = \alpha$. Zum Beispiel könnten die A_n von der Form $\Omega^+ \cup E_n$ mit $0 > \varphi(E_n) \to 0$ und $\bigcup_n E_n = \Omega^-$ sein. Dann wäre $A = \Omega^+ \cup \Omega^- = \Omega$ und nichts wäre gewonnen. Deshalb müssen wir genauer in die A_n hineinschauen.

Für jedes n betrachte die 2^n Mengen $\bigcap_{k=1}^n A'_k$, wobei jedes A'_k entweder gleich A_k oder $A \setminus A_k$ ist. Diese Mengen bilden eine Partition $\mathcal{P}_n$ von A. (Einige der Mengen können auch leer sein.) Setze

$$C_n := \bigcup_{P \in \mathcal{P}_n,\ \varphi(P) > 0} P .$$

Da A_n Vereinigung der 2^{n-1} Mengen $A_n \cap \bigcap_{k=1}^{n-1} A'_k$ ist, ist $\varphi(A_n) \leq \varphi(C_n)$.

Betrachte für $m \leq n$ die Mengen $E_m^n = C_m \cup \cdots \cup C_n$. Da jedes $P \in \mathcal{P}_n$ in genau einem $Q \in \mathcal{P}_{n-1}$ enthalten ist, ist für $m < n$ die Menge $E_m^n \setminus E_m^{n-1}$ disjunkte Vereinigung von Mengen $P \in \mathcal{P}_n$ mit $\varphi(P) \geq 0$, also $\varphi(E_m^n \setminus E_m^{n-1}) \geq 0$. Für $E_m := \bigcup_{n \geq m} C_n$ gilt dann $E_m^n \nearrow E_m$ $(n \to \infty)$ und

$$\begin{aligned}
\varphi(A_m) \leq \varphi(C_m) = \varphi(E_m^m) &\leq \varphi(E_m^m) + \sum_{n=m+1}^\infty \varphi(E_m^n \setminus E_m^{n-1}) \\
&= \varphi\left(E_m^m \cup \bigcup_{n=m+1}^\infty (E_m^n \setminus E_m^{n-1})\right) = \varphi\left(\bigcup_{n \geq m} E_m^n\right) = \varphi(E_m) .
\end{aligned}$$

Setze $\Omega^+ = \bigcap_{m=1}^\infty E_m$, also $E_m \searrow \Omega^+$. Dann ist

$$\varphi(E_m) = \varphi\left(\Omega^+ \cup \bigcup_{n \geq m} (E_n \setminus E_{n+1})\right) = \varphi(\Omega^+) + \sum_{n \geq m} \varphi(E_n \setminus E_{n+1}) \overset{\text{B.18.7}}{\to} \varphi(\Omega^+) \quad (m \to \infty) .$$

Also:

$$\alpha = \lim_{m\to\infty} \varphi(A_m) \leq \lim_{m\to\infty} \varphi(E_m) = \varphi(\Omega^+) ,$$

was zu zeigen war. □

Der Satz von Hahn führt direkt zur Zerlegung eines signierten Maßes in einen positiven und einen negativen Teil, die *singulär* zu einander sind.

18.10 Definition (Singularität von Maßen) *Zwei Maße μ und ν auf $\mathcal{A}$ sind singulär zu einander, kurz $\mu \perp \nu$, falls es ein $A \in \mathcal{A}$ mit $\mu(A) = 0 = \nu(\Omega \setminus A)$ gibt.*

18.11 Korollar (Zerlegungssatz von Jordan) *Sei φ ein signiertes Maß auf $(\Omega, \mathcal{A})$. Dann gibt es endliche Maße φ^+ und φ^- auf $\mathcal{A}$ mit $\varphi = \varphi^+ - \varphi^-$ und $\varphi^+ \perp \varphi^-$.*

Beweis: Sei $\Omega = \Omega^+ \uplus \Omega^-$ die Zerlegung aus dem Hahn'schen Satz. Setze $\varphi^+(A) := \varphi(A \cap \Omega^+)$, $\varphi^-(A) := -\varphi(A \cap \Omega^-)$. □

Es folgt einer der beiden wichtigsten Sätze dieses Kapitels.

18.12 Satz (Zerlegungssatz von Lebesgue)
Sei $(\Omega, \mathcal{A}, \mu)$ σ-endlich und sei ν ein weiteres σ-endliches Maß auf $(\Omega, \mathcal{A})$. Dann lässt sich ν wie folgt in zwei σ-endliche Maße zerlegen:

$$\nu = \nu_{abs} + \nu_{sing}, \quad \nu_{abs} \ll \mu, \quad \nu_{sing} \perp \mu$$

ν_{abs} hat Dichte bzgl. μ, und $\frac{d\nu_{abs}}{d\mu}$ ist $\mathcal{A}$-messbar und μ-f.s. endlich.

Eine sofortige Konsequenz daraus ist der bekannteste Satz dieses Kapitels.

18.13 Korollar (Satz von Radon-Nikodym) *Sei $(\Omega, \mathcal{A}, \mu)$ σ-endlich und sei ν ein weiteres σ-endliches Maß auf $(\Omega, \mathcal{A})$. Dann hat ν Dichte bzgl. μ genau dann, wenn $\nu \ll \mu$. Ist das der Fall, so ist $\frac{d\nu}{d\mu}$ $\mathcal{A}$-messbar und μ-f.s. endlich. $\frac{d\nu}{d\mu}$ heißt auch die Radon-Nikodym-Ableitung von ν nach μ.*

Beweis: "$\Rightarrow$" ist Bemerkung 18.2. Für "$\Leftarrow$" wende Satz 18.12 an. Zu zeigen bleibt dann $\nu = \nu_{\text{abs}}$. Sei dazu $A \in \mathcal{A}$ so, dass $\mu(A) = 0$ und $\nu_{\text{sing}}(\Omega \setminus A) = 0$. Da $\nu \ll \mu$, folgt daraus $\nu_{\text{sing}}(A) \leq \nu(A) = 0$, also insgesamt $\nu_{\text{sing}}(\Omega) = 0$, d.h. $\nu = \nu_{\text{abs}}$. □

Zum Beweis des Lebesgue'schen Zerlegungssatzes benötigen wir folgendes Lemma:

18.14 Lemma *Seien μ, ν endliche Maße auf $(\Omega, \mathcal{A})$, $\mu \not\perp \nu$. Dann gibt es ein $A \in \mathcal{A}$ mit $\mu(A) > 0$ und ein $\epsilon > 0$ derart, dass*

$$\epsilon \cdot \mu(E) \leq \nu(E) \quad \textit{für alle } E \subseteq A,\ E \in \mathcal{A}.$$

Beweis: Sei $\Omega = \Omega_n^+ \uplus \Omega_n^-$ eine Hahn-Zerlegung zum signierten Maß $(\nu - \frac{1}{n}\mu)$. Dann ist $M := \bigcap_{n>0} \Omega_n^-$ die Menge derjenigen Punkte in Ω, in denen "μ unendlich mal so viel Masse konzentriert wie ν". Da $M \subseteq \Omega_n^-$ für alle n, ist $(\nu - \frac{1}{n}\mu)(M) \leq 0$, d.h. $\nu(M) \leq \frac{1}{n}\mu(M)$ für alle n und daher $\nu(M) = 0$. Da $\mu \not\perp \nu$, folgt $\mu(\Omega \setminus M) = \mu\left(\bigcup_{n>0} \Omega_n^+\right) > 0$, d.h. $\mu(\Omega_n^+) > 0$ für mindestens ein $n = n_0$. Mit $A = \Omega_{n_0}^+$ und $\epsilon = \frac{1}{n_0}$ ist dann $\mu(A) > 0$ und $(\nu - \epsilon\mu)(E) \geq 0$ für alle $E \subseteq A$, $E \in \mathcal{A}$. □

Beweis von Satz 18.12: Wir betrachten zunächst den Spezialfall, dass μ und ν endliche Maße sind. Unser erstes Ziel ist es, ein maximales Element in der Familie

$$\mathcal{G} := \left\{ g : \Omega \to [0, \infty] \middle| g\ \mathcal{A}\text{-messbar},\ \int_A g\, d\mu \leq \nu(A)\, \forall A \in \mathcal{A} \right\} .$$

zu konstruieren. Von dem zeigen wir dann, dass es die gesuchte Dichte von ν_{abs} ist.

Natürlich ist $g = 0 \in \mathcal{G}$, also $\mathcal{G} \neq \emptyset$. Wir zeigen:

$$g_1, g_2 \in \mathcal{G} \quad \Rightarrow \quad \max\{g_1, g_2\} \in \mathcal{G}\ . \tag{*}$$

Setze dazu $E = \{g_1 \geq g_2\}$. Dann ist für $A \in \mathcal{A}$

$$\int_A \max\{g_1, g_2\}\, d\mu = \int_{A \cap E} g_1\, d\mu + \int_{A \setminus E} g_2\, d\mu \leq \nu(A \cap E) + \nu(A \setminus E) = \nu(A)\ .$$

Sei nun $\gamma := \sup\{\int g\, d\mu : g \in \mathcal{G}\}$. Wir konstruieren ein $f \in \mathcal{G}$ mit $\int f\, d\mu = \gamma$. Wähle dazu $g_n \in \mathcal{G}$ mit $\int g_n\, d\mu \to \gamma$ und setze $f_n := \max\{g_1, \ldots, g_n\}$. Wegen (*) ist $f_n \in \mathcal{G}$ für alle n, und es ist $0 \leq f_1 \leq f_2 \leq \ldots$. Aus dem Satz von der monotonen Konvergenz folgt daher für $f = \sup_n f_n$:

$$\int_A f\, d\mu = \sup_n \int_A f_n\, d\mu \leq \nu(A) \quad \forall A \in \mathcal{A}$$

und

$$\int f\,d\mu = \sup_n \int f_n\,d\mu \geq \sup_n \int g_n\,d\mu \geq \gamma\ .$$

Also ist $f \in \mathcal{G}$ und $\int f\,d\mu = \gamma \leq \nu(\Omega) < \infty$.

Definiere nun ν_{abs} und ν_{sing} durch

$$\nu_{\text{abs}}(A) := \int_A f\,d\mu, \quad \nu_{\text{sing}}(A) := \nu(A) - \nu_{\text{abs}}(A)\ .$$

Dann ist $\nu_{\text{abs}} \ll \mu$ ein endliches Maß mit $\mathcal{A}$-messbarer und μ-f.s. endlicher Dichte f bzgl. μ, und da $\nu_{\text{sing}}(A) = \nu(A) - \int_A f\,d\mu \geq 0$ (wegen $f \in \mathcal{G}$), ist auch ν_{sing} ein endliches Maß. Bleibt zu zeigen: $\nu_{\text{sing}} \perp \mu$.

Angenommen, es wäre $\nu_{\text{sing}} \not\perp \mu$. Wegen Lemma 18.14 gäbe es dann $A \in \mathcal{A}$ mit $\mu(A) > 0$ und $\epsilon > 0$ derart, dass $\epsilon \cdot \mu(E) \leq \nu_{\text{sing}}(E)\ \forall E \in \mathcal{A},\ E \subseteq A$. Also wäre für alle $B \in \mathcal{A}$

$$\int_B (f + \epsilon\,1_A)\,d\mu = \int_B f\,d\mu + \epsilon\,\mu(B \cap A) \leq \nu_{\text{abs}}(B) + \nu_{\text{sing}}(B \cap A) \leq \nu_{\text{abs}}(B) + \nu_{\text{sing}}(B) = \nu(B)\ ,$$

d.h. $(f + \epsilon\,1_A) \in \mathcal{G}$, im Widerspruch zu $\int (f + \epsilon\,1_A)\,d\mu = \gamma + \epsilon\,\mu(A) > \gamma$. Daher ist $\nu_{\text{sing}} \perp \mu$ und der Satz ist für endliche μ und ν bewiesen.

Um daraus die Aussage für σ-endliche μ und ν herzuleiten, benutzen wir die Charakterisierung der σ-Endlichkeit aus Lemma 11.10: Es gibt strikt positive $f \in \mathcal{L}^1_\mu$ und $g \in \mathcal{L}^1_\nu$. Sei $\tilde{\mu} = f\mu$, $\tilde{\nu} = g\nu$. $\tilde{\mu}$ und $\tilde{\nu}$ sind endliche Maße auf $(\Omega, \mathcal{A})$. Aus dem ersten Teil des Beweises folgt:

$$\tilde{\nu} = \tilde{\nu}_{\text{abs}} + \tilde{\nu}_{\text{sing}}, \quad \tilde{\nu}_{\text{abs}} \ll \tilde{\mu}, \quad \tilde{\nu}_{\text{sing}} \perp \tilde{\mu}\ .$$

Setze

$$\nu_{\text{abs}} := \frac{1}{g}\tilde{\nu}_{\text{abs}}, \quad \nu_{\text{sing}} := \frac{1}{g}\tilde{\nu}_{\text{sing}}\ .$$

Dann ist

$$\nu_{\text{abs}} + \nu_{\text{sing}} = \frac{1}{g}\tilde{\nu} = \nu\ ,$$

$$\nu_{\text{abs}} \approx \tilde{\nu}_{\text{abs}} \ll \tilde{\mu} \approx \mu \text{ und } \nu_{\text{abs}} = \frac{1}{g}\tilde{\nu}_{\text{abs}} = \frac{1}{g}\frac{d\tilde{\nu}_{\text{abs}}}{d\tilde{\mu}}\tilde{\mu} = \frac{f}{g}\frac{d\tilde{\nu}_{\text{abs}}}{d\tilde{\mu}}\mu\ ,$$

$$\nu_{\text{sing}} \perp \mu, \text{ da } \nu_{\text{sing}} \approx \tilde{\nu}_{\text{sing}} \text{ und } \mu \approx \tilde{\mu}\ .$$

□

Kapitel 19

Bedingte Wahrscheinlichkeit, bedingte Erwartung

Wir beginnen dieses Kapitel mit einer Diskussion des Begriffs der bedingten Wahrscheinlichkeit, die sich an die Vorstellungen der elementaren Stochastik anlehnt.

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $A, B \in \mathcal{A}$. Die *bedingte Wahrscheinlichkeit von A gegeben B* ist $P(A|B) = \frac{P(A\cap B)}{P(B)}$. Ist Ω höchstens abzählbar und $P(\{\omega\}) > 0$ für jedes $\omega \in \Omega$, so kann man auf diese Weise die bedingten Wahrscheinlichkeiten $P(A|\omega) := P(A|\{\omega\})$ definieren. Ist aber $P(\{\omega\}) = 0$ für jedes $\omega \in \Omega$, so muss man dazu etwas weniger naiv vorgehen. Wir illustrieren eine solche Vorgehensweise zunächst im Rahmen der elementaren Stochastik.

19.1 Beispiel Sei $\beta = (B_i : i \in I)$ eine endliche oder abzählbare Partition von Ω in messbare Mengen positiven Maßes, $\mathcal{F} = \sigma(\beta)$. ($\mathcal{F}$ ist die σ-Algebra aller Mengen, die sich als Vereinigung von Mengen B_i darstellen lassen.) Dann definieren wir für $A \in \mathcal{A}$

$$P[A\big|\mathcal{F}](\omega) := \sum_{i\in I} 1_{B_i}(\omega) \cdot P(A|B_i)\ .$$

$P[A\big|\mathcal{F}]$ ist also eine $\mathcal{F}$-messbare Funktion. Sie nimmt auf B_i den Wert $P(A|B_i)$ an und zeichnet sich durch folgende Eigenschaft aus: Für alle $B \in \mathcal{F}$ ist

$$\begin{aligned}
\int_B P[A\big|\mathcal{F}]\, dP &= \sum_{i\in I} P(A|B_i) \int_B 1_{B_i}\, dP = \sum_{i\in I} P(A|B_i) \underbrace{P(B\cap B_i)}_{=0 \text{ oder } =P(B_i)} \\
&= \sum_{i\in I, B_i\subseteq B} P(A|B_i)\, P(B_i) = \sum_{i\in I, B_i\subseteq B} P(A\cap B_i) = P(A\cap B) \\
&= \int_B 1_A\, dP = (1_A P)(B)\ .
\end{aligned}$$

Da $B \in \mathcal{F}$, folgt

$$\int_B P[A\big|\mathcal{F}]\, dP_{|\mathcal{F}} = (1_A P)_{|\mathcal{F}}(B)\ .$$

Mit den Bezeichnungen des letzten Kapitels ist daher:

$$P[A\big|\mathcal{F}] = \frac{d\big((1_A P)_{|\mathcal{F}}\big)}{d\big(P_{|\mathcal{F}}\big)}$$

eine Radon-Nikodym-Ableitung über dem messbaren Raum $(\Omega, \mathcal{F})$. Damit ist $P[A\big|\mathcal{F}](\omega)$ zwar nur noch für P-f.a. ω eindeutig definiert, aber dafür kann man diese Vorgehensweise auf sehr allgemeine Wahrscheinlichkeitsräume übertragen. Das ist das Ziel dieses Kapitels.

Sei nun $\mathcal{F}$ eine Teil-σ-Algebra von $\mathcal{A}$. Beachte, dass für messbares $f : \Omega \to \mathbb{R}$ gilt: $(f\,P)_{|\mathcal{F}} \ll P_{|\mathcal{F}}$.

19.2 Definition (Bedingte Wahrscheinlichkeit, bedingte Erwartung)

a) Sei $A \in \mathcal{A}$. Die $\mathcal{F}$-messbare Funktion

$$P[A\big|\mathcal{F}] := \frac{d\big((1_A\,P)_{|\mathcal{F}}\big)}{d\big(P_{|\mathcal{F}}\big)}$$

heißt bedingte Wahrscheinlichkeit von A gegeben $\mathcal{F}$.

b) Sei $0 \leq f \in \mathcal{L}^1_P$. Die $\mathcal{F}$-messbare Funktion

$$E[f\big|\mathcal{F}] := \frac{d\big((f\,P)_{|\mathcal{F}}\big)}{d\big(P_{|\mathcal{F}}\big)}$$

heißt bedingte Erwartung von f gegeben $\mathcal{F}$.

c) Sei $f = f^+ - f^- \in \mathcal{L}^1_P$. Setze $E[f\big|\mathcal{F}] := E[f^+\big|\mathcal{F}] - E[f^-\big|\mathcal{F}]$.

Wenn nötig schreibt man zur Verdeutlichung auch $E_P[f\big|\mathcal{F}]$.

19.3 Bemerkungen

a) $P[A\big|\mathcal{F}] = E[1_A\big|\mathcal{F}]$.

b) $P[\emptyset\big|\mathcal{F}] = 0$ und $P[\Omega\big|\mathcal{F}] = 1$.

c) $0 \leq E[f^\pm\big|\mathcal{F}]$ ist P-f.s. definiert und endlich, da $f^\pm P$ ein endliches Maß ist (Satz von Radon-Nikodym).

19.4 Lemma (Charakterisierung bedingter Erwartungen und bedingter Wahrscheinlichkeiten)
Seien $f, g \in \mathcal{L}^1_P$, g sei $\mathcal{F}$-messbar. Dann gilt:

a) Ist $g \cdot E[f\big|\mathcal{F}] \in \mathcal{L}^1_P$, so ist $\int g \cdot E[f\big|\mathcal{F}]\,dP = \int gf\,dP$. Äquivalent dazu: $E\big[g \cdot E[f\big|\mathcal{F}]\big] = E[gf]$. Insbesondere ist $\int_B E[f\big|\mathcal{F}]\,dP = \int_B f\,dP$ für alle $B \in \mathcal{F}$.

b) Es gilt $\int_B g\,dP \underset{(\leq)}{\overset{(\geq)}{=}} \int_B f\,dP$ für alle $B \in \mathcal{F}$ genau dann, wenn $g \underset{(\leq)}{\overset{(\geq)}{=}} E[f\big|\mathcal{F}]$.

c) Ist $\mathcal{C}$ ein $\cap$-stabiler Erzeuger von $\mathcal{F}$, $\Omega = \bigcup_{n=1}^\infty C_n$ für geeignete $C_n \in \mathcal{C}$ und ist $\int_C g\,dP = \int_C f\,dP$ für alle $C \in \mathcal{C}$, so ist $g = E[f\big|\mathcal{F}]$.

Beachte: *Alle Gleichungen und Ungleichungen, die bedingte Erwartungen (oder bedingte Wahrscheinlichkeiten) involvieren, können immer nur fast sicher verstanden werden, da Radon-Nikodym-Ableitungen nur fast sicher festgelegt sind! Wollen wir eine bedingte Erwartung oder Wahrscheinlichkeit als überall definiert verstehen, so wählen wir eine Version davon. Die ist dann messbar bzgl. der σ-Algebra, auf die bedingt wird.*

Beweis:

a) Da g $\mathcal{F}$-messbar ist, ist

$$\int g\, E[f|\mathcal{F}]\, dP = \int g\, E[f|\mathcal{F}]\, dP_{|\mathcal{F}} = \int g\, d\big((f\,P)_{|\mathcal{F}}\big) = \int g\, d(f\,P) = \int gf\, dP\,.$$

b) Wir beweisen die Äquivalenz für "$\leq$"; für "$\geq$" verläuft der Beweis genau so.

"$\Leftarrow$": Sei $B \in \mathcal{F}$. Dann ist $\int_B g\, dP \leq \int_B E[f|\mathcal{F}]\, dP \overset{a)}{=} \int_B f\, dP$.

"$\Rightarrow$": Sei $h := E[f|\mathcal{F}] - g$. Dann ist h $\mathcal{F}$-messbar und

$$0 \geq \int_{\{h<0\}} h\, dP = \int_{\{h<0\}} E[f|\mathcal{F}]\, dP - \int_{\{h<0\}} g\, dP = \int_{\{h<0\}} f\, dP - \int_{\{h<0\}} g\, dP \geq 0\,,$$

also $P\{h<0\} = 0$, d.h. $g \leq E[f|\mathcal{F}]$.

c) Man zeigt leicht: $\mathcal{D} := \{C \in \mathcal{F} : \int_C g\, dP = \int_C f\, dP\}$ ist ein Dynkin-System. Da $\mathcal{C} \subseteq \mathcal{D}$ ein $\cap$-stabiler Erzeuger für $\mathcal{F}$ ist, folgt aus Satz 1.15, dass $\mathcal{F} = \sigma(\mathcal{C}) = \mathcal{D}(\mathcal{C}) \subseteq \mathcal{D}$. Aus Teil b) des Lemmas folgt nun $g = E[f|\mathcal{F}]$.

□

19.5 Beispiele a) $\mathcal{F} = \mathcal{A}$. Dann ist $E[f|\mathcal{F}] = f$, da $\frac{d((fP)_{|\mathcal{A}})}{dP_{|\mathcal{A}}} = \frac{d(fP)}{dP} = f$.

b) $\mathcal{F}$ trivial (d.h. $P(A) = 0$ oder 1 für alle $A \in \mathcal{F}$). Dann ist $E[f|\mathcal{F}] = E[f]$, da $E[f]$ die einzige P-f.s. konstante Funktion ist, die den gleichen Erwartungswert wie f hat.

c) Ist $Y \in \mathcal{L}^1_P$ unabhängig von $\mathcal{F}$, so ist $E[Y|\mathcal{F}] = E[Y]$, denn für alle $B \in \mathcal{F}$ ist

$$\int 1_B\, Y\, dP = E[Y]\, P(B) = \int_B E[Y]\, dP$$

und daraus folgt $E[Y|\mathcal{F}] = E[Y]$ nach Lemma 19.4b. (Beachte, dass das vorhergehende Beispiel ein Spezialfall dieser Aussage ist.)

d) Sei $\mathcal{F} = \sigma(\beta)$, $\beta = (B_i|i \in I)$ eine höchstens abzählbare messbare Partition von Ω mit $P(B_i) > 0$. Dann ist

$$P[A|\mathcal{F}] = \frac{d\big((1_A P)_{|\mathcal{F}}\big)}{dP_{|\mathcal{F}}} = \sum_{i\in I} 1_{B_i}\, P(A|B_i)$$

wie in Beispiel 19.1. Allgemeiner ist in diesem Fall

$$E[f|\mathcal{F}] = \sum_{i\in I} \frac{1}{P(B_i)} \int_{B_i} f\, dP \cdot 1_{B_i}\,.$$

e) Sei $(\Omega, \mathcal{A}) = (\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$, P_1 ein Wahrscheinlichkeitsmaß auf $\mathcal{A}_1$ und K ein stochastischer Kern von Ω_1 nach $\mathcal{A}_2$. Setze $P := P_1 \times K$ (siehe Satz 11.7). Dann ist für $A \in \mathcal{A}$

$$P[A|\pi_1^{-1}\mathcal{A}_1](\omega_1, \omega_2) = K(\omega_1, A_{\omega_1})$$

wo $A_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}$. Das folgt aus Lemma 19.4c, denn für alle $B = B_1 \times \Omega_2 \in \pi_1^{-1}\mathcal{A}_1$ gilt:

$$\begin{aligned}
\int_B K(\omega_1, A_{\omega_1})\, dP(\omega_1, \omega_2) &= \int_{B_1} \left(\int_{\Omega_2} K(\omega_1, A_{\omega_1})\, K(\omega_1, d\omega_2)\right) dP_1(\omega_1) \\
&= \int_{B_1} K(\omega_1, A_{\omega_1})\, dP_1(\omega_1) \\
&= \int_{\Omega_1} K(\omega_1, (A \cap B)_{\omega_1})\, dP_1(\omega_1) \\
&= P(A \cap B)
\end{aligned}$$

wobei man beachte:

$$(A \cap B)_{\omega_1} = (A \cap (B_1 \times \Omega_2))_{\omega_1} = \begin{cases} A_{\omega_1}, & \text{falls } \omega_1 \in B_1 \\ \emptyset & \text{sonst} \end{cases}$$

19.6 Satz (Elementare Eigenschaften der bedingten Erwartung)
Für $X, Y, X_n \in \mathcal{L}^1_P$ und $\alpha, \beta \in \mathbb{R}$ gilt:

a) $E[\alpha X + \beta Y \big| \mathcal{F}] = \alpha\, E[X\big|\mathcal{F}] + \beta\, E[Y\big|\mathcal{F}]$.

b) $X \leq Y$ *P-f.s.* $\Rightarrow$ $E[X\big|\mathcal{F}] \leq E[Y\big|\mathcal{F}]$ *(Insbesondere:* $0 \leq P[A\big|\mathcal{F}] \leq 1$*)*

c) X *$\mathcal{F}$-messbar* $\Rightarrow$ $E[X\big|\mathcal{F}] = X$

d) $|E[X\big|\mathcal{F}]| \leq E[|X|\,\big|\mathcal{F}]$

e) Ist $\lim_{n\to\infty} X_n = X$ *P-f.s. und* $|X_n| \leq Y$ *P-f.s. für alle* n*,* $Y \in \mathcal{L}^1_P$*, so ist*

$$\lim_{n\to\infty} E[X_n\big|\mathcal{F}] = E[X\big|\mathcal{F}] \quad P\text{-f.s.}$$

Beweis:

a) Sei $\varphi := \alpha\, E[X\big|\mathcal{F}] + \beta\, E[Y\big|\mathcal{F}]$. φ ist $\mathcal{F}$-messbar und für $B \in \mathcal{F}$ ist

$$\begin{aligned}\int_B \varphi\, dP &= \alpha \int_B E[X\big|\mathcal{F}]\, dP + \beta \int_B E[Y\big|\mathcal{F}]\, dP = \alpha \int_B X\, dP + \beta \int_B Y\, dP \\ &= \int_B (\alpha X + \beta Y)\, dP\end{aligned}$$

Die Behauptung folgt aus Lemma 19.4b.

b) Für jedes $B \in \mathcal{F}$ ist $\int_B E[X\big|\mathcal{F}]\, dP = \int_B X\, dP \leq \int_B Y\, dP$, also nach Lemma 19.4b: $E[X\big|\mathcal{F}] \leq E[Y\big|\mathcal{F}]$.

c) Folgt aus Lemma 19.4b für $f = g = X$.

d) Folgt aus Aussage b), da $\pm X \leq |X|$.

e) Sei $Z_n := \sup_{k\geq n} |X_k - X|$. Dann gilt: $Z_n \searrow 0$ P-f.s. und $0 \leq Z_n \leq 2Y$, so dass $\int Z_n\, dP \to 0$ nach dem Satz von der majorisierten Konvergenz. Außerdem ist $|E[X_n\big|\mathcal{F}] - E[X\big|\mathcal{F}]| \leq E[Z_n\big|\mathcal{F}]$ wegen Aussage a) und d), und es fällt $E[Z_n\big|\mathcal{F}] \searrow Z \geq 0$ P-f.s. wegen Aussage b). Zu zeigen ist $Z = 0$ P-f.s., was wegen $Z \geq 0$ äquivalent ist zu $E[Z] = 0$. Das folgt aber aus $\int Z\, dP \leq \int E[Z_n\big|\mathcal{F}]\, dP = \int Z_n\, dP \to 0$.

$\square$

19.7 Satz
Seien $X \in \mathcal{L}^1_P$, Y $\mathcal{F}$-messbar und $XY \in \mathcal{L}^1_P$. Dann ist

$$E[XY\big|\mathcal{F}] = Y \cdot E[X\big|\mathcal{F}] \tag{*}$$

Beweis: Beide Seiten von (*) sind $\mathcal{F}$-messbar und für $B \in \mathcal{F}$ ist

$$\int_B Y \cdot E[X\big|\mathcal{F}]\, dP = \int \underbrace{1_B\, Y}_{\mathcal{F}\text{-mb.}} \cdot E[X\big|\mathcal{F}]\, dP \overset{\text{L.19.4a}}{=} \int 1_B Y X\, dP = \int_B Y X\, dP$$

Die Behauptung folgt aus Lemma 19.4b. □

19.8 Satz
Seien $\mathcal{F}_1 \subseteq \mathcal{F}_2$ Teil-σ-Algebren von $\mathcal{A}$, $X \in \mathcal{L}^1_P$. Dann ist $E[E[X|\mathcal{F}_2]|\mathcal{F}_1] = E[X|\mathcal{F}_1]$.

Beweis: Die linke Seite ist $\mathcal{F}_1$-messbar, und für $B \in \mathcal{F}_1 \subseteq \mathcal{F}_2$ gilt

$$\int_B E[E[X|\mathcal{F}_2]|\mathcal{F}_1]\,dP = \int_B E[X|\mathcal{F}_2]\,dP = \int_B X\,dP\,.$$

Die Behauptung folgt aus Lemma 19.4b. □

Sind $A_k \in \mathcal{A}$ paarweise disjunkt ($k > 0$), so ist

$$\begin{aligned} P\left[\biguplus_{k=1}^{\infty} A_k\Big|\mathcal{F}\right] &= E\left[\lim_{n\to\infty}\sum_{k=1}^{n} 1_{A_k}\Big|\mathcal{F}\right] \overset{\text{S.19.6e}}{=} \lim_{n\to\infty} E\left[\sum_{k=1}^{n} 1_{A_k}\Big|\mathcal{F}\right] \overset{\text{S.19.6a}}{=} \lim_{n\to\infty}\sum_{k=1}^{n} E\left[1_{A_k}|\mathcal{F}\right] \\ &= \sum_{k=1}^{\infty} P\left[A_k|\mathcal{F}\right] \end{aligned} \qquad (*)$$

Da außerdem $P[\emptyset|\mathcal{F}] = 0$ und $P[\Omega|\mathcal{F}] = 1$ (Bemerkung 19.3b), liegt die Vermutung nahe, dass die Abbildung $A \mapsto P[A|\mathcal{F}](\omega)$ für jedes (oder für P-fast jedes) ω ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$ ist.

In voller Allgemeinheit ist diese Vermutung leider falsch, denn für jede Folge $(A_n)_{n>0}$ paarweise disjunkter Mengen in $\mathcal{A}$ kann die σ-Additivität von $P[\,.\,|\mathcal{F}](\omega)$ in (*) auf einer anderen Nullmenge von ω's verletzt sein, und die (überabzählbare) Vereinigung solcher Ausnahmemengen muss keine Nullmenge sein. Ein Beispiel dafür, das auf der Existenz nicht Borel-messbarer Teilmengen von $[0,1]$ beruht, wird in [Billingsley, Problem 33.13] vorgestellt.

Es gilt jedoch:

19.9 Satz (Existenz bedingter Borel'scher Wahrscheinlichkeitsverteilungen auf $\mathbb{R}$)
Ist P ein Wahrscheinlichkeitsmaß auf den Borel-Mengen $\mathcal{B}$ von $\mathbb{R}$ und ist $\mathcal{F}$ eine Teil-σ-Algebra von $\mathcal{B}$, die alle Borel-Nullmengen enthält, so gibt es einen stochastischen Kern $K_{\mathcal{F}}$ von $(\mathbb{R},\mathcal{F})$ nach $(\mathbb{R},\mathcal{B})$ derart, dass für jedes $A \in \mathcal{B}$

$$P[A|\mathcal{F}](x) = K_{\mathcal{F}}(x,A)\,,$$

d.h. für jedes $A \in \mathcal{A}$ ist $K_{\mathcal{F}}(\,.\,,A)$ eine Version von $P[A|\mathcal{F}]$.
Die $K_{\mathcal{F}}(x,\,.\,)$ heißen bedingte Wahrscheinlichkeitsverteilungen.

Beweis: Für $r \in \mathbb{Q}$ sei $F(r,\,.\,)$ eine Version von $P[(-\infty,r]|\mathcal{F}]$. Für $r \le s$ ist $(-\infty,r] \subseteq (-\infty,s]$, also gibt es wegen Lemma 19.4b eine P-Nullmenge $A_{r,s}$ mit

$$F(r,x) \le F(s,x) \quad \forall x \notin A_{r,s}\,.$$

Wegen Satz 19.6e gibt es außerdem P-Nullmengen B_r ($r \in \mathbb{Q}$) und C derart, dass

$$\begin{aligned} &\lim_{n\to\infty} F(r+\frac{1}{n},x) = F(r,x) \quad \forall x \notin B_r \quad \text{und} \\ &\lim_{n\to\infty} F(-n,x) = 0, \quad \lim_{n\to\infty} F(n,x) = 1 \quad \forall x \notin C\,. \end{aligned}$$

$N := \bigcup_{r,s\in\mathbb{Q}} A_{r,s} \cup \bigcup_{r\in\mathbb{Q}} B_r \cup C$ ist eine P-Nullmenge, also $N \in \mathcal{F}$, und für $x \in \mathbb{R}\setminus N$ und $z \in \mathbb{R}$ ist

$$F(z,x) := \inf\{F(r,x) : r \in \mathbb{Q}, r \ge z\}$$

wohl definiert. Man prüft leicht nach, dass $F(\,.\,,x)$ eine Verteilungsfunktion ist. Für $x \in N$ setze $F(z,x) := F_0(z)$, wo F_0 eine beliebige fest gewählte Verteilungsfunktion ist. Dann ist $z \mapsto F(z,x)$ für jedes $x \in \mathbb{R}$ eine Verteilungsfunktion. Das durch sie bestimmte Wahrscheinlichkeitsmaß auf $\mathcal{B}$ werde mit $K_{\mathcal{F}}(x, \,.\,)$ bezeichnet.

Für $A = (-\infty, r]$, $r \in \mathbb{Q}$, ist

$$K_{\mathcal{F}}(x, A) = F(r, x) = P[A\big|\mathcal{F}](x) \cdot 1_{\mathbb{R}\setminus N}(x) + F_0(r) \cdot 1_N(x)$$

als Funktion von x $\mathcal{F}$-messbar. Wegen Bemerkung 11.4 ist dann $x \mapsto K_{\mathcal{F}}(x, A)$ für alle $A \in \mathcal{B}$ $\mathcal{F}$-messbar, d.h. $K_{\mathcal{F}}$ ist ein stochastischer Kern.

Betrachte nun ein $B \in \mathcal{F}$. Für $A = (-\infty, r]$, $r \in \mathbb{Q}$, ist

$$\int_B K_{\mathcal{F}}(x, A)\, dP(x) = \int_B P[A\big|\mathcal{F}]\, dP = P(A \cap B)\ .$$

Da beide Seiten dieser Gleichung (als Funktion von A) ein endliches Maß auf $\mathcal{A}$ beschreiben, folgt aus dem Eindeutigkeitssatz 2.12, dass diese Gleichheit für alle $A \in \mathcal{B}$ gilt. Aus Lemma 19.4c folgt dann, dass $K_{\mathcal{F}}(x, A) = P[A\big|\mathcal{F}](x)$ P-f.s. für jedes $A \in \mathcal{B}$. □

19.10 Bemerkung (Existenz bedingter Wahrscheinlichkeitsverteilungen auf polnischen Räumen)
Der letzte Satz bleibt richtig, wenn die Menge $\mathbb{R}$ durch einen polnischen Raum ersetzt wird, vergleiche Bemerkung 12.4.

Eine Art Umkehrung zu Satz 19.9 lieferte Beispiel 19.5 e.

19.11 Satz (Bedingte Erwartungen durch bedingte Wahrscheinlichkeitsverteilungen)
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subseteq \mathcal{A}$ eine Teil-σ-Algebra mit bedingten Wahrscheinlichkeitsverteilungen $K_{\mathcal{F}}(\omega, \,.\,)$. Dann ist für jedes $f \in \mathcal{L}^1_P$

$$E_P[f\big|\mathcal{F}](\omega) = \int_\Omega f(\omega')\, K_{\mathcal{F}}(\omega, d\omega') = E_{K_{\mathcal{F}}(\omega, \,.\,)}[f]\ .$$

Beweis: Die rechte Seite ist $\mathcal{F}$-messbar (Lemma 11.6). Für $f = 1_A$, $A \in \mathcal{A}$, ist

$$\int_\Omega f(\omega')\, K_{\mathcal{F}}(\omega, d\omega') = K_{\mathcal{F}}(\omega, A) = P[A\big|\mathcal{F}](\omega) = E[f\big|\mathcal{F}](\omega)\ .$$

Durch algebraische Induktion (siehe Beweis zu Satz 7.20) folgt die Gleichheit für $f \in \mathcal{L}^1_P$. Dabei wird zur Behandlung von $E_P[f\big|\mathcal{F}]$ mehrfach Satz 19.6 heran gezogen. □

Auch die Jensen'sche Ungleichung (Satz 10.9) kann auf bedingte Erwartungen verallgemeinert werden:

19.12 Satz (Bedingte Jensen'sche Ungleichung)
Sei $I \subseteq \mathbb{R}$ ein offenes Intervall (nicht notwendig endlich!), $\varphi : I \to \mathbb{R}$ konvex, $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, X eine Zufallsvariable mit Werten in I f.s., $X, \varphi(X) \in \mathcal{L}^1_P$, und sei $\mathcal{F}$ eine Teil-σ-Algebra von $\mathcal{A}$. Dann ist

$$\varphi\left(E[X\big|\mathcal{F}]\right) \le E\left[\varphi(X)\big|\mathcal{F}\right]\ .$$

Ist φ strikt konvex, so gilt fast sicher Gleichheit genau dann, wenn $X = E[X\big|\mathcal{F}]$, also genau dann wenn X $\mathcal{F}$-messbar ist.

Beweis: Nimmt man die Existenz bedingter Wahrscheinlichkeitsverteilungen an, so folgt der Satz direkt aus der einfachen Jensen'schen Ungleichung 10.9:

$$\varphi\left(E[X\big|\mathcal{F}](\omega)\right) = \varphi\left(E_{K_{\mathcal{F}}(\omega, \,.\,)}[X]\right) \le E_{K_{\mathcal{F}}(\omega, \,.\,)}[\varphi(X)] = E\left[\varphi(X)\big|\mathcal{F}\right](\omega)$$

und für strikt konvexes φ ist die Ungleichung mit positiver Wahrscheinlichkeit strikt genau dann, wenn $X \neq E[X|\mathcal{F}]$ mit positiver Wahrscheinlichkeit.

Der Vollständigkeit halber geben wir hier auch den Beweis für den allgemeinen Fall. Wir wiederholen aus Bemerkung 10.8:

Jedes konvexe $\varphi : I \to \mathbb{R}$ ist stetig im Inneren von I, also auch messbar, und für alle $x \in \overset{\circ}{I}$ existiert

$$\varphi^*(x) := \lim_{v \downarrow x} \frac{\varphi(v) - \varphi(x)}{v - x} = \lim_{n \to \infty} \frac{\varphi(x + \frac{1}{n}) - \varphi(x)}{\frac{1}{n}} \quad \text{(Monotonie!)}.$$

Mit φ ist auch φ^* messbar, und

$$\frac{\varphi(x) - \varphi(u)}{x - u} \leq \varphi^*(x) \leq \frac{\varphi(v) - \varphi(x)}{v - x} \quad \text{für alle } u < x < v \text{ in } I. \tag{*}$$

Wegen (*) gilt für $x, y \in I$

$$\varphi^*(x) \cdot (y - x) + \varphi(x) \leq \varphi(y) \ .$$

Da $X \in I$ f.s., ist auch $E[X|\mathcal{F}] \in I$ f.s. (Satz 19.6b), und es folgt mit $x = E[X|\mathcal{F}]$ und $y = X$

$$\varphi^*(E[X|\mathcal{F}]) \cdot \big(X - E[X|\mathcal{F}]\big) + \varphi(E[X|\mathcal{F}]) \leq \varphi(X) \ .$$

Nimmt $E[X|\mathcal{F}]$ seine Werte f.s. in einem kompakten Teilintervall $K \subseteq I$ an, so ist $\varphi^*(E[X|\mathcal{F}])$ wegen (*) f.s. beschränkt, und durch Anwendung von $E[\,.\,|\mathcal{F}]$ auf die Ungleichung erhält man unter Berücksichtigung von Satz 19.6 und Satz 19.7

$$\varphi^*\left(E[X|\mathcal{F}]\right) \cdot \underbrace{\left(E[X|\mathcal{F}] - E[X|\mathcal{F}]\right)}_{=0} + \varphi\left(E[X|\mathcal{F}]\right) \leq E\left[\varphi(X)|\mathcal{F}\right] \ , \tag{**}$$

also die Behauptung.

Andernfalls sei $K_n \nearrow I$ eine Folge kompakter Teilintervalle und

$$F_n := 1_{\{E[X|\mathcal{F}] \in K_n\}} \quad \text{für alle } n > 0.$$

Wähle $x_0 \in \bigcup_n K_n$. Da die F_n $\mathcal{F}$-messbar sind, ist

$$E\left[F_n\, X + (1 - F_n)\, x_0 \middle| \mathcal{F}\right] = F_n \cdot E[X|\mathcal{F}] + (1 - F_n) \cdot x_0 \in K_n \quad \text{f.s.},$$

also wegen (**)

$$\begin{aligned}\varphi\left(F_n \cdot E[X|\mathcal{F}] + (1 - F_n) \cdot x_0 \in K_n\right) &\leq E\left[\varphi\left(F_n\, X + (1 - F_n)\, x_0\right) \middle| \mathcal{F}\right] \\ &= E\left[F_n\, \varphi(X) + (1 - F_n)\, \varphi(x_0) \middle| \mathcal{F}\right] \\ = F_n \cdot E[\varphi(X)|\mathcal{F}] + (1 - F_n) \cdot \varphi(x_0) \ .\end{aligned}$$

Für alle ω mit $F_n(\omega) = 1$ folgt daraus $\varphi(E[X|\mathcal{F}])(\omega) \leq E[\varphi(X)|\mathcal{F}](\omega)$, und da $1_{F_n} \nearrow 1$ f.s., folgt die Behauptung. □

In der folgenden Bemerkung werden bedingte Erwartungen geometrisch als Orthogonalprojektoren in einem Hilbertraum interpretiert.

19.13 Bemerkung (Bedingte Erwartungen als Orthogonalprojektionen) Sei wieder $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subseteq \mathcal{A}$ eine Teil-σ-Algebra. Mit $L^2_P(\mathcal{F})$ bezeichnen wir die Menge der P-Äquivalenzklassen $\mathcal{F}$-messbarer, quadratintegrierbarer Funktionen $f : \Omega \to \mathbb{R}$. Durch das Skalarprodukt $\langle f, g\rangle = \int f g\, dP$ wird $L^2_P(\mathcal{F})$ zu einem Hilbertraum, siehe Bemerkung 10.6 mit daraus abgeleiteter Norm $\|f\|_2 = \langle f, f\rangle^{\frac{1}{2}}$.

Da das Skalarprodukt auf $L^2_P(\mathcal{F})$ die Einschränkung des Skalarprodukts auf dem größeren Raum $L^2_P(\mathcal{A})$ ist, und da der $L^2_P(\mathcal{F})$ mit seiner Norm vollständig ist, ist er ein abgeschlossener linearer Teilraum des $L^2_P(\mathcal{A})$.

Definiere nun

$$\Phi : L^2_P(\mathcal{A}) \to L^2_P(\mathcal{A}), \quad \Phi(f) := E_P[f|\mathcal{F}] \ .$$

Dann gilt

▷ Φ ist linear,

▷ $\Phi(f) \in L^2_P(\mathcal{F})$ für alle $f \in L^2_P(\mathcal{A})$ (Jensen'sche Ungleichung angewandt auf die konvexe Funktion $t \mapsto t^2$) und

▷ $\Phi(\Phi f) = E_P[E_p[f|\mathcal{F}]|\mathcal{F}] = \Phi(f)$, also $\Phi^2 = \Phi$,

▷ $\langle \Phi f, f - \Phi f\rangle = \int \Phi f \cdot (f - \Phi f)\, dP = \int E\left[\Phi f \cdot (f - \Phi f)\middle|\mathcal{F}\right] dP = \int \Phi f \cdot E\left[f - \Phi f\middle|\mathcal{F}\right] dP = 0$, da $\Phi f = E[f|\mathcal{F}]$.

Also ist Φ die Orthogonalprojektion auf $L^2_P(\mathcal{F})$ und $f = E[f|\mathcal{F}] + \big(f - E[f|\mathcal{F}]\big)$ die orthogonale Zerlegung von f bzgl. $L^2_P(\mathcal{F})$.

Wir beschließen dieses Kapitel mit einem Beweis des Ergodensatzes 9.3 für maßtreue Transformationen im nicht-ergodischen Fall. Sei wie dort $(X, \mathcal{F}, \mu)$ ein Wahrscheinlichkeitsraum, $T : X \to X$ messbar und maßerhaltend (d.h. $\mu \circ T^{-1} = \mu$). Wir bezeichnen mit

$$\mathcal{I}(T) = \{A \in \mathcal{A} : T^{-1}(A) = A)\}$$

die Familie der T-invarianten messbaren Teilmengen von X.

19.14 Lemma *$\mathcal{I}(T)$ ist eine σ-Algebra.*

Beweis: Einfache Übung. □

19.15 Satz (Birkhoffscher Ergodensatz für maßerhaltende Transformationen)
Sei $f \in \mathcal{L}^1_\mu$. Dann gibt es eine Version $\bar{f}$ der bedingten Erwartung $E[f|\mathcal{I}(T)]$ derart, dass

$$\bar{f} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f \circ T^k$$

μ-f.s. und in $\mathcal{L}^1_\mu$, und es gilt: $\int |\bar{f}|\, d\mu \le \int |f|\, d\mu < \infty$, also $\bar{f} \in \mathcal{L}^1_\mu$.

Beweis: Der Beweis ist eine Verallgemeinerung des Beweises zu Satz 9.3.

Seien $\bar{f}$ eine Version von $E[f|\mathcal{I}(T)]$, und sei $S_n f := \sum_{k=0}^{n-1} f \circ T^k$ $(n \ge 0)$. Wir werden zeigen, dass

$$F := \limsup_{n\to\infty} \frac{1}{n} S_n f \le \bar{f} \quad \mu\text{-f.s.}$$

Durch Anwendung des gleichen Arguments auf die Funktion $-f$ folgt dann $\limsup_{n\to\infty} \frac{1}{n} S_n(-f) \le -\bar{f}$ μ-f.s., d.h. $\liminf_{n\to\infty} \frac{1}{n} S_n f \ge \bar{f}$ μ-f.s. Beide Ungleichungen zusammen ergeben die Behauptung $\lim_{n\to\infty} \frac{1}{n} S_n f = \bar{f}$ μ-f.s. Die Ungleichung $\int |\bar{f}|\, d\mu \le \int |f|\, d\mu < \infty$ folgt schließlich direkt aus Satz 19.6d.

Für $\epsilon > 0$ seien $g_\epsilon := (f - \bar{f} - \epsilon) 1_{\{F > \bar{f} + \epsilon\}}$ und $G_\epsilon := \limsup_{n\to\infty} \frac{1}{n} S_n g_\epsilon$. Wie im Beweis zu Satz 9.3 folgt $G_\epsilon \circ T = G_\epsilon$, und genau so zeigt man $F \circ T = F$.

Da $\bar{f}$ eine Version von $E[f|\mathcal{I}(T)]$ ist, ist $\bar{f}$ messbar bzgl. $\mathcal{I}(T)$. Daraus folgt $\bar{f} \circ T = \bar{f}$, denn für alle $\alpha \in \mathbb{R}$ gilt

$$\{f \circ T < \alpha\} = T^{-1}\{f < \alpha\} = \{f < \alpha\}\,, \quad \text{da } \{f < \alpha\} \in \mathcal{I}(T)\,.$$

Daher ist $G_\epsilon = (F - \bar{f} - \epsilon) 1_{\{F > \bar{f} + \epsilon\}}$, und es reicht zu zeigen, dass $\mu\{G_\epsilon > 0\} = 0$ für jedes $\epsilon > 0$, denn wegen $\{F > \bar{f} + \epsilon\} = \{G_\epsilon > 0\}$ folgt daraus $\mu\{F > \bar{f} + \epsilon\} = 0$ für jedes $\epsilon > 0$, so dass

$$\mu\{F > \bar{f}\} = \mu\left(\bigcup_{k=1}^{\infty} \{F > \bar{f} + \frac{1}{k}\}\right) \le \sum_{k=1}^{\infty} \mu\{F > \bar{f} + \frac{1}{k}\} = 0\,,$$

also $F \leq \bar{f}$ μ-f.s.

Nun zum Beweis von $\mu\{G_\epsilon > 0\} = 0$: Da $G_\epsilon \circ T = G_\epsilon$, ist $\{G_\epsilon > 0\} \in \mathcal{I}(T)$.

Sei $M_n := \max\{0, S_1 g_\epsilon, \dots, S_n g_\epsilon\}$. Genau wie im Beweis zu Satz 9.3 zeigt man *Hopfs Maximalungleichung* und die daraus folgende Abschätzung

$$\int_{\{M_n > 0\}} g_\epsilon \, d\mu \geq 0 \; .$$

Nun ist aber $\{M_n > 0\} \nearrow \bigcup_{n=1}^\infty \{S_n g_\epsilon > 0\} = \{G_\epsilon > 0\}$, so dass aus Korollar 8.10 zum Satz von der majorisierten Konvergenz folgt

$$\int g_\epsilon \, d\mu = \lim_{n\to\infty} \int_{\{M_n > 0\}} g_\epsilon \, d\mu \geq 0$$

und daher

$$0 \leq \int g_\epsilon \, d\mu = \int_{\{F > \bar{f} + \epsilon\}} f \, d\mu - \int_{\{F > \bar{f} + \epsilon\}} \bar{f} \, d\mu - \epsilon = -\epsilon \cdot \mu\{G_\epsilon\} < 0 \; ,$$

wobei wir benutzt haben, dass $\bar{f} = E[f \big| \mathcal{I}(T)]$ und dass $\{F > \bar{f} + \epsilon\} = \{G_\epsilon > 0\} \in \mathcal{I}(T)$. Also ist $\mu\{G_\epsilon > 0\} = 0$ für jedes $\epsilon > 0$.

$\square$

Kapitel 20

Markov-Ketten

Dem Begriff der *Markov-Kette* (= Markov-Prozess mit Indexmenge $\mathbb{N}$ oder $\mathbb{Z}$, oft interpretiert als diskrete Zeit) kann man sich auf zwei leicht unterschiedliche Weisen nähern: Konstruktiv, indem man mit Hilfe von Übergangskernen Produktmaße mit speziellen Eigenschaften gewinnt, die dann als Verteilungen solcher Prozesse fungieren, oder analytisch, indem man mit Hilfe des Begriffs der bedingten Wahrscheinlichkeit gewisse Bedingungen an die Verteilung eines Prozesses formuliert. Wir werden beide Zugänge vorstellen und zeigen, dass sie zum gleichen Ergebnis führen.

Wir beginnen mit dem analytischen Zugang. Sei $T = \mathbb{N}$ oder $\mathbb{Z}$, $(X_t)_{t\in T}$ ein kanonischer Koordinatenprozess auf $(\mathbb{R}^T, \mathcal{B}^T, P)$. ($P$ ist also gleichzeitig die Verteilung von $(X_t)_{t\in T}$.) Wir führen folgende Notationen ein:

- ▷ Für $I \subseteq T$ sei $\mathcal{F}_I := \sigma(X_t : t \in I)$, siehe auch Definition 5.9.
- ▷ Für $t \in T$ seien $\mathcal{F}_t := \mathcal{F}_{\{t\}}$, $\mathcal{F}_{\leq t} := \mathcal{F}_{\{s\in T: s\leq t\}}$ und $\mathbb{R}^{\leq t} := \mathbb{R}^{\{s\in T: s\leq t\}}$ u.s.w.

20.1 Definition (Markov-Kette) *$(X_t)_{t\in T}$ ist eine Markov-Kette, falls für alle $B \in \mathcal{B}$ und für alle $s < t \in T$ gilt*

$$P[X_t \in B \big| \mathcal{F}_{\leq s}] = P[X_t \in B \big| \mathcal{F}_s] \quad \textit{(Markov-Eigenschaft)} \tag{ME}$$

20.2 Bemerkung Aus der Markov-Eigenschaft folgt sofort für jedes $\mathcal{F}_t$-messbare $f \in \mathcal{L}^1_P$

$$E[f \big| \mathcal{F}_{\leq s}] = E[f \big| \mathcal{F}_s] \ ,$$

denn für $f = 1_A$, $A \in \mathcal{F}_t$, ist das gerade (ME), und für allgemeine f wendet man algebraische Induktion an.

Für unsere weiteren Überlegungen ist folgendes Lemma nützlich.

20.3 Lemma *Sei $f : \mathbb{R}^T \to \bar{\mathbb{R}}$ $\mathcal{F}_I$-messbar, $I \subseteq T$. Dann gibt es eine $\mathcal{B}^I$-messbare Abbildung $g : \mathbb{R}^I \to \bar{\mathbb{R}}$ derart, dass $f = g \circ \pi^T_I$.*

Beweis: Ist $f = 1_F$, $F \in \mathcal{F}_I$, so gibt es ein $G \in \mathcal{B}^I$ mit $F = (\pi^T_I)^{-1}G$, also $f = 1_G \circ \pi^T_I$. Für allgemeine f folgt die Behauptung nun durch algebraische Induktion. □

20.4 Satz (Charakterisierung der Markov-Eigenschaft für $T = \mathbb{N}, \mathbb{Z}$)
Erfüllt der Prozess $(X_t)_{t\in T}$ die Eigenschaft (ME) für alle $t = s+1$, so auch für alle $t > s$.

Beweis: Induktion nach $n = t - s$: Für $n = 1$ ist das gerade die Voraussetzung. Gelte also (ME) für $t = s + n$, $s \in T$ beliebig. Dann ist

$$\begin{aligned}
P[X_{t+1} \in B \big| \mathcal{F}_{\leq s}] &\overset{\text{S. 19.8}}{=} E\left[P[X_{t+1} \in B \big| \mathcal{F}_{\leq s+1}] \big| \mathcal{F}_{\leq s}\right] \overset{\text{Induktion}}{=} E\left[P[X_{t+1} \in B \big| \mathcal{F}_{s+1}] \big| \mathcal{F}_{\leq s}\right] \\
&\overset{\text{B. 20.2 für } t = s+1}{=} E\left[P[X_{t+1} \in B \big| \mathcal{F}_{s+1}] \big| \mathcal{F}_{s}\right] \overset{\text{Induktion}}{=} E\left[P[X_{t+1} \in B \big| \mathcal{F}_{\leq s+1}] \big| \mathcal{F}_{s}\right] \\
&\overset{\text{S. 19.8}}{=} P[X_{t+1} \in B \big| \mathcal{F}_{s}]
\end{aligned}$$

□

In der Situation $t = s + 1$ wollen wir mit $\tilde{K}^{\leq s}$ und $\tilde{K}^s$ die stochastischen Kerne von $(\mathbb{R}^T, \mathcal{F}_{\leq s})$ bzw. $(\mathbb{R}^T, \mathcal{F}_s)$ nach $(\mathbb{R}, \mathcal{B})$ bezeichnen, die die bedingten Wahrscheinlichkeiten $P[X_{s+1} \in \boldsymbol{.} \big| \mathcal{F}_{\leq s}]$ und $P[X_{s+1} \in \boldsymbol{.} \big| \mathcal{F}_s]$ realisieren. Lemma 20.3 garantiert dann für jeden Prozess $(X_t)_{t \in T}$ (und nicht nur für Markov-Ketten) die Existenz von stochastischen Kernen

$$\begin{aligned}
K^{\leq s} : \mathbb{R}^{\leq s} \times \mathcal{B} \to [0, 1], \quad & \tilde{K}^{\leq s}(\omega, A) = K^{\leq s}(\pi^T_{\leq s}\omega, A) \quad \text{und} \\
K^s : \mathbb{R} \times \mathcal{B} \to [0, 1], \quad & \tilde{K}^s(\omega, B) = K^s(\omega_s, B) \,.
\end{aligned} \tag{*}$$

Ist allerdings $(X_t)_{t \in T}$ eine Markov-Kette, so können diese Kerne so gewählt werden, dass

$$K^{\leq s}(\omega_{\leq s}, A) = K^s(\omega_s, A) \,,$$

wobei $\omega_{\leq s}$ abkürzend für $\pi^T_{\leq s}\omega$ steht. In diesem und im nächsten Kapitel wollen wir den Begriff <u>Markovkern</u> für diese Situation reservieren.

Nun beginnen wir, den Blick von der analytischen in die konstruktive Richtung zu wenden.

20.5 Satz (Rekonstruktion der Prozess-Verteilung aus Kernen $K^{\leq s}$)
Die Verteilung P des Prozesses $(X_t)_{t \in T}$ lässt sich aus der Verteilung $P_{\leq 0} := P \circ (\pi^T_{\leq 0})^{-1}$ des Prozesses $(X_t)_{t \leq 0}$ und den Kernen $K^{\leq s}$, $s \geq 0$, zurückgewinnen: Sei $A^{\leq s} = \times_{i=0}^{\infty} A_i$ mit $A_0 \in \mathcal{B}^{\leq 0}$, $A_i \in \mathcal{B}$ $(i > 0)$ und $A_i = \mathbb{R}$ $(i > s)$. Durch

$$Q\left(A^{\leq s}\right) := \int_{A_0} \int_{A_1} \dots \int_{A_s} K^{\leq s-1}(\omega_{\leq s-1}, d\omega_s) \, \dots \, K^{\leq 0}(\omega_{\leq 0}, d\omega_1) \, dP_{\leq 0}(\omega_{\leq 0}) \tag{**}$$

wird ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^T, \mathcal{B}^T)$ definiert, und es ist $Q = P$. Im Fall einer Markov-Kette ist

$$Q(A^{\leq s}) = \int_{A_0} \int_{A_1} \dots \int_{A_s} K^{s-1}(\omega_{s-1}, d\omega_s) \, \dots \, K^0(\omega_0, d\omega_1) \, dP_{\leq 0}(\omega_{\leq 0}) \,. \tag{**'}$$

Beweis: Der Satz von Ionescu Tulcea (Satz 12.9) garantiert die Existenz eines Wahrscheinlichkeitsmaßes Q auf $(\mathbb{R}^T, \mathcal{B}^T)$ mit der Eigenschaft (**). Wir zeigen, dass $Q = P$, was äquivalent ist zu $Q_{\leq s} = P_{\leq s}$ für alle s. Dazu reicht es, $Q(A^{\leq s}) = P(A^{\leq s})$ für Mengen $A^{\leq s}$ des betrachteten Typs nachzuweisen (siehe Sätze 6.7 und 2.12): Für $s = 0$ ist $Q(A^{\leq 0}) = P_{\leq 0}(A_0) = P(A^{\leq 0})$, und durch Induktion nach s folgt

$$\begin{aligned}
&Q\left(A^{\leq s+1}\right) \\
&= \int_{A_0} \int_{A_1} \dots \int_{A_s} \int_{A_{s+1}} K^{\leq s}(\omega_{\leq s}, d\omega_{s+1}) K^{\leq s-1}(\omega_{\leq s-1}, d\omega_s) \, \dots \, K^{\leq 0}(\omega_{\leq 0}, d\omega_1) \, dP_{\leq 0}(\omega_{\leq 0}) \\
&= \int_{A_0} \int_{A_1} \dots \int_{A_s} P[A_{s+1} \big| \mathcal{F}_{\leq s}](\omega_{\leq s}) K^{\leq s-1}(\omega_{\leq s-1}, d\omega_s) \, \dots \, K^{\leq 0}(\omega_{\leq 0}, d\omega_1) \, dP_{\leq 0}(\omega_{\leq 0}) \\
&= \int 1_{A^{\leq s}} \cdot P[(\pi^T_{s+1})^{-1} A_{s+1} \big| \mathcal{F}_{\leq s}] \, dQ \\
&= \int 1_{A^{\leq s}} \cdot P[(\pi^T_{s+1})^{-1} A_{s+1} \big| \mathcal{F}_{\leq s}] \, dP \quad \text{(nach Induktionsannahme)} \\
&= P(A^{\leq s+1}) \quad \text{(nach Satz 19.7)}
\end{aligned}$$

□

Beim konstruktiven Zugang zum Begriff der Markov-Kette beschränken wir uns auf den Fall $T = \mathbb{N}$ und beginnen mit einer *Startverteilung* P_0 auf $(\mathbb{R}, \mathcal{B})$ und einer Folge von Markovkernen K^s, $s \geq 0$, von $(\mathbb{R}, \mathcal{B})$ nach $(\mathbb{R}, \mathcal{B})$ wie in (*). Damit definieren wir das Wahrscheinlichkeitsmaß Q auf $(\mathbb{R}^T, \mathcal{B}^T)$ wie in (**'). Wie in Beispiel 19.5e zeigt man

$$Q[X_{s+1} \in B \big| \mathcal{F}_{\leq s}](\omega_{\leq s}) = K^s(\omega_s, B)$$

ist $\mathcal{F}_s$-messbar, also

$$Q[X_{s+1} \in B \big| \mathcal{F}_{\leq s}] = E_Q\left[Q[X_{s+1} \in B \big| \mathcal{F}_{\leq s}] \big| \mathcal{F}_s\right] = Q[X_{s+1} \in B \big| \mathcal{F}_s]$$

für alle $B \in \mathcal{B}$. Mit Satz 20.4 folgt daraus die Markov-Eigenschaft (ME) für den Prozess $(X_t)_{t \in T}$ mit der Verteilung Q.

20.6 Übung (Partialsummenprozesse als Markov-Ketten) Seien $\xi_0, \xi_1, \xi_2, \ldots$ unabhängige Zufallsvariablen. Setze $X_n = \sum_{i=0}^n \xi_i$. Zeigen Sie, dass $(X_n)_{n \geq 0}$ eine Markov-Kette ist.

20.7 Bemerkung (Irrfahrten) Sind die ξ_i in der vorigen Übung $\mathbb{Z}$-wertig, so nennt man X_n auch eine *Irrfahrt* auf $\mathbb{Z}$. Sind die ξ_i $\mathbb{Z}^d$-wertig, so erhält man eine Irrfahrt auf dem d-dimensionalen Gitter $\mathbb{Z}^d$.

Wir spezialisieren uns nun auf zeitlich homogene Markov-Ketten auf $\mathbb{R}$: Seien dazu eine Startverteilung P_0 auf $\mathcal{B}$ und ein Markovkern K von $(\mathbb{R}, \mathcal{B})$ nach $(\mathbb{R}, \mathcal{B})$ gegeben. Die Verteilung Q auf $\mathcal{B}^{\mathbb{N}}$ werde mit den Kernen $K^1 = K^2 = \cdots = K$ konstruiert. Die folgende Sichtweise taucht in der Theorie der zeitlich homogenen Markov-Ketten häufig auf: Bezeichne $\mathcal{P}(\mathbb{R})$ den Raum der Borel'schen Wahrscheinlichkeitsmaße auf $\mathbb{R}$. Man betrachtet den (konvex linearen) Operator

$$V_K : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R}), \quad (V_K P)(A) = \int_{\mathbb{R}} K(x, A)\, dP(x) \;\left(= (P \times K)(\pi_1^{-1} A)\right) .$$

20.8 Lemma *In obiger Situation ist $P_{X_s} = V_K^s P_0$ für alle $s > 0$.*

Beweis: Da $P_{X_n}(A) = Q(\pi_n^{-1} A)$, folgt das sofort (durch "konzentriertes Hinsehen") aus (**) in Satz 20.5. Betrachte dort $A^{\leq s}$ mit $A_0 = \cdots = A_{s-1} = \mathbb{R}$, $A_s = A$. □

20.9 Satz (Stationäre Markov-Ketten)
In obiger Situation ist die zeitlich homogene Markov-Kette $(X_t)_{t \in \mathbb{N}}$ stationär genau dann, wenn

$$(V_K P_0)(A) = P_0(A) \tag{***}$$

für alle $A \in \mathcal{B}$. Hinreichend für die Stationarität ist sogar, dass (***) für alle A aus einem $\cap$-stabilen Erzeuger $\mathcal{C}$ von $\mathcal{B}$ gilt.

Beweis: Sei zunächst die Markov-Kette stationär. Dann ist

$$P_0(A) = Q(\pi_0^{-1} A) \overset{\text{Stationarität}}{=} Q(S_1^{-1} \pi_0^{-1} A)) = Q(\pi_1^{-1} A) = P_{X_1}(A) = (V_K P_0)(A)$$

für alle $A \in \mathcal{B}$ nach Lemma 20.8.

Gelte nun umgekehrt (***) für $A \in \mathcal{C}$. Das heißt, die Maße P_0 und $(P_0 \times K) \circ \pi_1^{-1}$ stimmen auf $\mathcal{C}$ und damit auf ganz $\mathcal{B}$ überein (beachte Satz 2.12). Wegen Satz 6.11 reicht es im Fall $T = \mathbb{N}$ zu zeigen, dass

$$Q\left(\bigcap_{k=0}^{n} \{X_{k+1} \in A_k\}\right) = Q\left(\bigcap_{k=0}^{n} \{X_k \in A_k\}\right)$$

für alle $A_0, \ldots, A_n \in \mathcal{B}$ und $n \geq 0$:

$$\begin{aligned}
Q\left(\bigcap_{k=0}^{n}\{X_{k+1} \in A_k\}\right) &= \int_{\mathbb{R}}\int_{A_0} \ldots \int_{A_n} K(\omega_n, d\omega_{n+1}) \,\ldots\, K(\omega_0, d\omega_1)\, dP_0(\omega_0) \\
&= \int_{A_0} \ldots \int_{A_n} K(\omega_n, d\omega_{n+1}) \,\ldots\, K(\omega_1, d\omega_2)\, d\underbrace{\left((P_0 \times K) \circ \pi_1^{-1}\right)}_{=P_0}(\omega_1) \\
&= Q\left(\bigcap_{k=0}^{n}\{X_k \in A_k\}\right) .
\end{aligned}$$

□

20.10 Übung (Autoregressiver Prozess 1. Ordnung) Seien $\xi_0, \xi_1, \xi_2, \ldots$ unabhängige Zufallsvariablen, $P_{\xi_0} = \mathcal{N}(m, \tau^2)$ und $P_{\xi_i} = \mathcal{N}(\mu, \sigma^2)$ für $i \geq 1$. Sei $\alpha \in (-1, 1)$. Definiere X_n rekursiv durch

$$X_0 = \xi_0, \quad X_{n+1} = \alpha X_n + \xi_{n+1}\ (n \geq 0) .$$

($(X_n)_{n\geq 0}$ ist ein *autoregressiver Prozess 1. Ordnung*.)

a) Zeigen Sie, dass $(X_n)_{n\geq 0}$ eine zeitlich homogene Markov-Kette ist und bestimmen Sie den zugehörigen Übergangskern.

b) Bestimmen Sie die gemeinsame Verteilung von $(X_0, \ldots, X_n)$.

c) Bestimmen Sie m und τ^2 so, dass $(X_n)_{n\geq 0}$ stationär ist.

d) Zeigen Sie, dass (für beliebige m und τ^2) gilt: $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} X_k = \frac{\mu}{1-\alpha}$ f.s.

Kapitel 21

Gibbs'sche Zufallsfelder

Im vorigen Abschnitt haben wir die Verteilung eines Prozesses, ausgehend von einer Startverteilung $P_{\leq 0}$ oder P_0 mit Hilfe stochastischer Kerne Schritt für Schritt aufgebaut. Hing dabei die bedingte Verteilung einer neu hinzu genommenen Koordinate (gegeben die bisher betrachteten Koordinaten) nur von der Realisierung der direkten Vorgängerkoordinate ab, dann haben wir von einer Markov-Kette gesprochen.

In diesem Kapitel untersuchen wir wieder eine Markov-Eigenschaft, ändern allerdings unseren Blickwinkel.

- ▷ Anstelle der Indexmengen $\mathbb{N}$ oder $\mathbb{Z}$ betrachten wir allgemeiner die Indexmenge $T = \mathbb{Z}^d$, $d \geq 1$. Das ist aber nicht der wesentliche neue Gesichtspunkt, wir hätten auch den Begriff der Markov-Kette in diese Richtung verallgemeinern können.
- ▷ Wir geben *keine* Startverteilung vor, sondern nur ein *verträgliches* System bedingter Verteilungen und fragen
  1. ob sich dieses System zur Verteilung eines stochastischen Prozesses über T "zusasmmensetzen" lässt,
  2. ob, falls 1. mit "ja" zu beantworten ist, die dort gefundene Verteilung eindeutig durch das System bedingter Verteilungen bestimmt ist.

Motiviert wird diese Betrachtungsweise durch Fragestellungen der statistischen Mechanik. Das werden wir im Anschluss an dieses Kapitel durch ein paar Untersuchungen am *Ising-Modell* illustrieren.

Um alle Messbarkeitsfragen so einfach wie möglich zu halten, beschränken wir uns in diesem Kapitel auf Produkträume

- ▷ $\Omega = \Sigma^T$, wo Σ eine endliche Menge ist und $T = \mathbb{Z}^d$. Ω ist wie üblich mit der von den Zylindermengen erzeugten σ-Algebra ausgerüstet, die in diesem Fall mit der Produkt-σ-Algebra der diskreten σ-Algebren $\mathcal{B} := \mathfrak{P}(\Sigma)$ übereinstimmt, siehe Satz 6.7.
- ▷ $(X_t)_{t \in T}$ bezeichne den kanonischen Koordinatenprozess, siehe Bemerkung 6.13.
- ▷ Für $S \subseteq T$ sei $\mathcal{F}_S := \sigma(X_t : t \in S)$, siehe Definition 5.9.
- ▷ Es sei $\mathcal{E}(T) := \{\Lambda \subseteq T : \Lambda \text{ endlich}\}$ wie in Definition 12.1.
- ▷ Für $\omega \in \Omega$ und $S \subseteq T$ sei $\omega_S = \pi_S^T \omega$.

Zunächst einmal müssen wir uns darüber klar werden, zu welcher Art von Systemen bedingter Verteilungen wir passende Prozessverteilungen finden wollen. Dazu gehen wir das Problem von rückwärts an, d.h. wir beschreiben Eigenschaften, die solche Systeme bedingter Verteilungen sicher haben, wenn sie von einer Prozessverteilung herrühren.

Sei nun P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B}^T)$. Wegen Satz 19.9 zusammen mit Bemerkung 19.10 gibt es für alle $\Lambda \in \mathcal{E}(T)$ einen stochastischen Kern K^Λ von $(\Sigma^{\Lambda^c}, \mathcal{B}^{\Lambda^c})$ nach $(\Sigma^\Lambda, \mathcal{B}^\Lambda)$ derart, dass

$$P\left[\pi_\Lambda^{-1} A \middle| \mathcal{F}_{\Lambda^c}\right](\omega) = K^\Lambda(\omega_{\Lambda^c}, A) \quad \text{für alle } A \in \mathcal{B}^\Lambda. \tag{*}$$

(Beachte, dass $\mathcal{B}^\Lambda = \mathfrak{P}(\Sigma^\Lambda)$, da Σ^Λ endlich ist.)

21.1 Satz (Verträgliche stochastische Kerne)
Die stochastischen Kerne K^Λ aus (*) können so gewählt werden, dass für alle $\Gamma \subseteq \Lambda \in \mathcal{E}(T)$, $A \in \mathcal{B}^\Gamma$ und $B \in \mathcal{B}^{\Lambda\setminus\Gamma}$

$$K^\Lambda\left(\omega_{\Lambda^c}, (\pi^\Lambda_\Gamma)^{-1}A \cap (\pi^\Lambda_{\Lambda\setminus\Gamma})^{-1}B\right) = \int_{(\pi^\Lambda_{\Lambda\setminus\Gamma})^{-1}B} K^\Gamma\left(\omega_{\Gamma^c}, A\right)\, K^\Lambda\left(\omega_{\Lambda^c}, d\omega_\Lambda\right)\ . \qquad (**)$$

Beweis: Für festes $\Gamma \subseteq \Lambda \in \mathcal{E}(T)$ und $A \in \mathcal{B}^\Gamma$, $B \in \mathcal{B}^{\Lambda\setminus\Gamma}$ ist

$$\begin{aligned} K^\Lambda\left(\omega_{\Lambda^c}, (\pi^\Lambda_\Gamma)^{-1}A \cap (\pi^\Lambda_{\Lambda\setminus\Gamma})^{-1}B\right) &= P\left[\pi_\Gamma^{-1}A \cap \pi^{-1}_{\Lambda\setminus\Gamma}B \middle| \mathcal{F}_{\Lambda^c}\right](\omega) \\ &= E\left[P\left[\pi_\Gamma^{-1}A \cap \pi^{-1}_{\Lambda\setminus\Gamma}B \middle| \mathcal{F}_{\Gamma^c}\right] \middle| \mathcal{F}_{\Lambda^c}\right](\omega) \\ &= E\left[P\left[\pi_\Gamma^{-1}A \middle| \mathcal{F}_{\Gamma^c}\right]\, 1_{\pi^{-1}_{\Lambda\setminus\Gamma}B} \middle| \mathcal{F}_{\Lambda^c}\right](\omega) \\ &= E\left[K^\Gamma\left(\omega_{\Gamma^c}, A\right)\, 1_{\pi^{-1}_{\Lambda\setminus\Gamma}B} \middle| \mathcal{F}_{\Lambda^c}\right](\omega) \\ &= \int_{(\pi^\Lambda_{\Lambda\setminus\Gamma})^{-1}B} K^\Gamma\left(\omega_{\Gamma^c}, A\right)\, K^\Lambda\left(\omega_{\Lambda^c}, d\omega_\Lambda\right) \end{aligned}$$

für P-f.a. ω. Da jedes $\mathcal{B}^\Lambda$, $\Lambda \in \mathcal{E}(T)$, endlich und $\mathcal{E}(T)$ abzählbar ist, gibt es eine P-Nullmenge $N \subseteq \Omega$ derart, dass (**) für alle $\Gamma \subseteq \Lambda \in \mathcal{E}(T)$, $A \in \mathcal{B}^\Gamma$, $B \in \mathcal{B}^{\Lambda\setminus\Gamma}$ und $\omega \in \Omega \setminus N$ gilt. Für $\omega \in N$ modifizieren wir die Kerne K^Λ wie folgt:

$$K^\Lambda(\omega_{\Lambda^c}, \,.\,) := P \circ \pi_\Lambda^{-1}\ ,$$

d.h. $K^\Lambda(\omega_{\Lambda^c}, \,.\,)$ wird nicht als bedingte Wahrsacheinlichkeit sondern als (unbedingte) Marginalwahrscheinlichkeit auf $\mathcal{F}_\Lambda$ definiert. Dann gilt (**) natürlich auch für $\omega \in N$. □

21.2 Definition (Verträgliches System von stochastischen Kernen) *a) Ein System $(K^\Lambda : \Lambda \in \mathcal{E}(T))$ von stochastischen Kernen heißt verträglich, falls (**) für alle $\Gamma \subseteq \Lambda \in \mathcal{E}(T)$, $A \in \mathcal{B}^\Gamma$, $B \in \mathcal{B}^{\Lambda\setminus\Gamma}$ und $\omega \in \Omega$ erfüllt ist. (Dabei müssen die Kerne K^Λ nicht wie in (*) aus einem gegebenen Wahrscheinlichkeitsmaß P gewonnen sein.)*

b) Die Kerne heißen stetig, falls die Abbildungen $\omega \mapsto K^\Lambda(\omega_{\Lambda^c}, A)$ für alle $\Lambda \in \mathcal{E}(T)$ und alle $A \in \mathcal{B}^\Lambda$ stetig sind.

Wurde ein System von stochastischen Kernen wie in (*) aus einem Wahrscheinlichkeitsmaß P auf Ω gewonnen, so folgt sofort, dass

$$P\left(\pi_\Lambda^{-1}A\right) = \int_\Omega K^\Lambda(\omega_{\Lambda^c}, A)\, dP(\omega) \quad \text{für alle } \Lambda \in \mathcal{E}(T) \text{ und } A \in \mathcal{B}^\Lambda. \qquad \text{(DLR)}$$

21.3 Definition (Gibbs-Maß) *Ein Wahrscheinlichkeitsmaß P, das die Bedingung (DLR)[1] erfüllt, heißt Gibbs-Maß zum System $(K^\Lambda : \Lambda \in \mathcal{E}(T))$.*

Obwohl die Bedingung (DLR) zunächst schwächer aussieht als (*), ist sie doch zu ihr äquivalent, wie wir im folgenden Satz zumindest im Fall stetiger Kerne zeigen.

[1](DLR) steht für *Dobrushin-Lanford-Ruelle*, drei der führenden Vertreter dieser Forschungsrichtung.

21.4 Satz
Sei P ein Gibbs-Maß zum System $(K^\Lambda : \Lambda \in \mathcal{E}(T))$ stetiger Kerne. Dann gilt (*) für alle $\Lambda \in \mathcal{E}(T)$, d.h.

$$P\left[\pi_\Lambda^{-1} A \middle| \mathcal{F}_{\Lambda^c}\right](\omega) = K^\Lambda(\omega_{\Lambda^c}, A) \quad \text{für alle } A \in \mathcal{B}^\Lambda.$$

Beweis: Seien $\Lambda_n \in \mathcal{E}(T)$, $\Lambda_n \nearrow T$. Da $\mathcal{F}_{\Lambda^c} = \sigma\left(\bigcup_n \left\{\pi_{\Lambda_n \setminus \Lambda}^{-1} B : B \in \mathcal{B}^{\Lambda_n \setminus \Lambda}\right\}\right)$, reicht es zu zeigen, dass für große n (so groß, dass $\Lambda \subseteq \Lambda_n$)

$$P\left(\pi_\Lambda^{-1} A \cap \pi_{\Lambda_n \setminus \Lambda}^{-1} B\right) = \int_{\pi_{\Lambda_n \setminus \Lambda}^{-1} B} K^\Lambda(\omega_{\Lambda^c}, A)\, dP(\omega)$$

für alle $B \in \mathcal{B}^{\Lambda_n \setminus \Lambda}$: Da $\omega \mapsto K^\Lambda(\omega_{\Lambda^c}, A)$ stetig ist, gibt es zu $\delta > 0$ ein $N \geq n$, so dass diese Funktion auf jedem Elementarzylinder $\pi_{\Lambda_N}^{-1}\{w\}$ mit $w \in \Sigma^{\Lambda_N}$ um höchstens δ oszilliert. Insbesondere gibt es eine Funktion $F_N : \Sigma^{\Lambda_N} \to \mathbb{R}$ mit $|K^\Lambda(\omega_{\Lambda^c}, A) - F_N(\omega_{\Lambda_N})| \leq \delta$. Es folgt

$$\begin{aligned}
P\left(\pi_\Lambda^{-1} A \cap \pi_{\Lambda_n \setminus \Lambda}^{-1} B\right) &\overset{\text{(DLR)}}{=} \int_\Omega K^{\Lambda_N}\left(\omega_{\Lambda_N^c}, (\pi_\Lambda^{\Lambda_N})^{-1} A \cap (\pi_{\Lambda_n \setminus \Lambda}^{\Lambda_N})^{-1} B\right) dP(\omega) \\
&\overset{(**)}{=} \int_\Omega \int_{(\pi_{\Lambda_n \setminus \Lambda}^{\Lambda_N})^{-1} B} K^\Lambda(\omega_{\Lambda^c}, A)\, K^{\Lambda_N}(\omega_{\Lambda_N^c}, d\omega_{\Lambda_N})\, dP(\omega) \\
&= \int_\Omega \int_{(\pi_{\Lambda_n \setminus \Lambda}^{\Lambda_N})^{-1} B} F_N(\omega_{\Lambda_N})\, K^{\Lambda_N}(\omega_{\Lambda_N^c}, d\omega_{\Lambda_N})\, dP(\omega) + \delta_1 \\
&\overset{\text{(DLR)}}{=} \int_{\pi_{\Lambda_n \setminus \Lambda}^{-1} B} F_N(\omega_{\Lambda_N})\, dP(\omega) + \delta_1 \\
&= \int_{\pi_{\Lambda_n \setminus \Lambda}^{-1} B} K^\Lambda(\omega_{\Lambda^c}, A)\, dP(\omega) + \delta_1 + \delta_2
\end{aligned}$$

mit $|\delta_1|, |\delta_2| \leq \delta$. □

In Analogie zum Satz 12.3 von Kolmogorov stellt sich nun die Frage, ob zu jedem verträglichen System stetiger stochastischer Kerne genau ein Gibbs-Maß existiert. Die Antwort ist in diesem Fall leider viel schwieriger. Die Existenz eines solchen P kann zwar gezeigt werden, die Eindeutigkeit ist aber oft nicht gegeben – in der Tat werden wir im nachfolgenden Abschnitt sehen, dass gerade dadurch diese Systeme zur Modellierung von Phasenübergängen geeignet sind.

Um diesen Fragen systematisch nachgehen zu können, führen wir folgende Bezeichnungen bei fest gegebenem verträglichem System $(K^\Lambda : \Lambda \in \mathcal{E}(T))$ stetiger stochastischer Kerne ein:

- ▷ $\mathcal{S}$ ist der Raum aller signierten Maße auf $(\Omega, \mathcal{B}^T)$. Durch $(\alpha\varphi_1 + \beta\varphi_2)(A) := \alpha\varphi_1(A) + \beta\varphi_2(A)$ wird $\mathcal{S}$ zu einem reellen Vektorraum. Für $\varphi = \varphi^+ - \varphi^-$ definiert man $\int f\, d\varphi := \int f\, d\varphi^+ - \int f\, d\varphi^-$ und zeigt durch algebraische Induktion, dass $\int f\, d(\alpha\varphi_1 + \beta\varphi_2) = \alpha \int f\, d\varphi_1 + \beta \int f\, d\varphi_2$.

- ▷ Für $\Lambda \in \mathcal{E}(T)$ definiere

$$V_\Lambda : \mathcal{S} \to \mathcal{S}, \quad (V_\Lambda\varphi)(C) = \int_{\Sigma^{\Lambda^c}} \int_{\Sigma^\Lambda} 1_C(\omega_\Lambda \vee \omega_{\Lambda^c})\, K^\Lambda(\omega_{\Lambda^c}, d\omega_\Lambda)\, d(\varphi \circ \pi_{\Lambda^c}^{-1})(\omega_{\Lambda^c})$$

  ($\omega_\Lambda \vee \omega_{\Lambda^c}$ bezeichnet diejenige Konfiguration aus Ω, die auf Λ mit ω_Λ und auf Λ^c mit ω_{Λ^c} übereinstimmt.) Ist $\varphi \geq 0$, also ein endliches Maß, so wird dadurch ein endliches Maß $V_\Lambda\varphi$ auf $\mathcal{B}^T$ definiert. Für allgemeine $\varphi \in \mathcal{S}$ setzt man $V_\Lambda\varphi := V_\Lambda\varphi^+ - V_\Lambda\varphi^-$.

- ▷ Sei $\mathcal{G} := \{\varphi \in \mathcal{S} : V_\Lambda\varphi = \varphi\, \forall \Lambda \in \mathcal{E}(T)\}$, $\mathcal{G}^+ := \{\varphi \in \mathcal{G} : \varphi \geq 0\}$, $\mathcal{G}_1^+ := \{\mu \in \mathcal{G}^+ : \mu(\Omega) = 1\}$.

21.5 Lemma *$\mathcal{G}_1^+$ ist gerade die Menge aller Gibbs-Maße.*

Beweis: Ist $V_\Lambda\,\varphi = \varphi$ für alle $\Lambda \in \mathcal{E}(T)$, so ist

$$\begin{aligned}
\varphi(\pi_\Lambda^{-1}A) &= (V_\Lambda\varphi)(\pi_\Lambda^{-1}A)\\
&= \int_{\Sigma^{\Lambda^c}}\int_{\Sigma^\Lambda} 1_A(\omega_\Lambda)\,K^\Lambda(\omega_{\Lambda^c}, d\omega_\Lambda)\,d(\varphi\circ\pi_{\Lambda^c}^{-1})(\omega_{\Lambda^c})\\
&= \int_\Omega K^\Lambda(\omega_{\Lambda^c}, A)\,d\varphi(\omega)
\end{aligned}$$

für alle Zylindermengen der Form $\pi_\Lambda^{-1}A$ mit $\Lambda \in \mathcal{E}(T)$ und $A \in \mathcal{B}^\Lambda$, d.h. es gilt (DLR) für φ.

Ist umgekehrt φ ein Gibbs-Maß, so folgt aus Satz 21.4 für $C = \pi_\Lambda^{-1}A \cap \pi_{\Lambda^c}^{-1}B$

$$\begin{aligned}
(V_\Lambda\varphi)(C) &= \int_{\Sigma^{\Lambda^c}}\int_{\Sigma^\Lambda} 1_A(\omega_\Lambda)\,1_B(\omega_{\Lambda^c})\,K^\Lambda(\omega_{\Lambda^c}, d\omega_\Lambda)\,d(\varphi\circ\pi_{\Lambda^c}^{-1})(\omega_{\Lambda^c})\\
&= \int_{\pi_{\Lambda^c}^{-1}B} \varphi\left[\pi_\Lambda^{-1}A \middle| \mathcal{F}_{\Lambda^c}\right](\omega)\,d\varphi(\omega)\\
&= \varphi(C)
\end{aligned}$$

□

Die Frage nach der Existenz und Eindeutigkeit eines Gibbs-Maßes lässt sich nun so formulieren:

- ▷ *Existenz*: $\operatorname{card}\mathcal{G}_1^+ \geq 1$?
- ▷ *Eindeutigkeit*: $\operatorname{card}\mathcal{G}_1^+ \leq 1$?

Um uns diesen Fragen nähern zu können, sammeln wir einige Eigenschaften von $\mathcal{G}$ und V im folgenden Lemma.

21.6 Lemma *Für alle $\Lambda \in \mathcal{E}(T)$ gilt:*

1) $V_\Lambda : \mathcal{S} \to \mathcal{S}$ ist linear.

2) $V_\Lambda(\mathcal{G}^+) \subseteq \mathcal{G}^+$.

3) $(V_\Lambda\varphi)(\Omega) = \varphi(\Omega)$ für alle $\varphi \in \mathcal{S}$.

4) $\varphi \in \mathcal{G} \Rightarrow \varphi^+, \varphi^- \in \mathcal{G}^+$.

Beweis:

1) Folgt aus der Definition.

2) Folgt aus der Positivität der Kerne K^Λ und der Positivität des Integrals.

3) Folgt aus $K^\Lambda(\omega_{\Lambda^c}, \Sigma^\Lambda) = 1$.

4) Sei $\varphi \in \mathcal{G}$, $\Omega = \Omega^+ \cup \Omega^-$ eine Hahn-Zerlegung von Ω (Satz 18.9) und $\varphi = \varphi^+ - \varphi^-$ die Jordan-Zerlegung von φ, also $\varphi^+(\Omega^-) = \varphi^-(\Omega^+) = 0$ (Korollar 18.11). Sei B messbar, $B^+ := B \cap \Omega^+$. Dann ist für alle $\Lambda \in \mathcal{E}(T)$

$$\varphi^+(B) = \varphi(B^+) \overset{\varphi\in\mathcal{G}}{=} (V_\Lambda\varphi)(B^+) \overset{V_\Lambda\,\text{linear}}{=} (V_\Lambda\varphi^+)(B^+) - (V_\Lambda\varphi^-)(B^+) \overset{V_\Lambda\varphi^-\geq 0}{\leq} (V_\Lambda\varphi^+)(B)\;.$$

Es folgt

$$\varphi^+(B) = \varphi^+(\Omega) - \varphi^+(B^c) \geq (V_\Lambda\varphi^+)(\Omega) - (V_\Lambda\varphi^+)(B^c) = (V_\Lambda\varphi^+)(B)\;,$$

also $\varphi^+(B) = (V_\Lambda\varphi^+)(B)$, und da $V_\Lambda\varphi^- = V_\Lambda(\varphi^+ - \varphi) \overset{V_\Lambda\,\text{linear}}{=} \varphi^+ - \varphi = \varphi^-$, sind $\varphi^+, \varphi^- \in \mathcal{G}^+$.

□

Die letzte Aussage dieses Lemmas erlaubt es, das folgende Eindeutigkeitskriterium zu beweisen: Für $n \geq 0$ sei

$$\Lambda_n := \{t \in \mathbb{Z}^d : |t_i| \leq n \ (i = 1, \ldots, d)\} .$$

21.7 Satz (Eindeutigkeitskriterium für Gibbs-Maße)
Sei $(K^\Lambda : \Lambda \in \mathcal{E}(T))$ ein verträgliches System stetiger stochastischer Kerne. Gibt es eine Konstante $C > 0$ derart, dass

$$\frac{K^{\Lambda_n}(\omega_{\Lambda_n^c}, A)}{K^{\Lambda_n}(\tilde{\omega}_{\Lambda_n^c}, A)} \leq C \quad \text{für alle } n \geq 0,\ A \in \mathcal{B}^{\Lambda_n} \text{ und } \omega, \tilde{\omega} \in \Omega , \tag{***}$$

so ist $\operatorname{card} \mathcal{G}_1^+ \leq 1$.

Beweis: Angenommen, es gibt $\mu, \nu \in \mathcal{G}_1^+$, $\mu \neq \nu$. Dann ist $\varphi := \mu - \nu \in \mathcal{G} \setminus \{0\}$, und wegen Lemma 21.6(4) sind $\varphi^+, \varphi^- \in \mathcal{G}^+ \setminus \{0\}$. Außerdem ist $\varphi^+(\Omega) - \varphi^-(\Omega) = \mu(\Omega) - \nu(\Omega) = 0$. Daher ist für $n \geq 0$ und $A \in \mathcal{B}^{\Lambda_n}$

$$\varphi^+(\pi_{\Lambda_n}^{-1} A) = \int K^{\Lambda_n}(\omega_{\Lambda_n^c}, A)\, d\varphi^+(\omega) \leq \sup_\omega K^{\Lambda_n}(\omega_{\Lambda_n^c}, A)\, \varphi^+(\Omega) ,$$
$$\varphi^-(\pi_{\Lambda_n}^{-1} A) = \int K^{\Lambda_n}(\omega_{\Lambda_n^c}, A)\, d\varphi^-(\omega) \geq \inf_\omega K^{\Lambda_n}(\omega_{\Lambda_n^c}, A)\, \varphi^-(\Omega) ,$$

also

$$\frac{\varphi^+(\pi_{\Lambda_n}^{-1} A)}{\varphi^-(\pi_{\Lambda_n}^{-1} A)} \leq C \, \frac{\varphi^+(\Omega)}{\varphi^-(\Omega)} = C .$$

Da $\{\pi_{\Lambda_n}^{-1} A : n \geq 0, A \in \mathcal{B}^{\Lambda_n}\}$ eine $\mathcal{B}^T$ erzeugende Algebra ist, folgt z.B. aus dem Approximationssatz 3.13, dass $\varphi^+ \leq C\varphi^-$, im Widerspruch zu $\varphi^+ \perp \varphi^-$ (siehe Korollar 18.11). □

Um die *Existenz* von Gibbs-Maßen zu zeigen, machen wir von den sehr günstigen topologischen Eigenschaften von $\Omega = \Sigma^T$ Gebrauch. Die wichtigsten sind hier zusammengestellt:

- ▷ Auf dem Gitter $T = \mathbb{Z}^d$ führen wir den Abstand $|t - s|_1 := \sum_{i=1}^d |t_i - s_i|$ ein.
- ▷ Auf Ω betrachten wir die Metrik $d(\omega, \omega') := \sum_{t \in T} \delta_{\omega_t \neq \omega'_t} 2^{-|t|_1}$. (Beachte, dass $\sum_{t \in T} 2^{-|t|_1} \leq \sum_{n=0}^\infty (2n+1)^d 2^{-n} < \infty$.) Für $T = \mathbb{N}$ ist das gerade die in Bemerkung 1.3 diskutierte Metrik. Man sieht leicht, dass alle dort verwendeten Argumente nur die Abzählbarkeit von $T = \mathbb{N}$ benutzen, sodass (Ω, d) auch im Fall $T = \mathbb{Z}^d$ ein kompakter metrischer Raum ist.
- ▷ Insbesondere ist die Menge aller Borel'schen Wahrscheinlichkeitsmaße auf Ω straff, also relativ folgenkompakt in der Topologie der schwachen Konvergenz (siehe Satz 15.14).
- ▷ Sei $\mathcal{Z} := \bigcup_{\Lambda \in \mathcal{E}(T)} \{Z_w := (\pi_\Lambda^T)^{-1}\{w\} : w \in \Sigma^\Lambda\}$ die Familie der (elementaren) Zylindermengen in Ω. Dann sind alle $Z \in \mathcal{Z}$ offen und abgeschlossen. Daher sind alle Indikatorfunktionen 1_Z, $Z \in \mathcal{Z}$, und damit auch alle bzgl. einer der endlichen σ-Algebren $\mathcal{F}_\Lambda$, $\Lambda \in \mathcal{E}(T)$ messbaren Funktionen stetig.

Ausgerüstet mit diesem Wissen können wir nun den folgenden Existenzsatz beweisen.

21.8 Satz (Existenz von Gibbs-Maßen)
Sei $T = \mathbb{Z}^d$ und $(K^\Lambda : \Lambda \in \mathcal{E}(T))$ ein verträgliches System stetiger stochastischer Kerne. Dann ist $\mathcal{G}_1^+ \neq \emptyset$, d.h. es existiert ein Gibbsmaß zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$.

Beweis: Wie vorher sei für $n \geq 0$

$$\Lambda_n := \{t \in \mathbb{Z}^d : |t_i| \leq n \ (i = 1, \dots, d)\} .$$

Sei P ein *beliebiges* Wahrscheinlichkeitsmaß auf Ω, z.B. $P = \delta_\omega$ für ein $\omega \in \Omega$. Für $n \geq 0$ betrachten wir die Wahrscheinlichkeitsmaße

$$P_n := (P \circ \pi_{\Lambda_n^c}^{-1}) \times K^{\Lambda_n}$$

auf Ω. Zur Erinnerung: Für $A \in \mathcal{B}^{\Lambda_n}$ ist dann

$$P_n(\pi_{\Lambda_n}^{-1} A) = \int K^{\Lambda_n}(\omega_{\Lambda_n^c}, A) \, d(P \circ \pi_{\Lambda_n^c}^{-1})(\omega_{\Lambda_n^c}) . \tag{§}$$

Da die Folge $(P_n)_{n \geq 0}$ straff ist, gibt es eine Teilfolge $(P_{n_i})_i$ und ein Wahrscheinlichkeitsmaß μ auf Ω derart, dass $P_{n_i} \Longrightarrow \mu$.

Wir zeigen, dass $\mu \in \mathcal{G}_1^+$: Sei $A \in \mathcal{B}^\Lambda$ für ein $\Lambda \in \mathcal{E}(T)$. Wähle i_0 so, dass $\Lambda \subseteq \Lambda_{n_i}$ für alle $n_i \geq N := n_{i_0}$. Zu zeigen ist

$$\mu\left(\pi_\Lambda^{-1} A\right) = \int K^\Lambda(\omega_{\Lambda^c}, A) \, d\mu(\omega) . \tag{DLR}$$

Sei $\tilde{A} = (\pi_\Lambda^{\Lambda_N})^{-1} A$. Wegen der Verträglichkeitseigenschaft (**) ist (DLR) äquivalent zu

$$\mu\left(\pi_{\Lambda_N}^{-1} \tilde{A}\right) = \int K^{\Lambda_N}(\omega_{\Lambda_N^c}, \tilde{A}) \, d\mu(\omega) ,$$

und wir können ohne Einschränkung annehmen, dass $\Lambda = \Lambda_N$.

Da die Funktionen $\omega \mapsto K^\Lambda(\omega_{\Lambda^c}, A)$ und $\omega \mapsto 1_A(\omega)$ nach Voraussetzung stetig sind, ist

$$\begin{aligned}
&\int_\Omega K^\Lambda(\omega_{\Lambda^c}, A) \, d\mu(\omega) \\
=\ & \lim_{i \to \infty} \int_\Omega K^\Lambda(\omega_{\Lambda^c}, A) \, dP_{n_i}(\omega) \\
\overset{(§)}{=}\ & \lim_{i \to \infty} \int_\Omega \int_{\Sigma^{\Lambda_{n_i}}} K^\Lambda(\omega_{\Lambda_{n_i}^c} \vee \tilde{\omega}_{\Lambda_{n_i} \setminus \Lambda}, A) \, K^{\Lambda_{n_i}}(\omega_{\Lambda_{n_i}^c}, d\tilde{\omega}_{\Lambda_{n_i}}) \, d(P \circ \pi_{\Lambda_{n_i}^c}^{-1})(\omega_{\Lambda_{n_i}^c}) \\
\overset{\text{D.21.2}}{=}\ & \lim_{i \to \infty} \int_\Omega K^{\Lambda_{n_i}}(\omega_{\Lambda_{n_i}^c}, (\pi_\Lambda^{\Lambda_{n_i}})^{-1} A) \, d(P \circ \pi_{\Lambda_{n_i}^c}^{-1})(\omega_{\Lambda_{n_i}^c}) \\
\overset{(§)}{=}\ & \lim_{i \to \infty} P_{n_i}\left(\pi_{\Lambda_{n_i}}^{-1} (\pi_\Lambda^{\Lambda_{n_i}})^{-1} A\right) \\
=\ & \lim_{i \to \infty} P_{n_i}\left(\pi_\Lambda^{-1} A\right) \\
=\ & \mu(\pi_\Lambda^{-1} A) ,
\end{aligned}$$

d.h. es gilt (DLR). □

Die Stetigkeitsvoraussetzung von Satz 21.8 ist insbesondere dann erfüllt, wenn die Kerne K^Λ die folgende Markov-Eigenschaft haben. Für $\Lambda \in \mathcal{E}(T)$ sei

$$\mathcal{U}(\Lambda) := \{t \in T : \ \exists s \in \Lambda \text{ so dass } |s - t|_1 \leq 1\} .$$

$\mathcal{U}(\Lambda)$ besteht also aus der Menge Λ und all ihren direkten "Nachbarn". Insbesondere ist mit Λ auch $\mathcal{U}(\Lambda)$ endlich.

21.9 Definition (Markovkern) *Der stochastische Kern K^Λ ist ein Markovkern, falls die Abbildungen $\omega \mapsto K^\Lambda(\omega_{\Lambda^c}, A)$ für alle $A \in \mathcal{B}^\Lambda$ messbar sind bzgl. $\mathcal{F}_{\mathcal{U}(\Lambda)}$.*

21.10 Lemma (Stetigkeit von Markovkernen) *Ist K^Λ ein Markovkern, so sind die Abbildungen $\omega \mapsto K^\Lambda(\omega_{\Lambda^c}, A)$ für alle $A \in \mathcal{B}^\Lambda$ stetig.*

Beweis: Mit Λ ist auch $\mathcal{U}(\Lambda)$ endlich, so dass $\omega \mapsto K^\Lambda(\omega_{\Lambda^c}, A)$ für jedes A konstant auf Zylindern über $\mathcal{U}(\Lambda)$ und damit stetig ist. □

Beispiel: Ising-Modell

Wir schließen an die Notation des vorigen Kapitels an, betrachten aber nur die Situation $\Sigma = \{+1, -1\}$, $T = \mathbb{Z}^d$. Es ist also $\Omega = \{-1, +1\}^{\mathbb{Z}^d}$. Ein $\omega \in \Omega = \Sigma^T$ wird nun interpretiert als *Konfiguration* von positiven oder negativen "spins", die auf einem "Gitter" T verteilt sind. Man könnte z.B. an eine idealisierte Kristallstruktur denken. Modelliert werden soll die Situation, dass auf dem Gitter benachbarte "spins" sich dahin gehend gegenseitig beeinflussen, dass sie eine mehr oder weniger starke Tendenz haben, sich gleich auszurichten, d.h. den gleichen Wert anzunehmen. Eine völlig homogene Ausrichtung (nur $+1$ oder nur -1) wird aber durch "thermische Fluktuationen" verhindert, die eine um so größere Abweichung von der Homogenität verursachen, je höher die "Temperatur" ist.

Um Missverständnissen vorzubeugen: Es geht hier nicht darum, den zeitlichen Verlauf solcher "thermisch induzierter" Fluktuationen zu modellieren, sondern es ist unser Ziel, "Momentaufnahmen" typischer Konfigurationen ω zu beschreiben, also mehr oder weniger zufällige $(-1, +1)$ - Muster als typische Realisierungen eines Wahrscheinlichkeitsmaßes auf Ω zu identifizieren. Wir bewegen uns daher im Rahmen der sogenannten *statistischen Mechanik von Gleichgewichtszuständen*.

Wir beginnen mit einigen idealisierenden Annahmen, die es uns ermöglichen zu einem mathematisch handhabbaren Modell zu gelangen.

- ▷ *Wechselwirkungen* bestehen nur zwischen benachbarten "spins", d.h. zwischen Gitterpositionen $s, t \in T$ mit $|s-t|_1 = 1$. Sei daher

$$\mathcal{N} := \{\{s,t\} : s,t \in T, |s-t|_1 = 1\} \;\text{ und }\; \mathcal{N}^\Lambda := \{N \in \mathcal{N} : N \cap \Lambda \neq \emptyset\} \text{ für } \Lambda \in \mathcal{E}(T).$$

- ▷ Das *Wechselwirkungspotential* einer Konfiguration $\omega \in \Omega$ im Bereich $\Lambda \in \mathcal{E}(T)$ wird definiert als

$$H^\Lambda(\omega_\Lambda | \omega_{\Lambda^c}) := - \sum_{\{s,t\} \in \mathcal{N}^\Lambda} \omega_s \omega_t$$

  Das Wechselwirkungspotential ist also hoch, wenn es viele benachbarte Paare nicht gleich ausgerichteter "spins" in Λ gibt. Es ist am geringsten, wenn ω konstant $+1$ oder konstant -1 ist. Das sind Konfigurationen minimaler Energie. Beachte, dass $H^\Lambda(\omega_\Lambda | \omega_{\Lambda^c})$ nur von den ω_t mit $t \in \mathcal{U}(\Lambda)$ abhängt, vergleiche Definition 21.9. Die H^Λ werden auch als *Hamilton-Funktionen* bezeichnet.

Nun wollen wir zunächst für eine feste *Randbedingung* ω_{Λ^c} und $\eta \in \mathbb{R}$ ein Wahrscheinlichkeitsmaß μ_η^Λ auf Σ^Λ bestimmen, das dem totalen Zufall (d.h. der Gleichverteilung m^Λ auf Σ^Λ) möglichst nahe kommt, das aber $\int H^\Lambda(\omega_\Lambda | \omega_{\Lambda^c})\, d\mu_\eta^\Lambda(\omega_\Lambda) = \eta$ erfüllt. In einem Satz über Informationsminimierung unter Nebenbedingungen (Satz 13.7) haben wir gesehen, dass eine solche Verteilung μ Dichte

$$f_\beta^\Lambda(\omega_\Lambda | \omega_{\Lambda^c}) = \exp\left(-\gamma_\beta(\omega_{\Lambda^c}) - \beta\, H^\Lambda(\omega_\Lambda | \omega_{\Lambda^c})\right) = \frac{\exp\left(\beta \sum_{\{s,t\} \in \mathcal{N}^\Lambda} \omega_s \omega_t\right)}{\exp \gamma_\beta(\omega_{\Lambda^c})}$$

bzgl. m^Λ hat, wobei $\beta = \beta(\eta)$. Dort hatten wir ϑ an Stelle von $-\beta$ geschrieben, aber wegen des physikalischen Hintergrunds des jetzt betrachteten Modells benutzen wir den Parameter β, der als *inverse Temperatur* interpretiert wird.[2]

[2] In der Physik wird er noch mit der sogenannten *Boltzmann-Konstanten* skaliert, um alle Größen in den üblichen Einheiten ausdrücken zu können. Darauf verzichten wir hier aber und integrieren die Boltzmann-Konstante in die inverse Temperatur.

Damit haben wir Verteilungen

$$K_\beta^\Lambda(\omega_{\Lambda^c}, A) := \int_A f_\beta^\Lambda(\omega_\Lambda | \omega_{\Lambda^c})\, dm^\Lambda(\omega_\Lambda) \tag{*}$$

auf Σ^Λ gefunden, die

- ▷ bei sehr hoher Temperatur ($\beta \searrow 0$) die Gleichverteilung approximieren, also fast sicher völlig ungeordnete Konfigurationen produzieren,
- ▷ bei sehr niedriger absoluter Temperatur ($\beta \nearrow +\infty$) Konfigurationen minimaler Energie, also (nahezu) konstante Konfigurationen bevorzugen.

Dieser Effekt ist in Abbildung 21.1 recht deutlich zu sehen.[3]

Allerdings ist dieser Übergang bei endlichem Λ kontinuierlich, ja sogar völlig glatt, in β. Insbesondere fällt es schwer, das folgende physikalische Phänomen eines *Phasenübergangs* in diesem Modell wieder zu finden: Bei niedriger Temperatur gibt es zwei stabile Gleichgewichtszustände: einen, in dem die $+1$'en stark überwiegen, und einen, in dem die -1'en stark überwiegen, während es bei hohen Temperaturen nur einen Gleichgewichtszustand gibt, bei dem $+1$ und -1 je zur Hälfte vertreten sind.

Die Lösung dieses Problems liegt in der Beobachtung, dass deutliche Phasenübergänge nur auf "großen" Gitterbereichen Λ auftreten. Sie sind daher angemessener auf einem unendlichen Gitter, z.B. auf ganz $T = \mathbb{Z}^d$, als auf einem sehr großen endlichen Gitter zu modellieren.[4] Allerdings wird durch (*) nun nicht die Verteilung selbst angegeben, sondern wir können (*) nur als bedingte Wahrscheinlichkeit interpretieren, dass im endlichen "Fenster" Λ der Ausschnitt ω_Λ einer Konfiguration zu sehen ist, falls der Rand $\partial\Lambda := \mathcal{U}(\Lambda) \setminus \Lambda$ mit $\omega_{\partial\Lambda}$ belegt ist. Dahinter sollte natürlich eine Verteilung μ auf Ω stehen, für die

$$\mu\left[\pi_\Lambda^{-1} A \middle| \mathcal{F}_{\Lambda^c}\right](\omega) = K_\beta^\Lambda(\omega_{\Lambda^c}, A)\ .$$

Nun scheint die Frage nach der Koexistenz von Gleichgewichtszuständen vor dem Hintergrund der Diskussion in Kapitel 21 auch in unserem sehr einfachen mathematischen Modell sinnvoll.

A) Wir müssen zeigen, dass $(K_\beta^\Lambda : \Lambda \in \mathcal{E}(T))$ für jedes $\beta \geq 0$ ein verträgliches System von stochastischen Kernen bildet, und

B) wir müssen zeigen, dass für ausreichend große β (sehr niedrige Temperatur) mindestens zwei Gibbs-Maße existieren.

(Wir sehen also Gibbs-Maße als die adäquate mathematische Beschreibung von Gleichgewichtszuständen an.) Damit ist das mathematische Programm für den Rest dieses Abschnitts umrissen. Wir formulieren zunächst die Resultate, illustrieren einige von ihnen durch Simulationsergebnisse, und liefern die Beweise am Ende nach.

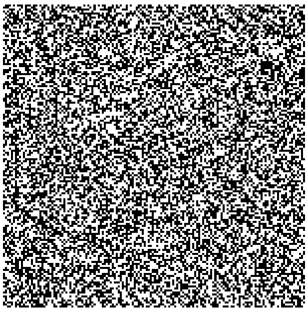
21.11 Lemma *Für jedes $\beta \geq 0$ bildet $(K_\beta^\Lambda : \Lambda \in \mathcal{E}(T))$ ein verträgliches System von stochastischen Kernen.*

Bezeichne nun $\mathcal{G}_1^+(\beta)$ die Familie der Gibbs-Maße zu $(K_\beta^\Lambda : \Lambda \in \mathcal{E}(T))$.

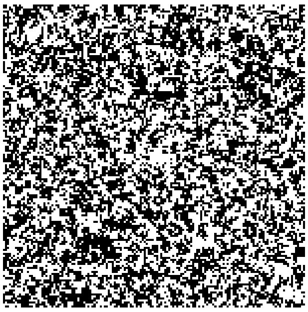
21.12 Satz (Eindeutigkeit des Gleichgewichtszustands im Ising-Modell in Dimension $d = 1$)
Auf dem Gitter $T = \mathbb{Z}$ ist $\operatorname{card} \mathcal{G}_1^+(\beta) = 1$ für jedes $\beta \geq 0$, d.h. es gibt keine koexistierenden Gleichgewichtszustände.

[3] Die Abbildungen 21.1 und 21.3 wurden mir von A. Klenke zur Verfügung gestellt.

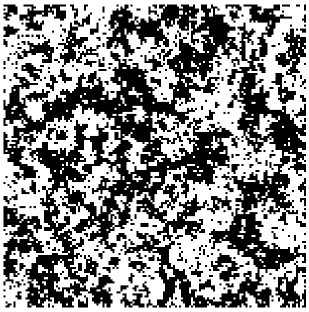
[4] Diese Situation ist vergleichbar damit, dass die Gesetze der Mechanik besser durch Sätze, der reellen Analysis - also mit den idealisierten reellen Zahlen - als durch Sätze einer diskreten Analysis, deren kleinste Einheit aus der Quantentheorie entlehnt ist, zu beschreiben sind.



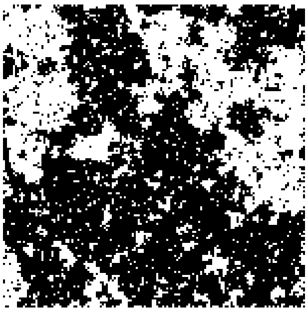
$\beta = 0$



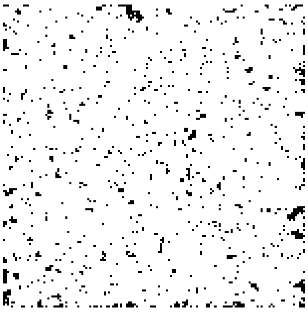
$\beta = 0.25$



$\beta = 0.375$



$\beta = 0.4407$



$\beta = 0.5$

Abbildung 21.1: Typische Ising Konfigurationen auf einem 150×150-Gitter. $\beta = 0.4407$ ist ein "kritischer" Parameter; für größere β überwiegt eine Art von "spins". Wegen der $+/-$ Symmetrie des Models kann das $+$ or $-$ mit gleicher Wahrscheinlichkeit sein.

21.13 Satz (Phasenkoexistenz im Ising-Modell bei niedriger Temperatur in Dimension $d = 2$)
Auf dem Gitter $T = \mathbb{Z}^2$ ist $\operatorname{card} \mathcal{G}_1^+(\beta) \geq 2$ für hinreichend großes β. (In der Tat gibt es $\mu^+, \mu^- \in \mathcal{G}_1^+(\beta)$ mit $\mu^+ \perp \mu^-$.)

21.14 Bemerkung Dieser Satz bleibt auch in Dimension $d > 2$ richtig. Darüber hinaus kann man zeigen, dass

$$\forall\, d \geq 2\, \exists \beta_c(d) > 0\, \forall \beta \in [0, \beta_c(d))\,:\ \operatorname{card} \mathcal{G}_1^+(\beta) = 1.$$

Beweis von Lemma 21.11: Um die recht aufwendige Notation etwas zu entlasten, setzen wir $\beta = 1$ im Beweis und unterdrücken es in allen Bezeichnungen. Der Beweis für allgemeine β verläuft ganz genau so. Seien $\Lambda_1 \subseteq \Lambda_2 \in \mathcal{E}(T)$. Zu zeigen ist:

$$\int_{\Sigma^{\Lambda_2}} K^{\Lambda_1}\left(\omega_{\Lambda_1^c}, A\right)\, K^{\Lambda_2}\left(\omega_{\Lambda_2^c}, d\omega_{\Lambda_2}\right) = K^{\Lambda_2}\left(\omega_{\Lambda_2^c}, (\pi^{\Lambda_2}_{\Lambda_1})^{-1} A\right) \qquad (*)$$

für alle $A \subseteq \Sigma^{\Lambda_1}$, siehe Satz 21.1. Da Σ^{Λ_1} endlich ist, reicht es (*) für einelementige Mengen $A = \{u\}$, $u \in \Sigma^{\Lambda_1}$, zu zeigen.

Sei nun $\tilde{\omega} \in \Sigma^{\Lambda_2^c}$ fest gewählt. Wir betrachten Konfigurationen $\omega = \omega_1 \vee \omega_2 \vee \tilde{\omega}$ wobei ω_1 die Menge Σ^{Λ_1} durchläuft und ω_2 die Menge $\Sigma^{\Lambda_2 \setminus \Lambda_1}$. Dann ist (*) für $A = \{u\}$ gleichbedeutend mit

$$\frac{e^{-\gamma(\tilde{\omega})}}{2^{|\Lambda_2|}} \sum_{\omega_2 \in \Sigma^{\Lambda_2 \setminus \Lambda_1}} e^{-H^{\Lambda_2}(u \vee \omega_2 | \tilde{\omega})} = \frac{e^{-\gamma(\tilde{\omega})}}{2^{|\Lambda_2|}} \sum_{\omega_1 \vee \omega_2 \in \Sigma^{\Lambda_2}} \left(\frac{e^{-\gamma(\omega_2 \vee \tilde{\omega})}}{2^{|\Lambda_1|}} e^{-H^{\Lambda_1}(u | \omega_2 \vee \tilde{\omega})} \right) e^{-H^{\Lambda_2}(\omega_1 \vee \omega_2 | \tilde{\omega})}$$

und es ist zu zeigen, dass für jedes $\omega_2 \in \Sigma^{\Lambda_2 \setminus \Lambda_1}$

$$e^{-\gamma(\omega_2 \vee \tilde{\omega})} e^{-H^{\Lambda_1}(u | \omega_2 \vee \tilde{\omega})} \sum_{\omega_1 \in \Sigma^{\Lambda_1}} e^{-H^{\Lambda_2}(\omega_1 \vee \omega_2 | \tilde{\omega})} = 2^{|\Lambda_1|}\, e^{-H^{\Lambda_2}(u \vee \omega_2 | \tilde{\omega})} \qquad (**)$$

Da $u \in \Sigma^{\Lambda_1}$, ist

$$\begin{aligned} & H^{\Lambda_1}(u|\omega_2 \vee \tilde{\omega}) + H^{\Lambda_2}(\omega_1 \vee \omega_2|\tilde{\omega}) \\ &= H^{\Lambda_1}(u|\omega_2 \vee \tilde{\omega}) + H^{\Lambda_1}(\omega_1|\omega_2 \vee \tilde{\omega}) + \sum_{\{s,t\} \in \mathcal{N}^{\Lambda_2} \setminus \mathcal{N}^{\Lambda_1}} \omega_s \omega_t \\ &= H^{\Lambda_1}(\omega_1|\omega_2 \vee \tilde{\omega}) + H^{\Lambda_2}(u \vee \omega_2|\tilde{\omega}) \ , \end{aligned}$$

und es folgt

$$\begin{aligned} e^{-H^{\Lambda_1}(u|\omega_2 \vee \tilde{\omega})} \sum_{\omega_1 \in \Sigma^{\Lambda_1}} e^{-H^{\Lambda_2}(\omega_1 \vee \omega_2|\tilde{\omega})} &= e^{-H^{\Lambda_2}(u \vee \omega_2|\tilde{\omega})} \sum_{\omega_1 \in \Sigma^{\Lambda_1}} e^{-H^{\Lambda_1}(\omega_1|\omega_2 \vee \tilde{\omega})} \\ &= 2^{|\Lambda_1|} e^{\gamma(\omega_2 \vee \tilde{\omega})} e^{-H^{\Lambda_2}(u \vee \omega_2|\tilde{\omega})} \ , \end{aligned}$$

also (**). □

Beweis von Satz 21.12: Wir weisen die Gültigkeit des Eindeutigkeitskriteriums aus Satz 21.7 nach. Sei $\Lambda = \Lambda_n = \{-n, \ldots, n\}$, $u \in \Sigma^\Lambda$, und seien $\omega, \tilde{\omega} \in \Omega$ so, dass $\omega_\Lambda = u = \tilde{\omega}_\Lambda$. Dann ist

$$\begin{aligned} \exp\left(-\beta\left(H^\Lambda(u|\omega_{\Lambda^c}) - (H^\Lambda(u|\tilde{\omega}_{\Lambda^c})\right)\right) &= \exp \beta \sum_{\{s,t\} \in \mathcal{N}^\Lambda} (\omega_s \omega_t - \tilde{\omega}_s \tilde{\omega}_t) \\ &= \exp \beta \left(\omega_{-(n+1)}\omega_{-n} - \tilde{\omega}_{-(n+1)}\tilde{\omega}_{-n} + \omega_{n+1}\omega_n - \tilde{\omega}_{n+1}\tilde{\omega}_n\right) \\ &\leq e^{4\beta} \ . \end{aligned}$$

Also ist

$$\gamma_\beta(\omega_{\Lambda^c}) = \sum_{v \in \Sigma^\Lambda} \exp\left(-\beta H^\Lambda(v|\omega_{\Lambda^c})\right) \leq e^{4\beta} \sum_{v \in \Sigma^\Lambda} \exp\left(-\beta H^\Lambda(v|\tilde{\omega}_{\Lambda^c})\right) = e^{4\beta} \gamma_\beta(\tilde{\omega}_{\Lambda^c})$$

und das gilt natürlich auch bei vertauschten Rollen von ω und $\tilde{\omega}$. Daher ist

$$K^\Lambda(\omega_{\Lambda^c}, \{u\}) = f_\beta^\Lambda(u|\omega_{\Lambda^c}) \leq e^{8\beta} f_\beta^\Lambda(u|\tilde{\omega}_{\Lambda^c}) = e^{8\beta} K^\Lambda(\tilde{\omega}_{\Lambda^c}, \{u\}) \ ,$$

und es ist $K^\Lambda(\omega_{\Lambda^c}, A) \leq e^{8\beta} \cdot K^\Lambda(\tilde{\omega}_{\Lambda^c}, A)$ für $A \subseteq \Sigma^\Lambda$. □

21.15 Bemerkung Die 8 im Exponenten von e^4 ergab sich im Beweis als "$2 \times 2 \times \mathrm{card}(\partial \Lambda_n)$". In Dimension $d > 1$ ist die entsprechende Größe $2 \times 2 \times 2d(2n+1)^{d-1}$ nicht gleichmäßig in n beschränkt, so dass Satz 21.7 nicht anwendbar ist.

Beweis von Satz 21.13: Der Beweis, dass es in Dimension $d = 2$ bei niedriger Temperatur mindestens zwei verschiedene Gibbs-Maße gibt, beruht auf einem kombinatorischen Argument, das dem zum Beweis der Existenz einer kritischen Perkolationswahrscheinlichkeit (Satz 6.19) ähnelt. Es geht auf R. Peierls zurück und wiard daher oft als *Peierls-Argument* bezeichnet.

Wir stellen uns eine Konfiguration ω als ein unendliches (unregelmäßiges) Schachbrett vor, dessen Felder mit $+1$'en und -1'en bedeckt sind. Zwei benachbarte Felder (d.h. zwei Gitterpunkte $s, t \in \mathbb{Z}^2$ mit $|s - t|_1 = 1$) werden durch eine Kante, den gemeinsamen Rand der beiden Felder, getrennt. Formal bezeichnen wir die Menge aller solcher Kanten mit

$$L := \{\{s,t\} : |s - t|_1 = 1\} \ .$$

Genau wie auf einem richtigen Schachbrett können Kanten zu einem Kantenzug aneinander gehängt werden. Einen einfach geschlossenen Kantenzug wollen wir mit dem Buchstaben C bezeichnen. Der

Jordan'sche Kurvensatz lehrt uns, dass ein solcher Kantenzug ein zusammenhängendes Inneres $I(C)$ hat. Daher können wir zu jedem einfach geschlossenen C eine Abbildung $\tau_C : \Omega \to \Omega$ definieren, $(\tau_C\omega)_s = -\omega_s$ für $s \in I(C)$ und $(\tau_C\omega)_s = \omega_s$ für $s \notin I(C)$. Für ein "Feld" $s \in \mathbb{Z}^2$ sei $\mathcal{C}(s)$ die Menge aller einfach geschlossenen Kantenzüge C mit $s \in I(C)$.
Sei $E_n^+ := \{\omega \in \Omega : \omega_0 = -1;\, \omega_s = +1\,\forall\, s \notin \Lambda_n\}$. Mit $A(\omega)$ bezeichnen wir diejenige "Zusammenhangskomponente" von -1-Feldern, die die Koordinate 0 enthält. $A(\omega)$ wird von einem einfach geschlossenen Kantenzug $C(\omega)$ umrandet. Es ist $\omega_s\omega_t = -1$ für alle $\{s,t\} \in C(\omega)$, d.h. jede Kante in $C(\omega)$ trennt ein $+1$-Feld von einem -1-Feld. Eine solche Kurve nennt man eine *Kontur* für ω. Abbildung 21.2 illustriert diese Bezeichnungen.
Ist umgekehrt ein einfach geschlossener Kantenzug C gegeben, so bezeichne Ω_C die Menge aller $\omega \in \Omega$ mit $\omega_0 = -1$, für die $A(\omega)$ durch C umrandet wird. Für $\omega \in E_n^+$ ist dann $A(\omega) \subset \Lambda_n$, und falls $\{s,t\} \in C$, so ist $\{s,t\} \cap \Lambda_n \neq \emptyset$.

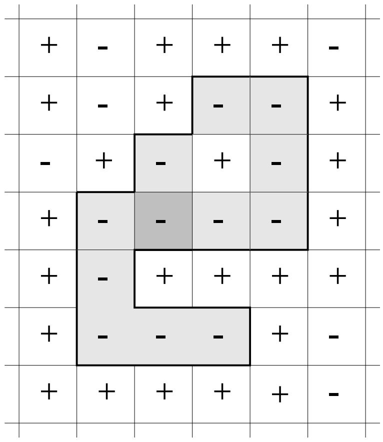


Abbildung 21.2: $A(\omega)$ ist das hellgraue Gebiet, 0 ist das dunklere Feld, $C(\omega)$ ist die fett gezeichnete Kontur.

Sei nun $\omega \in E_n^+$. Wie im Beweis zu Satz 21.12 schätzen wir ab

$$\begin{aligned} H^{\Lambda_n}(\omega_{\Lambda_n}|\omega_{\Lambda_n^c}) - H^{\Lambda_n}((\tau_C\omega)_{\Lambda_n}|(\tau_C\omega)_{\Lambda_n^c}) &= -\frac{1}{2}\sum_{s\in\Lambda_n}\sum_{t\in\mathbb{Z}^2,\{s,t\}\in C}\big(\underbrace{\omega_s\omega_t}_{=-1} - \underbrace{(\tau_C\omega)_s(\tau_C\omega)_t}_{=+1}\big) \\ &\geq 2\cdot\operatorname{card}(C) \end{aligned}$$

Sei nun C ein einfach geschlossener Kantenzug mit $0 \in I(C) \subseteq \Lambda_n$. Dann ist $(\tau_C\omega)_{\Lambda_n^c} = \omega_{\Lambda_n^c} = (+1)^{\Lambda_n^c}$, und da τ_C injektiv ist, folgt

$$\begin{aligned} &\sum_{\omega\in E_n^+\cap\Omega_C} \exp\left(-\gamma_\beta\big(\omega_{\Lambda_n^c}\big) - \beta H^{\Lambda_n}\big(\omega_{\Lambda_n}|\omega_{\Lambda_n^c}\big)\right) \\ &\leq \sum_{\omega\in E_n^+\cap\Omega_C} \exp\Big(-\gamma_\beta\big((+1)^{\Lambda_n^c}\big) - 2\beta\operatorname{card}(C) - \beta H^{\Lambda_n}\big((\tau_C\omega)_{\Lambda_n}|(+1)^{\Lambda_n^c}\big)\Big) \\ &\leq \exp\left(-2\beta\operatorname{card}(C)\right)\cdot\exp\Big(-\gamma_\beta\big((+1)^{\Lambda_n^c}\big)\Big)\cdot\sum_{w\in\Sigma^{\Lambda_n}}\exp\Big(-\beta H^{\Lambda_n}\big(w|(+1)^{\Lambda_n^c}\big)\Big) \\ &= \exp\left(-2\beta\operatorname{card}(C)\right)\ . \end{aligned} \tag{§}$$

Wie im Beweis zu Satz 21.8 betrachten wir nun Maße

$$P_n = (\delta_{(+1)^{\mathbb{Z}^2}}\circ\pi^{-1}_{\Lambda_n^c})\times K_\beta^{\Lambda_n} = \delta_{(+1)^{\Lambda_n^c}}\times K_\beta^{\Lambda_n}\ .$$

Wir schätzen ab, wie sich die rein positiven äußeren Randbedingungen auf die Verteilung des Werts von ω_0 auswirken. Es ist

$$\begin{aligned} P_n\{\omega_0 = -1\} &= \exp\Big(-\gamma_\beta\big((+1)^{\Lambda_n^c}\big)\Big)\cdot\sum_{\omega\in E_n^+}\exp\left(-\beta H^{\Lambda_n}\big(\omega_{\Lambda_n}|\omega_{\Lambda_n^c}\big)\right) \\ &= \sum_{C:\,0\in I(C)}\ \sum_{\omega\in E_n^+\cap\Omega_C}\exp\left(-\gamma_\beta\big(\omega_{\Lambda_n^c}\big) - \beta H^{\Lambda_n}\big(\omega_{\Lambda_n}|\omega_{\Lambda_n^c}\big)\right) \\ &\overset{(§)}{\leq} \sum_{C:\,0\in I(C)}\exp\left(-2\beta\operatorname{card}(C)\right) \\ &\leq \sum_{l=4}^{\infty}\ \sum_{C:\,0\in I(C),\operatorname{card}(C)=l}\exp\left(-2\beta\operatorname{card}(C)\right) \\ &\leq \sum_{l=4}^{\infty} l\cdot 3^{l-1}\,e^{-2\beta l}\ . \end{aligned}$$

Für die letzte Ungleichung mussten wir die Anzahl einfach geschlossener Kantenzüge C der Länge l mit $0 \in I(C)$ abschätzen: Offensichtlich enthält jeder solche Kantenzug mindestens ein horizontales Segment "oberhalb" des Ursprungs mit Abstand höchstens $\frac{l}{2}$ vom Ursprung. Die Anzahl solcher Segmente ist höchstens l. Ausgehend von einem solchen Segment ist der Rest der Kurve durch $(l-1)$-malige Auswahl zwischen höchstens drei Richtungen bestimmt.

Konstruieren wir nun ein Gibbs-Maß μ^+ wie im Beweis zu Satz 21.8 als schwachen Häufungspunkt der Maße P_n, so überträgt sich obige Abschätzung - die ja unabhängig von n ist - auf das Maß μ^+, und wir können schließen

$$\mu^+\{\omega_0 = -1\} \leq \sum_{l=4}^{\infty} l \cdot 3^{l-1}\, e^{-2\beta l} \to 0 \quad \text{mit } \beta \to \infty\ .$$

Also gibt es zu $\epsilon > 0$ ein $\beta(\epsilon) < \infty$ derart, dass $\mu^+\{\omega_0 = -1\} < \epsilon$ für $\beta > \beta(\epsilon)$. Da die H^Λ symmetrisch unter dem Übergang von einer Konfiguration ω zu ihrem "negativen Spiegelbild" $\bar{\omega}$ sind, ist mit μ^+ auch μ^-, definiert durch $\mu^-(A) := \mu^+\{\bar{\omega} : \omega \in A\}$ ein Gibbs-Maß, und es ist $\mu^-\{\omega_0 = +1\} < \epsilon$, also $\mu^-\{\omega_0 = -1\} > 1-\epsilon$. Daraus folgt, dass $\mu^+ \neq \mu^-$. Genau wie im Beweis zu Satz 21.7 folgt daraus unter Benutzung von Lemma 21.6(4) die Existenz zweier zueinander singulärer Gibbs-Maße. $\square$

21.16 Bemerkung Der Beweis des Satzes zeigt, dass bei ausreichend großem β ein geschlossener Kantenzug, auf dem alle "spins" den gleichen Wert annehmen, mit hoher Wahrscheinlichkeit auch in seinem Inneren ganz überwiegend diesen Wert annimmt. Das begünstigt eine "Inselbildung" wie in Abbildung 21.3. Das Bild kann man interpretieren als einen weißen "Ozean" mit schwarzen "Inseln". Die großen schwarzen Inseln spielen auf kleinerer Größenordnung dann die Rolle von schwarzen "Ozeanen" mit kleineren weißen "Inseln", und bei einigen dieser weißen "Inseln" kann man diese Beobachtung noch eine Stufe kleiner fortsetzen. Auf einem unendlichen Gitter findet man diese Struktur in jeder Größenordnung wieder.

21.17 Bemerkung Mit etwas mehr Aufwand (Übungsblatt 5) kann man zeigen, dass die im Beweis konstruierten Maße μ^+ und μ^- bereits singulär zueinander sind. Im Fall $d = 2$ gilt sogar: Jedes $\mu \in \mathcal{G}_1^+$ ist eine Konvexkombination von μ^+ und μ^-.

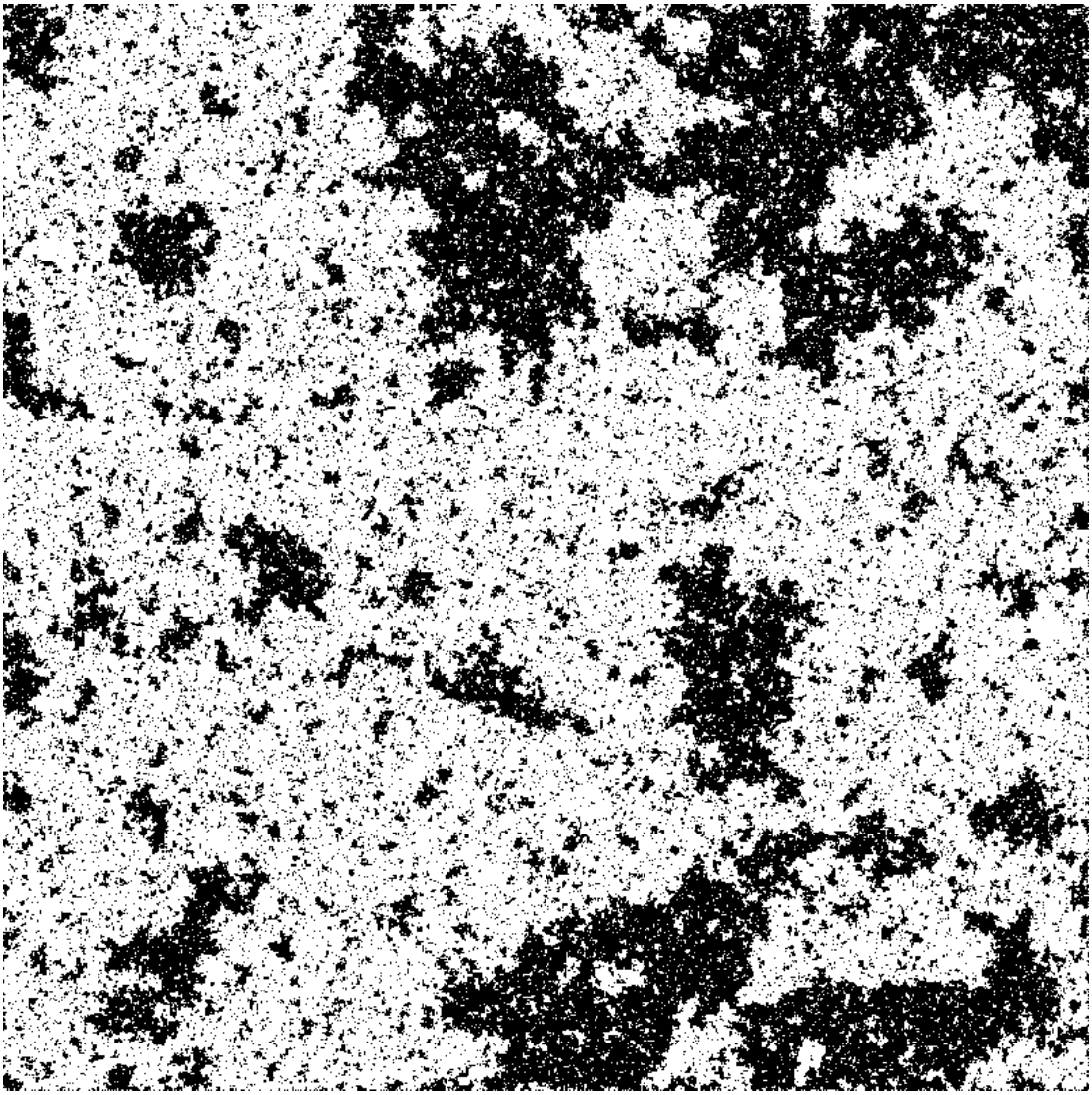


Abbildung 21.3: Typische Ising Konfiguration auf einem 800×800-Gitter. $\beta = 0.4410$ ist etwas größer als der kritische Parameter $\beta_c = \frac{1}{2}\sinh^{-1}(1)$; daher überwiegt eine Art von "spins". Die in Bemerkung 21.16 diskutierte geschachtelte "Ozean-Insel-Struktur" ist deutlich sichtbar.

Kapitel 22

Die Brown'sche Bewegung

Wir beginnen dieses Kapitel mit ein paar einfachen Überlegungen dazu, wie ein Prozess in stetiger Zeit aussehen müsste, um als Grenzfall einer Folge immer schneller ablaufender symmetrischer Bernoulli-Irrfahrten angesehen werden zu können.

Sei $(S_k)_{k\geq 0}$, $S_k := \sum_{j=1}^{k} \xi_j$, die *symmetrische Bernoulli-Irrfahrt* auf $\mathbb{Z}$, d.h. die ξ_j seien unabhängig und $P(\xi_j = -1) = P(\xi_j = 1) = \frac{1}{2}$. Lassen wir die Zeit "schneller laufen" indem wir den Prozess um den Faktor $n \in \mathbb{N}$ "beschleunigen", also das Teilchen zu den Zeitpunkten $0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots$ "springen" lassen, so erhalten wir die Familie $(S_{[t\cdot n]})_{t\geq 0}$, $n \in \mathbb{N}$, von stochastischen Prozessen mit stetiger Zeit. Da $E\xi_j = 0$ und $V(\xi_j) = 1$, gilt für $t \geq 0$

$$\begin{aligned} S_{[0\cdot n]} &= 0 && (22.1)\\ ES_{[t\cdot n]} &= 0 \\ V(S_{[t\cdot n]}) &= [t\cdot n] \approx n\cdot t\,. && (*) \end{aligned}$$

Will man nun die Zeit immer mehr beschleunigen und dabei für $n \to \infty$ zu einem "Grenzprozess" gelangen, so muss man die $S_{[t\cdot n]}$ wegen (*) normieren. Daher betrachtet man die Prozesse

$$X_t^{(n)} := \frac{1}{\sqrt{n}} S_{[t\cdot n]}, \quad t \geq 0.$$

Es folgt sofort, dass

$$\begin{aligned} X_0^{(n)} &= 0 \\ EX_t^{(n)} &= 0 \\ V(X_t^{(n)}) &= \frac{[t\cdot n]}{n} \to t \quad (n \to \infty). \end{aligned}$$

Diese Normierung scheint also vernünftig zu sein.

Wegen der Unabhängigkeit der ξ_j folgt aus dem Zentralen Grenzwertsatz für $0 = t_0 < t_1 < \cdots < t_r$ und $A_1, \ldots, A_r \in \mathcal{B}$

$$\begin{aligned} & P\left(X_{t_i}^{(n)} - X_{t_{i-1}}^{(n)} \in A_i \ (i = 1, \ldots, r)\right) \\ &= P\left(\frac{1}{\sqrt{n}}(S_{[t_i\cdot n]} - S_{[t_{i-1}\cdot n]}) \in A_i \ (i = 1, \ldots, r)\right) \\ &= \prod_{i=1}^{r} P\left(\frac{1}{\sqrt{n}}(\underbrace{S_{[t_i\cdot n]} - S_{[t_{i-1}\cdot n]}}_{\text{Summe von} \approx (t_i - t_{i-1})n \text{ u.i.v. } \xi_j}) \in A_i\right) \\ &\to \prod_{i=1}^{r} \mathcal{N}(0, t_i - t_{i-1})(A_i) \quad (n \to \infty) \end{aligned}$$

Mit anderen Worten: Sei $D := \operatorname{diag}(t_1 - t_0, \dots, t_r - t_{r-1})$. Dann konvergiert

$$P_{(X^{(n)}_{t_1} - X^{(n)}_{t_0}, \dots, X^{(n)}_{t_r} - X^{(n)}_{t_{r-1}})} \Longrightarrow \mathcal{N}(0, D) \quad \text{für } n \to \infty.$$

Durch Transformation mit der Matrix A, $a_{ij} := 1_{i \geq j}$, folgt

$$P_{(X^{(n)}_{t_1}, \dots, X^{(n)}_{t_r})} \Longrightarrow \mathcal{N}(0, ADA^T) \quad \text{für } n \to \infty$$

und

$$(ADA^T)_{ij} = \sum_{k=1}^{r} \sum_{\ell=1}^{r} a_{ik} d_{k\ell} a_{j\ell} = \sum_{k=1}^{i} \sum_{\ell=1}^{j} d_{k\ell} = t_{i \wedge j} = t_i \wedge t_j \ ,$$

wo $u \wedge v := \min\{u, v\}$, siehe Satz 17.3.

Daher suchen wir einen "Grenzprozess" $(B_t)_{t \geq 0}$ mit $B_0 = 0$, dessen Marginalverteilungen auf Indexmengen $I = \{t_0 = 0, t_1, \dots, t_r\}$ die Gestalt $P_I := \delta_0 \times \mathcal{N}(0, (t_i \wedge t_j)_{i,j \in I \setminus \{0\}})$ haben.

Sei $\mathcal{E}_0([0, \infty)) := \{I \subseteq [0, \infty) : 0 \in I, I \text{ endlich}\}$. Ist $I \subseteq J \in \mathcal{E}_0([0, \infty))$, so entsteht $(t_i \wedge t_j)_{i,j \in I \setminus \{0\}}$ aus $(t_i \wedge t_j)_{i,j \in J \setminus \{0\}}$ durch Streichen der Zeilen und Spalten mit Indizes aus $J \setminus I$. Da mehrdimensionale Normalverteilungen eindeutig durch ihre Kovarianzmatrizen bestimmt sind, folgt $P_I = P_J \circ (\pi^J_I)^{-1}$, d.h. $(P_I : I \in \mathcal{E}_0([0, \infty)))$ ist ein verträgliches System von Randverteilungen. So gelangen wir zum folgenden Satz.

22.1 Satz (Brown'sche Bewegung - erster Schritt)
Es gibt ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)})$ mit endlich-dimensionalen Randverteilungen $P \circ \pi_I^{-1} = \delta_0 \times \mathcal{N}(0, (t_i \wedge t_j)_{i,j \in I \setminus \{0\}})$.

Für $0 = t_0 < t_1 < \dots < t_r$ sind die Zufallsvariablen $B_{t_i} - B_{t_{i-1}}$, $i = 1, \dots, r$, unabhängig und nach $\mathcal{N}(0, t_i - t_{i-1})$ verteilt.

Beweis: Die Existenz von P folgt aus dem Fortsetzungssatz 12.3 von Kolmogorov. Für den Rest beachte, dass mit den oben eingeführten Matrizen A und D

$$(B_{t_1} - B_{t_0}, \dots, B_{t_r} - B_{t_{r-1}}) = A^{-1}(B_1, \dots, B_r)$$

nach $\mathcal{N}(0, A^{-1}(ADA^T)(A^{-1})^T) = \mathcal{N}(0, D)$ verteilt ist, siehe Satz 17.3. □

22.2 Bemerkung (Konstruktion von Gauß-Maßen) Auf genau die gleiche Art und Weise kann man zu jeder Familie von verträglichen Kovarianzmatrizen ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)})$ konstruieren, das gerade die durch die Kovarianzmatrizen bestimmten Normalverteilungen als endlich-dimensionale Randverteilungen hat. Solche Maße nennt man *Gauß-Maße*.

Mit der Konstruktion aus Satz 22.1 können wir noch nicht ganz zufrieden sein, denn da die Sprünge der stückweise konstanten Prozesse $S_{[t \cdot n]}$ Größe $\frac{1}{\sqrt{n}}$ haben, sollte ein vernünftiger Limes-Prozess stetige Pfade haben, d.h. die Abbildungen $t \mapsto B_t(\omega)$ sollten für P-f.a. ω stetig sein. Das kann man aber nicht erwarten, wie die folgende Diskussion zeigen soll. Wir benötigen folgendes Lemma.[1]

22.3 Lemma *Sei $A \in \mathcal{B}^T$, T eine beliebige Indexmenge. Dann gibt es eine höchstens abzählbare Teilmenge $S \subseteq T$ derart, dass $A \in (\pi^T_S)^{-1}(\mathcal{B}^S)$.*

Beweis: Sei

$$\mathcal{A} := \bigcup_{\substack{S \subseteq T \\ S \text{ abzählbar}}} (\pi^T_S)^{-1}(\mathcal{B}^S) \ .$$

$\mathcal{A}$ enthält alle Zylindermengen, und man zeigt leicht, dass $\mathcal{A}$ eine σ-Algebra ist. Also ist $\mathcal{B}^T \subseteq \mathcal{A}$. □

22.4 Korollar (Nichtmessbarkeit von $C(T; \mathbb{R})$ in $\mathcal{B}^T$) *Sei $T \subseteq \mathbb{R}$ ein nicht triviales Intervall.*

[1] Die Darstellung hier und in den weiteren Kapiteln folgt im Wesentlichen dem Buch von Revuz und Yor [Revuz-Yor].

a) $\{\omega \in \mathbb{R}^T : t \mapsto \omega_t$ *stetig*$\} \notin \mathcal{B}^T$

b) $\{\omega \in \mathbb{R}^T : t \mapsto \omega_t$ *monoton wachsend*$\} \notin \mathcal{B}^T$.

Beweis: Andernfalls würden die Eigenschaften "$t \mapsto \omega_t$ ist stetig" bzw. "$t \mapsto \omega_t$ ist monoton wachsend" durch die Werte ω_t für t aus einer abzählbaren Teilmenge von T determiniert, was aber nicht der Fall ist. □

22.5 Bemerkung Ähnliche Probleme treten auf, wenn man aus dem Prozess $(B_t)_{t\geq 0}$ hergeleitete zufällige Größen wie

$$T(\omega) := \inf\{t > 0 : B_t(\omega) > 1\} \quad \text{oder} \quad M_t(\omega) := \sup\{|B_s(\omega)| : s \leq t\}$$

betrachten will. Die bisherige Konstruktion, die $T \subseteq \mathbb{R}$ wie eine völlig strukturlose überabzählbare Indexmenge behandelt, erlaubt es nicht, die Messbarkeit dieser Größen nachzuweisen. Wir müssen die Konstruktion daher modifizieren.

22.6 Definition (Stetige Pfade) *Sei $T \subseteq \mathbb{R}$. Ein stochastischer Prozess $(X_t)_{t\in T}$ hat stetige Pfade, falls die Abbildung $t \mapsto X_t(\omega)$ für jedes ω auf ganz T stetig ist.*

22.7 Definition (Modifikation eines Prozesses) *Zwei auf dem gleichen Wahrscheinlichkeitsraum definierte stochastische Prozesse $(X_t)_{t\in T}$ und $(X'_t)_{t\in T}$ sind Modifikationen voneinander, falls*

$$\forall\, t \in T: \; X_t = X'_t \text{ f.s.}$$

22.8 Bemerkung (Modifikation ⇒ gleiche Verteilung) Sind $(X_t)_{t\in T}$ und $(X'_t)_{t\in T}$ Modifikationen voneinander, so ist auch $(X_{t_1}, \ldots, X_{t_r}) = (X'_{t_1}, \ldots, X'_{t_r})$ f.s. für beliebige $t_1, \ldots, t_r \in T$. Insbesondere haben die beiden Prozesse die gleichen endlich-dimensionalen Randverteilungen und damit die gleiche Verteilung (als Prozess).

Hinreichende Bedingungen, unter denen ein Prozess eine Modifikation mit stetigen, ja sogar Hölder-stetigen[2] Pfaden besitzt, gibt der folgende Satz an.

22.9 Satz (Kriterium von Kolmogorov)
Sei $T \subseteq \mathbb{R}$ ein Intervall. Für den reellwertigen stochastischen Prozess $(X_t)_{t\in T}$ gebe es Konstanten $\alpha, \beta, C > 0$ derart, dass für alle $s, t \in T$

$$E\left[|X_t - X_s|^\alpha\right] \leq C\,|t-s|^{1+\beta}\,.$$

Dann hat $(X_t)_{t\in T}$ eine Modifikation $(X'_t)_{t\in T}$ mit stetigen Pfaden.

Es gilt sogar mehr: Für jedes $\gamma \in (0, \frac{\beta}{\alpha})$ und jedes $N \in \mathbb{N}$ ist

$$E\left[\left(\sup_{\substack{s,t\in T\cap[0,N]\\ s\neq t}} \frac{|X'_t - X'_s|}{|t-s|^\gamma}\right)^\alpha\right] < \infty\,.$$

Insbesondere ist fast jeder Pfad von $(X'_t)_{t\in T}$ γ-Hölder stetig auf $T \cap [0, N]$, und durch Modifikation des Prozesses auf einer Nullmenge erhält man einen Prozess, bei dem sämtliche Pfade lokal γ-Hölder stetig sind.

Wir stellen den Beweis zurück und wenden den Satz zunächst auf den in Satz 22.1 konstruierten Prozess $(B_t)_{t\geq 0}$ an.

---

[2] Sei $\gamma > 0$, $T \subseteq \mathbb{R}$ ein Intervall. Eine Funktion $f : T \to \mathbb{R}$ heißt γ-Hölder-stetig, falls $H_\gamma(f) := \sup\{\frac{|f(x)-f(y)|}{|x-y|^\gamma} : x, y \in T, x \neq y\} < \infty$. f heißt lokal γ-Hölder-stetig, falls jedes $x \in T$ eine Umgebung besitzt, auf der f γ-Hölder-stetig ist.

22.10 Satz (Existenz der Brown'schen Bewegung)
Der in Satz 22.1 konstruierte Prozess $(B_t)_{t\geq 0}$ besitzt eine Modifikation mit stetigen Pfaden. Sie ist f.s. eindeutig bestimmt. Sie kann so gewählt werden, dass ihre Pfade für jedes $\gamma \in (0, \frac{1}{2})$ lokal γ-Hölder-stetig *sind. (Diese Modifikation wird wieder mit $(B_t)_{t\geq 0}$ bezeichnet.)*

22.11 Definition (Brown'sche Bewegung) *Ein Prozess $(B_t)_{t\geq 0}$ heißt Brown'sche Bewegung, wenn er wie die in Satz 22.10 konstruierte Modifikation folgende Eigenschaften hat:*

i) $B_0 = 0$,

ii) $P_{(B_{t_1},\ldots,B_{t_n})} = \mathcal{N}(0, (t_i \wedge t_j)_{i,j=1,\ldots,n})$ *für alle* $t_1, \ldots, t_n > 0$,

iii) $(B_t)_{t\geq 0}$ *hat stetige Pfade.*

22.12 Bemerkung (Gauß-Prozess) Fordert man in dieser Definition in ii) andere Kovarianzmatrizen, so gelangt man zu allgemeineren *Gauß-Prozessen*, siehe die Beispiele 22.23 und 22.24.

Beweis von Satz 22.10: Wir wenden das Kolmogorov-Kriterium mit $\alpha = 4$ und $\beta = 1$ an: Für $s < t$ ist $B_t - B_s$ nach $\mathcal{N}(0, t-s)$ verteilt. Daher ist $E[|B_t - B_s|^4] = (t-s)^2\, E[B_1^4]$. (Man kann nachrechnen, dass $E[B_1^4] = 3$.) Also gibt es eine Modifikation mit stetigen Pfaden.

Sind $(B'_t)_{t\geq 0}$ und $(B''_t)_{t\geq 0}$ zwei Modifikationen von $(B_t)_{t\geq 0}$ mit stetigen Pfaden, so gilt fast sicher: $B'_t = B''_t$ für alle $t \in \mathbb{Q} \cap [0, \infty)$, und aus der Stetigkeit der Pfade folgt, dass fast sicher gilt: $B'_t = B''_t$ für alle $t \geq 0$.

Wendet man das Kolmogorov-Kriterium mit $\alpha = 2k$ und $\beta = k - 1$, $k \in \mathbb{N}$, an, so erhält man $E[|B_t - B_s|^{2k}] = (t-s)^k\, E[B_1^{2k}] = E[B_1^{2k}]\, |t-s|^{1+\beta}$. Die Forderung $\gamma < \frac{\beta}{\alpha}$ in Satz 22.9 wird dann zu $\gamma < \frac{1}{2} - \frac{1}{2k}$, und da man k beliebig groß wählen kann, folgt, dass die Brown'schen Pfade für jedes $\gamma \in [0, \frac{1}{2})$ fast sicher γ-Hölder-stetig sind. □

22.13 Bemerkung Aus der Stetigkeit der Pfade folgt leicht, dass die beiden Zufallsvariablen T und M_t aus Bemerkung 22.5 messbar sind:

$$T(\omega) = \inf\{t \in \mathbb{Q} \cap [0, \infty) : B_t(\omega) > 1\}\,, \quad M_t(\omega) = \sup\{|B_s(\omega)| : s \in \mathbb{Q} \cap [0, t]\}$$

22.14 Korollar *Sei $(B_t)_{t\geq 0}$ eine Brown'sche Bewegung. Dann ist jeder der folgenden Prozesse $(X_t)_{t\geq 0}$ ebenfalls eine Brown'sche Bewegung:*

a) *Sei* $s > 0$, $X_t := B_{s+t} - B_s$, $t \geq 0$.

b) $X_t := -B_t$, $t \geq 0$.

c) *Sei* $c > 0$, $X_t := cB_{t/c^2}$, $t \geq 0$.

d) $X_0 := 0$, $X_t := tB_{1/t}$ $(t > 0)$. *(In diesem Fall muss der Prozess eventuell noch modifiziert werden.)*

Beweis: In allen vier Situationen ist $X_0 = 0$ und $P_{(X_{t_1},\ldots,X_{t_n})}$ für $t_1, \ldots, t_n > 0$ eine n-dimensionale Normalverteilung (Satz 17.3). Außerdem hat $(X_t)_{t\geq 0}$ in den Fällen a)-c) stetige Pfade.

a) $Cov(X_t, X_{t'}) = (t+s) \wedge (t'+s) - s - s + s = t \wedge t'$.

b) $Cov(X_t, X_{t'}) = Cov(B_t, B_{t'}) = t \wedge t'$.

c) $Cov(X_t, X_{t'}) = c^2 \cdot (\frac{t}{c^2} \wedge \frac{t'}{c^2}) = t \wedge t'$.

d) $Cov(X_t, X_{t'}) = tt' \cdot (\frac{1}{t} \wedge \frac{1}{t'}) = t \wedge t'$. Also haben die Prozesse $(X_t)_{t>0}$ und $(B_t)_{t>0}$ identische Verteilungen. Außerdem hat der Prozess $(X_t)_{t>0}$ stetige Pfade, so dass nur zu zeigen bleibt $P(\lim_{t\to 0} X_t = 0) = 1$. Da $(X_t)_{t>0}$ und $(B_t)_{t>0}$ identische Verteilungen haben, folgt das aus $P(\lim_{t\to 0} B_t = B_0 = 0) = 1$, sobald man gezeigt hat, dass das Ereignis $\{\lim_{t\to 0} X_t = 0\}$ messbar ist. Wegen der Stetigkeit der Pfade für $t > 0$ ist aber $\{\lim_{t\to 0} X_t = 0\} = \{\lim_{t\to 0, t\in\mathbb{Q}} X_t = 0\}$ messbar.

□

22.15 Bemerkung Aus Korollar 22.14d folgt sofort, dass $\lim_{t\to\infty} t^{-1}B_t = \lim_{t\to 0} X_t = 0$ fast sicher.

Beweis von Satz 22.9: Wir betrachten zunächst nur das Parameterintervall $[-2^N, 2^N]$ für ein festes $N \in \mathbb{N}$. Sei

$$D_n := \left\{ \frac{k}{2^n} \in T \cap [-2^N, 2^N] : k \in \mathbb{Z} \right\} \; (n \geq 0), \quad D := \bigcup_{n\geq 0} D_n \;,$$
$$\Delta_n := \{(s,t) : s,t \in D_n, |s-t| = 2^{-n}\} \;.$$

Beachte, dass $\operatorname{card} \Delta_n \leq 2^{n+N+2}$. Wir betrachten die Zufallsvariablen $K_n := \sup_{(s,t)\in\Delta_n} |X_t - X_s|$. Es ist

$$E\left[K_n^\alpha\right] \leq \sum_{(s,t)\in\Delta_n} E\left[|X_t - X_s|^\alpha\right] \leq 2^{n+N+2} \cdot C2^{-n(1+\beta)} = C2^{N+2}2^{-\beta n} \;. \tag{*}$$

Zu $s, t \in D$ und $m \geq 0$ definiere

$$s_m := \max\{u \in D_m : u \leq s\} \;, \quad t_m := \max\{u \in D_m : u \leq t\} \;.$$

Es gilt $s_m \nearrow s$ und $s_m = s$ falls $s \in D_n$ mit $m \geq n$. Analog für t_m. Daher ist

$$\begin{aligned} X_t - X_s &= (X_{t_m} - X_{s_m}) + (X_t - X_{t_m}) + (X_s - X_{s_m}) \\ &= (X_{t_m} - X_{s_m}) + \sum_{i=m}^{\infty} (X_{t_{i+1}} - X_{t_i}) - \sum_{i=m}^{\infty} (X_{s_{i+1}} - X_{s_i}) \;, \end{aligned}$$

Ist $|s-t| \leq 2^{-m}$, so ist entweder $s_m = t_m$ oder $(s_m, t_m) \in \Delta_m$. In jedem Fall ist also

$$|X_t - X_s| \leq K_m + 2\sum_{i=m}^{\infty} K_{i+1} \leq 2\sum_{i=m}^{\infty} K_i \;.$$

Setzt man nun

$$M_\gamma := \sup\left\{ \frac{|X_t - X_s|}{|t-s|^\gamma} : s,t \in D, s \neq t \right\} \;,$$

so folgt

$$\begin{aligned} M_\gamma &\leq \sup_{m\in\mathbb{N}} \left\{ 2^{(m+1)\gamma} \cdot \sup\left\{ |X_t - X_s| : s,t \in D, 2^{-(m+1)} < |t-s| \leq 2^{-m} \right\} \right\} \\ &\leq \sup_{m\in\mathbb{N}} \left\{ 2 \cdot 2^{(m+1)\gamma} \cdot \sum_{i=m}^{\infty} K_i \right\} \\ &\leq 2^{1+\gamma} \cdot \sum_{i=0}^{\infty} 2^{i\gamma} K_i \;. \end{aligned}$$

Für $\alpha \geq 1$ folgt daraus für $\gamma \in (0, \frac{\beta}{\alpha})$ unter Beachtung von (*)

$$\|M_\gamma\|_\alpha \leq 2^{1+\gamma} \cdot \sum_{i=0}^{\infty} 2^{i\gamma} \|K_i\|_\alpha \leq 2^{1+\gamma} \left(C2^{N+2}\right)^{\frac{1}{\alpha}} \cdot \sum_{i=0}^{\infty} 2^{i(\gamma - \frac{\beta}{\alpha})} < \infty \;,$$

also $E[M_\gamma^\alpha] < \infty$. Ist $\alpha < 1$, kann man (wegen der Konkavität von $x \mapsto x^\alpha$) auf die gleiche Weise $E[M_\gamma^\alpha]$ an Stelle von $\|M_\gamma\|_\alpha$ direkt abschätzen.

In jedem Fall ist $M_\gamma < \infty$ f.s., und es folgt, dass $D \ni t \mapsto X_t(\omega)$ für fast jedes ω gleichmäßig stetig ist. Daher kann man für solche ω diese Abbildung stetig fortsetzen zu

$$X'_t := \lim_{\substack{s\to t \\ s\in D}} X_s \;.$$

Schließlich ist für jedes $t \in T \cap [0, N]$

$$E\left[|X_t - X'_t|^\alpha\right] \le E\left[\lim_{\substack{s\to t\\ s\in D}} |X_t - X_s|^\alpha\right] \le \liminf_{\substack{s\to t\\ s\in D}} E\left[|X_t - X_s|^\alpha\right] \le \liminf_{\substack{s\to t\\ s\in D}} C\,|t-s|^{1+\beta} = 0\,,$$

so dass $X_t = X'_t$ fast sicher.

Die Hölderstetigkeit der Pfade von $(X'_t)_{t\in T}$ folgt nun leicht: Da wir schon wissen, dass diese Pfade stetig sind, ist

$$\sup_{\substack{s,t\in T\cap[-2^N,2^N]\\ s\neq t}} \frac{|X'_t - X'_s|}{|t-s|^\gamma} = \sup_{\substack{s,t\in D\\ s\neq t}} \frac{|X'_t - X'_s|}{|t-s|^\gamma} = M_\gamma < \infty \text{ f.s.},$$

wobei wir für die zweite Gleichung benutzt haben, dass $X_t = X'_t$ f.s. für $t \in D$.

Noch kann es eine Nullmenge $U \subseteq \Omega$ geben, für die $t \mapsto X'_t(\omega)$ nicht Hölder-stetig ist. Wir modifizieren daher X'_t auf U, indem wir für ein geeignetes $t_0 \in T$ setzen: $X'_t(\omega) := X'_{t_0}(\omega)$ $(\omega \in U)$. □

Nun stellt sich die Frage, wie es mit der γ-Hölder Stetigkeit Brown'scher Pfade für $\gamma \ge \frac{1}{2}$ aussieht. Wir werden sehen, dass sie für $\gamma > \frac{1}{2}$ fast sicher nirgends γ-Hölder stetig sind. Folgendes Konzept erweist sich dazu (und auch für spätere Überlegungen) als nützlich.

22.16 Definition (Unterteilung) *Eine Unterteilung Δ des Intervalls $[0,t]$ ist eine endliche, aufsteigende Folge $0 = t_0 < t_1 < \cdots < t_n = t$. Man setzt*

$$|\Delta| := \sup\{|t_i - t_{i-1}| : i = 1,\ldots,n\}$$

und, für einen stochastischen Prozess $(X_t)_{t\ge0}$,

$$Q_t^\Delta := \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2\,.$$

22.17 Definition (Endliche quadratische Variation) *Ein stochastischer Prozess $\mathcal{X} = (X_t)_{t\ge0}$ hat endliche quadratische Variation, falls es einen stochastischen Prozess $\langle\mathcal{X},\mathcal{X}\rangle = (\langle X,X\rangle_t)_{t\ge0}$ mit Werten in $[0,\infty)$ gibt, so dass für jedes $t \ge 0$ und für jede Folge $(\Delta_n)_{n>0}$ von Unterteilungen mit $\lim_{n\to\infty}|\Delta_n| = 0$ gilt:*

$$Q_t^{\Delta_n} \to \langle X,X\rangle_t \quad \text{stochastisch.} \tag{*}$$

Der Prozess $\langle\mathcal{X},\mathcal{X}\rangle$ heißt die quadratische Variation von $\mathcal{X}$.

22.18 Satz (Die quadratische Variation der Brown'schen Bewegung)
Für die Brown'sche Bewegung $(B_t)_{t\ge0}$ ist $\langle B,B\rangle_t = t$ für jedes $t \ge 0$. Die Konvergenz in (*) gilt sogar im L^2-Sinn.

Beweis: Da die $(B_{t_i} - B_{t_{i-1}})^2$ unabhängig sind und Erwartungswert $t_i - t_{i-1}$ haben (siehe Satz 22.1), ist

$$\begin{aligned}
E\left[\left(Q_t^\Delta - t\right)^2\right] &= E\left[\left(\sum_{i=1}^n \left((B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\right)\right)^2\right] \\
&= \sum_{i=1}^n E\left[\left((B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\right)^2\right] \\
&= \sum_{i=1}^n (t_i - t_{i-1})^2 E[(B_1^2 - 1)^2] \\
&\le |\Delta| \cdot E[(B_1^2-1)^2] \to 0 \quad (|\Delta| \to 0)
\end{aligned}$$

□

22.19 Bemerkung Ist $\sum_{n>0} |\Delta_n| < \infty$, so folgt im Fall der Brown'schen Bewegung aus dem Borel-Cantelli Lemma sogar fast sichere Konvergenz in (*).

22.20 Korollar (Unendliche Variation Brown'scher Pfade) *Sei* $\epsilon > 0$. *Für fast jeden Pfad* $t \mapsto B_t(\omega)$ *gilt:*

$$\forall\, 0 \le s < t: \quad \sup \left\{ \sum_{i=1}^{n} |B_{t_i} - B_{t_{i-1}}|^{2-\epsilon} : s = t_0 < t_1 < \cdots < t_n = t \right\} = \infty\,.$$

Beweis: Zunächst reicht es, rationale s und t zu betrachten. Damit kann man sich auf ein festes Paar $s < t$ beschränken. Wegen Korollar 22.14a reicht es dann, den Fall $s = 0$ zu betrachten. Sei daher $(\Delta_k)_{k>0}$ eine Folge von Unterteilungen von $[0,t]$ mit $\sum_{k>0} |\Delta_k| < \infty$. Dann ist

$$\sum_{i=1}^{n_k} |B_{t_i^{(k)}}(\omega) - B_{t_{i-1}^{(k)}}(\omega)|^{2-\epsilon} \ge \frac{Q_t^{\Delta_k}}{\sup_i |B_{t_i^{(k)}}(\omega) - B_{t_{i-1}^{(k)}}(\omega)|^{\epsilon}} \approx \frac{t}{\sup_i |B_{t_i^{(k)}}(\omega) - B_{t_{i-1}^{(k)}}(\omega)|^{\epsilon}} \quad (k \to \infty)$$

für fast jedes ω, und da $u \mapsto B_u(\omega)$ auf dem kompakten Intervall $[0,t]$ für jedes ω gleichmäßig stetig ist und $|\Delta_k| \to 0$ geht, konvergiert der Nenner dieses Ausdrucks für $k \to \infty$ gegen 0, so dass $\sum_{i=1}^{n_k} |B_{t_i^{(k)}}(\omega) - B_{t_{i-1}^{(k)}}(\omega)|^{2-\epsilon} \to \infty$ fast sicher. □

Daraus folgt direkt (mit $\epsilon = 2 - \frac{1}{\gamma}$):

22.21 Korollar *Ist* $\gamma > \frac{1}{2}$, *so ist* $t \mapsto B_t(\omega)$ *fast sicher auf keinem nicht trivialen Intervall* γ*-Hölder stetig. Insbesondere ist* $t \mapsto B_t(\omega)$ *fast sicher auf keinem Intervall differenzierbar. (Man kann sogar zeigen, dass das auch noch für* $\gamma = \frac{1}{2}$ *gilt; der Beweis ist aber aufwendiger.)*

22.22 Bemerkung (Wiener Maß auf $C([0,\infty);\mathbb{R})$) Sei $\mathcal{X} = (X_t)_{t\ge 0}$ ein stochastischer Prozess auf $(\Omega, \mathcal{F}, P)$. Wie in Definition 6.9 Betrachte $\mathcal{X} : \Omega \to \mathbb{R}^{[0,\infty)}$, $(\mathcal{X}(\omega))_t := X_t(\omega)$. Die Verteilung von $\mathcal{X}$ ist $P \circ \mathcal{X}^{-1}$, ein Wahrscheinlichkeitsmaß auf $\mathcal{B}^{[0,\infty)}$. Hat der Prozess $\mathcal{X}$ stetige Pfade, so ist $\mathcal{X}(\Omega) \subseteq C([0,\infty);\mathbb{R})$, aber $C([0,\infty);\mathbb{R}) \notin \mathcal{B}^{[0,\infty)}$, siehe Korollar 22.4. Trotzdem kann man auf $C([0,\infty);\mathbb{R})$ die σ-Algebra

$$\mathcal{B}_C^{[0,\infty)} := \mathcal{B}^{[0,\infty)} \cap C([0,\infty);\mathbb{R}) = \left\{ A \cap C([0,\infty);\mathbb{R}) : A \in \mathcal{B}^{[0,\infty)} \right\}$$

betrachten. Dann ist $\mathcal{X} : \Omega \to C([0,\infty);\mathbb{R})$ natürlich $\mathcal{F} - \mathcal{B}_C^{[0,\infty)}$-messbar und wir können $P \circ \mathcal{X}^{-1}$ als Maß auf $(C([0,\infty);\mathbb{R}), \mathcal{B}_C^{[0,\infty)})$ interpretieren. Dadurch haben wir eine *kanonische Version* von $\mathcal{X}$ auf dem Wahrscheinlichkeitsraum $(C([0,\infty);\mathbb{R}), \mathcal{B}_C^{[0,\infty)}, P \circ \mathcal{X}^{-1})$ gewonnen.

Ist $\mathcal{X}$ eine Brown'sche Bewegung, so heißt dieser kanonische Prozess Wiener-Prozess[3] und das Maß $P \circ \mathcal{X}^{-1}$ Wiener-Maß.

Die beiden folgenden wichtigen Gauß-Prozese lassen sich auf elementare Weise aus der Brown'schen Bewegung gewinnen.

22.23 Beispiel (Ornstein-Uhlenbeck Prozess) Seien $\alpha, \beta > 0$, $(B_t)_{t\ge 0}$ eine Brown'sche Bewegung. Für $t \in \mathbb{R}$ setze

$$U_t := e^{-\alpha t}\, B_{\beta e^{2\alpha t}}\,.$$

$(U_t)_{t\in\mathbb{R}}$ ist ein stationärer Gauß-Prozess mit stetigen Pfaden. Die Stetigkeit der Pfade folgt direkt aus der Stetigkeit Brown'scher Pfade. Wir zeigen, dass $(U_t)_{t\in\mathbb{R}}$ ein stationärer Gauß-Prozess ist.

Seien $t_1 < \cdots < t_n, s \in \mathbb{R}$. Mit den Bezeichnungen

$$\tilde\beta := \beta e^{2\alpha s}, \quad A := \operatorname{diag}\left(e^{-\alpha t_1}, \ldots, e^{-\alpha t_n}\right)$$

[3] nach Norbert Wiener (1894-1964)

Der zufällige Vektor

$$(U_{t_1+s}, \ldots, U_{t_n+s})^T = e^{-\alpha s} A \left(B_{\tilde{\beta}e^{2\alpha t_1}}, \ldots, B_{\tilde{\beta}e^{2\alpha t_n}}\right)^T$$

hat die Verteilung

$$\begin{aligned}
&\mathcal{N}\left(0, e^{-\alpha s} A V e^{-\alpha s} A^T\right) \qquad \left(\text{mit } V_{ij} = (\tilde{\beta}e^{2\alpha t_i} \wedge \tilde{\beta}e^{2\alpha t_j})_{ij} = (\tilde{\beta}e^{2\alpha(t_i \wedge t_j)})_{ij}\right) \\
=&\mathcal{N}\left(0, \left(e^{-2\alpha s}\tilde{\beta}e^{-\alpha t_i} e^{2\alpha(t_i \wedge t_j)} e^{-\alpha t_j}\right)_{ij}\right) \\
=&\mathcal{N}\left(0, \left(\beta e^{-\alpha|t_i - t_j|}\right)_{ij}\right) \ ,
\end{aligned}$$

also eine von der Verschiebung s nicht abhängende Normalverteilung. Insbesondere ist

$$Cov(U_s, U_t) = \beta\, e^{-\alpha|s-t|} \ .$$

22.24 Beispiel (Brown'sche Brücke) Für $y \in \mathbb{R}$ definiere den Prozess $(B_t^y)_{0 \le t \le 1}$ durch

$$B_t^y := B_t - t(B_1 - y), \quad \text{also } B_t^y = B_t^0 + ty \ .$$

Dieser Prozess hat ebenfalls stetige Pfade, und es ist $B_0^y = 0$, $B_1^y = y$, $E[B_t^y] = ty$. Außerdem ist $(B_t^0)_{0 \le t \le 1}$ ein zentrierter Gauß-Prozess, denn für $0 < t_1 < \cdots < t_n < 1$ ist

$$\begin{pmatrix} B_{t_1}^0 \\ \vdots \\ B_{t_n}^0 \end{pmatrix} = \begin{pmatrix} 1 & & & -t_1 \\ & \ddots & & \vdots \\ & & 1 & -t_n \end{pmatrix} \begin{pmatrix} B_{t_1} \\ \vdots \\ B_{t_n} \\ B_1 \end{pmatrix}$$

Die Kovarianzstruktur berechnet sich zu

$$\begin{aligned}
Cov(B_s^0, B_t^0) &= Cov(B_s, B_t) - s\, Cov(B_t, B_1) - t\, Cov(B_1, B_s) + st\, Cov(B_1, B_1) \\
&= s \wedge t - st - ts + st = s \wedge t - st
\end{aligned}$$

Es folgt sofort, dass auch für beliebige y

$$Cov(B_s^y, B_t^y) = s \wedge t - st \ .$$

Es ist möglich, die Verteilung von $(B_t^y)_{0 \le t \le 1}$ als "bedingte Verteilung von $(B_t)_{0 \le t \le 1}$ gegeben $B_1 = y$" zu interpretieren. Das wird in den Übungen genauer beschrieben.

Kapitel 23

Der Poisson-Prozess

In diesem kurzen Kapitel konstruieren wir den Prototyp eines Prozesses mit stückweise konstanten (also nicht stetigen) Pfaden, den *Poisson-Prozess*. Dabei spielen zwei klassische Verteilungen eine herausragende Rolle.

Zunächst einmal benötigt man die Exponentialverteilungen zum Parameter $\alpha > 0$, d.h. die Verteilungen auf $[0, \infty)$ mit Dichte $f_\alpha(x) = \alpha e^{-\alpha x}$. Insbesondere ist $P(X \geq z) = \int_z^\infty \alpha e^{-\alpha x}\, dx = e^{-\alpha z}$ und $E[X] = \int_0^\infty x\alpha e^{-\alpha x}\, dx = \frac{1}{\alpha}$ für nach $\mathcal{E}_\alpha$ verteilte X. Exponentialverteilungen haben die folgende charakteristische Eigenschaft: Ist X nach $\mathcal{E}_\alpha$ verteilt und sind $a, b \geq 0$, so ist

$$P\{X \geq a + b \big| X \geq a\} = \frac{P\{X \geq a + b\}}{P\{X \geq a\}} = \frac{e^{-\alpha(a+b)}}{e^{-\alpha a}} = e^{-\alpha b} = P\{X \geq b\}\ .$$

Diese Eigenschaft nennt man auch die *Gedächtnislosigkeit* der Exponentialverteilungen.

Des Weiteren benötigt man die *Poisson-Verteilungen* $\mathcal{P}_\alpha$ zum Parameter α, d.h. Wahrscheinlichkeitsmaße auf $\mathbb{N}_0$ mit $\mathcal{P}_\alpha(k) = e^{-\alpha}\frac{\alpha^k}{k!}$. Für nach $\mathcal{P}_\alpha$ verteilte X ist $E[X] = e^{-\alpha}\sum_{k=0}^\infty k\frac{\alpha^k}{k!} = \alpha$.

Seien nun $X_1, X_2, \ldots$ unabhängige nach $\mathcal{E}_\alpha$ verteilte Zufallsvariablen. Setze $S_n := X_1 + \cdots + X_n$. Da $\frac{1}{n}S_n \to E[X_1] = \frac{1}{\alpha}$ f.s. nach dem Gesetz der großen Zahl, geht $S_n \to \infty$ f.s., und da $X_k > 0$ f.s. für jedes k, folgt

$$0 =: S_0(\omega) < S_1(\omega) < S_2(\omega) < \cdots \nearrow \infty$$

fast sicher, und durch Modifikation auf einer Nullmenge kann man annehmen, dass das für alle $\omega \in \Omega$ gilt. Wir interpretieren die $S_n(\omega)$ als eine Folge zufälliger Zeitpunkte, zu denen gewisse Ereignisse realisiert werden, (z.B. atomare Zerfallsereignisse, Ankunft eines neuen Kunden an einem Bedienungsschalter, Anfrage an einen Server, u.s.w.). Die Zeitintervalle zwischen zwei Ereignissen sind dabei so "zufällig" wie möglich (bei gegebenem Erwartungswert $\frac{1}{\alpha}$), siehe Übung 13.10.

Setze nun für jedes $t \geq 0$

$$N_t(\omega) := \max\{n \geq 0 : S_n(\omega) \leq t\}\ , \quad \text{d.h.} S_{N_t} \leq t < S_{N_t+1}\ .$$

$N_t(\omega)$ ist also die Anzahl der zufälligen Ereignisse bis zum Zeitzpunkt t. Dann ist

$$\{N_t \geq k\} = \{S_k \leq t\} \quad \text{für alle } k \in \mathbb{N}_0\ . \tag{*}$$

23.1 Lemma (Verteilung von N_t) *Die Zufallsvariablen N_t sind nach $\mathcal{P}_{\alpha t}$ verteilt.*

Beweis: Sei $f_n(x)$ die Dichte von $S_n = S_{n-1} + X_n$. Da $P_{X_n} = P_{X_1} = P_{S_1}$, ist

$$f_1(x) = \begin{cases} \alpha e^{-\alpha x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

$$f_n(x) = (f_{n-1} * f_1)(x) = \int_0^x f_{n-1}(y)\, f_1(x-y)\, dy \quad \text{für } n \geq 2 \quad \text{(siehe Satz 11.19).}$$

Durch Induktion zeigt man: $f_n(x) = \alpha e^{-\alpha x} \cdot \frac{(\alpha x)^{n-1}}{(n-1)!}$.[1] Es folgt:

$$\begin{aligned}
\frac{d}{dt}P(N_t = k) &= \frac{d}{dt}\big(P(N_t \geq k) - P(N_t \geq k+1)\big) \\
&= \frac{d}{dt}\big(P(S_k \leq t) - P(S_{k+1} \leq t)\big) = f_k(t) - f_{k+1}(t) \\
&= \alpha e^{-\alpha t} \cdot \left(\frac{(\alpha t)^{k-1}}{(k-1)!} - \frac{(\alpha t)^k}{k!}\right) ,
\end{aligned}$$

so dass

$$P(N_t = k) = e^{-\alpha t}\frac{(\alpha t)^k}{k!} + C_k ,$$

und da $\sum_{k=0}^{\infty} P(N_t = k) = 1$, muss $C_k = 0$ sein.

□

23.2 Satz (Poisson-Prozess)
Der Prozess $(N_t)_{t\geq 0}$ heißt <u>*Poisson-Prozess*</u> *zum Parameter α. Er hat folgende Eigenschaften*

i) $N_0 = 0$ *und alle* N_t *sind* $\mathbb{N}_0$*-wertig.*

ii) Für $0 = t_0 < t_1 < \cdots < t_k$ *sind die Zuwächse* $N_{t_i} - N_{t_{i-1}}$ $(i = 1, \ldots, k)$ *unabhängig und nach* $\mathcal{P}_{\alpha(t_i - t_{i-1})}$ *verteilt.*

iii) Für alle $\omega \in \Omega$ *ist* $t \mapsto N_t(\omega)$ *nicht fallend und rechtsseitig stetig mit Sprunghöhe* 1.

Beweis: i) $N_0 = 0$, da $S_1 > 0$. $N_t(\omega) \in \mathbb{N}_0$ folgt direkt aus der Definition von N_t.
iii) Die Monotonie folgt aus der Definition, die Sprunghöhe 1 aus der Beobachtung, dass $S_n < S_{n+1}$, und für die rechtsseitige Stetigkeit beachte man

$$N_t = n \quad \Longrightarrow \quad S_n \leq t < S_{n+1} \quad \Longrightarrow \quad N'_t = n \;\forall\, t' \in [t, S_{n+1})$$

ii) Zum Beweis der Verteilungsaussage führen wir folgende Hilfsprozesse ein: Sei $t \geq 0$. Definiere

$$X_1^{(t)} := S_{N_t+1} - t \text{ (“erstes Sprungereignis nach } t\text{”)}, \quad X_k^{(t)} := X_{N_t+k}\ (k \geq 2) .$$

Weiter unten werden wir zeigen:

Für alle $n \in \mathbb{N}_0$ und $t \geq 0$ ist die bedingte Verteilung von $(X_k^{(t)})_{k\geq 1}$ gegeben $\{N_t = n\}$ gleich der Verteilung von $(X_k)_{k\geq 1}$. (**)

Zu $(X_k^{(t)})_{k\geq 1}$ definieren wir wie vorher $S_n^{(t)} = X_1^{(t)} + \cdots + X_n^{(t)}$ und

$$N_s^{(t)} := \max\left\{n \geq 0 : S_n^{(t)} \leq s\right\} .$$

Dann ist $S_n^{(t)} = S_{N_t+n} - t$ und

$$N_s^{(t)} = \max\{n \geq 0 : S_{N_t+n} \leq s+t\} = \max\{m \geq 0 : S_m \leq s+t\} - N_t = N_{s+t} - N_t ,$$

und wegen (**) hat $(N_s^{(t)})_{s\geq 0}$ die gleiche Verteilung wie $(N_s)_{s\geq 0}$. Insbesondere sind die Zuwächse $N_{s+t} - N_t$ nach $\mathcal{P}_{\alpha s}$ verteilt.

Seien nun $0 = t_0 < t_1 < \cdots < t_k$ und $n_i \in \mathbb{N}_0$ $(i = 1, \ldots, k)$. Dann ist

$$N_{t_i} - N_{t_{i-1}} = (N_{t_i} - N_{t_1}) - (N_{t_{i-1}} - N_{t_1}) = N^{(t_1)}_{t_i - t_1} - N^{(t_1)}_{t_{i-1}-t_1}$$

[1] f_n ist die Dichte der Γ-Verteilung mit Parametern α und n.

und daher

$$\begin{aligned}
&P\left\{N_{t_i}-N_{t_{i-1}}=n_i\,(i=1,\dots,k)\right\}\\
&\quad=P\Big\{N_{t_1}=n_1,\underbrace{N^{(t_1)}_{t_i-t_1}-N^{(t_1)}_{t_{i-1}-t_1}=n_i\,(i=2,\dots,k)}_{\sigma\left(X_j^{(t_1)}:j=1,2,\dots\right)\text{-messbar}}\Big\}\\
&\quad=P\left\{N_{t_1}=n_1\right\}\cdot P\left\{N_{t_i-t_1}-N_{t_{i-1}-t_1}=n_i\,(i=2,\dots,k)\right\}\quad\text{wegen (**)}\\
&\quad=P\left\{N_{t_1-t_0}=n_1\right\}\cdot P\left\{N_{t_i-t_1}-N_{t_{i-1}-t_1}=n_i\,(i=2,\dots,k)\right\}\ .
\end{aligned}$$

Durch Induktion folgt

$$P\left\{N_{t_i}-N_{t_{i-1}}=n_i\,(i=1,\dots,k)\right\}=\prod_{i=1}^k P\left\{N_{t_i-t_{i-1}}=n_i\right\}=\prod_{i=1}^k P\left\{N_{t_i}-N_{t_{i-1}}=n_i\right\}$$

da $P_{N_{t_i-t_{i-1}}}=\mathcal{P}_{\alpha(t_i-t_{i-1})}=P_{N_{t_i}-N_{t_{i-1}}}$, das heißt die Zuwächse sind unabhängig.

Es bleibt (**) zu beweisen: Zunächst ist

$$\begin{aligned}
P\left\{N_t=n,S_{n+1}>y+t\right\}&=P\left\{S_n\le t,X_{n+1}>y+t-S_n\right\}\\
&=E\left[1_{\{S_n\le t\}}\cdot P[X_{n+1}>y+(t-S_n)\big|\sigma(S_n)]\right]\\
&=E\Big[1_{\{S_n\le t\}}\cdot\underbrace{P[X_{n+1}>y\big|\sigma(S_n)]}_{=e^{-\alpha y}}\cdot P[X_{n+1}>t-S_n\big|\sigma(S_n)]\Big]
\end{aligned}$$

da X_{n+1} von S_n unabhängig, also die bedingte Verteilung von X_{n+1} gegeben S_n gleich $P_{X_{n+1}}=\mathcal{E}_\alpha$ ist (beachte die Gedächtnislosigkeit der Exponentialverteilung!). Daher ist

$$P\left\{N_t=n,S_{n+1}>y+t\right\}=e^{-\alpha y}\cdot P\left\{S_n\le t<S_{n+1}\right\}=P\left\{X_{n+1}>y\right\}\cdot P\left\{N_t=n\right\}\ .$$

Es folgt für $y_1,\dots,y_k\ge 0$:

$$\begin{aligned}
&P\left\{N_t=n,X_1^{(t)}>y_1,X_2^{(t)}>y_2,\dots,X_k^{(t)}>y_k\right\}\\
&\quad=P\Big\{\underbrace{N_t=n,S_{n+1}>y_1+t}_{\sigma(X_1,\dots,X_{n+1})\text{-messbar}},X_{n+2}>y_2,\dots,X_{n+k}>y_k\Big\}\\
&\quad=P\left\{N_t=n,S_{n+1}>y_1+t\right\}\cdot\prod_{i=2}^k P\left\{X_{n+i}>y_i\right\}\\
&\quad=P\left\{N_t=n\right\}\cdot\prod_{i=1}^k\underbrace{P\left\{X_{n+i}>y_i\right\}}_{=P\{X_i>y_i\}}\\
&\quad=P\left\{N_t=n\right\}\cdot P\left\{X_1>y_1,\dots,X_k>y_k\right\}\ ,
\end{aligned}$$

also die Behauptung (**). □

Kapitel 24

Martingale, Submartingale, Stoppzeiten

In Übung 20.6 hatten wir gesehen, dass der Partialsummenprozess unabhängiger Zufallsvariablen insbesondere eine Markov-Kette ist. In diesem Kapitel lernen wir eine andere, sehr weit reichende Klasse von Prozessen kennen, die die Partialsummenprozesse unabhängiger Zufallsvariablen mit Erwartung 0 verallgemeinern.

Allen Definitionen und Sätzen dieses Kapitels liegt ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ zu Grunde.

24.1 Definition (Filtration, adaptierter Prozess) *Sei $T \subseteq \bar{\mathbb{R}}$.*

a) *Eine Familie $(\mathcal{F}_t)_{t \in T}$ von Teil-σ-Algebren von $\mathcal{A}$ heißt eine Filtration, falls $\mathcal{F}_s \subseteq \mathcal{F}_t$ für $s \leq t$.*

b) *Eine Familie $\mathcal{X} = (X_t, \mathcal{F}_t)_{t \in T}$ heißt ein adaptierter Prozess, falls X_t für jedes $t \in T$ eine $\mathcal{F}_t$-messbare Zufallsvariable ist.*

24.2 Definition (Martingal) *Ein adaptierter Prozess $\mathcal{X}$ heißt ein Submartingal, falls für alle $s, t \in T$ mit $s < t$ gilt*

i) *$E|X_s| < \infty$ und*

ii) *$E[X_t \big| \mathcal{F}_s] \geq X_s$.*

Ein Prozess $\mathcal{X}$ ist ein Supermartingal, falls $-\mathcal{X} = (-X_t, \mathcal{F}_t)_{t \in T}$ ein Submartingal ist, und ein Martingal, falls er sowohl ein Submartingal als auch ein Supermartingal ist.

24.3 Bemerkung Ein Martingal ist also eine adaptierte Familie $(X_t, \mathcal{F}_t)_{t \in T}$ integrierbarer Zufallsvariablen mit $E[X_t \big| \mathcal{F}_s] = X_s$ falls $s < t$.

Wir sagen auch, dass $(X_t)_{t \in T}$ ein (Sub-, Super-) Martingal ist, falls $(X_t, \mathcal{F}_t)_{t \in T}$ mit $\mathcal{F}_t := \sigma(X_s : s \leq t)$ eines ist. (Man beachte den Unterschied zur Notation in Kapitel 20, wo wir diese σ-Algebra mit $\mathcal{F}_{\leq t}$ bezeichnet hätten.)

Zum Ursprung der Bezeichnung "Martingal" werden wir in Beispiel 25.2 noch ein paar Anmerkungen machen.

24.4 Bemerkungen

a) Die Summe endlich vieler (Sub-, Super-) Martingale ist ein (Sub-, Super-) Martingal.

b) Ist T ein Intervall in $\mathbb{Z}$, so folgt Eigenschaft 24.2ii aus: $E[X_{n+1} \big| \mathcal{F}_n] = X_n$ für alle n mit $n, n+1 \in T$.

24.5 Beispiele Wir beschränken uns zunächst auf $T = \mathbb{N}_0$ als Indexmenge.

A) Seien $\xi_0, \xi_1, \xi_2, \ldots$ unabhängige Zufallsvariablen, $E|\xi_i| < \infty$ für alle $i \geq 0$, $\mathcal{F}_n = \sigma(\xi_0, \ldots, \xi_n)$.

a) Sei $E\xi_i = 0$ (bzw. ≥ 0) für alle i, $X_n = \xi_0 + \cdots + \xi_n$. Dann ist $(X_n, \mathcal{F}_n)_{n\geq 0}$ ein Martingal (bzw. Submartingal), denn für $n \geq 0$ ist

$$E|X_n| \leq E|\xi_0| + \cdots + E|\xi_n| < \infty \,,$$

$$E[X_{n+1}\big|\mathcal{F}_n] = E[X_n\big|\mathcal{F}_n] + E[\xi_{n+1}\big|\mathcal{F}_n] \overset{S.19.6c, B.19.5c}{=} X_n + E\xi_{n+1} = (\geq) X_n \,.$$

b) Sei $E\xi_i = 1$ für alle i, $X_n = \xi_0 \cdot \xi_1 \cdot \cdots \cdot \xi_n$. Dann ist $(X_n, \mathcal{F}_n)_{n\geq 0}$ ein Martingal, denn für $n \geq 0$ ist

$$E|X_n| = E|\xi_0| \cdot \cdots \cdot E|\xi_n| < \infty \,,$$

$$E[X_{n+1}\big|\mathcal{F}_n] = E[X_n \cdot \xi_{n+1}\big|\mathcal{F}_n] \overset{S.19.7}{=} X_n \cdot E[\xi_{n+1}\big|\mathcal{F}_n] \overset{B.19.5c}{=} X_n \cdot E\xi_{n+1} = X_n \,.$$

B) Sei $\xi \in \mathcal{L}^1_P$, $(\mathcal{F}_n)_{n\geq 0}$ eine Filtration, $X_n := E[\xi\big|\mathcal{F}_n]$. Dann ist $(X_n, \mathcal{F}_n)_{n\geq 0}$ ein Martingal, denn für $n \geq 0$ ist

$$E|X_n| \overset{S.19.6d}{\leq} E\left[E[|\xi|\big|\mathcal{F}_n]\right] = E|\xi| < \infty \,,$$

$$E[X_{n+1}\big|\mathcal{F}_n] = E\left[E[\xi\big|\mathcal{F}_{n+1}]\big|\mathcal{F}_n\right] \overset{S.19.8}{=} E[\xi\big|\mathcal{F}_n] = X_n \,.$$

C) Sei $(X_n)_{n\geq 0}$ eine zeitlich homogene (aber nicht notwendig stationäre) Markov-Kette auf $(\Omega, \mathcal{A}, P)$ mit Werten in $[0, \infty)$ und $X_0 \in \mathcal{L}^1_P$. Sie habe den Übergangskern $K(x, A)$, siehe Kapitel 20, und es gebe eine Konstane $\gamma > 0$ derart, dass $\int_0^\infty y \cdot K(x, dy) = \gamma\, x$ für P_{X_0}-f.a. x. Setze $Y_n := \gamma^{-n} X_n$, $\mathcal{F}_n := \sigma(X_0, \ldots, X_n)$. Dann ist $(Y_n, \mathcal{F}_n)_{n\geq 0}$ ein Martingal, denn aus der Markov-Eigenschaft der X_n folgt

$$EX_n = E\left[\int y \cdot K(X_{n-1}, dy)\right] = \gamma\, EX_{n-1} = \cdots = \gamma^n EX_0 < \infty \,,$$

also $E|Y_n| = \gamma^n E[X_n] < \infty$, und

$$\begin{aligned} E[Y_{n+1}\big|\mathcal{F}_n] &= \gamma^{-(n+1)}\, E[X_{n+1}\big|\sigma(X_0, \ldots, X_n)] = \gamma^{-(n+1)}\, E[X_{n+1}\big|\sigma(X_n)] \\ &= \gamma^{-(n+1)} \int y \cdot K(X_n, dy) = \gamma^{-(n+1)}\, \gamma\, X_n = Y_n \,. \end{aligned}$$

Ein Spezialfall dieser Situation wird im folgenden Beispiel genauer untersucht.

24.6 Beispiel (Galton-Watson Prozess) Seien $N_{n,k}$, $(n, k = 1, 2, 3, \ldots)$ unabhängig identisch verteilte $\mathbb{N}_0$-wertige Zufallsvariablen. Definiere

$$X_0 := 1, \quad X_{n+1} := \sum_{k=1}^{X_n} N_{n,k} \,.$$

Die klassische Interpretation für diesen Prozess ist

X_n ist die Größe einer Population mit nicht überlappenden Generationen zur Zeit n,

$N_{n,k}$ ist die Zahl der Nachkommen des k-ten Individuums in der Population zur Zeit n.

Durch Induktion zeigt man leicht: Für jedes n sind die $N_{n,k}$ unabhängig von X_n.

Sei ν die Verteilung der $N_{n,k}$ auf $\mathbb{N}_0$. Da für $r \in \mathbb{N}_0$ und $A \subseteq \mathbb{N}_0$

$$P(X_{n+1} \in A | X_n = r) = P\left(\sum_{k=1}^{r} N_{n,k} \in A\right) = \underbrace{(\nu * \cdots * \nu)}_{r\text{-mal}}(A) =: K(r, A) \,,$$

ist $(X_n)_{n\geq 0}$ eine zeitlich homogene Markovkette mit Übergangskern K, und es ist

$$\int_0^\infty y\,K(r,dy) = \sum_{s=0}^\infty s\cdot \underbrace{(\nu * \cdots * \nu)}_{r\text{-mal}}\{s\} = r\cdot \underbrace{EN_{n,k}}_{=:\gamma}$$

$(X_n)_{n\geq 0}$ heißt auch Galton-Watson-Prozess.

24.7 Beispiel (Martingaleigenschaft des Poisson-Prozesses) Sei $(N_t)_{t\geq 0}$ der Poisson-Prozess mit Parameter α aus Satz 23.2, $\mathcal{F}_t = \sigma(N_s : s \leq t)$. Dann ist $(N_t - \alpha t, \mathcal{F}_t)_{t\geq 0}$ ein Martingal, denn für $0 \leq s < t$ ist wegen der Unabhängigkeit der Zuwächse $N_t - N_s$ von der σ-Algebra $\mathcal{F}_s$

$$\begin{aligned} E[N_t - \alpha t\big|\mathcal{F}_s] &= E[N_t - N_s\big|\mathcal{F}_s] - \alpha(t-s) + (N_s - \alpha s) \\ &= E[N_t - N_s] - \alpha(t-s) + (N_s - \alpha s) = N_s - \alpha s\ . \end{aligned}$$

24.8 Satz (Martingaleigenschaften der Brown'schen Bewegung)
Sei $(B_t)_{t\geq 0}$ eine Brown'sche Bewegung. Dann sind die folgenden Prozesse Martingale:

i) B_t *selbst,*

ii) $B_t^2 - t$,

iii) $M_t^\alpha := \exp\left(\alpha B_t - \frac{\alpha^2}{2}t\right)$ *für jedes* $\alpha \in \mathbb{R}$.

Beweis: Sei $0 \leq s < t$.

i) $E[B_t\big|\mathcal{F}_s] = E[B_t - B_s\big|\mathcal{F}_s] + E[B_s\big|\mathcal{F}_s] = B_s$, da $B_t - B_s$ von $\mathcal{F}_s$ unabhängig ist (vergleiche Beispiel 19.5c).

ii) Ähnlich ist

$$\begin{aligned} E[B_t^2\big|\mathcal{F}_s] &= E[(B_t - B_s)^2\big|\mathcal{F}_s] + 2E[(B_t - B_s)B_s\big|\mathcal{F}_s] + E[B_s^2\big|\mathcal{F}_s] \\ &= E[(B_t - B_s)^2] + 2B_s E[B_t - B_s\big|\mathcal{F}_s] + B_s^2 = t - s + B_s^2\ , \end{aligned}$$

so dass $E[B_t^2 - t\big|\mathcal{F}_s] = B_s^2 - s$.

iii) Es ist $M_t^\alpha = M_s^\alpha \cdot \exp\left(\alpha(B_t - B_s) - \frac{\alpha^2}{2}(t-s)\right)$, und da M_s^α bzgl. $\mathcal{F}_s$ messbar, der Exponent von $\mathcal{F}_s$ unabhängig und $P_{B_t - B_s} = \mathcal{N}(0, t-s)$ ist, folgt

$$\begin{aligned} E[M_t^\alpha\big|\mathcal{F}_s] &= M_s^\alpha \cdot E\left[\exp\left(\alpha(B_t - B_s) - \frac{\alpha^2}{2}(t-s)\right)\right] \\ &= M_s^\alpha \cdot e^{-\frac{\alpha^2}{2}(t-s)}\,\frac{1}{\sqrt{2\pi(t-s)}}\int_{\mathbb{R}} e^{\alpha x - \frac{x^2}{2(t-s)}}\,dx \\ &= M_s^\alpha\ . \end{aligned}$$

□

24.9 Lemma (Konvexe Transformation von Submartingalen) *Sei $I \subseteq \mathbb{R}$ ein (nicht notwendig endliches) Intervall, $\varphi : I \to \mathbb{R}$ konvex. Sei $\mathcal{X} = (X_t, \mathcal{F}_t)_{t\in T}$ ein Submartingal mit Werten in I.*

Ist a) φ nicht fallend, oder b) $\mathcal{X}$ ein Martingal, so ist auch $\varphi(\mathcal{X}) = (\varphi(X_t), \mathcal{F}_t)_{t\in T}$ ein Submartingal, falls die $\varphi(X_t)$ integrierbar sind.

Beweis: Mit X_t ist auch $\varphi(X_t)$ bzgl. $\mathcal{F}_t$ messbar. Außerdem folgt aus der bedingten Jensen'schen Ungleichung (Satz 19.12) für $s < t$

$$E[\varphi(X_t)\big|\mathcal{F}_s] \geq \varphi\left(E[X_t\big|\mathcal{F}_s]\right) \begin{cases} \geq \varphi(X_s) & \text{im Fall a)} \\ = \varphi(X_s) & \text{im Fall b).} \end{cases}$$

□

24.10 Beispiel Ist $\mathcal{X} = (X_t, \mathcal{F}_t)_{t\in T}$ ein Martingal, so sind $|\mathcal{X}|^p = (|X_t|^p, \mathcal{F}_t)_{t\in T}$ für $p \geq 1$, $\mathcal{X}^+ = (X_t^+, \mathcal{F}_t)_{t\in T}$ und $\mathcal{X}^- = (X_t^-, \mathcal{F}_t)_{t\in T}$ Submartingale. Das folgt aus dem vorhergehenden Lemma für $\varphi(x) = |x|^p, x^+, x^-$. (Für $\mathcal{X}^+$ bleibt die Aussage sogar richtig, wenn $\mathcal{X}$ nur ein Submartingal ist.)

Die Bedeutung von Martingalen in der Stochastik liegt zu einem großen Teil darin, dass man aus einem gegebenen Martingal auf vielfältige Weise neue Martingale und Submartingale gewinnen kann. Lemma 24.9 war ein erstes einfaches Beispiel dafür. Für die Theorie und die Anwendungen von Martingalen von zentraler Bedeutung ist die *Transformation durch Stoppzeiten*.

24.11 Definition (Stoppzeit, Abbruchregel) *Sei $(\mathcal{F}_t)_{t\in T}$ eine Filtration. Eine Abbildung $\tau : \Omega \to T$ heißt Stoppzeit (bzgl. $(\mathcal{F}_t)_{t\in T}$), falls*

$$\{\tau \leq t\} := \{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \forall\, t \in T\,.$$

Ist $\tau < \infty$ P-f.s., so heißt τ auch Abbruchregel.

24.12 Bemerkung Ist $T = \mathbb{N}$ oder $T = \mathbb{N} \cup \{\infty\}$, so sind

$$\{\tau = n\} \in \mathcal{F}_n \quad \text{und} \quad \{\tau < n\}, \{\tau \geq n\} \in \mathcal{F}_{n-1} \text{ für alle } n,$$

denn

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}\,,$$
$$\{\tau < n\} = \{\tau \leq n-1\} \quad \text{und} \quad \{\tau \geq n\} = \{\tau < n\}^c\,.$$

24.13 Beispiel Ist $(X_n, \mathcal{F}_n)_{n\geq 0}$ adaptiert und $B \subseteq \bar{\mathbb{R}}$ messbar, so ist

$$\tau_B := \inf\{n \geq 0 : X_n \in B\} \quad (\tau_B = +\infty, \text{ falls } \forall\, n \geq 0 :\ X_n \notin B)$$

eine Stoppzeit, die Zeit des ersten Eintritts in B, denn $\{\tau_B \leq n\} = \bigcup_{k=0}^n \{X_k \in B\} \in \mathcal{F}_n$. Das ist für überabzählbare T nicht immer richtig!

24.14 Lemma *Sind τ_1, τ_2 Stoppzeiten, so auch $\tau_1 \wedge \tau_2 := \min\{\tau_1, \tau_2\}$.*

Beweis: $\{\tau_1 \wedge \tau_2 \leq n\} = \{\tau_1 \leq n\} \cup \{\tau_2 \leq n\} \in \mathcal{F}_n$. □

24.15 Definition *Ist τ eine Stoppzeit bzgl. $(\mathcal{F}_t)_{t\in T}$, so heißt*

$$\mathcal{F}_\tau := \{A \in \mathcal{A} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ für alle } t \in T\}$$

die σ-Algebra der τ-Vergangenheit.

24.16 Bemerkung $\mathcal{F}_\tau$ ist tatsächlich eine σ-Algebra (*Übung*).

24.17 Satz (Diskretes stochastisches Integral)
Sei $\mathcal{X} = (X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein (Sub-)Martingal.

a) *Seien H_n beschränkte, nicht negative $\mathcal{F}_{n-1}$-messbare Zufallsvariablen $(n = 1, 2, \dots)$. Setze*

$$Y_0 = X_0,\ Y_{n+1} = Y_n + H_{n+1}(X_{n+1} - X_n)\ , \quad \text{also } Y_n = X_0 + \sum_{k=1}^{n} H_k(X_k - X_{k-1})\ .$$

Dann ist $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein (Sub-)Martingal. Y_n ist das diskrete stochastische Integral *von H bzgl. X und wird mit $(H \cdot X)_n$ bezeichnet. Ist $\mathcal{X}$ ein Martingal, so kann auf die Nichtnegativitätsannahme an die H_n verzichtet werden.*

b) *Sei τ eine Stoppzeit bzgl. $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Dann ist der gestoppte Prozess $(X_{\tau\wedge n}, \mathcal{F}_{\tau\wedge n})_{n\in\mathbb{N}_0}$ ein (Sub-) Martingal.*

Beweis:

a) Die Y_n sind offensichtlich adaptiert, und da die H_n beschränkt sind, sind die Y_n integrierbar. Schließlich ist

$$E[Y_{n+1}|\mathcal{F}_n] = Y_n + H_{n+1}(E[X_{n+1}|\mathcal{F}_n] - X_n) \overset{H_{n+1}\geq 0}{\geq} Y_n$$

mit Gleichheit, falls $\mathcal{X}$ ein Martingal ist.

b) Setze $H_n := 1_{\{\tau\geq n\}}$. Dann ist $H_n = 1 - 1_{\{\tau\leq n-1\}}$ $\mathcal{F}_{n-1}$-messbar und $(H \cdot X)_n = X_{\tau\wedge n}$, denn $(H \cdot X)_0 = X_0 = X_{\tau\wedge 0}$, und induktiv folgt

$$\begin{aligned}(H \cdot X)_{n+1} &= X_{\tau\wedge n} + 1_{\{\tau\geq n+1\}}(X_{n+1} - X_n) = \begin{cases} X_\tau & \text{falls } \tau \leq n \\ X_{n+1} & \text{falls } \tau \geq n+1 \end{cases} \\ &= X_{\tau\wedge(n+1)}\ .\end{aligned}$$

□

24.18 Satz (Optionales Stoppen, einfachste Version)
Seien $\sigma \leq \tau$ beschränkte *Stoppzeiten bzgl. $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$.*

a) *Sei $\mathcal{X} = (X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Submartingal. Dann ist $X_\sigma \leq E[X_\tau|\mathcal{F}_\sigma]$ mit Gleichheit, falls $\mathcal{X}$ ein Martingal ist.*

b) *Ein adaptierter integrierbarer Prozess $(X_n)_{n\in\mathbb{N}_0}$ ist ein Martingal genau dann, wenn $E[X_\sigma] = E[X_\tau]$ für jedes solche Paar von Stoppzeiten.*

Beweis: Wir beginnen mit dem Beweis folgender Hilfsbehauptung:

Ist $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein integrierbarer adaptierter Prozess und ist $E[X_\sigma] \overset{(\leq)}{=} E[X_\tau]$ für alle beschränkten Stoppzeiten $\sigma \leq \tau \leq M$, so ist sogar $X_\sigma \overset{(\leq)}{=} E[X_\tau|\mathcal{F}_\sigma]$ für alle solche Stoppzeiten. (*)

Für $B \in \mathcal{F}_\sigma$ sei $\sigma^B := \sigma 1_B + M 1_{B^c}$, $\tau^B := \tau 1_B + M 1_{B^c}$. Dann sind $\sigma^B \leq \tau^B \leq M$ Stoppzeiten, denn

$$\{\sigma^B \leq n\} = \begin{cases} B \cap \{\sigma \leq n\} & (n < M) \\ \Omega & (n \geq M) \end{cases} \qquad \text{gehört zu } \mathcal{F}_n.$$

Daher ist

$$E\left[X_\sigma 1_B + X_M 1_{B^c}\right] = E\left[X_{\sigma^B}\right] \overset{(\leq)}{=} E\left[X_{\tau^B}\right] = E\left[X_\tau 1_B + X_M 1_{B^c}\right]\ ,$$

so dass $E\left[X_\sigma 1_B\right] \overset{(\leq)}{=} E\left[X_\tau 1_B\right]$. Da das für alle $B \in \mathcal{F}_\sigma$ gilt, folgt $X_\sigma \overset{(\leq)}{=} E\left[X_\tau \middle| \mathcal{F}_\sigma\right]$, siehe Lemma 19.4b.

a) Sei nun zunächst $\mathcal{X}$ ein (Sub-)Martingal und $\sigma \leq \tau \leq M$. Dann sind

$$|X_\sigma|, |X_\tau| \leq \max\{|X_1|, \ldots, |X_M|\} \in \mathcal{L}^1_P \ .$$

Setze $H^\sigma_n = 1_{\{\sigma \geq n\}}$, $H^\tau_n = 1_{\{\tau \geq n\}}$. Es ist $0 \leq H^\tau_n - H^\sigma_n = 1_{\{\sigma < n \leq \tau\}} \leq 1$, also

$$\left((H^\tau - H^\sigma) \cdot X\right)_n - X_0 = (H^\tau \cdot X)_n - (H^\sigma \cdot X)_n = X_{\tau \wedge n} - X_{\sigma \wedge n} = X_\tau - X_\sigma$$

falls $n \geq M$. (Es ist $(H^\tau \cdot X)_n = X_{\tau \wedge n}$ wie im Beweis von Satz 24.17b.) Da $((H^\tau - H^\sigma) \cdot X)$ nach Satz 24.17 ein (Sub-)Martingal ist, erhält man

$$\begin{aligned} E[X_\tau] - E[X_\sigma] &= E\left[\left((H^\tau - H^\sigma) \cdot X\right)_n\right] - E[X_0] \overset{(\geq)}{=} E\left[\left((H^\tau - H^\sigma) \cdot X\right)_0\right] - E[X_0] \\ &= E[X_0] - E[X_0] = 0 \ , \end{aligned} \qquad (**)$$

und aus (*) folgt $X_\sigma \overset{(\leq)}{=} E\left[X_\tau \middle| \mathcal{F}_\sigma\right]$.

b) Die zweite Aussage folgt nun sofort aus (*) und (**).

□

Eine erste Anwendung dieses Satzes ist der Beweis der folgenden *Maximalungleichung*.

24.19 Satz (Maximalungleichung)
Sei $(X_n, \mathcal{F}_n)_{n=0,\ldots,N}$ ein Submartingal. Dann gilt für jedes $\lambda > 0$

$$\lambda\, P\left\{\sup_n X_n \geq \lambda\right\} \leq E\left[X_N\, 1_{\{\sup_n X_n \geq \lambda\}}\right] \leq E\left[|X_N|\, 1_{\{\sup_n X_n \geq \lambda\}}\right] \ .$$

Beweis: Die zweite Ungleichung ist trivial. Wir zeigen die erste. Sei dazu

$$\tau = \begin{cases} \min\{n : X_n \geq \lambda\} & \text{falls diese Menge nicht leer ist,} \\ N & \text{sonst.} \end{cases}$$

$\tau \leq N$ ist eine Stoppzeit (siehe Beispiel 24.13). Daher folgt aus Satz 24.18, dass

$$\begin{aligned} E\left[X_N\right] \geq E\left[X_\tau\right] &= E\left[X_\tau\, 1_{\{\sup_n X_n \geq \lambda\}}\right] + E\left[X_\tau\, 1_{\{\sup_n X_n < \lambda\}}\right] \\ &\geq \lambda\, P\left\{\sup_n X_n \geq \lambda\right\} + E\left[X_N\, 1_{\{\sup_n X_n < \lambda\}}\right] \ . \end{aligned}$$

Durch Subtraktion von $E\left[X_N 1_{\{\sup_n X_n < \lambda\}}\right]$ auf beiden Seiten der Ungleichung folgt die Behauptung.
□

24.20 Korollar (Maximalungleichung) *Sei $(X_t, \mathcal{F}_t)_{t \in T}$ ein Martingal oder ein positives Submartingal. Dann gilt für höchstens abzählbare $D \subseteq T$, $p \geq 1$ und jedes $\lambda > 0$*

$$\lambda^p\, P\left\{\sup_{t \in D} |X_t| \geq \lambda\right\} \leq \sup_{t \in D} E\left[|X_t|^p\right] \ .$$

Beweis: Sei D_n eine aufsteigende Folge endlicher Teilmengen von D mit $\bigcup_n D_n = D$. O.B.d.A. sei $\sup_{t \in D} E\left[|X_t|^p\right] < \infty$. Dann ist $(|X_t|^p, \mathcal{F}_t)_{t \in D_n}$ für jedes n ein Submartingal (siehe Beispiel 24.10), und aus Satz 24.19 folgt

$$\lambda^p\, P\left\{\sup_{t \in D_n} |X_t| \geq \lambda\right\} = \lambda^p\, P\left\{\sup_{t \in D_n} |X_t|^p \geq \lambda^p\right\} \leq E\left[|X_{\max D_n}|^p\right] \leq \sup_{t \in D_n} E\left[|X_t|^p\right] \ .$$

Da $\left\{\sup_{t\in D_n}|X_t|>\lambda\right\}\nearrow\{\sup_{t\in D}|X_t|>\lambda\}$ (für "$\geq$" an Stelle von "$>$" gilt das leider nicht!), kann man schließen
$$\lambda^p\,P\left\{\sup_{t\in D}|X_t|>\lambda\right\}\leq\sup_n\sup_{t\in D_n}E\left[|X_t|^p\right]=\sup_{t\in D}E\left[|X_t|^p\right]\ .$$
Wendet man diese Abschätzung auf eine Folge $\lambda_j\nearrow\lambda$ an, erhält man die gleiche Abschätzung mit "$\geq$" an Stelle von "$>$", also das behauptete Ergebnis. □

Nun übertragen wir diese Maximalungleichung auf überabzählbare Indexmengen. Das folgende, wie auch viele andere grundlegende Aussagen über Martingale, gehen auf J.L. Doob zurück.

24.21 Satz (Maximalungleichung von Doob)
Sei $T\subseteq\mathbb{R}$ ein Intervall, $(X_t,\mathcal{F}_t)_{t\in T}$ ein Martingal oder ein positives Submartingal mit rechtsseitig stetigen Pfaden. Dann ist $X^*:=\sup_{t\in T}|X_t|$ messbar und für $p\geq 1$ und $\lambda>0$ gilt
$$\lambda^pP\left\{X^*\geq\lambda\right\}\leq\sup_{t\in T}E\left[|X_t|^p\right]\ .$$

Beweis: Sei D eine abzählbare dichte Teilmenge von T (die den rechten Endpunkt von T enthält, falls ein solcher existiert). Dann ist $X^*=\sup_{t\in D}|X_t|$, also X^* messbar und aus Korollar 24.20 folgt
$$\lambda^pP\left\{X^*\geq\lambda\right\}=\lambda^pP\left\{\sup_{t\in D}|X_t|\geq\lambda\right\}\leq\sup_{t\in D}E\left[|X_t|^p\right]\leq\sup_{t\in T}E\left[|X_t|^p\right]\ .$$
□

Wir wenden die Doob'sche Maximalungleichung nun auf die Brown'sche Bewegung an.

24.22 Korollar (Maximalungleichung für die Brown'sche Bewegung) *Sei $(B_t)_{t\geq 0}$ eine Brown'sche Bewegung. Für $a>0$ ist*
$$P\left\{\sup_{0\leq s\leq t}B_s\geq at\right\}\leq\exp\left(-\frac{a^2t}{2}\right)\ .$$

Beweis: Wir benutzen das Martingal $(M_s^\alpha)_{0\leq s\leq t}$, $M_s^\alpha=\exp\left(\alpha B_s-\frac{\alpha^2 s}{2}\right)$, siehe Satz 24.8. Da
$$\exp\left(\alpha\cdot\sup_{0\leq s\leq t}B_s-\frac{\alpha^2t}{2}\right)=\sup_{0\leq s\leq t}\exp\left(\alpha B_s-\frac{\alpha^2t}{2}\right)\leq\sup_{0\leq s\leq t}M_s^\alpha\ ,$$
folgt für jedes $\alpha>0$
$$\begin{aligned}P\left\{\sup_{0\leq s\leq t}B_s\geq at\right\}&=P\left\{\alpha\cdot\sup_{0\leq s\leq t}B_s-\frac{\alpha^2t}{2}\geq\alpha at-\frac{\alpha^2t}{2}\right\}\\&\leq P\left\{\sup_{0\leq s\leq t}M_s^\alpha\geq\exp\left(\alpha at-\frac{\alpha^2t}{2}\right)\right\}\\&\leq\exp\left(-\alpha at+\frac{\alpha^2t}{2}\right)\sup_{0\leq s\leq t}\underbrace{E\left[M_s^\alpha\right]}_{=E[M_0^\alpha]=1}\qquad\text{(Satz 24.21)}\ .\end{aligned}$$

Da $\inf_{\alpha>0}\left(-\alpha at+\frac{\alpha^2t}{2}\right)=-\frac{a^2t}{2}$ (setze $\alpha=a$), folgt die Behauptung.

Man beachte die Ähnlichkeit dieser Abschätzung zur Abschätzung (**) im Beweis des Satzes über große Abweichungen 14.1. □

24.23 Satz (Satz vom iterierten Logarithmus für die Brown'sche Bewegung $(t \searrow 0)$)
Sei $(B_t)_{t\geq 0}$ *eine Brown'sche Bewegung. Dann ist*

$$\limsup_{t\searrow 0} \frac{B_t}{\sqrt{2t\log(\log\frac{1}{t})}} = 1 \quad \textit{f.s.} \quad \textit{und} \quad \liminf_{t\searrow 0} \frac{B_t}{\sqrt{2t\log(\log\frac{1}{t})}} = -1 \quad \textit{f.s.}$$

Beweis: Die zweite Aussage folgt aus der ersten, da auch $(-B_t)_{t\geq 0}$ eine Brown'sche Bewegung ist. Wir beweisen die erste Aussage.

Sei $h(t) = \sqrt{2t\log(\log\frac{1}{t})}$ und seien $\delta, \theta \in (0,1)$ zwei Konstanten, die später bestimmt werden. Setze

$$\alpha_n := (1+\delta)\theta^{-n}h(\theta^n), \quad \beta_n := \frac{h(\theta^n)}{2}\ .$$

Dann ist $\alpha_n\beta_n = (1+\delta)\log(-n\log\theta)$. Genau wie im Beweis der Maximalungleichung zeigt man

$$\begin{aligned} P\left\{\sup_{0\leq s\leq 1}\left(B_s - \frac{\alpha_n s}{2}\right) \geq \beta_n\right\} &= P\left\{\sup_{0\leq s\leq 1}\left(\alpha_n B_s - \frac{\alpha_n^2 s}{2}\right) \geq \alpha_n\beta_n\right\} \\ &= P\left\{\sup_{0\leq s\leq 1} M_s^{\alpha_n} \geq e^{\alpha_n\beta_n}\right\} \\ &\leq e^{-\alpha_n\beta_n} = (-\log\theta)^{-1-\delta}\cdot n^{-1-\delta}\ . \end{aligned}$$

Daher folgt aus dem Borel-Cantelli Lemma, dass es zu fast jedem ω ein $n_0(\omega) \in \mathbb{N}$ gibt, so dass für alle $n \geq n_0(\omega)$ und alle $s \in [0, \theta^{n-1}]$ gilt

$$B_s(\omega) \leq \frac{\alpha_n s}{2} + \beta_n \leq \frac{\alpha_n\theta^{n-1}}{2} + \beta_n = \left(\frac{1+\delta}{2\theta} + \frac{1}{2}\right)h(\theta^n)\ .$$

Da die Funktion $t \mapsto h(t)$ auf einem Intervall $[0, \theta^{n_1})$, $n_1 > 0$, monoton wächst, gilt für diese ω, $n \geq n_0(\omega) + n_1$ und $s \in (\theta^n, \theta^{n-1}]$

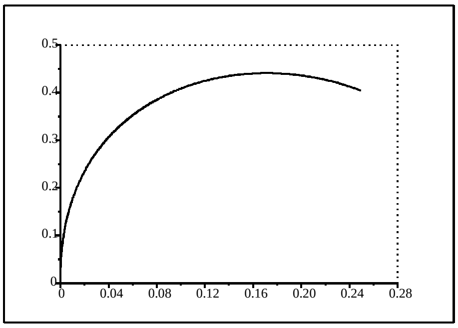
$$B_s(\omega) \leq \left(\frac{1+\delta}{2\theta} + \frac{1}{2}\right)h(s)\ .$$

Es folgt, dass

$$\limsup_{s\searrow 0}\frac{B_s}{h(s)} \leq \frac{1+\delta}{2\theta} + \frac{1}{2} \quad \text{f.s.}$$

Lässt man nun $\theta \nearrow 1$ und $\delta \searrow 0$ gehen, erhält man

$$\limsup_{s\searrow 0}\frac{B_s}{h(s)} \leq 1 \quad \text{f.s.}$$



Die Funktion $h(t)$

Es bleibt die umgekehrte Ungleichung zu zeigen. Für $\theta \in (0,1)$ sind die Ereignisse

$$\Gamma_n = \left\{B_{\theta^n} - B_{\theta^{n+1}} \geq (1-\sqrt{\theta})h(\theta^n)\right\}$$

unabhängig. Wir zeigen am Ende des Beweises, dass $\sum_{n=1}^\infty P(\Gamma_n) = +\infty$. Dann folgt aus dem Borel-Cantelli-Lévi Lemma[1], dass für fast alle ω gilt:

$$\text{Für unendlich viele } n \text{ ist } \omega \in \Gamma_n, \quad \text{d.h. } B_{\theta^n} \geq (1-\sqrt{\theta})\left(h(\theta^n) + B_{\theta^{n+1}}\right)\ .$$

[1] Sind die Ereignissse $A_1, A_2, \ldots$ unabhängig und ist $\sum_n P(A_n) = \infty$, so ist $\sum_n 1_{A_n} = \infty$ f.s. *Beweis:* $E[e^{-\sum_n 1_{A_n}}] = \prod_n E[e^{-1_{A_n}}] = \prod_n(1-(1-e^{-1})P(A_n)) = 0$, so dass $e^{-\sum_n 1_{A_n}} = 0$ f.s., also $\sum_n 1_{A_n} = \infty$ f.s.

Da $(-B_t)_{t\geq 0}$ ebenfalls eine Brown'sche Bewegung ist, folgt aus dem ersten Teil des Beweises, dass für fast alle ω ein $n_2(\omega)$ existiert, so dass $B_{\theta^{n+1}} \leq 2h(\theta^{n+1})$ für $n \geq n_2(\omega)$. Da $h(\theta^{n+1}) \leq 2\sqrt{\theta}h(\theta^n)$ für hinreichend große n, können wir folgern

$$B_{\theta^n} > (1-\sqrt{\theta})\left(h(\theta^n) - 2h(\theta^{n+1})\right) > h(\theta^n)\left(1-\sqrt{\theta}\right)\left(1-4\sqrt{\theta}\right) \quad \text{f.s. für unendlich viele } n.$$

Daher ist

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} \geq \limsup_{n\to\infty} \frac{B_{\theta^n}}{h(\theta^n)} \geq 1 - 5\sqrt{\theta} \quad \text{f.s.,}$$

und im Limes $\theta \searrow 0$ folgt die Behauptung.

Bleibt zu zeigen, dass $\sum_{n=1}^{\infty} P(\Gamma_n) = +\infty$. Setze

$$a_n := \frac{1-\sqrt{\theta}}{\sqrt{1-\theta}} \cdot \frac{h(\theta^n)}{\sqrt{\theta^n}}\,, \quad \text{also } \frac{a_n^2}{2} = \frac{1-2\sqrt{\theta}+\theta}{1-\theta} \log(\log\theta^{-n}) =: C(\theta)\ \log(n|\log\theta|)$$

mit $C(\theta) \nearrow 1$ für $\theta \searrow 0$. Beachte nun, dass $\frac{B_{\theta^n} - B_{\theta^{n+1}}}{\sqrt{\theta^n(1-\theta)}}$ nach $\mathcal{N}(0,1)$ verteilt ist. Daher ist für hinreichend große n

$$P(\Gamma_n) = \frac{1}{\sqrt{2\pi}} \int_{a_n}^{\infty} e^{-\frac{t^2}{2}}\,dt \overset{L.24.24}{>} \frac{1}{\sqrt{2\pi}} \frac{a_n}{1+a_n^2}\, e^{-\frac{a_n^2}{2}} > \frac{C_1(\theta)}{\sqrt{\log n}}\, n^{-C(\theta)}$$

mit einer geeigneten Konstanten $C_1(\theta) > 0$ und $C(\theta) \in (0,1)$ wie oben. Es folgt $\sum_{n=1}^{\infty} P(\Gamma_n) = +\infty$.
□

Das in der letzten Abschätzung zitierte Lemma liefern wir nach:

24.24 Lemma *Für jedes $a > 0$ ist*

$$\int_a^{\infty} e^{-\frac{t^2}{2}}\,dt > \frac{a}{1+a^2}\, e^{-\frac{a^2}{2}}\,.$$

Beweis: Die Behauptung folgt aus

$$a^{-2}\int_a^{\infty} e^{-\frac{t^2}{2}}\,dt > \int_a^{\infty} t^{-2} e^{-\frac{t^2}{2}}\,dt = -t^{-1}e^{-\frac{t^2}{2}}\Big|_a^{\infty} - \int_a^{\infty} e^{-\frac{t^2}{2}}\,dt = a^{-1}e^{-\frac{a^2}{2}} - \int_a^{\infty} e^{-\frac{t^2}{2}}\,dt\,.$$

□

24.25 Korollar (Satz vom iterierten Logarithmus für die Brown'sche Bewegung $(t \to \infty)$)
Sei $(B_t)_{t\geq 0}$ eine Brown'sche Bewegung. Dann ist

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log(\log t)}} = 1 \quad \textit{f.s.} \quad \textit{und} \quad \liminf_{t\to\infty} \frac{B_t}{\sqrt{2t\log(\log t)}} = -1 \quad \textit{f.s.}$$

Beweis: Wende Satz 24.23 auf die Brown'sche Bewegung $(tB_{1/t})_{t>0}$ an (siehe Korollar 22.14d). □

24.26 Bemerkung Da die Brown'sche Bewegung stetige Pfade hat, folgt aus Satz 24.23, dass $t = 0$ für fast alle ω ein Häufungspunkt der Nullstellen von $t \mapsto B_t(\omega)$ ist, und aus dem Korollar, dass die Menge der Nullstellen von $t \mapsto B_t(\omega)$ für fast alle ω unbeschränkt ist.

Kapitel 25

Der Martingal-Konvergenzsatz

Das Hauptergebnis dieses Kapitels ist Doobs Konvergenzsatz für (Sub-)Martingale mit Indexmenge $\mathbb{N}_0$.

25.1 Satz (Martingal-Konvergenzsatz)
Sei $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Submartingal und gelte

$$\sup_{n\in\mathbb{N}_0} E\left[X_n^+\right] < +\infty\,.$$

Dann existiert $X_\infty := \lim_{n\to\infty} X_n$ fast sicher und $E|X_\infty| < \infty$, also insbesondere $|X_\infty| < \infty$ f.s. Genauer: $X_\infty := \lim_{n\to\infty} X_n$, wo immer der Limes existiert und $X_\infty := 0$ sonst. Damit ist X_∞ messbar bzgl. $\mathcal{F}_\infty := \left(\bigcup_{n\in\mathbb{N}_0} \mathcal{F}_n\right)$.

Man beachte, dass zwar $X_\infty \in \mathcal{L}_P^1$, aber trotzdem die Konvergenz $X_n \to X_\infty$ nur fast sicher und nicht in $\mathcal{L}_P^1$ gilt. Die damit verbundenen Schwierigkeiten bei der Interpretation des Martingal-Konvergenzsatzes soll das folgende Beispiel illustrieren.

25.2 Beispiel Seien $\xi_1, \xi_2, \dots$ unabhängige Zufallsvariablen, die die Werte $2 + \frac{2\alpha}{n^2}$ und 0 je mit Wahrscheinlichkeit $\frac{1}{2}$ annehmen, wobei $\alpha \geq 0$. Sei $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$ und $X_n = \prod_{k=1}^n \xi_k$. X_n kann als Guthaben eines Spielers zur Zeit n interpretiert werden, der bei einem fairen Münzwurfspiel alles setzt und diesen Einsatz entweder verliert oder ihn etwas mehr als verdoppelt. Dann ist $(X_n, \mathcal{F}_n)_{n\geq 1}$ ein Submartingal. In der Tat ist $E[X_{n+1}\big|\mathcal{F}_n] = X_n \cdot (1 + \frac{\alpha}{n^2}) \geq X_n$ mit strikter Ungleichung, falls $\alpha > 0$ und $X_n > 0$. Ein Spieler, der im Fall $\alpha > 0$ immer nur einen Schritt in die Zukunft schaut, mag versucht sein, endlos zu spielen, da das erwartete Guthaben $E[X_{n+1}\big|\mathcal{F}_n]$ nach dem nächsten Spiel immer echt größer als sein gegenwärtiges Guthaben ist (solange er überhaupt noch ein Guthaben hat). Trotzdem ist natürlich $X_\infty = \lim_{n\to\infty} X_n = 0$ f.s., d.h. der Spieler verliert mit dieser Strategie fast sicher sein gesamtes Guthaben. Beachte auch, dass

$$1 \leq E[X_n^+] = E[X_n] = \prod_{k=1}^n (1 + \frac{\alpha}{n^2}) \leq \prod_{k=1}^\infty (1 + \frac{\alpha}{n^2}) < \infty$$

für alle n ist, so dass die Voraussetzungen des Martingal-Konvergenzsatzes erfüllt sind.

Im Fall $\alpha = 0$ liegt ein Martingal vor, und man hat es mit einer fairen Spielstrategie zu tun, die den Martingalen ihren Namen gegeben haben mag. Im Provenzalischen heißt diese Strategie beim Roulettespiel “jouga a la martegalo”, wie man im Buch *Wahrscheinlichkeitstheorie* von Bauer auf Seite 144 nachlesen kann, wo auch noch weitere Hinweise zu finden sind.

Den Beweis des Martingal-Konvergenzsatzes bereiten wir mit einer Bemerkung und zwei Lemmas vor.

25.3 Lemma *Sei T endlich, $(X_t, \mathcal{F}_t)_{t\in T}$ ein adaptierter Prozess, τ eine Stoppzeit, $B \subseteq \bar{\mathbb{R}}$ messbar. Dann ist auch τ^B eine Stoppzeit, wo*

$$\tau^B := \inf\{t > \tau : X_t \in B\} \quad (\tau^B = \max(T),\ \textit{falls}\ \{t > \tau : X_t \in B\} = \emptyset)\ .$$

Beweis:

$$\{\tau^B \le t\} = \bigcup_{\substack{s\in T\\ \tau<s\le t}} \{X_s \in B\} = \bigcup_{r\in T} \bigcup_{\substack{s\in T\\ r<s\le t}} \underbrace{\{\tau = r\}}_{\in\mathcal{F}_r\subseteq\mathcal{F}_t} \cap \underbrace{\{X_s \in B\}}_{\in\mathcal{F}_s\subseteq\mathcal{F}_t} \in \mathcal{F}_t$$

□

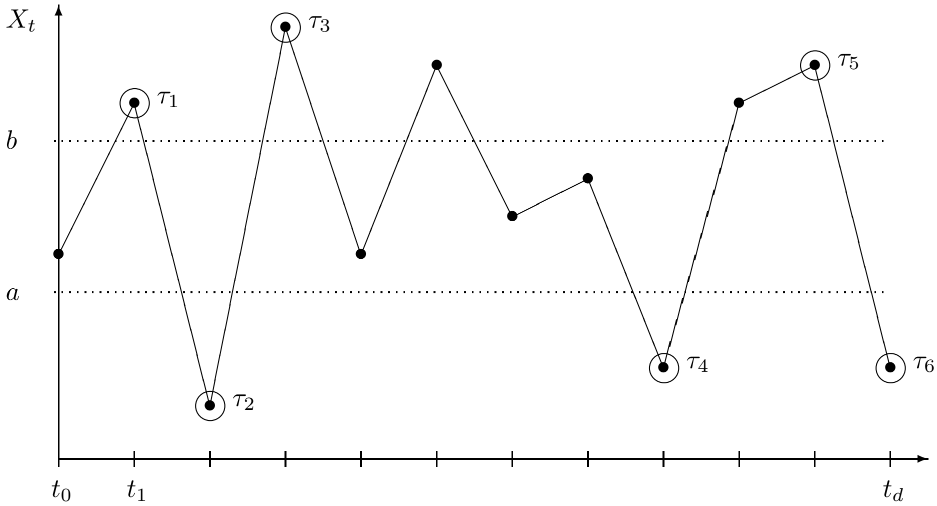
25.4 Bemerkung (Absteigende Überquerungen (“downcrossings”)) Sei $T \subseteq \mathbb{R}$ und $\mathcal{X} = (X_t)_{t\in T}$ ein stochastischer Prozess. Für $a, b \in \mathbb{R}$ mit $a < b$ und eine endliche Teilmenge $F = \{t_1 < \cdots < t_d\} \subseteq T$ definieren wir induktiv Zufallsvariablen

$$\tau_1 = \inf\{t_i : X_{t_i} > b\}, \qquad \tau_2 = \inf\{t_i > \tau_1 : X_{t_i} < a\},$$
$$\tau_{2n+1} = \inf\{t_i > \tau_{2n} : X_{t_i} > b\}, \qquad \tau_{2n+2} = \inf\{t_i > \tau_{2n+1} : X_{t_i} < a\},$$

wobei $\inf\emptyset = t_d$ gesetzt wird. Da F endlich ist, gibt es ein k_0 derart, dass $\tau_k = t_d$ für alle $k \ge k_0$. Setze

$$D(\mathcal{X}, F, [a, b]) := \sup\{n : \tau_{2n} < t_d\}\ .$$

$D(\mathcal{X}, F, [a, b])$ ist die Anzahl der Male, die der durch $(t_1, X_{t_1}), \ldots, (t_{d-1}, X_{t_{d-1}})$ definierte zufällige Polygonzug das Intervall $[a, b]$ “von oben nach unten” überspringt (“downcrossings”).



In diesem Beispiel ist $D(\mathcal{X}, F, [a, b]) = 2$ (nicht =3!).

Ist $(X_t, \mathcal{F}_t)_{t\in T}$ adaptiert, so sind die τ_k Stoppzeiten - siehe Lemma 25.3.

Schließlich setzt man

$$D(\mathcal{X}, T, [a, b]) := \sup\{D(\mathcal{X}, F, [a, b]) : F \subseteq T, F \text{ endlich}\}\ .$$

Ist T höchstens abzählbar, so gibt es endliche $D_n \nearrow T$, und es folgt, dass

$$D(\mathcal{X}, T, [a, b]) = \sup_n D(\mathcal{X}, D_n, [a, b])\ .$$

Daher ist in diesem Fall $D(\mathcal{X}, T, [a, b])$ messbar, also eine Zufallsvariable.

25.5 Lemma ("Downcrossing"-Ungleichung) *Sei $\mathcal{X} = (X_t, \mathcal{F}_t)_{t\in T}$ ein Submartingal mit höchstens abzählbarer Indexmenge T. Dann gilt für jedes Intervall $[a, b]$*

$$(b-a)\cdot E\left[D(\mathcal{X}, T, [a,b])\right] \le \sup_{t\in T} E\left[(X_t - b)^+\right] \ .$$

Beweis: Es reicht offensichtlich, diese Ungleichung für endliche T zu zeigen. Sei daher nun $T = F = \{t_1 < \cdots < t_d\}$ und die τ_k wie in Bemerkung 25.4. Setze $A_k := \{\tau_k < t_d\}$. Dann ist $A_k \in \mathcal{F}_{\tau_k}$, $A_{k+1} \subseteq A_k$ und $A_k = \emptyset$ für $k \ge k_0$. Da $X_{\tau_{2n-1}} > b$ auf A_{2n-1} und $X_{\tau_{2n}} < a$ auf A_{2n}, können wir folgendermaßen abschätzen:

$$\begin{aligned} 0 &\le \int_{A_{2n-1}} (X_{\tau_{2n-1}} - b)\,dP \overset{\text{S.24.18}}{\le} \int_{A_{2n-1}} E\left[X_{\tau_{2n}} - b \middle| \mathcal{F}_{\tau_{2n-1}}\right] dP \\ &= \int_{A_{2n-1}} (X_{\tau_{2n}} - b)\ dP \le (a-b)\,P(A_{2n}) + \int_{A_{2n-1}\setminus A_{2n}} (X_{\tau_{2n}} - b)\ dP \end{aligned}$$

Da $\tau_{2n} = t_d$ auf A_{2n}^c, folgt

$$(b-a)\,P(A_{2n}) \le \int_{A_{2n-1}\setminus A_{2n}} (X_{t_d} - b)\ dP$$

und daher

$$\begin{aligned} E\left[D(\mathcal{X}, F, [a,b])\right] &= \sum_{n=1}^{\infty} P(\underbrace{D(\mathcal{X}, F, [a,b]) \ge n}_{\Leftrightarrow\ \tau_{2n} < t_d}) = \sum_{n=1}^{\infty} P(A_{2n}) \\ &\le \frac{1}{b-a} \sum_{n=1}^{\infty} \int_{A_{2n-1}\setminus A_{2n}} (X_{t_d} - b)\ dP \\ &\le \frac{1}{b-a} \int (X_{t_d} - b)^+\ dP \ , \end{aligned}$$

da die Mengen $A_{2n-1} \setminus A_{2n}$ paarweise disjunkt sind. □

Beweis von Satz 25.1: Es ist

$$\{X_n \text{ konvergiert nicht}\} = \left\{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\right\} = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\}$$

und

$$\left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\} \subseteq \{D(\mathcal{X}, \mathbb{N}, [a,b]) = \infty\} \ .$$

Dabei ist $P(\{D(\mathcal{X}, \mathbb{N}, [a,b]) = \infty\}) = 0$, da

$$E\left[D(\mathcal{X}, \mathbb{N}, [a,b])\right] \overset{\text{L.25.5}}{\le} \frac{1}{b-a} \sup_n E\left[(X_n - a)^+\right] \le \frac{1}{b-a}\left(|a| + \sup_n E[X_n^+]\right) < \infty \ .$$

Also existiert $X_\infty = \lim_{n\to\infty} X_n$ f.s., und aus dem Lemma von Fatou folgt

$$E|X_\infty| \le \liminf_{n\to\infty} E|X_n| \le \sup_n E\big[\underbrace{2X_n^+ - X_n}_{=|X_n|}\big] \le 2\sup_n E[X_n^+] - E[X_0] < \infty \ ,$$

denn $E[X_n] \ge E[X_0]$, weil $\mathcal{X}$ ein Submartingal ist. □

25.6 Korollar *a) Ein nach oben beschränktes Submartingal $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ konvergiert fast sicher.*

b) Ein nach unten beschränktes Martingal $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ konvergiert fast sicher.

Beweis: a) folgt direkt aus Satz 25.1, und für b) beachte, dass $(-X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein nach oben beschränktes Submartingal ist. $\square$

Als erste Anwendung betrachten wir das Martingal $E[\xi|\mathcal{F}_n]$ aus Beispiel 24.5B.

25.7 Satz (Konvergenz bedingter Erwartungen)
Sei $\xi \in \mathcal{L}^1_P$, $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ eine Filtration, $\mathcal{F}_\infty = \sigma\left(\bigcup_{n\in\mathbb{N}_0} \mathcal{F}_n\right)$. Dann konvergiert $E[\xi\big|\mathcal{F}_n] \to E[\xi\big|\mathcal{F}_\infty]$ fast sicher und in $\mathcal{L}^1_P$.

Zum Beweis benötigen wir folgendes Lemma.

25.8 Lemma (Gleichgradige Integrierbarkeit bedingter Erwartungen) *Sei $\xi \in \mathcal{L}^1_P$. Dann ist die Familie*

$$\left\{X_\mathcal{G} := E[\xi\big|\mathcal{G}] : \mathcal{G} \subseteq \mathcal{F} \text{ Teil-}\sigma\text{-Algebra}\right\}$$

gleichgradig integrierbar.

Beweis: Sei $\mathcal{G}$ eine Teil-σ-Algebra von $\mathcal{F}$ und $a > 0$.

$$\begin{aligned}
\int_{\{|X_\mathcal{G}|\geq a^2\}} |X_\mathcal{G}|\,dP &\overset{S.19.6d}{\leq} \int_{\{|X_\mathcal{G}|\geq a^2\}} E\left[|\xi|\big|\mathcal{G}\right]\,dP = \int_{\{|X_\mathcal{G}|\geq a^2\}} |\xi|\,dP \\
&= \int_{\{|X_\mathcal{G}|\geq a^2\}\cap\{|\xi|\leq a\}} |\xi|\,dP + \int_{\{|X_\mathcal{G}|\geq a^2\}\cap\{|\xi|>a\}} |\xi|\,dP \\
&\leq \underbrace{a\cdot P\{|X_\mathcal{G}|\geq a^2\}}_{\leq \frac{a}{a^2}E|X_\mathcal{G}|} + \int_{\{|\xi|>a\}} |\xi|\,dP \\
&\leq \frac{1}{a}E|\xi| + \int_{\{|\xi|>a\}} |\xi|\,dP \\
&\to 0 \quad (a\to\infty)
\end{aligned}$$

gleichmäßig in $\mathcal{G}$. $\square$

Beweis von Satz 25.7: Sei $X_n := E[\xi\big|\mathcal{F}_n]$. Nach Beispiel 24.5B ist $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Martingal, und wegen des vorhergehenden Lemmas sind die X_n gleichgradig integrierbar. Insbesondere ist auch $\sup_n E\left[X_n^+\right] < \infty$. Daher kann man den Martingal-Konvergenzsatz 25.1 anwenden, so dass $X_\infty := \lim_{n\to\infty} X_n$ fast sicher existiert. Setze $X_\infty = 0$, wo dieser Limes nicht existiert. Dann ist X_∞ bzgl. $\mathcal{F}_\infty$ messbar. Die $\mathcal{L}^1_P$-Konvergenz folgt aus Satz 8.20. Bleibt zu zeigen: $X_\infty = E[\xi\big|\mathcal{F}_\infty]$. Für $A \in \mathcal{F}_n$ und $m \geq n \geq 0$ ist

$$\int_A X_m\,dP = \int_A E[\xi\big|\mathcal{F}_m]\,dP = \int_A \xi\,dP = \int_A E[\xi\big|\mathcal{F}_\infty]\,dP\,,$$

also

$$\int_A X_\infty\,dP = \lim_{m\to\infty}\int_A X_m\,dP = \int_A E[\xi\big|\mathcal{F}_\infty]\,dP\,,$$

und da $\mathcal{F}_\infty = \sigma\left(\bigcup_{n\in\mathbb{N}_0}\mathcal{F}_n\right)$, folgt $X_\infty = E[\xi\big|\mathcal{F}_\infty]$ aus Lemma 19.4c. $\square$

Der Martingal-Konvergenzsatz garantiert, dass der fast sichere Grenzwert $X_\infty = \lim_{n\to\infty} X_n$ integrierbar ist. Trotzdem liegt, anders als im gerade bewiesenen Satz, im Allgemeinen keine $\mathcal{L}^1$-Konvergenz vor, wie schon Beispiel 25.2 zeigte. Die Frage, wann Submartingale in $\mathcal{L}^1$ konvergieren, klärt der folgende Satz. Seine Aussage iii) zeigt, dass in Satz 25.7 schon der allgemeine Fall eines $\mathcal{L}^1$-konvergenten Martingals behandelt worden ist.

25.9 Satz ($\mathcal{L}^1$-Konvergenz und gleichgradige Integrierbarkeit)
a) Sei $\mathcal{X} = (X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Martingal und gelte $\sup_{n\in\mathbb{N}_0} E\left[X_n^+\right] < +\infty$. Seien X_∞ und $\mathcal{F}_\infty$ wie in Satz 25.1. Dann sind folgende Aussagen äquivalent:

i) $(X_n)_{n\in\mathbb{N}_0}$ ist gleichgradig integrierbar.

ii) $\lim_{n\to\infty} E|X_n - X_\infty| = 0$.

iii) $\bar{\mathcal{X}} := (X_n, \mathcal{F}_n)_{0\le n\le\infty}$ ist ein Martingal.

b) Ist $\mathcal{X}$ nur ein Submartingal, so gilt i) $\Leftrightarrow$ ii) $\Rightarrow$ "$\bar{\mathcal{X}}$ ist ein Submartingal".

Beweis: i) $\Leftrightarrow$ ii): Folgt aus Satz 8.20.
ii) $\Rightarrow$ iii): Aus Satz 25.1 folgt, dass $E|X_\infty| < \infty$ und dass X_∞ bzgl. $\mathcal{F}_\infty$ messbar ist. Außerdem ist für jedes $k \in \mathbb{N}_0$

$$E\left[\left(E[X_\infty|\mathcal{F}_k] - X_k\right)^-\right] = E\left[\lim_{n\to\infty}\left(E[X_n|\mathcal{F}_k] - X_k\right)^-\right] \overset{\text{Fatou}}{\le} \liminf_{n\to\infty} E\left[\left(E[X_n|\mathcal{F}_k] - X_k\right)^-\right]$$

und $\left(E[X_n|\mathcal{F}_k] - X_k\right)^- = 0$ $(n \ge k)$ gilt sogar für Submartingale. Also ist $E[X_\infty|\mathcal{F}_k] \ge X_k$, und für Martingale folgt die umgekehrte Ungleichung durch Betrachtung des Submartingals $-\mathcal{X}$.
iii) $\Rightarrow$ ii): Da $X_\infty \in \mathcal{L}^1_P$ und $X_n = E[X_\infty|\mathcal{F}_n]$ nach Voraussetzung, folgt das aus Satz 25.7. □

25.10 Korollar *Jedes gleichgradig integrierbare Submartingal $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ konvergiert fast sicher und in $\mathcal{L}^1_P$.*

25.11 Bemerkung Ist $\sup_n E[|X_n|^p] < \infty$ für ein $p > 1$, so ist $(X_n)_{n\in\mathbb{N}_0}$ gleichgradig integrierbar, denn $\sup_n \int_{\{|X_n|>a\}} |X_n|\, dP \le a^{-(p-1)} \sup_n \int |X_n|^p\, dP \to 0$ mit $a \to \infty$.

Dieser Satz hilft uns, einen erweiterten Satz über optionales Stoppen zu beweisen. Anders als in Satz 24.18 müssen die Stoppzeiten nicht mehr beschränkt sein.

25.12 Satz (Optionales Stoppen)
Seien $\sigma \le \tau$ fast sicher endliche Stoppzeiten bzgl. $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ und sei $\mathcal{X} = (X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein gleichgradig integrierbares Submartingal. Dann ist $X_\sigma \le E[X_\tau|\mathcal{F}_\sigma]$ mit Gleichheit, falls $\mathcal{X}$ ein Martingal ist.

25.13 Bemerkung Ohne die Annahme der gleichgradigen Integrierbarkeit wäre dieser Satz falsch. Betrachte z.B. das Martingal $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$, wo $X_n := \prod_{i=1}^n \xi_i$, $\mathcal{F}_n := \sigma(\xi_1, \ldots, \xi_n)$ mit unabhängigen, identisch verteilten ξ_i, $P\{\xi_i = 0\} = P\{\xi_i = 2\} = \frac{1}{2}$. Dann ist $\tau := \inf\{n > 0 : X_n = 0\} \ge 1$ eine fast sicher endliche Stoppzeit, aber $E[X_\tau] = 0 \ne 1 = E[X_1]$.

Beweis von Satz 25.12: Wir beweisen den Satz hier nur für Martingale. Einen Beweis für Submartingale findet man z.B. in [Širjaev, Abschnitt VII.2.1].

Da $\mathcal{X}$ gleichgradig integrierbar ist, ist $X_n = E[X_\infty|\mathcal{F}_n]$ (Satz 25.9). Sei $Y_n := E[X_\infty|\mathcal{F}_{\tau\wedge n}]$. Dann ist auch $(Y_n)_{n\in\mathbb{N}_0}$ gleichgradig integrierbar (Lemma 25.8) und $Y_n = X_{\tau\wedge n}$, denn für $A \in \mathcal{F}_{\tau\wedge n}$ ist

$$\int_A X_{\tau\wedge n}\, dP = \sum_{k=0}^n \int_{\underbrace{A\cap\{\tau\wedge n = k\}}_{\in\mathcal{F}_k}} X_k\, dP = \sum_{k=0}^n \int_{A\cap\{\tau\wedge n=k\}} X_\infty\, dP = \int_A X_\infty\, dP\,.$$

Da $X_{\tau\wedge n} \to X_\tau$ f.s. und da $X_{\tau\wedge n} = Y_n$ f.s. und in $\mathcal{L}^1$ konvergiert (Satz 25.7), konvergiert auch $X_{\tau\wedge n} \to X_\tau$ in $\mathcal{L}^1$.

Sei nun $A \in \mathcal{F}_\sigma$ und $A_m := A \cap \{\sigma \le m\}$. Dann ist $A_m \in \mathcal{F}_{\sigma \wedge m}$, denn für $n \ge 0$ ist

$$A_m \cap \{\sigma \wedge m \le n\} = \begin{cases} A \cap \{\sigma \le m\} \in \mathcal{F}_m \subseteq \mathcal{F}_n, & \text{falls } n \ge m \\ A \cap \{\sigma \le m\} \cap \{\sigma \le n\} = A \cap \{\sigma \le n\} \in \mathcal{F}_n, & \text{falls } n < m. \end{cases}$$

Es folgt

$$\int_{A_m} X_\tau \, dP = \lim_{n\to\infty} \int_{A_m} X_{\tau \wedge n} \, dP \overset{\text{S.24.18}}{=} \lim_{n\to\infty} \int_{A_m} X_{\sigma\wedge n} \, dP = \int_{A_m} X_\sigma \, dP$$

und da $\lim_{m\to\infty} P(A \setminus A_m) = 0$, folgt $\int_A X_\sigma \, dP = \int_A X_\tau \, dP$, also $X_\sigma = E[X_\tau \big| \mathcal{F}_\sigma]$. □

Eine andere Verallgemeinerung des optionalen Stoppens mit beschränkten Stoppzeiten ist die *Wald'sche Gleichung*.

25.14 Satz (Wald'sche Gleichung)
Seien $\xi_1, \xi_2, \ldots$ unabhängige identisch verteilte integrierbare Zufallsvariablen, $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ und τ eine Stoppzeit zur Filtration $(\mathcal{F}_n)_{n\ge 0}$. Ist $E[\tau] < \infty$, so folgt

$$E\left[\xi_1 + \cdots + \xi_\tau\right] = E[\tau] \cdot E[\xi_1] \, .$$

Beweis: Sei $X_n := \xi_1 + \cdots + \xi_n - nE[\xi_1]$. Dann ist $(X_n, \mathcal{F}_n)_{n\ge 0}$ ein Martingal (Beispiel 24.5A). Aus Satz 24.18 folgt $EX_{\tau\wedge n} = EX_1 = 0$, also

$$E\left[\xi_1 + \cdots + \xi_{\tau\wedge n}\right] = E[\tau \wedge n] \cdot E[\xi_1] \qquad (*)$$

Sind die $\xi_i \ge 0$, so folgt aus dem Satz von der monotonen Konvergenz

$$E\left[\xi_1 + \cdots + \xi_\tau\right] = \sup_n E\left[\xi_1 + \cdots + \xi_{\tau\wedge n}\right] = \sup_n E[\tau \wedge n] \cdot E[\xi_1] = E[\tau] \cdot E[\xi_1] < \infty \, .$$

Für allgemeine ξ_i folgt insbesondere, dass $|\xi_1 + \cdots + \xi_{\tau\wedge n}| \le |\xi_1| + \cdots + |\xi_\tau| \in \mathcal{L}^1_P$ für alle n, so dass aus (*) und dem Satz von der majorisierten Konvergenz folgt

$$E\left[\xi_1 + \cdots + \xi_\tau\right] = \lim_{n\to\infty} E\left[\xi_1 + \cdots + \xi_{\tau\wedge n}\right] = \lim_{n\to\infty} E[\tau \wedge n] \cdot E[\xi_1] = E[\tau] \cdot E[\xi_1] \, .$$

□

Bevor wir dieses Kapitel mit einer Anwendung der "Downcrossing"-Ungleichung auf die Regularisierung von Pfaden von Submartingalen mit kontinuierlicher Indexmenge beschließen, beweisen wir noch eine Variante des Martingal-Konvergenzsatzes für Rückwärtsmartingale.

25.15 Definition (Rückwärtsmartingal) $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ *ist ein* Rückwärtsmartingal*, falls der adaptierte Prozess $(X_{-n}, \mathcal{F}_{-n})_{n\in -\mathbb{N}_0}$ ein Martingal ist. (Insbesondere: $E[X_n \big| \mathcal{F}_{n+1}] = X_{n+1}$.)*

25.16 Bemerkung (Gleichgradige Integrierbarkeit von Rückwärtsmartingalen) Ist $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Rückwärtsmartingal, so ist nach Definition $X_n = E[X_0 \big| \mathcal{F}_n]$. Aus Lemma 25.8 folgt, dass $(X_n)_{n\in\mathbb{N}_0}$ gleichgradig integrierbar ist.

25.17 Satz (Konvergenzsatz für Rückwärtsmartingale)
Ist $\mathcal{X} = (X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ ein Rückwärtsmartingal, so existiert $X_\infty := \lim_{n\to\infty} X_n$ P-f.s. und in $\mathcal{L}^1_P$. X_∞ hat eine Version, die bzgl. $\mathcal{F}_{-\infty} := \bigcap_{n\in\mathbb{N}_0} \mathcal{F}_n$ messbar ist.

Beweis: Wegen der gleichgradigen Integrierbarkeit (Bemerkung 25.16) ist nur die fast sichere Konvergenz zu zeigen. Betrachte dazu für $N > 0$ die Martingale $\mathcal{Y}^{(N)} = (X_{N-n}, \mathcal{F}_{N-n})_{0\le n\le N}$. Bezeichnet

$U(\mathcal{Y}^{(N)}, \{0,\dots,N\}, [a,b])$ die Anzahl der “Upcrossings” des Martingals $\mathcal{Y}^{(N)}$ über $[a,b]$ im Intervall $\{0,\dots,N\}$, so ist

$$D(\mathcal{X}, \{0,\dots,N\}, [a,b]) = U(\mathcal{Y}^{(N)}, \{0,\dots,N\}, [a,b]) = D(-\mathcal{Y}^{(N)}, \{0,\dots,N\}, [-b,-a]) \ ,$$

also nach Lemma 25.5 (und Bemerkung 25.16)

$$\begin{aligned} E\left[D(\mathcal{X}, \mathbb{N}_0, [a,b])\right] &= \sup_N E\left[D(\mathcal{X}, \{0,\dots,N\}, [a,b])\right] \\ &\le \frac{1}{-a-(-b)} \sup_N \sup_{0\le n\le N} E\left[(-X_{N-n}+a)^+\right] \\ &\le \frac{1}{b-a}\left(\sup_n E|X_n| + |a|\right) \\ &\le \frac{E|X_0|+|a|}{b-a} < \infty \ . \end{aligned}$$

Daraus folgt die fast sichere Konvergenz wie im Beweis des Martingal-Konvergenzsatzes 25.1. Die Messbarkeitsaussage ist dann sofort klar. □

Im nächsten Kapitel werden wir diesen Satz bei der Untersuchung verschiedener 0-1-*Gesetze* anwenden. Dieses Kapitel schließen wir mit folgendem Satz ab, der nur exemplarisch darstellen soll, wie Martingaltechniken auch bei Submartingalen mit kontinuierlicher Indexmenge angewandt werden können.

25.18 Satz (Regularisierung von Pfaden von Submartingalen)
Sei $\mathcal{X} = (X_t, \mathcal{F}_t)_{t\in\mathbb{R}_+}$ ein Submartingal. Dann gilt für fast jedes ω:

$$\lim_{\substack{r\uparrow t\\ r\in\mathbb{Q}}} X_r(\omega) =: X_t^l(\omega) \quad \textit{existiert für jedes } t\in(0,\infty)\ ,$$

$$\lim_{\substack{s\downarrow t\\ s\in\mathbb{Q}}} X_s(\omega) =: X_t^r(\omega) \quad \textit{existiert für jedes } t\in[0,\infty)\ .$$

Nach Konstruktion sind für fast alle ω die Pfade $t\mapsto X_t^r(\omega)$ rechtsseitig und die Pfade $t\mapsto X_t^l(\omega)$ linksseitig stetig auf ganz $\mathbb{R}_+$.

Beweis: Es reicht, den Satz für $t\in I_N := [0,N]$, $N\in\mathbb{N}$, zu beweisen, da $\mathbb{R}_+ = \bigcup_N I_N$. Für jedes $t\in I_N$ ist

$$E\left[(X_t-b)^+\right] \le E[X_N^+] + |b| < \infty\ ,$$

so dass, wie im Beweis des Martingal-Konvergenzsatzes aus der “Downcrossing”-Ungleichung folgt: Es gibt eine Menge $\Omega_0\subseteq\Omega$ mit $P(\Omega_0)=1$ derart, dass für alle $\omega\in\Omega_0$ und alle $a,b\in\mathbb{Q}$ mit $a<b$

$$D(\mathcal{X}, I_N\cap\mathbb{Q}, [a,b])(\omega) < \infty\ . \tag{*}$$

Gäbe es nun zu einem $\omega\in\Omega_0$ ein $t\in I_N$ für das einer der beiden einseitigen Limiten nicht existiert, so gäbe es auch wieder ein Intervall $[a,b]$, das unendlich oft übersprungen wird, im Widerspruch zu (*). □

25.19 Bemerkung Die Ausnahmemenge $\Omega\setminus\Omega_0$ ist $\mathcal{F}_N$-messbar, aber nicht notwendig $\mathcal{F}_t$-messbar für $t<N$. Definiert man nun

$$\tilde{X}_t(\omega) := \begin{cases} X_t^r(\omega) & (\omega\in\Omega_0) \\ 0 & \text{sonst,} \end{cases}$$

so ist $\tilde{X}_t$ nicht notwendig $\mathcal{F}_{t+\epsilon}$-messbar für $\epsilon>0$. Setzt man allerdings voraus, dass die Filtration $(\mathcal{F}_t)_{t\ge0}$ *vollständig* ist (d.h. $A\subseteq B\in\mathcal{F}_t$ und $P(B)=0$ impliziert $A\in\mathcal{F}_0$), so ist $\tilde{X}_t$ messbar bzgl. $\bigcap_{s>t}\mathcal{F}_s$, denn dann ist $\Omega_0\in\mathcal{F}_s$ für alle s und auf dieser Menge ist $\lim_{s\downarrow t, s\in\mathbb{Q}} X_s$ messbar bzgl. $\bigcap_{s>t}\mathcal{F}_s$.

25.20 Korollar *Sei* $\mathcal{X} = (X_t, \mathcal{F}_t)_{t\in\mathbb{R}_+}$ *ein Martingal.*

a) Für jedes $t \in [0,\infty)$ *ist* $E|X_t^r| < \infty$ *und* $X_t = E\left[X_t^r \middle| \mathcal{F}_t\right]$*.*

b) Ist die Filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ *vollständig und* rechtsseitig stetig*, d.h. ist* $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$*, so ist* $\tilde{\mathcal{X}} := (\tilde{X}_t, \mathcal{F}_t)_{t\in\mathbb{R}_+}$ *ein Martingal. Außerdem ist in diesem Fall* $\tilde{\mathcal{X}}$ *eine Modifikation von* $\mathcal{X}$ *(siehe Definition 22.7) mit rechtsseitig stetigen Pfaden.*

Beweis:

a) Sei $t \in \mathbb{R}$, $t_n \searrow t$, $t_n \in \mathbb{Q}$. Dann ist $(X_{t_n}, \mathcal{F}_{t_n})_{n\in\mathbb{N}_0}$ ein Rückwärtsmartingal und $X_{t_n} \to X_t^r$ in $\mathcal{L}^1$. Insbesondere ist $X_t^r \in \mathcal{L}_P^1$ und

$$E\left[X_t^r \middle| \mathcal{F}_t\right] = \lim_{n\to\infty} \underbrace{E\left[X_{t_n} \middle| \mathcal{F}_t\right]}_{=X_t \text{ für alle } n} = X_t\ . \tag{**}$$

b) Wegen Bemerkung 25.19 ist $\tilde{X}_t$ messbar bzgl. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$. Insbesondere ist $E[X_t^r \big| \mathcal{F}_t] = E[\tilde{X}_t \big| \mathcal{F}_t] = \tilde{X}_t$, und aus (**) folgt $\tilde{X}_t = X_t$ f.s., d.h. $\tilde{\mathcal{X}}$ ist eine Modifikation von $\mathcal{X}$. Schließlich ist für $u < t$

$$E\left[\tilde{X}_t \middle| \mathcal{F}_u\right] = E\left[X_t \middle| \mathcal{F}_u\right] = X_u = \tilde{X}_u \quad f.s.$$

□

25.21 Bemerkung Wendet man in der Situation von Korollar 25.20b den Satz 25.18 auf das rechtsseitig stetige Martingal $\tilde{\mathcal{X}}$ an, so sieht man, dass $\tilde{\mathcal{X}}$ auch linksseitige Limiten besitzt. Man sagt: $\tilde{\mathcal{X}}$ ist *cadlag* (continue à droite, limites à gauche). *Beweis:* Sei

$$\Omega_1 := \left\{ \omega : \tilde{X}_t^l(\omega) = \lim_{\substack{r\uparrow t\\ r\in\mathbb{Q}}} \tilde{X}_r(\omega) \text{ existiert für jedes } t \in (0,\infty) \right\}\ .$$

Zu $\omega \in \Omega_1$, $t > 0$ und $\epsilon > 0$ gibt es ein $\delta > 0$ derart, dass $|\tilde{X}_t^l(\omega) - \tilde{X}_r(\omega)| < \epsilon$ für $r \in \mathbb{Q} \cap (t-\delta, t)$. Sei nun $t' \in (t-\delta, t)$ beliebig, $r_n \downarrow t'$, $r_n \in \mathbb{Q} \cap (t', t)$. Dann ist

$$\left|\tilde{X}_t^l(\omega) - \tilde{X}_{t'}(\omega)\right| \le \underbrace{\lim_{n\to\infty} \left|\tilde{X}_t^l(\omega) - \tilde{X}_{r_n}(\omega)\right|}_{=0} + \underbrace{\limsup_{n\to\infty} \left|\tilde{X}_{r_n}(\omega) - \tilde{X}_{t'}(\omega)\right|}_{\le\epsilon}$$

25.22 Bemerkung Vergleichbare Aussagen zur Regularisierung kann man mit nur wenig mehr Aufwand auch für Submartingale beweisen, siehe z.B. [Revuz-Yor, Abschnitt II.2].

Kapitel 26

0-1-Gesetze

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, T eine Indexmenge und $\mathcal{X} = (X_t)_{t\in T}$ ein stochastischer Prozess mit Verteilung $P_{\mathcal{X}}$ auf $(\mathbb{R}^T, \mathcal{B}^T)$. Unter einem 0-1-*Gesetz* für $\mathcal{X}$ versteht man eine Aussage der Art "Eine σ-Algebra $\mathcal{G} \subseteq \mathcal{B}^T$ enthält nur Mengen vom $P_{\mathcal{X}}$-Maß 0 oder 1" (oder auch "Eine σ-Algebra $\mathcal{A}_0 \subseteq \mathcal{A}$ enthält nur Mengen vom P-Maß 0 oder 1"). Ist z.B. $\mathcal{X} = (X_n)_{n\in\mathbb{N}_0}$ ein stationärer Prozess und ist $\mathcal{G} = \mathcal{I}$ die σ-Algebra der translationsinvarianten Ereignisse in $\mathcal{B}^{\mathbb{N}_0}$ (vergl. Lemma 19.14), so kann die Aussage "$\mathcal{X}$ ist ergodisch" als 0-1-Gesetz für $\mathcal{X}$ interpretiert werden. Normalerweise denkt man bei einem 0-1-Gesetz jedoch an Aussagen, die die *σ-Algebra im Unendlichen* oder die *σ-Algebra der austauschbaren Ereignisse* betrifft. Mit diesen beiden Fällen wollen wir uns in diesem Kapitel beschäftigen. Dabei spielen die Martingal-Konvergenzsätze des letzten Kapitels eine wichtige Rolle.

Für dieses Kapitel vereinbaren wir:

- ▷ $\mathcal{X} = (X_t)_{t\in T}$ sei ein stochastischer Prozess mit <u>höchstens abzählbarer</u> Indexmenge T.
- ▷ Für $S \subseteq T$ sei $\mathcal{F}_S := \sigma(X_t : t \in S)$, siehe Definition 5.9.
- ▷ Es sei $\mathcal{E}(T) := \{\Lambda \subseteq T : \Lambda \text{ endlich}\}$ wie in Definition 12.1.
- ▷ Erinnerung: Eine Teil-σ-Algebra eines Wahrscheinlichkeitsraumes heißt *trivial*, wenn sie nur Mengen vom Maß 0 oder 1 enthält.

26.1 Definition (Asymptotische σ-Algebra) $\quad \mathcal{A}_\infty := \bigcap_{\Lambda\in\mathcal{E}(T)} \mathcal{F}_{\Lambda^c}$ *heißt <u>asymptotische σ-Algebra</u> oder auch <u>σ-Algebra im Unendlichen</u> (und englisch auch* tail field*) des Prozesses $(X_t)_{t\in T}$.*

26.2 Bemerkung Da T höchstens abzählbar ist, gilt für jede aufsteigende Folge $\Lambda_n \nearrow T$, $\Lambda_n \in \mathcal{E}(T)$: $\mathcal{A}_\infty = \bigcap_{n\in\mathbb{N}} \mathcal{F}_{\Lambda_n^c}$. Eine solche Folge gibt es immer, und wenn nichts anderes vermerkt wird, bezeichnet $(\Lambda_n)_n$ eine beliebige solche Folge. Beachte auch dass $\mathcal{A}_\infty \subseteq \mathcal{F}_T = \sigma\left(\bigcup_{n>0} \mathcal{F}_{\Lambda_n}\right)$.

26.3 Satz (0-1-Gesetz von Kolmogorov)
Sind die X_t, $t \in T$, unabhängig, so ist $\mathcal{A}_\infty$ trivial.

Beweis: Sei $A \in \mathcal{A}_\infty$. Da A für jedes $n > 0$ von $\mathcal{F}_{\Lambda_n}$ unabhängig ist, folgt aus Beispiel 19.5c und Satz 25.7

$$P(A) = P\left[A \middle| \mathcal{F}_{\Lambda_n}\right] \to P\left[A \middle| \mathcal{F}_T\right] = 1_A \quad (n \to \infty)$$

P-fast sicher, also $P(A) \in \{0, 1\}$. □

Hier ist ein einfaches Beispiel für einen Prozess, dessen asymptotische σ-Algebra nicht trivial ist.

26.4 Beispiel Sei ν ein Wahrscheinlichkeitsmaß auf $[0,1]$. Für $p \in [0,1]$ bezeichne P_p das Bernoulli-Maß zum Parameter p auf $\Omega := \{0,1\}^{\mathbb{N}}$. Dann wird durch

$$P(A) := \int_0^1 P_p(A)\, d\nu(p)$$

ein Wahrscheinlichkeitsmaß auf Ω definiert. Wie üblich sei $X_n(\omega) = \omega_n$.

Betrachte $h : \Omega \to [0,1]$, $h = \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k$. Da $h = \limsup_{n\to\infty} \frac{1}{n}\sum_{k=m}^n X_k$ für jedes $m > 0$, ist h bzgl. $\mathcal{A}_\infty$ messbar, d.h. $h^{-1}(B) \in \mathcal{A}_\infty$ für jedes $B \in \mathcal{B}([0,1])$. Es gilt sogar:

$$P(h^{-1}B) = \int_0^1 P_p\left\{\limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k \in B\right\} d\nu(p) = \int_0^1 1_B(p)\, d\nu(p) = \nu(B)\ .$$

Wir zeigen noch mehr:

> Der Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}_\infty, P)$ ist zu $([0,1], \mathcal{B}([0,1]), \nu)$ isomorph in dem Sinn dass $B \mapsto h^{-1}B$ eine maßtreue Bijektion zwischen $\mathcal{B}([0,1])$ und $\mathcal{A}_\infty$ (modulo Nullmengen) ist.

Dazu bleibt noch zu zeigen, dass es zu jedem $A \in \mathcal{A}_\infty$ ein $U_A \in \mathcal{B}([0,1])$ mit $P(A \triangle h^{-1}U_A) = 0$ gibt: Beachte zunächst, dass $P_p(A) \in \{0,1\}$ für jedes p (Satz 26.3). Da $p \mapsto P_p(A)$ für jedes messbare A messbar ist (siehe Bemerkung 11.4), ist $U_A := \{p \in [0,1] : P_p(A) = 1\} \in \mathcal{B}([0,1])$, und es folgt

$$\begin{aligned}
P_p(h^{-1}U_A) = 1 &\iff P_p\left\{\omega : \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \omega_k \in U_A\right\} = 1\\
&\iff p = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \omega_k \in U_A \quad P_p\text{-f.s.}\\
&\iff p \in U_A\\
&\iff P_p(A) = 1\ .
\end{aligned}$$

Da auch $h^{-1}U_A \in \mathcal{A}_\infty$, folgt $P_p(A) = P_p(h^{-1}U_A) \in \{0,1\}$, also $P_p(A \triangle h^{-1}U_A) = 0$ für jedes p. Insbesondere ist $P(A \triangle h^{-1}U_A) = 0$.

Wir halten noch die folgende Eigenschaft von P fest: Seien $a_1, \dots, a_n \in \{0,1\}$, $(\pi(1), \dots, \pi(n))$ eine Permutation von $(1, \dots, n)$. Dann ist

$$\begin{aligned}
P\left(X_1 = a_1, \dots, X_n = a_n\right) &= \int_0^1 P_p\left(X_1 = a_1, \dots, X_n = a_n\right)\, d\nu(p)\\
&= \int_0^1 P_p\left(X_{\pi(1)} = a_1, \dots, X_{\pi(n)} = a_n\right)\, d\nu(p)\\
&= P\left(X_{\pi(1)} = a_1, \dots, X_{\pi(n)} = a_n\right)\ ,
\end{aligned}$$

d.h. die Verteilung von $(X_n)_{n\in\mathbb{N}}$ ist unter beliebigen Permutationen endlich vieler Koordinaten invariant.

Die zuletzt beobachtete Eigenschaft motiviert folgende Definition:

26.5 Definition (Vertauschbarkeit) *Der stochastische Prozess $(X_t)_{t\in T}$ hat eine <u>vertauschbare Verteilung</u> (oder kurz: ist <u>vertauschbar</u>), falls für jedes $\Lambda \in \mathcal{E}(T)$ und jede Permutation π von Λ gilt: Die zufälligen Vektoren $(X_t)_{t\in\Lambda}$ und $(X_{\pi(t)})_{t\in\Lambda}$ haben die gleiche Verteilung.*

26.6 Definition (Bedingte Unabhängigkeit, bedingt identische Verteilung) *Sei $\mathcal{F}$ eine Teil-σ-Algebra von $\mathcal{A}$.*

a) Der stochastische Prozess $(X_t)_{t\in T}$ ist bedingt unabhängig gegeben $\mathcal{F}$, falls für jedes $\Lambda \in \mathcal{E}(T)$ und jede Familie $A_t \in \mathcal{B}(\mathbb{R})$ $(t \in \Lambda)$ gilt:

$$P\left[X_t \in A_t\,(t \in \Lambda)\middle|\mathcal{F}\right] = \prod_{t\in\Lambda} P\left[X_t \in A_t\middle|\mathcal{F}\right] \ . \qquad (*)$$

b) Die Zufallsvariablen X_t, $t \in T$, sind bedingt identisch verteilt gegeben $\mathcal{F}$, falls $P\left[X_s \in A\middle|\mathcal{F}\right] = P\left[X_t \in A\middle|\mathcal{F}\right]$ für alle $s, t \in T$ und $A \in \mathcal{B}(\mathbb{R})$.

Die Struktur vertauschbarer Verteilungen beschreibt der folgende Satz

26.7 Satz (de Finetti – über die Struktur vertauschbarer Verteilungen)
Sei $(X_t)_{t\in T}$ ein stochastischer Prozess. Die folgenden Aussagen sind äquivalent:

i) $(X_t)_{t\in T}$ ist vertauschbar.

ii) Es gibt eine σ-Algebra $\mathcal{F} \subseteq \mathcal{A}$ derart, dass $(X_t)_{t\in T}$ - bedingt auf $\mathcal{F}$ - unabhängig und identisch verteilt ist.

iii) $(X_t)_{t\in T}$ ist - bedingt auf $\mathcal{A}_\infty$ - unabhängig und identisch verteilt.

Beweis: iii) $\Rightarrow$ ii) Trivial.

ii) $\Rightarrow$ i) Der Beweis ist eine einfache Verallgemeinerung der Rechnung am Ende von Beispiel 26.4. Für $\Lambda \in \mathcal{E}(T)$, $A_t \in \mathcal{B}(\mathbb{R})$ $(t \in \Lambda)$ und eine Permutation π von Λ ist

$$\begin{aligned} P\left(X_{\pi(t)} \in A_t\,(t \in \Lambda)\right) &= E\left[P\left[X_{\pi(t)} \in A_t\,(t \in \Lambda)\right)\middle|\mathcal{F}\right]\right] = E\left[\prod_{t\in\Lambda} P\left[X_{\pi(t)} \in A_t\middle|\mathcal{F}\right]\right] \\ &= E\left[\prod_{t\in\Lambda} P\left[X_t \in A_t\middle|\mathcal{F}\right]\right] = E\left[P\left[X_t \in A_t\,(t \in \Lambda)\middle|\mathcal{F}\right]\right] \\ &= P\left(X_t \in A_t\,(t \in \Lambda)\right) \end{aligned}$$

i) $\Rightarrow$ iii) Wir halten eine Folge von Mengen $\Lambda_n \in \mathcal{E}(T)$ mit $\Lambda_n \nearrow T$ fest. Zu beliebigem $\Lambda \in \mathcal{E}(T)$ gibt es ein n_0, so dass $\Lambda \subseteq \Lambda_n$ für $n \geq n_0$. Sei π eine Permutation von Λ. Seien $A_t \in \mathcal{B}(\mathbb{R})$ für $t \in \Lambda$. Für $B \in \mathcal{F}_{\Lambda_m\setminus\Lambda_n}$ und $m \geq n \geq n_0$ folgt dann aus der Vertauschbarkeit

$$P\left(X_t \in A_t\,(t \in \Lambda), (X_t)_{t\in\Lambda_m\setminus\Lambda_n} \in B\right) = P\left(X_{\pi(t)} \in A_t\,(t \in \Lambda), (X_t)_{t\in\Lambda_m\setminus\Lambda_n} \in B\right)$$

und daher

$$P\left[X_t \in A_t\,(t \in \Lambda)\middle|\mathcal{F}_{\Lambda_m\setminus\Lambda_n}\right] = P\left[X_{\pi(t)} \in A_t\,(t \in \Lambda)\middle|\mathcal{F}_{\Lambda_m\setminus\Lambda_n}\right] \ .$$

Im Limes $m \to \infty$ folgt aus Satz 25.7

$$P\left[X_t \in A_t\,(t \in \Lambda)\middle|\mathcal{F}_{\Lambda_n^c}\right] = P\left[X_{\pi(t)} \in A_t\,(t \in \Lambda)\middle|\mathcal{F}_{\Lambda_n^c}\right] \ ,$$

und im Limes $n \to \infty$ aus Satz 25.17

$$P\left[X_t \in A_t\,(t \in \Lambda)\middle|\mathcal{A}_\infty\right] = P\left[X_{\pi(t)} \in A_t\,(t \in \Lambda)\middle|\mathcal{A}_\infty\right] \ . \qquad (\#)$$

Insbesondere folgt daraus, dass die X_t - bedingt auf $\mathcal{A}_\infty$ - identisch verteilt sind.

Sei nun $\Lambda = \{t_1, \ldots, t_k\}$. Durch Induktion nach k zeigen wir (*) für $\mathcal{F} = \mathcal{A}_\infty$: Für $k = 1$ ist das trivial. Sei nun $k > 1$. Wir betrachten $A_{t_1} \in \mathcal{B}(\mathbb{R})$ und wählen n so groß dass $\Lambda \subseteq \Lambda_n$. Seien $t'_2, \ldots, t'_k$ verschiedene Elemente von Λ_n^c. Zur Abkürzung führen wir noch ein:

$$\zeta_n := P\left[X_{t_1} \in A_{t_1} \middle| \mathcal{F}_{\Lambda_n^c}\right] - P\left[X_{t_1} \in A_{t_1} \middle| \mathcal{A}_\infty\right] \ .$$

Wegen Satz 25.17 geht $\zeta_n \to 0$ f.s. Dann folgt aus (#), angewandt auf die Indexmenge $\Lambda \cup \{t'_2, \ldots, t'_k\}$:

$$\begin{aligned} & P\left[X_{t_i} \in A_{t_i}\ (i = 1, \ldots, k) \middle| \mathcal{A}_\infty\right] \\ \overset{(\#)}{=} & P\left[X_{t_1} \in A_1, X_{t'_i} \in A_{t_i}\ (i = 2, \ldots, k) \middle| \mathcal{A}_\infty\right] \\ = & E\left[P\left[X_{t_1} \in A_1 \middle| \mathcal{F}_{\Lambda_n^c}\right] 1_{A_{t_2}} \circ X_{t'_2} \ldots 1_{A_{t_k}} \circ X_{t'_k} \middle| \mathcal{A}_\infty\right] \\ = & E\left[\left(P\left[X_{t_1} \in A_1 \middle| \mathcal{A}_\infty\right] + \zeta_n\right) 1_{A_{t_2}} \circ X_{t'_2} \ldots 1_{A_{t_k}} \circ X_{t'_k} \middle| \mathcal{A}_\infty\right] \\ = & P\left[X_{t_1} \in A_{t_1} \middle| \mathcal{A}_\infty\right] \cdot P\left[X_{t_i} \in A_{t_i}\ (i = 2, \ldots, k) \middle| \mathcal{A}_\infty\right] + E\left[\zeta_n \cdot \psi \middle| \mathcal{A}_\infty\right] \end{aligned}$$

wo $|\psi| \leq 1$. Da $|\zeta_n \psi| \leq 2$ und $\zeta_n \to 0$ f.s., folgt im Limes $n \to \infty$

$$P\left[X_{t_i} \in A_{t_i}\ (i = 1, \ldots, k) \middle| \mathcal{A}_\infty\right] = P\left[X_{t_1} \in A_{t_1} \middle| \mathcal{A}_\infty\right] \cdot P\left[X_{t_i} \in A_{t_i}\ (i = 2, \ldots, k) \middle| \mathcal{A}_\infty\right] \ ,$$

und die bedingte Unabhängigkeit folgt nun induktiv. □

Sei nun $(\Omega, \mathcal{A}) = (\mathbb{R}^T, \mathcal{B}^T)$. Um zu einer Verallgemeinerung des Kolmogorovschen 0-1-Gesetzes zu gelangen, führen wir folgende Notation ein: Ist $\Lambda \in \mathcal{E}(T)$ und $\pi \in \Pi(\Lambda)$ eine Permutation von Λ, so sei

$$\hat{\pi} : \Omega \to \Omega, \quad (\hat{\pi}\omega)_t = \begin{cases} \omega_{\pi(t)} & (t \in \Lambda) \\ \omega_t & (t \in T \setminus \Lambda) \end{cases}$$

26.8 Definition (Vertauschbare Ereignisse) *Mit*

$$\mathcal{V}_\infty := \left\{A \in \mathcal{A} : \hat{\pi}^{-1}A = A \text{ für alle } \pi \in \Pi(\Lambda), \Lambda \in \mathcal{E}(T)\right\}$$

bezeichnen wir die σ-Algebra der vertauschbaren Ereignisse.

Da für festes $\Lambda \in \mathcal{E}(T)$

$$\mathcal{F}_{\Lambda^c} \subseteq \left\{A \in \mathcal{A} : \hat{\pi}^{-1}A = A \text{ für alle } \pi \in \Pi(\Lambda)\right\} \ ,$$

ist

$$\mathcal{A}_\infty \subseteq \mathcal{V}_\infty \ .$$

Die Umkehrung ist nicht richtig, da z.B. für jedes nichttriviale $A \in \mathcal{B}$ das Ereignis $\{\exists t \in T \text{ so dass } \omega_t \in A\}$ zu $\mathcal{V}_\infty$ aber nicht zu $\mathcal{A}_\infty$ gehört. Es gilt aber:

26.9 Satz (0-1-Gesetz von Hewitt und Savage)
a) Sei $(X_t)_{t \in T}$ austauschbar mit Verteilung P. Dann ist $\mathcal{V}_\infty \subseteq \mathcal{A}_\infty \mod P$, d.h.

$$\forall V \in \mathcal{V}_\infty\ \exists V' \in \mathcal{A}_\infty : \ P(V \triangle V') = 0 \ .$$

b) Sei $(X_t)_{t \in T}$ unabhängig mit Verteilung P. Dann ist $\mathcal{V}_\infty$ trivial für P.

Beweis: Da jeder unabhängige Prozess austauschbar ist, folgt b) direkt aus a) und dem Kolmogorovschen 0-1-Gesetz. Wir beweisen a): Seien $\Lambda_n \in \mathcal{E}(T)$, $\Lambda_n \nearrow T$. Sei $V \in \mathcal{V}_\infty$. Es gibt $A_n \in \mathcal{F}_{\Lambda_n}$ derart, dass $P(V \triangle A_n) \to 0$ $(n \to \infty)$, siehe Satz 3.13b. Zu jedem n wählen wir ein $k(n) > n$ derart, dass $\mathrm{card}(\Lambda_{k(n)}) > 2\ \mathrm{card}(\Lambda_n)$. Dann gibt es ein $\pi_n \in \Pi(\Lambda_{k(n)})$ mit $\pi_n(\Lambda_n) \cap \Lambda_n = \emptyset$. Insbesondere ist $\hat{\pi}_n^{-1}(A_n) \in \mathcal{F}_{\Lambda_{k(n)} \setminus \Lambda_n} \subseteq \mathcal{F}_{\Lambda_n^c}$, und es folgt

$$P(V \triangle \hat{\pi}_n^{-1} A_n) = P\left(\hat{\pi}_n^{-1}(V \triangle A_n)\right) = P(V \triangle A_n) \to 0 \quad (n \to \infty).$$

Es folgt, dass für jedes feste $n > 0$ gilt: $V \in \mathcal{F}_{\Lambda_n^c} \mod P$, und man schließt leicht, dass auch $V \in \mathcal{A}_\infty \mod P$. $\square$

Nun wenden wir uns Gibbs-Maßen zu, wie wir sie in Kapitel 21 eingeführt haben. Wir setzen also voraus:

- $\triangleright$ $T = \mathbb{Z}^d$, $\Lambda_n \in \mathcal{E}(T)$, $\Lambda_n \nearrow T$, $\Omega = \Sigma^T$, Σ endlich.
- $\triangleright$ $(K^\Lambda : \Lambda \in \mathcal{E}(T))$ ist ein verträgliches System stochastischer Kerne (siehe Definition 21.2), d.h.

$$K^\Lambda\left(\omega_{\Lambda^c}, (\pi_\Gamma^\Lambda)^{-1} A \cap (\pi_{\Lambda\setminus\Gamma}^\Lambda)^{-1} B\right) = \int_{(\pi_{\Lambda\setminus\Gamma}^\Lambda)^{-1} B} K^\Gamma\left(\omega_{\Gamma^c}, A\right)\ K^\Lambda\left(\omega_{\Lambda^c}, d\omega_\Lambda\right) \qquad (**)$$

  für alle $\Gamma \subseteq \Lambda \in \mathcal{E}(T)$, $A \in \mathcal{B}^\Gamma$ und $B \in \mathcal{B}^{\Lambda\setminus\Gamma}$.
- $\triangleright$ P ist ein Gibbs-Maß zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$, d.h.

$$P\left(\pi_\Lambda^{-1} A\right) = \int_\Omega K^\Lambda(\omega_{\Lambda^c}, A)\, dP(\omega) \quad \text{für } \Lambda \in \mathcal{E}(T) \text{ und } A \in \mathcal{B}^\Lambda. \qquad \text{(DLR)}$$

Wir beginnen mit einer Martingaleigenschaft der Kerne K^{Λ_n}.

26.10 Satz (Martingaleigenschaft der Übergangskerne)
Sei P ein Gibbs-Maß zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$.

a) *Sei $A \in \mathcal{F}_{\Lambda_{n_0}}$. Für $n \geq n_0$ betrachte die Zufallsvariablen*

$$X_n(\omega) := K^{\Lambda_n}\left(\omega_{\Lambda_n^c}, (\pi_{\Lambda_{n_0}}^{\Lambda_n})^{-1} A\right).$$

Dann ist $(X_n, \mathcal{F}_{\Lambda_n^c})_{n \geq n_0}$ ein Rückwärtsmartingal auf $(\Omega, \mathcal{B}^T, P)$. Der Limes $X_\infty := \lim_{n\to\infty} X_n$ existiert P-f.s., und es ist $X_\infty = P\left[\pi_{\Lambda_{n_0}}^{-1} A \middle| \mathcal{A}_\infty\right]$. Die Zufallsvariable X_∞ kann $\mathcal{A}_\infty$-messbar gewählt werden.

b) *Für P-f.a. ω konvergieren die Wahrscheinlichkeitsmaße $\delta_{\omega_{\Lambda_n^c}} \times K^{\Lambda_n}(\omega_{\Lambda_n^c}, .)$ schwach gegen ein Wahrscheinlichkeitsmaß $K^\infty(\omega, .)$ auf $(\Omega, \mathcal{B}^T)$. K^∞ ist ein Übergangskern von $(\Omega, \mathcal{A}_\infty)$ nach $(\Omega, \mathcal{B}^T)$ und für $U \in \mathcal{B}^T$ ist*

$$K^\infty(\omega, U) = P\left[U \middle| \mathcal{A}_\infty\right](\omega).$$

c) *Falls die Abbildungen $\omega \mapsto K^\Lambda(\omega, A)$ für alle Λ und A stetig sind, sind die $K^\infty(\omega, .)$ für P-f.a. ω Gibbs-Maße zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$.*

d) *Die Identität (DLR) bleibt im Limes erhalten, d.h.*

$$P(U) = \int_\Omega K^\infty(\omega, U)\, dP(\omega) \quad \text{für alle } U \in \mathcal{B}^T.$$

Beweis:

a) Nach Satz 21.4 ist $X_n(\omega) = P\left[\pi_{\Lambda_{n_0}}^{-1} A \middle| \mathcal{F}_{\Lambda_n^c}\right](\omega)$. Daraus folgt die Behauptung, siehe Satz 25.17.

b) Für jedes ω ist die Folge $(\delta_{\omega_{\Lambda_n^c}} \times K^{\Lambda_n}(\omega_{\Lambda_n^c}, .))_{n\in\mathbb{N}}$ straff, so dass jede Teilfolge eine schwach konvergente Teilfolge hat. Für P-f.a. ω folgt aus Teil a), dass alle solche schwachen Limiten auf Zylindermengen übereinstimmen. Also konvergiert die Folge für P-f.a. ω schwach gegen ein Wahrscheinlichkeitsmaß $K^\infty(\omega, .)$ auf $(\Omega, \mathcal{B}^T)$, und für jede Zylindermenge $Z = \pi_{\Lambda_{n_0}}^{-1} A$ ist $\omega \mapsto K^\infty(\omega, Z) = \lim_{n\to\infty} K^{\Lambda_n}(\omega_{\Lambda_n^c}, (\pi_{\Lambda_{n_0}}^{\Lambda_n})^{-1}A) = P\left[Z\middle|\mathcal{A}_\infty\right](\omega)$ messbar bzgl. $\mathcal{A}_\infty$. Wegen Bemerkung 11.4 folgt daraus die $\mathcal{A}_\infty$-Messbarkeit von $\omega \mapsto K^\infty(\omega, U)$ für beliebige $U \in \mathcal{B}^T$, ebenso Identität $K^\infty(\omega, U) = P\left[U\middle|\mathcal{A}_\infty\right](\omega)$.

c) Das folgt aus der schwachen Konvergenz in b) und dem Beweis zu Satz 21.8.

d) Sei zunächst $U = \pi_\Lambda^{-1} A$, $A \in \mathcal{B}^\Lambda$, $\Lambda \in \mathcal{E}(T)$. Für hinreichend große n ist $\Lambda \subseteq \Lambda_n$. Also ist

$$P(U) = E\left[P[\pi_\Lambda^{-1}A|\mathcal{F}_{\Lambda_n^c}]\right] \to E\left[P[\pi_\Lambda^{-1}A|\mathcal{A}_\infty]\right] = \int_\Omega K^\infty(\omega, U)\, dP(\omega)\ .$$

Da die Abbildungen $U \mapsto P(U \cap V)$ und $U \mapsto \int_V K^\infty(\omega, U)\, dP(\omega)$ endliche Maße sind, überträgt sich diese Identität von Zylindermengen U auf biliebige $U \in \mathcal{B}^T$.

□

26.11 Satz (Gibbs-Maße: Eindeutigkeit und triviale asymptotische σ-Algebra)

a) Seien P, Q Gibbs-Maße zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$. Es ist $P = Q$ genau dann, wenn $P_{|\mathcal{A}_\infty} = Q_{|\mathcal{A}_\infty}$.

b) Die folgenden Aussagen sind äquivalent:

i) Es gibt höchstens ein Gibbs-Maß zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$.

ii) Ist P ein Gibbs-Maß zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$, so ist $\mathcal{A}_\infty$ trivial für P.

Beweis:

a) Das folgt aus Satz 26.10d.

b) i) $\Rightarrow$ ii) Sei $V \in \mathcal{A}_\infty$ und P ein Gibbs-Maß zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$. Angenommen, $0 < P(V) < 1$. Setze $P_V(U) := P(V \cap U)/V(B)$. Dann ist P_V ein Wahrscheinlichkeitsmaß, und für $A \in \mathcal{B}^\Lambda$, $\Lambda \in \mathcal{E}(T)$, ist

$$\int K^\Lambda(\omega_{\Lambda^c}, A)\, dP_V(\omega) \overset{\text{S.21.4}}{=} \frac{1}{P(V)} \int 1_V\, P\left[\pi_\Lambda^{-1}A\middle|\mathcal{F}_{\Lambda^c}\right]\, dP = \frac{P(\pi_\Lambda^{-1}A \cap V)}{P(V)} = P_V(\pi_\Lambda^{-1}A)\ .$$

Also ist P_V ein Gibbs-Maß zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$ und $P \neq P_V$, Widerspruch!.

ii) $\Rightarrow$ i) Seien P, Q Gibbs-Maße zu $(K^\Lambda : \Lambda \in \mathcal{E}(T))$. $R := \frac{1}{2}(P + Q)$ ist dann ebenfalls ein Gibbs-Maß (siehe Lemma 21.6). Nach Voraussetzung ist $\mathcal{A}_\infty$ trivial für P, Q, R. Dann ist aber für jedes $A \in \mathcal{A}_\infty$

$$P(A) = 1 \iff R(A) = 1 \iff Q(A) = 1\ ,$$

also $P_{|\mathcal{A}_\infty} = Q_{|\mathcal{A}_\infty}$, so dass $P = Q$ wegen Teil a) dieses Satzes.

□

26.12 Bemerkung Die Frage, ob $\mathcal{A}_\infty$ für die Maße $K^\infty(\omega, .)$ trivial ist, muss hier offen bleiben. Beachte aber, dass für jedes $V \in \mathcal{A}_\infty$ gilt: $K^\infty(\omega, V) = P\left[V\middle|\mathcal{A}_\infty\right](\omega) = 1_V(\omega) \in \{0, 1\}$ P-f.s. Das Problem ist, dass die Ausnahme-Nullmenge von V abhängt, und dass $\mathcal{A}_\infty$ in der Regel nicht abzählbar erzeugt ist.

Kapitel 27

Quadratintegrierbare Martingale

Dieses Kapitel dient unter anderem zur Vorbereitung der stochastischen Integration. Wir beginnen mit einem einfachen Zerlegungssatz für Submartingale mit diskreter Zeit.

27.1 Satz (Doob-Zerlegung)
Sei $\mathcal{X} = (X_n, \mathcal{F}_n)_{n\geq 0}$ ein Submartingal. Dann lässt sich X_n eindeutig zerlegen als

$$X_n = M_n + A_n \quad (n \geq 0)$$

wobei $(M_n, \mathcal{F}_n)_{n\geq 0}$ ein Martingal ist, $0 = A_0 \leq A_1 \leq A_2 \leq \ldots$ und die A_n bzgl. $\mathcal{F}_{n-1}$ messbar sind $(n \geq 1)$. Der Prozess $(A_n)_{n\geq 0}$ heißt der Kompensator *von $\mathcal{X}$.*

Beweis: Setze $M_0 := X_0$, $A_0 := 0$, und für $n \geq 1$

$$A_n := \sum_{k=1}^{n} \left(E\left[X_k \middle| \mathcal{F}_{k-1}\right] - X_{k-1}\right) \qquad M_n := X_n - A_n\ . \tag{*}$$

Nach Konstruktion ist A_n bzgl. $\mathcal{F}_{n-1}$ messbar, und da $\mathcal{X}$ ein Submartingal ist, ist $A_n - A_{n-1} = E\left[X_n \middle| \mathcal{F}_{n-1}\right] - X_{n-1} \geq 0$. Außerdem ist

$$E|M_n| \leq E|X_n| + EA_n = E|X_n| + EX_n - EX_0 < \infty$$

und

$$\begin{aligned} E\left[M_n \middle| \mathcal{F}_{n-1}\right] &= E\left[X_n - A_{n-1} - \big(E[X_n|\mathcal{F}_{n-1}] - X_{n-1}\big) \middle| \mathcal{F}_{n-1}\right] \\ &= E\left[X_n \middle| \mathcal{F}_{n-1}\right] - A_{n-1} - E\left[X_n \middle| \mathcal{F}_{n-1}\right] + X_{n-1} \\ &= M_{n-1}\ . \end{aligned}$$

Ist $X_n = M'_n + A'_n$ eine weitere Zerlegung der gleichen Art, so ist

$$\begin{aligned} \left(A'_n - A'_{n-1}\right) - \left(A_n - A_{n-1}\right) &= E\left[\left(A'_n - A'_{n-1}\right) - \left(A_n - A_{n-1}\right) \middle| \mathcal{F}_{n-1}\right] \\ &= -E\left[\left(M'_n - M'_{n-1}\right) - \left(M_n - M_{n-1}\right) \middle| \mathcal{F}_{n-1}\right] \\ &= 0 \end{aligned}$$

und da $A_0 = 0 = A'_0$ folgt sofort, dass $A_n = A'_n$ und damit auch $M_n = M'_n$ für alle n. $\square$

27.2 Definition (Quadratische Charakteristik) *Ist $\mathcal{X} = (X_n, \mathcal{F}_n)_{n\geq 0}$ ein* quadratintegrierbares *Martingal, so sei*

$$X_n^2 = M_n + \langle X, X \rangle_n \quad (n \geq 0)$$

die Doob-Zerlegung des Submartingals $\mathcal{X}^2$ (siehe Beispiel 24.10). Der monoton wachsende Prozess $\langle \mathcal{X}, \mathcal{X} \rangle$ heißt die quadratische Charakteristik *des Martingals $\mathcal{X}$. Außerdem bezeichnen wir $\Delta X_n := X_n - X_{n-1}$.*

27.3 Lemma *Sei $\mathcal{X}$ ein quadratintegrierbares Martingal, $\mathcal{X}^2 = \mathcal{M} + \langle \mathcal{X}, \mathcal{X} \rangle$. Dann gilt*

a) $\langle X, X \rangle_n = \sum_{k=1}^n E\left[(\Delta X_k)^2 \middle| \mathcal{F}_{k-1}\right]$, *insbesondere* $\langle X, X \rangle_n - \langle X, X \rangle_{n-1} = E\left[(\Delta X_n)^2 \middle| \mathcal{F}_{n-1}\right]$.

b) Für $0 \le m \le n$ *ist*

$$E\left[(X_n - X_m)^2 \middle| \mathcal{F}_m\right] = E\left[\langle X, X \rangle_n \middle| \mathcal{F}_m\right] - \langle X, X \rangle_m = \sum_{k=m+1}^n E\left[(\Delta X_k)^2 \middle| \mathcal{F}_m\right] .$$

Beweis: Zunächst ist

$$\begin{aligned} E\left[(X_n - X_m)^2 \middle| \mathcal{F}_m\right] &= E\left[X_n^2 \middle| \mathcal{F}_m\right] - 2X_m E\left[X_n \middle| \mathcal{F}_m\right] + X_m^2 = E\left[X_n^2 \middle| \mathcal{F}_m\right] - X_m^2 \\ &= E\left[\langle X, X \rangle_n \middle| \mathcal{F}_m\right] - \langle X, X \rangle_m , \end{aligned} \tag{**}$$

also die erste Identität in b). Aus (*) und (**) folgt (setze $n = k$ und $m = k - 1$)

$$\langle X, X \rangle_n = \sum_{k=1}^n \left(E\left[X_k^2 \middle| \mathcal{F}_{k-1}\right] - X_{k-1}^2\right) = \sum_{k=1}^n E\left[(\Delta X_k)^2 \middle| \mathcal{F}_{k-1}\right] ,$$

also a). Daraus folgt weiterhin

$$E\left[\langle X, X \rangle_n \middle| \mathcal{F}_m\right] - \langle X, X \rangle_m = E\left[\langle X, X \rangle_n - \langle X, X \rangle_m \middle| \mathcal{F}_m\right] = \sum_{k=m+1}^n E\left[(\Delta X_k)^2 \middle| \mathcal{F}_m\right] ,$$

d.h. die zweite Identität in b). $\square$

27.4 Beispiel Seien $\xi_0, \xi_1, \xi_2, \dots$ unabhängige Zufallsvariablen, $E\xi_i = 0$, $E\xi_i^2 < \infty$. Betrachte das Martingal $X_n := \sum_{i=0}^n \xi_i$ zur Filtration $\mathcal{F}_n = \sigma(\xi_0, \dots, \xi_n)$. Dann ist $\langle X, X \rangle_n = \sum_{i=1}^n E\left[\xi_i^2\right]$, siehe Lemma 27.3a.

27.5 Beispiel (Quadratische Charakteristik diskreter stochastischer Integrale)
Sei $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}_0}$ ein Martingal, und seien H_n beschränkte, $\mathcal{F}_{n-1}$-messbare Zufallsvariablen $(n = 1, 2, \dots)$. In Satz 24.17 wurde gezeigt, dass durch

$$(H \cdot X)_n = X_0 + \sum_{k=1}^n H_k (X_k - X_{k-1})$$

ein Martingal bzgl. $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ definiert wird. Es ist

$$\begin{aligned} \langle H \cdot X, H \cdot X \rangle_n &= \sum_{k=1}^n E\left[\left(\Delta (H \cdot X)_k\right)^2 \middle| \mathcal{F}_{k-1}\right] = \sum_{k=1}^n E\left[H_k^2 \left(X_k - X_{k-1}\right)^2 \middle| \mathcal{F}_{k-1}\right] \\ &= \sum_{k=1}^n H_k^2 \, E\left[(\Delta X_k)^2 \middle| \mathcal{F}_{k-1}\right] = \left(H^2 \cdot \langle X, X \rangle\right)_n \end{aligned}$$

Insbesondere ist

$$E\left[(H \cdot X)_n^2\right] = E\left[\langle H \cdot X, H \cdot X \rangle_n\right] = \sum_{k=1}^n E\left[H_k^2 (\Delta X_k)^2\right] = E\left[\left(H^2 \cdot \langle X, X \rangle\right)_n\right] . \tag{*}$$

Sei nun

$$\mathcal{T}_N := \{\mathcal{H} = (H_n)_{1 \le n \le N} : H_n \text{beschränkt und messbar bzgl. } \mathcal{F}_{n-1} \text{ für } 1 \le n \le N\} .$$

Man rechnet leicht nach, dass durch

$$(\mathcal{H}, \tilde{\mathcal{H}})_{\mathcal{X}} := \sum_{k=1}^n E\left[H_k \tilde{H}_k (\Delta X_k)^2\right]$$

ein Skalarprodukt auf $\mathcal{T}_N$ gegeben ist. Bei fest gehaltenem "Integrator" $\mathcal{X}$ erhält man also eine *isometrische lineare* Abbildung

$$I_{\mathcal{X}} : \mathcal{T}_N \to \mathcal{L}^2_P, \quad (H_n)_{1\leq n\leq N} \mapsto (H\cdot X)_N \,.$$

Diese Abbildung hat eine eindeutige lineare stetige Fortsetzung auf die $(\,.\,,\,.\,)_{\mathcal{X}}$-Vervollständigung von $\mathcal{T}_N$. Insbesondere kann damit die Einschränkung auf beschränkte H_n überwunden werden, wie man an Hand des Spezialfalls von Beispiel 27.5 sieht: Ist $X_n = \xi_0 + \cdots + \xi_n$, so folgt

$$\langle H\cdot X, H\cdot X\rangle_n = \sum_{k=1}^n H_k^2\, E[\xi_k^2] \quad \text{und}$$

$$(\mathcal{H},\tilde{\mathcal{H}})_{\mathcal{X}} = \sum_{k=1}^n E[H_k\tilde{H}_k]\, E[\xi_k^2] \leq \sum_{k=1}^n E[\xi_k^2]\cdot \|H_k\|_{L^2_P}\cdot\|\tilde{H}_k\|_{L^2_P} \,.$$

Im Allgemeinen kann man $(\mathcal{H},\tilde{\mathcal{H}})_{\mathcal{X}}$ interpretieren als

$$(\mathcal{H},\tilde{\mathcal{H}})_{\mathcal{X}} = \int_{\{1,\ldots,N\}\times\Omega} H_n(\omega)\tilde{H}_n(\omega)\, d\mu_{\mathcal{X}}(n,\omega)$$

wo für $I\subseteq\{1,\ldots,N\}$ und messbares $A\subseteq\Omega$ durch

$$\mu_{\mathcal{X}}(I\times A) := \sum_{k\in I}\int_A (\Delta X_k)^2\, dP$$

ein endliches Maß auf $\{1,\ldots,N\}\times\Omega$ definiert wird. $(\,.\,,\,.\,)_{\mathcal{X}}$ kann also mit dem kanonischen Skalarprodukt des $\mathcal{L}^2_{\mu_{\mathcal{X}}}$ identifiziert werden und die Vervollständigung von $\mathcal{T}_N$ mit dem Raum $L^2_{\mu_{\mathcal{X}}}$. Dieses ist auch die Grundstruktur der Argumentation, mit der wir im letzten Kapitel kontinuierliche stochastische Integrale einführen.

Vorher schauen wir uns noch ein paar Eigenschaften diskreter quadratintegrierbarer Martingale an. Als erstes beweisen wir einen L^2-Konvergenzsatz, ohne auf die bisherigen Konvergenzsätze zurückzugreifen.

27.6 Satz (Konvergenzsatz für quadratintegrierbare Martingale)
Sei $(X_n,\mathcal{F}_n)_{n\geq 0}$ ein Martingal mit $\sup_{n\geq 0} E\left[X_n^2\right] < \infty$. Dann existiert ein $X_\infty\in\mathcal{L}^2_P$ derart, dass $\lim_{n\to\infty}\|X_n - X_\infty\|_2 = 0$.

Beweis: Aus Lemma 27.3b folgt

$$E\left[(X_n - X_m)^2\right] = E\langle X,X\rangle_n - E\langle X,X\rangle_m \,. \qquad (***)$$

Da $0\leq\langle X,X\rangle_n\leq\langle X,X\rangle_{n+1}$ für alle n, existiert $\langle X,X\rangle_\infty := \sup_n\langle X,X\rangle_n$ und es ist

$$E\langle X,X\rangle_\infty = \sup_n E\langle X,X\rangle_n = \sup_n E\left[X_n^2\right] < \infty \,.$$

Also konvergiert $E\langle X,X\rangle_n \to E\langle X,X\rangle_\infty$, und aus (***) folgt, dass $(X_n)_{n\geq 0}$ eine $\mathcal{L}^2_P$-Cauchy-Folge ist. Die Existenz des Limes X_∞ folgt dann aus der Vollständigkeit des $\mathcal{L}^2_P$ (Satz 10.5). □

27.7 Beispiel Sei $(X_n,\mathcal{F}_n)_{n\geq 0}$ der Galton-Watson Prozess aus Beispiel 24.6, also $X_{n+1} = \sum_{j=1}^{X_n} N_{n,j}$. Wir haben gesehen, dass $Y_n = \gamma^{-n}X_n$, wo $\gamma = EN_{n,j}$, ein Martingal bildet. Bezeichne $\sigma^2 := V(N_{n,j})$. Da alle die $(N_{n-1,j} : j\in\mathbb{N})$ unabhängig untereinander und von X_{n-1} sind, rechnet man leicht nach, dass

$$
\begin{aligned}
E\left[Y_n^2|\mathcal{F}_{n-1}\right] &= \gamma^{-2n} E\left[\left(\sum_{j=1}^{X_{n-1}} N_{n-1,j}\right)^2 |\mathcal{F}_{n-1}\right] \\
&= \gamma^{-2n} X_{n-1} E\left[N_{1,1}^2\right] + \gamma^{-2n} X_{n-1}(X_{n-1}-1) E\left[N_{1,1}N_{1,2}\right] \\
&= \gamma^{-2n}\left(X_{n-1}\sigma^2 + X_{n-1}^2\gamma^2\right) \\
&= \gamma^{-n-1}\sigma^2 Y_{n-1} + Y_{n-1}^2
\end{aligned}
$$

Wir betrachten den Fall $\gamma > 1$. Da $EY_{n-1} = EY_0 = EX_0$, folgt

$$
E\left[Y_n^2\right] = \gamma^{-(n-1)}\frac{\sigma^2}{\gamma^2} EX_0 + E\left[Y_{n-1}^2\right] = \cdots = EY_0^2 + \frac{1-\gamma^{-n}}{1-\gamma^{-1}}\frac{\sigma^2}{\gamma^2} EX_0 \le EY_0^2 + \frac{\sigma^2}{\gamma^2-\gamma} EX_0 \, .
$$

Also konvergiert $Y_n = \gamma^{-n} X_n \to Y_\infty$ in $\mathcal{L}^2$. Insbesondere ist $EY_\infty = EY_0 = EX_0$ und der Prozess stirbt nicht fast sicher aus.

Ist $\gamma \le 1$, so kann man folgendermaßen zeigen, dass der Prozess fast sicher ausstirbt, falls $\sigma^2 > 0$: Angenommen es wäre $P\{\lim_{n\to\infty} X_n = 0\} \le 1 - \frac{1}{S}$ für ein $S > 0$. Dann wäre natürlich $P\{X_n \ge 1\} \ge \frac{1}{S}$ für alle n. Sei $A_n := \{1 \le X_n \le 2S, X_{n+1} = 0\}$. Da $A_n \supseteq \{1 \le X_n \le 2S\} \cap \{N_{n,1} = \cdots = N_{n,2S} = 0\}$, ist

$$
\begin{aligned}
P(A_n) &\ge (P\{1 \le X_n\} - P\{X_n > 2S\}) \cdot (P\{N_{1,1} = 0\})^{2S} \\
&\ge (\frac{1}{S} - \frac{1}{2S} E[X_n]) \cdot (P\{N_{1,1} = 0\})^{2S} \\
&= \frac{1}{2S} E[X_0] \cdot (P\{N_{1,1} = 0\})^{2S} \, .
\end{aligned}
$$

Da die A_n, $n \ge 0$, paarweise disjunkt sind und da $E[X_0] > 0$ ist, folgt $N_{n,j} \ge 1$ f.s. für alle n, j. Da $\gamma = E[N_{n,j}] \le 1$, folgt $N_{n,j} = 1$ f.s. im Widerspruch zu $\sigma^2 = V(N_{n,j}) > 0$.

Als weitere Anwendung beweisen wir eine Version der Waldschen Gleichung für zweite Momente.

27.8 Korollar *Sei $(X_n, \mathcal{F}_n)_{n\ge 0}$ ein quadratintegrierbares Martingal und τ eine fast sicher endliche Stoppzeit zu $(\mathcal{F}_n)_{n\ge 0}$. Ist $E\langle X, X\rangle_\tau < \infty$, so gilt*

$$
E\left[X_\tau\right] = E\left[X_0\right] \quad \textit{und} \quad E\left[X_\tau^2\right] = E\left[X_0^2\right] + E\langle X, X\rangle_\tau \, .
$$

Beweis: Sei $M_n = X_n^2 - \langle X, X\rangle_n$, also $M_0 = X_0^2$. Da $\mathcal{M}$ ein Martingal ist, gilt für alle $n \ge 0$ nach Satz 25.12: $EM_{\tau\wedge n} = EM_0 = E[X_0^2]$, also

$$
\sup_n EX_{\tau\wedge n}^2 = \sup_n E\langle X, X\rangle_{\tau\wedge n} + E[X_0^2] \le E\langle X, X\rangle_\tau + E[X_0^2] < \infty \, .
$$

Daher kann man Satz 27.6 auf das Martingal $(X_{\tau\wedge n}, \mathcal{F}_{\tau\wedge n})_{n\ge 0}$ anwenden und erhält, dass $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ nicht nur f.s. sondern auch in $\mathcal{L}^2$ gilt. Insbesondere folgt

$$
\begin{aligned}
E\left[X_\tau\right] &= \lim_{n\to\infty} E\left[X_{\tau\wedge n}\right] = E\left[X_0\right] \\
E\left[X_\tau^2\right] &= \lim_{n\to\infty} E\left[X_{\tau\wedge n}^2\right] = \lim_{n\to\infty} E\langle X, X\rangle_{\tau\wedge n} + E\left[X_0^2\right] = E\langle X, X\rangle_\tau + E\left[X_0^2\right] \, .
\end{aligned}
$$

□

Bevor wir im nächsten Kapitel das, was wir hier über die quadratische Charakteristik diskreter stochastischer Integrale gelernt haben, auf die stochastische Integration anwenden, wollen wir uns exemplarisch noch mit Konvergenzfragen bei quadratintegrierbaren Martingalen befassen. Viele der Aussagen lassen sich verallgemeinern, siehe z.B. das Buch von Širjaev [Širjaev].

27.9 Lemma (Die Konvergenzmenge nichtnegativer Submartingale) *Sei $(X_n, \mathcal{F}_n)_{n\geq 0}$ ein nichtnegatives Submartingal mit Doob-Zerlegung $X_n = M_n + A_n$. Sei $A_\infty := \sup_n A_n$. Dann ist*

$$\{A_\infty < \infty\} \subseteq \{X_n \text{ konvergiert}\} \quad P\text{-f.s.}$$

Beweis: Für $a > 0$ sei $\tau_a := \inf\{n \geq 0 : A_{n+1} > a\}$. Man beachte, dass τ_a eine Stoppzeit ist, weil A_{n+1} bzgl. $\mathcal{F}_n$ messbar ist. Aus Satz 24.18 folgt

$$E[X_{\tau_a \wedge n}] = E[M_{\tau_a \wedge n}] + E[A_{\tau_a \wedge n}] \leq E[M_0] + a < \infty\ .$$

Der Martingalkonvergenzsatz 25.1 garantiert dann die fast sichere Konvergenz von $(X_{\tau_a \wedge n})_{n\geq 0}$, so dass $(X_n)_{n\geq 0}$ auf der Menge $\{\tau_a = \infty\} = \{A_\infty \leq a\}$ fast sicher konvergiert. Die Behauptung folgt nun aus der Beobachtung, dass $\{A_\infty < \infty\} = \bigcup_{a\in\mathbb{N}}\{A_\infty \leq a\}$. □

Eine leichte Folgerung ist nun der folgende Satz:

27.10 Satz (Die Konvergenzmenge quadratintegrierbarer Martingale)
Sei $\mathcal{M} = (M_n, \mathcal{F}_n)$ ein quadratintegrierbares Martingal, $\langle M, M\rangle_\infty := \sup_n \langle M, M\rangle_n$. Dann ist

$$\{\langle M, M\rangle_\infty < \infty\} \subseteq \{M_n \text{ konvergiert}\} \quad P\text{-f.s.}$$

Beweis: $\{\langle M, M\rangle_\infty < \infty\} \subseteq \{M_n^2 \text{ konvergiert}\}$ P-f.s. folgt aus Lemma 27.9. Betrachte nun das Martingal $\mathcal{M} + 1 = (M_n + 1, \mathcal{F}_n)_{n\geq 0}$. Da $\Delta(M_n + 1) = \Delta M_n$, ist $\langle M + 1, M + 1\rangle_\infty = \langle M, M\rangle_\infty$, also auch

$$\{\langle M, M\rangle_\infty < \infty\} = \{\langle M+1, M+1\rangle_\infty < \infty\} \subseteq \{(M_n + 1)^2 \text{ konvergiert}\} \quad P\text{-f.s.}$$

Da $\{M_n^2 \text{ konvergiert}\} \cap \{(M_n+1)^2 \text{ konvergiert}\} = \{M_n \text{ konvergiert}\}$, folgt die Behauptung des Satzes. □

Die bisher gewonnenen Informationen über die Konvergenzmenge quadratintegrierbarer Martingale $\mathcal{M}$ wollen wir nun zur Untersuchung der Frage benutzen, unter welchen Voraussetzungen die Folge $\frac{M_n}{\langle M,M\rangle_n}$ fast sicher konvergiert. Dazu benötigen wir u.a. das folgende einfache Lemma über reelle Zahlenfolgen:

27.11 Lemma (Kronecker Lemma) *Sei $(z_n)_{n\in\mathbb{N}}$ eine Folge reeller Zahlen und $0 < c_n \nearrow \infty$. Ist $b := \sum_{n=1}^\infty \frac{z_n}{c_n} < \infty$, so ist $\lim_{n\to\infty} \frac{1}{c_n}\sum_{i=1}^n z_i = 0$.*

Beweis: Setze $c_0 := 0$, $b_0 := 0$, $b_n := \sum_{i=1}^n \frac{z_i}{c_i}$ $(n \geq 1)$, also $\lim_{n\to\infty} b_n = b$. Da $z_n = (b_n - b_{n-1})c_n$, folgt für $n \geq 1$:

$$\frac{1}{c_n}\sum_{i=1}^n z_i = \frac{1}{c_n}\sum_{i=1}^n (b_i - b_{i-1})c_i = b_n + \sum_{i=1}^{n-1} b_i \frac{c_i - c_{i+1}}{c_n} = \sum_{i=0}^{n-1}(b_n - b_i)\frac{c_{i+1} - c_i}{c_n}\ ,$$

da $\sum_{i=0}^{n-1}\frac{c_{i+1}-c_i}{c_n} = \frac{c_n - c_0}{c_n} = 1$. Sei $\epsilon > 0$ und wähle $N = N(\epsilon)$ so, dass $\sup_{i,j\geq N}|b_j - b_i| < \epsilon$ für $n \geq N$. Für solche n ist dann

$$\left|\frac{1}{c_n}\sum_{i=1}^n z_i\right| \leq \frac{1}{c_n}\sum_{i=0}^{N-1}|b_n - b_i|\cdot|c_{i+1} - c_i| + \sum_{i=N}^{n-1}|b_n - b_i|\frac{|c_{i+1} - c_i|}{c_n} \leq \frac{c_N - c_0}{c_n}\cdot 2\sup_{i\geq 0}|b_i| + \epsilon$$

und da und $(b_i)_{i\geq 0}$ als konvergente Folge beschränkt ist, geht der erste Summand auf der rechten Seite gegen 0 mit $n \to \infty$. □

27.12 Satz
Sei $(M_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ ein quadratintegrierbares Martingal. Dann ist $\lim_{n\to\infty}\frac{M_n}{\langle M,M\rangle_n} = 0$ P-f.s. auf der Menge $\{\langle M, M\rangle_\infty = \infty\}$.

Beweis: Da $\lim_{n\to\infty} \frac{\langle M,M\rangle_n}{\langle M,M\rangle_n+1} = 1$ auf $\{\langle M,M\rangle_\infty = \infty\}$, reicht es, die fast sichere Konvergenz von $\frac{M_n}{\langle M,M\rangle_n+1}$ auf dieser Mengezu zeigen. Wegen des Kronecker Lemmas, angewandt auf $z_n = \Delta M_n$ und $c_n = \langle M,M\rangle_n$, reicht es dazu zu zeigen, dass $X_n := \sum_{k=1}^n \frac{\Delta M_k}{\langle M,M\rangle_k+1}$ auf $\{\langle M,M\rangle_\infty = \infty\}$ P-f.s. konvergiert. Beachte nun, dass $X_n = (\frac{1}{\langle M,M\rangle_n+1} \cdot M)$ ein diskretes stochastisches Integral ist. Wegen Satz 27.10 bleibt zu zeigen, dass $\langle X,X\rangle_\infty < \infty$ P-f.s. In Beispiel 27.5 hatten wir gesehen, dass

$$\langle X,X\rangle_n = \left(\frac{1}{(\langle M,M\rangle+1)^2} \cdot \langle M,M\rangle\right)_n = \sum_{k=1}^n \frac{\langle M,M\rangle_k - \langle M,M\rangle_{k-1}}{(\langle M,M\rangle_k+1)^2}\,.$$

Unter Beachtung der Monotonie von $\langle M,M\rangle_k$ folgt nun

$$\begin{aligned}\langle X,X\rangle_\infty &\le \sum_{k=1}^\infty \frac{\langle M,M\rangle_k - \langle M,M\rangle_{k-1}}{(\langle M,M\rangle_k+1)(\langle M,M\rangle_{k-1}+1)}\\ &= \sum_{k=1}^\infty \left(\frac{1}{\langle M,M\rangle_{k-1}+1} - \frac{1}{\langle M,M\rangle_k+1}\right)\\ &\le \frac{1}{\langle M,M\rangle_0+1} < \infty\end{aligned}$$

□

27.13 Beispiel (Parameterschätzung bei autoregressiven Prozessen 1.Ordnung) Seien $\xi_1,\xi_2,\xi_3,\dots$ unabhängige Zufallsvariablen mit $E\xi_i = 0$, $E\xi_i^2 = 1$. Betrachte den autoregressiven Prozess 1.Ordnung $X_n = \theta X_{n-1} + \xi_n$ $(n \ge 1)$ mit von den ξ_i unabhängigem X_0. $\theta \in \mathbb{R}$ wird als unbekannter Parameter angesehen und soll auf Basis der zufälligen Beobachtungen $X_0, X_1, \dots, X_n$ geschätzt werden. Der Schätzer $\hat\theta_n$ wird als *Kleinster Quadrate Schätzer* so bestimmt, dass der quadratische Fehler $\sum_{k=1}^n (X_k - \theta X_{k-1})^2$ als Funktion von θ für $\theta = \hat\theta_n$ minimal wird. Man kann leicht zeigen, dass $\hat\theta_n = \theta + \frac{M_n}{\langle M,M\rangle_n}$, wo $M_n = \sum_{k=1}^n X_{k-1}\xi_k$ ein Martingal vom Typ "diskretes stochastisches Integral" ist, also $\langle M,M\rangle_n = \sum_{k=1}^n X_{k-1}^2 E[\xi_k^2] = \sum_{k=1}^n X_{k-1}^2$. Um Satz 27.12 anwenden zu können, bleibt zu zeigen, dass $\langle M,M\rangle_\infty = \infty$ P-f.s. Da s folgt aber aus der fast sicheren Identität

$$\begin{aligned}\infty &= \sum_{k=1}^\infty \xi_k^2 = \sum_{k=1}^\infty (X_k - \theta X_{k-1})^2\\ &\le \sum_{k=1}^\infty 2\,(X_k^2 + \theta^2 X_{k-1}^2) \le \theta^2 X_0^2 + 2\sum_{k=1}^\infty (1+\theta^2) X_k^2\\ &= \theta^2 X_0^2 + 2(1+\theta^2)\langle M,M\rangle_\infty\,.\end{aligned}$$

Kapitel 28

Das stochastische Integral

Ziel dieses Kapitels ist es, einem Ausdruck der Gestalt $\int_0^t H_s\,dX_s$ einen Sinn zu geben, wenn $\mathcal{H}$ und $\mathcal{X}$ stochastische Prozesse sind. Ganz allgemein geht das nicht, aber man kann doch eine recht große Klasse von Prozessen dabei zulassen. Das offensichtlich größte Problem liegt darin, dass die Pfade von $t \mapsto X_t(\omega)$ unbeschränkte Variation haben können, wie das z.B. bei der Brownschen Bewegung der Fall ist. Wir werden uns diesem Problem in mehreren Schritten nähern.

Integratoren mit Pfaden von endlicher Variation
Zunächst sei $G : [0,T] \to \mathbb{R}$ eine monoton wachsende und rechtsseitig stetige Funktion mit $G(0) = 0$. Dann ist G die Verteilungsfunktion eines endlichen Maßes m_G auf $[0,T]$, siehe Beispiel 2.4. Man schreibt auch

$$\int_0^t f(s)\,dG(s) := \int_0^t f\,dm_G$$

und erhält auf diese Weise eine "G-Stammfunktion" zu f. Ist G stetig, d.h. hat m_G keine Punktmasse, so ist die Stammfunktion stetig, sonst ist die Stammfunktion zumindest für beschränkte f noch rechtsseitig stetig. Dieses Integral ist wie üblich zunächst für messbare $f \geq 0$ und dann für m_G-integrierbare f definiert. So werden wir die Existenz eines solchen Integrals für messbare f auch im Folgenden verstehen, ohne das explizit zu erwähnen.

Ist nun $(H_t)_{0\leq t\leq T}$ ein messbarer stochastischer Prozess (d.h. ist die Abbildung $(t,\omega) \mapsto H_t(\omega)$ messbar bzgl. $\mathcal{B}([0,T]) \times \mathcal{F}$), so kann man das Integral $\int_0^t H_s\,dG(s)$ noch Pfadweise definieren und gelangt zu der Zufallsvariablen

$$\int_0^t H_s(\omega)\,dG(s)\ .$$

Diese Überlegungen übertragen sich sofort auf den Fall, dass G eine rechtsseitig stetige Funktion von *beschränkter Variation* ist, d.h.

$$\sup\left\{\sum_{i=1}^n |G(t_i) - G(t_{i-1})| : 0 = t_0 < t_1 < \cdots < t_n = T\right\} < \infty\ .$$

Denn dann lässt sich G schreiben als $G = G_1 - G_2$ mit monoton wachsenden und rechtsseitig stetigen Funktionen G_1 und G_2, und man setzt wenn immer diese Differenz wohldefiniert ist[1]

$$\int_0^t H_s\,dG(s) := \int_0^t H_s\,dG_1(s) - \int_0^t H_s\,dG_2(s)$$

Grundsätzlich lässt sich diese Vorgehensweise auf den Fall übertragen, wo $G(t)$ ein Pfad eines messbaren stochastischen Prozesses $(X_t)_{0\leq t\leq T}$ ist. Das Integral $\int_0^t H_s\,dX_s$ definiert für jedes t eine Zufallsvariable – der Zufall kommt dabei sowohl durch den Integranden H_s als auch durch den Integrator X_s ins Spiel.

[1] Diese Zerlegung von G ist völlig analog zur Jordan-Zerlegung eines signierten Maßes; es gilt nämlich $(m_G)^+ = m_{G_1}$ und $(m_G)^- = m_{G_2}$.

28.1 Beispiel (Ein Poisson-Prozess als Integrator) Sei $(N_t)_{t\geq 0}$ ein Poisson-Prozess zum Parameter α, also $N_t(\omega) = \max\{n \geq 0 : S_n(\omega) \leq t\}$, wo $S_n = \xi_1 + \cdots + \xi_n$ mit unabhängigen $\mathcal{E}_\alpha$-verteilten ξ_i. Dann ist

$$\int_0^t H_s\, dN_s = \sum_{\substack{n=1 \\ S_n(\omega)\leq t}}^{\infty} H_{S_n(\omega)}(\omega)$$

Integratoren mit Pfaden von unendlicher Variation

Der bisherige Ansatz findet sein Ende, wenn der Integrator Pfade mit lokal unendlicher Variation hat, wie z.B. die Brownsche Bewegung. Dass die Brownsche Bewegung dabei kein ungünstiger Sonderfall ist, belegt der folgende Satz.

28.2 Satz (Stetige Martingale haben Pfade von unendlicher Variation)
Sei $\mathcal{X}$ ein Martingal mit stetigen Pfaden. Dann sind entweder fast alle Pfade von $\mathcal{X}$ konstant, oder $\mathcal{X}$ hat mit positiver Wahrscheinlichkeit Pfade von unendlicher Variation (im Sinne von Korollar 22.20).

Einen *Beweis* findet man in [Revuz-Yor, Proposition IV.1.2].

Aber auch durch Integratoren von unendlicher pfadweiser Variation lassen sich stückweise konstante Integranden im Sinne eines Riemann-Stieltjes Integrals noch integrieren. Sei $\mathcal{U} = (0 = t_0 < t_1 < \cdots < t_n = T)$, seien $\tilde{H}_i$ beschränkte Zufallsvariablen $(i = 1, \ldots, n)$ und sei

$$H_t := \sum_{i=1}^{n} \tilde{H}_i\, 1_{(t_{i-1},t_i]}(t)\ . \tag{*}$$

$(H_t)_{0\leq t\leq T}$ ist ein stochastischer Prozess mit $H_t = \tilde{H}_i$ für $t_{i-1} < t \leq t_i$ $(i = 1 \ldots, n)$. Wir setzen nun

$$\mathcal{U}\text{-}\int_0^t H_s\, dX_s := \sum_{i=1}^{n} \tilde{H}_i \left(X_{t_i\wedge t} - X_{t_{i-1}\wedge t}\right)$$

Haben die Pfade von $\mathcal{X}$ endliche Variation, so stimmt das mit dem pfadweisen Riemann-Stieltjes Integral $\int H_s\, dX_s$ überein. Um von diesem Punkt an eine schlagkräftige Theorie der stochastischen Integration aufbauen zu können, setzen wir nun voraus:

$$\mathcal{X} = (X_t, \mathcal{F}_t)_{0\leq t\leq T} \text{ ist ein Martingal.}$$

Dann ist auch $\mathcal{X}^{\mathcal{U},t} := (X_{t_i\wedge t}, \mathcal{F}_{t_i\wedge t})_{i=0,\ldots,n}$ für jedes $t > 0$ ein Martingal und es ist

$$\mathcal{U}\text{-}\int_0^t H_s\, dX_s = (\tilde{H} \cdot \mathcal{X}^{\mathcal{U},t})_n$$

ein diskretes stochastisches Integral, vorausgesetzt,

$$\text{die } \tilde{H}_i \text{ sind } \mathcal{F}_{t_{i-1}\wedge t}\text{-messbar.}$$

Aus Beispiel 27.5 folgt

$$\begin{aligned} E\left[\mathcal{U}\text{-}\left(\int_0^t H_s\, dX_s\right)^2\right] &= E\left[(\tilde{H} \cdot X^{\mathcal{U},t})_n^2\right] = E\left[\langle \tilde{H} \cdot X^{\mathcal{U},t}, \tilde{H} \cdot X^{\mathcal{U},t}\rangle_n\right] \\ &= E\left[\sum_{i=1}^n H_i^2 (\Delta X_i^{\mathcal{U},t})^2\right] = E\left[\sum_{i=1}^n H_i^2 E[(\Delta X_i^{\mathcal{U},t})^2 | \mathcal{F}_{k-1}]\right] \qquad (**) \\ &= E\left[\sum_{i=1}^n H_{t_i}^2 \Delta\langle X^{\mathcal{U},t}, X^{\mathcal{U},t}\rangle_i\right] = E\left[\int_0^t H_s^2\, d\langle X^{\mathcal{U},t}, X^{\mathcal{U},t}\rangle_s\right] \end{aligned}$$

wo $X_s^{\mathcal{U},t} := X_i^{\mathcal{U},t}$ für $t_{i-1} < s \le t_i$. Wir erinnern daran, dass $\Delta\langle X^{\mathcal{U},t}, X^{\mathcal{U},t}\rangle_i = E[(\Delta X_i^{\mathcal{U},t})^2|\mathcal{F}_{i-1}]$ gemäß unserer Festlegung in Lemma 27.3, beobachten aber gleichzeitig, dass (**) auch richtig bleibt, wenn wir $\Delta\langle X^{\mathcal{U},t}, X^{\mathcal{U},t}\rangle_i$ als $(\Delta X_i^{\mathcal{U},t})^2$ interpretieren. Das werden wir von jetzt an tun. Insbesondere wird Satz 28.3 diese Interpretation benutzen.

Als nächstes wollen wir zeigen, dass dieses stochastische Integral von der Darstellung des Prozesses H_t in (*) unabhängig ist. Zu zeigen ist, dass Darstellungen von $\mathcal{H}$ durch zwei Unterteilungen $\mathcal{U}'$ und $\mathcal{U}''$ zum selben Wert $\mathcal{U}'\text{-}\int_0^t H_s\,dX_s = \mathcal{U}''\text{-}\int_0^t H_s\,dX_s$ führen. Es reicht, beide Integrale mit dem entsprechenden Integral der Verfeinerung $\mathcal{U}$ von $\mathcal{U}'$ und $\mathcal{U}''$ zu vergleichen, und da die Unterteilungen endlich sind, können wir uns schließlich auf den Fall zurückziehen, wo $\mathcal{U}'$ aus $\mathcal{U}$ durch Streichung eines t_{i_0} entsteht. Dann ist $\tilde{H}_{i_0} = \tilde{H}_{i_0+1}$ und daher

$$\begin{aligned}\mathcal{U}'\text{-}\int_0^t H_s\,dX_s &= \sum_{i=1}^{i_0-1} \tilde{H}_i(X_{t_i\wedge t} - X_{t_{i-1}\wedge t})\\ &\quad + \tilde{H}_{i_0+1}(X_{t_{i_0}\wedge t} - X_{t_{i_0-1}\wedge t}) + \sum_{i=i_0+1}^{n} \tilde{H}_i(X_{t_i\wedge t} - X_{t_{i-1}\wedge t})\\ &= \sum_{i=1}^{n} \tilde{H}_i(X_{t_i\wedge t} - X_{t_{i-1}\wedge t}) = \mathcal{U}\text{-}\int_0^t H_s\,dX_s\ .\end{aligned}$$

Daher können wir bei "stückweise konstanten" Integranden $\mathcal{H}$ wie in (*) von nun an $\int_0^t H_s\,dX_s$ an Stelle von $\mathcal{U}\text{-}\int_0^t H_s\,dX_s$ schreiben.

Um zu allgemeineren Integranden überzugehen, soll wieder ein L^2-Isometrie-Argument wie in Beispiel 27.5 zum Einsatz kommen. Dazu muss der Ausdruck $E\left[\int_0^t H_s^2\,d\langle X^{\mathcal{U},t}, X^{\mathcal{U},t}\rangle_s\right]$ in (**) ohne Bezug auf eine endliche Unterteilung geschrieben werden. Es stellt sich also die Frage, ob man sinnvoller Weise von einem Limes der $\langle \mathcal{X}^{\mathcal{U},t}, \mathcal{X}^{\mathcal{U},t}\rangle$ sprechen kann, wenn die "Maschenweite" $|\mathcal{U}|$ von $\mathcal{U}$ gegen null geht. Unter einer recht schwachen lokalen gleichgradigen Integrabilitätsannahme, die zum Begriff des *lokalen Martingals* führt, ist das tatsächlich der Fall.

28.3 Satz (Quadratische Charakteristik lokaler Martingale)
Ist $\mathcal{X}$ ein stetiges lokales Martingal, so gibt es genau einen wachsenden stetigen Prozess $\langle \mathcal{X}, \mathcal{X}\rangle$ mit $\langle X, X\rangle_0 = 0$ derart, dass $\mathcal{X}^2 - \langle \mathcal{X}, \mathcal{X}\rangle$ ein stetiges lokales Martingal ist und dass für jede Folge $\mathcal{U}_n$ von Unterteilungen mit $|\mathcal{U}_n| \to 0$ gilt

$$\sup_{0\le s\le t} \left|\langle X^{\mathcal{U}_n,t}, X^{\mathcal{U}_n,t}\rangle_s - \langle X, X\rangle_s\right| \to 0 \quad (n\to\infty)$$

stochastisch. Dabei ist $X_s^{\mathcal{U}_n,t} = X_i^{\mathcal{U}_n,t}$ für $t_{i-1} < s \le t_i$ wie oben.

Der *Beweis* ist aber etwas länglich und kann in [Revuz-Yor, Theoreme IV.1.3und IV.1.8] gefunden werden. Zur Erläuterung des Begriffs "lokales Martingal" merken wir an:

28.4 Bemerkung (Lokales Martingal) Jedes rechtsseitig stetige Martingal ist ein lokales Martingal.

28.5 Beispiel (Brownsche Bewegung als Integrator) Im Spezialfall der Brownschen Bewegung hatten wir das und noch mehr schon in Satz 22.18 bewiesen, wo gezeigt wurde, dass (in der jetzigen Notation)

$$\begin{aligned}\langle B^{\mathcal{U},t}, B^{\mathcal{U},t}\rangle_n &= \sum_{i=1}^{n} E\left[(\Delta B_i^{\mathcal{U},t})^2\middle|\mathcal{F}_{t_{i-1}\wedge t}\right] = \sum_{i=1}^{n} E\left[(\Delta B_i^{\mathcal{U},t})^2\right]\\ &= \sum_{i=1}^{n}(t_i\wedge t - t_{i-1}\wedge t) = t_n\wedge t = t\end{aligned}$$

für alle $\mathcal{U}$. (Die zweite Identität beruht auf der Unabhängigkeit der Zuwächse im Fall der Brownschen Bewegung.) Dass dabei auch die Kompensatoreigenschaft im Limes erhalten bleibt, wurde bereits in Satz 24.8 gezeigt: $(B_t^2 - t, \mathcal{F}_t)_{t\geq 0}$ ist ein Martingal.

Der Satz gestattet, auch auf der rechten Seite von (**) zum Limes $|\mathcal{U}| \to 0$ überzugehen, und man erhält

$$E\left[\left(\int_0^t H_s\,dX_s\right)^2\right] = E\left[\int_0^t H_s^2\,d\langle X,X\rangle_s\right] \tag{***}$$

Dadurch wird nahe gelegt, welche Klasse von Prozessen mit dem Integrator $\mathcal{X}$ integriert werden kann.

28.6 Definition (Previsibel) *Sei $(\mathcal{F}_t)_{0\leq t\leq T}$ eine Filtration.*

a)

$$\mathcal{R} := \{\{0\}\times F_0 : F_0 \in \mathcal{F}_0\} \cup \{(s,t]\times F_s : 0 \leq s < t < \infty, F_s \in \mathcal{F}_s\}$$

ist $\cap$-stabil und wird als System der previsiblen Rechtecke bezeichnet.

b) $\mathcal{P} := \sigma(\mathcal{R})$ heißt die previsible σ-Algebra.

c) Ein Prozess $\mathcal{H} = (H_t)_{0\leq t\leq T}$ heißt previsibel, falls die Abbildung

$$(t,\omega) \mapsto H_t(\omega) \text{ von } [0,T]\times\Omega \text{ nach } \mathbb{R}$$

$\mathcal{P}$-messbar ist.

28.7 Bemerkung Jeder adaptierte, linksseitig stetige Prozess ist previsibel, denn es gilt

$$H_t(\omega) = H_0(\omega)\,1_{\{0\}}(t) + \lim_{n\to\infty}\sum_{k=1}^{n2^n} H_{k2^{-n}}(\omega)\,1_{(k2^{-n},(k+1)2^{-n}]}(t)\ .$$

Für den Rest des Kapitels nehmen wir an:

> $\mathcal{X} = (X_t, \mathcal{F}_t)_{0\leq t\leq T}$ ist ein quadratintegrierbares Martingal, und es gibt einen wachsenden, adaptierten Prozess $(\langle X,X\rangle_t)_{0\leq t\leq T}$ mit $\langle X,X\rangle_0 = 0$, für den $\mathcal{X}^2 - \langle\mathcal{X},\mathcal{X}\rangle$ ein Martingal ist.

Sei $\mathcal{L}^2(\mathcal{X})$ der Raum der (bzgl. $(\mathcal{F}_t)_{0\leq t\leq T}$) previsiblen Prozesse $\mathcal{H}$ mit

$$\|\mathcal{H}\|_{\mathcal{X}}^2 := E\left[\int_0^T H_s^2\,d\langle X,X\rangle_s\right] < \infty\ .$$

Dieser Raum lässt sich wie in Beispiel 27.5 wieder als $\mathcal{L}^2$-Raum zu einem Maß auf $[0,T]\times\Omega$ darstellen: Für $\Gamma \in \mathcal{B}([0,T])\times\mathcal{F}_T$ sei

$$\mu_{\mathcal{X}}(\Gamma) := \int_\Omega\int_0^T 1_\Gamma(s,\omega)\,d\langle X,X\rangle_s(\omega)\,dP(\omega)\ .$$

Dann ist $\mathcal{L}^2(\mathcal{X})$ auf triviale Weise isomorph zum Raum $\mathcal{L}^2_{\mu_{\mathcal{X}}|\mathcal{P}}$ aller previsiblen, $\mu_{\mathcal{X}}$-quadratintegrierbaren Funktionen auf $[0,T]\times\Omega$, und wir werden diese beiden Räume von nun an identifizieren.

Bezeichne $\mathcal{T} \subset \mathcal{L}^2(\mathcal{X})$ den reellen Vektorraum aller Prozesse $\mathcal{H}$, die sich mit einer geeigneten Unterteilung $\mathcal{U}$ wie in (*) als $H_t = \sum_{i=1}^n \tilde{H}_i \cdot 1_{(t_{i-1},t_i]}(t)$ darstellen lassen. Da die $\tilde{H}_i$ beschränkt und $\mathcal{F}_{t_{i-1}}$-messbar sind, ist $\mathcal{T} \subseteq \mathcal{L}^2_{\mu_{\mathcal{X}}|\mathcal{P}}$. Durch

$$I_{\mathcal{X}} : \mathcal{T} \to \mathcal{L}^2_P, \quad \mathcal{H} \mapsto \int_0^T H_s\,dX_s$$

wird eine lineare isometrische Abbildung definiert; siehe (***) für die Isometrie-Eigenschaft. $I_{\mathcal{X}}$ lässt sich stetig linear auf den Abschluss $\overline{\mathcal{T}}$ von $\mathcal{T}$ im $\mathcal{L}^2_{\mu_{\mathcal{X}}|\mathcal{P}}$ fortsetzen, und man kann zeigen, dass $\overline{\mathcal{T}} = \mathcal{L}^2_{\mu_{\mathcal{X}}|\mathcal{P}} = \mathcal{L}^2(\mathcal{X})$.

28.8 Definition (Stochastisches Integral) *Die Fortsetzung von $I_{\mathcal{X}}$ auf $\mathcal{L}^2(\mathcal{X})$ heißt stochastisches Integral. Für $\mathcal{H} \in \mathcal{L}^2(\mathcal{X})$ schreibt man $I_{\mathcal{X}}(\mathcal{H}) = \int_0^T H_s\, dX_s$.*

Ohne Beweis listen wir einige Eigenschaften des stochastischen Integrals auf:

- ▷ $\int_0^t H_s\, dX_s$ ist für jedes $t \in [0, T]$ definiert.
- ▷ $(\mathcal{H} \cdot \mathcal{X}) := \left(\int_0^t H_s\, dX_s, \mathcal{F}_t\right)_{0 \le t \le T}$ ist ein Martingal, vergleiche Satz 24.17.
- ▷ Das Martingal $(\mathcal{H} \cdot \mathcal{X})$ hat eine Modifikation mit rechtsseitig stetigen Pfaden. Es heißt auch das *Itô Integral* von $\mathcal{H}$ bzgl. $\mathcal{X}$.

Wir beschließen dieses Kapitel mit einem sehr einfachen Beispiel für den Kalkül mit stochastischen Integralen.

28.9 Bemerkung Sei $\mathcal{X}$ ein quadratintegrierbares stetiges Martingal. Dann ist

$$X_t^2 = X_0^2 + 2 \int_0^t X_s\, dX_s + \langle X, X \rangle_t \,.$$

Denn sei $\mathcal{U} = (0 = t_0 < t_1 < \cdots < t_n = t)$. Dann ist

$$2 \sum_{i=1}^n X_{t_{i-1}} (X_{t_i} - X_{t_{i-1}}) + \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2 = \sum_{i=1}^n X_{t_i}^2 - \sum_{i=1}^n X_{t_{i-1}}^2 = X_t^2 - X_0^2 \,.$$

Mit der vorher eingeführten Notation wird das zu

$$2 \int_0^t X_s^{\mathcal{U},t}\, dX_s + \sum_{i=1}^n \left(\Delta X_i^{\mathcal{U},t}\right)^2 = X_t^2 - X_0^2 \,.$$

Im Limes $|\mathcal{U}| \to 0$ folgt die Behauptung.

28.10 Beispiel Im Fall der Brownschen Bewegung folgt:

$$B_t^2 - t = 2 \int_0^t B_s\, dB_s \,.$$

Eine Verallgemeinerung der vorangehenden Bemerkung ist

28.11 Satz (Itô Formel)
Sei $F \in \mathcal{C}^2(\mathbb{R}; \mathbb{R})$ und $\mathcal{X}$ ein stetiges quadratintegrierbares Martingal. Dann ist

$$F(X_t) = F(X_0) + \int_0^t F'(X_s)\, dX_s + \frac{1}{2} \int_0^t F''(X_s)\, d\langle X, X \rangle_s \,.$$

Diese Formel bleibt für Prozesse richtig, die sich als Summe eines stetigen lokalen Martingals und eines stetigen adaptierten Prozesses von endlicher Variation schreiben lassen.

Die Heuristik hinter dem Beweis zu diesem Satz ist folgende: Sei $\mathcal{U} = (0 = t_0 < \cdots < t_n = t)$.

$$F(X_t) - F(X_0) = \sum_{i=1}^n F(X_{t_i}) - F(X_{t_{i-1}})$$

$$= \sum_{i=1}^n F'(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) + \frac{1}{2} \sum_{i=1}^n F''(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}})^2 + o(1) \cdot \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2$$

Wegen der Stetigkeit von F' und F'' kann man zeigen, dass der erste Term mit $|\mathcal{U}| \to 0$ gegen $\int_0^t F'(X_s)\, dX_s$ und der zweite gegen $\frac{1}{2} \int_0^t F''(X_s)\, d\langle X, X \rangle_s$ konvergiert, während der dritte gegen null geht.

Index