

Wahrscheinlichkeitstheorie

VON PETER PFAFFELHUBER

Version: 25. Januar 2019

Vorbemerkung

Dieses Manuskript ist parallel zu Vorlesungen entstanden, die ich in den Wintersemestern 2010 und 2018 an der Universität Freiburg gehalten habe. Es stellt eine Grundlage der modernen Stochastik dar und wird ergänzt werden durch Skripte zu den Vorlesungen Stochastische Prozesse und Stochastische Integration.

Die Stochastik ist ein Gebiet der Mathematik, das sich bis heute ständig weiterentwickelt. Sie ist weder ein Gebiet der reinen noch der angewandten Mathematik, sondern beides. Auf der einen Seite stehen oftmals intuitive Ideen im Vordergrund, deren Formalisierung entscheidende Fortschritte bringt. Auf der anderen Seite führt der so entwickelte Formalismus ein Eigenleben, und ist unerlässlich im formalen Umgang mit der Stochastik.

Die moderne Wahrscheinlichkeitstheorie basiert auf der *Maßtheorie*, die wir zu Beginn des Skriptes wiederholen. Der Teil über *Wahrscheinlichkeitstheorie* umfasst die verschiedenen Konvergenzarten (fast sicher, stochastik, schwach, L^p), die in der Stochastik vorkommen. Weiter wird der Begriff der bedingten Erwartung eingeführt, der die Behandlung von Martingalen einläutet. Ausblicke werden zu den Themen zufällige Graphen, Populationsgenetik und Finanzmathematik gegeben.

Es gibt viele Lehrbücher der Stochastik. Besonders erwähnen möchte ich hier allerdings jedoch nur ein paar, die mir beim Schreiben des Skriptes stets wertvolle Dienste erwiesen:

- Kallenberg, Olaf. Foundations of Modern Probability Theory. Springer, Second edition, 2010
- Klenke, Achim. Wahrscheinlichkeitstheorie, Springer, 2. Auflage, 2008
- Lamberton, Damien; Lapeyre, Bernard. Introduction to Stochastic Calculus Applied to Finance. Chapman and Hall/CRC Financial Mathematics. 2. Auflage, 2007
- Øksendal, Bernt. Stochastic Differential Equations: An Introduction with Applications. Springer, 6. Auflage, 2003
- Rogers, L.C.D.; Williams, David. Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus. Cambridge Mathematical Library, 2. Auflage, 2000

Inhaltsverzeichnis

I	Maßtheorie	5
1	Wiederholung Topologie	5
1.1	Grundlagen	5
1.2	Kompakte Mengen	8
2	Mengensysteme	11
2.1	Halbringe, Ringe und σ -Algebren	11
2.2	Erzeuger und Erweiterungen	13
2.3	Dynkin-Systeme	15
2.4	Kompakte Systeme	16
3	Maße	17
3.1	Mengenfunktionen	17
3.2	σ -Additivität	21
3.3	Eindeutigkeit und Fortsetzung von Maßen	23
3.4	Maße auf $\mathcal{B}(\mathbb{R})$	28
3.5	Bildmaße	30
4	Messbare Funktionen und das Integral	31
4.1	Messbare Funktionen	31
4.2	Definition	34
4.3	Eigenschaften des Integrals	37
4.4	Konvergenzsätze	39
5	\mathcal{L}^p-Räume	40
5.1	Grundlagen	41
5.2	\mathcal{L}^p -Konvergenz	42
5.3	Der Raum \mathcal{L}^2	43
5.4	Satz von Radon-Nikodým	44
6	Produkträume	48
6.1	Topologie	48
6.2	Mengensysteme	49
6.3	Maße und Integrale	50
6.4	Faltung von Maßen	53
6.5	Projektive Familien von Wahrscheinlichkeitsmaßen	55

II Grundlagen der Wahrscheinlichkeitstheorie	58
7 Zufallsvariable	58
7.1 Wiederholung	58
7.2 Momente	62
7.3 Charakteristische Funktionen	63
8 Fast sichere, stochastische und \mathcal{L}^p-Konvergenz	66
8.1 Definition und Beispiele	66
8.2 Fast sichere und stochastische Konvergenz	68
8.3 Stochastische und \mathcal{L}^p -Konvergenz	69
9 Unabhängigkeit und das starke Gesetz	73
9.1 Definition und einfache Eigenschaften	73
9.2 Das Kolmogorov'sche 0-1-Gesetz	76
9.3 Summen unabhängiger Zufallsvariable	78
9.4 Das starke Gesetz der großen Zahlen	80
10 Schwache Konvergenz	84
10.1 Definition und einfache Eigenschaften	84
10.2 Der Satz von Prohorov	91
10.3 Separierende Funktionenklassen	95
10.4 Der Satz von Lévy	97
11 Grenzwertsätze in Verteilung	101
11.1 Poisson-Konvergenz	101
11.2 Der zentrale Grenzwertsatz	104
11.3 Mehrdimensionale Grenzwertsätze	108
12 Die bedingte Erwartung	110
12.1 Motivation	110
12.2 Definition und Eigenschaften	111
12.3 Der Fall $\mathcal{G} = \sigma(X)$	115
12.4 Bedingte Unabhängigkeit	117
12.5 Reguläre Version der bedingten Verteilung	119
13 Martingale	122
13.1 Definition und ein paar Beispiele	122
13.2 Das stochastische Integral als Martingal	126
13.3 Gestoppte Martingale	129
13.4 Martingalkonvergenzsätze	134
13.5 Der zentrale Grenzwertsatz für Martingale	144
14 Ausblicke	148
14.1 Zufällige Graphen	148
14.2 Populationsgenetik	152
14.3 Finanzmathematik	159

Teil I

Maßtheorie

Die Maßtheorie war bereits Teil der Vorlesung *Analysis 3*. Es mag zunächst verwundern, dass ein Gebiet, in dem es um Messen von Teilmengen eines Grundraumes geht, etwas mit Stochastik zu tun hat. Insbesondere sind die in der Analysis betrachteten Grundräume meist Teilmengen von reellen Zahlen, Zufallsvariablen können jedoch Werte in ganz anderen Mengen annehmen. (Man denke etwa an die Menge {Kopf, Zahl}, die beim Münzwurf auftritt.) Wie sich jedoch herausstellt, gelten für Flächenmessungen ähnliche Rechenregeln wie für Wahrscheinlichkeiten. Klar ist beispielsweise, dass man Flächen zweier disjunkter Mengen addieren muss, wenn man deren gemeinsamen Inhalt berechnen will. Genauso muss man die Wahrscheinlichkeit für zwei disjunkte Ereignisse addieren, wenn man die Wahrscheinlichkeit ausrechnen will, dass eines der beiden Ereignisse eintritt.

In diesem Kapitel bezeichnet $\Omega \neq \emptyset$ (irgend)eine Menge und alle aufgeführten Mengensysteme¹ sind Teilmengen der Potenzmenge von Ω , die wir mit 2^Ω bezeichnen. Alle betrachteten Mengensysteme seien o.E. nicht leer. Zentrale Begriffe werden σ -Algebra und (Wahrscheinlichkeits-)Maß sein.

1 Wiederholung Topologie

Topologien werden in der Mathematik immer dann gebraucht, wenn ein Konvergenzbegriff eingeführt wird. Auch wenn Topologien in den bisherigen Vorlesungen nur am Rande behandelt wurden, sind doch einige Konvergenzbegriffe bekannt. Auch zwischen Maßtheorie und Topologie gibt es viele Verbindungen. Insbesondere werden wir Maße auf topologischen Räumen betrachten. Dabei werden die zugrunde liegenden σ -Algebren von Topologie erzeugt (siehe Definition 2.7). Deshalb wiederholen wir zunächst Grundbegriffe der Topologie.

1.1 Grundlagen

Unter einer *Topologie* versteht man eine Familie offener Teilmengen eines Grundraumes Ω . Was offene Mengen sind, ist etwa schon aus der Analysis bekannt. In metrischen Räumen nennt man eine Menge A genau dann offen, wenn für jedes $\omega \in A$ auch ein offener Ball $B_\varepsilon(\omega) \subseteq A$ für ein $\varepsilon > 0$. Dieser Fall von metrischen Räumen ist in der Praxis auch am wichtigsten. In der Maßtheorie kommt dem Fall von separablen Topologien, die von vollständigen Metriken erzeugt werden, eine besondere Bedeutung zu. Solche Räume heißen *polnisch*.

Definition 1.1. 1. Eine Funktion $r : \Omega \times \Omega \rightarrow \mathbb{R}_+$ heißt Metrik, falls (i) $r(\omega, \omega') \neq 0$ für $\omega \neq \omega'$, (ii) $r(\omega, \omega') = r(\omega', \omega)$ für alle $\omega, \omega' \in \Omega$ und (iii) $r(\omega, \omega'') \leq r(\omega, \omega') + r(\omega', \omega'')$ für alle $\omega, \omega', \omega'' \in \Omega$. Das Paar (Ω, r) ist ein metrischer Raum.

Für $\omega \in \Omega$ und $\varepsilon > 0$ bezeichnen wir mit $B_\varepsilon(\omega) := \{\omega' \in \Omega : r(\omega, \omega') < \varepsilon\}$ den offenen Ball um ω mit Radius ε .

¹Als *Mengensystem* bezeichnet man eine Menge, deren Elemente wieder Mengen sind.

2. Eine Metrik r auf Ω heißt vollständig, wenn jede Cauchy-Folge konvergiert. Ist also $\omega_1, \omega_2, \dots \in \Omega$ mit

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n, m \geq N : r(\omega_n, \omega_m) < \varepsilon,$$

so gibt es ein $\omega \in \Omega$ mit $r(\omega_n, \omega) \xrightarrow{n \rightarrow \infty} 0$.

3. Ein Mengensystem $\mathcal{O} \subseteq 2^\Omega$ heißt Topologie, falls (i) $\emptyset, \Omega \in \mathcal{O}$, (ii) ist $A, B \in \mathcal{O}$, so ist auch $A \cap B \in \mathcal{O}$ (iii) ist I beliebig und ist $A_i \in \mathcal{O}, i \in I$, so ist auch $\bigcup_{i \in I} A_i \in \mathcal{O}$. Das Paar (Ω, \mathcal{O}) heißt topologischer Raum. Mengen $A \in \mathcal{O}$ heißen offen, Mengen $A \subseteq \Omega$ mit $A^c \in \mathcal{O}$ heißen abgeschlossen.
4. Ist (Ω, \mathcal{O}) ein topologischer Raum und $A \subseteq \Omega$. Dann heißt

$$A^\circ := \bigcup \{O \subseteq A : O \in \mathcal{O}\}$$

das Innere von A und

$$\bar{A} := \bigcap \{F \supseteq A : F^c \in \mathcal{O}\}$$

den Abschluss von A .

5. Ein topologischer Raum (Ω, \mathcal{O}) heißt separabel, wenn es eine abzählbare Menge $\Omega' \subseteq \Omega$ gibt mit $\bar{\Omega}' = \Omega$.
6. Sei (Ω, \mathcal{O}) ein topologischer Raum und $\mathcal{B} \subseteq \mathcal{O}$. Dann heißt \mathcal{B} eine Basis von \mathcal{O} , falls

$$\forall A \in \mathcal{O} \forall \omega \in A \exists B \in \mathcal{B} : \omega \in B \subseteq A.$$

Dies ist genau dann der Fall, wenn

$$\mathcal{O} = \{A \subseteq \Omega : \forall \omega \in A \exists B \in \mathcal{B} : \omega \in B \subseteq A\}. \quad (1.1)$$

oder (äquivalent dazu)

$$\mathcal{O} = \left\{ \bigcup_{i \in I} B_i : B_i \in \mathcal{B}, i \in I, I \text{ beliebig} \right\}. \quad (1.2)$$

7. Sei $\mathcal{B} \subseteq 2^\Omega$. Dann wird durch (1.1) oder (1.2) eine Topologie $\mathcal{O}(\mathcal{B})$ definiert, die von \mathcal{B} erzeugte Topologie.

8. Sei (Ω, r) ein metrischer Raum und

$$\mathcal{B} := \{B_\varepsilon(\omega) : \varepsilon > 0, \omega \in \Omega\}. \quad (1.3)$$

Dann ist $\mathcal{O}(\mathcal{B})$ die von r erzeugte Topologie. Falls speziell $\Omega \subseteq \mathbb{R}^d$ und r der euklidische Abstand ist, heißt die in (1.1) oder (1.2) definierte Topologie die euklidische Topologie.

9. Der Raum (Ω, \mathcal{O}) heißt (vollständig) metrisierbar, wenn es eine (vollständige) Metrik r auf Ω gibt, so dass (1.1) gilt. Der Raum (Ω, \mathcal{O}) heißt polnisch, falls er separabel und vollständig metrisierbar ist.

10. Seien (Ω, \mathcal{O}) und (Ω', \mathcal{O}') topologische Räume. Dann heißt eine Abbildung $f : \Omega \rightarrow \Omega'$ stetig, falls $f^{-1}(A') \in \mathcal{O}$ für alle $A' \in \mathcal{O}'$ gilt.

Beispiel 1.2 (Der Raum $\overline{\mathbb{R}}$). Wir werden oftmals Funktionen mit Werten in

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\} \quad \text{oder} \quad \overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$$

betrachten.² Um diese Räume als topologische Räume betrachten zu können, setzen wir

$$\varphi : \begin{cases} \overline{\mathbb{R}} & \rightarrow [-1, 1], \\ x & \mapsto \begin{cases} \frac{2}{\pi} \arctan(x), & x \in \mathbb{R}, \\ 1, & x = \infty, \\ -1, & x = -\infty \end{cases} \end{cases}$$

und definieren die Metrik

$$r_{\overline{\mathbb{R}}}(x, y) := |\varphi(x) - \varphi(y)|, \quad x, y \in \overline{\mathbb{R}}.$$

Der von $r_{\overline{\mathbb{R}}}$ definierte topologische Raum $(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ erweitert die euklidische Topologie $(\mathbb{R}, \mathcal{O})$ auf \mathbb{R} insofern, als dass $\{A \cap \mathbb{R} : A \in \overline{\mathcal{O}}\} = \mathcal{O}$. Dies gilt deshalb, weil φ stetig auf \mathbb{R} ist mit stetiger Umkehrfunktion. Weiter gilt, dass $(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ separabel ist und $r_{\overline{\mathbb{R}}}$ ist eine vollständige Metrik.

Auf $\overline{\mathbb{R}}$ kann man wie in der Analysis gewohnt rechnen. Etwa ist $a \cdot \infty = \infty$ für $a > 0$. Allerdings sind die Ausdrücke wie $\infty - \infty$ und ∞/∞ nicht definiert.

Bemerkung 1.3 (Zusammenhang zwischen Metrik und Topologie). Sei (Ω, \mathcal{O}) ein topologischer Raum und $\omega, \omega_1, \omega_2, \dots \in \Omega$. Wir definieren

$$\omega_n \xrightarrow{n \rightarrow \infty} \omega : \iff \forall O \in \mathcal{O} : \omega \in O \Rightarrow \omega_n \in O \text{ für fast alle } n \in \mathbb{N}. \quad (1.4)$$

Insbesondere ergibt so jede Topologie auf Ω einen Konvergenzbegriff für Folgen in Ω .

Dieser Konvergenzbegriff stimmt mit dem auf metrischen Räumen bekannten Begriff überein: ist nämlich r eine Metrik auf Ω , die \mathcal{O} erzeugt, dann gilt die rechte Seite aus (1.4) genau dann, wenn für alle $\varepsilon > 0$ gilt, dass $r(\omega_n, \omega) < \varepsilon$ für fast alle $n \in \mathbb{N}$.

Mit Hilfe des Konvergenzbegriffes aus (1.4) sehen wir nun folgende bekannte Eigenschaft ein:

Lemma 1.4 (Abschluss bei metrischen Räumen). Sei (Ω, r) ein metrischer Raum und \mathcal{O} die von r erzeugte Topologie. Für $F \subseteq \Omega$ sind äquivalent:

1. F ist abgeschlossen.
2. Für alle $\omega_1, \omega_2, \dots \in F$ und $\omega \in \Omega$ mit $\omega_n \xrightarrow{n \rightarrow \infty} \omega$ gilt $\omega \in F$.

Insbesondere gilt: für jedes $A \subseteq \Omega$ besteht der Abschluss \overline{A} genau aus den Häufungspunkten von A .

²Die Schreibweise $\overline{\mathbb{R}}$ suggeriert, dass hier der Abschluss von \mathbb{R} gemeint ist. Dies stimmt nicht, da die hinzugekommenen Elemente $-\infty, \infty$ nicht in \mathbb{R} liegen, Abschlüsse von Mengen aber immer höchstens die Elemente des Grundraumes enthalten können. Topologisch gesehen ist $\overline{\mathbb{R}}$ die Zwei-Punkte-Kompaktifizierung von \mathbb{R} .

Beweis. '1. \Rightarrow 2.': Angenommen, es gäbe $\omega_1, \omega_2, \dots \in F$, konvergent gegen $\omega \in F^c$, dann wäre, da $F^c \in \mathcal{O}$ auch $\omega_n \in F^c$ für fast alle n . Dies ist im Widerspruch zur Voraussetzung.

'2. \Rightarrow 1.': Angenommen, F wäre nicht abgeschlossen, F^c also nicht offen. Dann gibt es $\omega \in F^c$, so dass für alle $\varepsilon > 0$ gilt, dass $B_\varepsilon(\omega) \not\subseteq F^c$. Wähle $\varepsilon_1, \varepsilon_2, \dots > 0$ mit³ $\varepsilon_n \downarrow 0$ und $\omega_n \in B_{\varepsilon_n}(\omega) \cap F$. Dann gilt $\omega_1, \omega_2, \dots \in F$, $\omega_n \xrightarrow{n \rightarrow \infty} \omega$, aber $\omega \in F^c$. \square

Lemma 1.5 (Abzählbare Basis und separable Räume). Sei (Ω, r) ein separabler, metrischer Raum, \mathcal{O} die von r erzeugte Topologie, Ω' abzählbar mit $\overline{\Omega'} = \Omega$ und

$$\tilde{\mathcal{B}} := \{B_\varepsilon(\omega) : \varepsilon \in \mathbb{Q}_+, \omega \in \Omega'\}.$$

Dann ist $\tilde{\mathcal{B}}$ abzählbar und $\mathcal{O}(\tilde{\mathcal{B}}) = \mathcal{O}$.

Beweis. Klar ist, dass $\tilde{\mathcal{B}}$ abzählbar ist und $\mathcal{O}(\tilde{\mathcal{B}}) \subseteq \mathcal{O}$. Sei \mathcal{B} wie in (1.3). Dann ist für $B_\varepsilon(\omega) \in \mathcal{B}$

$$B_\varepsilon(\omega) = \bigcup_{\tilde{B} \ni B \subseteq B_\varepsilon(\omega)} B,$$

also $\mathcal{B} \subseteq \mathcal{O}(\tilde{\mathcal{B}})$ und damit $\mathcal{O} = \mathcal{O}(\mathcal{B}) \subseteq \mathcal{O}(\tilde{\mathcal{B}})$. \square

Beispiel 1.6 (Zwei polnische Räume). 1. Sei \mathcal{O} die euklidische Topologie auf \mathbb{R}^d . Diese wird nach Definition 1.1.9 durch die euklidische Metrik definiert. Bekannt ist, dass diese vollständig ist. Weiter ist \mathbb{Q}^d abzählbar und jedes $\omega \in \mathbb{R}^d$ ist Häufungspunkt einer Folge in \mathbb{Q}^d . Insbesondere ist also $\overline{\mathbb{Q}^d} = \mathbb{R}^d$ nach Lemma 1.4, also ist \mathbb{R}^d separabel. Insgesamt ist also $(\mathbb{R}^d, \mathcal{O})$ polnisch.

2. Sei $K \subseteq \mathbb{R}$ kompakt und $\Omega = \mathcal{C}_{\mathbb{R}}(K)$ die Menge der stetigen Funktionen $\omega : K \rightarrow \mathbb{R}$. Auf Ω sei

$$r(\omega_1, \omega_2) := \sup_{x \in K} |\omega_1(x) - \omega_2(x)|$$

der Supremumsabstand. Bekannt ist wieder, dass r vollständig ist. Außerdem lässt sich jedes $\omega \in \Omega$ nach dem Weierstraß'schen Approximationssatz gleichmäßig durch Polynome approximieren. Sei Ω' die abzählbare Menge der Polynome mit rationalen Koeffizienten. Dann gilt auch, dass $\overline{\Omega'} = \Omega$. Damit ist (Ω, \mathcal{O}) separabel, also polnisch.

1.2 Kompakte Mengen

Topologische Räume können sehr groß sein. Man denke schon an den Raum \mathbb{R} , in dem es Folgen gibt, die divergieren. Nun werden *kompakten Menge* als kleinere Teilmengen eines topologischen Raumes betrachtet. In solchen kompakten Mengen gibt es immer konvergente Teilfolgen.

Definition 1.7 (Relativ kompakt, kompakt, relativ folgenkompakt, total beschränkt). Sei (Ω, \mathcal{O}) ein topologischer Raum und $K \subseteq \Omega$.

1. Die Menge K heißt *kompakt*, falls jede offene Überdeckung eine endliche Teilüberdeckung besitzt. Das bedeutet: sind $O_i \in \mathcal{O}, i \in I$ und $K \subseteq \bigcup_{i \in I} O_i$, dann gibt es⁴ $J \in I$ mit $K \subseteq \bigcup_{i \in J} O_i$.

³Wir schreiben $\varepsilon_n \downarrow 0$ falls $\varepsilon_1 \geq \varepsilon_2 \geq \dots$ und $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$

⁴Wir schreiben $J \in I$, falls $J \subseteq I$ und J endlich ist.

2. Die Menge K heißt relativ kompakt, falls \overline{K} kompakt ist.
3. Die Menge K heißt relativ folgenkompakt, falls gilt: für jede Folge $\omega_1, \omega_2, \dots \in K$ gibt es eine konvergente Teilfolge, d.h. $\omega_{k_1}, \omega_{k_2}, \dots \in K$ und $\omega \in \Omega$ mit $\omega_{k_n} \xrightarrow{n \rightarrow \infty} \omega$ wie in (1.4).
4. Sei r eine Metrik, die \mathcal{O} erzeugt. Dann heißt $K \subseteq \Omega$ total beschränkt, wenn es für jedes $\varepsilon > 0$ ein $N \in \mathbb{N}$ und $\omega_1, \dots, \omega_N \in K$ gibt, so dass $K \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$.

Lemma 1.8 (Kompakte Mengen sind abgeschlossen). Sei (Ω, r) ein metrischer Raum und \mathcal{O} die von r erzeugte Topologie. Ist $K \subseteq \Omega$ kompakt, dann ist K auch abgeschlossen.

Beweis. Wir zeigen, dass K^c offen ist. Sei hierzu $\omega \in K^c$. Für alle $\omega' \in K$ wählen wir $\delta_{\omega'}$ und $\varepsilon_{\omega'}$, so dass $B_{\delta_{\omega'}}(\omega) \cap B_{\varepsilon_{\omega'}}(\omega') = \emptyset$. Dann ist offenbar $\bigcup_{\omega' \in K} B_{\varepsilon_{\omega'}}(\omega') \supseteq K$, also gibt es $J \Subset K$ mit $K \subseteq \bigcup_{\omega' \in J} B_{\varepsilon_{\omega'}}(\omega')$. Setze $\delta := \min_{\omega' \in J} \delta_{\omega'} > 0$. Dann ist $B_\delta(\omega) \cap K \subseteq B_\delta(\omega) \cap \bigcup_{\omega' \in J} B_{\varepsilon_{\omega'}}(\omega') = \emptyset$, also $B_\delta(\omega) \subseteq K^c$. Da $\omega \in K^c$ beliebig war, ist K^c offen, K also abgeschlossen. \square

Folgender Satz über kompakte Mengen unterstreicht zum ersten Mal, warum polnischen Räumen eine besondere Bedeutung zukommt.

Proposition 1.9 (Charakterisierung relativ kompakter Mengen). Sei (Ω, r) ein metrischer Raum, \mathcal{O} die von r erzeugte Topologie und $K \subseteq \Omega$. Betrachte folgende Aussagen:

1. K ist relativ kompakt.
2. Es gilt: sind $F_i \subseteq \overline{K}$ abgeschlossen, $i \in I$ und $\bigcap_{i \in I} F_i = \emptyset$, dann gibt es $J \Subset I$ mit $\bigcap_{i \in J} F_i = \emptyset$.
3. K ist relativ folgenkompakt.
4. K ist total beschränkt.

Dann gilt

$$4. \iff 1. \iff 2. \implies 3.$$

Außerdem gilt auch $3. \implies 2.$ falls (Ω, \mathcal{O}) separabel ist und $4. \implies 3.$ falls (Ω, r) vollständig ist. Insbesondere sind alle vier Aussagen äquivalent, falls (Ω, \mathcal{O}) polnisch ist.

Korollar 1.10. Sei (Ω, r) ein metrischer Raum, \mathcal{O} die von r erzeugte Topologie. Dann sind abgeschlossene Teilmengen kompakter Mengen wieder kompakt.

Beweis. Sei $K \subseteq \Omega$ kompakt und $A \subseteq K$ abgeschlossen. Eine abgeschlossene Menge ist genau dann kompakt, wenn sie relativ kompakt ist. Aus Proposition 1.9.2 liest man wegen der Relativkompaktheit von K ab, dass für $F_i \in \mathcal{O}^c$, $i \in I$ mit $F_i \subseteq A \subseteq K$ und $\bigcap_{i \in I} F_i = \emptyset$ ein $J \Subset I$ existiert mit $\bigcap_{i \in J} F_i = \emptyset$. Wieder mit Proposition 1.9.2 folgt daraus, dass A relativ kompakt, also kompakt ist. \square

Beweis von Proposition 1.9. '1. \implies 4.': Sei \overline{K} kompakt und $\varepsilon > 0$. Offenbar ist $\bigcup_{\omega \in K} B_\varepsilon(\omega) \supseteq \overline{K}$ eine offene Überdeckung. Damit gibt es eine endliche Teilüberdeckung, d.h. es gibt $\omega_1, \dots, \omega_N$ mit $\overline{K} \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$. Da $\varepsilon > 0$ beliebig war, folgt die Behauptung.

'1. \Rightarrow 2.': Sei nun $F_i, i \in I$ wie angegeben. Dann ist $\bigcup_{i \in I} F_i^c = \left(\bigcap_{i \in I} F_i\right)^c = \Omega \supseteq \bar{K}$. Da \bar{K} kompakt ist, gibt es $J \Subset I$ mit $\bar{K} \subseteq \bigcup_{i \in J} F_i^c$. Damit ist $\bigcap_{i \in J} F_i = \left(\bigcup_{i \in J} F_i^c\right)^c \subseteq \bar{K}^c$. Da aber $F_i \subseteq \bar{K}$ vorausgesetzt war, ist $\bigcap_{i \in J} F_i = \emptyset$.

'2. \Rightarrow 1.': Sei $O_i \in \mathcal{O}, i \in I$ mit $\bar{K} \subseteq \bigcup_{i \in I} O_i$. Setze $F_i = O_i^c \cap \bar{K}$, dann ist $F_i^c \in \mathcal{O}$ und $\bigcap_{i \in I} F_i = \bar{K} \cap \left(\bigcup_{i \in I} O_i\right)^c = \emptyset$. Also gibt es $J \Subset I$ mit $\bigcap_{i \in J} F_i = \emptyset$. Damit ist $\bar{K}^c \cup \bigcup_{i \in J} O_i = \bigcup_{i \in J} F_i^c = \Omega$, also $\bigcup_{i \in J} O_i \supseteq \bar{K}$. Mit anderen Worten ist \bar{K} kompakt.

'2. \Rightarrow 3.': Sei $\omega_1, \omega_2, \dots \in K$. Wir setzen $F_n = \overline{\{\omega_n, \omega_{n+1}, \dots\}} \subseteq \bar{K}$. Angenommen, es gibt keine konvergente Teilfolge von $\omega_1, \omega_2, \dots$. Dann ist $\bigcap_{n=1}^{\infty} F_n = \emptyset$. Aus 2. folgt dann, dass es ein $N \in \mathbb{N}$ gibt mit $\emptyset = \bigcap_{n=1}^N F_n = F_N$. Dies ist ein Widerspruch, da F_N nach Konstruktion nicht leer ist; also gibt es eine konvergente Teilfolge.

'3. \Rightarrow 1.' falls (Ω, \mathcal{O}) separabel ist: Sei Ω' abzählbar mit $\bar{\Omega}' = \Omega$ und $\mathcal{B} := \{B_{1/n}(\omega) : \omega \in \Omega', n \in \mathbb{N}\}$. Damit ist \mathcal{B} eine Basis von \mathcal{O} und auch abzählbar. Wir schreiben $\mathcal{B} = \{B_1, B_2, \dots\}$.

Angenommen \bar{K} ist nicht kompakt. Das heißt, es gibt $A_i \in \mathcal{O}, i \in I$ mit $\bar{K} \subseteq \bigcup_{i \in I} A_i$ und es gibt keine endliche Teilüberdeckung. Wir setzen für $i \in I$

$$J_i = \{j \in \mathbb{N} : B_j \subseteq A_i\} \subseteq \mathbb{N}$$

sowie $J := \bigcup_{i \in I} J_i \subseteq \mathbb{N}$. Damit ist $A_i = \bigcup_{j \in J_i} B_j$, also

$$K \subseteq \bigcup_{i \in I} A_i = \bigcup_{i \in I} \bigcup_{j \in J_i} B_j = \bigcup_{j \in J} B_j.$$

Klar ist, dass nun $B_j \in \mathcal{O}, j \in J$ eine abzählbare Überdeckung von \bar{K} darstellt. Da es keine endliche Teilüberdeckung für die $A_i, i \in I$ gibt, kann es auch keine endliche Teilüberdeckung für $B_j, j \in J$ geben. (Jedes B_j ist schließlich Teilmenge eines A_i 's.) Für $n \in \mathbb{N}$ setzen wir $\omega_n \in \bar{K} \setminus \bigcup_{j \in J, j \leq n} B_j$. Nach Voraussetzung hat die Folge $\omega_1, \omega_2, \dots \in K$ einen Häufungspunkt $\omega \in \bar{K}$. Da $\bar{K} \subseteq \bigcup_{j \in J} B_j$, gibt es $k \in J \subseteq \mathbb{N}$ mit $\omega \in B_k$. Damit liegen einerseits (da B_k offen ist) unendlich viele der ω_n in B_k , andererseits ist $\omega_i \notin B_k$ für alle $i \geq k$ nach Konstruktion. Dies ist ein Widerspruch, also ist \bar{K} kompakt.

'4. \Leftarrow 3.' falls (Ω, r) vollständig ist: Sei $\omega_1, \omega_2, \dots \in K$. Wir konstruieren eine Teilfolge, die Cauchy-Folge ist. Diese konvergiert, da (Ω, r) vollständig ist, und K ist als relativ folgenkompakt erkannt. Wähle eine Folge $\varepsilon_1, \varepsilon_2, \dots > 0$ mit $\varepsilon_n \downarrow 0$. Da K total beschränkt ist, gibt es endlich viele ε_1 -Bälle, die K überdecken. Mindestens einer dieser Bälle muss unendlich viele der ω_n enthalten. Diese haben jeweils höchstens Abstand $2\varepsilon_1$. Wähle ω_{k_1} als einer dieser unendlich vielen Punkte. Da dieser ε_1 -Ball durch endlich viele ε_2 -Bälle überdeckt wird, gibt es einen dieser ε_2 -Bälle, der unendlich viele der ω_n enthält. Diese haben jeweils höchstens Abstand $2\varepsilon_2$. Wähle $\omega_{k_2} \neq \omega_{k_1}$ als einen dieser unendlich vielen Punkten. Durch weiteres Vorgehen erhalten wir eine Folge $\omega_{k_1}, \omega_{k_2}, \dots \in K$, so dass $r(\omega_{k_n}, \omega_{k_m}) \leq 2\varepsilon_{m \wedge n}$. Mit anderen Worten haben wir wie angekündigt eine Cauchy-Folge in K gefunden. \square

Beispiel 1.11 (Kompakte metrische Räume sind polnisch). Sei (Ω, r) ein metrischer Raum und \mathcal{O} die von r erzeugte Topologie. Falls Ω kompakt ist, dann ist (Ω, \mathcal{O}) polnisch.

Wir zeigen, dass r vollständig ist. Sei $\omega_1, \omega_2, \dots \in \Omega$ eine Cauchy-Folge. Da K relativ folgenkompakt nach Proposition 1.9 ist, gibt es eine konvergente Teilfolge $\omega_{k_1}, \omega_{k_2}, \dots$. Sei $\omega \in \Omega$ der Grenzwert der konvergenten Teilfolge. Sei $\varepsilon > 0$ und $N \in \mathbb{N}$, so dass $r(\omega_m, \omega_n) < \varepsilon/2$ für $m, n > N$ und $r(\omega_{k_n}, \omega) < \varepsilon/2$ für $k_n > N$. Dann gilt für $m > N$, dass $r(\omega_m, \omega) \leq r(\omega_m, \omega_{k_n}) + r(\omega_{k_n}, \omega) \leq \varepsilon$. Daraus folgt, dass $\omega_n \xrightarrow{n \rightarrow \infty} \omega$.

Für die Separabilität von (Ω, \mathcal{O}) sei $\varepsilon_1, \varepsilon_2, \dots > 0$ mit $\varepsilon_n \downarrow 0$. Da K total beschränkt ist, gibt es für alle $n \in \mathbb{N}$ ein k_n und $\omega_{n1}, \dots, \omega_{nk_n}$ mit $K \subseteq \bigcup_{k=1}^{k_n} B_{\varepsilon_n}(\omega_{nk_n})$. Sei $\Omega' = \{\omega_{nk} : n \in \mathbb{N}, k = 1, \dots, k_n\}$. Dann ist Ω' abzählbar und für jedes $\omega \in \Omega$ und jedes $n \in \mathbb{N}$ gibt es ein $k(\omega, n) \in \{1, \dots, k_n\}$ mit $r(\omega_{k(\omega, n)}, \omega) < \varepsilon_n$. Damit gilt $(\omega_{k(\omega, n)}, \omega) \xrightarrow{n \rightarrow \infty} \omega$. Also ist $\overline{\Omega'} = \Omega$.

2 Mengensysteme

Die Wahrscheinlichkeitstheorie formalisiert das umgangssprachlich verwendete Wort *wahrscheinlich*. Dies ist (im weitesten Sinne) eine Eigenschaft eines möglichen Ausgangs eines Experimentes. Grundlegend in der Wahrscheinlichkeitstheorie ist der Begriff des *Ereignisses*, der alles beschreiben soll, was bei dem Experiment passieren kann. Ereignisse werden durch Teilmengen eines abstrakten Grundraumes, der immer Ω genannt wird, dargestellt. Ziel dieses Abschnittes ist es, möglichst vielen Teilmengen von Ω eine Wahrscheinlichkeit zuzuordnen. Dies führt auf den Begriff der σ -Algebra, denn diese enthalten genau die Teilmengen des Grundraumes, denen dann im nächsten Abschnitt Wahrscheinlichkeiten zugeordnet werden. Mit anderen Worten werden Elemente von σ -Algebren Ereignisse im obigen Sinne sein. Die anderen in diesem Abschnitt eingeführten Mengensysteme werden dazu dienen, geeignete σ -Algebren zu definieren.

2.1 Halbringe, Ringe und σ -Algebren

Die in diesem Abschnitt eingeführten Begriffe haben einen einfachen Zusammenhang. Ist nämlich $\mathcal{C} \subseteq 2^\Omega$, so gilt

$$\mathcal{C} \text{ ist } \sigma\text{-Algebra} \implies \mathcal{C} \text{ ist Ring} \implies \mathcal{C} \text{ ist Halbring.}$$

Beziehungen zwischen den Mengensystemen sind in Tabelle 2.1 festgehalten.

Definition 2.1 (Halbring, Ring, σ -Algebra).

1. Ein Mengensystem \mathcal{H} heißt *schnittstabil*, falls mit $A, B \in \mathcal{H}$ auch $A \cap B \in \mathcal{H}$ gilt. Es heißt *vereinigungsstabil*, falls mit $A, B \in \mathcal{H}$ auch $A \cup B \in \mathcal{H}$ gilt. Es heißt *komplementstabil*, wenn mit $A \in \mathcal{H}$ auch $A^c \in \mathcal{H}$ gilt. Es heißt *differenzmengenstabil*, falls mit $A, B \in \mathcal{H}$ auch $B \setminus A \in \mathcal{H}$ gilt.
2. Ein nicht-leeres Mengensystem \mathcal{H} ist ein Halbring (auf Ω), falls es (i) *schnittstabil* ist, und (ii) wenn es für $A, B \in \mathcal{H}$ Mengen $C_1, \dots, C_n \in \mathcal{H}$ gibt mit⁵ $B \setminus A = \biguplus_{i=1}^n C_i$.
3. Ein Mengensystem \mathcal{R} heißt Ring (auf Ω), falls (i) mit $A, B \in \mathcal{R}$ ist auch $A \cup B \in \mathcal{R}$ und (ii) mit $A, B \in \mathcal{R}$ ist auch $B \setminus A \in \mathcal{R}$.

⁵Wir schreiben $A \uplus B$ für $A \cup B$, falls $A \cap B = \emptyset$.

	\mathcal{C} Halbring	\mathcal{C} Ring	\mathcal{C} σ -Algebra
\mathcal{C} schnittstabil	•	◦	◦
\mathcal{C} σ -schnittstabil			◦
\mathcal{C} vereinigungsstabil		•	◦
\mathcal{C} σ -vereinigungsstabil			•
\mathcal{C} differenzmengenstabil		•	◦
\mathcal{C} komplementstabil			•
$B \setminus A = \bigsqcup_{i=1}^n C_i$	•	◦	◦
$\Omega \in \mathcal{C}$			◦

Tabelle 2.1: Für $\mathcal{C} \subseteq 2^\Omega$ wird hier die Beziehung zwischen Halbringen, Ringen und σ -Algebren dargestellt. Ein • bedeutet, dass in der Definition des Mengensystems (Spalte) die entsprechende Eigenschaft (Zeile) gefordert ist. Ein ◦ bedeutet, dass die entsprechende Eigenschaft aus den definierenden Eigenschaften des Mengensystems folgt.

4. Ein Mengensystem \mathcal{F} heißt σ -Algebra (auf Ω), falls mit $A, A_1, A_2, \dots \in \mathcal{F}$ auch $A^c, \bigcup_{n=1}^\infty A_n \in \mathcal{F}$ gilt. Dann heißt (Ω, \mathcal{F}) ein Messraum.

Bemerkung 2.2 (Beziehungen zwischen den und weitere Mengensystemen).

1. *Jeder Ring ist ein Halbring:* Um dies einzusehen, ist nur nachzuprüfen, dass jeder Ring \mathcal{R} schnittstabil ist. Dies folgt aus der Differenzmengenstabilität, da mit $A, B \in \mathcal{R}$ auch $A \cap B = A \setminus (A \setminus B) \in \mathcal{R}$.
2. *Jede σ -Algebra ist schnittstabil und damit ein Ring:* Das ist klar, da $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c$ und $B \setminus A = B \cap A^c$.
3. *Algebra und Verband:* Eine *Algebra* ist ein vereinigungs- und komplementstabiles Mengensystem. Ein schnitt- und vereinigungsstabiles Mengensystem heißt *Verband*. Unsere Darstellung kommt jedoch ohne diese Begriffe aus.

Beispiel 2.3 (Halbringe, σ -Algebren).

1. *Halboffene Intervalle bilden einen Halbring:* Sei $\Omega = \mathbb{R}$. Dann definiert

$$\mathcal{H} := \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$$

einen Halbring. Denn es gilt für $a_1 \leq b_1, a'_1 \leq b'_1$, dass⁶ $(a_1, b_1] \cap (a'_1, b'_1] = (a_1 \vee a'_1, b_1 \wedge b'_1]$ und $(a_1, b_1] \setminus (a'_1, b'_1] = (a_1, a'_1 \wedge b_1] \sqcup (b'_1, b_1]$, wobei $(a, b] = \emptyset$, falls $a \geq b$.

⁶Wie üblich bezeichnet $x \wedge y := \min(x, y)$ und $x \vee y := \max(x, y)$

2. *Einfache Beispiele für σ -Algebren:* Die einfachsten Beispiele für σ -Algebren sind $\{\emptyset, \Omega\}$ und 2^Ω . (Beides sind übrigens auch Topologien.) Für überabzählbares Ω ist ein nicht ganz so triviales Beispiel für eine σ -Algebra das Mengensystem

$$\{A \subseteq \Omega : A \text{ abzählbar oder } A^c \text{ abzählbar}\}.$$

(Dieses Mengensystem ist kein Topologie.)

Ein weiteres Beispiel werden wir in Abschnitt 4.1 antreffen: Falls \mathcal{F}' eine σ -Algebra auf Ω' ist und $f : \Omega \rightarrow \Omega'$. Dann ist

$$\sigma(f) := \{f^{-1}(A') : A' \in \mathcal{F}'\} \subseteq 2^\Omega \quad (2.1)$$

eine σ -Algebra auf Ω . Ist nämlich $A', A'_1, A'_2, \dots \in \sigma(f)$, so ist $(f^{-1}(A'))^c = f^{-1}((A')^c) \in \sigma(f)$ und $\bigcup_{n=1}^{\infty} f^{-1}(A'_n) = f^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) \in \sigma(f)$.

In der Praxis am weitaus häufigsten benötigt man Borel'sche σ -Algebren; siehe Definition 2.7.

2.2 Erzeuger und Erweiterungen

Auf der einen Seite ist es Ziel der Maßtheorie, Mengenfunktionen auf σ -Algebren zu definieren. Auf der anderen Seite kann man oftmals nur Halbringe konkret angeben, nicht jedoch σ -Algebren; siehe Beispiel 2.3.1. Allerdings gibt es für jeden Halbring \mathcal{H} einen kleinsten Ring, der \mathcal{H} enthält, $\mathcal{R}(\mathcal{H})$. Genauso gibt es eine kleinste σ -Algebra \mathcal{F} , die einen Halbring \mathcal{H} (oder einen Ring \mathcal{R}) enthält, $\sigma(\mathcal{H})$. Damit erzeugt jeder Halbring \mathcal{H} eine σ -Algebra.

Bemerkung 2.4 (Erzeugte Mengensysteme). Leicht überzeugt man sich, dass der Schnitt von Ringen (σ -Algebren) wieder ein Ring (σ -Algebra) ist. Insbesondere benötigen wir folgende Begriffe:

Sei $\mathcal{C} \subseteq 2^\Omega$, dann ist

$$\mathcal{R}(\mathcal{C}) := \bigcap \left\{ \mathcal{R} \supseteq \mathcal{C} : \mathcal{R} \text{ Ring} \right\}$$

der von \mathcal{C} erzeugte Ring und

$$\sigma(\mathcal{C}) := \bigcap \left\{ \mathcal{F} \supseteq \mathcal{C} : \mathcal{F} \text{ } \sigma\text{-Algebra} \right\}$$

die von \mathcal{C} erzeugte σ -Algebra. Klar ist, dass $\mathcal{R}(\mathcal{R}(\mathcal{H})) = \mathcal{R}(\mathcal{H})$ und $\sigma(\sigma(\mathcal{H})) = \sigma(\mathcal{H})$.

Lemma 2.5 (Von Halbring erzeugter Ring). Sei \mathcal{H} ein Halbring. Dann ist

$$\mathcal{R}(\mathcal{H}) = \left\{ \biguplus_{k=1}^n A_k : A_1, \dots, A_n \in \mathcal{H} \text{ disjunkt}, n \in \mathbb{N} \right\}$$

der von \mathcal{H} erzeugte Ring.

Beispiel 2.6 (Von Intervallen erzeugter Ring). Sei \mathcal{H} der von halboffenen Intervallen erzeugte Halbring aus Beispiel 2.3. Dann ist also

$$\mathcal{R}(\mathcal{H}) = \left\{ \biguplus_{k=1}^n (a_k, b_k] : a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{Q}, \right. \\ \left. a_k < b_k, k = 1, \dots, n \text{ und } a_k < b_{k+1}, k = 1, \dots, n-1 \right\}$$

der von \mathcal{H} erzeugte Ring.

Beweis von Lemma 2.5. Klar ist, dass $\mathcal{R}(\mathcal{H})$ schnittstabil ist. Um zu zeigen, dass $\mathcal{R}(\mathcal{H})$ ein Ring ist, zeigen wir zunächst die Differenzmengenstabilität. Sei $A_1, \dots, A_n \in \mathcal{H}$ und $B_1, \dots, B_m \in \mathcal{H}$, jeweils disjunkt. Dann gilt

$$\left(\biguplus_{i=1}^n A_i \right) \setminus \left(\biguplus_{j=1}^m B_j \right) = \biguplus_{i=1}^n \bigcap_{j=1}^m A_i \setminus B_j \in \mathcal{R}(\mathcal{H}).$$

Um die Vereinigungsstabilität von $\mathcal{R}(\mathcal{H})$ zu zeigen sei $A, B \in \mathcal{R}(\mathcal{H})$. Dann ist auch $A \cup B = (A \cap B) \uplus (A \setminus B) \uplus (B \setminus A) \in \mathcal{R}(\mathcal{H})$, da Schnitt- und Differenzmengenstabilität schon gezeigt sind.

Es gibt keinen kleineren Ring, der \mathcal{H} enthält. Schließlich müsste dieser vereinigungsstabil sein, und ist damit eine Obermenge von $\mathcal{R}(\mathcal{H})$. \square

Definition 2.7 (Borel'sche σ -Algebra). Sei (Ω, \mathcal{O}) ein topologischer Raum. Dann bezeichnet man mit $\mathcal{B}(\Omega) := \sigma(\mathcal{O})$ die Borel'sche σ -Algebra auf Ω . Ist $\Omega \subseteq \mathbb{R}^d$, so bezeichnen wir mit $\mathcal{B}(\Omega)$ die Borel'sche σ -Algebra, die von der euklidischen Topologie auf \mathbb{R}^d erzeugt wird. Ist $\Omega \subseteq \overline{\mathbb{R}}$, so ist $\mathcal{B}(\Omega)$ die Borel'sche σ -Algebra, die von der Topologie aus Beispiel 1.2 erzeugt wird. Mengen in $\mathcal{B}(\overline{\mathbb{R}})$ bezeichnet man auch als (Borel-)messbare Mengen.

Lemma 2.8 (Abzählbare Basis und Borel'sche σ -Algebra). Sei (Ω, \mathcal{O}) ein topologischer Raum mit abzählbarer Basis $\mathcal{C} \subseteq \mathcal{O}$. Dann gilt $\sigma(\mathcal{O}) = \sigma(\mathcal{C})$. Insbesondere ist Ω separabel.

Beweis. Wir müssen nur zeigen, dass $\mathcal{O} \subseteq \sigma(\mathcal{C})$. Dies ist jedoch klar, da $A \in \mathcal{O}$ als abzählbare Vereinigung von Mengen aus \mathcal{C} dargestellt werden kann. Siehe Lemma 1.5. \square

Lemma 2.9 (Borel'sche σ -Algebra wird von Intervallen erzeugt). Sei

$$\begin{aligned} \mathcal{C}_1 &= \{[-\infty, b] : b \in \mathbb{Q}\} \text{ oder} \\ \mathcal{C}_2 &= \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \\ \mathcal{C}_3 &= \{(a, b) : a, b \in \mathbb{Q}, a \leq b\} \\ \mathcal{C}_4 &= \{[a, b) : a, b \in \mathbb{Q}, a \leq b\}. \end{aligned}$$

Dann ist $\sigma(\mathcal{C}_i) = \mathcal{B}(\overline{\mathbb{R}})$, $i = 1, \dots, 4$.

Beweis. Das Mengensystem \mathcal{C}_3 ist eine abzählbare Basis der euklidischen Topologie auf $\overline{\mathbb{R}}$. Für diesen Fall folgt die Aussage aus Lemma 2.8.

Wir zeigen die Aussage nur für \mathcal{C}_1 und \mathcal{C}_2 , die für \mathcal{C}_4 folgt analog. Zunächst ist $\mathcal{C}_2 := \{A \setminus B : A, B \in \mathcal{C}_1\} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \subseteq \sigma(\mathcal{C}_2)$ der von \mathcal{C}_1 erzeugte Halbring aus Beispiel 2.3. Damit ist $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$ und es genügt zu zeigen, dass $\sigma(\mathcal{C}_2) = \mathcal{B}(\overline{\mathbb{R}})$.

Sei hierzu \mathcal{O} wie in Definition 1.1.8 mit $\Omega = \mathbb{R}$. Wir zeigen (i) aus $A \in \mathcal{O}$ folgt, dass $A \in \sigma(\mathcal{C}_2)$, und (ii) aus $A \in \mathcal{C}_2$ folgt, dass $A \in \sigma(\mathcal{O})$. Daraus folgt dann $\mathcal{O} \subseteq \sigma(\mathcal{C}_2) \subseteq \sigma(\mathcal{O})$, also $\sigma(\mathcal{O}) = \sigma(\mathcal{C}_2)$.

Für (i) sei also $A \in \mathcal{O}$. Wir behaupten

$$A = \bigcup \{(a, b] : (a, b] \subseteq A, a, b \in \mathbb{Q}\}, \quad (2.2)$$

und bemerken, dass die rechte Seite ein Element von $\sigma(\mathcal{C}_2)$ ist. In (2.2) ist '⊇' klar. Um '⊆' einzusehen, wählen wir $x \in A$. Dann gibt es nach Definition von \mathcal{O} ein $\varepsilon > 0$, so dass

$B_\varepsilon(x) \subseteq A$. Allerdings gibt es auch $a, b \in \mathbb{Q}$ mit $a \leq b$ und $x \in (a, b) \subseteq B_\varepsilon(x)$. Damit ist '⊆' gezeigt und (i) folgt.

Für (ii) gehen wir ähnlich vor; sei $A \in \mathcal{C}_2$. Dann ist offenbar

$$A = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right).$$

Da $(a, b + \frac{1}{n}) \in \mathcal{O}$, ist also $A \in \sigma(\mathcal{O})$. □

Beispiel 2.10 (Borel-messbare Mengen). Natürlich sind alle abzählbaren Durchschnitte und Vereinigungen von Intervallen nach Lemma 2.9 in $\mathcal{B}(\overline{\mathbb{R}})$. Sei beispielsweise

$$\begin{aligned} A_1 &= [0, \frac{1}{3}] \cup [\frac{2}{3}, 1], \\ A_2 &= [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{3}{9}] \cup [\frac{6}{9}, \frac{7}{9}] \cup [\frac{8}{9}, 1], \\ A_3 &= [0, \frac{1}{27}] \cup [\frac{2}{27}, \frac{3}{27}] \cup [\frac{6}{27}, \frac{7}{27}] \cup [\frac{8}{27}, \frac{9}{27}] \cup [\frac{18}{27}, \frac{19}{27}] \cup [\frac{20}{27}, \frac{21}{27}] \cup [\frac{24}{27}, \frac{25}{27}] \cup [\frac{26}{27}, 1], \\ &\dots \end{aligned}$$

dann bezeichnet $A = \bigcap_{n=1}^{\infty} A_n$ das Cantor'sche Diskontinuum. Diese Menge ist als abzählbarer Schnitt von endlichen Vereinigungen von Intervallen messbar. Wir werden in Beispiel 3.26 ein Beispiel für eine nicht Borel-messbare Menge kennen lernen.

2.3 Dynkin-Systeme

Oftmals muss man in der Maßtheorie zeigen, dass ein bestimmtes Mengensystem \mathcal{F} eine σ -Algebra ist und einen Halbring \mathcal{H} beinhaltet. Die in diesem Abschnitt behandelten Dynkin-Systeme sind dabei sehr hilfreich. Es genügt nämlich wegen Theorem 2.13 zu zeigen, dass \mathcal{F} ein schnittstabiles Dynkin-System mit $\mathcal{H} \subseteq \mathcal{F}$ ist. Dies ist oftmals einfacher, als direkt zu zeigen, dass \mathcal{F} eine σ -Algebra ist.

Definition 2.11 (Dynkin-System). Ein Mengensystem \mathcal{D} heißt Dynkin-System (auf Ω), falls (i) $\Omega \in \mathcal{D}$, (ii) mit $A, B \in \mathcal{D}$ und $A \subseteq B$ ist auch $B \setminus A \in \mathcal{D}$ und (iii) mit $A_1, A_2, \dots \in \mathcal{D}$ mit $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ ist auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$.

Beispiel 2.12 (σ -Algebren sind Dynkin-Systeme). Jede σ -Algebra ist ein Dynkin-System: Sei \mathcal{F} eine σ -Algebra. Dann ist mit $A, B \in \mathcal{F}$ auch $A^c \in \mathcal{F}$ und damit $\Omega = A \cup A^c \in \mathcal{F}$ sowie $B \setminus A = B \cap A^c \in \mathcal{F}$.

Theorem 2.13 (Schnittstabile Dynkin-Systeme). Sei \mathcal{D} ein Dynkin-System und $\mathcal{C} \subseteq \mathcal{D}$ schnittstabil. Dann ist $\sigma(\mathcal{C}) \subseteq \mathcal{D}$. Insbesondere ist jedes schnittstabile Dynkin-System eine σ -Algebra.

Beweis. Genau wie in Lemma 2.4 definiert man das von \mathcal{C} erzeugte Dynkin-System $\lambda(\mathcal{C})$. Da offenbar der Schnitt von Dynkin-Systemen wieder ein Dynkin-System ist, gilt $\lambda(\mathcal{C}) \subseteq \mathcal{D}$. Wir werden zeigen, dass $\lambda(\mathcal{C})$ eine σ -Algebra ist, denn dann ist $\sigma(\mathcal{C}) \subseteq \sigma(\lambda(\mathcal{C})) = \lambda(\mathcal{C}) \subseteq \mathcal{D}$.

Wir zeigen hierzu, dass $\lambda(\mathcal{C})$ schnittstabil ist. Dann ist nämlich für $A_1, A_2, \dots \in \lambda(\mathcal{C})$

$$\bigcup_{i=1}^n A_i = \left(\bigcap_{i=1}^n A_i^c \right)^c \in \lambda(\mathcal{C})$$

und, da $\lambda(\mathcal{C})$ ein Dynkin-System ist $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \bigcup_{i=1}^n A_i \in \lambda(\mathcal{C})$.

Wir müssen also zeigen, dass mit $A, B \in \lambda(\mathcal{C})$ auch $A \cap B \in \lambda(\mathcal{C})$ gilt. Falls $A, B \in \mathcal{C}$, so ist dies wegen der Schnittstabilität von \mathcal{C} klar. Für $B \in \mathcal{C}$ setzen wir

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

Dann ist \mathcal{D}_B ein Dynkin-System, da (i) $\Omega \in \mathcal{D}_B$, (ii) für $A, C \subseteq \mathcal{D}_B$ ist $A \cap B, C \cap B \in \lambda(\mathcal{C})$ also mit $A \subseteq C$ auch $A \cap B \subseteq C \cap B$, also $(C \setminus A) \cap B = (C \cap B) \setminus (A \cap B) \in \lambda(\mathcal{C})$ und (iii) für $A_1, A_2, \dots \in \mathcal{D}_B$ ist $A_1 \cap B, A_2 \cap B, \dots \in \lambda(\mathcal{C})$ und falls $A_1 \subseteq A_2 \subseteq \dots$ ist $A_1 \cap B \subseteq A_2 \cap B \subseteq \dots$, woraus $\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B = \left(\bigcup_{n=1}^{\infty} A_n \cap B\right) \in \lambda(\mathcal{C})$ folgt.

Da $\mathcal{C} \subseteq \mathcal{D}_B$ und \mathcal{D}_B ein Dynkin-System ist, ist auch $\lambda(\mathcal{C}) \subseteq \mathcal{D}_B$. Dies bedeutet, dass für $A \in \lambda(\mathcal{C})$ und $B \in \mathcal{C}$ auch $A \cap B \in \lambda(\mathcal{C})$. Wir setzen nun für ein $A \in \lambda(\mathcal{C})$

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

Wie oben zeigt man, dass \mathcal{B}_A ein Dynkin-System mit $\mathcal{C} \subseteq \mathcal{B}_A$ ist. Damit ist wieder $\lambda(\mathcal{C}) \subseteq \mathcal{B}_A$. Insbesondere ist mit $A, B \in \lambda(\mathcal{C})$ auch $A \cap B \in \lambda(\mathcal{C})$ und die erste Behauptung ist gezeigt. Die zweite Behauptung folgt, wenn man $\mathcal{C} := \mathcal{D}$ setzt. \square

2.4 Kompakte Systeme

Nachdem wir im Abschnitt 1.2 kompakte Teilmengen von Ω kennen gelernt haben, stellen wir nun einen wichtigen Zusammenhang zwischen Systemen kompakter Mengen und der Maßtheorie vor. Solche kompakten Systeme spielen eine wichtige Rolle beim Beweis von Theorem 3.9. Hier wird gezeigt, dass aus der Additivität einer Mengenfunktion und einer Approximations-eigenschaft bzgl. eines kompakten Systems die σ -Additivität der Mengenfunktion folgt.

Definition 2.14 (Kompaktes System). *Ein schnittstabiles Mengensystem \mathcal{K} heißt kompaktes System (auf Ω), falls aus $\bigcap_{n=1}^{\infty} K_n = \emptyset$ mit $K_1, K_2, \dots \in \mathcal{K}$ folgt, dass es ein $N \in \mathbb{N}$ gibt mit $\bigcap_{n=1}^N K_n = \emptyset$.*

Beispiel 2.15 (Kompakte Mengen). *Kompakte Mengen bilden ein kompaktes System:* Sei (Ω, r) ein metrischer Raum und \mathcal{O} die von r erzeugte Topologie. Dann ist jedes schnittstabile $\mathcal{K} \subseteq \{K \subseteq \Omega : K \text{ kompakt}\}$ ein kompaktes System. Denn: sei $\bigcap_{n=1}^{\infty} K_n = \emptyset$. Dann ist sowohl K_1 , als auch $L_n := K_1 \cap K_n \subseteq K_1$ für $n = 1, 2, \dots$ nach Lemma 1.8 abgeschlossen und wegen der Kompaktheit von K_1 gibt es ein N mit $\bigcap_{n=1}^N K_n = \emptyset$ nach Proposition 1.9.

Lemma 2.16 (Erweiterung kompakter Systeme). *Sei \mathcal{K} ein kompaktes System. Dann ist auch*

$$\mathcal{K}_{\cup} := \left\{ \bigcup_{i=1}^n K_i : K_1, \dots, K_n \in \mathcal{K}, n \in \mathbb{N} \right\}$$

ein kompaktes System.

Beweis. Klar ist, dass \mathcal{K}_{\cup} schnittstabil ist. Seien $L_1, L_2, \dots \in \mathcal{K}_{\cup}$ mit $\bigcap_{n=1}^N L_n \neq \emptyset$ für alle $N \in \mathbb{N}$. Zu zeigen ist, dass dann auch $\bigcap_{n=1}^{\infty} L_n \neq \emptyset$. Dazu zeigen wir mit Induktion über N folgende Aussage:

Es gibt für jedes $N \in \mathbb{N}$ Mengen $K_1, \dots, K_N \in \mathcal{K}$ mit $K_n \subseteq L_n$, $n = 1, \dots, N$, so dass für alle $k \in \mathbb{N}_0$ gilt dass $K_1 \cap \dots \cap K_N \cap L_{N+1} \cap \dots \cap L_{N+k} \neq \emptyset$.

Sei $N = 1$. Da $L_1 = \bigcup_{j=1}^{m_1} K'_j$ für $K'_1, \dots, K'_{m_1} \in \mathcal{K}$ und $\bigcap_{n=1}^N L_n = \bigcup_{j=1}^{m_1} K'_j \cap \bigcap_{n=1}^N L_n \neq \emptyset$ für jedes $N \in \mathbb{N}$, gibt es ein $j = 1, \dots, m_1$, so dass $K'_j \cap \bigcap_{n=1}^N L_n \neq \emptyset$ für jedes $N \in \mathbb{N}$. Setze $K_1 := K'_j$, und die Behauptung ist für $N = 1$ gezeigt.

Gelte die Behauptung für $N - 1$. Wieder ist $L_N = \bigcup_{j=1}^{m_N} K''_j$ für $K''_1, \dots, K''_{m_N} \in \mathcal{K}$. Damit gilt nach der Induktionsvoraussetzung für alle $k \in \mathbb{N}$, dass

$$\begin{aligned} K_1 \cap \dots \cap K_{N-1} \cap \left(\bigcup_{j=1}^{m_N} K''_j \right) \cap L_{N+1} \cap \dots \cap L_{N+k} \\ = \bigcup_{j=1}^{m_N} K_1 \cap \dots \cap K_{N-1} \cap K''_j \cap L_{N+1} \cap \dots \cap L_{N+k} \neq \emptyset. \end{aligned}$$

Damit gibt es ein j , so dass $K_1 \cap \dots \cap K_{N-1} \cap K''_j \cap L_{N+1} \cap \dots \cap L_{N+k} \neq \emptyset$ für alle $k \in \mathbb{N}$. Setze $K_N := K''_j$, was den Induktionsbeweis abschließt.

Setzt man $k = 0$ in obiger Behauptung, sieht man, dass es $K_1, K_2, \dots \in \mathcal{K}$ und $K_n \subseteq L_n$, $n \in \mathbb{N}$ gibt mit $\bigcap_{n=1}^N K_n \neq \emptyset$ für alle $N \in \mathbb{N}$. Da \mathcal{K} ein kompaktes System ist, gilt insbesondere

$$\emptyset \neq \bigcup_{n=1}^{\infty} K_n \subseteq \bigcup_{n=1}^{\infty} L_n$$

und die Behauptung ist gezeigt. \square

3 Maße

Will man Mengen *messen*, also etwa jeder Menge ihren Inhalt zuordnen, scheinen einige Forderungen natürlich. Dies ist aus der Stochastik bekannt, wo 'messen' bedeutet, dass einer Menge eine Wahrscheinlichkeit zugeordnet wird. Etwa sollte die leere Menge (ein Ereignis, das nie eintritt) *Masse* 0 zugeordnet werden, oder die Massen sollten sich abzählbar additiv verhalten, siehe (3.1). Zentral in der Wahrscheinlichkeitstheorie ist deshalb der Begriff des Wahrscheinlichkeitsmaßes. Wie sich herausstellt, sind Maße auf σ -Algebren zu definieren, damit die Forderung der abzählbaren Additivität erfüllt werden kann. In diesem Abschnitt geben wir die wichtigsten Schritte an, solche Maße zu konstruieren. Als wichtigstes Beispiel eines Maßes dient dabei das Lebesgue-Maß. Dies ist zwar kein Wahrscheinlichkeitsmaß, jedoch funktioniert seine Konstruktion analog zur Konstruktion von Wahrscheinlichkeitsmaßen.

3.1 Mengenfunktionen

Wir werden nun Funktionen $\mu : \mathcal{C} \rightarrow \overline{\mathbb{R}}_+$ betrachten, wenn \mathcal{C} ein Halbring, Ring oder eine σ -Algebra ist. Für die Stochastik am wichtigsten ist dabei sicherlich der Begriff des *Wahrscheinlichkeitsmaßes*, das den Spezialfall $\mu(\Omega) = 1$ beschreibt.

Definition 3.1 (Maß und äußeres Maß). Sei $\mathcal{F} \subseteq 2^\Omega$ und $\mu : \mathcal{F} \rightarrow \overline{\mathbb{R}}_+$.

1. Die Abbildung μ heißt endlich additiv, falls für disjunkte $A_1, \dots, A_n \in \mathcal{F}$ mit $\biguplus_{k=1}^n A_k \in \mathcal{F}$ gilt, dass

$$\mu\left(\biguplus_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k). \quad (3.1)$$

Sie heißt subadditiv, falls für (beliebige) $A_1, \dots, A_n \in \mathcal{F}$ mit $\bigcup_{k=1}^n A_k \in \mathcal{F}$ gilt, dass

$$\mu\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n \mu(A_k). \quad (3.2)$$

2. Eine Abbildung $\mu : \mathcal{F} \rightarrow \mathbb{R}_+$ heißt σ -additiv, falls (3.1) auch für $n = \infty$ gilt. Sie heißt σ -sub-additiv, wenn (3.2) auch für $n = \infty$ gilt. Sie heißt monoton, wenn aus $A, B \in \mathcal{F}$ mit $A \subseteq B$ folgt, dass $\mu(A) \leq \mu(B)$.
3. Falls es eine Folge $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ mit $\bigcup_{n=1}^{\infty} \Omega_n = \Omega$ und $\mu(\Omega_n) < \infty$ für alle $n = 1, 2, \dots$ gilt, so heißt μ σ -endlich.
4. Sei \mathcal{F} eine σ -Algebra und $\mu : \mathcal{F} \rightarrow \mathbb{R}_+$. Ist μ auch σ -additiv, so heißt μ ein Maß (auf \mathcal{F}) und $(\Omega, \mathcal{F}, \mu)$ ist ein Maßraum. Gilt $\mu(\Omega) < \infty$, so heißt μ endliches Maß und falls $\mu(\Omega) = 1$, so heißt μ ein Wahrscheinlichkeitsmaß oder eine Wahrscheinlichkeitsverteilung oder auch einfach eine Verteilung. Außerdem heißt $(\Omega, \mathcal{F}, \mu)$ dann ein Wahrscheinlichkeitsraum.
5. Sei (Ω, \mathcal{O}) ein topologischer Raum und μ ein Maß auf $\mathcal{B}(\mathcal{O})$. Dann heißt die kleinste abgeschlossene Menge F (d.h. $F^c \in \mathcal{O}$) mit $\mu(F^c) = 0$ der Träger von μ .
6. Eine σ -subadditiv, monotone Abbildung $\mu^* : 2^\Omega \rightarrow \mathbb{R}_+$ heißt äußeres Maß, falls $\mu^*(\emptyset) = 0$. Eine Menge $A \subseteq \Omega$ heißt μ^* -messbar, falls

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^c) \quad (3.3)$$

für alle $E \subseteq \Omega$ gilt.

7. Sei \mathcal{F} schnittstabil und $\mathcal{K} \subseteq \mathcal{F}$ ein kompaktes System. Dann heißt μ von innen \mathcal{K} -regulär, falls für alle $A \in \mathcal{K}$

$$\mu(A) = \sup_{\mathcal{K} \ni K \subseteq A} \mu(K).$$

Beispiel 3.2 (Beispiele von Mengenfunktionen). 1. Wir werden uns oft mit Mengenfunktionen auf $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ aus 2.3 beschäftigen. Etwa definiert $\mu((a, b]) = b - a$ eine additive, σ -endliche Mengenfunktion. Wir werden diese Funktion in eindeutiger Weise auf die Borel'sche σ -Algebra $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{H})$ (siehe Lemma 2.9) erweitern und so das Lebesgue-Maß konstruieren, siehe Korollar 3.17.

2. Einfache Beispiele für Mengenfunktionen sind die Dirac-Maße. Ist $\omega' \in \Omega$, so ist

$$\delta_{\omega'} : \begin{cases} 2^\Omega & \rightarrow \{0, 1\} \\ A & \mapsto 1_{\{\omega' \in A\}} \end{cases}$$

ein (Wahrscheinlichkeits-)Maß.

3. Für $\mu_i = \delta_{\omega_i}$, $i \in I$ heißt $\sum_{i \in I} \delta_{\omega_i}$ ein Zählmaß.

4. Sind $\mu_i, i \in I$ Maße auf einer σ -Algebra \mathcal{F} . Dann ist für $a_i \in \mathbb{R}_+, i \in I$, auch $\sum_{i \in I} a_i \mu_i$ ein Maß. Beispiele hierfür sind aus der Vorlesung *Stochastik* wohlbekannt. Dort war etwa mit $\mathcal{F} = 2^{\mathbb{N}_0}$ und δ_k wie in 2.

$$\mu_{\text{Poi}(\gamma)} := \sum_{k=0}^{\infty} e^{-\gamma} \frac{\gamma^k}{k!} \cdot \delta_k$$

die Poisson-Verteilung auf $2^{\mathbb{N}_0}$ mit Parameter γ ,

$$\mu_{\text{geo}(p)} := \sum_{k=1}^{\infty} (1-p)^{k-1} p \cdot \delta_k$$

die geometrische Verteilung mit Erfolgsparameter p und

$$\mu_{B(n,p)} := \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \cdot \delta_k$$

die Binomialverteilung $B(n, p)$.

Bemerkung 3.3 (Inhalte und Prämaße). Endlich additive Mengenfunktionen heißen oft *Inhalt*, σ -additive Mengenfunktionen, die nicht auf σ -Algebren definiert sind, heißen oft *Prämaße*. Die auf einer Borel'schen σ -Algebra definierten, bezüglich der kompakten Mengen von innen regulären Maße heißen Radon-Maße. Wir werden diese Begriffe nicht verwenden.

Lemma 3.4 (Mengenfunktionen auf Halbringen). Sei \mathcal{H} ein Halbring und $\mu : \mathcal{H} \rightarrow \overline{\mathbb{R}}_+$ endlich additiv. Dann ist μ monoton und subadditiv. Außerdem ist μ genau dann σ -additiv wenn es σ -subadditiv ist.

Beweis. Wir beginnen mit einer Argumentation, die wir häufiger brauchen werden. Für $A_1, A_2, \dots \in \mathcal{H}$ und $B_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i \right)$ und $n \geq 2$ gibt es $D_1, \dots, D_k \in \mathcal{H}$ disjunkt mit $B_n = \bigsqcup_{i=1}^k D_i$. Die Behauptung ist klar für $n = 2$, da \mathcal{H} ein Halbring ist. Gilt die Behauptung für ein n , so ist $B_{n+1} = A_{n+1} \setminus \left(\bigcup_{i=1}^n A_i \right) = \left(A_{n+1} \setminus \left(\bigcup_{i=2}^n A_i \right) \right) \setminus A_1$. Nach Voraussetzung gibt es D_1, \dots, D_k mit $A_{n+1} \setminus \left(\bigcup_{i=2}^n A_i \right) = \bigsqcup_{i=1}^k D_i$. Damit ist $B_{n+1} = \left(\bigsqcup_{i=1}^k D_i \right) \cap A_1^c = \bigsqcup_{i=1}^k (D_i \setminus A_1)$. Für alle $i = 1, \dots, k$ gibt es Mengen $E_{i1}, \dots, E_{ik_i} \in \mathcal{H}$ mit $D_i \setminus A_1 = \bigsqcup_{j=1}^{k_i} E_{ij}$. Damit gilt $B_{n+1} = \bigsqcup_{i=1}^k \bigsqcup_{j=1}^{k_i} E_{ij}$ und die Behauptung ist gezeigt.

Wir zeigen nun, dass μ monoton ist. Sei $A, B \in \mathcal{H}$ mit $A \subseteq B$. Dann gibt es disjunkte $D_1, \dots, D_k \in \mathcal{H}$ mit $B \setminus A = \bigsqcup_{i=1}^k D_i$, also $B = A \sqcup \bigsqcup_{i=1}^k D_i$ und damit wegen der Additivität von μ auch $\mu(B) = \mu(A) + \sum_{i=1}^k \mu(D_i) \geq \mu(A)$. Genauso zeigt man, dass aus $E_1, \dots, E_k \in \mathcal{H}$ disjunkt mit $\bigsqcup_{i=1}^k E_i \subseteq A \in \mathcal{H}$ folgt, dass $\sum_{i=1}^k \mu(E_i) \leq \mu(A)$.

Für die Subadditivität seien $A_1, \dots, A_n \in \mathcal{H}$ mit $\bigcup_{i=1}^n A_i \in \mathcal{H}$. Wir setzen $B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2), \dots$ Nach oben Gesagtem gibt es für jedes $i = 1, \dots, k$ ein $k_i \in \mathbb{N}$ und Mengen $D_{ij} \in \mathcal{H}, j = 1, \dots, k_i$ mit $B_i = \bigsqcup_{j=1}^{k_i} D_{ij}$. Damit gilt $\bigcup_{i=1}^n A_i = \bigsqcup_{i=1}^n B_i = \bigsqcup_{i=1}^n \bigsqcup_{j=1}^{k_i} D_{ij}$. Weiter ist $A_i = B_i \sqcup \bigcup_{j=1}^{i-1} A_j$, also

$$\mu \left(\bigcup_{i=1}^n A_i \right) = \sum_{i=1}^n \sum_{j=1}^{k_i} \mu(D_{ij}) \leq \sum_{i=1}^n \mu(A_i).$$

Wir zeigen nun μ ist σ -additiv $\iff \mu$ ist σ -sub-additiv.

' \implies ' folgt genauso wie die Subadditivität von μ . Für '←' sei $A_1, A_2, \dots \in \mathcal{H}$ disjunkt mit $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$. Da μ monoton ist, gilt

$$\sum_{n=1}^{\infty} \mu(A_n) = \lim_{N \rightarrow \infty} \sum_{n=1}^N \mu(A_n) = \lim_{N \rightarrow \infty} \mu\left(\biguplus_{n=1}^N A_n\right) \leq \mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

wegen der σ -Subadditivität. \square

Lemma 3.5 (Fortsetzung von Mengenfunktionen auf Halbringen). Sei \mathcal{H} ein Halbring, \mathcal{R} der von \mathcal{H} erzeugte Ring aus Lemma 2.5 und μ eine endlich additive Funktion auf \mathcal{H} . Definiere die Mengenfunktion $\tilde{\mu}$ auf \mathcal{R} durch

$$\tilde{\mu}\left(\biguplus_{i=1}^n A_i\right) := \sum_{i=1}^n \mu(A_i)$$

für $A_1, \dots, A_n \in \mathcal{H}$ disjunkt. Dann ist $\tilde{\mu}$ die einzige additive Fortsetzung von μ auf \mathcal{R} , die auf \mathcal{H} mit μ übereinstimmt. Außerdem ist $\tilde{\mu}$ genau dann σ -additiv, wenn μ σ -additiv ist.

Beweis. Es ist nur zu zeigen, dass $\tilde{\mu}$ wohldefiniert ist. Sei hierzu $A_1, \dots, A_m, B_1, \dots, B_n \in \mathcal{H}$ mit $\biguplus_{i=1}^m A_i = \biguplus_{j=1}^n B_j$. Da

$$A_i = \biguplus_{j=1}^n A_i \cap B_j, \quad B_j = \biguplus_{i=1}^m A_i \cap B_j,$$

gilt wegen der endlichen Additivität von $\tilde{\mu}$

$$\sum_{i=1}^m \mu(A_i) = \sum_{i=1}^m \sum_{j=1}^n \mu(A_i \cap B_j) = \sum_{j=1}^n \sum_{i=1}^m \mu(A_i \cap B_j) = \sum_{j=1}^n \mu(B_j). \quad \square$$

Proposition 3.6 (Einschluss-/Ausschlussformel). Sei μ eine additive Mengenfunktion auf einem Ring \mathcal{R} und I endlich. Dann gilt für $A_i \in \mathcal{R}$, $i \in I$, dass

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{J \subseteq I} (-1)^{|J|+1} \mu\left(\bigcap_{j \in J} A_j\right)$$

Für $I = \{1, 2\}$ gilt also

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2)$$

und für $I = \{1, 2, 3\}$

$$\begin{aligned} \mu(A_1 \cup A_2 \cup A_3) &= \mu(A_1) + \mu(A_2) + \mu(A_3) \\ &\quad - \mu(A_1 \cap A_2) - \mu(A_1 \cap A_3) - \mu(A_2 \cap A_3) + \mu(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Beweis. Wir verwenden Induktion uber $|I|$. Fur $|I| = 2$ ist die Behauptung wegen $A_1 \cup A_2 = A_1 \uplus (A_2 \setminus A_1)$ und $(A_2 \setminus A_1) \uplus (A_1 \cap A_2) = A_2$ klar. Gilt sie fur $I = \{1, \dots, n\}$, so ist wegen der Additivitat von μ

$$\begin{aligned}
\mu\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mu\left(\bigcup_{i=1}^n (A_i \cup A_{n+1})\right) \\
&= \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \mu\left(A_{n+1} \cup \bigcap_{j \in J} A_j\right) \\
&= \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \left(\mu(A_{n+1}) + \mu\left(\bigcap_{j \in J} A_j\right) - \mu\left(\bigcap_{j \in J} A_j \cap A_{n+1}\right)\right) \\
&= \mu(A_{n+1}) + \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \left(\mu\left(\bigcap_{j \in J} A_j\right) - \mu\left(\bigcap_{j \in J} A_j \cap A_{n+1}\right)\right) \\
&= \sum_{J \subseteq \{1, \dots, n+1\}} (-1)^{|J|+1} \mu\left(\bigcap_{j \in J} A_j\right). \quad \square
\end{aligned}$$

3.2 σ -Additivitat

Die endliche Additivitat von Mengenfunktionen ist eine Forderung, die man oft nachprufen kann. Anders sieht es mit der σ -Additivitat aus. Wir werden nun alternative Formulierungen fur σ -Additivitat kennen lernen.

Proposition 3.7 (Stetigkeit von unten und von oben). *Sei \mathcal{R} ein Ring und $\mu : \mathcal{R} \rightarrow \overline{\mathbb{R}}_+$ endlich additiv. Betrachte folgende Eigenschaften:*

1. μ ist σ -additiv
2. μ ist σ -subadditiv
3. μ ist stetig von unten, d.h. fur $A, A_1, A_2, \dots \in \mathcal{R}$ und $A_1 \subseteq A_2 \subseteq \dots$ mit $A = \bigcup_{n=1}^{\infty} A_n$ ist $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$.
4. μ ist stetig von oben in \emptyset , d.h. fur $A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ und $A_1 \supseteq A_2 \supseteq \dots$ mit $\bigcap_{n=1}^{\infty} A_n = \emptyset$ ist $\lim_{n \rightarrow \infty} \mu(A_n) = 0$.
5. μ ist stetig von oben, d.h. fur $A, A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ und $A_1 \supseteq A_2 \supseteq \dots$ mit $A = \bigcap_{n=1}^{\infty} A_n$ ist $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$.

Dann gilt

$$1. \iff 2. \iff 3. \implies 4. \iff 5..$$

Auerdem gilt $4. \implies 3.$, falls $\mu(A) < \infty$ fur alle $A \in \mathcal{R}$.

Beweis. $1. \iff 2.$ folgt aus Lemma 3.5, da \mathcal{R} ein Halbring ist.

$1. \implies 3.$: Sei μ also σ -additiv und $A, A_1, A_2, \dots \in \mathcal{R}$ wie in 3. Dann gilt mit $A_0 = \emptyset$

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n-1}) = \lim_{N \rightarrow \infty} \sum_{n=1}^N \mu(A_n \setminus A_{n-1}) = \lim_{N \rightarrow \infty} \mu(A_N).$$

3.⇒1.: Sei $B_1, B_2, \dots \in \mathcal{R}$ paarweise disjunkt und $B = \biguplus_{n=1}^{\infty} B_n \in \mathcal{R}$. Dann gilt für $A_N = \biguplus_{n=1}^N B_n$

$$\mu(B) = \lim_{N \rightarrow \infty} \mu(A_N) = \sum_{n=1}^{\infty} \mu(B_n).$$

4.⇒5.: Sei $A, A_1, A_2, \dots \in \mathcal{R}$ wie in 5. vorausgesetzt. Weiter sei $B_n := A_n \setminus A$. Dann erfüllt B_1, B_2, \dots die Voraussetzungen von 4., also gilt $\mu(B_n) \xrightarrow{n \rightarrow \infty} 0$, also $\mu(A_n) = \mu(B_n) + \mu(A) \xrightarrow{n \rightarrow \infty} \mu(A)$.

5.⇒4.: ist klar.

3.⇒4.: Sei $A_1, A_2, \dots \in \mathcal{R}$ wie in 4. vorausgesetzt. Setze $B_n := A_1 \setminus A_n, n \in \mathbb{N}$. Dann erfüllt $B = A_1, B_1, B_2, \dots \in \mathcal{R}$ die Voraussetzungen von 3., und damit gilt $\mu(A_1) = \lim_{n \rightarrow \infty} \mu(B_n) = \mu(A_1) - \lim_{n \rightarrow \infty} \mu(A_n)$, woraus 4. folgt.

4.⇒3. falls $\mu(A) < \infty$ für alle $A \in \mathcal{R}$. Sei $A, A_1, A_2, \dots \in \mathcal{R}$ wie in 3. vorausgesetzt. Setze $B_n := A \setminus A_n \in \mathcal{R}, n \in \mathbb{N}$. Dann gilt $\bigcap_{n=1}^{\infty} B_n = \emptyset$, also $0 = \lim_{n \rightarrow \infty} \mu(B_n) = \mu(A) - \lim_{n \rightarrow \infty} \mu(A_n)$, woraus 3. folgt. Hier geht in der letzten Gleichheit die Voraussetzung ein, dass $\mu(A) < \infty$. \square

Wir wollen nun von innen bezüglich eines kompakten Systems reguläre Mengenfunktionen näher beleuchten. In vielen Situationen ist die Forderung, dass ein Maß bezüglich eines kompakten Systems von innen regulär ist, erfüllt. Diese Tatsache, die durch das folgende Lemma gezeigt wird, wird für die Theorie schwacher Konvergenz eine wichtige Rolle spielen.

Lemma 3.8. *Ist (Ω, \mathcal{O}) polnisch und μ ein endliches Maß auf $\mathcal{B}(\mathcal{O})$, so existiert zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subseteq \Omega$ mit $\mu(\Omega \setminus K) < \varepsilon$.*

Beweis. Zuerst bemerken wir, dass alle kompakten Mengen nach Lemma 1.8 auch abgeschlossen sind, also sind alle kompakten Mengen in $\mathcal{B}(\mathcal{O})$ und damit ist $\mu(\Omega \setminus K)$ im Lemma wohldefiniert.

Sei $\varepsilon > 0$. Da Ω separabel ist, gibt es für jedes $n = 1, 2, \dots$ Punkte $\omega_1^n, \omega_2^n, \dots \in \Omega$ mit $\Omega = \bigcup_{k=1}^{\infty} B_{1/n}(\omega_k^n)$. Da μ von oben stetig ist (Proposition 3.7), gilt

$$0 = \mu\left(\Omega \setminus \bigcup_{k=1}^{\infty} B_{1/n}(\omega_k^n)\right) = \lim_{N \rightarrow \infty} \mu\left(\Omega \setminus \bigcup_{k=1}^N B_{1/n}(\omega_k^n)\right).$$

Damit gibt es ein $N_n \in \mathbb{N}$ mit $\mu\left(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\right) < \varepsilon/2^n$. Nun ist

$$A := \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)$$

nach Definition total beschränkt, also relativ kompakt nach Lemma 1.9. Außerdem gilt für die kompakte Menge \bar{A}

$$\mu(\Omega \setminus \bar{A}) \leq \mu(\Omega \setminus A) \leq \mu\left(\bigcup_{n=1}^{\infty} \left(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\right)\right) \leq \sum_{n=1}^{\infty} \mu\left(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\right) < \varepsilon.$$

Damit ist die Behauptung gezeigt. \square

Theorem 3.9 (Von innen reguläre additive Mengenfunktionen sind σ -additiv). *Sei \mathcal{H} ein Halbring und $\mu : \mathcal{H} \rightarrow \mathbb{R}_+$ endlich, endlich additiv und für ein kompaktes System $\mathcal{K} \subseteq \mathcal{H}$ von innen regulär. Dann ist μ auch σ -additiv.*

Beweis. Wie in Lemma 3.5 kann man die Mengenfunktion μ auf eindeutige Weise auf den von \mathcal{H} erzeugten Ring $\mathcal{R}(\mathcal{H})$ (siehe Lemma 2.5) erweitern. Weiter ist nach Lemma 2.16 das System $\mathcal{K}_\cup \subseteq \mathcal{R}(\mathcal{H})$, das durch Vereinigungsbildung aus Mengen von \mathcal{K} besteht, ebenfalls kompakt. Wählt man nun $\varepsilon > 0$ und $A = \bigcup_{i=1}^n A_i \in \mathcal{R}(\mathcal{H})$ mit $A_1, \dots, A_n \in \mathcal{H}$, so gibt es kompakte Mengen $K_1, \dots, K_n \in \mathcal{K} \subseteq \mathcal{H}$ mit $\mu(A_i) \leq \mu(K_i) + \frac{\varepsilon}{n}$ für $i = 1, \dots, n$. Damit gilt für die Erweiterung von μ auf den Ring $\mathcal{R}(\mathcal{H})$

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i) \leq \left(\sum_{i=1}^n \mu(K_i)\right) + \varepsilon = \mu\left(\bigcup_{i=1}^n K_i\right) + \varepsilon.$$

Damit ist μ von innen \mathcal{K}_\cup -regulär. Also ist o.E. der Halbring \mathcal{H} ein Ring und \mathcal{K} vereinigungsstabil.

Wir zeigen nun, dass μ stetig von oben in \emptyset ist. Dies genügt nach Proposition 3.7 wegen der Endlichkeit von μ auf \mathcal{H} . Seien $A_1, A_2, \dots \in \mathcal{H}$ mit $A_1 \supseteq A_2 \supseteq \dots$ und $\bigcap_{n=1}^\infty A_n = \emptyset$ und $\varepsilon > 0$. Wähle $K_1, K_2, \dots \in \mathcal{K}$ mit $K_n \subseteq A_n, n \in \mathbb{N}$ und

$$\mu(A_n) \leq \mu(K_n) + \varepsilon 2^{-n}.$$

Es gilt $\bigcap_{n=1}^\infty K_n \subseteq \bigcap_{n=1}^\infty A_n = \emptyset$. Deswegen gibt es ein $N \in \mathbb{N}$ mit $\bigcap_{n=1}^N K_n = \emptyset$. Damit gilt

$$A_N = A_N \cap \left(\bigcup_{n=1}^N K_n^c\right) = \bigcup_{n=1}^N A_N \setminus K_n \subseteq \bigcup_{n=1}^N A_n \setminus K_n.$$

Daraus folgt wegen der Subadditivität und der Monotonie von μ für alle $m \geq N$

$$\mu(A_m) \leq \mu(A_N) \leq \sum_{n=1}^N \mu(A_n \setminus K_n) \leq \varepsilon \sum_{n=1}^N 2^{-n} \leq \varepsilon.$$

Damit ist die Behauptung gezeigt, da $\varepsilon > 0$ beliebig war. \square

3.3 Eindeutigkeit und Fortsetzung von Maßen

Angenommen, eine additive Mengenfunktion $\mu : \mathcal{H} \rightarrow \overline{\mathbb{R}}_+$ ist gegeben. Wir beschäftigen uns damit, wann es möglich ist, diese Mengenfunktion zu einem Maß (d.h. einer σ -additiven Mengenfunktion) auf $\sigma(\mathcal{H})$ auszuweiten. Ziel ist es dabei, Bedingungen aufzustellen, wann das Maß durch μ schon eindeutig gegeben ist. Das Resultat ist in Theorem 3.15 zusammengefasst. Siehe auch Tabelle 3.1 für eine Übersicht, wie die Resultate vergangener Kapitel damit zusammenhängen.

Proposition 3.10 (Eindeutigkeit von Maßen). *Sei \mathcal{F} eine σ -Algebra und $\mu, \nu : \mathcal{F} \rightarrow \mathbb{R}_+$ Maße. Sei \mathcal{C} ein durchschnittstabiles Mengensystem, so dass $\sigma(\mathcal{C}) = \mathcal{F}$ und $\mu|_{\mathcal{C}}, \nu|_{\mathcal{C}}$ sind σ -endlich. Dann ist $\mu = \nu$ genau dann, wenn $\mu(A) = \nu(A)$ für alle $A \in \mathcal{C}$ gilt.*

	Lemma 3.4	Theorem 3.9	Theorem 3.15
μ endlich additiv	○	○	
μ endlich		○	
μ σ -endlich			○
μ definiert auf Halbring	○	○	○
μ σ -additiv	○/●	●	○
μ σ -subadditiv	●/○		
μ von innen \mathcal{K} -regulär		○	
μ eindeutig auf $\sigma(\mathcal{H})$ fortsetzbar			●

Tabelle 3.1: Um den Carathéodory'schen Fortsetzungssatz anwenden zu können, spielen Lemma 3.4 und Theorem 3.9 eine Rolle. In der Tabelle bedeuten die ○'s jeweils die Voraussetzungen des Satzes und ● die Folgerungen. Wie man leicht ablesen kann, gilt der Carathéodory'sche Fortsetzungssatz z.B. dann, wenn μ endlich und von innen \mathcal{K} -regulär ist.

Korollar 3.11 (Eindeutigkeit von Wahrscheinlichkeitsmaßen). Sei \mathcal{F} eine σ -Algebra und $\mu, \nu : \mathcal{F} \rightarrow [0, 1]$ seien Wahrscheinlichkeitsmaße. Sei \mathcal{C} ein durchschnittstabiles Mengensystem mit $\sigma(\mathcal{C}) = \mathcal{F}$. Dann ist $\mu = \nu$ genau dann, wenn $\mu(A) = \nu(A)$ für alle $A \in \mathcal{C}$ gilt.

Beweis. O.E. ist $\Omega \in \mathcal{C}$, da $\mu(\Omega) = \nu(\Omega) = 1$ gilt. Damit sind μ und ν insbesondere σ -endlich und die Aussage folgt aus Proposition 3.10. \square

Beweis von Proposition 3.10. Die 'genau dann'-Richtung ist klar. Für die 'wenn'-Richtung setzen wir für ein $C \in \mathcal{C}$ mit $\mu(C) = \nu(C) < \infty$

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

Wir zeigen, dass \mathcal{D}_C ein Dynkin System ist. Klar ist, dass $\Omega \in \mathcal{D}_C$. Ist weiter $A, B \in \mathcal{D}$ und $A \subseteq B$, so ist $\mu((B \setminus A) \cap C) = \mu(B \cap C) - \mu(A \cap C) = \nu(B \cap C) - \nu(A \cap C) = \nu((B \setminus A) \cap C)$, also $B \setminus A \in \mathcal{D}_C$. Ist $A_1, A_2, \dots \in \mathcal{D}$ mit $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \in \mathcal{D}_C$ und $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, so ist wegen Proposition 3.7.

$$\mu(A \cap C) = \lim_{n \rightarrow \infty} \mu(A_n \cap C) = \lim_{n \rightarrow \infty} \nu(A_n \cap C) = \nu(A \cap C),$$

also $A \in \mathcal{D}_C$. Damit ist \mathcal{D}_C für alle $C \in \mathcal{C}$ mit $\mu(C) < \infty$ ein Dynkin-System und damit gilt wegen Theorem 2.13, dass $\mathcal{F} = \sigma(\mathcal{C}) = \mathcal{D}_C$. Sei $\Omega_1, \Omega_2, \dots \in \mathcal{C}$ mit $\Omega_n \uparrow \Omega$ und $\mu(\Omega_n), \nu(\Omega_n) < \infty, n = 1, 2, \dots$. Dann gilt für alle $n = 1, 2, \dots$, dass $\mu(A \cap \Omega_n) = \nu(A \cap \Omega_n)$ für alle $A \in \mathcal{F}$. Daraus folgt dann für $A \in \mathcal{F}$, da μ und ν stetig von unten sind,

$$\mu(A) = \lim_{n \rightarrow \infty} \mu(A \cap \Omega_n) = \lim_{n \rightarrow \infty} \nu(A \cap \Omega_n) = \nu(A),$$

d.h. $\mu = \nu$. \square

Das folgende Theorem erklärt, warum der Begriff der σ -Algebra zentral ist.

Theorem 3.12 (μ^* -messbare Mengen sind eine σ -Algebra). Sei μ^* ein äußeres Maß auf Ω und \mathcal{F}^* die Menge der μ^* -messbaren Mengen. Dann ist \mathcal{F}^* eine σ -Algebra und $\mu := \mu^*|_{\mathcal{F}^*}$ ein Maß. Außerdem ist $\mathcal{N} := \{N \subseteq \Omega : \mu^*(N) = 0\} \subseteq \mathcal{F}^*$.

Bemerkung 3.13 (Nullmengen und Eigenschaften fast überall). Mengen N mit $\mu(N) = 0$ heißen auch (μ -)Nullmengen. Weiter sagen wir, dass $A \subseteq \Omega$ (μ -)fast überall gilt, wenn $A^c \subseteq N$ und N eine μ -Nullmenge ist. Ist μ ein Wahrscheinlichkeitsmaß, sagen wir anstatt 'fast überall' auch *fast sicher*.

Beweis von Theorem 3.12. Wir zeigen zunächst, dass \mathcal{F}^* eine σ -Algebra ist. Klar ist, dass

$$\mu^*(E) = \mu^*(\emptyset) + \mu^*(E) = \mu^*(E \cap \emptyset) + \mu^*(E \cap \Omega),$$

d.h. $\emptyset \in \mathcal{F}^*$. Weiter ist klar, dass aus $A \in \mathcal{F}^*$ auch $A^c \in \mathcal{F}^*$ folgt. Mit $A, B \in \mathcal{F}^*$ ist wegen $(E \cap A \cap B^c) \uplus (E \cap A^c) = E \cap (A \cap B)^c$ und der Subadditivität von μ^*

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*((E \cap A) \cap B) + \mu^*((E \cap A) \cap B^c) + \mu^*(E \cap A^c) \\ &\geq \mu^*(E \cap (A \cap B)) + \mu^*(E \cap (A \cap B)^c) \geq \mu^*(E). \end{aligned}$$

Damit ist $A \cap B \in \mathcal{F}^*$. Seien nun $A_1, A_2, \dots \in \mathcal{F}^*$ disjunkt und $B_n = \biguplus_{k=1}^n A_k$ und $B = \bigcup_{n=1}^{\infty} B_n = \biguplus_{k=1}^{\infty} A_k$. Da \mathcal{F}^* Durchschnitt- und komplementstabil ist, ist auch $B_1, B_2, \dots \in \mathcal{F}^*$. Weiter zeigen wir, dass $\mu^*(E \cap B_n) = \sum_{k=1}^n \mu^*(E \cap A_k)$ für alle $E \subseteq \Omega$ gilt. Für $n = 1$ ist dies klar, und gilt es für ein n , so folgt, da $B_n \in \mathcal{F}^*$,

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

Damit gilt wegen der Monotonie von μ^*

$$\mu^*(E \cap B) \leq \sum_{k=1}^{\infty} \mu^*(E \cap A_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu^*(E \cap A_k) = \lim_{n \rightarrow \infty} \mu^*(E \cap B_n) \leq \mu^*(E \cap B),$$

also

$$\mu^*(E \cap B) = \lim_{n \rightarrow \infty} \mu^*(E \cap B_n) = \sum_{k=1}^{\infty} \mu^*(E \cap A_k). \quad (3.4)$$

Als nächstes zeigen wir, dass $B \in \mathcal{F}^*$ ist, woraus folgt, dass \mathcal{F}^* eine σ -Algebra ist. Es gilt nämlich für $E \subseteq \Omega$, wegen (3.4) und der Subadditivität von μ^*

$$\mu^*(E) = \lim_{n \rightarrow \infty} \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E),$$

woraus $B \in \mathcal{F}^*$ folgt. Weiter folgt aus (3.4), dass μ^* auf \mathcal{F}^* sogar σ -additiv ist, d.h. dass $\mu = \mu^*|_{\mathcal{F}^*}$ ein Maß ist.

Sei nun $N \subseteq \Omega$ so, dass $\mu^*(N) = 0$ und $E \subseteq \Omega$. Dann ist wegen der Monotonie von μ^* auch $\mu^*(E \cap N) = 0$, also

$$\mu^*(E) \geq \mu^*(E \cap N^c) = \mu^*(E \cap N^c) + \mu^*(E \cap N)$$

und damit $N \in \mathcal{F}^*$. □

Proposition 3.14 (Von endlich additiver Mengenfunktion erzeugtes äußeres Maß).

Sei \mathcal{H} ein Halbring und $\mu : \mathcal{H} \rightarrow \mathbb{R}_+$ endlich additiv. Für $A \subseteq \Omega$ sei

$$\mu^*(A) := \inf_{\mathcal{G} \in \mathcal{U}(A)} \sum_{G \in \mathcal{G}} \mu(G)$$

wobei

$$\mathcal{U}(A) := \left\{ \mathcal{G} \subseteq \mathcal{H} \text{ höchstens abzählbar, } A \subseteq \bigcup_{G \in \mathcal{G}} G \right\}$$

die Menge der höchstens abzählbaren Überdeckungen von A ist und $\mu^*(A) = \infty$ falls $\mathcal{U}(A) = \emptyset$. Dann ist μ^* ein äußeres Maß.

Beweis. Die Abbildung μ^* ist monoton, und da $\emptyset \in \mathcal{H}$ ist $\mu^*(\emptyset) = 0$ wegen der endlichen Additivität von μ ; siehe Lemma 3.4. (Es ist nämlich $\mu(\emptyset) = \mu(\emptyset) + \mu(\emptyset)$, woraus $\mu(\emptyset) = 0$ folgt.) Um die σ -Subadditivität von μ^* zu prüfen, wählen wir $A_1, A_2, \dots \subseteq \Omega$. Für $n = 1, 2, \dots$ und $\epsilon > 0$ gibt es Mengen $G_{nk} \in \mathcal{H}$, $k \in \mathcal{K}_n$ höchstens abzählbar mit

$$\begin{aligned} A_n &\subseteq \bigcup_{k \in \mathcal{K}_n} G_{nk}, \\ \mu^*(A_n) &\geq \sum_{k \in \mathcal{K}_n} \mu(G_{nk}) - \epsilon 2^{-n}. \end{aligned}$$

Daraus folgt wegen $\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} \bigcup_{k \in \mathcal{K}_n} G_{nk}$, der Monotonie von μ^* und der Definition von μ^*

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \sum_{k \in \mathcal{K}_n} \mu(G_{nk}) \leq \epsilon + \sum_{n=1}^{\infty} \mu^*(A_n).$$

Mit $\epsilon \rightarrow 0$ folgt die σ -Subadditivität von μ^* , d.h. μ^* ist ein äußeres Maß. \square

Theorem 3.15 (Fortsetzung einer σ -additiven Mengenfunktion). Sei \mathcal{H} ein Halbring und $\mu : \mathcal{H} \rightarrow \mathbb{R}_+$ σ -endlich und σ -additiv. Weiter sei $\tilde{\mu} = \mu^*|_{\mathcal{F}^*}$ mit μ^* aus Proposition 3.14 und \mathcal{F}^* aus Theorem 3.12. Dann ist $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ und $\tilde{\mu}|_{\sigma(\mathcal{H})}$ ist das einzige Maß, das auf \mathcal{H} mit μ übereinstimmt.

Beweis. Zunächst stellen wir fest, dass μ sowohl endlich additiv als auch σ -subadditiv ist nach Lemma 3.4. Nach Proposition 3.14 ist μ^* ein äußeres Maß und nach Theorem 3.12 ist \mathcal{F}^* eine σ -Algebra.

Schritt 1: μ^ stimmt auf \mathcal{H} mit μ überein:* Sei $H \in \mathcal{H}$. Wähle \mathcal{K} höchstens abzählbar und $H_k \in \mathcal{H}$, $k \in \mathcal{K}$ mit $H \subseteq \bigcup_{k \in \mathcal{K}} H_k$ und

$$\mu^*(H) \geq \sum_{k \in \mathcal{K}} \mu(H_k) - \epsilon.$$

Damit gilt wegen $H = \bigcup_{k \in \mathcal{K}} H_k \cap H$ und der σ -Subadditivität von μ

$$\mu^*(H) \leq \mu(H) \leq \sum_{k \in \mathcal{K}} \mu(H_k \cap H) \leq \sum_{k \in \mathcal{K}} \mu(H_k) \leq \mu^*(H) + \epsilon,$$

wobei wir im zweiten ' \leq ' die endliche Additivität von μ verwendet haben. Mit $\varepsilon \rightarrow 0$ folgt, dass $\mu^*(H) = \mu(H)$.

Schritt 2: $\sigma(\mathcal{H}) \subseteq \mathcal{F}^$:* Sei $E \subseteq \Omega, H \in \mathcal{H}$ und $\varepsilon > 0$. Wähle \mathcal{K} höchstens abzählbar und $H_k \in \mathcal{H}, k \in \mathcal{K}$ mit $\mu^*(E) \geq \sum_{k \in \mathcal{K}} \mu(H_k) - \varepsilon$. Dann gilt wegen der endlichen Additivität von μ

$$\mu^*(E) + \varepsilon \geq \sum_{k \in \mathcal{K}} \mu(H_k) = \sum_{k \in \mathcal{K}} \mu(H_k \cap H) + \sum_{k \in \mathcal{K}} \mu(H_k \cap H^c) \geq \mu^*(E \cap H) + \mu^*(E \cap H^c).$$

Mit $\varepsilon \rightarrow 0$ und der σ -Subadditivität von μ^* folgt $\mu^*(E) = \mu^*(E \cap H) + \mu^*(E \cap H^c)$, d.h. H ist μ^* -messbar und damit $\mathcal{H} \subseteq \mathcal{F}^*$. Da \mathcal{F}^* nach Theorem 3.12 eine σ -Algebra ist, folgt auch $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$.

Schritt 3: Eindeutigkeit: Nach Theorem 3.12 ist $\tilde{\mu}$ ein Maß, also auch $\tilde{\mu}|_{\sigma(\mathcal{H})} = \mu^*|_{\sigma(\mathcal{H})}$. Sei $\nu : \sigma(\mathcal{H}) \rightarrow \mathbb{R}_+$ ein weiteres Maß, das auf \mathcal{H} mit μ übereinstimmt. Da $\mu = \tilde{\mu}|_{\mathcal{H}}$ als σ -endlich vorausgesetzt war, ist auch $\nu|_{\mathcal{H}}$ σ -endlich. Mit Proposition 3.10 folgt wegen der Durchschnittsstabilität von \mathcal{H} , dass $\tilde{\mu} = \nu$ auf $\sigma(\mathcal{H})$ gilt.

Nun sind alle Behauptungen bewiesen. \square

In obigem Theorem wird nur klar, dass $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$. Das nächste Resultat zeigt, wodurch sich Mengen in \mathcal{F}^* von Mengen in $\sigma(\mathcal{H})$ unterscheiden.

Proposition 3.16 (Charakterisierung von \mathcal{F}^* aus Proposition 3.14). *Sei \mathcal{H} ein Halb-ring, $\mu : \mathcal{H} \rightarrow \mathbb{R}_+$ σ -endlich und σ -additiv, μ^* wie in Proposition 3.14 und $\mathcal{F}^*, \mathcal{N}$ wie in Theorem 3.12. Dann ist*

$$\mathcal{F}^* = \{A \setminus N : A \in \sigma(\mathcal{H}), N \in \mathcal{N}\}.$$

Insbesondere ist die rechte Seite eine σ -Algebra.

Beweis. ' \supseteq ': Nach Theorem 3.15 ist $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ und nach Theorem 3.12 ist $\mathcal{N} \subseteq \mathcal{F}^*$. Daraus folgt bereits ' \supseteq ', da \mathcal{F}^* komplementstabil ist.

' \subseteq ': Sei $B \in \mathcal{F}^*$. Weiter seien $\Omega_1, \Omega_2, \dots \in \mathcal{H}$ mit $\mu(\Omega_n) < \infty, n = 1, 2, \dots$ und $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$. Sei $\varepsilon_1, \varepsilon_2, \dots > 0$ mit $\varepsilon_i \downarrow 0$. Für $B_n := B \cap \Omega_n$ wählen wir \mathcal{K}_{ni} höchstens abzählbar, $A_{nik} \in \mathcal{H}, n \in \mathbb{N}, k \in \mathcal{K}_{ni}, B_n \subseteq \bigcup_{k \in \mathcal{K}_{ni}} A_{nik}$ und

$$\mu^*(B_n) \geq \sum_{k \in \mathcal{K}_{ni}} \mu(A_{nik}) - 2^{-n} \varepsilon_i.$$

Klar ist $A_i := \bigcup_{n=1}^{\infty} \bigcup_{k \in \mathcal{K}_{ni}} A_{nik} \in \sigma(\mathcal{H}), B \subseteq A_i, i \in \mathbb{N}$ und $A_i \setminus B = \bigcup_{n=1}^{\infty} \bigcup_{k \in \mathcal{K}_{ni}} A_{nik} \setminus B_n$. Damit gilt

$$\mu^*(A_i \setminus B) \leq \sum_{n=1}^{\infty} 2^{-n} \varepsilon_i = \varepsilon_i.$$

Setze $A = \bigcap_{i=1}^{\infty} A_i \in \sigma(\mathcal{H})$. Dann ist $B \subseteq A, N := A \setminus B \subseteq A_n \setminus B, n \in \mathbb{N}$ und

$$\mu^*(N) = \mu^*(A \setminus B) \leq \limsup_{i \rightarrow \infty} \mu^*(A_i \setminus B) \leq \limsup_{i \rightarrow \infty} \varepsilon_i = 0.$$

Damit folgt die Behauptung, da $B = A \setminus N$. \square

3.4 Maße auf $\mathcal{B}(\mathbb{R})$

Aus der Vorlesung *Stochastik* kennt man bereits Wahrscheinlichkeitsverteilungen mit Dichte. Diese sind Maße auf $\mathcal{B}(\mathbb{R})$. Wir werden hier die in den letzten Kapiteln erarbeitete allgemeine Theorie anwenden, um solche Maße zu charakterisieren.

Proposition 3.17 (Lebesgue-Maß auf \mathbb{R}). *Es gibt genau ein Maß λ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so dass*

$$\lambda((a, b]) = b - a \quad (3.5)$$

für $a, b \in \mathbb{Q}$ mit $a \leq b$.

Beweis. Nach Beispiel 2.3 ist $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ ein Halbring. Genauso ist $\tilde{\mathcal{H}} := \{(a, b], [a, b), (a, b), [a, b] : a, b \in \mathbb{Q}, a \leq b\}$ ein Halbring mit $\sigma(\tilde{\mathcal{H}}) = \mathcal{B}(\mathbb{R})$. Wir definieren $\tilde{\lambda}$ auf $\tilde{\mathcal{H}}$ durch

$$\tilde{\lambda}((a, b]) = \tilde{\lambda}([a, b)) = \tilde{\lambda}((a, b)) = \tilde{\lambda}([a, b]) = b - a.$$

Dann ist $\tilde{\lambda}$ klar σ -endlich. Es ist $\mathcal{K} = \{[a, b] : a, b \in \mathbb{Q}, a \leq b\} \subseteq \tilde{\mathcal{H}}$ ein kompaktes System nach Beispiel 2.3. Außerdem ist klar $\tilde{\lambda}$ von innen \mathcal{K} -regulär und damit σ -additiv nach Theorem 3.9. Nach Theorem 3.15 gibt es damit genau eine Fortsetzung von $\tilde{\lambda}$ auf $\sigma(\tilde{\mathcal{H}}) = \mathcal{B}(\mathbb{R})$. \square

Proposition 3.18 (Charakterisierung der σ -endlichen Maße auf \mathbb{R}). *Eine Funktion $\mu : \mathcal{B}(\mathbb{R}) \rightarrow \overline{\mathbb{R}}_+$ ist genau dann ein σ -endliches Maß, wenn es eine nicht-fallende und rechtsstetige Funktion $G : \mathbb{R} \rightarrow \mathbb{R}$ gibt mit*

$$\mu((a, b]) = G(b) - G(a) \quad (3.6)$$

für $a, b \in \mathbb{Q}$ mit $a \leq b$. Ist \tilde{G} eine weitere Funktion $\tilde{G} : \mathbb{R} \rightarrow \mathbb{R}$ für die (3.6) gilt, so ist $\tilde{G} = G + c$ für ein $c \in \mathbb{R}$.

Korollar 3.19 (Charakterisierung der Wahrscheinlichkeitsmaße auf \mathbb{R}). *Eine Funktion $\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ ist genau dann ein Wahrscheinlichkeitsmaß, wenn es eine nicht-fallende und rechtsstetige Funktion $F : \mathbb{R} \rightarrow [0, 1]$ gibt mit $\lim_{b \rightarrow \infty} F(b) = 1$ und*

$$\mu((a, b]) = F(b) - F(a) \quad (3.7)$$

für $a, b \in \mathbb{Q}$ mit $a \leq b$. In diesem Fall ist F eindeutig durch μ festgelegt.

Beweis. Die Behauptung folgt direkt aus Proposition 3.18, da der Bildbereich von F nur für eine rechtsstetige, nicht-fallende Funktion genau $[0, 1]$ sein kann. \square

Beweis von Proposition 3.18. '⇒': Sei also μ ein σ -endliches Maß auf $\mathcal{B}(\mathbb{R})$. Definiere dann $G(0) = 0$, sowie $G(x) := \mu((0, x])$ für $x > 0$ und $G(x) := \mu((x, 0])$ für $x < 0$. Dann ist G rechtsstetig, nicht fallend, und (etwa für $0 < a < b$) $\mu((a, b]) = \mu((0, b]) - \mu((0, a]) = G(b) - G(a)$.

'⇐': Der Beweis verläuft ähnlich zu dem Beweis von Korollar 3.17. Sei $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ der Halbring der halboffenen Intervalle mit Enden in rationalen Zahlen. Wir zeigen, dass durch (3.6) eine σ -additive Mengenfunktion μ auf \mathcal{H} definiert wird. Dann sieht man mit Theorem 3.15, dass μ auf eindeutige Weise zu einem σ -endlichen Maß auf $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$ erweitert werden kann. Sei also a_1, a_2, \dots disjunkt so, dass $\bigcup_{n=1}^{\infty} (a_n, a_{n+1}] = (a, b] \in \mathcal{H}$.

Ohne Einschränkung ist $a_1 \geq a_2 \geq \dots$. Dann ist $b = a_1$ und $a_n \xrightarrow{n \rightarrow \infty} a$. Es gilt wegen der Rechtsstetigkeit von G , dass

$$\mu(a, b] = G(b) - G(a) = G(a_1) - \lim_{N \rightarrow \infty} G(a_N) = \sum_{n=1}^{\infty} G(a_n) - G(a_{n+1}) = \sum_{n=1}^{\infty} \mu((a_n, a_{n+1}]),$$

und damit ist die σ -Additivität von μ gezeigt.

Sei nun \tilde{G} eine weitere Funktion, für die (3.6) gilt. Dann gilt für alle $a \in \mathbb{R}$

$$\tilde{G}(b) = \tilde{G}(a) + \mu((a, b]) = G(b) + \tilde{G}(a) - G(a)$$

und die Behauptung folgt mit $c = \tilde{G}(a) - G(a)$. \square

Definition 3.20 (Lebesgue-Maß und Verteilungsfunktion). 1. Das eindeutig definierte Maß λ aus Korollar 3.17 heißt ein-dimensionales Lebesgue-Maß.

2. Für ein Wahrscheinlichkeitsmaß μ auf $\mathcal{B}(\mathbb{R})$ heißt die Funktion F aus Korollar 3.19 Verteilungsfunktion.

Beispiel 3.21 (Bekannte Verteilungsfunktionen). Sei $f : \mathbb{R} \rightarrow \mathbb{R}_+$ stückweise stetig, so dass⁷ $\int_{-\infty}^{\infty} f(x) dx = 1$ gilt. Wie aus der Vorlesung *Stochastik* bekannt, heißt eine solche Funktion eine *Dichte*. Solche Dichtefunktionen definieren vermöge

$$F(x) := \int_{-\infty}^x f(a) da$$

eine Verteilungsfunktion. Auf der anderen Seite definiert jede dieser Verteilungsfunktionen wegen Korollar 3.19 auf eindeutige Art und Weise ein Wahrscheinlichkeitsmaß. Verteilungen mit Dichten werden wir genauer im Satz von Radon-Nikodým (siehe Abschnitt 5.4) beleuchten.

Wie bereits bekannt, ist

$$F_{U(0,1)}(x) = \int_{-\infty}^x 1_{[0,1]}(a) da = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x \leq 1, \\ 1, & x > 1 \end{cases} \quad (3.8)$$

die Verteilungsfunktion der Gleichverteilung auf $[0, 1]$. Weiter ist für $x \geq 0$

$$F_{\text{exp}(\lambda)}(x) = \int_{-\infty}^x 1_{[0,\infty)}(a) \lambda e^{-\lambda a} da = 1 - e^{-\lambda x} \quad (3.9)$$

die Verteilungsfunktion der Exponentialverteilung mit Parameter λ . Außerdem ist

$$F_{N(\mu, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{(a-\mu)^2}{2\sigma^2}\right) da =: \Phi(x) \quad (3.10)$$

die Verteilungsfunktion der Normalverteilung $N(\mu, \sigma^2)$ mit dem Erwartungswert μ und der Varianz σ^2 .

⁷Wir setzen hier das Riemann-Integral $\int_a^b f(x) dx$ als bekannt voraus. (Siehe auch Definition 4.21.) Einen weiteren Integralbegriff, das Lebesgue-Integral, werden wir in Kapitel 4 kennen lernen.

3.5 Bildmaße

Sei μ ein Maß auf \mathcal{F} . Wenn man nun den Grundraum vermöge einer Funktion $f : \Omega \rightarrow \Omega'$ transformiert, kann man auf Ω' ein zu der Transformation korrespondierendes Maß definieren, das sogenannte Bildmaß. Sei etwa $\Omega := [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$ und $f : u \mapsto 1 - u$. Wir werden dann sehen, dass das Bildmaß von $\mu_{U(0,1)}$ unter f wieder $\mu_{U(0,1)}$ ist. Wir erinnern zunächst an die Situation aus Beispiel 2.3.2 und definieren das Bildmaß.

Definition 3.22 (Bildmaß). *Ist $(\Omega, \mathcal{F}, \mu)$ ein Maßraum, (Ω', \mathcal{F}') ein Messraum und $f : \Omega \rightarrow \Omega'$, so dass $\sigma(f) \subseteq \mathcal{F}$ für $\sigma(f)$ aus (2.1). Dann definieren wir eine Mengenfunktion $f_*\mu : \mathcal{F}' \rightarrow \mathbb{R}_+$ durch*

$$f_*\mu(A') := \mu(f^{-1}(A')) = \mu(f \in A'), \quad A' \in \mathcal{F}'.$$

Hier heißt $f_*\mu$ auch Bildmaß von μ unter f .

Ist μ ein Wahrscheinlichkeitsmaß, so heißt $f_*\mu$ auch Verteilung von f .

Bemerkung 3.23 (Messbare Funktionen). Ist $\sigma(f) \subseteq \mathcal{F}$ wie in obiger Definition, so sagen wir, dass f (nach \mathcal{F}/\mathcal{F}') messbar ist. Dieses Konzept wird uns im nächsten Abschnitt beschäftigen.

Proposition 3.24 (Bildmaß ist ein Maß). *Sei $(\Omega, \mathcal{F}, \mu)$, (Ω', \mathcal{F}') , $f : \Omega \rightarrow \Omega'$ und $f_*\mu$ wie in Definition 3.22. Dann ist $f_*\mu$ ein Maß auf \mathcal{F}' .*

Beweis. Seien $A'_1, A'_2, \dots \in \mathcal{F}'$ disjunkt, so gilt

$$f_*\mu\left(\biguplus_{n=1}^{\infty} A'_n\right) = \mu\left(f^{-1}\left(\biguplus_{n=1}^{\infty} A'_n\right)\right) = \mu\left(\biguplus_{n=1}^{\infty} (f^{-1}(A'_n))\right) = \sum_{n=1}^{\infty} \mu(f^{-1}(A'_n)) = \sum_{n=1}^{\infty} f_*\mu(A'_n).$$

Damit ist $f_*\mu$ also σ -additiv und die Behauptung ist gezeigt. \square

Beispiel 3.25 (Bekannte Transformationen). 1. Für $\Omega = [0, 1]$ ist $\{[0, b] : 0 \leq b \leq 1\}$ ein schnittstabiles Erzeugendensystem von $\mathcal{B}([0, 1])$. Sei $\mu = \mu_{U(0,1)}$ die Gleichverteilung auf $[0, 1]$ mit Verteilungsfunktion $F_{U(0,1)}$ aus (3.8) und $f : u \mapsto 1 - u$. Dann ist $f_*\mu = \mu$, denn

$$f_*\mu([0, b]) = \mu(f^{-1}([0, b])) = \mu([1 - b, 1]) = F_{U(0,1)}(1) - F_{U(0,1)}(1 - b) = b.$$

Damit stimmen μ und $f_*\mu$ auf einem schnittstabilen Erzeuger überein und die Aussage folgt mit Proposition 3.10.

2. Sei $\Omega = \mathbb{R}$, $f_y : x \mapsto x + y$ für ein $y \in \mathbb{R}$ und λ das Lebesgue-Maß aus Korollar 3.17. Dann ist $(f_y)_*\lambda = \lambda$, denn

$$(f_y)_*\lambda([a, b]) = \lambda(f_y^{-1}([a, b])) = \lambda([a - y, b - y]) = b - a.$$

Man sagt auch, dass das Lebesgue-Maß translationsinvariant ist.

3. Sei $\Omega = [0, 1]$, $\Omega' = \mathbb{R}_+$, jeweils versehen mit der Borel'schen σ -Algebra. Weiter sei $\mu = \mu_{U(0,1)}$ mit Verteilungsfunktion $F_{U(0,1)}$ aus (3.8) und $f : x \mapsto -\frac{1}{\lambda} \log(x)$ für ein $\lambda > 0$. Dann ist $f_*\mu = \mu_{\exp(\lambda)}$, wobei $\mu_{\exp(\lambda)}$ die Verteilungsfunktion $F_{\exp(\lambda)}$ aus (3.9) hat. Denn es gilt für $x \geq 0$

$$f_*\mu([0, x]) = \mu(f^{-1}([0, x])) = \mu([e^{-\lambda x}, 1]) = 1 - e^{-\lambda x}.$$

Beispiel 3.26 (Beispiel einer nicht Borel-messbaren Menge, Vitali-Menge).

Bisher gab es noch kein Beispiel für eine Menge, die nicht in $\mathcal{B}(\mathbb{R})$ liegt. Eine solche werden wir nun konstruieren. Sie ist als *Vitali-Menge* bekannt. Hierzu definieren wir auf \mathbb{R} eine Äquivalenzrelation durch $x \sim y \iff y - x \in \mathbb{Q}$. Bezüglich dieser Äquivalenzrelation zerfällt \mathbb{R} in Äquivalenzklassen der Form $\{x + q : q \in \mathbb{Q}\}$. Wir wählen aus jeder Äquivalenzklasse eine Zahl aus $[0, 1]$ aus und bilden damit die Menge V . (Es sei hier angemerkt, dass dieses Auswählen mittels des Auswahlaxioms der Mengenlehre geschieht und damit keinen trivialen Schritt darstellt.) Weiter sei nun für $q \in \mathbb{Q} \cap [-1, 1]$

$$V_q := \{x + q : x \in V\}.$$

Dann ist $[0, 1] \subseteq \bigsqcup_{q \in \mathbb{Q} \cap [-1, 1]} V_q \subseteq [-1, 2]$.

Angenommen, die Menge V wäre messbar. Dann wären auch die Mengen V_q messbar und, wegen der Translationsinvarianz des Lebesgue-Maßes aus Beispiel 3.25.2 wäre $\lambda(V_q)$ nicht von q abhängig. Sei also $\lambda(V_q) =: a \geq 0$. Außerdem wäre dann wegen der Monotonie des Lebesgue-Maßes

$$1 \leq \sum_{q \in \mathbb{Q} \cap [-1, 1]} \lambda(V_q) = \sum_{q \in \mathbb{Q} \cap [-1, 1]} a \leq 3.$$

Dies ist jedoch nicht möglich, weder für $a = 0$ noch für $a > 0$. Wegen diesem Widerspruch muss also $V \notin \mathcal{B}(\mathbb{R})$ gelten.

4 Messbare Funktionen und das Integral

In diesem Kapitel sei immer $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Wir können mittlerweile mit dem Maß μ Maßen den Inhalt von Mengen aus \mathcal{F} messen. Ziel der Einführung des Integrals ist es, bei einer solchen Messung die Elemente von Ω verschieden zu gewichten. Diese Gewichtung wird mit einer Funktion $f : \Omega \rightarrow \mathbb{R}$ vorgenommen. Solche Funktionen müssen die Minimalforderung der Messbarkeit genügen. Das Resultat dieser Gewichtung führt auf den Begriff des Integrals.

4.1 Messbare Funktionen

Wir kennen bereits den Begriff der (bezüglich der σ -Algebra \mathcal{F}) messbaren Menge, der nun erweitert wird. Sei $A \in \mathcal{F}$ also eine messbare Menge, so definiert diese mit $\omega \mapsto 1_A(\omega)$ eine messbare Funktion in dem in Definition 4.3 eingeführte Sinne. Die Linearkombination von solchen Indikatorfunktionen werden wir einfache Funktion nennen. Diesen kommt wegen Theorem 4.8 eine besondere Bedeutung zu, da jede nicht-negative messbare Funktion (im Sinne der punktweisen Konvergenz) von unten durch einfache Funktionen approximiert werden kann.

Bemerkung 4.1 (Notation). Seien Ω, Ω' Mengen, $f : \Omega \rightarrow \Omega'$ und I beliebig.

- Wir schreiben $f(A) := \{f(\omega) : \omega \in A\}$ für $A \subseteq \Omega$ sowie $f^{-1}(A') := \{f^{-1}(\omega') : \omega' \in A'\}$ für $A' \subseteq \Omega'$. Wir bemerken, dass folgende Regeln für $A', A'_i \subseteq \Omega', i \in I$ gelten:

$$f^{-1}((A')^c) = (f^{-1}(A'))^c, \quad f^{-1}\left(\bigcap_{i \in I} A'_i\right) = \bigcap_{i \in I} f^{-1}(A'_i), \quad f^{-1}\left(\bigcup_{i \in I} A'_i\right) = \bigcup_{i \in I} f^{-1}(A'_i).$$

Etwas Vorsicht ist jedoch geboten, da für $A, A_i \subseteq \Omega, i \in I$ nur $f(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f(A_i)$, im allgemeinen jedoch $f(A^c) \neq (f(A))^c$ und $f(\bigcap_{i \in I} A_i) \neq \bigcap_{i \in I} f(A_i)$.

2. Für $\mathcal{C} \subseteq 2^{\Omega'}$ schreiben wir ganz analog

$$f^{-1}(\mathcal{C}) := \{f^{-1}(A') : A' \in \mathcal{C}\}.$$

Lemma 4.2 (Urbilder von σ -Algebren). *Sei Ω eine Menge und (Ω', \mathcal{F}') ein Messraum, $f : \Omega \rightarrow \Omega'$ und $\mathcal{C}' \subseteq \mathcal{F}'$ mit $\sigma(\mathcal{C}') = \mathcal{F}'$. Dann ist $\sigma(f^{-1}(\mathcal{C}')) = f^{-1}(\sigma(\mathcal{C}'))$. Insbesondere ist $f^{-1}(\mathcal{F}')$ eine σ -Algebra auf Ω .*

Beweis. '⊆': Nach Bemerkung 4.1 ist klar, dass $f^{-1}(\sigma(\mathcal{C}'))$ eine σ -Algebra ist. Damit ist $\sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(f^{-1}(\sigma(\mathcal{C}'))) = f^{-1}(\sigma(\mathcal{C}'))$.

'⊇': Wir definieren

$$\tilde{\mathcal{F}}' := \{A' \in \sigma(\mathcal{C}') : f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))\} \subseteq \sigma(\mathcal{C}').$$

Dann ist, wieder wegen Bemerkung 4.1, $\tilde{\mathcal{F}}'$ eine σ -Algebra und $\mathcal{C}' \subseteq \tilde{\mathcal{F}}' \subseteq \sigma(\mathcal{C}')$, also $\tilde{\mathcal{F}}' = \sigma(\mathcal{C}')$. Für $A' \in \sigma(\mathcal{C}')$ ist also $f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))$, was gleichbedeutend ist mit $f^{-1}(\sigma(\mathcal{C}')) \subseteq \sigma(f^{-1}(\mathcal{C}'))$. \square

Definition 4.3 (Messbare Funktionen). *Seien (Ω, \mathcal{F}) und (Ω', \mathcal{F}') Messräume und $f : \Omega \rightarrow \Omega'$.*

1. *Die Funktion f heißt \mathcal{F}/\mathcal{F}' -messbar, falls $f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$. Die σ -Algebra $f^{-1}(\mathcal{F}')$ heißt die von f erzeugte σ -Algebra auf Ω und wird auch mit $\sigma(f)$ bezeichnet.*
2. *Ist $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \Omega'$ messbar, so heißt X eine Ω' -wertige Zufallsvariable. Das Bildmaß $X_*\mathbf{P}$ aus Definition 3.22 heißt dann die Verteilung von X .*
3. *Ist $(\Omega', \mathcal{F}') = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, so heißt f eine reell(wertig)e Funktion. Ist f nach $\mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$ -messbar, so sagen wir, die Funktion f sei (Borel-)messbar.*
4. *Ist $\Omega' = \overline{\mathbb{R}}$ und $f = 1_A$ für $A \subseteq \Omega$, so heißt f auch Indikatorfunktion. Ist $f = \sum_{k=1}^n c_k 1_{A_k}$ für $c_1, \dots, c_n \in \overline{\mathbb{R}}$ paarweise verschieden und $A_1, \dots, A_n \subseteq \Omega$, so heißt f einfach.*

Beispiel 4.4. Sei (Ω, \mathcal{F}) ein Messraum.

1. Die allermeisten Funktionen $f : \Omega \rightarrow \overline{\mathbb{R}}$, die man sich vorstellen kann, sind (Borel-)messbar. Etwa ist die Identitätsabbildung $f : \omega \mapsto \omega$ messbar, da $f^{-1}(\mathcal{F}) = \mathcal{F}$.
2. Es ist wichtig zu sehen, dass für viele messbaren Funktionen f gilt, dass $\sigma(f) \subsetneq \mathcal{F}$, siehe etwa das nächste Beispiel.
3. Eine Funktion $f : \Omega \rightarrow \{0, 1\}$ ist genau dann messbar, wenn $f^{-1}(\{1\}) \in \mathcal{F}$. Es ist nämlich in diesem Fall $\sigma(f) = \{\emptyset, f^{-1}(\{1\}), (f^{-1}(\{1\}))^c, \Omega\}$.
4. Sei $\mathcal{F} = \mathcal{B}(\mathbb{R})$. Um eine nicht \mathcal{F} -messbare Funktion anzugeben, muss man sich genauso anstrengen, wie eine nicht Borel-messbare Menge zu konstruieren. Es ist etwa für die Vitali-Menge V aus Beispiel 3.26 die Funktion 1_V nicht messbar.

Lemma 4.5 (Eigenschaften von Messbarkeit). *Seien (Ω, \mathcal{F}) , (Ω', \mathcal{F}') und $(\Omega'', \mathcal{F}'')$ Messräume.*

1. Ist $\mathcal{C}' \subseteq \mathcal{F}'$ mit $\mathcal{F}' = \sigma(\mathcal{C}')$, so ist $f : \Omega \rightarrow \Omega'$ genau dann nach \mathcal{F}/\mathcal{F}' -messbar, wenn $f^{-1}(\mathcal{C}') \subseteq \mathcal{F}$.
2. Ist $f : \Omega \rightarrow \Omega'$ nach \mathcal{F}/\mathcal{F}' und $g : \Omega' \rightarrow \Omega''$ nach $\mathcal{F}'/\mathcal{F}''$ messbar. Dann ist $g \circ f : \Omega \rightarrow \Omega''$ nach $\mathcal{F}/\mathcal{F}''$ -messbar.
3. Seien (Ω, \mathcal{O}) und (Ω', \mathcal{O}') topologische Räume, $f : \Omega \rightarrow \Omega'$ stetig sowie $\mathcal{F} = \sigma(\mathcal{O})$ und $\mathcal{F}' = \sigma(\mathcal{O}')$ die Borel'schen σ -Algebren. Dann ist f nach \mathcal{F}/\mathcal{F}' -messbar.
4. Eine reelle Funktion f (d.h. $f : \Omega \rightarrow \mathbb{R}$) ist genau dann messbar, wenn $\{\omega : f(\omega) \leq x\} \in \mathcal{F}$ für alle $x \in \mathbb{Q}$.
5. Eine einfache Funktion $f = \sum_{k=1}^n c_k 1_{A_k}$ mit paarweise verschiedenen $c_1, \dots, c_n \in \overline{\mathbb{R}}$ und $A_1, \dots, A_n \subseteq \Omega$ ist genau dann messbar, wenn $A_1, \dots, A_n \in \mathcal{F}$.

Beweis. 1. Die 'genau dann'-Richtung ist klar. Für die 'wenn'-Richtung verwenden wir Lemma 4.2 und erhalten $f^{-1}(\mathcal{F}') = f^{-1}(\sigma(\mathcal{C}')) = \sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(\mathcal{F}) = \mathcal{F}$. Damit ist f nach \mathcal{F}/\mathcal{F}' -messbar.

2. Wir schreiben direkt $(g \circ f)^{-1}(\mathcal{F}'') = f^{-1}(g^{-1}(\mathcal{F}'')) \subseteq f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$, was die Behauptung bereits zeigt.

3. Nach Definition der Borel'schen σ -Algebra ist \mathcal{O}' ein Erzeuger für $\mathcal{B}(\Omega')$. Da f stetig ist (d.h. $f^{-1}(\mathcal{O}') \subseteq \mathcal{O}$) folgt $f^{-1}(\mathcal{O}') \subseteq \mathcal{O} \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega)$. Nach 1. ist f also $\mathcal{B}(\Omega)/\mathcal{B}(\Omega')$ -messbar.

4. Wir verwenden 1. Ist nämlich $\Omega' = \mathbb{R}$, so wird nach Lemma 4.2 $\mathcal{B}(\Omega')$ von $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{Q}\}$ erzeugt. Also ist ein reelles f genau dann messbar, wenn $f^{-1}(\mathcal{C}') = \{\{\omega : f(\omega) \leq x\} : x \in \mathbb{R}\} \subseteq \mathcal{F}$.

5. Sei $A := \left(\bigcup_{k=1}^n A_k\right)^c \in \mathcal{F}$. Dann ist $f^{-1}(\mathcal{B}(\overline{\mathbb{R}})) = \left\{A \cup \bigcup_{j \in J} A_j, \bigcup_{j \in J} A_j : J \subseteq \{1, \dots, n\}\right\}$, woraus die Behauptung folgt. \square

Lemma 4.6 (Messbarkeit und Rechenregeln). Sei (Ω, \mathcal{F}) ein Messraum.

1. Sind $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ messbar. Dann sind auch fg , sowie $af + bg$ für $a, b \in \mathbb{R}$ und f/g messbar, falls $g(\omega) \neq 0$ für alle $\omega \in \Omega$.
2. Sind $f_1, f_2, \dots : \Omega \rightarrow \overline{\mathbb{R}}$ messbar, so sind auch

$$\sup_{n \in \mathbb{N}} f_n, \quad \inf_{n \in \mathbb{N}} f_n, \quad \limsup_{n \rightarrow \infty} f_n, \quad \liminf_{n \rightarrow \infty} f_n$$

messbar. Falls existent, ist auch $\lim_{n \rightarrow \infty} f_n$ messbar.

Beweis. 1. Betrachte die Abbildung $\psi : \Omega \rightarrow \mathbb{R}^2$, definiert durch $\psi(\omega) = (f(\omega), g(\omega))$. Man überlegt sich leicht, dass ψ nach $\mathcal{F}/\mathcal{B}(\mathbb{R}^2)$ -messbar ist. Weiter ist $(x, y) \mapsto ax + by$ und $(x, y) \mapsto xy$ auf \mathbb{R} und $(x, y) \mapsto x/y$ auf $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$ stetig, also messbar nach Lemma 4.5.3. Damit folgen die Behauptungen nach Lemma 4.5.2.

2. Wir zeigen nur die Messbarkeit von $\sup_{n \in \mathbb{N}} f_n$. Die anderen Aussagen folgen dann mittels $\inf_{n \in \mathbb{N}} f_n = -\sup_{n \in \mathbb{N}}(-f_n)$, sowie $\limsup_{n \rightarrow \infty} f_n = \inf_{n \in \mathbb{N}} \sup_{k \geq n} f_k$ und $\liminf_{n \rightarrow \infty} f_n = \sup_{n \in \mathbb{N}} \inf_{k \geq n} f_k$. Es gilt für $x \in \mathbb{R}$ nach Lemma 4.5.4

$$\left\{\omega : \sup_{n \in \mathbb{N}} f_n(\omega) \leq x\right\} = \bigcap_{n=1}^{\infty} \underbrace{\left\{\omega : f_n(\omega) \leq x\right\}}_{\in \mathcal{F}} \in \mathcal{F}$$

und die Behauptung ist gezeigt. \square

Korollar 4.7 (Messbarkeit von Positiv- und Negativteil). Sei (Ω, \mathcal{F}) ein Messraum und $f : \Omega \rightarrow \overline{\mathbb{R}}$. Dann ist f genau dann messbar, wenn $f^+ := f \vee 0$ und $f^- := (-f) \vee 0$ messbar sind. Dann ist auch $|f|$ messbar.

Beweis. Es gilt $f = f^+ - f^-$ und $|f| = f^+ + f^-$. Damit folgt die Behauptung aus Lemma 4.6.2. \square

Theorem 4.8 (Approximation mit messbaren Funktionen). Sei (Ω, \mathcal{F}) ein Messraum und $f : \Omega \rightarrow \overline{\mathbb{R}}$ messbar. Dann gibt es eine Folge $f_1, f_2, \dots : \Omega \rightarrow \overline{\mathbb{R}}$ von einfachen Funktionen mit⁸ $f_n \uparrow f$.

Beweis. Wir schreiben einfach für⁹ $\omega \in \Omega, n \in \mathbb{N}$

$$f_n(\omega) = n \wedge 2^{-n} \lceil 2^n f(\omega) \rceil,$$

und bemerken, dass $f_n \uparrow f$ per Konstruktion gilt. Außerdem ist $\omega \mapsto \lceil 2^n f(\omega) \rceil$ nach Lemma 4.5.4 messbar, wenn f messbar ist. \square

4.2 Definition

Die Konstruktion des Integrals einer Funktion f nach einem Maß erfolgt nun in mehreren Schritten. Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Wir bemerken, dass wir für das Integral von $f : \Omega \rightarrow \mathbb{R}$ nach μ verschiedene synonyme Schreibweisen verwenden, nämlich

$$\mu[f] = \int f d\mu = \int f(\omega) \mu(d\omega). \quad (4.1)$$

Die Definition des Integrals erfolgt zunächst für einfache Funktionen und anschließend (siehe Theorem 4.8) durch Approximation für allgemeine nicht-negative messbare Funktionen. Das Integral für (nicht notwendig nicht-negative) messbare Funktionen wird dann durch das Integral des Positiv- und Negativteils definiert.

Später werden wir Erwartungswerte von Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$ berechnen. Dann ist $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und wir verwenden die Schreibweise

$$\mathbf{E}[X] := \mathbf{P}[X],$$

wobei $\mathbf{P}[X]$ wie in (4.1) definiert ist.

Definition 4.9 (Integral von einfachen Funktionen). Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f = \sum_{k=1}^m c_k 1_{A_k}$ eine einfache Funktion mit $c_1, \dots, c_m \geq 0, A_1, \dots, A_m \in \mathcal{F}$. Dann heißt

$$\mu[f] := \int f d\mu := \sum_{k=1}^m c_k \mu(A_k)$$

das Integral von f bezüglich μ .

⁸Analog zu '↓' schreiben wir für $x, x_1, x_2, \dots \in \mathbb{R}$, dass $x_n \uparrow x$, falls $x_1 \leq x_2 \leq \dots$ und $x_n \xrightarrow{n \rightarrow \infty} x$. Für Funktionen $f, f_1, f_2, \dots : \Omega \rightarrow \overline{\mathbb{R}}$ bedeutet $f_n \uparrow f$, dass $f_n(\omega) \uparrow f(\omega)$ für alle $\omega \in \Omega$.

⁹Hier ist $[x]$ für $x \in \mathbb{R}$ die größte ganze Zahl, die kleiner als x ist, also $[x] := \sup\{n \in \mathbb{Z} : n \leq x\}$.

Bemerkung 4.10 (Wohldefiniertheit des Integrals). Wir müssen sicherstellen, dass obiges Integral wohldefiniert ist. Sei dazu $f = \sum_{l=1}^n d_l 1_{B_l}$ eine weitere Darstellung von f mit $d_1, \dots, d_n \geq 0$ und $B_1, \dots, B_n \in \mathcal{F}$. Dann gilt

$$\sum_{k=1}^m c_k \mu(A_k) = \sum_{k=1}^m \sum_{l=1}^n c_k \mu(A_k \cap B_l) = \sum_{k=1}^m \sum_{l=1}^n d_l \mu(A_k \cap B_l) = \sum_{l=1}^n d_l \mu(B_l),$$

woraus die Wohldefiniertheit folgt.

Lemma 4.11 (Einfache Eigenschaften). Seien f, g nicht-negative, einfache Funktionen und $\alpha \geq 0$. Dann gilt¹⁰

$$\mu[\alpha f + bg] = \alpha \mu[f] + b \mu[g], \quad f \leq g \Rightarrow \mu[f] \leq \mu[g].$$

Beweis. Klar. □

Beispiel 4.12 (Das Integral von Indikatorfunktionen und Riemann-Integral). Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A \in \mathcal{F}$. Dann ist $f = 1_A$ eine einfache Funktion und es gilt

$$\mu[f] = \mu(A)$$

nach Definition 4.9. Es sei hierbei schon bemerkt, dass die Funktion $f = 1_A$ nicht mehr stückweise stetig sein muss. (Sei etwa A das in Beispiel 2.10 betrachtete Cantor'sche Kontinuum.) Deswegen ist es nicht klar, dass die Funktion 1_A integrierbar bezüglich des Riemann-Integrals ist.

Definition 4.13 (Das Integral von messbaren, nicht-negativen Funktionen). Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \rightarrow \overline{\mathbb{R}}_+$ messbar. Das Integral von f bezüglich μ ist gegeben durch

$$\mu[f] := \int f d\mu := \int f(\omega) \mu(d\omega) := \sup\{\mu[g] : g \text{ einfach, nicht-negativ, } g \leq f\}. \quad (4.2)$$

Bemerkung 4.14 (Das Integral als Fortsetzung). Nach Lemma 4.11 ist klar, dass die Definition von $\mu[f]$ für einfache, nicht-negative Funktionen aus Definition 4.9 und Definition 4.13 übereinstimmt. Damit ist obige Definition eine Fortsetzung von $\mu[f]$ auf den Raum der nicht-negativen, messbaren Funktionen.

Weiter ist wichtig zu sehen, dass nach Theorem 4.8 jede der in Definition 4.13 auftretenden Funktionen beliebig genau (punktweise) durch einfache Funktionen approximierbar ist. Insbesondere beinhaltet die Menge, deren Supremum in (4.2) gesucht wird, auch einfache Funktionen g , die beliebig dicht an f liegen.

Proposition 4.15 (Eigenschaften des Integrals). Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f, g, f_1, f_2, \dots : \Omega \rightarrow \overline{\mathbb{R}}_+$ messbar. Dann gilt:

1. Für $f \leq g$ gilt $\mu[f] \leq \mu[g]$.
2. Es gilt monotone Konvergenz, d.h.

$$f_n \uparrow f \quad \Rightarrow \quad \mu[f_n] \uparrow \mu[f].$$

¹⁰Für Funktionen $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ schreiben wir $f \leq g$, falls $f(\omega) \leq g(\omega)$ für alle $\omega \in \Omega$ gilt.

3. Für $a, b \geq 0$ gilt $\mu[af + bg] = a\mu[f] + b\mu[g]$.

Beweis. 1. ist klar nach der Definition des Integrals. Für 2. ist zunächst, da $f_n \leq f$ für $n = 1, 2, \dots$

$$\lim_{n \rightarrow \infty} \mu[f_n] = \sup_{n \in \mathbb{N}} \mu[f_n] \leq \mu[f].$$

Für ' \geq ' genügt es zu zeigen, dass

$$\mu[g] \leq \sup_{n \in \mathbb{N}} \mu[f_n] \quad (4.3)$$

für alle einfachen Funktionen $g \leq f$. Sei also $g = \sum_{k=1}^m c_k 1_{A_k} \leq f$ für disjunkte Mengen A_1, \dots, A_m und $c_1, \dots, c_m > 0$. Für $\varepsilon > 0$ und $n = 1, 2, \dots$ sei $B_n^\varepsilon := \{f_n \geq (1 - \varepsilon)g\}$. Da $f_n \uparrow f$ und $g \leq f$ gilt $\bigcup_{n=1}^\infty B_n^\varepsilon = \Omega$ für alle $\varepsilon > 0$. Also ist

$$\begin{aligned} \mu[f_n] &\geq \mu[(1 - \varepsilon)g 1_{B_n^\varepsilon}] = \sum_{k=1}^m (1 - \varepsilon)c_k \mu(A_k \cap B_n^\varepsilon) \\ &\xrightarrow{n \rightarrow \infty} \sum_{k=1}^m (1 - \varepsilon)c_k \mu(A_k) = (1 - \varepsilon)\mu[g]. \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt (4.3).

Für 3. seien $f_1, g_1, f_2, g_2, \dots$ einfache Funktionen mit $f_n \uparrow f$ und $g_n \uparrow g$. Dann ist $af_n + bg_n \uparrow af + bg$ und es folgt

$$\mu[af + bg] = \lim_{n \rightarrow \infty} \mu[af_n + bg_n] = \lim_{n \rightarrow \infty} a\mu[f_n] + b\mu[g_n] = a\mu[f] + b\mu[g]$$

aus 3. wegen Lemma 4.11. □

Wir können nun das Integral für messbare Funktionen definieren. Hierzu sei zunächst bemerkt, dass für $f : \Omega \rightarrow \overline{\mathbb{R}}$ messbar gilt, dass $f^+, f^- \leq |f|$. Ist insbesondere $\mu[|f|] < \infty$, so ist auch $\mu[f^+], \mu[f^-] < \infty$.

Definition 4.16 (Integral messbarer Funktionen). Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \rightarrow \overline{\mathbb{R}}$ messbar. Dann heißt f nach μ integrierbar, falls $\mu[|f|] < \infty$ und wir definieren

$$\mu[f] := \int f(\omega) \mu(d\omega) := \int f d\mu := \mu[f^+] - \mu[f^-]. \quad (4.4)$$

Außerdem setzen wir

$$\mathcal{L}^1(\mu) := \left\{ f : \Omega \rightarrow \overline{\mathbb{R}} : \mu[|f|] < \infty \right\}$$

Für $A \in \mathcal{F}$ schreiben wir außerdem

$$\mu[f, A] := \int_A f d\mu := \mu[f 1_A].$$

Bemerkung 4.17 (Erweiterung des Integrals und \mathcal{L}^p -Räume). 1. Ist höchstens einer der beiden Terme $\mu[f^+]$ oder $\mu[f^-]$ unendlich, so definieren wir das Integral $\mu[f]$ weiterhin mittels (4.4). In anderen Fällen bleibt das Integral undefiniert.

2. Die Funktionenräume $\mathcal{L}^p(\mu) := \left\{ f : \Omega \rightarrow \overline{\mathbb{R}} : \mu[|f|^p] < \infty \right\}$, $p > 0$, werden in Abschnitt 5 eine besondere Rolle spielen.

4.3 Eigenschaften des Integrals

Wir stellen zunächst einige Eigenschaften des eingeführten Integralbegriffes fest. Diese sind etwa die Monotonie und Linearität. Weiter werden wir sehen, dass sich das Integral einer Funktion nicht ändert, wenn man sie auf einer Nullmenge abändert; siehe Proposition 4.20.

Proposition 4.18 (Einfache Eigenschaften des Integrals). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f, g \in \mathcal{L}^1(\mu)$. Dann gilt:*

1. Das Integral ist monoton, d.h.

$$f \leq g \text{ fast überall} \implies \mu[f] \leq \mu[g].$$

2. Es gilt die Dreiecksungleichung

$$|\mu[f]| \leq \mu[|f|].$$

3. Das Integral ist linear, seien also $a, b \in \mathbb{R}$, dann ist $af + bg \in \mathcal{L}^1(\mu)$ und

$$\mu[af + bg] = a\mu[f] + b\mu[g].$$

Beweis. Alle Eigenschaften folgen aus Proposition 4.15.1 und 3., sowie aus der Definition des Integrals für messbare Funktionen. \square

Proposition 4.19 (Substitutionssatz). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum, (Ω', \mathcal{F}') ein Messraum, $f : \Omega \rightarrow \Omega'$ messbar und $f_*\mu$ das Bildmaß von f aus Definition 3.22. Dann gilt für $g \in \mathcal{L}^1(f_*\mu)$, dass $g \circ f \in \mathcal{L}^1(\mu)$ und*

$$\mu[g \circ f] = f_*\mu[g].$$

Beweis. Es genügt, die Behauptung für einfache, nicht-negative Funktionen g zu zeigen. Der allgemeine Fall folgt dann mittels Approximation durch einfache Funktionen. Sei also $g = \sum_{k=1}^m c_k 1_{A'_k}$ mit $A'_k \in \mathcal{F}'$. Dann ist $g \circ f = \sum_{k=1}^m c_k 1_{f \in A'_k}$ und

$$\mu[g \circ f] = \sum_{k=1}^m c_k \mu(f \in A'_k) = \sum_{k=1}^m c_k f_*\mu(A'_k) = f_*\mu[g].$$

\square

Proposition 4.20 (Integrale und Eigenschaften fast überall). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \rightarrow \overline{\mathbb{R}}_+$ messbar.*

1. Es ist $f = 0$ fast überall genau dann, wenn $\mu[f] = 0$.

2. Falls $\mu[f] < \infty$, so ist $f < \infty$ fast überall.

Beweis. 1. Sei $N := \{f > 0\} \in \mathcal{F}$.

' \Rightarrow ': Es gilt $\mu(N) = 0$. Dann ist $f \leq \infty \cdot 1_N$, also gilt wegen Proposition 4.15.2

$$0 \leq \mu[f] \leq \mu[\infty, N] = \lim_{n \rightarrow \infty} \mu[n, N] = 0.$$

Für ' \Leftarrow ' sei $N_n := \{f \geq 1/n\}$ und damit $N_n \uparrow N$ und $nf \geq 1_{N_n}$, also

$$0 = \mu[f] \geq \frac{1}{n} \mu(N_n).$$

Damit ist $\mu(N_n) = 0$ und $\mu(N) = \mu(\bigcup_{n=1}^{\infty} N_n) = 0$.

Für 2. sei $A := \{f = \infty\}$. Wegen $f1_{f \geq n} \geq n1_{f \geq n}$ ist

$$\mu(A) = \mu[1_A] \leq \mu[1_{f \geq n}] \leq \frac{1}{n} \mu[f, 1_{f \geq n}] \leq \frac{1}{n} \mu[f] \xrightarrow{n \rightarrow \infty} 0.$$

Damit ist $\mu(f = \infty) = 0$, also $f < \infty$ fast überall; siehe Bemerkung 3.13. \square

Zum Abschluss dieses Abschnittes stellen wir noch die Beziehung des (Lebesgue-)Integrals zum Riemann-Integral dar.

Definition 4.21 (Treppenfunktion und Riemann-Integral). Eine Treppenfunktion $f : \mathbb{R} \rightarrow \mathbb{R}$, für die es eine Darstellung

$$f(t) = \sum_{j=-\infty}^{\infty} a_j 1_{t_{j-1}, t_j}(t)$$

mit $t_{j-1} \leq t_j, j \in \mathbb{Z}$ gibt, wobei $a_j \in \mathbb{R}, j \in \mathbb{Z}$. Eine messbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt eigentlich Riemann-integrierbar, falls $\lambda[|f|] < \infty$ und es Treppen-Funktionen $f_1^+, f_1^-, f_2^+, f_2^-, \dots$ gibt mit $f_n^- \leq f \leq f_n^+$ und $\lambda[f_n^+ - f_n^-] \xrightarrow{n \rightarrow \infty} 0$. Das Riemann-Integral von f ist dann durch $\lambda[f]$ definiert. (Insbesondere stimmen dann Riemann- und Lebesgue-Integral überein.

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt uneigentlich Riemann-integrierbar, wenn $f1_K$ für alle kompakten Intervalle $K \subseteq \mathbb{R}$ eigentlich Riemann-integrierbar ist und $\lambda[f1_{[-n, n]}]$ konvergiert. Dieser Grenzwert ist dann das Riemann-Integral von f bezüglich λ .

Proposition 4.22 (Riemann-Integrierbarkeit). Die Funktion $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ habe eine diskrete Menge von Sprungstellen. Dann ist f genau dann eigentlich Riemann-integrierbar, wenn f Lebesgue-integrierbar ist und es gilt

$$\lambda[f] = \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} f(s_{n,k})(t_{n,k} - t_{n,k-1}) \quad (4.5)$$

für $0 = t_{n,0} \leq \dots \leq t_{n,k_n} = t$ mit $\max_k |t_{n,k} - t_{n,k-1}| \xrightarrow{n \rightarrow \infty} 0$ und beliebiges $t_{n,k-1} \leq s_{n,k} \leq t_{n,k}$.

Beweis. Es genügt, die Behauptung für stetiges f zu zeigen. Andernfalls kann man f in die stetigen Teilabschnitte zerlegen. Weiter genügt es, die Behauptung für f mit kompaktem Träger K zu zeigen. Da f auf K gleichmäßig stetig ist, wählt man zunächst $\varepsilon_n \downarrow 0$ und $t_{n,0} \leq \dots \leq t_{n,k_n}$ so, dass $K \subseteq [t_{n,0}, t_{n,k_n}]$ und $\max_{t_{n,k-1} \leq s < t_{n,k}} |f(t_{n,k-1}) - f(s)| < \varepsilon_n$. Nun ist es leicht, Treppenfunktionen f_n^+ und f_n^- zu finden, so dass $f_n^- \leq f \leq f_n^+$ und $\|f_n^+ - f_n^-\| \leq \varepsilon_n$. Daraus folgt die erste Behauptung. Die Formel (4.5) gilt wegen der gleichmäßigen Approximation der Funktion f . \square

Beispiel 4.23 (Unterschiede zwischen Riemann- und Lebesgue-Integral). 1.

Wir beginnen mit einer Funktion, die Lebesgue- aber nicht Riemann-integrierbar ist. Sei $f = 1_{[0,1] \cap \mathbb{Q}}$. Dann ist $1_{[0,1]} \leq f^+$ für jede reguläre Treppenfunktion $f^+ \geq f$ und $f^- \leq 0$ für jede reguläre Treppenfunktion $f^- \leq f$. Insbesondere ist f nicht Riemann-integrierbar.

2. Wie aus der Definition des Riemann-Integrals hervorgeht, ist jede eigentlich Riemann-integrierbare Funktion auch Lebesgue-integrierbar. Anders verhält es sich mit uneigentlich Riemann-integrierbaren Funktionen. Sei hierzu f gegeben durch $f(t) = \frac{(-1)^{\lceil t \rceil} + 1}{\lceil t \rceil}$. Dann ist

$$\lambda[f 1_{[0,2n]}] = \sum_{k=1}^{2n} \frac{(-1)^{k+1}}{k} = \sum_{k=1}^n \frac{1}{k} - \frac{1}{k+1} = 1 - \frac{1}{n+1} \xrightarrow{n \rightarrow \infty} 1.$$

Damit ist f uneigentlich Riemann-integrierbar. Allerdings gilt

$$\lambda[|f|] = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

Also ist f nicht Lebesgue-integrierbar.

4.4 Konvergenzsätze

Man kann sich fragen, ob es wirklich so wichtig ist, dass man mehr Funktionen bezüglich des Lebesgue-Integrals als bezüglich des Riemann Integrals integrieren kann. Es kommen schließlich in Anwendungen meistens Riemann-integrierbare Funktionen vor. Es gibt allerdings einen weiteren Vorteil des Lebesgue-Integrals, auf den wir nun eingehen werden. Es gelten nämlich Vertauschungsgesetze von Grenzwerten und Integralen (die hier Konvergenzsätze genannt werden), die relativ wenige Voraussetzungen benötigen. Die Wichtigsten sind der Satz von der monotonen und der von der majorisierten Konvergenz.

Theorem 4.24 (Monotone Konvergenz). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum, $f_1, f_2, \dots \in \mathcal{L}^1(\mu)$ und $f : \Omega \rightarrow \overline{\mathbb{R}}$ messbar mit $f_n \uparrow f$ fast überall. Dann gilt*

$$\lim_{n \rightarrow \infty} \mu[f_n] = \mu[f],$$

wobei beide Seiten den Wert ∞ annehmen können.

Beweis. Sei $N \in \mathcal{F}$ so, dass $\mu(N) = 0$ und $f_n(\omega) \uparrow f(\omega)$ für $\omega \notin N$. Setze $g_n := (f_n - f_1)1_{N^c} \geq 0$. Damit ist $g_n \uparrow (f - f_1)1_{N^c} =: g$ und mit Proposition 4.18, Proposition 4.20 und Proposition 4.15.2 gilt

$$\mu[f_n] = \mu[f_1] + \mu[g_n] \xrightarrow{n \rightarrow \infty} \mu[f_1] + \mu[g] = \mu[f].$$

□

Theorem 4.25 (Lemma von Fatou). *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f_1, f_2, \dots : \Omega \rightarrow \overline{\mathbb{R}}_+$ messbar. Dann gilt*

$$\liminf_{n \rightarrow \infty} \mu[f_n] \geq \mu[\liminf_{n \rightarrow \infty} f_n].$$

Beweis. Für alle $k \geq n$ gilt $f_k \geq \inf_{k \geq n} f_k$ und damit

$$\inf_{k \geq n} \mu[f_k] \geq \mu[\inf_{k \geq n} f_k]$$

wegen Proposition 4.15.1. Deshalb gilt mit $n \rightarrow \infty$

$$\liminf_{n \rightarrow \infty} \mu[f_n] \geq \sup_{n \in \mathbb{N}} \mu[\inf_{k \geq n} f_k] = \mu[\liminf_{n \rightarrow \infty} f_n]$$

wegen der monotonen Konvergenz, da $\inf_{k \geq n} f_k \uparrow \liminf_{n \rightarrow \infty} f_n$. □

Theorem 4.26 (Satz von der majorisierten Konvergenz). Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f, g, f_1, f_2, \dots : \Omega \rightarrow \overline{\mathbb{R}}$ messbar mit $|f_n| \leq g$ fast überall, $\lim_{n \rightarrow \infty} f_n = f$ und $g \in \mathcal{L}^1(\mu)$. Dann gilt

$$\lim_{n \rightarrow \infty} \mu[f_n] = \mu[f].$$

Beweis. Ohne Einschränkung gilt $|f_n| \leq g$ überall. Wir verwenden das Lemma von Fatou, sowie $g - f_n, g + f \geq 0$, also

$$\begin{aligned} \mu[g + f] &\leq \liminf_{n \rightarrow \infty} \mu[g + f_n] = \mu[g] + \liminf_{n \rightarrow \infty} \mu[f_n], \\ \mu[g - f] &\leq \liminf_{n \rightarrow \infty} \mu[g - f_n] = \mu[g] - \limsup_{n \rightarrow \infty} \mu[f_n]. \end{aligned}$$

Nach Subtrahieren von $\mu[g]$ ist also

$$\mu[f] \leq \liminf_{n \rightarrow \infty} \mu[f_n] \leq \limsup_{n \rightarrow \infty} \mu[f_n] \leq \mu[f]. \quad \square$$

Beispiel 4.27 (Beispiele zu den Konvergenzsätzen). 1. Im Lemma von Fatou ist nicht vorausgesetzt, dass eines der f_n integrierbar ist. Wir geben nun ein Beispiel dafür, dass im Lemma von Fatou wirklich ' $<$ ' und nicht '=' gilt. Sei hierzu λ das Lebesgue-Maß und $f_n = 1/n$ (also insbesondere f_n konstant), $n = 1, 2, \dots$. Dann gilt $f_n \downarrow 0$, also

$$\liminf_{n \rightarrow \infty} \mu[f_n] = \infty > 0 = \mu[0] = \mu[\liminf_{n \rightarrow \infty} f_n].$$

2. Im Satz von der majorisierten Konvergenz kann auf die Voraussetzung, dass $|f_n| \leq g$ und $g \in \mathcal{L}^1(\mu)$ nicht verzichtet werden. Sei etwa λ das Lebesgue-Maß auf $[0, 1]$ und $f_n = n \cdot 1_{[0, 1/n]}$. Dann ist $\sup_{n \in \mathbb{N}} f_n(x) = \sup\{n : x \leq 1/n\} = \left\lceil \frac{1}{x} \right\rceil$ ¹¹. Also existiert kein $g \in \mathcal{L}^1(\lambda)$ mit $f_n \leq g$. Außerdem ist $\lim_{n \rightarrow \infty} f_n = 0$ fast überall (wobei $\{0\}$ die Ausnahmemeenge ist) und

$$\lim_{n \rightarrow \infty} \mu[f_n] = 1 \neq 0 = \mu[\lim_{n \rightarrow \infty} f_n].$$

Anders verhält es sich für $f_n = n \cdot 1_{[0, 1/n^2]}$. Hier ist

$$\sup_{n \in \mathbb{N}} f_n(x) = \sup\{n : x \leq 1/n^2\} = \left\lceil \frac{1}{\sqrt{x}} \right\rceil \leq \frac{1}{\sqrt{x}} =: g(x).$$

Hier ist einerseits $g \in \mathcal{L}^1(\lambda)$, der Satz von der majorisierten Konvergenz ist also anwendbar, andererseits ist $\lim_{n \rightarrow \infty} f_n = 0$ fast überall und

$$\lim_{n \rightarrow \infty} \mu[f_n] = \lim_{n \rightarrow \infty} \frac{1}{n} = 0 = \mu[0] = \mu[\lim_{n \rightarrow \infty} f_n].$$

5 \mathcal{L}^p -Räume

Im ganzen folgenden Abschnitt sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Wir werden uns nun mit den messbaren Funktionen $f : \Omega \rightarrow \overline{\mathbb{R}}$ beschäftigen, für die $\mu[|f|^p] < \infty$ gilt. Die resultierenden Funktionenräume $\mathcal{L}^p(\mu)$ werden wir als normierte, vollständige Räume (Proposition 5.7) erkennen, was automatisch zu einem neuen Konvergenzbegriff führt. Weiter wird der Raum $\mathcal{L}^2(\mu)$ eine große Rolle spielen. Da man diesen mit einem Skalarprodukt (nämlich $\langle f, g \rangle := \mu[fg]$) ausstatten kann, stehen hier allgemeine Aussagen zur Verfügung, etwa der Satz von Riesz-Fréchet (Proposition 5.10). Dies werden wir verwenden, um σ -endliche Maße mit Dichte durch den Satz von Radon-Nikodým (Korollar 5.16) zu charakterisieren.

¹¹Mit $\lceil x \rceil := \sup\{n \in \mathbb{Z} : n \leq x\}$ bezeichnen wir die Abrundungsfunktion.

5.1 Grundlagen

Bereits in Bemerkung 4.17 haben wir die Räume $\mathcal{L}^p(\mu)$ erwähnt. Durch die Definition des Integrals im letzten Abschnitt können wir diese nun näher beleuchten. Vor allem zeigen wir die wichtige Hölder- und die Minkowski-Ungleichung, siehe Proposition 5.2.

Definition 5.1 ($\mathcal{L}^p(\mu)$ -Räume). Sei $0 < p \leq \infty$. Wir setzen

$$\mathcal{L}^p := \mathcal{L}^p(\mu) := \{f : \Omega \rightarrow \overline{\mathbb{R}} \text{ messbar mit } \|f\|_p < \infty\}$$

für

$$\|f\|_p := (\mu[|f|^p])^{1/p}, \quad 0 < p < \infty \quad (5.1)$$

und

$$\|f\|_\infty := \inf\{K : \mu(|f| > K) = 0\}.$$

Auf den Räumen \mathcal{L}^p , $p \geq 1$ zeigen wir nun eine Dreiecksungleichung, die Minkowski-Ungleichung. Außerdem sei bemerkt, dass die Hölder-Ungleichung im Spezialfall $p = q = 2$ auch Cauchy-Schwartz-Ungleichung heißt.

Proposition 5.2 (Hölder und Minkowski's Ungleichung). Seien f, g messbar.

1. Sei $0 < p, q, r \leq \infty$ so, dass $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Dann gilt

$$\|fg\|_r \leq \|f\|_p \|g\|_q \quad (\text{Hölder-Ungleichung})$$

2. Für $1 \leq p \leq \infty$ ist

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \quad (\text{Minkowski-Ungleichung})$$

Beweis. Wir starten mit dem Beweis der Hölder-Ungleichung. Im Fall $p = \infty$ oder $q = \infty$ ist die Aussage klar, sei also $p, q < \infty$. Ist entweder $\|f\|_p = 0$, $\|f\|_p = \infty$, $\|g\|_q = 0$ oder $\|g\|_q = \infty$, ist die Aussage ebenso klar. Sei also o.E. $f, g \geq 0$ und $0 < \|f\|_p, \|g\|_q < \infty$ und

$$\tilde{f} := \frac{f}{\|f\|_p}, \quad \tilde{g} = \frac{g}{\|g\|_q}.$$

Dann ist zu zeigen, dass $\|\tilde{f}\tilde{g}\|_r \leq 1$. Wegen der Konvexität der Exponentialfunktion ist

$$(xy)^r = \exp\left(\frac{r}{p}p \log x + \frac{r}{q}q \log y\right) \leq \frac{r}{p}x^p + \frac{r}{q}y^q$$

und damit

$$\|\tilde{f}\tilde{g}\|_r^r = \mu[(\tilde{f}\tilde{g})^r] \leq \frac{r}{p}\mu[\tilde{f}^p] + \frac{r}{q}\mu[\tilde{g}^q] = 1$$

und die Behauptung folgt.

Zum Beweis der Minkowski-Ungleichung bemerken wir zunächst, dass in den Fällen $p = 1$ und $p = \infty$ die Behauptung klar ist. Im Falle $1 < p < \infty$ gilt mit $q = p/(p-1)$ und $r = 1/p + 1/q = 1$ mit der Hölder-Ungleichung

$$\begin{aligned} \|f + g\|_p^p &\leq \mu[|f| \cdot |f + g|^{p-1}] + \mu[|g| \cdot |f + g|^{p-1}] \\ &\leq \|f\|_p \cdot \|(f + g)^{p-1}\|_q + \|g\|_p \cdot \|(f + g)^{p-1}\|_q \\ &= (\|f\|_p + \|g\|_p) \cdot \|f + g\|_p^{p-1}, \end{aligned}$$

da $\|(f + g)^{p-1}\|_q = \|(f + g)^{q(p-1)}\|_1^{1/q} = \|(f + g)^p\|_1^{(p-1)/p} = \|f + g\|_p^{p-1}$. Dividieren durch $\|f + g\|_p^{p-1}$ bringt das Resultat. \square

Proposition 5.3 (Zusammenhang zwischen \mathcal{L}^r und \mathcal{L}^q). Sei μ endlich und $1 \leq r < q \leq \infty$. Dann ist $\mathcal{L}^q(\mu) \subseteq \mathcal{L}^r(\mu)$.

Beweis. Die Behauptung ist klar für $q = \infty$. Sei also $q < \infty$. Wir verwenden die Hölder-Ungleichung. Es gilt für $f \in \mathcal{L}^q$, da $\|1\|_p < \infty$ wegen der Endlichkeit von μ ,

$$\|f\|_r = \|1 \cdot f\|_r \leq \|1\|_p \cdot \|f\|_q < \infty \quad (5.2)$$

für $\frac{1}{p} = \frac{1}{r} - \frac{1}{q} > 0$, woraus die Behauptung sofort folgt. \square

Bemerkung 5.4 (Gegenbeispiel für σ -endliches μ). Sicherlich gilt Proposition 5.3 nicht, falls μ nicht endlich ist. Sei etwa λ das eindimensionale Lebesgue-Maß und $f : x \mapsto \frac{1}{x} \cdot 1_{x>1}$. Dann ist $f \in \mathcal{L}^2(\lambda)$, aber $f \notin \mathcal{L}^1(\lambda)$.

5.2 \mathcal{L}^p -Konvergenz

Wir haben im Satz von der majorisierten Konvergenz (Theorem 4.26) gesehen, dass für eine Folge von Funktionen, die fast überall konvergiert, oft auch deren Integrale konvergieren. Die hier betrachtete \mathcal{L}^p -Konvergenz geht jetzt von Konvergenz von Integralen aus. Wir werden sehen, dass der resultierende Konvergenzbegriff zur Folge hat, dass jede Cauchy-Folge konvergiert (Proposition 5.7).

Definition 5.5 (Konvergenz im p -ten Mittel). Eine Folge f_1, f_2, \dots in $\mathcal{L}^p(\mu)$ konvergiert in $\mathcal{L}^p(\mu)$ (oder im p -ten Mittel) gegen $f \in \mathcal{L}^p(\mu)$, falls

$$\|f_n - f\|_p \xrightarrow{n \rightarrow \infty} 0.$$

Wir schreiben dann auch $f_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p f$.

Proposition 5.6 (Konvergenz im p -ten und q -ten Mittel). Sei $\mu(\Omega) < \infty$, $1 \leq r < q \leq \infty$ und $f, f_1, f_2, \dots \in \mathcal{L}^q$. Falls $f_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^q f$, so auch $f_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^r f$.

Beweis. Die Behauptung ist klar für $q = \infty$, sei also $q < \infty$. Aus (5.2) folgt sofort dass $\|f - g\|_r \leq \|f - g\|_q$, woraus die Behauptung bereits folgt. \square

Proposition 5.7 (Vollständigkeit von \mathcal{L}^p). Sei $p \geq 1$ und f_1, f_2, \dots eine Cauchy-Folge in \mathcal{L}^p . Dann gibt es ein $f \in \mathcal{L}^p$ mit $\|f_n - f\|_p \xrightarrow{n \rightarrow \infty} 0$.

Beweis. Sei $\varepsilon_1, \varepsilon_2, \dots$ summierbar, also z.B. $\varepsilon_n := 2^{-n}$. Da f_1, f_2, \dots eine Cauchy-Folge ist, gibt es für jedes k einen Index n_k mit $\|f_m - f_n\|_p \leq \varepsilon_k$ für alle $m, n \geq n_k$. Insbesondere gilt

$$\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p \leq \sum_{k=1}^{\infty} \varepsilon_k < \infty.$$

Nach monotoner Konvergenz und der Minkowski'schen Ungleichung gilt damit

$$\left\| \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \right\|_p \leq \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty.$$

Damit ist insbesondere $\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty$ fast überall, also $f_{n_1}(\omega), f_{n_2}(\omega), \dots$ für fast alle ω eine Cauchy-Folge in \mathbb{R} . Damit gibt es eine messbare Abbildung f mit $f_{n_k} \xrightarrow{k \rightarrow \infty} f$ fast überall. Nach Fatou's Lemma gilt

$$\|f_n - f\|_p \leq \liminf_{k \rightarrow \infty} \|f_{n_k} - f_{n_k}\|_p \leq \sup_{m \geq n} \|f_m - f_n\|_p \xrightarrow{n \rightarrow \infty} 0,$$

d.h. $f_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p f$. \square

5.3 Der Raum \mathcal{L}^2

Für jedes $1 \leq p \leq \infty$ gilt, dass $\|af\|_p = |a| \cdot \|f\|_p$ für $a \in \mathbb{R}$ gilt. Zusammen mit der Minkowski'schen Ungleichung bedeutet das, dass $\mathcal{L}^p(\mu)$ ein reeller Vektorraum ist. Es ist entscheidend zu bemerken, dass die Abbildung $f \mapsto \|f\|_p$ eine Pseudo-Norm ist. Sie kann keine volle Norm¹² sein, weil $\|f\|_p = 0$ nach Proposition 4.20 nur impliziert, dass $\mu(f \neq 0) = 0$, aber nicht, dass $f = 0$. Wir werden deswegen im Folgenden Funktionen f und g miteinander identifizieren, wenn $f = g$ μ -fast überall gilt. Nach dem eben gesagtem ist dann $(\mathcal{L}^p, \|\cdot\|_p)$ ein normierter Raum. Da $\|\cdot\|_p$ nach Proposition 5.7 vollständig ist, ist $(\mathcal{L}^p, \|\cdot\|_p)$ sogar ein Banachraum für jedes $1 \leq p \leq \infty$. Als nächstes betrachten wir den Spezialfall $p = 2$. Wir definieren eine Abbildung $\mathcal{L}^2 \times \mathcal{L}^2 \rightarrow \mathbb{R}$ durch

$$\langle f, g \rangle := \mu[fg].$$

Dann ist $\langle \cdot, \cdot \rangle$ offensichtlich linear, symmetrisch und positiv semi-definit, also ein Skalarprodukt¹³. Wir schreiben konsequenterweise $f \perp g$ genau dann, wenn $\mu[fg] = 0$. Da $\|f\| := \|f\|_2 = \langle f, f \rangle^{1/2}$, ist $(\mathcal{L}^2, \langle \cdot, \cdot \rangle)$ also ein Hilbertraum.

Lemma 5.8 (Parallelogrammidentität). *Seien $f, g \in \mathcal{L}^2$. Dann gilt*

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

Beweis. Aus der Definition von $\|\cdot\|$ und der Symmetrie und Bilinearität von $\langle \cdot, \cdot \rangle$ folgt

$$\|f + g\|^2 + \|f - g\|^2 = \langle f + g, f + g \rangle + \langle f - g, f - g \rangle = 2\langle f, f \rangle + 2\langle g, g \rangle = 2\|f\|^2 + 2\|g\|^2.$$

□

Proposition 5.9 (Zerlegung von $f \in \mathcal{L}^2$). *Sei M ein abgeschlossener, linearer Teilraum von \mathcal{L}^2 . Dann hat jede Funktion $f \in \mathcal{L}^2$ eine fast sicher eindeutige Zerlegung $f = g + h$ mit $g \in M, h \perp M$.*

Beweis. Für $f \in \mathcal{L}^2$ definieren wir

$$d_f := \inf_{g \in M} \{\|f - g\|\}.$$

Wähle g_1, g_2, \dots mit $\|f - g_n\| \xrightarrow{n \rightarrow \infty} d_f$. Nach der Parallelogrammidentität gilt

$$4d_f^2 + \|g_m - g_n\|^2 \leq \|2f - g_m - g_n\|^2 + \|g_m - g_n\|^2 = 2\|f - g_m\|^2 + 2\|f - g_n\|^2 \xrightarrow{m, n \rightarrow \infty} 4d_f^2.$$

Also ist $\|g_m - g_n\|^2 \xrightarrow{m, n \rightarrow \infty} 0$, d.h. g_1, g_2, \dots ist eine Cauchy-Folge. Nach Proposition 5.7 gibt es damit ein $g \in \mathcal{L}^2$ mit $\|g_n - g\|_p \xrightarrow{n \rightarrow \infty} 0$, und wegen der Abgeschlossenheit von M

¹²Sei V ein reeller Vektorraum. Eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ heißt *Norm*, wenn (i) $\|x\| = 0$ gilt genau dann, wenn $x = 0$, (ii) $\|a \cdot x\| = |a| \cdot \|x\|$ für alle $x \in \mathbb{R}$ und (iii) $\|x + y\| \leq \|x\| + \|y\|$. Dann heißt das Paar $(V, \|\cdot\|)$ ein normierter Raum.

¹³Sei V ein reeller Vektorraum. Dann heißt eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ ein *Skalarprodukt*, falls (i) $\langle x, \alpha y + z \rangle = \alpha \langle x, y \rangle + \langle x, z \rangle$ für alle $x, y, z \in V$ und $\alpha \in \mathbb{R}$ (Linearität), (ii) $\langle x, y \rangle = \langle y, x \rangle$ (Symmetrie) und (iii) $\langle x, x \rangle > 0$ für jedes $x \in V \setminus \{0\}$ (Positive Definitheit). Durch ein Skalarprodukt wird die Norm $\|x\| := \langle x, x \rangle^{1/2}$ auf V definiert. Ist $(V, \|\cdot\|)$ vollständig, so heißt $(V, \langle \cdot, \cdot \rangle)$ ein *Hilbertraum*.

auch $g \in M$. Da $\|h\| = d_f$ für $h := f - g$, folgt für alle $t > 0, l \in M$, wegen der Definition von d_f ,

$$d_f^2 \leq \|h + tl\|^2 = d_f^2 + 2t\langle h, l \rangle + t^2\|l\|^2.$$

Da dies für alle t gilt, folgt $\langle h, l \rangle = 0$, also $h \perp M$.

Zum Beweis der Eindeutigkeit sei $g' + h'$ eine weitere Zerlegung von f . Dann ist aufgrund der Linearität von M einerseits $g - g' \in M$, andererseits ist fast sicher $g - g' = h - h' \perp M$, also $g - g' \perp g - g'$. Dies bedeutet $\|g - g'\| = \langle g - g', g - g' \rangle = 0$, also $g = g'$ fast überall. \square

Proposition 5.10 (Riesz-Fréchet). *Eine Abbildung $F : \mathcal{L}^2 \rightarrow \mathbb{R}$ ist genau dann stetig und linear wenn es ein $h \in \mathcal{L}^2$ gibt, so dass für alle $f \in \mathcal{L}^2$*

$$F(f) = \langle f, h \rangle.$$

Dann ist $h \in \mathcal{L}^2$ fast sicher eindeutig bestimmt.

Beweis. ' \Leftarrow ': Die Linearität von $f \mapsto \langle f, h \rangle$ folgt aus der Bilinearität von $\langle \cdot, \cdot \rangle$. Die Stetigkeit folgt aus der Cauchy-Schwartz Ungleichung mittels

$$|\langle f - f', h \rangle| \leq \|f - f'\| \cdot \|h\|.$$

' \Rightarrow ': Falls $F \equiv 0$ wähle $h = 0$. Falls $F \not\equiv 0$, ist $M = F^{-1}\{0\}$ ein (wegen der Stetigkeit von F) abgeschlossener und (wegen der Linearität von F) linearer Unterraum von \mathcal{L}^2 . Wähle $f' \in \mathcal{L}^2 \setminus M$ mit der (nach Proposition 5.9 fast sicher eindeutigen) orthogonalen Zerlegung $f' = g' + h'$ mit $g' \in M$ und $h' \perp M$. Da $f' \notin M$, ist $h' \neq 0$ und $F(h') = F(f') - F(g') = F(f') \neq 0$. Wir setzen $h'' = \frac{h'}{F(h')}$, so dass $h'' \perp M$ und $F(h'') = 1$ und es gilt für alle $f \in \mathcal{L}^2$

$$F(f - F(f)h'') = F(f) - F(f)F(h'') = 0.$$

d.h. $f - F(f)h'' \in M$, also insbesondere $\langle F(f)h'', h'' \rangle = \langle f, h'' \rangle$ und

$$F(f) = \frac{1}{\|h''\|^2} \cdot \langle F(f)h'', h'' \rangle = \frac{1}{\|h''\|^2} \cdot \langle f, h'' \rangle = \langle f, \frac{h''}{\|h''\|^2} \rangle.$$

Nun folgt die Behauptung mit $h = \frac{h''}{\|h''\|^2}$.

Zur Eindeutigkeit sei $\langle f, h_1 - h_2 \rangle = 0$ für alle $f \in \mathcal{L}^2$; insbesondere ist mit $f = h_1 - h_2$

$$\|h_1 - h_2\|^2 = \langle h_1 - h_2, h_1 - h_2 \rangle = 0,$$

also $h_1 = h_2$ μ -fast sicher. \square

Bemerkung 5.11 (Allgemeingültigkeit der letzten Aussagen). Lemma 5.8, sowie die Propositionen 5.9 und 5.10 gelten ebenso, falls man \mathcal{L}^2 durch einen anderen Hilbert-Raum ersetzt.

5.4 Satz von Radon-Nikodým

Aus der Vorlesung *Stochastik* sind bereits Verteilungen mit Dichte bekannt. Dieses Konzept wird nun aufgegriffen und in den Kontext von Integralen eingebettet. Sei hierzu ν ein weiteres Maß auf \mathcal{F} . Ziel ist es, Bedingungen anzugeben, wann das Maß ν durch eine Dichte darstellbar ist. Die Antwort findet sich im Satz von Radon-Nikodým (Korollar 5.16). Er ist ein Spezialfall des Lebesgue'schen Zerlegungssatzes, Theorem 5.15. Dieser zeigt, dass für je zwei σ -endliche Maße μ, ν das Maß ν (additiv) in zwei Anteile zerlegt werden kann: einen absolutstetigen bzgl. μ und einen zu μ singulären. Der absolutstetige Anteil hat dabei eine Dichte bzgl. μ . Zunächst müssen wir alle Begriffe erklären.

Definition 5.12 (Absolutstetige Maße). 1. Wir sagen, ν besitzt eine Dichte f bzgl. μ , falls für alle $A \in \mathcal{F}$

$$\nu(A) = \mu[f; A].$$

gilt. Wir schreiben dann $f = \frac{d\nu}{d\mu}$ und $\nu = f \cdot \mu$.

2. Das Maß ν heißt absolutstetig bzgl. μ , falls alle μ -Nullmengen auch ν -Nullmengen sind. Wir schreiben dann $\nu \ll \mu$. Ist sowohl $\nu \ll \mu$ als auch $\mu \ll \nu$, so heißen μ und ν äquivalent.

3. Die Maße μ und ν heißen singulär, falls es ein $A \in \mathcal{F}$ gibt mit $\mu(A) = 0$ und $\nu(A^c) = 0$. Wir schreiben dann $\mu \perp \nu$.

Lemma 5.13 (Kettenregel und Eindeutigkeit). Sei μ ein Maß auf \mathcal{F} .

1. Sei ν ein σ -endliches Maß. Sind g_1 und g_2 Dichten von ν bzgl. μ , so ist $g_1 = g_2$ μ -fast überall.

2. Sei $f : \Omega \rightarrow \mathbb{R}_+$ und $g : \Omega \rightarrow \mathbb{R}$ messbar. Dann gilt

$$(f \cdot \mu)[g] = \mu[fg],$$

falls eine der beiden Seiten existiert.

Beweis. 1. Sei $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ so, dass $\Omega_n \uparrow \Omega$ und $\nu(\Omega_n) < \infty$. Setze $A_n := \Omega_n \cap \{g_1 > g_2\}$. Da sowohl g_1 als auch g_2 Dichten von ν bzgl. μ sind, folgt

$$\mu[g_1 - g_2; A_n] = 0.$$

Da auf A_n nur $g_1 > g_2$ möglich ist, ist $g_1 = g_2 \cdot 1_{A_n}$ μ -fast überall. Außerdem ist

$$\mu\{g_1 > g_2\} = \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = 0.$$

Analog folgt $\mu\{g_1 < g_2\} = 0$ und damit $g_1 = g_2$ μ -fast überall.

2. gilt per Definition für $g = 1_A$ mit $A \in \mathcal{F}$. Damit erweitert man die Aussage schrittweise für einfach Funktionen, positive messbare Funktionen und schlussendlich auf den allgemeinen Fall. \square

Beispiel 5.14 (Bekannte Dichten). 1. Aus der Vorlesung *Stochastik* sind bereits einige Dichtefunktionen bekannt. Sei etwa für $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$

$$f_{N(\mu, \sigma^2)}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

und λ das eindimensionale Lebesgue-Maß. Dann heißt das Wahrscheinlichkeitsmaß $f_{N(\mu, \sigma^2)} \cdot \lambda$ auch *Normalverteilung mit Erwartungswert μ und Varianz σ^2* .

Für $\gamma \geq 0$ und

$$f_{\text{exp}(\gamma)}(x) := 1_{x \geq 0} \cdot \gamma e^{-\gamma x}$$

heißt $f_{\text{exp}(\gamma)} \cdot \lambda$ auch *Exponentialverteilung zum Parameter γ* . Man kann nun beispielsweise mit Lemma 5.13 berechnen, dass

$$f_{\text{exp}(\gamma)} \cdot \lambda[\text{id}] = \int_0^\infty \gamma e^{-\gamma x} x dx = -e^{-\gamma x} x \Big|_0^\infty + \int_0^\infty e^{-\gamma x} dx = \frac{1}{\gamma}.$$

Den Wert $1/\gamma$ haben wir bereits als Erwartungswert der Exponentialverteilung zum Parameter γ interpretiert.

2. Es gibt natürlich nicht nur Dichten bezüglich des Lebesgue-Maßes. Sei beispielsweise

$$\mu = \sum_{n=0}^{\infty} \delta_n$$

das Zählmaß auf \mathbb{N}_0 (siehe Beispiel 3.2) und $f : \mathbb{N}_0 \rightarrow \mathbb{R}_+$, gegeben für ein $\gamma \geq 0$ durch

$$f(k) = e^{-\gamma} \frac{\gamma^k}{k!}.$$

Dann ist $f \cdot \mu$ die Poisson-Verteilung zum Parameter γ auf $2^{\mathbb{N}_0}$ nach Beispiel 3.2.

Theorem 5.15 (Lebesgue'scher Zerlegungssatz). *Seien μ, ν σ -endliche Maße auf (Ω, \mathcal{F}) . Dann lässt sich ν eindeutig schreiben als*

$$\nu = \nu_a + \nu_s \quad \text{mit} \quad \nu_a \ll \mu, \nu_s \perp \mu.$$

Das Maß ν_a hat eine Dichte bzgl. μ , die μ -fast überall endlich ist.

Beweis. Durch eine Ausschöpfung $\Omega_1, \Omega_2, \dots \subseteq \Omega$ mit $\Omega_n \uparrow \Omega$ und $\nu(\Omega_n), \mu(\Omega_n) < \infty$ können wir uns auf den Fall endlicher Maße zurückziehen. Mit Proposition 5.6 ist die lineare Abbildung

$$\begin{cases} \mathcal{L}^2(\mu + \nu) & \rightarrow \mathbb{R} \\ f & \mapsto \nu[f] \end{cases}$$

stetig. Nach Proposition 5.10 gibt es also ein $h \in \mathcal{L}^2(\mu + \nu)$ mit

$$\nu[f] = (\mu + \nu)[fh], \tag{5.3}$$

also

$$\nu[f(1-h)] = \mu[fh] \tag{5.4}$$

für jedes $f \in \mathcal{L}^2(\mu + \nu)$. Wählt man $f = 1_{\{h < 0\}}$ in (5.3) folgt

$$0 \leq \nu\{h < 0\} = (\mu + \nu)[h; h < 0] \leq 0,$$

also $h \geq 0$ $(\mu + \nu)$ -fast überall. Analog kann man mittels $f = 1_{\{h > 1\}}$ aus (5.4) folgern, dass

$$0 \leq \mu[h; \{h > 1\}] = \nu[1-h; \{h > 1\}] \leq 0,$$

also $h \leq 1$ $(\mu + \nu)$ -fast sicher. Sei nun $f \geq 0$ messbar und $f_1, f_2, \dots \in \mathcal{L}^2(\mu + \nu)$ mit $f_n \uparrow f$. Mit dem Satz von der monotonen Konvergenz folgt

$$\nu[f(1-h)] = \lim_{n \rightarrow \infty} \nu[f_n(1-h)] = \lim_{n \rightarrow \infty} \mu[f_n h] = \mu[fh],$$

d.h. (5.4) gilt für alle messbaren $f \geq 0$.

Sei nun $E := h^{-1}\{1\}$. Aus (5.4) folgt mit $f = 1_E$, dass

$$\mu(E) = \mu[h; E] = \nu[1 - h; E] = 0.$$

Wir definieren für $A \in \mathcal{F}$ zwei Maße ν_a und ν_s durch

$$\nu_a(A) = \nu(A \setminus E), \quad \nu_s(A) = \nu(A \cap E),$$

so dass $\nu = \nu_a + \nu_s$ und $\nu_s \perp \mu$. Um zu zeigen, dass $\nu_a \ll \mu$ wähle $A \in \mathcal{F}$ mit $\mu(A) = 0$. Damit ist nach (5.4)

$$\nu[1 - h; A \setminus E] = \mu[h; A \setminus E] = 0.$$

Da $h < 1$ auf $A \setminus E$ ist, gilt damit $\nu_a(A) = \nu(A \setminus E) = 0$, also $\nu_a \ll \mu$.

Um die Dichte von ν_a bzgl. μ zu bestimmen, setzen wir $g := \frac{h}{1-h} 1_{\Omega \setminus E}$ und verwenden (5.4), so dass

$$\mu[g; A] = \mu\left[\frac{h}{1-h}; A \setminus E\right] = \nu(A \setminus E) = \nu_a(A).$$

Also ist $g = \frac{d\nu_a}{d\mu}$.

Um die Eindeutigkeit der Zerlegung zu zeigen, sei $\nu = \nu_a + \nu_s = \tilde{\nu}_a + \tilde{\nu}_s$ für $\nu_a, \tilde{\nu}_a \ll \mu$, $\nu_s, \tilde{\nu}_s \perp \mu$. Wähle $A, \tilde{A} \in \mathcal{A}$ mit $\nu_s(A) = \mu(A^c) = \tilde{\nu}_s(\tilde{A}) = \mu(\tilde{A}^c) = 0$. Dann gilt

$$\nu_s(A \cap \tilde{A}) = \tilde{\nu}_s(A \cap \tilde{A}) = \nu_a(A^c \cup \tilde{A}^c) = \tilde{\nu}_a(A^c \cup \tilde{A}^c) = 0$$

und deshalb

$$\begin{aligned} \nu_a &= 1_{A \cap \tilde{A}} \cdot \nu_a = 1_{A \cap \tilde{A}} \cdot \nu = 1_{A \cap \tilde{A}} \cdot \tilde{\nu}_a = \tilde{\nu}_a, \\ \nu_s &= \nu - \nu_a = \nu - \tilde{\nu}_a = \tilde{\nu}_s. \end{aligned}$$

□

Korollar 5.16 (Satz von Radon-Nikodým). *Seien μ und ν σ -endliche Maße. Dann hat ν genau dann eine Dichte bzgl. μ , wenn $\nu \ll \mu$.*

Beweis. '⇒': klar.

'⇐': Nach Theorem 5.15 gibt es eine eindeutige Zerlegung $\nu = \nu_a + \nu_s$ mit $\nu_a \ll \mu$, $\nu_s \perp \mu$. Da $\nu \ll \mu$, muss $\nu_s = 0$ gelten und damit $\nu = \nu_a$. Insbesondere existiert die Dichte von ν bzgl. μ . □

Beispiel 5.17. Im Zerlegungssatz von Lebesgue 5.15 und im Satz von Radon-Nikodým 5.16 darf man die Voraussetzung, dass μ und ν σ -endlich sind nicht weglassen, wie folgendes Beispiel zeigt:

Sei (Ω, \mathcal{F}) ein Messraum mit überabzählbarem Ω und

$$\mathcal{F} := \{A : A \text{ oder } A^c \text{ abzählbar}\}.$$

Seien μ und ν unendliche Maße auf (Ω, \mathcal{F}) , gegeben durch

$$\nu(A) := \begin{cases} 0, & A \text{ abzählbar,} \\ \infty, & \text{sonst,} \end{cases} \quad \mu(A) := \begin{cases} |A|, & A \text{ endlich,} \\ \infty, & \text{sonst.} \end{cases}$$

Dann ist offenbar $\nu \ll \mu$. Angenommen, es gäbe eine \mathcal{F} -messbare Dichte von ν bzgl. μ , so wäre für alle $\omega \in \Omega$

$$0 = \nu\{\omega\} = \mu[f; \{\omega\}] = f(\omega)\mu(\{\omega\}) = f(\omega).$$

Damit wäre $f = 0$ und $\nu = 0$ im Widerspruch zur Definition von ν .

6 Produkträume

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

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

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

$$\pi_H^J : \prod_{i \in J} \Omega_i \rightarrow \prod_{i \in H} \Omega_i.$$

sowie $\pi_H := \pi_H^I$ und $\pi_i := \pi_{\{i\}}$, $i \in I$. In diesem Kapitel werden wir alle bisher eingeführten Konzepte auf solche Produkträume anwenden. Für die Anwendung der Maßtheorie auf stochastische Prozesse besonders wichtig wird Satz über projektive Limiten von Wahrscheinlichkeitsmaßen, Theorem 6.24.

6.1 Topologie

Wir beginnen mit der Definition der Topologie auf Produkträumen. Diese ist gerade so gemacht, dass Projektionen stetige Funktionen sind.

Definition 6.1 (Produktraum und Produkt-Topologie). Ist $(\Omega_i, \mathcal{O}_i)_{i \in I}$ eine Familie von topologischen Räumen, dann heißt die von

$$\mathcal{B} := \left\{ \prod_{i \in J} A_i \times \prod_{i \in I \setminus J} \Omega_i : J \subseteq I, A_i \in \mathcal{O}_i \right\}$$

erzeugte Topologie (siehe Definition 1.1.7) \mathcal{O} die Produkttopologie auf Ω .

Bemerkung 6.2 (Stetigkeit der Koordinatenabbildungen). Es sei bemerkt, dass bezüglich der Produkttopologie \mathcal{O} alle Projektionen $\pi_i, i \in I$ stetig sind. Es ist nämlich

$$\pi_i^{-1}(A_i) = A_i \times \prod_{I \ni j \neq i} \Omega_j \in \mathcal{O}$$

für $A_i \in \mathcal{O}_i$. Damit ist die Projektion stetig (siehe Definition 1.1.10).

Proposition 6.3 (Abzählbare Produkte polnischer Räume sind polnisch). Sei $(\Omega_i, \mathcal{O}_i)_{i \in \mathbb{N}}$ eine Familie polnischer Räume. Dann ist der Produktraum (Ω, \mathcal{O}) aus Definition 6.1 polnisch.

Beweis. Sei $\Omega'_i \subseteq \Omega_i$ abzählbar mit $\overline{\Omega'_i} = \Omega_i$ und r_i eine vollständige Metrik, die \mathcal{O}_i erzeugt, $i \in \mathbb{N}$. Weiter sei $(\omega'_i)_{i \in \mathbb{N}} \in \prod_{i \in \mathbb{N}} \Omega'_i$.

Zunächst bemerken wir, dass für $\omega^1 = (\omega_i^1)_{i \in \mathbb{N}}, \omega^2 = (\omega_i^2)_{i \in \mathbb{N}} \in \Omega$ durch

$$r(\omega^1, \omega^2) := \sum_{i=1}^{\infty} 2^{-i} (r_i(\omega_i^1, \omega_i^2) \wedge 1)$$

eine vollständige Metrik auf Ω definiert wird, die \mathcal{O} erzeugt. Außerdem ist damit die Menge

$$\left\{ (\omega_i)_{i \in \mathbb{N}} \in \prod_{i \in \mathbb{N}} \Omega'_i : \omega_i \neq \omega'_i \text{ für endlich viele } i \in \mathbb{N} \right\}$$

abzählbar und dicht in Ω . Also ist (Ω, \mathcal{O}) polnisch. \square

6.2 Mengensysteme

Analog zur Topologie ist die Produkt- σ -Algebra gerade so, dass Projektionen messbare Funktionen sind.

Definition 6.4 (Produkt- σ -Algebra). Ist $(\Omega_i, \mathcal{F}_i)_{i \in I}$ eine Familie von Messräumen, dann heißt die σ -Algebra

$$\bigotimes_{i \in I} \mathcal{F}_i := \sigma \left(\left\{ \prod_{i \in J} A_i \times \prod_{i \in I \setminus J} \Omega_i : J \subseteq I, A_i \in \mathcal{F}_i \right\} \right) \quad (6.1)$$

die Produkt- σ -Algebra auf Ω . Ist $(\Omega_i, \mathcal{F}_i) = (\Omega, \mathcal{F}), i \in I$, so setzen wir $\mathcal{F}^I := \bigotimes_{i \in I} \mathcal{F}$.

Bemerkung 6.5 (Messbarkeit der Koordinatenabbildung). Analog zur Produkttopologie gilt, dass die Projektionen π_i messbar bzgl. $\bigotimes_{i \in I} \mathcal{F}_i$ sind. Es ist nämlich für $A_i \in \mathcal{F}_i$

$$\pi_i^{-1}(A_i) = A_i \times \prod_{I \ni j \neq i} \Omega_j \in \bigotimes_{i \in I} \mathcal{F}_i.$$

Außerdem gilt

$$\bigotimes_{i \in I} \mathcal{F}_i = \sigma \left(\left\{ A_i \times \prod_{j \neq i} \Omega_j : A_i \in \mathcal{F}_i, i \in I \right\} \right). \quad (6.2)$$

Lemma 6.6 (Produkt- σ -Algebra bei abzählbaren Produkten). Sei I abzählbar und $(\Omega_i, \mathcal{O}_i)_{i \in I}$ eine Familie polnischer Räume sowie (Ω, \mathcal{O}) der Produktraum, versehen mit der Produkttopologie aus Definition 6.1. Dann ist $\mathcal{B}(\Omega) = \bigotimes_{i \in I} \mathcal{B}(\Omega_i)$. Insbesondere ist $\mathcal{B}(\mathbb{R}^d) = \bigotimes_{i=1}^d \mathcal{B}(\mathbb{R})$.

Beweis. Da alle $(\Omega_i, \mathcal{O}_i)$, $i \in I$, separabel sind, gibt es nach Lemma 1.5 für jedes $i \in I$ eine abzählbare Basis \mathcal{C}_i von \mathcal{O}_i . Damit ist

$$\mathcal{C} := \left\{ \prod_{i \in J} A_i \times \prod_{i \in I \setminus J} \Omega_i : J \subseteq I, A_i \in \mathcal{C}_i \right\}$$

eine abzählbare Basis von (Ω, \mathcal{O}) . Nach Lemma 2.8 gilt $\sigma(\mathcal{C}) = \mathcal{B}(\Omega)$. Außerdem ist offenbar $\mathcal{C} \subseteq \bigotimes_{i \in I} \mathcal{B}(\Omega_i)$ und damit $\mathcal{B}(\Omega) \subseteq \bigotimes_{i \in I} \mathcal{B}(\Omega_i)$. Andersherum liefert (6.2) und, dass für $A_i \in \mathcal{F}_i$

$$A_i \times \prod_{j \neq i} \Omega_j \in \sigma \left(\left\{ A_i \times \prod_{j \neq i} \Omega_j : A_i \in \mathcal{O}_i \right\} \right) = \pi_i^{-1}(\mathcal{F}_i) \subseteq \mathcal{B}(\Omega),$$

woraus die Behauptung folgt. \square

Lemma 6.7 (Produkte von Erzeugern/Halbringen sind Erzeuger/Halbringe). Seien $(\Omega_i, \mathcal{F}_i)$ Messräume und $\Omega = \prod_{i \in I} \Omega_i$.

1. Sei I endlich und \mathcal{H}_i ein Halbring mit $\sigma(\mathcal{H}_i) = \mathcal{F}_i$. Dann ist

$$\mathcal{H} := \left\{ \prod_{i \in I} A_i : A_i \in \mathcal{H}_i, i \in I \right\} \quad (6.3)$$

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

2. Sei I beliebig und \mathcal{H}_i ein schnittstabiler Erzeuger von \mathcal{F}_i , $i \in I$. Dann ist

$$\mathcal{H} := \left\{ \prod_{i \in J} A_i \times \prod_{i \in I \setminus J} \Omega_i : J \subseteq I, A_i \in \mathcal{H}_i, i \in J \right\}$$

ein schnittstabiler Erzeuger von $\bigotimes_{i \in I} \mathcal{F}_i$.

Beweis. Für 1. ist o.E. $I = \{1, \dots, d\}$. Es ist zunächst klar, dass \mathcal{H} schnittstabil ist. Die Eigenschaft (ii) für Halbringe zeigt man durch Induktion über d . Die Behauptung ist klar für $d = 1$, da \mathcal{H}_1 ein Halbring ist. Gilt sie für $d - 1$, so gilt

$$\begin{aligned} & (A_1 \times \dots \times A_d) \setminus (B_1 \times \dots \times B_d) \\ &= (A_1 \times \dots \times A_{d-1} \times (A_d \setminus B_d)) \uplus ((A_1 \times \dots \times A_{d-1}) \setminus (B_1 \times \dots \times B_{d-1})) \times (A_d \cap B_d) \end{aligned}$$

Der erste Term der letzten Zeile ist als disjunkte Vereinigung von Mengen aus \mathcal{H} darstellbar, da \mathcal{H}_d ein Halbring ist. Der zweite Term ist als disjunkte Vereinigung darstellbar, da nach Induktionsvoraussetzung $(A_1 \times \dots \times A_{d-1}) \setminus (B_1 \times \dots \times B_{d-1})$ als disjunkte Vereinigung aus Mengen der Form $A_1 \times \dots \times A_{d-1}$ mit $A_i \in \mathcal{H}_i$, $i = 1, \dots, d - 1$ darstellbar ist.

Für 2. ist wieder klar, dass \mathcal{H} schnittstabil ist. Aus (6.1) folgt sofort, dass $\mathcal{H} \subseteq \bigotimes_{i \in I} \mathcal{F}_i$, also $\sigma(\mathcal{H}) \subseteq \bigotimes_{i \in I} \mathcal{F}_i$. Andersherum ist klar, dass für $A_i \in \mathcal{F}_i$

$$A_i \times \prod_{j \neq i} \Omega_j \in \sigma\left(\left\{A_i \times \prod_{j \neq i} \Omega_j : A_i \in \mathcal{H}_i\right\}\right) \subseteq \sigma(\mathcal{H}),$$

woraus $\bigotimes_{i \in I} \mathcal{F}_i \subseteq \sigma(\mathcal{H})$ wegen (6.2) und damit die Behauptung folgt. \square

Korollar 6.8 (Borel'sche σ -Algebra auf \mathbb{R}^d wird von Quadern erzeugt). Sei $\Omega = \mathbb{R}^d$. Für $\underline{a} = (a_1, \dots, a_d), \underline{b} = (b_1, \dots, b_d) \in \mathbb{R}^d$ setzen wir $\underline{a} \leq \underline{b}$ genau dann, wenn $a_i \leq b_i$, $i = 1, \dots, d$, sowie mit

$$(\underline{a}, \underline{b}] := (a_1, b_1] \times \dots \times (a_d, b_d]$$

den halboffenen Quader. Dann definiert

$$\mathcal{H} := \{(\underline{a}, \underline{b}] : \underline{a}, \underline{b} \in \mathbb{Q}, \underline{a} \leq \underline{b}\}$$

einen Halbring mit $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R}^d)$.

Beweis. Nach Beispiel 2.3.1 und Lemma 6.7.1 ist \mathcal{H} ein Halbring, der $\bigotimes_{i=1}^d \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$ erzeugt; siehe Lemma 6.6. \square

6.3 Maße und Integrale

Mehrfachintegrale sind schon aus der Analysis bekannt. Wir definieren nun zunächst Maße auf Produkträumen und im gleichen Atemzug auch die dazu gehörigen (Mehrfach-)Integrale. Mit dem Satz von Fubini (Theorem 6.13) können dann Integrale nach Maßen auf Produkträumen als Mehrfachintegrale interpretiert und ausgewertet werden. Hierzu ist es notwendig, dass die in den Mehrfachintegralen auftauchenden Integranden messbar sind. Dies wird in Lemma 6.11 sichergestellt. Um Maße auf Produkträumen in genügender Allgemeinheit definieren zu können, benötigen wir zunächst den Begriff des Übergangskernes.

Definition 6.9 (Übergangskern). Seien $(\Omega_i, \mathcal{F}_i), i = 1, 2$ Messräume. Eine Abbildung $\kappa : \Omega_1 \times \mathcal{F}_2 \rightarrow \mathbb{R}_+$ heißt ein Übergangskern von $(\Omega_1, \mathcal{F}_1)$ nach $(\Omega_2, \mathcal{F}_2)$, wenn (i) für alle $\omega_1 \in \Omega_1$ ist $\kappa(\omega_1, \cdot)$ ein Maß auf \mathcal{F}_2 und (ii) für alle $A_2 \in \mathcal{F}_2$ ist $\kappa(\cdot, A_2)$ nach \mathcal{F}_1 -messbar.

Ein Übergangskern heißt σ -endlich, wenn es eine Folge $\Omega_{21}, \Omega_{22}, \dots \in \mathcal{F}_2$ gibt mit $\Omega_{2n} \uparrow \Omega_2$ und $\sup_{\omega_1} \kappa(\omega_1, \Omega_{2n}) < \infty$ für alle $n = 1, 2, \dots$. Er heißt stochastischer Kern oder Markov'scher Kern, falls für alle $\omega_1 \in \Omega_1$ gilt, dass $\kappa(\omega_1, \Omega_2) = 1$.

Beispiel 6.10 (Markov-Kette). Sei $\Omega = \{\omega_1, \dots, \omega_n\}$ endlich und $P = (p_{ij})_{1 \leq i, j \leq n}$ mit $p_{ij} \in [0, 1]$ und $\sum_{j=1}^n p_{ij} = 1$. Dann definiert für $A \subset \Omega$

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

einen Markov-Kern von $(\Omega, 2^\Omega)$ nach $(\Omega, 2^\Omega)$. Hier ist P als stochastische Matrix die Übergangsmatrix einer homogenen, Ω -wertigen Markov-Kette.

Lemma 6.11 (Messbarkeit integrierbarer Schnitte). Seien $(\Omega_i, \mathcal{F}_i), i = 1, 2$ Messräume, κ ein σ -endlicher Übergangskern von $(\Omega_1, \mathcal{F}_1)$ nach $(\Omega_2, \mathcal{F}_2)$ und $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}_+$ nach $\mathcal{F}_1 \otimes \mathcal{F}_2$ messbar. Dann ist

$$\omega_1 \mapsto \kappa(\omega_1, \cdot)[f] := \int \kappa(\omega_1, d\omega_2) f(\omega_1, \omega_2)$$

nach \mathcal{F}_1 -messbar.

Beweis. Wir nehmen an, dass $\kappa(\omega_1, \Omega_2) < \infty$ für alle $\omega_1 \in \Omega_1$ gilt. (Der allgemeine Fall erfolgt dann mittels einer Folge $\Omega_{11}, \Omega_{12}, \dots \in \mathcal{F}_1$ mit $\Omega_{1n} \uparrow \Omega_1$.) Sei

$$\mathcal{D} := \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 : \omega_1 \mapsto \kappa(\omega_1, \cdot)[1_A] \text{ ist } \mathcal{F}_1\text{-messbar}\}.$$

Dann prüft man leicht nach, dass \mathcal{D} ein schnittstabiles Dynkin-System ist. Weiter ist sicherlich $\mathcal{H} \subseteq \mathcal{D}$, wobei \mathcal{H} wie in (6.3) definiert ist. Damit ist nach Theorem 2.13 $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{H}) \subseteq \mathcal{D} \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$. Also ist $\omega_1 \mapsto \kappa(\omega_1, \cdot)[1_A]$ für alle $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ nach \mathcal{F}_1 -messbar. Diese Aussage lässt sich sofort erweitern, indem man anstatt 1_A eine Treppenfunktionen einsetzt. Durch monotone Konvergenz folgt dann auch, dass $\omega_1 \mapsto \kappa(\omega_1, \cdot)[f]$ für alle messbaren, nicht-negativen Funktionen nach \mathcal{F}_1 -messbar ist. \square

Theorem 6.12 (Satz von Ionescu-Tulcea). Seien $(\Omega_i, \mathcal{F}_i), i = 0, \dots, n$ Messräume, μ ein σ -endliches Maß auf \mathcal{F}_0 und κ_i ein σ -endlicher Übergangskern von $\left(\times_{j=0}^{i-1} \Omega_j, \otimes_{j=0}^{i-1} \mathcal{F}_j\right)$ nach $(\Omega_i, \mathcal{F}_i), i = 1, \dots, n$. Dann gibt es genau ein σ -endliches Maß $\mu \otimes_{i=1}^n \kappa_i$ auf $\left(\times_{i=0}^n \Omega_i, \otimes_{i=0}^n \mathcal{F}_i\right)$ mit

$$\left(\mu \otimes_{i=1}^n \kappa_i\right)(A_0 \times \dots \times A_n) = \int_{A_0} \mu(d\omega_0) \left(\int_{A_1} \kappa_1(\omega_0, d\omega_1) \cdots \left(\int_{A_n} \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) \right) \cdots \right). \quad (6.4)$$

Beweis. Wir zeigen das Theorem nur für $n = 1$, der allgemeine Fall erfolgt dann durch Induktion.

Der Beweis ist eine Anwendung von Theorem 3.15. Zunächst stellen wir fest, dass nach Lemma 6.7 das in (6.3) definierte Mengensystem \mathcal{H} ein Halbring auf $\times_{i=1}^n \Omega_i$ ist. Wir zeigen zunächst, dass die angegebene Mengenfunktion σ -endlich auf \mathcal{H} ist. Es gibt nämlich $\Omega_{i1}, \Omega_{i2}, \in \mathcal{F}_i$ mit $\Omega_{in} \uparrow \Omega_i, i = 0, 1$ mit $\mu(\Omega_{0n}) < \infty, \kappa_1(\omega_0, \Omega_{1n}) < \infty, n = 1, 2, \dots, \omega_0 \in \Omega_0$ und $\sup_{\omega_0 \in \Omega_0} \kappa_1(\omega_0, \Omega_{1n}) =: C_n < \infty$. Damit ist $\mu \otimes \kappa_1(\Omega_{0n} \times \Omega_{1n}) \leq C_n \cdot \mu(\Omega_{0n}) < \infty$ und $\Omega_{0n} \times \Omega_{1n} \uparrow \Omega_0 \times \Omega_1$. Damit ist also $\mu \otimes \kappa_1$ auch σ -endlich. Definiert man $\tilde{\mu}$ auf \mathcal{H} mittels (6.4), so ist dies also eine σ -endliche Mengenfunktion.

Wir zeigen nun, dass $\tilde{\mu}$ σ -subadditiv und endlich additiv auf \mathcal{H} ist. Für $A_1, \dots, A_n \in \mathcal{H}$ und $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$ ist wegen der σ -Subadditivität von $\kappa_1(\omega_0, \cdot)$ für alle $\omega_0 \in \Omega_0$

$$\begin{aligned} \tilde{\mu}(A) &= \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_A(\omega_0, \omega_1) \\ &\leq \sum_{n=1}^{\infty} \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_{A_n}(\omega_0, \omega_1) = \sum_{n=1}^{\infty} \tilde{\mu}(A_n). \end{aligned}$$

Analog zeigt man die endliche Additivität. Nach Lemma 3.4 ist $\tilde{\mu}$ damit σ -additiv. Aus Theorem 3.15 folgt nun, dass es genau eine Erweiterung von $\tilde{\mu}$ auf $\sigma(\mathcal{H}) = \bigotimes_{i=1}^n \sigma(\mathcal{H}_i)$ gibt, welche die im Theorem angegebene ist. \square

Wir beschäftigen uns nun mit dem in Theorem 6.12 definierten Maß.

Theorem 6.13 (Satz von Fubini). *Seien $(\Omega_i, \mathcal{F}_i)$, μ, κ_i und $\mu \bigotimes_{i=1}^n \kappa_i$ wie in Theorem 6.12. Weiter sei $f : \times_{i=0}^n \Omega_i \rightarrow \mathbb{R}_+$ messbar bezüglich $\bigotimes_{i=0}^n \mathcal{F}_i$. Dann gilt*

$$\int f d(\mu \bigotimes_{i=0}^n \kappa_i) = \int \mu(d\omega_0) \left(\int \kappa_1(\omega_1, d\omega_2) \cdots \left(\int \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) f(\omega_0, \dots, \omega_n) \right) \cdots \right). \quad (6.5)$$

Diese Gleichheit gilt auch, falls $f : \times_{i=0}^n \Omega_i \rightarrow \mathbb{R}$ messbar ist mit $\int |f| d(\mu \bigotimes_{i=0}^n \kappa_i) < \infty$.

Beweis. Betrachte die Mengenfunktion $\tilde{\mu}$ auf $\bigotimes_{i=0}^n \mathcal{F}_i$, gegeben durch

$$\tilde{\mu} : A \mapsto \int \mu(d\omega_0) \left(\int \kappa_1(\omega_1, d\omega_2) \cdots \left(\int \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) 1_A(\omega_0, \dots, \omega_n) \right) \cdots \right).$$

Man sieht, dass $\tilde{\mu}$ auf \mathcal{H} aus (6.3) mit $\mu \bigotimes_{i=1}^n \kappa_i$ übereinstimmt. Da \mathcal{H} schnittstabil ist, folgt die Gleichheit (6.5) für Indikatorfunktionen wegen Proposition 3.10. Mittels Linearität des Integrals erweitert man (6.5) zunächst auf einfache Funktionen und dann mit Monotonie auf beliebige, nicht-negative, messbare Funktionen. Man beachte hierbei, dass alle vorkommenden Integranden nach Lemma 6.11 messbar sind. \square

Korollar 6.14 (Produktmaße). *Sei $\Omega = \times_{i=1}^n \Omega_i$ und $\mathcal{H}_i \subseteq 2^{\Omega_i}$ ein Halbring, $i = 1, \dots, n$, sowie $\mu_i : \mathcal{H}_i \rightarrow \mathbb{R}_+$ σ -endlich und, σ -additiv, $i = 1, \dots, n$. Dann gibt es genau ein Maß $\mu_1 \otimes \cdots \otimes \mu_n$ auf $\bigotimes_{i=1}^n \sigma(\mathcal{H}_i)$ mit*

$$\mu_1 \otimes \cdots \otimes \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n). \quad (6.6)$$

Für eine Funktion $f : \Omega \rightarrow \mathbb{R}_+$ und jede Permutation π auf $\{1, \dots, n\}$ gilt dann

$$\int f d\mu_1 \otimes \cdots \otimes \mu_n = \int \left(\cdots \left(\int f(\omega_1, \dots, \omega_n) \mu_{\pi(1)}(d\omega_{\pi(1)}) \right) \cdots \right) \mu_{\pi(n)}(d\omega_{\pi(n)}).$$

Diese Formel gilt auch für $f : \Omega \rightarrow \mathbb{R}$, falls $\int |f| d\mu_1 \otimes \cdots \otimes \mu_n < \infty$.

Beweis. Das Korollar folgt direkt aus Theorem 6.12 und Theorem 6.13, wenn man $\kappa_i(\omega_0, \dots, \omega_{i-1}, \cdot) = \mu_i(\cdot)$ für alle $\omega_0, \dots, \omega_{i-1}$ setzt. \square

Definition 6.15 (Endliches Produktmaß). Betrachte dieselbe Situation wie in Korollar 6.14. Dann heißt das eindeutige Maß $\mu_1 \otimes \dots \otimes \mu_n$ aus Korollar 6.14 das Produktmaß der μ_1, \dots, μ_n . Wir schreiben auch

$$\bigotimes_{i=1}^n \mu_i := \mu_1 \otimes \dots \otimes \mu_n.$$

Gilt $(\Omega_i, \mathcal{H}_i, \mu_i) = (\Omega_0, \mathcal{H}_i, \mu_0)$, $i = 1, \dots, n$, sind also alle Räume gleich, so bezeichnen wir es auch mit

$$\mu_0^{\otimes n} := \mu_1 \otimes \dots \otimes \mu_n.$$

Beispiel 6.16 (Mehrdimensionales Lebesgue-Maß). 1. Sei λ das ein-dimensionale Lebesgue-Maß auf $\mathcal{B}(\mathbb{R})$ aus Proposition 3.17. Dann heißt $\lambda^{\otimes d}$ das d -dimensionale Lebesgue-Maß.

2. Sei $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ gegeben durch

$$f(x, y) = \frac{xy}{(x^2 + y^2)^2}.$$

Dann ist für jedes $x \in \mathbb{R}$

$$\int \lambda(dy) f(x, y) = 0,$$

da $f(x, \cdot) \in \mathcal{L}^1(\lambda)$ und $f(x, y) = -f(x, -y)$. Also gilt insbesondere

$$\int \lambda(dx) \left(\int \lambda(dy) f(x, y) \right) = \int \lambda(dy) \left(\int \lambda(dx) f(x, y) \right) = 0,$$

allerdings ist $|f|$ nicht nach $\lambda^{\otimes 2}$ integrierbar, weil f in $(0, 0)$ eine nicht integrierbare Polstelle besitzt. Wie dieses Beispiel zeigt, muss man mit Mehrfachintegralen aufpassen. Insbesondere folgt aus der Gleichheit und Endlichkeit der Mehrfachintegrale nicht, dass der Integrand integrierbar ist.

6.4 Faltung von Maßen

Wir betrachten nun eine einfache Verknüpfung von Produktmaßen und Bildmaßen. Zur Faltung von Maßen μ, ν auf $\mathcal{B}(\mathbb{R})$ betrachten wir zunächst das Produktmaß $\mu \otimes \nu$. Das Bildmaß unter Summenbildung ist dann die Faltung aus μ, ν . Diese Faltung werden wir später als die Verteilung von $X + Y$ identifizieren, wenn X, Y unabhängige Zufallsvariablen mit Verteilung μ und ν sind. Manchmal, etwa bei Poisson-Verteilungen und bei Normalverteilungen, ist die Faltung wieder eine Poisson- bzw. Normalverteilung.

Definition 6.17 (Faltung von Maßen). Seien μ_1, \dots, μ_n σ -endliche Maße auf $\mathcal{B}(\mathbb{R})$ und $\mu_1 \otimes \dots \otimes \mu_n$ deren Produktmaß. Weiter sei $S(x_1, \dots, x_n) := x_1 + \dots + x_n$. Dann heißt das Bildmaß $S_*(\mu_1 \otimes \dots \otimes \mu_n)$ die Faltung der Maße μ_1, \dots, μ_n und wird mit $\mu_1 * \dots * \mu_n$ oder $*_{i=1}^n \mu_i$ bezeichnet.

Beispiel 6.18 (Faltung von Poisson- und geometrischen Verteilungen). 1. Für $\gamma_1, \gamma_2 \geq 0$ seien $\mu_{\text{Poi}(\gamma_1)}$ und $\mu_{\text{Poi}(\gamma_2)}$ zwei Poisson-Verteilungen aus Beispiel 3.2. Wir berechnen die Faltung der beiden Verteilungen durch

$$\begin{aligned} \mu_{\text{Poi}(\gamma_1)} * \mu_{\text{Poi}(\gamma_2)} &= \sum_{m,n} 1_{m+n=k} e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^n}{m!n!} \cdot \delta_k \\ &= \sum_{m=0}^k e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^{k-m}}{m!(k-m)!} \cdot \delta_k \\ &= e^{-(\gamma_1+\gamma_2)} \frac{(\gamma_1 + \gamma_2)^k}{k!} \cdot \delta_k \sum_{m=0}^k \binom{k}{m} \frac{\gamma_1^m \gamma_2^{k-m}}{(\gamma_1 + \gamma_2)^k} \\ &= \mu_{\text{Poi}(\gamma_1+\gamma_2)}. \end{aligned}$$

2. Die geometrische Verteilung zum Parameter $p \in [0, 1]$ ist bereits aus Beispiel 3.2 bekannt. Die Faltung zweier Maße $\mu_{\text{geom}(p)}$ ist gegeben durch

$$\begin{aligned} \mu_{\text{geom}(p)} * \mu_{\text{geom}(p)} &= \sum_{m=2}^k (1-p)^{m-1} p (1-p)^{k-m-1} p \cdot \delta_k \\ &= (k-1)(1-p)^{k-2} p^2 \cdot \delta_k. \end{aligned}$$

Dies ist eine negative Binomialverteilung zu den Parametern p und 2.

Lemma 6.19 (Faltung von Verteilungen mit Dichten). Sei λ ein Maß auf $\mathcal{B}(\mathbb{R})$, $\mu = f_\mu \cdot \lambda$ und $\nu = f_\nu \cdot \lambda$ für messbare Dichten $f_\mu, f_\nu : \mathbb{R} \rightarrow \mathbb{R}_+$. Dann gilt $\mu * \nu = f_{\mu*\nu} \cdot \lambda$ mit

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

Beweis. Der Beweis ist eine einfache Anwendung des Satzes von Fubini, Theorem 6.13. \square

Beispiel 6.20 (Faltung von Normalverteilungen). Seien $f_{N(\mu_1, \sigma_1^2)}$ und $f_{N(\mu_2, \sigma_2^2)}$ die Dichtefunktionen zweier Normalverteilungen mit Erwartungswert μ_1, μ_2 und Varianz σ_1^2 und σ_2^2 . Sei weiter $\mu := \mu_1 + \mu_2$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Dann ist die Dichte der Faltung gegeben durch

$$\begin{aligned} x \mapsto & \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \int \exp\left(-\frac{(y-\mu_1)^2}{2\sigma_1^2} - \frac{(x-y-\mu_2)^2}{2\sigma_2^2}\right) dy \\ & \stackrel{y \rightarrow (y-\mu_1)\sigma/(\sigma_1\sigma_2)}{=} \frac{1}{2\pi\sigma} \int \exp\left(-\frac{\sigma_2^2 y^2}{2\sigma^2} - \frac{\left((x-\mu) - y\frac{\sigma_1\sigma_2}{\sigma}\right)^2}{2\sigma_2^2}\right) dy \\ & = \frac{1}{2\pi\sigma} \int \exp\left(-\frac{\sigma_2^2 y^2 + \left((x-\mu)\frac{\sigma}{\sigma_2} - \sigma_1 y\right)^2}{2\sigma^2}\right) dy \\ & = \frac{1}{2\pi\sigma} \int \exp\left(-\frac{(\sigma y - \frac{\sigma_1}{\sigma_2}(x-\mu))^2}{2\sigma^2} - \frac{(x-\mu)^2\left(\frac{\sigma^2}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma_2^2}\right)}{2\sigma^2}\right) dy \\ & = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \end{aligned}$$

also ist die Faltung wieder eine Normalverteilung. Diese hat nun Erwartungswert μ und Varianz σ^2 .

6.5 Projektive Familien von Wahrscheinlichkeitsmaßen

Bisher haben wir σ -endliche Maße auf endlichen Produkträumen definiert. Dies ist für die zu behandelnde Wahrscheinlichkeitstheorie nicht ausreichend. Um das einzusehen, sei an den unendlichen Münzwurf erinnert, der schon in der Vorlesung *Stochastik* betrachtet wurde. Hier würde man sagen, dass $\Omega = \{\text{Kopf, Zahl}\}^{\mathbb{N}}$ und das dazu gehörige Wahrscheinlichkeitsmaß das Produktmaß $\mathbf{P}^{\otimes \infty}$ von $\mathbf{P} = \frac{1}{2}\delta_{\text{Kopf}} + \frac{1}{2}\delta_{\text{Zahl}}$ ist. Dies ist jedoch ein unendliches (aber immerhin doch abzählbares) Produktmaß, dessen Existenz wir noch nicht gezeigt haben. Oftmals ist es auch notwendig, dass wir Maße auf überabzählbaren Produkten betrachten. Ein Großteil der Vorlesung *Stochastische Prozesse* wird solche enthalten. Wir geben hier nun die allgemeine Konstruktion von Wahrscheinlichkeitsmaßen auf Produktmaßen, die auf Kolmogorov zurück geht. Es sei hier erwähnt, dass in dem resultierenden Satz von Kolmogorov (Theorem 6.24) die Voraussetzung getroffen wird, dass Ω polnisch ist.

Definition 6.21 (Projektiver Limes). 1. Sei (Ω, \mathcal{F}) ein Messraum, I eine beliebige Indexmenge und $(\Omega^I, \mathcal{F}^I)$ wie in Definition 6.4. Eine Familie von Wahrscheinlichkeitsmaßen $(\mathbf{P}_J)_{J \in I}$ heißt projektive Familie auf \mathcal{F} , falls \mathbf{P}_J für alle $J \in I$ ein Wahrscheinlichkeitsmaß auf \mathcal{F}^J ist und

$$\mathbf{P}_H = (\pi_H^J)_* \mathbf{P}_J$$

für alle $H \subseteq J \in I$.

2. Existiert für eine projektive Familie $(\mathbf{P}_J)_{J \in I}$ von Wahrscheinlichkeitsmaßen auf \mathcal{F} ein Wahrscheinlichkeitsmaß \mathbf{P}_I auf \mathcal{F}^I mit $\mathbf{P}_J = (\pi_J)_* \mathbf{P}_I$ für alle $J \in I$, so heißt \mathbf{P}_I projektiver Limes der projektiven Familie. Wir schreiben dann

$$\mathbf{P}_I = \varprojlim_{J \in I} \mathbf{P}_J.$$

Beispiel 6.22 (Projektive Limiten und stochastische Prozesse). In mindestens zwei Situationen spielen projektive Familien eine große Rolle.

1. Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und I eine unendliche Indexmenge. In Definition 6.15 haben wir für jedes $J \in I$ das Produktmaß $\mathbf{P}^{\otimes J}$ auf \mathcal{F}^J definiert. Die Familie $(\mathbf{P}^{\otimes J})_{J \in I}$ ist projektiv. Ist nämlich $H \subseteq J \in I$, so ist für $A_i \in \mathcal{F}, i \in H$,

$$\begin{aligned} (\pi_H^J)_* \mathbf{P}^{\otimes J} \left(\prod_{i \in H} A_i \right) &= \mathbf{P}^{\otimes J} \left((\pi_H^J)^{-1} \left(\prod_{i \in H} A_i \right) \right) \\ &= \mathbf{P}^{\otimes J} \left(\prod_{i \in H} A_i \times \prod_{i \in J \setminus H} \Omega \right) \\ &= \prod_{i \in H} \mathbf{P}(A_i) \cdot \prod_{i \in J \setminus H} \mathbf{P}(\Omega) \\ &= \prod_{i \in H} \mathbf{P}(A_i) \\ &= \mathbf{P}^{\otimes H} \left(\prod_{i \in H} A_i \right). \end{aligned}$$

Allerdings haben wir noch nicht gezeigt, wann es den projektiven Limes von $(\mathbf{P}^{\otimes J})_{J \in I}$ gibt. Diesen würden wir dann das unendliche Produktmaß $\mathbf{P}^{\otimes I}$ nennen.

2. Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge, $(\tilde{\Omega}, \tilde{\mathcal{F}})$ ein Messraum und $X_i : \Omega \rightarrow \tilde{\Omega}, i \in I$ Zufallsvariable. Wir werden die Familie $\mathcal{X} := (X_i)_{i \in I}$ einen stochastischen Prozess nennen. Also ist $\mathcal{X} : \Omega \rightarrow \tilde{\Omega}^I$ mit $\mathcal{X}(\omega) = (X_i(\omega))_{i \in I}$. Man kann sich nun fragen, ob die Verteilung von \mathcal{X} (d.h. das Bildmaß $\mathcal{X}_* \mathbf{P}$) als Verteilung auf $\tilde{\mathcal{F}}^I$ existiert.

Hierzu sei bemerkt, dass $\tilde{\mathbf{P}}_J := ((X_j)_{j \in J})_* \mathbf{P}, J \in I$ eine projektive Familie ist. Ist nämlich $H \subset J \in I$ und $\tilde{A}_i \in \tilde{\mathcal{F}}, i \in H$, dann ist

$$\begin{aligned} (\pi_H^J)_* \tilde{\mathbf{P}}_J \left(\prod_{j \in H} \tilde{A}_j \right) &= \tilde{\mathbf{P}}_J \left((\pi_H^J)^{-1} \prod_{j \in H} \tilde{A}_j \right) \\ &= \tilde{\mathbf{P}}_J \left(\prod_{j \in H} \tilde{A}_j \times \prod_{j \in J \setminus H} \tilde{\Omega} \right) \\ &= \mathbf{P} \left(X_j \in \tilde{A}_j, j \in H \text{ und } X_j \in \tilde{\Omega}, j \in J \setminus H \right) \\ &= \mathbf{P} \left(X_j \in \tilde{A}_j, j \in H \right) \\ &= \tilde{\mathbf{P}}_H \left(\prod_{j \in H} \tilde{A}_j \right). \end{aligned}$$

Wie Theorem 6.24 zeigt, gibt es die Verteilung $\mathcal{X}_* \mathbf{P}$ (was dann der projektive Limes der $(\tilde{\mathbf{P}}_J)_{J \in I}$ ist) zumindest dann, wenn $\tilde{\mathcal{F}}$ die Borel'sche σ -Algebra eines polnischen Raumes ist.

Bemerkung 6.23 (Eindeutigkeit des projektiven Limes). Zu jeder projektiven Familie $(\mathbf{P}_J)_{J \in I}$ gibt es höchstens einen projektiven Limes. Denn: seien \mathbf{P}_I und $\tilde{\mathbf{P}}_I$ zwei projektive Limiten, so ist für $A := \times_{i \in J} A_i \times \times_{i \in I \setminus J} \Omega_i \in \mathcal{H}$ mit \mathcal{H} aus Lemma 6.7 und $J \in I$

$$\mathbf{P}_I(A) = \mathbf{P}_J \left(\prod_{i \in J} A_i \right) = \tilde{\mathbf{P}}_J \left(\prod_{i \in J} A_i \right) = \tilde{\mathbf{P}}_I(A).$$

Damit stimmen \mathbf{P}_I und $\tilde{\mathbf{P}}_I$ auf dem schnittstabilen Erzeuger überein und nach Proposition 3.10 gilt $\mathbf{P}_I = \tilde{\mathbf{P}}_I$. Inhalt des nächsten Theorems ist, dass es bei polnischen Räumen genau einen projektiven Limes gibt.

Theorem 6.24 (Existenz von Prozessen, Kolmogorov). Sei (Ω, \mathcal{O}) polnisch, $\mathcal{F} = \mathcal{B}(\mathcal{O})$ und $(\mathbf{P}_J)_{J \in I}$ eine projektive Familie von Wahrscheinlichkeitsmaßen auf \mathcal{F} . Dann gibt es den projektiven Limes $\varprojlim_{J \in I} \mathbf{P}_J$.

Beweis. Sei \mathcal{H} wie in Lemma 6.7 und μ eine endlich additive Mengenfunktion auf \mathcal{H} , definiert durch die projektive Familie mittels

$$\mu \left(\prod_{j \in J} A_j \times \prod_{i \in I \setminus J} \Omega \right) := \mathbf{P}_J \left(\prod_{j \in J} A_j \right).$$

Nach Lemma 6.7 ist \mathcal{H} ein Halbring und μ ein wohldefinierter Inhalt auf \mathcal{H} ist. Weiter ist

$$\mathcal{K} := \left\{ \prod_{j \in J} K_j \times \prod_{i \in I \setminus J} \Omega : J \in I, K_j \text{ kompakt} \right\} \subseteq \mathcal{H}$$

ein kompaktes System.

Wir zeigen nun, dass μ von innen \mathcal{K} -regulär ist. Sei $\varepsilon > 0$, $\times_{i \in J} A_i \times \times_{i \in I \setminus J} \Omega \in \mathcal{H}$ für $J \Subset I$ und $A_i \in \mathcal{F}$, $i \in J$. Da \mathbf{P}_j für $j \in I$ ein Maß ist, gibt es nach Lemma 3.8 kompakte Mengen $K_j \in \mathcal{F}$ mit $K_j \subseteq A_j$ und $\mathbf{P}_j(A_j \setminus K_j) \leq \varepsilon$. Damit ist

$$\begin{aligned}
\mu\left(\left(\times_{i \in J} A_i \times \times_{i \in I \setminus J} \Omega\right) \setminus \left(\times_{i \in J} K_i \times \times_{i \in I \setminus J} \Omega\right)\right) &= \mu\left(\left(\times_{i \in J} A_i\right) \setminus \left(\times_{i \in J} K_i\right) \times \times_{i \in I \setminus J} \Omega\right) \\
&= \mathbf{P}_J\left(\left(\times_{j \in J} A_j\right) \setminus \left(\times_{j \in J} K_j\right)\right) \\
&\leq \mathbf{P}_J\left(\bigcup_{j \in J} (A_j \setminus K_j) \times \times_{i \neq j} \Omega\right) \\
&\leq \sum_{j \in J} \mathbf{P}_J\left((A_j \setminus K_j) \times \times_{i \neq j} \Omega\right) \\
&= \sum_{j \in J} \mathbf{P}_j(A_j \setminus K_j) \\
&\leq |J|\varepsilon.
\end{aligned}$$

Da J endlich und $\varepsilon > 0$ beliebig war, ist also μ von innen \mathcal{K} -regulär. Nach Theorem 3.9 ist μ σ -additiv. Weiter ist $\mu(\Omega^I) = 1$, also lässt μ sich nach Theorem 3.15 in eindeutiger Weise auf ein Maß \mathbf{P} auf $\sigma(\mathcal{H}) = \mathcal{F}^I$ fortsetzen. Dies muss der projektive Limes von $(\mathbf{P}_J)_{J \Subset I}$ sein. \square

Teil II

Grundlagen der Wahrscheinlichkeitstheorie

Wir kommen nun zum eigentlichen Teil der Vorlesung, der Stochastik. Hierzu sei im Folgenden immer $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Das Integral bezüglich \mathbf{P} bezeichnen wir mit $\mathbf{E}[\cdot] := \mathbf{P}[\cdot]$. Weiter kürzen wir $\mathcal{L}^p := \mathcal{L}^p(\mathbf{P})$ ab, falls das nicht zu Verwechslungen führt.

Ziel dieses Abschnittes ist es, die wichtigsten probabilistischen Aussagen bereit zu stellen. Grundlegend hierfür ist sicherlich der Begriff der Zufallsvariable, den wir in Kapitel 7 beleuchten wollen. Oft werden wir den Fall von E -wertigen Zufallsvariablen betrachten, wobei E ein polnischer Raum ist. Die wichtigsten Sätze der Stochastik sind das starke Gesetz der großen Zahlen (Theorem 9.21) und der zentrale Grenzwertsatz (Theorem 11.8). Diese beiden Sätze sind Grenzwertaussagen für Zufallsvariable, wobei wichtig ist, dass die Art der Konvergenz in beiden Sätzen grundverschieden ist. Während das starke Gesetz der großen Zahlen eine fast sichere Konvergenz beschreibt, ist der zentrale Grenzwertsatz eine Aussage über Konvergenz in Verteilung (d.h. über die schwache Konvergenz der Verteilungen der Zufallsvariablen). Konsequenterweise wird es unter anderem darum gehen, Zusammenhänge zwischen verschiedenen Konvergenzarten einzusehen, siehe Kapitel 8 und 10.

7 Zufallsvariable

Meist werden wir reellwertige Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ (d.h. Borel-messbare Funktionen) betrachten. Wir wiederholen nun alles, was wir bereits über Zufallsvariablen wissen und im Folgenden direkt benötigen werden.

7.1 Wiederholung

Viele Begriffe fielen bereits im Abschnitt *Maßtheorie*. Die Wichtigsten wiederholen wir kurz. Außerdem steht nun der Zusammenhang zwischen der Maßtheorie und der Vorlesung *Stochastik* im Vordergrund.

Bemerkung 7.1 (Zufallsvariable und deren Verteilung). Sei (Ω', \mathcal{F}') ein Maßraum.

1. Jede \mathcal{F}/\mathcal{F}' -messbare Funktion X heißt $(\Omega'$ -wertige) *Zufallsvariable*. Ist $(\Omega', \mathcal{F}') = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, so heißt sie *reellwertig*. Die σ -Algebra $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{F}'\}$ ist die von X erzeugte σ -Algebra.
2. Das Wahrscheinlichkeitsmaß $X_*\mathbf{P}$ auf \mathcal{F}' heißt *Verteilung von X* . Ist weiter Y eine Zufallsvariable und $X_*\mathbf{P} = Y_*\mathbf{P}$ (d.h. $\mathbf{P}(X \in A') = \mathbf{P}(Y \in A')$ für alle $A' \in \mathcal{F}'$), so heißen X und Y *identisch verteilt* und wir schreiben $X \stackrel{d}{=} Y$. Diese Schreibweise ist jedoch mit Vorsicht zu genießen, da man die Gleichheit $X \stackrel{d}{=} Y$ nicht durch Äquivalenzumformungen zu anderen Aussagen erweitern kann. (Etwa gilt $X - Y \stackrel{d}{=} 0$ im Allgemeinen nicht, wenn X und Y identisch verteilt sind.)
3. Für eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen heißt $((X_i)_{i \in I})_*\mathbf{P}$ die *gemeinsame Verteilung von $(X_i)_{i \in I}$* . (Dies ist das Bildmaß unter der Abbildung $(X_i)_{i \in I} : \omega \mapsto (X_i(\omega))_{i \in I}$.)

4. Wir werden folgende Redewendung verwenden: *Sei X eine nach $N(\mu, \sigma^2)$ -verteilte Zufallsvariable...* Damit ist gemeint, dass $X : \Omega \mapsto \mathbb{R}$ eine messbare Abbildung ist und $X_*\mathbf{P} = \mu_{N(\mu, \sigma^2)}$, siehe Beispiel 3.21. In dieser Situation schreiben wir auch $X \sim N(\mu, \sigma^2)$. Hier bedeutet ' \sim ' *ist so verteilt wie*.
5. Sei μ ein weiteres Maß auf \mathcal{F} und $f : \Omega \rightarrow \mathbb{R}$ mit $f \geq 0$ fast überall und $\mu[f] = 1$. Dann hat X genau dann die Dichte f bezüglich μ , wenn $X_*\mathbf{P} = f \cdot \mu$ (siehe Definition 5.12). Dann gilt also für $A \in \mathcal{F}$

$$\mathbf{P}(X \in A) = \mu[f, A].$$

In diesem Fall gilt für $g : \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$, dass (siehe Lemma 5.13)

$$\mathbf{E}[g(X)] = (X_*\mathbf{P})[g] = (f \cdot \mu)[g] = \mu[fg],$$

falls die rechte Seite existiert.

6. Die Monotonie und Linearität des Integrals bedeutet nun für Zufallsvariable $X, Y \in \mathcal{L}^1$ und $a, b \in \mathbb{R}$:

$$\begin{aligned} X \leq Y \text{ fast sicher} &\implies \mathbf{E}[X] \leq \mathbf{E}[Y], \\ \mathbf{E}[aX + bY] &= a\mathbf{E}[X] + b\mathbf{E}[Y]. \end{aligned}$$

Außerdem gilt nach Proposition 4.20

$$\mathbf{E}[X] < \infty \implies \mathbf{P}(X < \infty) = 1.$$

Obwohl die σ -Algebra \mathcal{F} gegeben ist, wird im weiteren Verlauf der Vorlesung, insbesondere bei der Einführung der bedingten Erwartung in Kapitel 12, die von X erzeugte σ -Algebra eine besondere Rolle spielen. Einfach gesagt ist eine reellwertige Zufallsvariable Y genau dann $\sigma(X)$ -messbar, wenn $Y = \varphi(X)$ für eine Borel-messbare Abbildung φ . Anders ausgedrückt heißt das, dass man den Wert von $Y(\omega)$ kennt, falls man $X(\omega)$ kennt, obwohl man nicht weiß, welchen Wert ω angenommen hat.

Lemma 7.2 (Messbarkeit bezüglich $\sigma(X)$). *Sei (Ω', \mathcal{F}') ein Messraum und X ein Zufallsvariable mit Werten in Ω' und $Z : \Omega \rightarrow \overline{\mathbb{R}}$. Genau dann ist Z $\sigma(X)$ -messbar, wenn es eine $\mathcal{F}'/\mathcal{B}(\overline{\mathbb{R}})$ -messbare Abbildung $\varphi : \Omega' \rightarrow \overline{\mathbb{R}}$ gibt mit $\varphi \circ X = Z$.*

Beweis. ' \Leftarrow ': klar

' \Rightarrow ': Es genügt, den Fall $Z \geq 0$ zu betrachten; ansonsten teilt man $Z = Z^+ - Z^-$ auf. Sei zunächst $Z = 1_A$ für $A \in \sigma(X)$. Dann gibt es ein $A' \in \mathcal{F}'$ mit $X^{-1}(A') = A$, d.h. $Z = 1_{X^{-1}(A')} = 1_{A'} \circ X$, d.h. $\varphi = 1_{A'}$ erfüllt die Aussage. Durch Linearität ist die Aussage auch für einfache Funktionen, d.h. endliche Linearkombinationen von Indikatorfunktionen erfüllt. Im allgemeinen Fall gibt es einfache Funktionen $Z_1, Z_2, \dots \geq 0$ mit $Z_n \uparrow Z$. Hierzu gibt es \mathcal{F}' -messbare Funktionen φ_n mit $Z_n = \varphi_n \circ X$. Dann ist $\varphi = \sup_n \varphi_n$ wieder \mathcal{F}' -messbar und, da $Z \geq 0$ ist,

$$\varphi \circ X = (\sup_n \varphi_n) \circ X = \sup_n (\varphi_n \circ X) = \sup_n Z_n = Z. \quad \square$$

Wir wiederholen nun kurz die Konvergenzsätze für Integrale im Kontext von Zufallsvariablen.

Proposition 7.3 (Integral-Konvergenzsätze). *Seien X, X_1, X_2, \dots reellwertige Zufallsvariablen.*

1. Lemma von Fatou, Theorem 4.25: *Es gilt*

$$\liminf_{n \rightarrow \infty} \mathbf{E}[X_n] \geq \mathbf{E}[\liminf_{n \rightarrow \infty} X_n].$$

2. Satz von der monotonen Konvergenz, Theorem 4.24: *Ist $X_1, X_2, \dots \in \mathcal{L}^1$ und gilt $X_n \uparrow X$ fast sicher, dann ist*

$$\lim_{n \rightarrow \infty} \mathbf{E}[X_n] = \mathbf{E}[X],$$

wobei beide Seiten den Wert ∞ annehmen können.

3. Satz von der majorisierten Konvergenz, Theorem 4.26: *Sei $X_n \xrightarrow{n \rightarrow \infty} X$ fast sicher und Y eine weitere reellwertige Zufallsvariable mit $|X_1|, |X_2|, \dots \leq Y$ fast sicher und $\mathbf{E}[Y] < \infty$. Dann gilt*

$$\mathbf{E}[X_n] \xrightarrow{n \rightarrow \infty} \mathbf{E}[X].$$

Wir sammeln nun bereits bekannte Ungleichungen. Sie helfen oft, Wahrscheinlichkeiten oder Erwartungswerte abzuschätzen. Die meisten Ungleichungen sind schon aus der Vorlesung *Stochastik* bekannt.

Proposition 7.4 (Markov- und Chebyshev-Ungleichung). *1. Sei X eine Zufallsvariable mit Werten in $\overline{\mathbb{R}}_+$ und $x \in \mathbb{R}_+$. Dann gilt die Markov-Ungleichung*

$$\mathbf{P}(X \geq x) \leq \frac{\mathbf{E}[X]}{x}.$$

- 2. Ist X eine reellwertige Zufallsvariable und $p, x \in \mathbb{R}_+$, dann gilt die Chebyshev-Ungleichung*

$$\mathbf{P}(|X| \geq x) \leq \frac{\mathbf{E}[|X|^p]}{x^p}.$$

Beweis. 1. Da X nicht-negativ ist, gilt $x \cdot 1_{X \geq x} \leq X$. Also ist auch

$$x \cdot \mathbf{P}(X \geq x) = \mathbf{E}[x \cdot 1_{X \geq x}] \leq \mathbf{E}[X],$$

woraus die Ungleichung folgt. Die Ungleichung in 2. folgt aus 1. durch

$$\mathbf{P}(|X| \geq x) = \mathbf{P}(|X|^p \geq x^p) \leq \frac{\mathbf{E}[|X|^p]}{x^p}. \quad \square$$

Proposition 7.5 (Minkowski und Hölder-Ungleichung). *Seien X, Y reellwertige Zufallsvariablen.*

1. *Ist $0 < p, q, r \leq \infty$ so, dass $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Dann gilt*

$$\mathbf{E}[|XY|^r]^{1/r} \leq \mathbf{E}[|X|^p]^{1/p} \cdot \mathbf{E}[|Y|^q]^{1/q} \quad (\text{Hölder-Ungleichung})$$

Speziell für $p = q = 2$ ergibt sich

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^2]^{1/2} \cdot \mathbf{E}[|Y|^2]^{1/2}. \quad (\text{Cauchy-Schwartz-Ungleichung})$$

2. Für $1 \leq p \leq \infty$ ist

$$\begin{aligned} \mathbf{E}[|X + Y|^p]^{1/p} &\leq \mathbf{E}[|X|^p]^{1/p} + \mathbf{E}[|Y|^p]^{1/p}, & 1 \leq p \leq \infty & \quad (\text{Minkowski-Ungleichung}) \\ \mathbf{E}[|X + Y|^p] &\leq \mathbf{E}[|X|^p] + \mathbf{E}[|Y|^p], & 0 < p < 1 \end{aligned} \quad (7.1)$$

Beweis. Siehe Proposition 5.2 für die Hölder-Ungleichung und die Minkowski-Ungleichung für $1 \leq p \leq \infty$. Für $0 < p < 1$ ist $x \mapsto x^p$ konkav, also $(x + y)^p = \left(\frac{2x+2y}{2}\right)^p \leq \frac{1}{2}(2x)^p + \frac{1}{2}(2y)^p \leq x^p + y^p$ für reelle x, y , woraus die Minkowski-Ungleichung auch im Fall $0 < p < 1$ folgt. \square

Proposition 7.6 (Jensen'sche Ungleichung). *Sei I ein offenes Intervall und $X \in \mathcal{L}^1$ mit Werten in I und $\varphi : I \rightarrow \mathbb{R}$ konvex.¹⁴ Dann gilt*

$$\mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X]).$$

Beweis. Da φ konvex ist, ist φ stetig und

$$t \mapsto \frac{\varphi(tx + (1-t)y) - \varphi(y)}{t(x-y)}$$

für $y \leq x$ monoton fallend. Insbesondere existiert für $y \in I$

$$\lambda(y) := \lim_{x \downarrow y} \frac{\varphi(x) - \varphi(y)}{x - y} = \lim_{t \downarrow 0} \frac{\varphi(tx + (1-t)y) - \varphi(y)}{t(x-y)} \quad (7.2)$$

und es gilt

$$\frac{\varphi(x) - \varphi(y)}{x - y} \geq \lambda(y) \quad \implies \quad \varphi(y) + \lambda(y)(x - y) \leq \varphi(x) \quad (7.3)$$

für alle $x \in I$. (Für $y > x$ argumentiert man analog wie oben.)

Nun also zum Beweis der Jensen'schen Ungleichung. Da I ein Intervall ist, ist $\mathbf{E}[X] \in I$. Nach (7.3) ist für $x \in I$ mit $y = \mathbf{E}[X]$

$$\varphi(x) \geq \varphi(\mathbf{E}[X]) + \lambda(\mathbf{E}[X])(x - \mathbf{E}[X])$$

und damit

$$\mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X]) + \lambda(\mathbf{E}[X])\mathbf{E}[X - \mathbf{E}[X]] = \varphi(\mathbf{E}[X]). \quad \square$$

Mit der Jensen'schen Ungleichung kann man etwa zeigen, dass $\mathcal{L}^q \subseteq \mathcal{L}^p$ für $p \leq q$. Alternativ liest man diese Eigenschaft aus Proposition 5.3 ab.

Lemma 7.7 (p -fach und q -fach integrierbare Zufallsvariable). *Sei $q > 0$ und $X \in \mathcal{L}^q$ eine reelwertige Zufallsvariable. Dann ist für $p \leq q$*

$$\mathbf{E}[|X|^p] \leq \mathbf{E}[|X|^q]^{p/q}.$$

Insbesondere ist $\mathcal{L}^q \subseteq \mathcal{L}^p$.

Beweis. Die Abbildung $y \mapsto y^{p/q}$ ist konkav auf \mathbb{R}_+ , also gilt mit der Jensen'schen Ungleichung

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

¹⁴Eine Abbildung $\varphi : I \rightarrow \mathbb{R}$ heißt konvex, falls $\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$ für alle $0 \leq t \leq 1$ und $x, y \in I$.

7.2 Momente

Aus der Vorlesung *Stochastik* sind bereits Begriffe wie Erwartungswert, Varianz und Kovarianz bekannt. Diese wiederholen wir nun. Es gelten alle schon bekannten Rechenregeln. Der einzige Unterschied ist, dass nun $\mathbf{E}[\cdot]$ das Integral bezüglich eines Wahrscheinlichkeitsmaßes ist.

Definition 7.8 (Momente). Seien X, Y reellwertige Zufallsvariable. Dann heißt, falls existent, $\mathbf{E}[X]$ der Erwartungswert der Zufallsvariable X . Außerdem ist, falls existent,

$$\mathbf{V}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2]$$

die Varianz von X und

$$\mathbf{COV}[X, Y] := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

die Kovarianz von X und Y . Ist $\mathbf{COV}[X, Y] = 0$, so heißen X und Y unkorreliert. Weiter heißt $\mathbf{E}[X^p]$ für $p > 0$ das p -te Moment von X und $\mathbf{E}[(X - \mathbf{E}[X])^p]$ das zentrierte p -te Moment von X .

Wir wiederholen hier nur ein paar Eigenschaften.

Proposition 7.9 (Eigenschaften der zweiten Momente). Seien $X, Y \in \mathcal{L}^2$ reellwertige Zufallsvariable. Dann ist $\mathbf{V}[X], \mathbf{V}[Y], \mathbf{COV}[X, Y] < \infty$ und es gilt

$$\begin{aligned}\mathbf{V}[X] &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2, \\ \mathbf{COV}[X, Y] &= \mathbf{E}[XY] - \mathbf{E}[X] \cdot \mathbf{E}[Y].\end{aligned}$$

Außerdem gilt die Cauchy-Schwartz-Ungleichung

$$\mathbf{COV}[X, Y]^2 \leq \mathbf{V}[X] \cdot \mathbf{V}[Y].$$

Sind $X_1, \dots, X_n \in \mathcal{L}^2$, so gilt die Gleichung von Bienamyé

$$\mathbf{V}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \mathbf{V}[X_k] + 2 \sum_{1 \leq k < l \leq n} \mathbf{COV}[X_k, X_l].$$

Beweis. Da $\mathbf{V}[X] = \mathbf{COV}[X, X]$ genügt es für die erste Aussage, die zweite Gleichung zu zeigen. Diese folgt aus der Linearität des Erwartungswertes mittels

$$\begin{aligned}\mathbf{COV}[X, Y] &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[XY] - \mathbf{E}[\mathbf{E}[X]Y] - \mathbf{E}[X\mathbf{E}[Y]] + \mathbf{E}[X]\mathbf{E}[Y] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].\end{aligned}$$

Die Cauchy-Schwartz-Ungleichung folgt durch Anwenden von Proposition 7.5 auf die Zufallsvariablen $X - \mathbf{E}[X]$ und $Y - \mathbf{E}[Y]$. Insbesondere ist $\mathbf{COV}[X, Y] < \infty$. Für die Gleichung von Bienamyé sei o.E. $\mathbf{E}[X_k] = 0$, $k = 1, \dots, n$ (ansonsten geht man zu den Zufallsvariablen $X_k - \mathbf{E}[X_k]$ über). Dann ist

$$\begin{aligned}\mathbf{V}\left[\sum_{k=1}^n X_k\right] &= \mathbf{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] = \sum_{k=1}^n \sum_{l=1}^n \mathbf{E}[X_k X_l] = \sum_{k=1}^n \mathbf{E}[X_k^2] + 2 \sum_{1 \leq k < l \leq n} \mathbf{E}[X_k X_l] \\ &= \sum_{k=1}^n \mathbf{V}[X_k] + 2 \sum_{1 \leq k < l \leq n} \mathbf{COV}[X_k, X_l].\end{aligned}$$

□

Proposition 7.10 (Momente nicht-negativer Zufallsvariable). *Sei X eine Zufallsvariable mit Werten in \mathbb{R}_+ . Dann gilt*

$$\mathbf{E}[X^p] = p \int_0^\infty \mathbf{P}(X > t)t^{p-1}dt = p \int_0^\infty \mathbf{P}(X \geq t)t^{p-1}dt.$$

Beweis. Wir verwenden den Satz von Fubini,

$$\mathbf{E}[X^p] = p\mathbf{E}\left[\int_0^X t^{p-1}dt\right] = p \int_0^\infty \mathbf{E}\left[1_{X>t}t^{p-1}\right]dt = p \int_0^\infty \mathbf{P}(X > t)t^{p-1}dt.$$

Der Beweis der zweiten Gleichung ist analog. \square

7.3 Charakteristische Funktionen und Laplace-Transformierte

Wir führen nun Erwartungswerte bestimmter Funktionen von Zufallsvariablen ein. Die daraus resultierenden Funktionen sind die charakteristische Funktion (der Verteilung reellwertiger Zufallsvariablen) und die Laplace-Transformierte (der Verteilung nicht-negativer Zufallsvariablen). Die Nützlichkeit dieser beiden Funktionen ist darauf zurückzuführen, dass man mit ihrer Hilfe leicht die Momente der Verteilungen der Zufallsvariablen berechnen kann (siehe Proposition 7.14). Außerdem werden wir später in Proposition 10.25 zeigen, dass diese Funktionen verteilungsbestimmend sind.

Definition 7.11 (Charakteristische Funktion und Laplace-Transformierte).

1. Die charakteristische Funktion einer Zufallsvariablen X mit Werten in \mathbb{R}^d ist gegeben durch

$$\psi_X := \psi_{X*\mathbf{P}} := \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{C}, \\ t & \mapsto \mathbf{E}[e^{itX}] := \mathbf{E}[\cos(tX)] + i\mathbf{E}[\sin(tX)], \end{cases}$$

wobei $tx := \langle t, x \rangle$ das Skalarprodukt in \mathbb{R}^d ist.

2. Die Laplace-Transformierte von X ist gegeben durch

$$\mathcal{L}_X := \mathcal{L}_{X*\mathbf{P}} := \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{R}, \\ t & \mapsto \mathbf{E}[e^{-tX}], \end{cases}$$

gegeben das Integral auf der rechten Seite existiert. Diese wird meistens für Wahrscheinlichkeitsmaße auf \mathbb{R}_+^d betrachtet.

In diesem einführenden Teil über charakteristische Funktionen und Laplace-Transformierte werden wir nur einige wichtige Eigenschaften herleiten.

Proposition 7.12 (Eigenschaften von charakteristischen Funktionen). *Seien X, Y Zufallsvariable mit Werten in \mathbb{R}^d und charakteristischen Funktionen ψ_X, ψ_Y . Dann gilt*

1. $|\psi_X(t)| \leq 1$ für jedes $t \in \mathbb{R}^d$ und $\psi_X(0) = 1$.
2. ψ_X ist gleichmäßig stetig.
3. $\psi_{aX+b}(t) = \psi_X(at)e^{ibt}$ für alle $a \in \mathbb{R}, b \in \mathbb{R}^d$.

Beweis. 1. ist klar. Für die gleichmäßige Stetigkeit sei erwähnt, dass

$$\begin{aligned} |e^{ihx} - 1| &= \sqrt{|\cos(hx) + i\sin(hx) - 1|^2} = \sqrt{(\cos(hx) - 1)^2 + \sin^2(hx)} \\ &= \sqrt{2(1 - \cos(hx))} = 2|\sin(hx/2)| \leq |hx| \wedge 2. \end{aligned}$$

Damit gilt 2. wegen

$$\begin{aligned} \sup_{t \in \mathbb{R}^d} |\psi_X(t+h) - \psi_X(t)| &= \sup_{t \in \mathbb{R}^d} |\mathbf{E}[e^{i(t+h)X} - e^{itX}]| = \sup_{t \in \mathbb{R}^d} |\mathbf{E}[e^{itX}(e^{ihX} - 1)]| \\ &\leq \mathbf{E}[|e^{ihX} - 1|] \leq \mathbf{E}[|hX| \wedge 2] \xrightarrow{h \rightarrow 0} 0. \end{aligned}$$

Für 3. berechnen wir

$$\mathbf{E}[e^{it(aX+b)}] = e^{itb} \mathbf{E}[e^{i(at)X}] = e^{itb} \psi_X(at). \quad \square$$

Beispiel 7.13 (Beispiele für charakteristische Funktionen). 1. Die charakteristische Funktion einer nach $B(n, p)$ verteilten Zufallsvariable X ist gegeben durch

$$\psi_{B(n,p)}(t) = (1 - p + pe^{it})^n.$$

Nach Definition gilt nämlich

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

2. Die charakteristische Funktion einer nach $\text{Poi}(\gamma)$ -verteilten Zufallsvariable ist gegeben durch

$$\psi_{\text{Poi}(\gamma)} = e^{\gamma(e^{it}-1)},$$

denn

$$\psi_{\text{Poi}(\gamma)} = e^{-\gamma} \sum_{n=1}^{\infty} \frac{\gamma^n e^{itn}}{n!} = e^{\gamma(e^{it}-1)}.$$

3. Die charakteristische Funktion einer nach $N(\mu, \sigma^2)$ -verteilten Zufallsvariable X ist gegeben durch

$$\psi_{N(\mu, \sigma^2)}(t) = e^{it\mu} e^{-\sigma^2 t^2/2}.$$

Nach Proposition 7.12.2 genügt es, diese Behauptung für $\mu = 0, \sigma^2 = 1$ nachzurechnen. Für diesen Fall gilt mittels partieller Integration

$$\frac{d}{dt} \psi_{N(0,1)}(t) = \frac{i}{\sqrt{2\pi}} \int x e^{-x^2/2} e^{itx} dx = \frac{i}{\sqrt{2\pi}} \int e^{-x^2/2} i t e^{itx} dx = -t \psi_{N(0,1)}(t).$$

Diese Differentialgleichung mit $\psi_{N(0,1)}(0) = 1$ hat die eindeutige Lösung $\psi_{N(0,1)}(t) = e^{-t^2/2}$.

4. Die Laplace-Transformierte einer nach $\exp(\gamma)$ -verteilten Zufallsvariablen X ist gegeben durch

$$\mathcal{L}_{\exp(\gamma)}(t) = \frac{\gamma}{\gamma + t}.$$

Es ist nämlich

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

Oftmals sind charakteristische Funktionen und Laplace-Transformierte ein einfaches Hilfsmittel, um Momente von Zufallsvariablen zu berechnen.

Proposition 7.14 (Charakteristische Funktion und Momente). *Sei X eine Zufallsvariable mit Werten in \mathbb{R} .*

1. Ist $X \in \mathcal{L}^p$, so ist ψ_X p -mal stetig differenzierbar und für $k = 0, \dots, p$ gilt

$$\psi_X^{(k)}(t) = \mathbf{E}[(iX)^k e^{itX}].$$

Insbesondere ist $\psi_X^{(k)}(0) = i^k \mathbf{E}[X^k]$.

2. Ist speziell $X \in \mathcal{L}^2$, so ist

$$\psi_X(t) = 1 + it\mathbf{E}[X] - \frac{t^2}{2}\mathbf{E}[X^2] + \varepsilon(t)t^2$$

mit $\varepsilon(t) \xrightarrow{t \rightarrow 0} 0$.

Beweis. 1. Mit $|X|^p$ ist auch $|X|^p \vee 1$ integrierbar. Damit haben alle $|X|^k$ eine integrierbare Majorante und die rechte Seite existiert. Da die Aussage offensichtlich für $k = 0$ gilt, nehmen wir an, dass sie für ein $k < n$ gilt. Dann ist

$$\left| \frac{d^{k+1}}{dt^{k+1}} e^{itX} \right| = \lim_{h \rightarrow 0} \left| \frac{(iX)^k e^{i(t+h)X} - (iX)^k e^{itX}}{h} \right| \leq |X^k \frac{e^{ihX} - 1}{h}| \leq |X^{k+1}|.$$

Wegen majorisierter Konvergenz darf man Ableitung und Integral vertauschen und es folgt

$$\psi_X^{(k+1)}(t) = \mathbf{E} \left[\frac{d}{dt} (iX)^k e^{itX} \right] = \mathbf{E}[(iX)^{k+1} e^{itX}].$$

Die Stetigkeit der Ableitung folgt ebenso mit majorisierter Konvergenz.

2. Für die Abschätzung benötigen wir die Taylorentwicklung von ψ_X mit Restglied. Es gilt

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

mit Zufallszahlen θ_1, θ_2 , so dass $|\theta_1|, |\theta_2| \leq 1$. Deshalb bekommen wir

$$\psi_X(t) = 1 + it\mathbf{E}[X] - \frac{t^2}{2}\mathbf{E}[X^2] + \varepsilon(t)t^2$$

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

Beispiel 7.15 (Momente der Exponential- und Normalverteilung). 1. Sei X eine nach $\exp(\gamma)$ -verteilte Zufallsvariable. Wir haben bereits die Laplace-Transformierte von X , $\mathcal{L}_{\exp(\gamma)}(t) = \gamma/(\gamma + t)$, berechnet. Daraus ergeben sich leicht alle Momente von X , nämlich

$$\mathbf{E}[X^n] = (-1)^n \frac{d^n}{dt^n} \mathbf{E}[e^{-tX}] \Big|_{t=0} = (-1)^n \frac{d^n}{dt^n} \frac{\gamma}{\gamma + t} \Big|_{t=0} = \frac{n! \gamma}{(\gamma + t)^{n+1}} \Big|_{t=0} = \frac{n!}{\gamma^n}.$$

2. Für eine nach $N(\mu, \sigma^2)$ -verteilte Zufallsvariable X kennen wir bereits die charakteristische Funktion $\psi_{N(\mu, \sigma^2)}(t) = e^{it\mu - \sigma^2 t^2/2}$. Für kleine t entwickeln wir dies mit

$$\psi_{N(\mu, \sigma^2)}(t) = 1 + it\mu - \sigma^2 t^2/2 - \mu^2 t^2/2 + \varepsilon(t)t^2$$

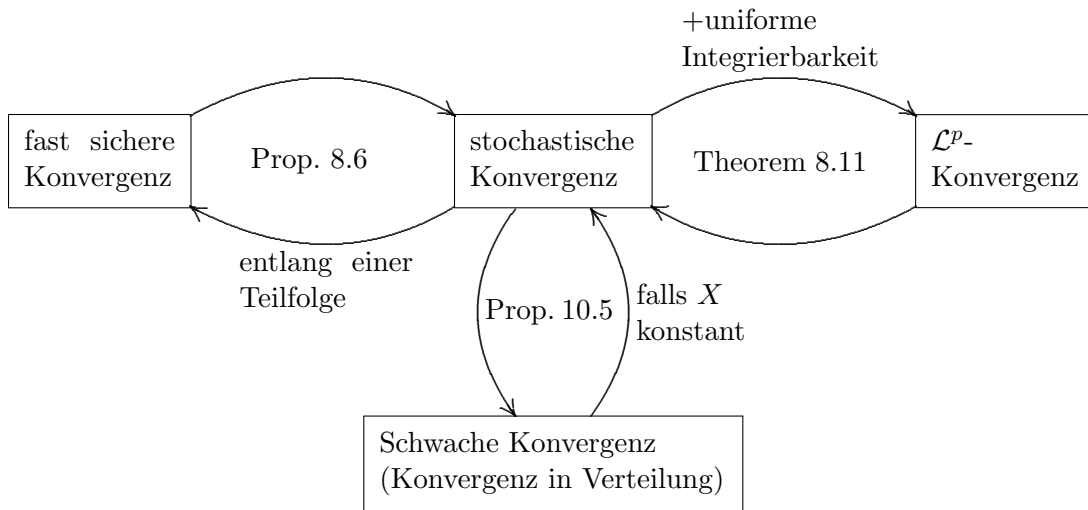
mit $\varepsilon(t) \xrightarrow{t \rightarrow 0} 0$. Daraus liest man mittels Proposition 7.14.2 ab, dass

$$\mathbf{E}[X] = \mu, \quad \mathbf{V}[X] = \mathbf{E}[X^2] - \mu^2 = \sigma^2.$$

8 Fast sichere, stochastische und \mathcal{L}^p -Konvergenz

Aus der Analysis ist bereits bekannt, dass es für Funktionenfolgen verschiedene Konvergenzarten gibt, etwa die gleichmäßige und die punktweise Konvergenz. Wir wollen nun die wichtigsten Konvergenzarten besprechen, bezüglich derer Zufallsvariable konvergieren können.

Neben der fast sicheren Konvergenz werden wir die stochastische Konvergenz und die \mathcal{L}^p -Konvergenz (siehe auch Abschnitt 5) kennenlernen. Im Abschnitt 10 werden wir außerdem die Konvergenz in Verteilung (was dasselbe ist wie die schwache Konvergenz der Verteilungen der Zufallsvariablen) kennen lernen. Folgendes Schaubild fasst alle Konvergenzarten zusammen:



8.1 Definition und Beispiele

Wir beginnen mit einigen Definitionen.

Definition 8.1 (Fast sichere und stochastische Konvergenz). Seien X, X_1, X_2, \dots Zufallsvariablen mit Werten in einem metrischen Raum (E, r) .

1. Ist

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} r(X_n, X) = 0\right) = 1,$$

sagen wir, dass die Folge X_1, X_2, \dots fast sicher gegen X konvergiert und schreiben $X_n \xrightarrow{n \rightarrow \infty}_{fs} X$.

2. Ist

$$\lim_{n \rightarrow \infty} \mathbf{P}(r(X_n, X) > \varepsilon) = 0,$$

für alle $\varepsilon > 0$, sagen wir, dass die Folge X_1, X_2, \dots in Wahrscheinlichkeit (oder stochastisch) gegen X konvergiert und schreiben $X_n \xrightarrow{n \rightarrow \infty}_p X$.

3. Sind die Zufallsvariablen reellwertig und ist für ein $p > 0$

$$\lim_{n \rightarrow \infty} \mathbf{E}[|X_n - X|^p] = 0,$$

sagen wir, dass die Folge X_1, X_2, \dots in \mathcal{L}^p (oder im p -ten Mittel) gegen X konvergiert und schreiben auch $X_n \xrightarrow{n \rightarrow \infty}_{\mathcal{L}^p} X$.

Bemerkung 8.2 (Eigenschaften der \mathcal{L}^p -Konvergenz). Wir wissen aus Abschnitt 5 schon einiges über die \mathcal{L}^p -Konvergenz. Ist etwa X, X_1, X_2, \dots so, dass $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^q X$ und $p < q$, dann gilt auch $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p X$ nach Proposition 5.6. Außerdem sind die Räume \mathcal{L}^p vollständig nach Proposition 5.7. Gibt es also für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$, so dass für alle $m, n \geq n$

$$\mathbf{E}[|X_n - X_m|^p] < \varepsilon,$$

so gibt es eine Zufallsvariable $X \in \mathcal{L}^p$ mit $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p X$.

Beispiel 8.3 (Gegenbeispiele). Betrachten wir uns das Schaubild am Anfang des Kapitels, stellen wir fest, dass aus der fast sicheren zwar die stochastische Konvergenz folgt, jedoch nicht umgekehrt. Außerdem folgt zwar aus der \mathcal{L}^1 -Konvergenz die stochastische, jedoch folgt nicht einmal aus der fast sicheren Konvergenz die \mathcal{L}^1 -Konvergenz. Wir geben zunächst zwei Beispiele für diese beiden Fälle.

1. Aus der stochastischen Konvergenz folgt nicht die fast sichere: Sei U eine auf $[0, 1]$ uniform verteilte Zufallsvariable. Weiter setzen wir

$$\begin{aligned} A_1 &= [0, \frac{1}{2}], & A_2 &= [\frac{1}{2}, 1], \\ A_3 &= [0, \frac{1}{4}], & A_4 &= [\frac{1}{4}, \frac{2}{4}], & A_5 &= [\frac{2}{4}, \frac{3}{4}], & A_6 &= [\frac{3}{4}, 1], \\ &\dots & & & & & & \end{aligned}$$

und $X_n := 1_{U \in A_n}$. Dann gilt klar für $0 < \varepsilon < 1$

$$\lim_{n \rightarrow \infty} \mathbf{P}(|X_n| > \varepsilon) = \lim_{n \rightarrow \infty} \mathbf{P}(U \in A_n) = 0,$$

d.h. $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p 0$, jedoch gibt es für jedes $n \in \mathbb{N}$ eine $m > n$ mit $X_m = 1$. Deshalb konvergiert die Folge X_1, X_2, \dots nicht fast sicher gegen 0.

2. Aus der fast sicheren folgt nicht die \mathcal{L}^1 -Konvergenz: Sei wieder U eine auf $[0, 1]$ uniform verteilte Zufallsvariable. Weiter ist $B_n = [0, \frac{1}{n}]$ und $Y_n = n \cdot 1_{U \in B_n}$. Dann ist $Y_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^1 Y$ fast sicher, also $Y = \infty \cdot 1_{U=0}$, also $Y = 0$ fast sicher. Andererseits ist

$$\mathbf{E}[Y_n] = n \cdot \mathbf{P}(U \in B_n) = 1,$$

also konvergiert Y_1, Y_2, \dots nicht in \mathcal{L}^1 gegen 0.

Lemma 8.4 (Stochastischer Limes eindeutig). Sei X, Y, X_1, X_2, \dots Zufallsvariable mit Werte in einem metrischen Raum (E, r) und $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p X$ sowie $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p Y$. Dann ist $X = Y$ fast sicher.

Beweis. Es gilt für alle $\varepsilon > 0$

$$\mathbf{P}(r(X, Y) > 2\varepsilon) \leq \mathbf{P}(r(X_n, X) > \varepsilon \text{ oder } r(X_n, Y) > \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

Also folgt

$$\mathbf{P}(X \neq Y) = \mathbf{P}\left(\bigcup_{k=1}^{\infty} \{r(X, Y) > 1/k\}\right) \leq \sum_{k=1}^{\infty} \mathbf{P}(r(X, Y) > 1/k) = 0,$$

woraus die Aussage folgt. □

8.2 Fast sichere und stochastische Konvergenz

Wir zeigen nun ein Resultat, das fast sichere und Konvergenz in Wahrscheinlichkeit in Beziehung setzt.

Lemma 8.5 (Charakterisierung von stochastischer Konvergenz). *Seien X, X_1, X_2, \dots Zufallsvariablen mit Werten in einem metrischen Raum (E, r) . Dann gilt*

$$X_n \xrightarrow{n \rightarrow \infty}_p X \iff \mathbf{E}[r(X_n, X) \wedge 1] \xrightarrow{n \rightarrow \infty} 0. \quad (8.1)$$

Beweis. Falls $X_n \xrightarrow{n \rightarrow \infty}_p X$, so gilt für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbf{E}[r(X_n, X) \wedge 1] \leq \lim_{n \rightarrow \infty} (\varepsilon + \mathbf{P}(r(X_n, X) > \varepsilon)) = \varepsilon,$$

womit die rechte Seite gezeigt ist. Gilt hingegen die rechte Seite, folgt mit der Chebyshev-Ungleichung für $0 < \varepsilon \leq 1$, dass

$$\mathbf{P}(r(X_n, X) > \varepsilon) \leq \frac{\mathbf{E}[r(X_n, X) \wedge 1]}{\varepsilon} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Proposition 8.6 (Konvergenz in Wahrscheinlichkeit und fast sichere Konvergenz).

Seien X, X_1, X_2, \dots Zufallsvariablen mit Werten in einem metrischen Raum (Ω', r) . Dann sind äquivalent:

1. $X_n \xrightarrow{n \rightarrow \infty}_p X$
2. Für jede Folge $(n_k)_{k=1,2,\dots}$ gibt es eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$ mit $X_{n_{k_\ell}} \xrightarrow{\ell \rightarrow \infty}_{fs} X$.

Insbesondere gilt

$$X_n \xrightarrow{n \rightarrow \infty}_{fs} X \implies X_n \xrightarrow{n \rightarrow \infty}_p X.$$

Beweis. 1. \Rightarrow 2.: Wegen (8.1) können wir für jede Folge $(n_k)_{k=1,2,\dots}$ eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$ wählen, so dass

$$\mathbf{E} \left[\sum_{\ell=1}^{\infty} (r(X_{n_{k_\ell}}, X) \wedge 1) \right] = \sum_{\ell=1}^{\infty} \mathbf{E}[r(X_{n_{k_\ell}}, X) \wedge 1] < \infty.$$

Das erste Gleichheitszeichen gilt dabei wegen monotoner Konvergenz. Damit ist

$$1 = \mathbf{P} \left(\sum_{\ell=1}^{\infty} (r(X_{n_{k_\ell}}, X) \wedge 1) < \infty \right) \leq \mathbf{P} \left(\limsup_{\ell \rightarrow \infty} r(X_{n_{k_\ell}}, X) = 0 \right) \leq 1,$$

d.h. $X_{n_{k_\ell}} \xrightarrow{\ell \rightarrow \infty}_{fs} X$.

2. \Rightarrow 1.: Nehmen wir an, dass 1. nicht gilt. Wegen (8.1) gibt es ein $\varepsilon > 0$ und eine Teilfolge $(n_k)_{k=1,2,\dots}$, so dass $\lim_{k \rightarrow \infty} \mathbf{E}[r(X_{n_k}, X) \wedge 1] > \varepsilon$. Angenommen, es gäbe nun eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$, so dass $X_{n_{k_\ell}} \xrightarrow{\ell \rightarrow \infty} X$ fast sicher. Dann wäre auch

$$\lim_{\ell \rightarrow \infty} \mathbf{E}[r(X_{n_{k_\ell}}, X) \wedge 1] = \mathbf{E} \left[\lim_{\ell \rightarrow \infty} r(X_{n_{k_\ell}}, X) \wedge 1 \right] = 0$$

wegen majorisierter Konvergenz, also ein Widerspruch. Also haben wir eine Folge $(n_k)_{k=1,2,\dots}$ gefunden, für die es keine weitere Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$ gibt mit $X_{n_{k_\ell}} \xrightarrow{\ell \rightarrow \infty}_{fs} X$, also haben wir gezeigt, dass 2. nicht gilt. \square

8.3 Stochastische und \mathcal{L}^p -Konvergenz

In Beispiel 8.3 hatten wir bereits gesehen, dass die fast sichere Konvergenz (und damit auch die stochastische Konvergenz) nicht die \mathcal{L}^1 -Konvergenz impliziert. Das erstaunt nicht, da ja der Satz von der majorisierten Konvergenz besagt, dass eine Folge X_1, X_2, \dots , die fast sicher gegen X konvergiert und eine integrierbare Majorante besitzt auch in \mathcal{L}^1 gegen X konvergiert. Würde die fast sichere Konvergenz die \mathcal{L}^1 -Konvergenz implizieren, bräuchte man die Forderung einer integrierbaren Majorante nicht zu machen. Wir wollen im Folgenden die Bedingung der integrierbaren Majorante für die \mathcal{L}^1 -Konvergenz abschwächen. Siehe Theorem 8.11 und Korollar 8.12. Der Begriff der gleichgradigen Integrierbarkeit ist hierfür zentral, siehe Definition 8.7.

Definition 8.7 (Gleichgradige Integrierbarkeit). Eine Familie $(X_i)_{i \in I}$ heißt gleichgradig integrierbar, falls

$$\inf_K \sup_{i \in I} \mathbf{E}[|X_i|; |X_i| > K] = 0$$

Beispiel 8.8 (Gleichgradige Integrierbarkeit). 1. Sei $Y \in \mathcal{L}^1$ und $(X_i)_{i \in I}$ mit $\sup_i |X_i| < |Y|$. Dann ist $(X_i)_{i \in I}$ gleichgradig integrierbar. Denn

$$\sup_{i \in I} \mathbf{E}[|X_i|; |X_i| > K] \leq \mathbf{E}[|Y|; |Y| > K] \xrightarrow{K \rightarrow \infty} 0$$

nach majorisierter Konvergenz. Insbesondere ist jedes $Y \in \mathcal{L}^1$ gleichgradig integrierbar.

2. Jede endliche Familie $(X_i)_{i=1, \dots, n}$ mit $X_i \in \mathcal{L}^1, i = 1, \dots, n$ ist gleichgradig integrierbar, denn $\sup_{1 \leq i \leq n} |X_i| \in \mathcal{L}^1$ und

$$\inf_K \sup_{1 \leq i \leq n} \mathbf{E}[|X_i|; |X_i| > K] \leq \inf_K \mathbf{E}[\sup_{1 \leq i \leq n} |X_i|; \sup_{1 \leq i \leq n} |X_i| > K] = 0$$

wegen majorisierter Konvergenz.

3. Betrachten wir das Beispiel 8.3.2. Hier ist für $n > K$

$$\mathbf{E}[|Y_n|; |Y_n| > K] = \mathbf{E}[Y_n] = 1.$$

Insbesondere ist $(Y_n)_{n=1,2,\dots}$ nicht gleichgradig integrierbar.

4. Sei $p > 1$. Dann ist $(X_i)_{i \in I}$ mit $X_i \in \mathcal{L}^p, i \in I$ gleichgradig integrierbar, falls $\sup_{i \in I} \|X_i\|_p < \infty$. Denn es gilt mit der Markov-Ungleichung (Proposition 7.4)

$$\sup_{i \in I} \mathbf{E}[|X_i|; |X_i| > K] \leq \sup_{i \in I} \frac{\mathbf{E}[|X_i|^p]}{K^{p-1}} \xrightarrow{K \rightarrow \infty} 0.$$

Lemma 8.9 (Charakterisierung von gleichgradiger Integrierbarkeit). Sei $(X_i)_{i \in I}$ eine Familie von Zufallsvariablen. Dann sind äquivalent:

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

2. Es gilt

$$\sup_{i \in I} \mathbf{E}[|X_i|] < \infty \quad \text{und} \quad \lim_{\varepsilon \rightarrow 0} \sup_{A: \mathbf{P}(A) < \varepsilon} \sup_{i \in I} \mathbf{E}[|X_i|; A] = 0.$$

3. Es gilt

$$\lim_{K \rightarrow \infty} \sup_{i \in I} \mathbf{E}[(|X_i| - K)^+] = 0.$$

4. Es gibt eine Funktion $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ so, dass $\frac{f(x)}{x} \xrightarrow{x \rightarrow \infty} \infty$ und $\sup_{i \in I} \mathbf{E}[f(|X_i|)] < \infty$.

Gilt eine der vier Aussagen, so kann die Funktion f in 4. monoton wachsend und konvex gewählt werden.

Beweis. '1. \Rightarrow 2.': Sei $\delta > 0$ gegeben und $K = K_\delta$ so, dass $\sup_{i \in I} \mathbf{E}[|X_i|; |X_i| > K] \leq \delta$. Dann ist für $A \in \mathcal{F}$

$$\mathbf{E}[|X_i|; A] = \mathbf{E}[|X_i|; A \cap \{|X_i| > K\}] + \mathbf{E}[|X_i|; A \cap \{|X_i| \leq K\}] \leq \delta + K \cdot \mathbf{P}(A).$$

Insbesondere ist

$$\sup_{i \in I} \mathbf{E}[|X_i|] = \sup_{i \in I} \mathbf{E}[|X_i|; \Omega] \leq \delta + K < \infty$$

und

$$\sup_{A: \mathbf{P}(A) < \varepsilon} \sup_{i \in I} \mathbf{E}[|X_i|; A] \leq \delta + K\varepsilon \xrightarrow{\varepsilon \rightarrow 0} \delta.$$

Da $\delta > 0$ beliebig war, muss

$$\lim_{\varepsilon \rightarrow 0} \sup_{A: \mathbf{P}(A) < \varepsilon} \sup_{i \in I} \mathbf{E}[|X_i|; A] = 0$$

gelten.

'2. \Rightarrow 3.': Zunächst bemerken wir, dass $(|X_i| - K)^+ \leq |X_i| 1_{|X_i| \geq K}$. Sei $\varepsilon > 0$. Wähle $K = K_\varepsilon$ groß genug, so dass nach der Markov-Ungleichung

$$\sup_{i \in I} \mathbf{P}(|X_i| > K) \leq \sup_{i \in I} \frac{\mathbf{E}[|X_i|]}{K} < \varepsilon$$

ist. Damit folgt 3. aus

$$\begin{aligned} \lim_{K \rightarrow \infty} \sup_{i \in I} \mathbf{E}[(|X_i| - K)^+] &= \lim_{\varepsilon \rightarrow 0} \sup_{i \in I} \mathbf{E}[(|X_i| - K_\varepsilon)^+] \leq \lim_{\varepsilon \rightarrow 0} \sup_{i \in I} \mathbf{E}[|X_i|; |X_i| > K_\varepsilon] \\ &\leq \lim_{\varepsilon \rightarrow 0} \sup_{A: \mathbf{P}(A) < \varepsilon} \sup_{i \in I} \mathbf{E}[|X_i|; A] = 0. \end{aligned}$$

'3. \Rightarrow 4.': Es gibt eine Folge K_1, K_2, \dots mit $K_n \uparrow \infty$ und $\sup_{i \in I} \mathbf{E}[(|X_i| - K_n)^+] \leq 2^{-n}$. Wir setzen

$$f(x) := \sum_{n=1}^{\infty} (x - K_n)^+.$$

Dann ist f monoton wachsend und als Summe konvexer Funktionen wieder konvex. Außerdem ist für $x \geq 2K_n$

$$\frac{f(x)}{x} \geq \sum_{k=1}^n \left(1 - \frac{K_k}{x}\right) \geq \frac{n}{2},$$

also $\frac{f(x)}{x} \xrightarrow{x \rightarrow \infty} \infty$. Wegen monotoner Konvergenz gilt außerdem

$$\mathbf{E}[f(|X_i|)] = \sum_{n=1}^{\infty} \mathbf{E}[(|X_i| - K_n)^+] \leq \sum_{n=1}^{\infty} 2^{-n} = 1.$$

'4. \Rightarrow 1.': Setze $a_K := \inf_{x \geq K} \frac{f(x)}{x}$, so dass $a_K \xrightarrow{K \rightarrow \infty} \infty$. Also ist

$$\sup_{i \in I} \mathbf{E}[|X_i|; |X_i| \geq K] \leq \frac{1}{a_K} \sup_{i \in I} \mathbf{E}[f(|X_i|); |X_i| \geq K] \leq \frac{1}{a_K} \sup_{i \in I} \mathbf{E}[f(|X_i|)] \xrightarrow{K \rightarrow \infty} 0. \quad \square$$

Beispiel 8.10 (Differenz und gleichgradige Integrierbarkeit). Für $X \in \mathcal{L}^1$ ist $(X_i)_{i \in I}$ genau dann gleichgradig integrierbar, wenn $(X_i - X)_{i \in I}$ gleichgradig integrierbar ist.

Um dies zu sehen, sei $(X_i)_{i \in I}$ gleichgradig integrierbar. Nach Beispiel 8.8.2 ist X gleichgradig integrierbar. Außerdem ist

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

und es gilt

$$\sup_{A: \mathbf{P}(A) < \varepsilon} \sup_{i \in I} \mathbf{E}[|X_i - X|; A] \leq \sup_{A: \mathbf{P}(A) < \varepsilon} \sup_{i \in I} \mathbf{E}[|X_i|; A] + \sup_{A: \mathbf{P}(A) < \varepsilon} \mathbf{E}[|X|; A] \xrightarrow{\varepsilon \rightarrow 0} 0,$$

d.h. nach Lemma 8.9 ist $(X_i - X)_{i \in I}$ gleichgradig integrierbar. Die Umkehrung folgt analog.

Theorem 8.11 (Stochastische und Konvergenz im p -ten Mittel). Sei X_1, X_2, \dots eine Folge in \mathcal{L}^p mit $0 < p < \infty$. Folgende Aussagen sind äquivalent:

1. Es gibt eine messbare Funktion $X \in \mathcal{L}^p$ mit $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^p X$.
2. Die Familie $(|X_i|^p)_{i=1,2,\dots}$ ist gleichgradig integrierbar und es gibt eine messbare Funktion X mit $X_n \xrightarrow{n \rightarrow \infty} X$ stochastisch.

Gilt 1. oder 2. dann stimmen die Limiten überein.

Beweis. 1. \Rightarrow 2.: Zunächst ist wegen der Chebyshev'schen Ungleichung für jedes $\varepsilon > 0$

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

d.h. die stochastische Konvergenz gilt. Für den Beweis der gleichgradigen Integrierbarkeit verwenden wir Lemma 8.9. Sei $\varepsilon > 0$ und $N = N_\varepsilon$ so, dass $\|X_n - X\|_p < \varepsilon$ für $n \geq N$. Dann ist mit der Minkowski-Ungleichung (7.1)

$$\begin{aligned} \sup_{n \in \mathbb{N}} (\mathbf{E}[|X_n|^p])^{1 \wedge 1/p} &= \sup_{n \in \mathbb{N}} \|X_n\|_p^{p \wedge 1} \leq \sup_{n < N} \|X_n\|_p^{p \wedge 1} + \sup_{n \geq N} \|X_n - X\|_p^{p \wedge 1} + \|X\|_p^{p \wedge 1} \\ &\leq \sup_{n < N} (\mathbf{E}[|X_n|^p])^{1 \wedge 1/p} + \varepsilon^{p \wedge 1} + (\mathbf{E}[|X|^p])^{1 \wedge 1/p} < \infty, \end{aligned}$$

und für $A \in \mathcal{F}$, wieder mit der Minkowski Ungleichung

$$\begin{aligned} \sup_{n \in \mathbb{N}} (\mathbf{E}[|X_n|^p; A])^{1 \wedge 1/p} &= \sup_{n \in \mathbb{N}} \|X_n 1_A\|_p^{p \wedge 1} \\ &\leq \sup_{n < N} \|X_n 1_A\|_p^{p \wedge 1} + \sup_{n \geq N} \|(X_n - X) 1_A\|_p^{p \wedge 1} + \|X 1_A\|_p^{p \wedge 1} \\ &\leq \sup_{n < N} (\mathbf{E}[|X_n|^p; A])^{1 \wedge 1/p} + \varepsilon^{p \wedge 1} + (\mathbf{E}[|X|^p; A])^{1 \wedge 1/p}. \end{aligned}$$

Da N endlich ist, folgt

$$\lim_{\delta \rightarrow 0} \sup_{A: \mathbf{P}(A) < \delta} \sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|^p; A] \leq \varepsilon^p.$$

Weil $\varepsilon > 0$ beliebig war, folgt die Behauptung.

2. \Rightarrow 1.: Da $X_n \xrightarrow[n \rightarrow \infty]{p} X$, gibt es nach Proposition 8.6 eine Teilfolge n_1, n_2, \dots mit $X_{n_k} \xrightarrow[k \rightarrow \infty]{} X$ fast sicher. Mit dem Lemma von Fatou ist

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

wegen Lemma 8.9. Insbesondere ist $X \in \mathcal{L}^p$. Genau wie in Beispiel 8.10 ist auch $\{|X_n - X|^p : n \in \mathbb{N}\}$ gleichgradig integrierbar. Für jedes $\delta > 0$ gilt wegen der stochastischen Konvergenz

$$\mathbf{P}(|X_n - X| > \delta) \xrightarrow{n \rightarrow \infty} 0.$$

Aus Lemma 8.9 folgt nun mit majorisierter Konvergenz

$$\lim_{n \rightarrow \infty} \mathbf{E}[|X_n - X|^p] = \lim_{n \rightarrow \infty} \mathbf{E}[|X_n - X|^p; |X_n - X| > \delta] + \mathbf{E}[|X_n - X|^p; |X_n - X| \leq \delta] \leq \delta^p.$$

Da $\delta > 0$ beliebig war, folgt $X_n \xrightarrow[n \rightarrow \infty]{} X$ in \mathcal{L}^p . \square

Korollar 8.12 (Erwartungswert-Konvergenz und gleichgradige Integrierbarkeit).
Sei $0 < p < \infty$ und $X_1, X_2, \dots \in \mathcal{L}^p$ und X messbar mit $X_n \xrightarrow[n \rightarrow \infty]{p} X$. Folgende Aussagen sind äquivalent:

1. $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^p} X$,
2. $\|X_n\|_p \xrightarrow[n \rightarrow \infty]{} \|X\|_p$,
3. Die Familie $(|X_n|^p)_{n=1,2,\dots}$ ist gleichgradig integrierbar.

Beweis. Die Äquivalenz 1. \Leftrightarrow 3. ist klar aus Theorem 8.11.

1. \Rightarrow 2.: folgt aus der Minkowski'schen Ungleichung mit

$$\left| \|X_n\|_p^{p \wedge 1} - \|X\|_p^{p \wedge 1} \right| \leq \|X_n - X\|_p^{p \wedge 1} \xrightarrow{n \rightarrow \infty} 0.$$

2. \Rightarrow 3.: Für festes K ist

$$\mathbf{E}[|X_n|^p; |X_n| > K] \leq \mathbf{E}[|X_n|^p - (|X_n| \wedge (K - |X_n|)^+)^p] \xrightarrow{n \rightarrow \infty} \mathbf{E}[|X|^p - (|X| \wedge (K - |X|)^+)^p].$$

Die Konvergenz folgt hierbei, da $\mathbf{E}[|X_n|^p] \xrightarrow{n \rightarrow \infty} \mathbf{E}[|X|^p]$, und $(|X_n| \wedge (K - |X_n|)^+)^p \xrightarrow{n \rightarrow \infty} |X| \wedge (K - |X|)^+)^p$, da die Konvergenz nach Proposition 8.6 stochastisch gilt und $((|X_n| \wedge (K - |X_n|)^+)^p)_{n=1,2,\dots}$ beschränkt, insbesondere gleichgradig integrierbar ist. Da $\mathbf{E}[|X|^p - (|X| \wedge (K - |X|)^+)^p] \xrightarrow{K \rightarrow \infty} 0$ nach majorisierter Konvergenz, ist $(|X_n|^p)_{n=1,2,\dots}$ gleichgradig integrierbar. \square

9 Unabhängigkeit und das starke Gesetz

Mit unserem Wissen über Wahrscheinlichkeitsmaße und σ -Algebren beleuchten wir nun den Begriff der stochastischen Unabhängigkeit. Insbesondere werden wir in diesem Kapitel das starke Gesetz der großen Zahlen beweisen, siehe Theorem 9.21. Auf dem Weg dahin beweisen wir das Borel-Cantelli Lemma (Theorem 9.8) und das Kolmogorov'sche 0-1-Gesetz (Theorem 9.15).

9.1 Definition und einfache Eigenschaften

Bereits in der Vorlesung *Stochastik* wurden unabhängige Zufallsvariablen betrachtet. Die intuitive Vorstellung von Unabhängigkeit ist oft richtig, manchmal jedoch mit Vorsicht zu genießen.

Definition 9.1 (Unabhängigkeit). 1. Eine Familie von Mengen $(A_i)_{i \in I}$ mit $A_i \in \mathcal{F}$ heißt unabhängig, falls

$$\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j) \quad (9.1)$$

für alle $J \subseteq I$.

2. Eine Familie $(C_i)_{i \in I}$ von Mengensystemen $C_i \subseteq \mathcal{F}$ heißt unabhängig, falls (9.1) für alle $J \subseteq I$ und $A_j \in C_j, j \in J$ gilt.
3. Eine Familie von Zufallsvariablen $(X_i)_{i \in I}$ heißt unabhängig, falls $(\sigma(X_i))_{i \in I}$ unabhängig ist.

Wir beschäftigen uns zuerst mit der Frage, ob es Wahrscheinlichkeitsräume gibt, auf denen es beliebig viele unabhängige Zufallsvariablen gibt. Hierbei kommt uns das Wissen über Produktmaße zu Gute.

Proposition 9.2 (Unabhängigkeit und Produktmaße). Eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen ist genau dann unabhängig, falls für jedes $J \subseteq I$

$$((X_i)_{i \in J})_* \mathbf{P} = \bigotimes_{i \in J} (X_i)_* \mathbf{P},$$

die gemeinsame Verteilung jeder endlichen Teilfamilie also gleich der Produktverteilung der einzelnen Verteilungen ist.

Beweis. Nach Definition ist die Familie $(X_i)_{i \in I}$ genau dann unabhängig, falls für jedes $J \subseteq I$ und $A_i \in \mathcal{F}, i \in J$,

$$\mathbf{P}(X_i \in A_i, i \in J) = \prod_{i \in J} \mathbf{P}(X_i \in A_i).$$

Die Behauptung folgt nun daraus, dass $\mathbf{P}(X_i \in A_i) = (X_i)_* \mathbf{P}(A_i)$ (siehe Definition 3.22) und $\mathbf{P}(X_i \in A_i, i \in J) = ((X_i)_{i \in J})_* \mathbf{P}(\times_{i \in J} A_i)$ (siehe Korollar 6.14). \square

Korollar 9.3 (Existenz von überabzählbar vielen unabhängigen Zufallsvariablen). Sei E ein polnischer Raum, I eine beliebige Indexmenge. Seien $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$ Wahrscheinlichkeitsräume und X_i E -wertige Zufallsvariable, $i \in I$. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und eine Familie $(Y_i)_{i \in I}$ E -wertiger, unabhängiger Zufallsvariable mit $Y_i \stackrel{d}{=} X_i$.

Beweis. Es sei bemerkt, dass $((X_i)_{i \in J})_* \otimes_{i \in J} \mathbf{P}_i)_{J \in I}$ eine projektive Familie von Wahrscheinlichkeitsmaßen auf $(E, \mathcal{B}(E))$ ist. Wegen Theorem 6.24 gibt es also den projektiven Limes \mathbf{P}_I . Dieser ist ein Wahrscheinlichkeitsmaß auf $(E^I, (\mathcal{B}(E))^I)$. Außerdem ist mit $\pi_i : E^I \rightarrow E$, der i -ten Projektion, $(\pi_i)_* \mathbf{P}_I = (X_i)_* \mathbf{P}_i$, d.h. $\pi_i \stackrel{d}{=} X_i$. \square

Lemma 9.4 (Funktionen unabhängiger Zufallsvariablen). *Seien $(\Omega'_i, \mathcal{F}'_i), (\Omega''_i, \mathcal{F}''_i), i \in I$, Messräume. Sei $(X_i)_{i \in I}$ eine Familie unabhängiger Zufallsvariablen, $X_i : \Omega \rightarrow \Omega'_i$, und $\varphi_i : \Omega'_i \rightarrow \Omega''_i$ messbar, $i \in I$. Dann ist auch die Familie $(\varphi_i(X_i))_{i \in I}$ unabhängig.*

Beweis. Nach Lemma 7.2 ist die Zufallsvariable $\varphi_i(X_i)$ nach $\sigma(X_i)$ messbar, $i \in I$, d.h. $\sigma(\varphi_i(X_i)) \subseteq \sigma(X_i)$. Da $(\sigma(X_i))_{i \in I}$ nach Voraussetzung unabhängig sind, folgt die Behauptung aus der Definition der Unabhängigkeit. \square

Proposition 9.5 (Unabhängigkeit und Unkorreliertheit). *Seien $X, Y \in \mathcal{L}^1$ unabhängige, reellwertige Zufallsvariablen. Dann ist $XY \in \mathcal{L}^1$ und es gilt*

$$\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Beweis. Zunächst bemerken wir: gilt die Behauptung für die Paare (X_i, Y_j) , $i, j = 1, \dots, n$, so auch für $\sum_{i=1}^n X_i$ und $\sum_{j=1}^n Y_j$. Wegen der Linearität des Erwartungswertes ist nämlich

$$\mathbf{E}\left[\sum_{i=1}^n X_i \cdot \sum_{j=1}^n Y_j\right] = \sum_{i=1}^n \sum_{j=1}^n \mathbf{E}[X_i Y_j] = \sum_{i=1}^n \sum_{j=1}^n \mathbf{E}[X_i] \mathbf{E}[Y_j] = \mathbf{E}\left[\sum_{i=1}^n X_i\right] \cdot \mathbf{E}\left[\sum_{j=1}^n Y_j\right].$$

Die Behauptung ist klar, wenn X und Y Indikatorfunktionen sind. Wegen oben gesagtem gilt sie damit auch für einfache Funktionen, und damit mit monotoner Konvergenz auch für nicht-negative messbare Funktionen. Der allgemeine Fall folgt mit der Zerlegung $X = X^+ - X^-$ und $Y = Y^+ - Y^-$. \square

Beispiel 9.6 (Unkorrelierte, nicht unabhängige Zufallsvariablen). Sei U eine auf $[0, 1]$ uniform verteilte Zufallsvariable, $X = \cos(2\pi U)$ und $Y = \sin(2\pi U)$. Dann ist $\mathbf{E}[X] = \mathbf{E}[Y] = 0$ und

$$\mathbf{E}[XY] = \int_0^1 \cos(2\pi u) \sin(2\pi u) du = \frac{1}{2} \int_0^1 \sin(4\pi u) du = 0$$

und damit sind X, Y unkorreliert. Allerdings ist $\{|X| < \varepsilon, |Y| < \varepsilon\} = \emptyset$ für $\varepsilon > 0$ klein genug und damit ist $\mathbf{P}(X^{-1}(-\varepsilon, \varepsilon), Y^{-1}(-\varepsilon, \varepsilon)) = 0 < \mathbf{P}(X^{-1}(-\varepsilon, \varepsilon)) \cdot \mathbf{P}(Y^{-1}(-\varepsilon, \varepsilon))$. Damit sind X und Y nicht unabhängig.

Hat man einen Wahrscheinlichkeitsraum und (abzählbar) viele Ereignisse, kann man sich fragen, wie viele dieser Ereignisse wohl eintreten. Das Borel-Cantelli Lemma gibt ein scharfes Kriterium dafür, wann nur endlich viele Ereignisse eintreten.

Definition 9.7 (limsup von Mengen). *Für $A_1, A_2, \dots \in \mathcal{F}$ ist*

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

das Ereignis 'unendlich viele der A_n treten ein'.

Theorem 9.8 (Borel-Cantelli Lemma). 1. Sei $A_1, A_2, \dots \in \mathcal{F}$. Dann gilt

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty \implies \mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 0.$$

2. Sind A_1, A_2, \dots unabhängig, so gilt auch

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty \implies \mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 1.$$

Beweis. Wir beginnen mit 1. Wegen der Stetigkeit von \mathbf{P} von oben (siehe Proposition 3.7) gilt

$$\mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} \mathbf{P}\left(\bigcup_{m \geq n} A_m\right) \leq \lim_{n \rightarrow \infty} \sum_{m=n}^{\infty} \mathbf{P}(A_m) = 0$$

nach Voraussetzung. Für 2. verwenden wir, dass $\log(1-x) \leq -x$ für $x \in [0, 1]$. Damit gilt nämlich, wegen der Stetigkeit von \mathbf{P} von unten und der Unabhängigkeit von $(A_n)_{n=1,2,\dots}$

$$\begin{aligned} \mathbf{P}((\limsup_{n \rightarrow \infty} A_n)^c) &= \mathbf{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m^c\right) \\ &= \lim_{n \rightarrow \infty} \mathbf{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) \\ &= \lim_{n \rightarrow \infty} \prod_{m=n}^{\infty} (1 - \mathbf{P}(A_m)) \\ &= \lim_{n \rightarrow \infty} \exp\left(\sum_{m=n}^{\infty} \log(1 - \mathbf{P}(A_m))\right) \\ &\leq \lim_{n \rightarrow \infty} \exp\left(-\sum_{m=n}^{\infty} \mathbf{P}(A_m)\right) \\ &= 0 \end{aligned}$$

und die Behauptung folgt. □

Beispiel 9.9 (Unendlicher Münzwurf und geometrische Verteilungen).

- Wir betrachten einen unendlichen Münzwurf. Das bedeutet, dass wir einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und unabhängige Zufallsvariablen X_1, X_2, \dots mit Werte in $\{\text{Kopf}, \text{Zahl}\}$ haben. Der Münzwurf sei fair, d.h. $\mathbf{P}(X_n = \text{Kopf}) = 1/2$. Wir betrachten die Ereignisse $A_n = \{X_n = \text{Kopf}\}$. Da

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{2} = \infty$$

und die Familie $(A_n)_{n \in \mathbb{N}}$ unabhängig ist, folgt aus dem Borel-Cantelli Lemma, dass fast sicher unendlich oft *Kopf* kommt.

2. Wir betrachten dieselbe Situation wie in 1., jedoch die Ereignisse $B_n := \{X_1 = \text{Kopf}\}$. Klar ist, dass die Familie $(B_n)_{n \in \mathbb{N}}$ nicht unabhängig ist. (Z.B. ist ja $\mathbf{P}(B_1 \cap B_2) = \mathbf{P}(B_1) = 1/2 \neq \frac{1}{4} = \mathbf{P}(B_1) \cdot \mathbf{P}(B_2)$.) Genau wie in 1. ist $\sum_{n=1}^{\infty} \mathbf{P}(B_n) = \infty$. Klar ist auch, dass $\mathbf{P}(\limsup_{n \rightarrow \infty} B_n) = \frac{1}{2}$. Daraus folgt, dass im Borel-Cantelli Lemma auf die Bedingung der Unabhängigkeit in 2. nicht verzichtet werden kann.
3. Seien X_1, X_2, \dots zum Erfolgsparameter p geometrisch verteilte Zufallsvariablen. Wir betrachten die Ereignisse $A_n := \{X_n \geq n\}$ und fragen uns, ob unendlich viele dieser Ereignisse eintreten können. Da

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \sum_{n=1}^{\infty} \mathbf{P}(X_n \geq n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{1}{p} < \infty.$$

Deshalb treten fast sicher nur endlich viele der Ereignisse $\{X_n \geq n\}$ ein.

9.2 Das Kolmogorov'sche 0-1-Gesetz

Bereits das Borel-Cantelli Lemma ist eine Aussage darüber, wann ein von unendlich vielen Ereignissen abhängiges Ereignis fast sicher eintritt. Diese Situation werden wir nun weiter beleuchten.

Proposition 9.10 (Unabhängigkeit erzeugter σ -Algebren). *Sei $(\mathcal{C}_i)_{i \in I}$ eine Familie unabhängiger, schnittstabiler Mengensysteme. Dann ist auch $(\sigma(\mathcal{C}_i))_{i \in I}$ eine unabhängige Familie.*

Beweis. Sei $J = \{i_1, \dots, i_n\} \in I$ und o.E. $|J| > 1$. Dann gilt (9.1) für beliebige A_{i_1}, \dots, A_{i_n} mit $A_{i_k} \in \mathcal{C}_{i_k}, k = 1, \dots, n$. Wir halten A_{i_2}, \dots, A_{i_n} fest und definieren

$$\mathcal{D} := \{A_{i_1} \in \mathcal{F} : (9.1) \text{ gilt}\}.$$

Wir werden nun zeigen, dass \mathcal{D} ein Dynkin-System ist. Ist nämlich $A \subseteq B \in \mathcal{D}$, so ist auch $B \setminus A \in \mathcal{D}$, weil

$$\begin{aligned} \mathbf{P}\left((B \setminus A) \cap \bigcap_{k=2}^n A_{i_k}\right) &= \mathbf{P}\left(B \cap \bigcap_{k=2}^n A_{i_k}\right) - \mathbf{P}\left(A \cap \bigcap_{k=2}^n A_{i_k}\right) \\ &= (\mathbf{P}(B) - \mathbf{P}(A)) \cdot \prod_{k=2}^n \mathbf{P}(A_{i_k}) \\ &= \mathbf{P}(B \setminus A) \cdot \prod_{k=2}^n \mathbf{P}(A_{i_k}). \end{aligned}$$

Ist außerdem $A_1, A_2, \dots \in \mathcal{D}$ mit $A_1 \subseteq A_2 \subseteq A_3 \dots$, so ist wegen der Stetigkeit von \mathbf{P} von unten

$$\begin{aligned} \mathbf{P}\left(\left(\bigcup_{j=1}^{\infty} A_j\right) \cap \bigcap_{k=2}^n A_{i_k}\right) &= \sup_{j \in \mathbb{N}} \mathbf{P}\left(A_j \cap \bigcap_{k=2}^n A_{i_k}\right) \\ &= \sup_{j \in \mathbb{N}} \mathbf{P}(A_j) \cdot \prod_{k=2}^n \mathbf{P}(A_{i_k}) \\ &= \mathbf{P}\left(\bigcup_{j=1}^{\infty} A_j\right) \cdot \prod_{k=2}^n \mathbf{P}(A_{i_k}). \end{aligned}$$

Da \mathcal{C}_{i_1} schnittstabil ist und $\mathcal{C}_{i_1} \subseteq \mathcal{D}$, ist $\sigma(\mathcal{C}_{i_1}) \subseteq \mathcal{D}$ nach Theorem 2.13. Insbesondere gilt (9.1) für $A_{i_1} \in \sigma(\mathcal{C}_{i_1}), A_{i_2} \in \mathcal{C}_{i_2}, \dots, A_{i_n} \in \mathcal{C}_{i_n}$. Iteriert man obiges Verfahren für $k = 2, \dots, n$, erhält man die Aussage. \square

Korollar 9.11 (Unabhängigkeit von Indikatorfunktionen). *Eine Familie von Mengen $(A_i)_{i \in I}$ ist genau dann unabhängig, wenn die Familie der Zufallsvariablen $(1_{A_i})_{i \in I}$ unabhängig ist. Insbesondere gilt*

$$\mathbf{P}\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} \mathbf{P}(B_j)$$

für $J \subseteq I, B_j \in \{A_j, A_j^c\}, j \in J$.

Beweis. Für $i \in I$ sei $\mathcal{C}_i = \{A_i\}$. Dann ist $\sigma(1_{A_i}) = \{\emptyset, A_i, A_i^c, \Omega\} = \sigma(\mathcal{C}_i)$. Da \mathcal{C}_i trivialerweise schnittstabil ist, folgt die Aussage aus Proposition 9.10. \square

Korollar 9.12 (Gruppierung). *Sei $(\mathcal{F}_i)_{i \in I}$ eine Familie unabhängiger σ -Algebren. Weiter sei \mathcal{I} eine Partition von I , d.h. $\mathcal{I} = \{I_k, k \in K\}$ mit $\bigsqcup_{k \in K} I_k = I$, die I_k sind also disjunkt und deren Vereinigung ist I . Dann ist auch $(\sigma(\mathcal{F}_i : i \in I_k))_{k \in K}$ ein unabhängiges System.*

Beweis. Das Mengensystem $\mathcal{C}_k := \{\bigcap_{i \in J_k} A_i : J_k \subseteq I_k, A_i \in \mathcal{F}_i\}$ ist schnittstabil und $\sigma(\mathcal{C}_k) = \sigma(\mathcal{F}_i : i \in I_k), k \in K$. Da nach Voraussetzung die Familie $(\mathcal{C}_k)_{k \in K}$ unabhängig ist, folgt die Behauptung aus Proposition 9.10. \square

Wir kommen nun zur Hauptaussage dieses Abschnittes, dem Kolmogorov'schen 0-1-Gesetz. Hierzu führen wir eine bestimmte σ -Algebra ein, die terminale σ -Algebra.

Definition 9.13 (Terminale und triviale σ -Algebren). 1. Sei $\mathcal{F}_1, \mathcal{F}_2, \dots \subseteq \mathcal{F}$ eine Folge von σ -Algebren. Dann ist

$$\mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots) = \bigcap_{n \geq 1} \sigma\left(\bigcup_{m > n} \mathcal{F}_m\right)$$

die σ -Algebra der terminalen Ereignisse von $\mathcal{F}_1, \mathcal{F}_2, \dots$

2. Eine σ -Algebra $\tilde{\mathcal{F}} \subseteq \mathcal{F}$ heißt **P-trivial**, falls $\mathbf{P}(A) \in \{0, 1\}$ für alle $A \in \tilde{\mathcal{F}}$.

Lemma 9.14 (Triviale σ -Algebren). 1. Eine σ -Algebra $\tilde{\mathcal{F}}$ ist genau dann **P-trivial**, wenn $\tilde{\mathcal{F}}$ von sich selbst unabhängig ist.

2. Sei $\tilde{\mathcal{F}}$ eine **P-triviale** σ -Algebra und X eine $\tilde{\mathcal{F}}$ -messbare Zufallsvariable mit Werten in einem separablen metrischen Raum E . Dann ist X fast sicher konstant.

Beweis. 1. Sei $\tilde{\mathcal{F}}$ zunächst **P-trivial** und $A, B \in \tilde{\mathcal{F}}$. Dann gilt $\mathbf{P}(A \cap B) = \mathbf{P}(A) \wedge \mathbf{P}(B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$, also ist $\tilde{\mathcal{F}}$ von sich selbst unabhängig. Ist andererseits $\tilde{\mathcal{F}}$ von sich selbst unabhängig und $A \in \tilde{\mathcal{F}}$, dann ist $\mathbf{P}(A) = \mathbf{P}(A \cap A) = (\mathbf{P}(A))^2$, also $\mathbf{P}(A) \in \{0, 1\}$.

2. Für $n \in \mathbb{N}$ sei $(B_{nj})_{j=1,2,\dots}$ eine abzählbare Überdeckung von E mit Bällen vom Radius $1/n$. Da $\tilde{\mathcal{F}}$ eine **P-triviale** σ -Algebra ist, gilt $\mathbf{P}(X \in B_{nj}) \in \{0, 1\}$ für alle n, j . Für $n \in \mathbb{N}$ sei $J_n := \{j \in \mathbb{N} : \mathbf{P}(X \in B_{nj}) = 1\} \neq \emptyset$. Damit ist wegen der Stetigkeit von oben $\mathbf{P}\left(X \in \bigcap_{n=1}^{\infty} \bigcap_{j \in J_n} B_{nj}\right) = 1$. Da $\bigcap_{n=1}^{\infty} \bigcap_{j \in J_n} B_{nj}$ höchstens ein Element hat, folgt die Behauptung. \square

Unter Unabhängigkeit ist die σ -Algebra der terminalen Ereignisse besonders einfach.

Theorem 9.15 (Kolmogorov'sches 0-1-Gesetz). *Sei $\mathcal{F}_1, \mathcal{F}_2, \dots \subseteq \mathcal{F}$ eine Folge unabhängiger σ -Algebren. Dann ist $\mathcal{T} := \mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots)$ \mathbf{P} -trivial.*

Beweis. Sei $\mathcal{T}_n := \sigma\left(\bigcup_{m>n} \mathcal{F}_m\right)$, $n = 1, 2, \dots$. Nach Korollar 9.12 sind $(\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{T}_n)$ unabhängig, $n = 1, 2, \dots$. Damit sind auch $(\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{T})$ unabhängig, $n = 1, 2, \dots$ und damit auch $(\mathcal{T}, \mathcal{F}_1, \mathcal{F}_2, \dots)$. Wieder mit Korollar 9.12 folgt, dass $(\mathcal{T}_0, \mathcal{T})$ unabhängig sind und, da $\mathcal{T} \subseteq \mathcal{T}_0$ auch, dass \mathcal{T} von sich selbst unabhängig ist. Deswegen folgt die Behauptung aus Lemma 9.14. \square

9.3 Summen unabhängiger Zufallsvariable

Viele wichtige Sätze der Wahrscheinlichkeitstheorie beschäftigen sich mit unabhängigen Zufallsvariablen. In dieser Vorlesung sind dies vor allem das starke Gesetz der großen Zahlen (Theorem 9.21) und der zentrale Grenzwertsatz (Theorem 11.8). Bereits hier geben wir wichtige Hilfsmittel zur Analyse von Summen unabhängiger Zufallsvariablen an. Das erste ist der Zusammenhang mit der Faltung von Wahrscheinlichkeitsmaßen (siehe Abschnitt 6.4).

Proposition 9.16 (Faltung ist Verteilung der Summe). *Seien X_1, \dots, X_n unabhängige, reellwertige Zufallsvariablen. Dann gilt*

$$(X_1 + \dots + X_n)_* \mathbf{P} = (X_1)_* \mathbf{P} * \dots * (X_n)_* \mathbf{P}.$$

Weiter gilt für die charakteristischen Funktionen

$$\psi_{X_1 + \dots + X_n} = \psi_{X_1} \cdots \psi_{X_n}$$

und, falls X_1, \dots, X_n Werte in \mathbb{R}_+ annehmen,

$$\mathcal{L}_{X_1 + \dots + X_n} = \mathcal{L}_{X_1} \cdots \mathcal{L}_{X_n}.$$

Beweis. Zunächst ist nach Proposition 9.2 $((X_1, \dots, X_n))_* \mathbf{P} = (X_1)_* \mathbf{P} \otimes \dots \otimes (X_n)_* \mathbf{P}$. Damit folgt die erste Behauptung bereits aus der Definition 6.17 der Faltung von Maßen. Die weiteren Behauptungen folgen aus Proposition 9.5, da etwa

$$\begin{aligned} \psi_{X_1 + \dots + X_n}(t) &= \mathbf{E}[e^{it(X_1 + \dots + X_n)}] = \mathbf{E}[e^{itX_1} \cdots e^{itX_n}] \\ &= \mathbf{E}[e^{itX_1}] \cdots \mathbf{E}[e^{itX_n}] = \psi_{X_1}(t) \cdots \psi_{X_n}(t). \end{aligned} \quad \square$$

Das Kolmogorov'sche 0-1-Gesetz stellt recht einfach eine Aussage zur Verfügung, wann Summen unabhängiger Zufallsvariable fast sicher konvergieren.

Proposition 9.17 (Konvergenz von Summen unabhängiger Zufallsvariablen). *Seien X_1, X_2, \dots unabhängige Zufallsvariablen und $S_n := X_1 + \dots + X_n$.*

1. Es gilt

$$\mathbf{P}(\omega : S_n(\omega) \text{ konvergiert für } n \rightarrow \infty) \in \{0, 1\}$$

2. Weiter ist

$$\mathbf{P}(\omega : S_n(\omega)/n \text{ konvergiert für } n \rightarrow \infty) \in \{0, 1\}.$$

Falls $\mathbf{P}(S_n/n \text{ konvergiert}) = 1$, ist der Grenzwert fast sicher konstant.

Beweis. Setze $\mathcal{F}_i := \sigma(X_i)$, $i = 1, 2, \dots$. Damit ist die Familie $(\mathcal{F}_i)_{i=1,2,\dots}$ unabhängig. Die Menge $\{\omega : S_n(\omega) \text{ konvergiert für } n \rightarrow \infty\}$ ist messbar bezüglich $\mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots)$ und damit folgt die erste Aussage aus Theorem 9.15. Genauso folgt, dass $\mathbf{P}(S_n/n \text{ konvergiert}) \in \{0, 1\}$. Sei $S = \lim_{n \rightarrow \infty} S_n(n)/n$. Damit gilt für alle $m = 1, 2, \dots$

$$S = \lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \lim_{n \rightarrow \infty} \frac{X_m + \dots + X_n}{n},$$

also ist S nach $\sigma\left(\bigcup_{k \geq m} \mathcal{F}_k\right)$ messbar. Damit ist S auch \mathcal{T} -messbar und damit fast sicher konstant nach Theorem 9.15 und Lemma 9.14. \square

Proposition 9.18 (Maximal-Ungleichung von Kolmogorov). Seien $X_1, X_2, \dots \in \mathcal{L}^2$ unabhängige Zufallsvariablen. Dann gilt für $K > 0$

$$\mathbf{P}\left(\sup_{n \in \mathbb{N}} \left| \sum_{k=1}^n X_k - \mathbf{E}[X_k] \right| > K\right) \leq \frac{\sum_{n=1}^{\infty} \mathbf{V}(X_n)}{K^2}.$$

Beweis. O.E. sei $\mathbf{E}[X_k] = 0$, $k = 1, 2, \dots$. Weiter setzen wir $S_n = X_1 + \dots + X_n$ und $T := \inf\{n : |S_n| > K\}$. Dann gilt $\mathbf{P}(\sup_n |S_n| > K) = \mathbf{P}(T < \infty)$. Wegen Korollar 9.12 sind $S_k \cdot 1_{T=k}$ und $S_n - S_k$ unabhängig für $k \leq n$. Deshalb gilt

$$\begin{aligned} \sum_{k=1}^n \mathbf{E}[X_k^2] &= \mathbf{E}[S_n^2] \geq \sum_{k=1}^n \mathbf{E}[S_n^2, T = k] \\ &= \sum_{k=1}^n \mathbf{E}[S_k^2 + (S_n - S_k + 2S_k)(S_n - S_k), T = k] \\ &\geq \sum_{k=1}^n \mathbf{E}[S_k^2, T = k] + 2\mathbf{E}[S_k(S_n - S_k), T = k] \\ &= \sum_{k=1}^n \mathbf{E}[S_k^2, T = k] \geq K^2 \mathbf{P}(T \leq n) \end{aligned}$$

Nun folgt die Behauptung mit $n \rightarrow \infty$. \square

Theorem 9.19 (Konvergenzkriterium für Reihen). Seien $X_1, X_2, \dots \in \mathcal{L}^2$ unabhängige Zufallsvariablen mit $\sum_{n=1}^{\infty} \mathbf{V}[X_n] < \infty$. Dann konvergiert $\sum_{k=1}^n X_k - \mathbf{E}[X_k]$ fast sicher.

Beweis. Wieder sei o.E. $\mathbf{E}[X_k] = 0$, $k = 1, 2, \dots$ und wir schreiben $S_n = X_1 + \dots + X_n$. Für $\varepsilon > 0$ gilt nach Proposition 9.18

$$\lim_{k \rightarrow \infty} \mathbf{P}(\sup_{n \geq k} |S_n - S_k| > \varepsilon) \leq \lim_{k \rightarrow \infty} \frac{\sum_{n=k+1}^{\infty} \mathbf{E}[X_n^2]}{\varepsilon^2} = 0.$$

Deswegen konvergiert $\sup_{n \geq k} |S_n - S_k| \xrightarrow{k \rightarrow \infty} 0$. Nach Proposition 8.6 gibt es also eine Teilfolge k_1, k_2, \dots mit $\sup_{n \geq k_i} |S_n - S_{k_i}| \xrightarrow{i \rightarrow \infty} 0$. Da aber $(\sup_{n \geq k} |S_n - S_k|)_{k=1,2,\dots}$ fallend ist, gilt $\sup_{n \geq k} |S_n - S_k| \xrightarrow{k \rightarrow \infty} 0$. Das bedeutet aber, dass $(S_n)_{n=1,2,\dots}$ konvergiert. \square

9.4 Das starke Gesetz der großen Zahlen

Bereits in der Vorlesung *Stochastik* haben wir das schwache Gesetz der großen Zahlen bewiesen: sind $X_1, X_2, \dots \in \mathcal{L}^2$ identisch verteilt und unkorreliert, dann ist für $\varepsilon > 0$

$$\mathbf{P}\left(\frac{1}{n}\left|\sum_{k=1}^n(X_k - \mathbf{E}[X_k])\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \mathbf{V}\left[\frac{1}{n}\sum_{k=1}^n X_k\right] = \frac{1}{\varepsilon^2 n^2} \sum_{k=1}^n \mathbf{V}[X_k] = \frac{\mathbf{V}[X_1]}{\varepsilon^2 n} \xrightarrow{n \rightarrow \infty} 0.$$

Diese Aussage wollen wir nun in zwei Richtungen verschärfen. Wir wollen einerseits die stochastische Konvergenz durch die fast sichere Konvergenz ersetzen, und außerdem nur die Existenz erster Momente (nicht jedoch die Existenz zweiter Momente) fordern. Zunächst jedoch definieren wir, was wir genau meinen, wenn wir sagen, dass eine Folge von Zufallsvariablen einem Gesetz großer Zahlen folgt.

Definition 9.20 (Gesetz der großen Zahlen). Sei $X_1, X_2, \dots \in \mathcal{L}^1$ eine Folge reellwertiger Zufallsvariablen. Wir sagen, dass die Folge dem schwachen Gesetz der großen Zahlen folgt, falls

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mathbf{E}[X_k]) \xrightarrow{n \rightarrow \infty}_p 0.$$

Die Folge genügt dem starken Gesetz der großen Zahlen, falls

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mathbf{E}[X_k]) \xrightarrow{n \rightarrow \infty}_{fs} 0.$$

Theorem 9.21 (Starkes Gesetz für unabhängige Zufallsvariablen). Eine Folge $X_1, X_2, \dots \in \mathcal{L}^1$ unabhängiger und identisch verteilter Zufallsvariablen genügt dem starken Gesetz der großen Zahlen, d.h. es gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{n \rightarrow \infty}_{fs} \mathbf{E}[X_1].$$

Bemerkung 9.22 (Schwachtes Gesetz). Da aus der fast sicheren Konvergenz die stochastische Konvergenz folgt (siehe Proposition 8.6), genügt die Folge X_1, X_2, \dots aus dem Theorem auch dem schwachen Gesetz der großen Zahlen. Weiter genügt auch die Folge X_1^+, X_2^+, \dots dem starken Gesetz und $\mathbf{E}[\frac{1}{n}(X_1^+ + \dots + X_n^+)] = \mathbf{E}[X_1^+]$. Damit ist die Folge $(\frac{1}{n}(X_1^+ + \dots + X_n^+))_{n=1,2,\dots}$ nach Korollar 8.12 gleichgradig integrierbar. Genauso ist die Folge der Partialsummen der Negativteile gleichgradig integrierbar. Es folgt aus Theorem 8.11, dass $\frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{n \rightarrow \infty}_{\mathcal{L}^1} \mathbf{E}[X_1]$.

Bemerkung 9.23 (Endliche vierte und zweite Momente). Die Schwierigkeit im Beweis des starken Gesetzes ist, dass nur verwendet werden darf, dass $X_1 \in \mathcal{L}^1$. Wesentlich einfacher wird der Beweis, wenn man $X_1 \in \mathcal{L}^4$ bzw. $X_1 \in \mathcal{L}^2$ voraussetzt. Diese beiden Beweise geben wir zunächst an. Es sei $S_n := X_1 + \dots + X_n$.

1. Der Fall $X_1 \in \mathcal{L}^4$: Hier kommt man ganz ohne weitere Hilfsmittel aus:

Aus der Linearität des Erwartungswertes ist klar, dass $\mathbf{E}[S_n/n] = \mathbf{E}[X_1]$. O.E. sei $\mathbf{E}[X_1] = 0$, ansonsten geht man zu den Zufallsvariablen $X_1 - \mathbf{E}[X_1], X_2 - \mathbf{E}[X_2], \dots \in \mathcal{L}^4$

über. Zunächst berechnen wir mit Hilfe der Unabhängigkeit von $(X_k)_{k=1,2,\dots}$

$$\mathbf{E}[S_n^4] = \sum_{k=1}^n \mathbf{E}[X_k^4] + 3 \sum_{\substack{k,l=1 \\ k \neq l}}^{\infty} \mathbf{E}[X_k^2 X_l^2] \leq (n + 6n^2) \mathbf{E}[X_1^4]$$

wegen der Cauchy-Schwartz-Ungleichung. Daraus folgt

$$\mathbf{E} \left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n} \right)^4 \right] \leq \sum_{n=1}^{\infty} \frac{n + 6n^2}{n^4} \mathbf{E}[X_1^4] < \infty.$$

Deswegen gilt $\sum_{n=1}^{\infty} \left(\frac{S_n}{n} \right)^4 < \infty$ fast sicher, insbesondere also $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$.

2. Der Fall $X_1 \in \mathcal{L}^2$: Hier ist das Konvergenzkriterium für Reihen, Theorem 9.19 von entscheidender Hilfe. Außerdem benötigen wir noch folgendes Resultat:

Lemma 9.24 (Kronecker Lemma). *Seien $x_1, x_2, \dots \in \mathbb{R}$, $y_1, y_2, \dots \in \mathbb{R}$ monoton mit $y_n \uparrow \infty$ und $\sum_{n=1}^{\infty} x_n/y_n < \infty$. Dann gilt $\sum_{k=1}^n x_k/y_n \xrightarrow[n \rightarrow \infty]{} 0$.*

Beweis. Sei $z_0 = 0$, $z_n := \sum_{k=1}^n x_k/y_k$. Dann gilt $z_n \xrightarrow[n \rightarrow \infty]{} z_{\infty} < \infty$ und $x_k = y_k(z_k - z_{k-1})$. Wir schreiben mit $y_0 = 0$

$$\begin{aligned} \frac{\sum_{k=1}^n x_k}{y_n} &= \frac{1}{y_n} \sum_{k=1}^n y_k(z_k - z_{k-1}) = z_n + \frac{1}{y_n} \left(\sum_{k=0}^{n-1} y_k z_k - \sum_{k=1}^n y_k z_{k-1} \right) \\ &= z_n - \frac{1}{y_n} \left(\sum_{k=1}^n y_k z_{k-1} - y_{k-1} z_{k-1} \right) \\ &\xrightarrow[n \rightarrow \infty]{} z_{\infty} - z_{\infty} \cdot \lim_{n \rightarrow \infty} \frac{1}{y_n} \sum_{k=1}^n y_k - y_{k-1} = 0. \end{aligned}$$

□

Zurück zum Beweis des starken Gesetzes im Fall $X_1 \in \mathcal{L}^2$. O.E. sei wieder $\mathbf{E}[X_1] = 0$. Betrachte die Folge $X_1/1, X_2/2, \dots$. Wegen $\sum_{n=1}^{\infty} \mathbf{V}[X_n/n] = \mathbf{V}[X_1] \sum_{n=1}^{\infty} 1/n^2$ gilt nach Theorem 9.19, dass $\sum_{k=1}^n X_k/k$ fast sicher konvergiert. Mit Lemma 9.24 folgt, dass $S_n/n \xrightarrow[n \rightarrow \infty]{f.s.} 0$.

Beweis von Theorem 9.21 falls $X_1 \in \mathcal{L}^1$. Es genügt, den Fall von nicht-negativen Zufallsvariablen zu betrachten. Im allgemeinen Fall sei bemerkt, dass $X_1^+, X_2^+, \dots \in \mathcal{L}^1$ und $X_1^-, X_2^-, \dots \in \mathcal{L}^1$ die Voraussetzungen des Satzes erfüllen, und aus $(X_1^+ + \dots + X_n^+)/n \xrightarrow[n \rightarrow \infty]{f.s.} \mathbf{E}[X_1^+]$ und $(X_1^- + \dots + X_n^-)/n \xrightarrow[n \rightarrow \infty]{f.s.} \mathbf{E}[X_1^-]$ die Aussage wegen Linearität des Erwartungswertes folgt.

Für $S_n = X_1 + \dots + X_n$ werden wir zeigen, dass

$$\mathbf{E}[\limsup_{n \rightarrow \infty} S_n/n] \leq \mathbf{E}[X_1]. \quad (9.2)$$

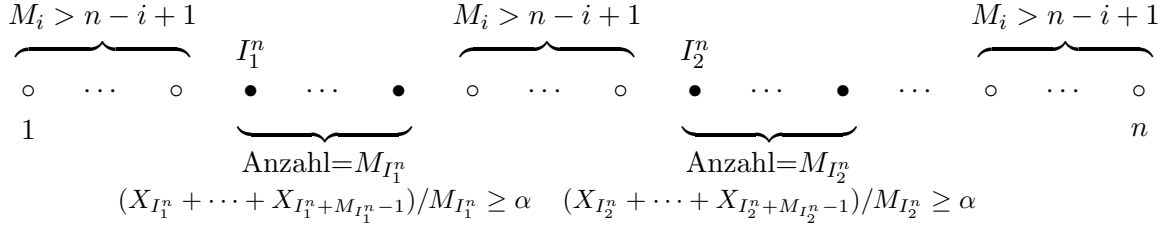


Abbildung 9.1: Illustration von M_i, I_j^n , eingeführt unterhalb von (9.3). Die Größe L_n ist die Anzahl der zusammenhängenden Bereiche von \bullet 's.

Gilt dies, so folgt erstens

$$\begin{aligned} \mathbf{E}[\liminf_{n \rightarrow \infty} S_n/n] &\geq \mathbf{E}[\liminf_{n \rightarrow \infty} (X_1 \wedge k + \dots + X_n \wedge k)/n] \\ &= k - \mathbf{E}[\limsup_{n \rightarrow \infty} ((k - X_1)^+ + \dots + (k - X_n)^+)/n] \\ &\geq \mathbf{E}[k - (k - X_1)^+] \xrightarrow{k \rightarrow \infty} \mathbf{E}[X_1] \end{aligned}$$

Zweitens ist dann $\mathbf{E}[\limsup_{n \rightarrow \infty} S_n/n - \liminf_{n \rightarrow \infty} S_n/n] = 0$, also $\limsup_{n \rightarrow \infty} S_n/n = \liminf_{n \rightarrow \infty} S_n/n = 0$ fast sicher, da sowohl $\liminf_{n \rightarrow \infty} S_n/n$ als auch $\limsup_{n \rightarrow \infty} S_n/n$ terminale Funktionen, und damit nach Theorem 9.15 und Lemma 9.14 fast sicher konstant sind. Außerdem ist damit

$$\liminf_{n \rightarrow \infty} S_n/n = \mathbf{E}[\liminf_{n \rightarrow \infty} S_n/n] \geq \mathbf{E}[X_1] \geq \mathbf{E}[\limsup_{n \rightarrow \infty} S_n/n] = \limsup_{n \rightarrow \infty} S_n/n,$$

woraus die Behauptung folgt.

Es bleibt also (9.2) zu zeigen. O.E. sei $\mathbf{E}[X_1] > 0$, ansonsten ist $X_k = 0$ fast sicher, $k = 1, 2, \dots$ und die Aussage ist trivial. Hierzu werden wir

$$0 < \alpha < \mathbf{E}[\limsup_{n \rightarrow \infty} S_n/n] \implies \alpha \leq \mathbf{E}[X_1] \quad (9.3)$$

beweisen. Nach Voraussetzung ist für $i = 0, 1, 2, \dots$

$$\alpha < \mathbf{E}[\limsup_{n \rightarrow \infty} S_n/n] = \limsup_{n \rightarrow \infty} S_n/n = \limsup_{n \rightarrow \infty} (X_{i+1} + \dots + X_{i+n})/n.$$

Damit ist

$$M_i := \inf\{n \in \mathbb{N} : (X_i + \dots + X_{i+n-1})/n \geq \alpha\}$$

fast sicher endlich, $i = 1, 2, \dots$. Die M_i 's sind identisch verteilt. Wir definieren für $n = 1, 2, \dots$ rekursiv (siehe auch Abbildung 9.1) $I_1^n = 0$ sowie für $j = 0, 1, 2, \dots$ (mit $M_0 := 0$)

$$I_{j+1}^n := \inf\{i \in \mathbb{N} : i \geq I_j^n + M_j^n, M_i \leq n - i + 1\}$$

mit $\inf \emptyset = \infty$ und $L_n := \sup\{n \in \mathbb{N}_0 : I_j^n < \infty\}$. Damit gilt für $1 \leq j \leq L_n$, dass

$I_j^n + M_{I_j^n} \leq n$, also $(X_{I_j^n} + \cdots + X_{I_j^n + M_{I_j^n} - 1})/M_{I_j^n} \geq \alpha$. Dies verwenden wir nun mittels

$$\begin{aligned} \mathbf{E}[X_1] &= \mathbf{E}[(X_1 + \cdots + X_n)/n] \\ &\geq \frac{1}{n} \mathbf{E} \left[\sum_{j=1}^{L_n} M_{I_j^n} \cdot (X_{I_j^n} + \cdots + X_{I_j^n + M_{I_j^n} - 1})/M_{I_j^n} \right] \\ &\geq \frac{\alpha}{n} \mathbf{E} \left[\sum_{j=1}^{L_n} M_{I_j^n} \right] = \alpha - \frac{\alpha}{n} \mathbf{E} \left[n - \sum_{j=1}^{L_n} M_{I_j^n} \right] \\ &\geq \alpha - \frac{\alpha}{n} \mathbf{E} \left[\sum_{i=1}^n 1_{M_i > n-i+1} \right] \\ &= \alpha \left(1 - \frac{1}{n} \sum_{i=1}^n \mathbf{P}(M_i > i) \right) \xrightarrow{n \rightarrow \infty} \alpha, \end{aligned}$$

da $(\frac{1}{n} \sum_{i=1}^n \mathbf{P}(M_i > i))_{n=1,2,\dots}$ als Cesàro-Limes von $(\mathbf{P}(M_i > i))_{i=1,2,\dots}$ wegen der Identität der Verteilungen der M_i 's gegen 0 konvergiert. Damit ist (9.3) gezeigt und die Behauptung ist bewiesen. \square

Wir geben nun eine einfache Anwendung des starken Gesetzes an. Gerade in der Statistik kommt es oft vor, dass man eine große Anzahl unabhängiger, identisch verteilter, reellwertiger Zufallsgrößen betrachtet. Der Satz von Glivenko-Cantelli (Theorem 9.26) besagt, dass die empirische Verteilung der Zufallsvariablen fast sicher gegen die zu Grunde liegende Verteilung konvergiert.

Definition 9.25 (Empirische Verteilung). Seien X_1, X_2, \dots Zufallsvariable. Für $n = 1, 2, \dots$ heißt die (zufällige) Wahrscheinlichkeitsverteilung

$$\hat{\mu}_n := \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$$

die empirische Verteilung von X_1, \dots, X_n . Sind die Zufallsvariablen reellwertig, ist außerdem

$$\hat{F}_n(x) := \frac{1}{n} \sum_{k=1}^n 1_{X_k \leq x}$$

die empirische Verteilungsfunktion von X_1, \dots, X_n .

Theorem 9.26 (Satz von Glivenko-Cantelli). Seien X_1, X_2, \dots unabhängige, reellwertige Zufallsvariablen mit identischer Verteilung mit Verteilungsfunktion F . Dann gilt

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{f_s} 0.$$

Beweis. Für $x \in \mathbb{R}$ und $n = 1, 2, \dots$ sei $Y_n(x) := 1_{X_n \leq x}$ und $Z_n(x) := 1_{X_n < x}$. Nach Theorem 9.21 gilt für jedes $x \in \mathbb{R}$

$$\begin{aligned} \hat{F}_n(x) &= \frac{1}{n} \sum_{k=1}^n Y_k(x) \xrightarrow[n \rightarrow \infty]{f_s} \mathbf{E}[Y_1(x)] = \mathbf{P}(X_1 \leq x) = F(x), \\ \hat{F}_n(x-) &= \frac{1}{n} \sum_{k=1}^n Z_k(x) \xrightarrow[n \rightarrow \infty]{f_s} \mathbf{E}[Z_1(x)] = \mathbf{P}(X_1 < x) = F(x-). \end{aligned}$$

Wir müssen zeigen, dass diese Konvergenzen auch uniform für alle $x \in \mathbb{R}$ gelten. Für $N = 1, 2, \dots$ und $j = 0, \dots, N$ setzen wir

$$x_j^N := \inf\{x \in \mathbb{R} : F(x) \geq j/N\}$$

und

$$R_n^N := \max_{j=1, \dots, N-1} (|\widehat{F}_n(x_j^N) - F(x_j^N)| + |\widehat{F}_n(x_j^N -) - F(x_j^N -)|).$$

Für $N = 1, 2, \dots$ gilt also $R_n^N \xrightarrow{n \rightarrow \infty}_{fs} 0$. Außerdem ist für $x \in (x_{j-1}^N, x_j^N)$

$$\begin{aligned} \widehat{F}_n(x) &\leq \widehat{F}_n(x_j^N) \leq \widehat{F}_n(x_j^N) + R_n^N \leq F(x) + R_n^N + \frac{1}{N}, \\ \widehat{F}_n(x) &\geq \widehat{F}_n(x_{j-1}^N) \geq F(x_{j-1}^N) - R_n^N \geq F(x) - R_n^N - \frac{1}{N}, \end{aligned}$$

also, für jedes $N = 1, 2, \dots$

$$\sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F(x)| \leq \frac{1}{N} + R_n^N \xrightarrow{n \rightarrow \infty}_{fs} \frac{1}{N}.$$

Da die linke Seite nicht von N abhängt, folgt die Behauptung mit $N \rightarrow \infty$. \square

10 Schwache Konvergenz

Für Maß- und Messräume haben wir bisher oft verwendet, dass die σ -Algebra die Borel'sche ist, d.h. die σ -Algebra, die von einer Topologie erzeugt wird. In diesem Abschnitt werden wir oft voraussetzen, dass der topologische Raum polnisch ist, also separabel und metrisierbar durch eine vollständige Metrik. Um uns Schreibarbeit zu ersparen, setzen wir nun gleich voraus, dass (E, r) ein metrischer Raum ist und manchmal werden wir voraussetzen, dass er vollständig und separabel ist.

Für eine messbare Abbildung $f : E \rightarrow \mathbb{R}$ und ein Maß μ auf $\mathcal{B}(E)$ (der Borel'schen σ -Algebra von E) werden wir in diesem und im nächsten Kapitel durchgehend $\mu[f] := \int f d\mu$ schreiben.

10.1 Definition und einfache Eigenschaften

Bisher haben wir uns mit verschiedenen Konvergenzarten von Zufallsvariablen beschäftigt. Die Konvergenz in Verteilung von Zufallsvariablen ist dasselbe wie die schwache Konvergenz der Verteilungen der Zufallsvariablen. Zur Motivation hinter den folgenden Definitionen sei an ein Faktum erinnert: in einem metrischen Raum (E, r) ist $x_n \xrightarrow{n \rightarrow \infty} x$ genau dann wenn $f(x_n) \xrightarrow{n \rightarrow \infty} f(x)$ für alle stetigen Funktionen auf E (d.h. $f \in \mathcal{C}(E, \mathbb{R})$).

Definition 10.1 (Schwache Konvergenz und Konvergenz in Verteilung).

1. Wir bezeichnen mit $\mathcal{P}(E)$ die Menge der Wahrscheinlichkeitsmaße auf $\mathcal{B}(E)$ und mit $\mathcal{P}_{\leq 1}(E)$ die Menge der endlichen Maße μ auf $\mathcal{B}(E)$ mit $\mu(E) \leq 1$. Weiter ist $\mathcal{C}_b(E)$ die Menge der reellwertigen, beschränkten, stetigen Funktionen auf E und $\mathcal{C}_c(E) \subseteq \mathcal{C}_b(E)$ die Menge der reellwertigen, beschränkten stetigen Funktionen auf E mit kompaktem Träger.

2. Eine Folge $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$ konvergiert schwach gegen $\mathbf{P} \in \mathcal{P}(E)$, falls

$$\mathbf{P}_n[f] \xrightarrow{n \rightarrow \infty} \mathbf{P}[f] \quad (10.1)$$

für alle $f \in \mathcal{C}_b(E)$. Wir schreiben dann

$$\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}.$$

3. Ist $\mu_1, \mu_2, \dots \in \mathcal{P}_{\leq 1}$ und μ ein Maß auf E . Gilt (10.1) nur für alle $f \in \mathcal{C}_c(E)$, so sagen wir, μ_n konvergiert vage gegen μ . Wir schreiben dann

$$\mu_n \xrightarrow{n \rightarrow \infty}_v \mu.$$

4. Seien X, X_1, X_2, \dots Zufallsvariablen auf Wahrscheinlichkeitsräumen $(\Omega, \mathcal{A}, \mathbf{P})$, $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1)$, $(\Omega_2, \mathcal{A}_2, \mathbf{P}_2), \dots$ mit Werten in E . Dann konvergiert X_1, X_2, \dots in Verteilung gegen X , falls $(X_n)_* \mathbf{P}_n \xrightarrow{n \rightarrow \infty} X_* \mathbf{P}$. Wir schreiben dann

$$X_n \xrightarrow{n \rightarrow \infty} X.$$

Bemerkung 10.2. 1. Man beachte, dass für Zufallsvariablen X, X_1, X_2, \dots mit Werten in E genau dann $X_n \xrightarrow{n \rightarrow \infty} X$ gilt, wenn

$$\mathbf{P}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbf{P}[f(X)]$$

für alle $f \in \mathcal{C}_b(E)$ gilt. Viele der folgenden Resultate lassen sich deswegen auf zwei Arten und Weisen formulieren: entweder mittels Wahrscheinlichkeitsverteilungen, oder mittels Zufallsvariablen. Der Zusammenhang ist dabei immer, dass die Aussage über die Wahrscheinlichkeitsverteilungen ebenfalls eine Aussage über die Verteilungen der Zufallsvariablen ist.

2. Der schwache Limes von Wahrscheinlichkeitsmaßen muss wieder ein Wahrscheinlichkeitsmaß sein, da $1 \in \mathcal{C}_b(E)$. Der vage Limes von Wahrscheinlichkeitsmaßen muss jedoch nicht unbedingt wieder ein Wahrscheinlichkeitsmaß sein, da $1 \notin \mathcal{C}_c(E)$, falls E nicht kompakt ist; siehe auch Beispiel 10.3.1. Immerhin sind vage Grenzwerte in $\mathcal{P}_{\leq 1}(E)$, wie Lemma 10.12.

3. Wir kennen bereits die fast sicher Konvergenz, die stochastische und die Konvergenz in \mathcal{L}^p von Zufallsvariablen X_1, X_2, \dots gegen X . Entscheidender Unterschied zur Konvergenz in Verteilung ist, dass letztere nicht voraussetzt, dass die Zufallsvariablen auf demselben Wahrscheinlichkeitsraum definiert sind.

4. Nach Definition 10.1 ist die Topologie der schwachen Konvergenz auf $\mathcal{P}(E)$ die schwächste (d.h. die kleinste) Topologie, für die $\mathbf{P} \mapsto \mathbf{P}[f]$ für alle $f \in \mathcal{C}_b(E)$ stetig ist.

Beispiel 10.3. 1. Sei $x, x_1, x_2, \dots \in \mathbb{R}$ mit $x_n \xrightarrow{n \rightarrow \infty} x$ sowie $\mathbf{P} = \delta_x, \mathbf{P}_1 = \delta_{x_1}, \mathbf{P}_2 = \delta_{x_2}, \dots$. Dann gilt $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}$, da

$$\mathbf{P}_n[f] = f(x_n) \xrightarrow{n \rightarrow \infty} f(x) = \mathbf{P}[f]$$

für alle $f \in \mathcal{C}_b(\mathbb{R})$.

Falls die Folge x_1, x_2, \dots divergiert, etwa $x_n = n$, so gilt $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \nu$ 0 (das ist das 0-Maß auf $\mathcal{B}(\mathbb{R})$), da

$$\mathbf{P}_n[f] = f(x_n) \xrightarrow{n \rightarrow \infty} 0 = 0[f]$$

für alle $f \in \mathcal{C}_c(\mathbb{R})$ gilt. Allerdings gilt die schwache Konvergenz nicht, da $\mathbf{P}_n[1] = 1 \neq 0 = 0[1]$.

2. Seien X, X_1, X_2, \dots identisch verteilt. Dann ist $X_n \xrightarrow{n \rightarrow \infty} X$, jedoch gilt die Konvergenz im allgemeinen weder fast sicher, noch stochastisch oder in \mathcal{L}^p .
3. Wie wir noch sehen werden, ist der zentrale Grenzwertsatz, den wir in Kapitel 11 kennenlernen werden (Theorem 11.8), ein Resultat über Konvergenz in Verteilung. In seiner einfachsten Form, dem Satz von deMoivre-Laplace (siehe auch Bemerkung 10.8 und Beispiel 10.34), besagt dieser: sei $p \in (0, 1)$, $X_n \sim B(n, p)$, $n = 1, 2, \dots$ und $X \sim N(0, 1)$. Dann gilt

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{n \rightarrow \infty} X.$$

4. Ebenso ist die Poisson-Approximation der Binomialverteilung eine Aussage über Konvergenz in Verteilung (siehe auch Theorem 11.5): sei $X_n \sim B(n, p_n)$, $n = 1, 2, \dots$ mit $n \cdot p_n \xrightarrow{n \rightarrow \infty} \lambda$ und $X \sim \text{Poi}(\lambda)$. Dann gilt

$$X_n \xrightarrow{n \rightarrow \infty} X.$$

Lemma 10.4 (Eindeutigkeit des schwachen Limes). Seien $\mu, \nu, \mu_1, \mu_2, \dots \in \mathcal{P}(E)$ mit $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ und $\mu_n \xrightarrow{n \rightarrow \infty} \nu$. Dann ist $\mu = \nu$.

Beweis. Nach Proposition 3.10 genügt es zu zeigen, dass $\mu(A) = \nu(A)$ für alle abgeschlossenen $A \subseteq E$. Die Menge aller abgeschlossenen Mengen ist nämlich ein schnittstabiler Erzeuger von $\mathcal{B}(E)$. Sei also $A \subseteq E$ abgeschlossen. Wir setzen

$$r(x, A) := \inf_{y \in A} r(x, y)$$

sowie

$$f_m(x) \mapsto (1 - m \cdot r(x, A))^+.$$

für $m = 1, 2, \dots$. Dann ist $f_m \xrightarrow{m \rightarrow \infty} 1_A$, da A abgeschlossen ist. Weiter gilt mit majorisierter Konvergenz

$$\mu(A) = \lim_{m \rightarrow \infty} \mu[f_m] = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \mu_n[f_m] = \lim_{m \rightarrow \infty} \nu[f_m] = \nu(A)$$

und die Behauptung folgt. □

Wir erinnern an die Eingangsgrafik von Kapitel 8. Eine Folge von Zufallsvariablen kann fast sicher, in Wahrscheinlichkeit, in \mathcal{L}^p oder in Verteilung konvergieren. Die Konvergenz in Verteilung ist der schwächste dieser Begriffe in folgendem Sinne.

Proposition 10.5 (Konvergenz in Wahrscheinlichkeit und in Verteilung). Seien X, X_1, X_2, \dots Zufallsvariable mit Werten in E . Falls $X_n \xrightarrow{n \rightarrow \infty} X$ in Wahrscheinlichkeit, so auch $X_n \xrightarrow{n \rightarrow \infty} X$ in Verteilung. Ist X konstant, so gilt auch die Umkehrung.

Beweis. Sei $X_n \xrightarrow{n \rightarrow \infty}_p X$. Angenommen, es gäbe ein $f \in \mathcal{C}_b(E)$ so, dass $\lim_{n \rightarrow \infty} \mathbf{P}[f(X_n)] \neq \mathbf{P}[f(X)]$. Dann gibt es eine Teilfolge $(n_k)_{k=1,2,\dots}$ und ein $\varepsilon > 0$ mit

$$\lim_{k \rightarrow \infty} |\mathbf{P}[f(X_{n_k})] - \mathbf{P}[f(X)]| > \varepsilon. \quad (10.2)$$

Wegen $X_n \xrightarrow{n \rightarrow \infty}_p X$ und Proposition 8.6 gibt es eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$, so dass $X_{n_{k_\ell}} \xrightarrow{\ell \rightarrow \infty} X$ fast sicher. Wegen majorisierter Konvergenz wäre damit auch

$$\lim_{\ell \rightarrow \infty} \mathbf{P}[f(X_{n_{k_\ell}})] = \mathbf{P}[f(X)]$$

im Widerspruch zu (10.2).

Für die Umkehrung sei $X = s \in E$. Es ist $x \mapsto r(x, s) \wedge 1$ eine beschränkte, stetige Funktion und damit

$$\mathbf{P}[r(X_n, s) \wedge 1] \xrightarrow{n \rightarrow \infty} \mathbf{P}[r(X, s) \wedge 1] = 0.$$

Also gilt $X_n \xrightarrow{n \rightarrow \infty}_p X$ wegen (8.1). \square

Theorem 10.6 (Portmanteau Theorem). *Seien X, X_1, X_2, \dots Zufallsvariablen mit Werten in E . Folgende Bedingungen sind äquivalent:*

(i) $X_n \xrightarrow{n \rightarrow \infty} X$

(ii) $\mathbf{P}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbf{P}[f(X)]$ für alle beschränkten, Lipschitz-stetigen Funktionen f .

(iii) $\liminf_{n \rightarrow \infty} \mathbf{P}(X_n \in G) \geq \mathbf{P}(X \in G)$ für alle offenen $G \subseteq E$.

(iv) $\limsup_{n \rightarrow \infty} \mathbf{P}(X_n \in F) \leq \mathbf{P}(X \in F)$ für alle abgeschlossenen $F \subseteq E$.

(v) $\lim_{n \rightarrow \infty} \mathbf{P}(X_n \in B) = \mathbf{P}(X \in B)$ für alle $B \in \mathcal{B}(E)$ mit¹⁵ $\mathbf{P}(X \in \partial B) = 0$.

Beweis. (i) \Rightarrow (ii): klar

(ii) \Rightarrow (iv) Sei $F \subseteq E$ abgeschlossen und f_1, f_2, \dots Lipschitz-stetig, so dass $f_k \downarrow 1_F$. (Beispielsweise wählt man $\varepsilon_k \downarrow 0$ und $f_k(x) = (1 - \frac{1}{\varepsilon_k} r(x, F))^+$, wobei $r(x, F) := \inf_{y \in F} r(x, y)$.) Damit gilt

$$\limsup_{n \rightarrow \infty} \mathbf{P}(X_n \in F) \leq \inf_{k=1,2,\dots} \limsup_{n \rightarrow \infty} \mathbf{P}[f_k(X_n)] = \inf_{k=1,2,\dots} \mathbf{P}[f_k(X)] = \mathbf{P}(X \in F).$$

(iii) \iff (iv) Das ist klar. Man muss in der Richtung (iii) \Rightarrow (iv) nur $F := E \setminus G$ und in der Richtung (iv) \Rightarrow (iii) analog $G := E \setminus F$ setzen.

(iii) \Rightarrow (i) Sei $f \geq 0$ stetig. Wegen Proposition 7.10 und Fatou's Lemma ist damit

$$\begin{aligned} \mathbf{P}[f(X)] &= \int_0^\infty \mathbf{P}(f(X) > t) dt \leq \int_0^\infty \liminf_{n \rightarrow \infty} \mathbf{P}(f(X_n) > t) dt \\ &\leq \liminf_{n \rightarrow \infty} \int_0^\infty \mathbf{P}(f(X_n) > t) dt = \liminf_{n \rightarrow \infty} \mathbf{P}[f(X_n)]. \end{aligned}$$

¹⁵Für den Abschluss \bar{B} und das Innere B° bezeichne hier $\partial B := \bar{B} \setminus B^\circ$ den Rand von B .

Für $-c < f < c$ gilt damit, da $-f + c \geq 0$ ist,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbf{P}[f(X_n)] &= c - \liminf_{n \rightarrow \infty} \mathbf{P}[-f(X_n) + c] \leq c - \mathbf{P}[-f(X) + c] = \mathbf{P}[f(X)] \\ &\leq \liminf_{n \rightarrow \infty} \mathbf{P}[f(X_n)], \end{aligned}$$

also $\mathbf{P}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbf{P}[f(X)]$.

(iii), (iv) \Rightarrow (v) Für $B \in \mathcal{B}(E)$ ist

$$\mathbf{P}(X \in B^\circ) \leq \liminf_{n \rightarrow \infty} \mathbf{P}(X_n \in B^\circ) \leq \limsup_{n \rightarrow \infty} \mathbf{P}(X_n \in \overline{B}) \leq \mathbf{P}(X \in \overline{B}).$$

Gegeben $\mathbf{P}(X \in \partial B) = \mathbf{P}(X \in \overline{B}) - \mathbf{P}(X \in B^\circ) = 0$, ist damit $\mathbf{P}(X_n \in B) \xrightarrow{n \rightarrow \infty} \mathbf{P}(X \in B)$.
(v) \Rightarrow (iv) Angenommen, (v) trifft zu und $F \subseteq E$ ist abgeschlossen. Wir schreiben $F^\varepsilon := \{x \in E : r(x, F) \leq \varepsilon\}$ für $\varepsilon > 0$. Die Mengen $\partial F^\varepsilon \subseteq \{x : r(x, F) = \varepsilon\}$ sind disjunkt, also gilt

$$\mathbf{P}(X \in \partial F^\varepsilon) = 0 \tag{10.3}$$

für Lebesgue-fast jedes ε . Sei $\varepsilon_1, \varepsilon_2, \dots$ eine Folge mit $\varepsilon_k \downarrow 0$, so dass (10.3) für alle $\varepsilon_1, \varepsilon_2, \dots$ gilt. Damit ist

$$\limsup_{n \rightarrow \infty} \mathbf{P}(X_n \in F) \leq \inf_{k=1,2,\dots} \limsup_{n \rightarrow \infty} \mathbf{P}(X_n \in F^{\varepsilon_k}) = \inf_{k=1,2,\dots} \mathbf{P}(X \in F^{\varepsilon_k}) = \mathbf{P}(X \in F).$$

□

Korollar 10.7 (Konvergenz von Verteilungsfunktionen). Seien $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ mit Verteilungsfunktionen F, F_1, F_2, \dots . Dann gilt $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}$ genau dann wenn $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$ für alle Stetigkeitsstellen x von F .

Beweis. ' \Rightarrow ': Ist x eine Stetigkeitsstelle von F , so ist $\mathbf{P}(\partial(-\infty; x]) = \mathbf{P}(\{x\}) = 0$. Damit gilt nach Theorem 10.6 (Richtung (i) \Rightarrow (v)), dass

$$F_n(x) = \mathbf{P}_n((-\infty; x]) \xrightarrow{n \rightarrow \infty} \mathbf{P}((-\infty; x]) = F(x).$$

' \Leftarrow ': Nach Theorem 10.6 (Richtung (ii) \Rightarrow (i)) genügt es zu zeigen, dass $\mathbf{P}_n[f] \xrightarrow{n \rightarrow \infty} \mathbf{P}[f]$ für alle beschränkten, Lipschitzfunktionen f . O.E. nehmen wir an, dass $|f| \leq 1$ und f Lipschitz-Konstante 1 hat. Für $\varepsilon > 0$ wähle $N \in \mathbb{N}$ und Stetigkeitspunkte $y_0 < \dots < y_N$ von F so, dass $F(y_0) < \varepsilon$, $F(y_N) > 1 - \varepsilon$ und $y_i - y_{i-1} < \varepsilon$ für $i = 1, \dots, N$. Dann ist $F_n(y_i) \xrightarrow{n \rightarrow \infty} F(y_i)$ und

$$f \leq 1_{(-\infty, y_0]} + 1_{(y_N, \infty)} + \sum_{i=1}^{N-1} (f(y_i) + \varepsilon) 1_{(y_i, y_{i+1}]},$$

also

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbf{P}_n[f] &\leq \limsup_{n \rightarrow \infty} F_n(y_0) + 1 - F_n(y_N) + \sum_{i=1}^N (f(y_i) + \varepsilon)(F_n(y_i) - F_n(y_{i-1})) \\ &\leq 3\varepsilon + \sum_{i=1}^N f(y_i)(F(y_i) - F(y_{i-1})) \leq 4\varepsilon + \mathbf{P}[f] \end{aligned}$$

Mit $\varepsilon \rightarrow 0$ und durch Ersetzen von f mit $1 - f$ folgt $\mathbf{P}_n[f] \xrightarrow{n \rightarrow \infty} \mathbf{P}[f]$. □

Bemerkung 10.8 (Der Satz von deMoivre-Laplace). In Beispiel 10.3 hatten wir behauptet, dass der Satz von deMoivre-Laplace eine Aussage über schwache Konvergenz macht. Man zeigt gewöhnlich, dass für $B(n, p)$ -verteilte Zufallsgrößen X_n gilt, dass

$$\mathbf{P}\left(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right) \xrightarrow{n \rightarrow \infty} \Phi(x),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist. Wie Korollar 10.7 zeigt, bedeutet das genau die Konvergenz in Verteilung gegen eine Standardnormalverteilung.

Korollar 10.9 (Satz von Slutsky). Seien $X, X_1, X_2, \dots, Y_1, Y_2, \dots$ Zufallsvariable mit Werten in E . Gilt $X_n \xrightarrow{n \rightarrow \infty} X$ und $r(X_n, Y_n) \xrightarrow{n \rightarrow \infty}_p 0$, so gilt auch $Y_n \xrightarrow{n \rightarrow \infty} X$.

Beweis. Siehe Übung. □

Theorem 10.10 (Continuous mapping theorem). Sei E separabel, (E', r') ein weiterer metrischer Raum und $\varphi : E \rightarrow E'$ messbar und $U_\varphi \subseteq E$ die Menge der Unstetigkeitsstellen von φ .

1. Sind $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$ und $\mathbf{P}(U_\varphi) = 0$ sowie $\mathbf{P}_i \xrightarrow{n \rightarrow \infty} \mathbf{P}$, dann ist auch $\varphi_* \mathbf{P}_n \xrightarrow{n \rightarrow \infty} \varphi_* \mathbf{P}$.
2. Sind X, X_1, X_2, \dots Zufallsvariable mit Werten in E und $\mathbf{P}(X \in U_\varphi) = 0$ und $X_n \xrightarrow{n \rightarrow \infty} X$, dann ist auch $\varphi(X_n) \xrightarrow{n \rightarrow \infty} \varphi(X)$.

Beweis. Zunächst bemerken wir, dass 2. eine Anwendung von 1. ist, falls man $\mathbf{P}_n = (X_n)_* \mathbf{P}$ setzt. Die Menge U_φ ist Borel-messbar, da

$$U_\varphi^{\delta, \varepsilon} := \{x \in E : \exists y, z \in B_\delta(x), r'(\varphi(y), \varphi(z)) > \varepsilon\}$$

Borel-messbar ist (hier geht die Separabilität von E ein) und

$$U_\varphi = \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} U_\varphi^{1/k, 1/n}.$$

Sei $G \subseteq E'$ offen und $x \in \varphi^{-1}(G) \cap U_\varphi^c$. Da φ in x also stetig ist, gibt es ein $\delta > 0$ mit $\varphi(y) \in G$ (d.h. $y \in \varphi^{-1}(G)$) für alle y mit $r(x, y) < \delta$. Also ist $\varphi^{-1}(G) \cap U_\varphi^c \subseteq (\varphi^{-1}(G))^\circ$. Daraus folgt mit Theorem 10.6 (Richtung (i) \Rightarrow (iii))

$$\begin{aligned} \varphi_* \mathbf{P}(G) &= \mathbf{P}(\varphi^{-1}(G)) = \mathbf{P}(\varphi^{-1}(G) \cap U_\varphi^c) \leq \mathbf{P}((\varphi^{-1}(G))^\circ) \\ &\leq \liminf_{n \rightarrow \infty} \mathbf{P}_n((\varphi^{-1}(G))^\circ) \leq \liminf_{n \rightarrow \infty} \mathbf{P}_n(\varphi^{-1}(G)) = \liminf_{n \rightarrow \infty} \varphi_* \mathbf{P}_n(G). \end{aligned}$$

Wieder wegen Theorem 10.6 (Richtung (iii) \Rightarrow (i)) folgt also $\varphi_* \mathbf{P}_n \xrightarrow{n \rightarrow \infty} \varphi_* \mathbf{P}$. □

Sieht man von der vagen Konvergenz einmal ab, so ist die Konvergenz in Verteilung der schwächste Konvergenzbegriff. Jedoch besteht ein Zusammenhang mit der fast sicheren Konvergenz, wie folgendes Theorem zeigt.

Theorem 10.11 (Schwache und fast sichere Konvergenz, Skorohod). Seien X, X_1, X_2, \dots Zufallsvariablen mit Werten in einem vollständigen und separablen Raum (E, r) . Dann gilt $X_n \xrightarrow{n \rightarrow \infty} X$ genau dann, wenn es einen Wahrscheinlichkeitsraum gibt, auf dem Zufallsgrößen Y, Y_1, Y_2, \dots definiert sind mit $Y_n \xrightarrow{n \rightarrow \infty}_{fs} Y$ und $Y \stackrel{d}{=} X, Y_1 \stackrel{d}{=} X_1, Y_2 \stackrel{d}{=} X_2, \dots$

Beweis. ' \Leftarrow ': klar, da aus der fast sicheren die schwache Konvergenz nach Proposition 10.5 folgt.

' \Rightarrow ': Als Wahrscheinlichkeitsraum erweitern wir den Raum, auf dem X definiert ist, setzen also insbesondere $Y = X$. Sei zunächst $E = \{1, \dots, m\}$ endlich. Sei U uniform auf $[0, 1]$ verteilt und unabhängig von Y , sowie W_1, W_2, \dots unabhängig mit

$$\mathbf{P}(W_n = k) = \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l=1}^m \mathbf{P}(X = l) \wedge \mathbf{P}(X_n = l)}.$$

Wir setzen $Y_n = k$ falls entweder

$$X = k \text{ und } U \leq \frac{\mathbf{P}(X_n = k)}{\mathbf{P}(X = k)}$$

oder

$$X = l \text{ und } U > \frac{\mathbf{P}(X_n = l)}{\mathbf{P}(X = l)} \text{ und } W_n = k.$$

Dann ist

$$\begin{aligned} \mathbf{P}(Y_n = k) &= \mathbf{P}(X = k) \cdot \frac{\mathbf{P}(X_n = k)}{\mathbf{P}(X = k)} \wedge 1 \\ &+ \sum_{l=1}^m \mathbf{P}(X = l) \cdot \left(1 - \frac{\mathbf{P}(X_n = l)}{\mathbf{P}(X = l)}\right) \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l'=1}^m \mathbf{P}(X = l') \wedge \mathbf{P}(X_n = l')} \\ &= \mathbf{P}(X_n = k) \wedge \mathbf{P}(X = k) \\ &+ \sum_{l=1}^m (\mathbf{P}(X = l) - \mathbf{P}(X_n = l) \wedge \mathbf{P}(X = l)) \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l'=1}^m \mathbf{P}(X = l') \wedge \mathbf{P}(X_n = l')} \\ &= \mathbf{P}(X_n = k). \end{aligned}$$

Damit ist $Y_n \stackrel{d}{=} X_n$. Da nach Voraussetzung $\mathbf{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbf{P}(X = k)$, folgt die fast sichere Konvergenz.

Für allgemeines E wählen wir für jedes p eine Partition von E in Mengen B_1, B_2, \dots in E mit $\mathbf{P}(Y \in \partial B_k) = 0$ und Durchmesser höchstens 2^{-p} . Wähle m groß genug, damit $\mathbf{P}(Y \notin B_0) < 2^{-p}$ mit $B_0 := E \setminus \bigcup_{k \leq m} B_k$. Für $k = 1, 2, \dots$, definiere Zufallsgrößen $\tilde{Z}, \tilde{Z}_1, \tilde{Z}_2, \dots$ so, dass $\tilde{Z} = k$ genau dann, wenn $Y \in B_k$ und $\tilde{Z}_n = k$ wenn $Y_n \in B_k$. Dann gilt $\tilde{Z}_n \xrightarrow{n \rightarrow \infty} \tilde{Z}$. Da $\tilde{Z}, \tilde{Z}_1, \dots$ nur Werte in einer endlichen Menge annehmen, können wir damit Zufallsgrößen Z, Z_1, Z_2, \dots definieren mit $Z_n \xrightarrow{n \rightarrow \infty}_{fs} Z$. Weiter seien $W_{n,k}$ Zufallsgrößen mit Verteilung $\mathbf{P}[X_n \in \cdot | X_n \in B_k]$ und $\tilde{Y}_{n,p} = \sum_k W_{n,k} 1_{Z_n=k}$, so dass $\tilde{Y}_{n,p} \stackrel{d}{=} X_n$ für alle n . Klar ist nun

$$\left\{r(\tilde{Y}_{n,p}, Y) > 2^{-p}\right\} \subseteq \{Z_n \neq Z\} \cup \{Y \in B_0\}.$$

Da $Z_n \xrightarrow{n \rightarrow \infty}_{fs} Z$ und $\mathbf{P}\{Y \in B_0\} < 2^{-p}$, gibt es für jedes p Zahlen $n_1 < n_2 < \dots$ mit

$$\mathbf{P}\left(\bigcup_{n \geq n_p} \{r(\tilde{Y}_{n,p}, Y) > 2^{-p}\}\right) < 2^{-p}$$

für alle p . Mit dem Borel-Cantelli-Lemma bekommen wir

$$\sup_{n \geq n_p} r(\tilde{Y}_{n,p}, Y) \leq 2^{-p}$$

für fast alle p . Wir definieren also $Y_n := \tilde{Y}_{n,p}$ für $n_p \leq n < n_{p+1}$ und beachten, dass $X_n \stackrel{d}{=} Y_n \xrightarrow{n \rightarrow \infty}_{f.s.} Y$. \square

10.2 Der Satz von Prohorov

In diesem Abschnitt beleuchten wir zunächst den Begriff der vagen Konvergenz. Dabei werden wir uns auf den Raum $E = \mathbb{R}$ beschränken. (Die meisten der hier gezeigten Aussagen gelten auch noch in lokal-kompakten Räumen.) Klar ist schon, dass aus der schwachen Konvergenz von Verteilungen die vage folgt, und dass die schwache Konvergenz äquivalent mit der Konvergenz der Verteilungsfunktionen ist (Korollar 10.7). Hauptresultat ist hier der Satz von Helly (Theorem 10.13), der angibt, dass jede Folge von Wahrscheinlichkeitsmaßen ein vage konvergente Teilfolge hat.

Anschließend untersuchen wir die Frage, inwieweit eine Folge von Wahrscheinlichkeitsmaßen auch schwach konvergente Teilfolgen besitzt. Dies führt uns auf den Begriff der Straffheit von Wahrscheinlichkeitsmaßen und den Satz von Prohorov (Theorem 10.19).

Wie wir bereits in Bemerkung 10.2.1 gesehen haben, kann es sein, dass das Grenzmaß bei vager Konvergenz kein Wahrscheinlichkeitsmaß ist. Allerdings zeigt nachfolgendes Resultat, dass das Grenzmaß zumindest Gesamtmasse höchstens 1 besitzt.

Lemma 10.12 (Massenverlust bei vager Konvergenz). *Sei $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ und $\mu \in \mathcal{P}(\mathbb{R})$ mit $\mathbf{P}_n \xrightarrow{n \rightarrow \infty}_v \mu$, so gilt $\mu(\mathbb{R}) \leq 1$.*

Beweis. Sei $f_1, f_2, \dots \in \mathcal{C}_c(\mathbb{R})$ mit $f_k \uparrow 1$. Dann gilt mit monotoner Konvergenz

$$\mu(\mathbb{R}) = \sup_{k \in \mathbb{N}} \mu[f_k] = \sup_{k \in \mathbb{N}} \limsup_{n \rightarrow \infty} \mathbf{P}_n[f_k] \leq 1.$$

\square

Theorem 10.13 (Satz von Helly). *Sei $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$. Dann gibt es eine Teilfolge $(n_k)_{k=1,2,\dots}$ und ein $\mu \in \mathcal{P}_{\leq 1}(\mathbb{R})$ mit $\mathbf{P}_{n_k} \xrightarrow{k \rightarrow \infty}_v \mu$.*

Beweis. Seien F_1, F_2, \dots die Verteilungsfunktionen von $\mathbf{P}_1, \mathbf{P}_2, \dots$. Weiter sei (x_1, x_2, \dots) eine Abzählung von \mathbb{Q} . Da $[0, 1]$ kompakt ist, gibt es zu jeder Folge $(F_n(x_i))_{n=1,2,\dots}$ eine konvergente Teilfolge. Mittels eines Diagonal-Argumentes gibt es eine Folge $(n_k)_{k=1,2,\dots}$, so dass $(F_{n_k}(x_i))_{k=1,2,\dots}$ für alle i gegen einen Limes $G(x_i)$ auf \mathbb{Q} konvergiert. Wir definieren

$$F(x) := \inf\{G(r) : r \in \mathbb{Q}, r > x\}.$$

Da alle F_n und damit G nicht-negative Zuwächse haben, gilt dasselbe für F . Aus der Definition von F und der Monotonie von G folgt weiter, dass F rechtsstetig ist. Nach Proposition 3.18 gibt es ein Maß μ auf \mathbb{R} mit $\mu((x, y]) = F(y) - F(x)$ für alle $x, y \in \mathbb{R}, x < y$. Es bleibt zu zeigen, dass $\mathbf{P}_n[f] \xrightarrow{n \rightarrow \infty} \mu[f]$ für alle $f \in \mathcal{C}_c(\mathbb{R})$. O.E. können wir annehmen, dass $f \geq 0$ ist.

Es ist $F_{n_k}(x) \xrightarrow{n \rightarrow \infty} F(x)$ an allen Stetigkeitsstellen x von F nach Konstruktion. Es gibt eine abzählbare Menge $D \subseteq \mathbb{R}$, so dass F auf D^c stetig ist. Damit ist $\mathbf{P}_n(U) \xrightarrow{n \rightarrow \infty} \mu(U)$ für alle endlichen Vereinigungen U von Intervallen mit Ecken in D^c . Sei nun $B \subseteq \mathbb{R}$ offen und beschränkt. Seien U_1, U_2, \dots und V_1, V_2, \dots Folgen von endlichen Vereinigungen von offenen Intervallen mit Ecken in D^c , so dass $U_k \uparrow B, V_k \downarrow \bar{B}$. Dann gilt

$$\begin{aligned} \mu(B) &= \lim_{k \rightarrow \infty} \mu(U_k) = \lim_{k \rightarrow \infty} \liminf_{n \rightarrow \infty} \mathbf{P}_n(U_k) \leq \liminf_{n \rightarrow \infty} \mathbf{P}_n(B) \\ &\leq \limsup_{n \rightarrow \infty} \mathbf{P}_n(B) \leq \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P}_n(V_k) = \lim_{k \rightarrow \infty} \mu(V_k) = \mu(\bar{B}). \end{aligned}$$

Da $\mu(f = t) > 0$ für höchstens abzählbar viele t , und da $\mathbf{P}_n(f > t) \leq 1_{t \geq \|f\|}$, folgt mit majorisierter Konvergenz

$$\begin{aligned} \mu[f] &= \int_0^\infty \mu(f > t) dt \leq \int_0^\infty \liminf_{n \rightarrow \infty} \mathbf{P}_n(f > t) dt = \liminf_{n \rightarrow \infty} \int_0^\infty \mathbf{P}_n(f > t) dt = \liminf_{n \rightarrow \infty} \mathbf{E}_n[f] \\ &\leq \limsup_{n \rightarrow \infty} \mathbf{P}_n[f] = \limsup_{n \rightarrow \infty} \int_0^\infty \mathbf{P}_n(f > t) dt = \int_0^\infty \limsup_{n \rightarrow \infty} \mathbf{P}_n(f > t) dt \leq \int_0^\infty \mu(f \geq t) \\ &= \mu[f]. \end{aligned} \quad \square$$

Wir kehren nun zurück zum Fall eines allgemeinen metrischen Raumes (E, r) . Um die Existenz von Häufungspunkten im Sinne der schwachen Konvergenz zu zeigen, muss sichergestellt sein, dass Grenzmaße wieder Wahrscheinlichkeitsmaße sind. Insbesondere darf also beim Grenzübergang keine Masse verloren gehen wie bei der vagen Konvergenz (siehe Lemma 10.12). Hierbei ist das Konzept der *Straffheit* zentral.

Definition 10.14 (Straffheit). Sei \mathcal{K} das System aller kompakten Mengen in E . Eine Familie $(\mathbf{P}_i)_{i \in I}$ in $\mathcal{P}(E)$ ist straff, falls

$$\sup_{K \in \mathcal{K}} \inf_{i \in I} \mathbf{P}_i(K) = 1.$$

Eine Familie $(X_i)_{i \in I}$ von E -wertigen Zufallsvariablen ist straff, falls $((X_i)_* \mathbf{P})_{i \in I}$ straff ist, d.h.

$$\sup_{K \in \mathcal{K}} \inf_{i \in I} \mathbf{P}(X_i \in K) = 1.$$

Bemerkung 10.15 (Äquivalente Formulierungen). 1. Die Definition der Straffheit einer Familie $(\mathbf{P}_i)_{i \in I}$ in $\mathcal{P}(E)$ ist äquivalent zu folgender Bedingung: für alle $\varepsilon > 0$ gibt es $K \subseteq E$ kompakt mit $\inf_{i \in I} \mathbf{P}_i(K) \geq 1 - \varepsilon$.

2. Falls $E = \mathbb{R}^d$, ist eine Familie $(\mathbf{P}_i)_{i \in I}$ genau dann straff, wenn

$$\sup_{r > 0} \inf_{i \in I} \mathbf{P}_i(B_r(0)) = 1,$$

wobei $B_r(0)$ die Kugel um 0 mit Radius r ist.

3. In Lemma 3.8 haben wir gezeigt, dass $\mathbf{P} \in \mathcal{P}(E)$ straff ist, falls (E, r) vollständig und separabel ist. Daraus folgt auch, dass jede endliche Familie von Wahrscheinlichkeitsmaßen auf der Borel'schen σ -Algebra eines polnischen Raumes straff ist.

4. Weiter ist eine abzählbare Familie $(\mathbf{P}_i)_{i=1,2,\dots}$ von Wahrscheinlichkeitsmaßen auf einem polnischen Raum (E, r) genau dann straff, falls

$$\sup_{K \in \mathcal{K}} \liminf_{i=1,2,\dots} \mathbf{P}_i(K) = 1.$$

Beweis. '⇒': klar, da $\liminf_{i=1,2,\dots} \mathbf{P}_i(K) \geq \inf_{i=1,2,\dots} \mathbf{P}_i(K) = 1$.

'⇐': Sei $\varepsilon > 0$ und K so, dass $\liminf_{i=1,2,\dots} \mathbf{P}_i(K) \geq 1 - \varepsilon/2$. Wähle N so, dass $\inf_{i=N+1, N+2, \dots} \mathbf{P}_i(K) \geq 1 - \varepsilon$ und K_1, \dots, K_N kompakt so, dass $\mathbf{P}_i(K_i) \geq 1 - \varepsilon$ für $i = 1, \dots, N$. Da $\tilde{K} = K \cup K_1 \cup \dots \cup K_N$ kompakt ist und $\inf_{i=1,2,\dots} \mathbf{P}_i(\tilde{K}) \geq 1 - \varepsilon$ folgt die Straffheit von $(\mathbf{P}_i)_{i=1,2,\dots}$. \square

Beispiel 10.16 (Straffe Mengen von Wahrscheinlichkeitsmaßen). 1. Ist E kompakt, so ist jede Familie von Wahrscheinlichkeitsmaßen auf $\mathcal{B}(E)$ straff.

2. Eine Familie $(X_i)_{i \in I}$ von reellwertigen Zufallsvariablen mit

$$\sup_{i \in I} \mathbf{P}[|X_i|] < \infty,$$

ist straff. Denn es gilt

$$\inf_{r>0} \sup_{i \in I} \mathbf{P}(|X_i| \geq r) \leq \inf_{r>0} \sup_{i \in I} \frac{\mathbf{P}[|X_i|]}{r} = 0.$$

3. Die Familie $(\delta_n)_{n=1,2,\dots}$, wobei δ_n das Diracmaß auf n ist, ist nicht straff.

Lemma 10.17 (Vage Konvergenz und Straffheit). Seien $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ und $\mu \in \mathcal{P}_{\leq 1}(\mathbb{R})$ mit

$$\mathbf{P}_n \xrightarrow[n \rightarrow \infty]{v} \mu.$$

Dann ist

$$\mu(\mathbb{R}) = 1 \quad \Longleftrightarrow \quad (\mathbf{P}_n)_{n=1,2,\dots} \text{ ist straff.}$$

In diesem Fall gilt $\mathbf{P}_n \xrightarrow[n \rightarrow \infty]{} \mu$.

Beweis. Für $r > 0$ wähle ein $g_r \in \mathcal{C}_c(\mathbb{R})$, $1_{B_r(0)} \leq g_r \leq 1_{B_{r+1}(0)}$. Dann ist $(\mathbf{P}_n)_{n=1,2,\dots}$ genau dann straff, wenn

$$\sup_{r>0} \liminf_{n \rightarrow \infty} \mathbf{P}_n[g_r] = 1.$$

' \Rightarrow ': Es gilt, da μ von unten stetig ist

$$1 = \sup_{r>0} \mu(B_r(0)) \leq \sup_{r>0} \mu[g_r] = \sup_{r>0} \liminf_{n \rightarrow \infty} \mathbf{P}_n[g_r] \leq 1.$$

' \Leftarrow ': Sei $(\mathbf{P}_n)_{n=1,2,\dots}$ straff. Dann folgt mit Lemma 10.12

$$1 \geq \mu(\mathbb{R}) = \sup_{r>0} \mu(B_r(0)) = \sup_{r>0} \mu[g_r] = \sup_{r>0} \liminf_{n \rightarrow \infty} \mathbf{P}_n[g_r] = 1.$$

Es bleibt, die schwache Konvergenz zu zeigen. Angenommen, $(\mathbf{P}_n)_{n=1,2,\dots}$ ist straff und $f \in \mathcal{C}_b(\mathbb{R})$. Dann gilt

$$\begin{aligned} \limsup_{n \rightarrow \infty} |\mathbf{P}_n[f] - \mu[f]| &\leq \inf_{r>0} \limsup_{n \rightarrow \infty} (|\mathbf{P}_n[f - fg_r]| + |\mathbf{P}_n[fg_r] - \mu[fg_r]| + |\mu[f - fg_r]|) \\ &\leq \|f\| \inf_{r>0} \limsup_{n \rightarrow \infty} \mathbf{P}_n(B_r(0)^c) + \inf_{r>0} \mu[B_r(0)^c] = 0, \end{aligned}$$

woraus $\mathbf{P}_n \xrightarrow[n \rightarrow \infty]{} \mu$ folgt. □

Korollar 10.18 (Schwache Konvergenz und Straffheit). Seien $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$. Falls $\mathbf{P}_n \xrightarrow[n \rightarrow \infty]{} \mathbf{P}$, so ist $(\mathbf{P}_n)_{n \in \mathbb{N}}$ straff.

Beweis. Da aus der schwachen Konvergenz von $\mathbf{P}_1, \mathbf{P}_2, \dots$ gegen \mathbf{P} , die vage Konvergenz folgt, gelten für $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots$ die Voraussetzungen von Lemma 10.17 und $\mathbf{P}(\mathbb{R}) = 1$. Also ist $(\mathbf{P}_n)_{n \in \mathbb{N}}$ straff. □

Um die schwache Konvergenz von Wahrscheinlichkeitsmaßen festzustellen, ist Theorem 10.6 hilfreich. Wir wenden uns nun der Frage zu, ob eine Folge von Wahrscheinlichkeitsmaßen überhaupt einen Häufungspunkt haben kann. Das bedeutet, dass es eine Teilfolge gibt, die schwach gegen ein Wahrscheinlichkeitsmaß konvergiert.

Theorem 10.19 (Satz von Prohorov). *Sei (E, r) vollständig und separabel und $(\mathbf{P}_i)_{i \in I}$ eine Familie in $\mathcal{P}(E)$. Dann sind äquivalent:*

1. Die Familie $(\mathbf{P}_i)_{i \in I}$ ist relativ kompakt bezüglich der Topologie der schwachen Konvergenz, d.h. jede Folge in $(\mathbf{P}_i)_{i \in I}$ hat eine schwach konvergente Teilfolge.
2. Zu jedem $\varepsilon > 0$ gibt es ein $N \in \mathbb{N}$ und $x_1, \dots, x_N \in E$, so dass

$$\inf_{i \in I} \mathbf{P}_i \left(\bigcup_{k=1}^N B_\varepsilon(x_k) \right) \geq 1 - \varepsilon.$$

3. Die Familie $(\mathbf{P}_i)_{i \in I}$ ist straff.

Beweis. Sei x_1, x_2, \dots eine dichte Teilfolge in E .

1. \Rightarrow 2.: Angenommen, 2. ist nicht wahr. Dann gibt es ein $\varepsilon > 0$ und für jedes $N = 1, 2, \dots$ ein \mathbf{P}_{i_N} mit $\mathbf{P}_{i_N} \left(\bigcup_{k=1}^N B_\varepsilon(x_k) \right) \leq 1 - \varepsilon$. Wegen der Relativkompaktheit gäbe es dann eine schwach gegen ein $\mathbf{P} \in \mathcal{P}(E)$ konvergente Teilfolge $(\mathbf{P}_{i_M})_{M=1,2,\dots}$. Damit gilt wegen Theorem 10.6 ((i) \Rightarrow (iii)), dass

$$1 = \mathbf{P}(E) = \sup_{N \in \mathbb{N}} \mathbf{P} \left(\bigcup_{k=1}^N B_\varepsilon(x_k) \right) \leq \sup_{N \in \mathbb{N}} \liminf_{M \rightarrow \infty} \mathbf{P}_{i_M} \left(\bigcup_{k=1}^N B_\varepsilon(x_k) \right) \leq 1 - \varepsilon,$$

also ein Widerspruch.

2. \Rightarrow 3.: Sei $\varepsilon > 0$. Für $j = 1, 2, \dots$ wählen wir x_{j1}, \dots, x_{jN_j} , so dass

$$\inf_{i \in I} \mathbf{P}_i \left(\bigcup_{k=1}^{N_j} B_{\varepsilon 2^{-j}}(x_{jk}) \right) > 1 - \varepsilon 2^{-j}.$$

Weiter setzen wir

$$K := \bigcap_{j=1}^{\infty} \bigcup_{k=1}^{N_j} B_{\varepsilon 2^{-j}}(x_{jk}).$$

Dann ist $K \subseteq E$ nach Konstruktion total beschränkt, nach Proposition 1.9 also relativ kompakt, \bar{K} also kompakt. Außerdem gilt

$$\sup_{i \in I} \mathbf{P}_i(\bar{K}^c) \leq \sup_{i \in I} \sum_{j=1}^{\infty} \mathbf{P}_i \left(\bigcap_{k=1}^{N_j} (B_{\varepsilon 2^{-j}}(x_{jk}))^c \right) \leq \varepsilon.$$

Damit ist die Familie $(\mathbf{P}_i)_{i \in I}$ straff.

3. \Rightarrow 1.: Sei $\mathbf{P}_1, \mathbf{P}_2, \dots$ eine Folge in der Familie $(\mathbf{P}_i)_{i \in I}$. Ziel ist es, eine konvergente Teilfolge zu bestimmen. Hierzu wählen wir kompakte Mengen $K_1 \subseteq K_2 \subseteq \dots \subseteq E$ aus mit $\inf_{n=1,2,\dots} \mathbf{P}_n(K_j) \geq 1 - 1/j$. Weiter wählen wir das System kompakter Mengen

$$\mathcal{K} := \left\{ \bigcup_{k=1}^N K_{j_k} \cap \overline{B_{\varepsilon_k}(x_k)} : N, j_k \in \mathbb{N}, \varepsilon_k \in \mathbb{Q}^+ \right\}.$$

Damit ist \mathcal{K} ein Halbring und auch ein schnittstabiler Erzeuger von $\sigma(E)$. Da \mathcal{K} abzählbar ist, können wir eine Teilfolge $\mathbf{P}_{n_1}, \mathbf{P}_{n_2}, \dots$ von $\mathbf{P}_1, \mathbf{P}_2, \dots$ finden, so dass $\mathbf{P}_{n_k}(A)$ für alle $A \in \mathcal{K}$ konvergiert. Wir setzen für $A \in \mathcal{K}$

$$\mu(A) = \lim_{k \rightarrow \infty} \mathbf{P}_{n_k}(A).$$

Damit ist nach Theorem 3.9 und Theorem 3.15 der Inhalt μ auf \mathcal{K} eindeutig zu einem Maß auf $\sigma(\mathcal{K}) = \mathcal{B}(E)$ fortsetzbar. Das so definierte μ ist ein Wahrscheinlichkeitsmaß, weil ja

$$1 \geq \mu(E) = \sup_{j=1,2,\dots} \mu(K_j) = \sup_{j=1,2,\dots} \lim_{k \rightarrow \infty} \mathbf{P}_{n_k}(K_j) \geq \sup_{j=1,2,\dots} 1 - 1/j = 1.$$

Außerdem gilt für jede offene Menge A

$$\mu(A) = \sup_{A \supseteq H \in \mathcal{K}} \mu(H) = \sup_{A \supseteq H \in \mathcal{K}} \lim_{k \rightarrow \infty} \mathbf{P}_{n_k}(H) \leq \liminf_{k \rightarrow \infty} \mathbf{P}_{n_k}(A).$$

Nun folgt $\mathbf{P}_{n_k} \xrightarrow{k \rightarrow \infty} \mu$ aus Theorem 10.6 ((iii) \Rightarrow (i)). □

10.3 Separierende Funktionenklassen

Nun werden wir das Konzept der separierenden Funktionenklasse vorstellen. Dabei wird klar werden, welche Nützlichkeit charakteristische Funktionen und Laplace-Transformierte von Verteilungen (siehe Definition 7.11) besitzen. Diese basieren nämlich auf zwei bestimmten Funktionenklassen, die separierend sind.

Definition 10.20 (Punktetrennende und separierende Funktionenklassen).

1. Eine Funktionenklasse $\mathcal{M} \subseteq \mathcal{C}(E)$ heißt *punktetrennend* in E , falls es für alle $x, y \in E$ mit $x \neq y$ ein $f \in \mathcal{M}$ gibt mit $f(x) \neq f(y)$.
2. Eine Funktionenklasse $\mathcal{M} \subseteq \mathcal{C}(E)$ heißt *separierend* in $\mathcal{P}(E)$, falls aus $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$ und

$$\mathbf{P}[f] = \mathbf{Q}[f] \text{ für alle } f \in \mathcal{M}$$

folgt, dass $\mathbf{P} = \mathbf{Q}$.

Beispiel 10.21. 1. Die Funktionenklasse $\mathcal{M} := \mathcal{C}_b(E)$ ist sowohl punktstrennend als auch separierend. Ist nämlich $x \neq y$, so ist $z \mapsto r(x, z) \wedge 1$ eine beschränkte, stetige Funktion, die x und y trennt. Ist außerdem $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$ und $\mathbf{P} \neq \mathbf{Q}$, so gibt es einen offenen Ball A mit $\mathbf{P}(A) \neq \mathbf{Q}(A)$. Sei f_1, f_2, \dots eine Folge in $\mathcal{C}_b(E)$ mit $f_n \uparrow 1_A$. Wäre $\mathbf{P}[f_n] = \mathbf{Q}[f_n]$ für alle $n = 1, 2, \dots$, so wäre auch

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}[f_n] = \lim_{n \rightarrow \infty} \mathbf{Q}[f_n] = \mathbf{Q}(A)$$

im Widerspruch zur Voraussetzung.

2. Die Funktionenklasse $\{x \mapsto cx : c \in \mathbb{R}\}$ aller linearen Funktionen ist zwar punktstrennend, aber nicht separierend.

Das nächste Resultat benötigt das Stone-Weierstrass-Theorem, das wir zunächst wiederholen.

Definition 10.22 (Algebra). Ein Mengensystem $\mathcal{M} \subseteq \mathcal{C}(E)$ heißt Algebra, falls $1 \in \mathcal{M}$, sowie mit f, g und $\alpha, \beta \in \mathbb{R}$ auch $\alpha f + \beta g$, sowie fg in \mathcal{M} enthalten sind.

Theorem 10.23 (Stone-Weierstrass). Sei (E, r) kompakt und $\mathcal{M} \subseteq \mathcal{C}_b(E)$ eine punktetrennende Algebra. Dann ist \mathcal{M} dicht in $\mathcal{C}_b(E)$ bzgl. der Supremumsnorm.

Beweis. Siehe Analysis. □

Theorem 10.24 (Punktetrennende und separierende Algebren).

Sei (E, r) vollständig und separabel. Ist $\mathcal{M} \subseteq \mathcal{C}_b(E)$ punktetrennend und so, dass mit $f, g \in \mathcal{M}$ auch $fg \in \mathcal{M}$. Dann ist \mathcal{M} separierend.

Beweis. Sei $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$. Ohne Einschränkung ist $1 \in \mathcal{M}$, da $\mathbf{P}[1] = \mathbf{Q}[1]$ immer gilt. Also ist \mathcal{M} o.E. eine Algebra. Sei $\varepsilon > 0$ und K kompakt, so dass $\mathbf{P}(K) > 1 - \varepsilon$, $\mathbf{Q}(K) > 1 - \varepsilon$. Für $g \in \mathcal{C}_b(E)$ gibt es nach dem Stone-Weierstrass Theorem 10.23 eine Folge $(g_n)_{n=1,2,\dots}$ in \mathcal{M} mit

$$\sup_{x \in K} |g_n(x) - g(x)| \xrightarrow{n \rightarrow \infty} 0. \quad (10.4)$$

Nun,

$$\begin{aligned} |\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}]| &\leq |\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{P}[ge^{-\varepsilon g^2}; K]| \\ &\quad + |\mathbf{P}[ge^{-\varepsilon g^2}; K] - \mathbf{P}[g_n e^{-\varepsilon g_n^2}; K]| \\ &\quad + |\mathbf{P}[g_n e^{-\varepsilon g_n^2}; K] - \mathbf{P}[g_n e^{-\varepsilon g_n^2}]| \\ &\quad + |\mathbf{P}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[g_n e^{-\varepsilon g_n^2}]| \\ &\quad + |\mathbf{Q}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[g_n e^{-\varepsilon g_n^2}; K]| \\ &\quad + |\mathbf{Q}[g_n e^{-\varepsilon g_n^2}; K] - \mathbf{Q}[ge^{-\varepsilon g^2}; K]| \\ &\quad + |\mathbf{Q}[ge^{-\varepsilon g^2}; K] - \mathbf{Q}[ge^{-\varepsilon g^2}]| \end{aligned}$$

Den ersten Term beschränken wir durch

$$|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{P}[ge^{-\varepsilon g^2}; K]| \leq \frac{C}{\sqrt{\varepsilon}} \mathbf{P}(K^c) \leq C\sqrt{\varepsilon}$$

mit $C = \sup_{x \geq 0} x e^{-x^2}$; analog den dritten, fünften und letzten. Der zweite und vorletzte Term konvergieren für $n \rightarrow \infty$ gegen 0 wegen (10.4). Da \mathcal{M} eine Algebra ist, kann $g_n e^{-\varepsilon g_n^2}$ durch Funktionen in \mathcal{M} approximiert werden, womit auch der vierte Term für $n \rightarrow \infty$ gegen 0 konvergiert. Damit gilt

$$|\mathbf{P}[g] - \mathbf{Q}[g]| = \lim_{\varepsilon \rightarrow 0} |\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}]| \leq 4C \lim_{\varepsilon \rightarrow 0} \sqrt{\varepsilon} = 0.$$

Da g beliebig war und $\mathcal{C}_b(E)$ separierend ist, folgt $\mathbf{P} = \mathbf{Q}$. □

Wir kommen nun auf die charakteristische Funktion und die Laplace-Transformierte zurück. Wie bereits erwähnt ist die Nützlichkeit der charakteristischen Funktion und der Laplace-Transformierten darauf zurückzuführen, dass sie verteilungsbestimmend sind.

Proposition 10.25 (Charakteristische Funktion verteilungsbestimmend).

Ein Wahrscheinlichkeitsmaß $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ ($\mathbf{P} \in \mathcal{P}(\mathbb{R}_+^d)$) wird eindeutig durch die charakteristische Funktion $\psi_{\mathbf{P}}$ (die Laplace-Transformierte $\mathcal{L}_{\mathbf{P}}$) bestimmt.

Beweis. Wir zeigen die Aussage nur für charakteristische Funktionen, die für Laplace-Transformierte beweist man analog. Wir stellen fest, dass die Menge $\mathcal{M} := \{x \mapsto e^{itx}; t \in \mathbb{R}^d\}$ in \mathbb{R}^d Punkte trennt. Da $\mathcal{M} \subseteq \mathcal{C}_b(\mathbb{R}^d)$ und abgeschlossen unter Produktbildung ist, ist sie nach Theorem 10.24 auch separierend. Dies zeigt die Aussage. \square

Korollar 10.26 (Unabhängigkeit und charakteristische Funktion). 1. Eine Familie $(X_j)_{j \in I}$ von reellwertigen Zufallsvariablen ist genau dann unabhängig, falls für alle $J \in I$

$$\mathbf{E} \left[\prod_{j \in J} e^{it_j X_j} \right] = \prod_{j \in J} \mathbf{E} [e^{it_j X_j}] \quad (10.5)$$

für alle $(t_j)_{j \in J} \in \mathbb{R}^J$ gilt.

2. Eine Familie $(X_j)_{j \in I}$ von Zufallsvariablen mit Werten in \mathbb{R}_+ ist genau dann unabhängig, falls für alle $J \in I$

$$\mathbf{E} \left[\prod_{j \in J} e^{-t_j X_j} \right] = \prod_{j \in J} \mathbf{E} [e^{-t_j X_j}]$$

für alle $(t_j)_{j \in J} \in \mathbb{R}^J$ gilt.

Beweis. Wir zeigen nur die erste Aussage, die zweite folgt analog. Ist $(X_j)_{j \in I}$ unabhängig, so sind nach Lemma 9.4 auch die Zufallsvariablen $(e^{it_j X_j})_{j \in I}$ für alle $(t_j)_{j \in J} \in \mathbb{R}^J$ unabhängig. Damit folgt (10.5) aus Proposition 9.5. Umgekehrt gelte (10.5). Einerseits stellt die linke Seite von (10.5) die charakteristische Funktion der Verteilung $((X_j)_{j \in J})_* \mathbf{P}$ dar. Andererseits ist die rechte Seite von (10.5) die charakteristische Funktion von $\bigotimes_{j \in J} (X_j)_* \mathbf{P}$. Da die charakteristische Funktion nach Proposition 10.25 die gemeinsame Verteilung von $(X_j)_{j \in J}$ eindeutig bestimmt, gilt also $((X_j)_{j \in J})_* \mathbf{P} = \bigotimes_{j \in J} (X_j)_* \mathbf{P}$. Damit folgt die Unabhängigkeit von $(X_j)_{j \in I}$ aus Proposition 9.2. \square

10.4 Der Satz von Lévy

Wir wollen nun den Zusammenhang zwischen schwacher Konvergenz und der Konvergenz der charakteristischen Funktionen der zu Grunde liegenden Verteilungen beleuchten. Seien hierzu $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}^d)$. Wie man aus Proposition 10.27 sieht, folgt die schwache Konvergenz $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}$ aus der punktweisen Konvergenz der charakteristischen Funktionen, $\psi_{\mathbf{P}_n}(t) \xrightarrow{n \rightarrow \infty} \psi_{\mathbf{P}}(t)$, $t \in \mathbb{R}^d$, gegeben $(\mathbf{P}_n)_{n \in \mathbb{N}}$ ist straff. Entscheidend ist nun, dass man die Straffheit der Familie $(\mathbf{P}_n)_{n \in \mathbb{N}}$ ebenfalls aus den charakteristischen Funktionen ablesen kann, wie wir in Proposition 10.32 zeigen werden. Dies führt zu der Aussage des Lévy'schen Stetigkeitssatzes (Theorem 10.33), der angibt, wann der punktweise Limes von charakteristischen Funktionen wieder charakteristische Funktion eines Wahrscheinlichkeitsmaßes ist.

Proposition 10.27 (Separierende Funktionenklasse und schwache Konvergenz). Sei (E, r) vollständig und separabel und $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$. Dann sind äquivalent:

1. $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}$.

2. $(\mathbf{P}_n)_{n=1,2,\dots}$ ist straff und es gibt eine separierende Familie $\mathcal{M} \subseteq \mathcal{C}_b(E)$ mit

$$\mathbf{P}_n[f] \xrightarrow{n \rightarrow \infty} \mathbf{P}[f] \text{ für alle } f \in \mathcal{M}.$$

Beweis. 1. \Rightarrow 2. Nach Korollar 10.18 ist $(\mathbf{P}_n)_{n=1,2,\dots}$ straff. Der zweite Teil von 2. gilt wegen der Definition der schwachen Konvergenz.

2. \Rightarrow 1. Angenommen, $(\mathbf{P}_n)_{n=1,2,\dots}$ ist straff und $\mathbf{P}_1, \mathbf{P}_2, \dots$ konvergiert nicht schwach gegen \mathbf{P} . Dann gibt es $\varepsilon > 0$, ein $f \in \mathcal{C}_b(E)$ und eine Teilfolge $(n_k)_{k=1,2,\dots}$, so dass

$$|\mathbf{P}_{n_k}[f] - \mathbf{P}[f]| > \varepsilon \text{ für alle } k. \quad (10.6)$$

Nach Theorem 10.19 gibt es eine Teilfolge $(n_{k_\ell})_{\ell=1,2,\dots}$ und ein $\mathbf{Q} \in \mathcal{P}(E)$, so dass $\mathbf{P}_{n_{k_\ell}} \xrightarrow{\ell \rightarrow \infty} \mathbf{Q}$. Wegen (10.6) ist

$$|\mathbf{P}[f] - \mathbf{Q}[f]| \geq |\liminf_{\ell \rightarrow \infty} (\mathbf{P}[f] - \mathbf{P}_{n_{k_\ell}}[f]) + \liminf_{\ell \rightarrow \infty} (\mathbf{P}_{n_{k_\ell}}[f] - \mathbf{Q}[f])| > \varepsilon,$$

insbesondere also $\mathbf{P} \neq \mathbf{Q}$. Andererseits haben wir für alle $g \in \mathcal{M}$

$$\mathbf{P}[g] = \lim_{\ell \rightarrow \infty} \mathbf{P}_{n_{k_\ell}}[g] = \mathbf{Q}[g].$$

Da \mathcal{M} separierend ist, ist dies ein Widerspruch und 1. ist gezeigt. \square

Sei $\mathbf{P} \in \mathcal{P}(\mathbb{R})$ und $\psi_{\mathbf{P}}$ deren charakteristische Funktion. Wir zeigen zunächst eine Abschätzung, die wichtig ist, um Straffheit und die $\psi_{\mathbf{P}}$ in Verbindung zu bringen.

Lemma 10.28 (Straffheit und die charakteristische Funktion). *Sei $\mathbf{P} \in \mathcal{P}(\mathbb{R})$. Dann gilt für alle $r > 0$*

$$\mathbf{P}((-\infty; -r] \cup [r; \infty)) \leq \frac{r}{2} \int_{-2/r}^{2/r} (1 - \psi_{\mathbf{P}}(t)) dt, \quad (10.7)$$

Beweis. Es ist $\sin(x)/x \leq 1$ für $x \leq 2$ und $\sin x \leq x/2$ für $x \geq 2$. Sei X eine Zufallsvariable mit Verteilung \mathbf{P} . Also gilt für jedes $c > 0$ nach Fubini

$$\begin{aligned} \int_{-c}^c (1 - \psi_{\mathbf{P}}(t)) dt &= \mathbf{P} \left[\int_{-c}^c (1 - e^{itX}) dt \right] = \mathbf{P} \left[2c - \frac{1}{iX} e^{itX} \Big|_{t=-c}^c \right] \\ &= 2c \mathbf{P} \left[1 - \frac{\sin(cX)}{cX} \right] \\ &\geq 2c \mathbf{P} \left[1 - \frac{\sin(cX)}{cX}; |cX| \geq 2 \right] \\ &\geq c \cdot \mathbf{P}(|cX| \geq 2) = c \mathbf{P}((-\infty; -\frac{2}{c}] \cup [\frac{2}{c}; \infty)), \end{aligned}$$

und die Behauptung folgt mit $c = 2/r$. \square

Definition 10.29 (Gleichgradige Stetigkeit). *Wir wiederholen eine Definition aus der Analysis. Eine Menge $\mathcal{M} \subseteq \mathcal{C}(\mathbb{R}^d)$ heißt in $x \in \mathbb{R}^d$ gleichgradig stetig, falls*

$$\sup_{f \in \mathcal{M}} |f(y) - f(x)| \xrightarrow{y \rightarrow x} 0.$$

Bemerkung 10.30 (Äquivalente Bedingung für Folgen). Falls $\mathcal{M} = \{f_1, f_2, \dots\}$, so ist die Bedingung

$$\limsup_{n \rightarrow \infty} |f_n(y) - f_n(x)| \xrightarrow{y \rightarrow x} 0$$

äquivalent.

Lemma 10.31 (Gleichgradige Stetigkeit und Konvergenz). Seien $f_1, f_2, \dots \in \mathcal{C}(\mathbb{R}^d)$, so dass $f_n \xrightarrow{n \rightarrow \infty} f$ punktweise für eine Funktion $f : \mathbb{R}^d \rightarrow \mathbb{R}$ gilt. Genau dann ist f in 0 stetig, wenn $(f_n)_{n=1,2,\dots}$ in 0 gleichgradig stetig ist.

Beweis. Ist $(f_n)_{n=1,2,\dots}$ gleichgradig stetig in 0, so folgt

$$|f(t) - f(0)| = \left| \lim_{n \rightarrow \infty} (f_n(t) - f_n(0)) \right| \leq \limsup_{n \rightarrow \infty} |f_n(t) - f_n(0)| \xrightarrow{t \rightarrow 0} 0.$$

Ist andersherum f stetig in 0, so gilt

$$\limsup_{n \rightarrow \infty} |f_n(t) - f_n(0)| \leq \limsup_{n \rightarrow \infty} |f_n(t) - f(t)| + |f(t) - f(0)| + |f(0) - f_n(0)| = |f(t) - f(0)| \xrightarrow{t \rightarrow 0} 0.$$

□

Proposition 10.32 (Straffheit und gleichgradige Stetigkeit). Sei $(\mathbf{P}_i)_{i \in I}$ eine Familie in $\mathcal{P}(\mathbb{R}^d)$. Ist $(\psi_{\mathbf{P}_i})_{i \in I}$ gleichgradig stetig in 0, so ist $(\mathbf{P}_i)_{i \in I}$ straff.

Beweis. Es genügt zu zeigen, dass $((\pi_k)_* \mathbf{P}_i)_{i \in I}$ für alle Projektionen π_1, \dots, π_d straff ist. Es gilt offenbar $\psi_{(\pi_k)_* \mathbf{P}_i}(t) = \psi_{\mathbf{P}_i}(te_k)$, falls e_k der k -te Einheitsvektor ist. Damit genügt es, die Behauptung im Fall $d = 1$ zu zeigen. Da $\psi_{\mathbf{P}_i}(0) = 1$ für alle $i \in I$ gilt, folgern wir aus der gleichgradigen Stetigkeit, dass

$$\sup_{i \in I} |1 - \psi_{\mathbf{P}_i}(t)| \xrightarrow{t \rightarrow 0} 0,$$

also, siehe Bemerkung 10.15,

$$\begin{aligned} \sup_{r > 0} \inf_{i \in I} \mathbf{P}_i([-r; r]) &\geq 1 - \inf_{r > 0} \sup_{i \in I} \frac{r}{2} \int_{-2/r}^{2/r} (1 - \psi_{\mathbf{P}_i}(t)) dt \\ &\geq 1 - \inf_{r > 0} \frac{r}{2} \int_{-2/r}^{2/r} \sup_{i \in I} |1 - \psi_{\mathbf{P}_i}(t)| dt \\ &\geq 1 - 2 \inf_{r > 0} \sup_{t \in [0; 2/r]} \sup_{i \in I} |1 - \psi_{\mathbf{P}_i}(t)| = 1. \end{aligned}$$

Damit ist die Behauptung gezeigt. □

Theorem 10.33 (Lévy'scher Stetigkeitssatz). Seien $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}^d)$ und $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$, so dass $\psi_{\mathbf{P}_n}(t) \xrightarrow{n \rightarrow \infty} \psi(t)$ für alle $t \in \mathbb{R}^d$. Falls ψ stetig in 0 ist, so ist $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}$ für ein $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ mit $\psi_{\mathbf{P}} = \psi$.

Beweis. Da $\psi_{\mathbf{P}_n}$ punktweise gegen die in 0 stetige Funktion ψ konvergieren, folgt mit Lemma 10.31, dass $(\psi_{\mathbf{P}_n})_{n=1,2,\dots}$ in 0 gleichgradig stetig ist. Mit Proposition 10.32 folgt, dass $(\mathbf{P}_n)_{n=1,2,\dots}$ straff ist. Sei $(n_k)_{k=1,2,\dots}$ eine Teilfolge und $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$, so dass $\mathbf{P}_{n_k} \xrightarrow{k \rightarrow \infty} \mathbf{P}$. Da $x \mapsto e^{itx}$ eine stetige, beschränkte Funktion ist, folgt $\psi_{\mathbf{P}_{n_k}}(t) \xrightarrow{k \rightarrow \infty} \psi_{\mathbf{P}}(t)$ für alle $t \in \mathbb{R}^d$. Andererseits ist nach Voraussetzung auch $\psi_{\mathbf{P}_n}(t) \xrightarrow{n \rightarrow \infty} \psi(t)$, woraus $\psi_{\mathbf{P}} = \psi$ folgt. Damit ist ψ als charakteristische Funktion von \mathbf{P} identifiziert, und da diese \mathbf{P} eindeutig bestimmt, gilt damit $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}$. □

Beispiel 10.34 (Satz von deMoivre-Laplace). Seien $S_n \sim B(n, p)$. Der Satz von deMoivre-Laplace besagt, dass

$$S_n^* := \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{n \rightarrow \infty} N(0, 1). \quad (10.8)$$

Wir wollen dies nun nochmal mit Hilfe von charakteristischen Funktionen zeigen, also $\psi_{S_n^*} \xrightarrow{n \rightarrow \infty} \psi_{N(0,1)}$ punktweise. Dazu verwenden wir Proposition 7.12.3 und schreiben mit $q := 1 - p$ und $C_1, C_2, \dots \in \mathbb{C}$ mit $\limsup_{n \rightarrow \infty} |C_n| < \infty$

$$\begin{aligned} \psi_{S_n^*}(t) &= \exp\left(-it\sqrt{\frac{np}{q}}\right) \cdot \psi_{B(n,p)}\left(\frac{t}{\sqrt{npq}}\right) \\ &= \exp\left(-it\sqrt{\frac{np}{q}}\right) \left(q + p \exp\left(\frac{it}{\sqrt{npq}}\right)\right)^n \\ &= \left(q \exp\left(-it\sqrt{\frac{p}{nq}}\right) + p \exp\left(it\sqrt{\frac{q}{np}}\right)\right)^n \\ &= \left(1 - qit\sqrt{\frac{p}{nq}} - q\frac{t^2}{2}\frac{p}{nq} + pit\sqrt{\frac{q}{np}} - p\frac{t^2}{2}\frac{q}{np} + \frac{C_n}{n^{3/2}}\right)^n \\ &= \left(1 - \frac{t^2}{2}\frac{1}{n} + \frac{C_n}{n^{3/2}}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}} = \psi_{N(0,1)}(t). \end{aligned}$$

Aus Theorem 10.33 folgt nun (10.8).

Der Lévy-sche Stetigkeitssatz lässt sich auch mit Laplace-Transformierten formulieren. Wir geben den Satz ohne Beweis an:

Theorem 10.35 (Lévy'scher Stetigkeitssatz für Laplace-Transformierte). Seien $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}_+^d)$ und $\mathcal{L} : \mathbb{R}^d \rightarrow [0, 1]$, so dass $\mathcal{L}_{\mathbf{P}_n}(t) \xrightarrow{n \rightarrow \infty} \mathcal{L}(t)$ für alle $t \in \mathbb{R}^d$. Falls \mathcal{L} stetig in 0 ist, so ist $\mathbf{P}_n \xrightarrow{n \rightarrow \infty} \mathbf{P}$ für ein $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ mit $\mathcal{L}_{\mathbf{P}} = \mathcal{L}$.

Beispiel 10.36 (Konvergenz der geometrischen zur Exponentialverteilung). Sei etwa $X_n \sim \mu_{\text{geo}(p_n)}$ verteilt und $n \cdot p_n \xrightarrow{n \rightarrow \infty} \lambda$. Dann ist

$$\begin{aligned} \mathcal{L}_{X_n/n}(t) &= \mathbf{P}[e^{-tX_n/n}] = \sum_{k=1}^{\infty} (1-p_n)^{k-1} p_n e^{-tk/n} \\ &= p_n e^{-t/n} \frac{1}{1 - (1-p_n)e^{-t/n}} \\ &= \frac{\lambda}{n(1 - (1 - \lambda/n)(1 - t/n))} + o(1/n) \\ &\xrightarrow{n \rightarrow \infty} \frac{\lambda}{\lambda + t}. \end{aligned}$$

Also gilt $\frac{X_n}{n} \xrightarrow{n \rightarrow \infty} Y$, wobei $Y \sim \mu_{\text{exp}(\lambda)}$, da

$$\mathcal{L}_{\text{exp}(\lambda)}(t) = \int_0^{\infty} \lambda e^{-\lambda a} e^{-ta} da = \frac{\lambda}{\lambda + t}.$$

11 Grenzwertsätze in Verteilung

Wir werden nun unsere Kenntnisse über schwache Konvergenz und charakteristische Funktionen in speziellen Situationen einsetzen. In Abschnitt 11.1 geht es um Aussagen, wann die Summe von Zufallsvariablen gegen eine Poissonverteilte Zufallsvariable konvergiert. In Abschnitt 11.2 werden wir den zentralen Grenzwertsatz von Lindeberg-Feller kennenlernen, der eine Charakterisierung für die schwache Konvergenz gegen eine Normalverteilung darstellt. In Abschnitt 11.3 geht es schließlich um Erweiterungen für den Fall von mehrdimensionalen Zufallsvariablen.

11.1 Poisson-Konvergenz

Bereits bekannt ist die Aussage, dass $B(n, p_n)$ für $n \cdot p_n \xrightarrow{n \rightarrow \infty} \lambda$ für große n schwach gegen die $\text{Poi}(\lambda)$ -Verteilung konvergiert, siehe Beispiel 11.1. In diesem Abschnitt verallgemeinern wir diese Aussage, siehe Theorem 11.5.

Beispiel 11.1 (Poisson-Approximation der Binomialverteilung). Sei $p_1, p_2, \dots \in [0, 1]$ so, dass $n \cdot p_n \xrightarrow{n \rightarrow \infty} \lambda$. Dann ist bereits aus der Vorlesung Stochastik bekannt, dass

$$B(n, p_n)(\{k\}) \xrightarrow{n \rightarrow \infty} \text{Poi}(\lambda)(\{k\}).$$

Anders ausgedrückt ist das eine Aussage über schwache Konvergenz:

$$B(n, p_n) \xrightarrow{n \rightarrow \infty} \text{Poi}(\lambda). \quad (11.1)$$

Der Satz von Lévy gibt eine weitere Möglichkeit an, dieses Resultat zu beweisen. Wir erinnern an die charakteristischen Funktionen der Binomial- und Poisson-Verteilung aus Beispiel 7.13. Wir schreiben direkt

$$\begin{aligned} \psi_{B(n, p_n)}(t) &= \left(1 - p_n(1 - e^{it})\right)^n \\ &= \left(1 - \frac{n \cdot p_n}{n}(1 - e^{it})\right)^n \\ &\xrightarrow{n \rightarrow \infty} \exp(-\lambda(1 - e^{it})) = \psi_{\text{Poi}(\lambda)}(t). \end{aligned}$$

Insbesondere konvergieren die charakteristischen Funktionen der Binomialverteilungen punktweise gegen eine in 0 stetige Funktion, nämlich die charakteristische Funktion der Poisson-Verteilung. Aus Theorem 10.33 folgt (11.1).

Im folgenden werden wir sehen, dass die schwache Konvergenz gegen eine Poisson-Verteilung noch allgemeiner gilt. Hierzu werden wir Erzeugendenfunktionen verwenden.

Bemerkung 11.2 (Erzeugendenfunktion). Betrachte eine Zufallsvariable X mit Werten in \mathbb{Z}_+ und definiere die *Erzeugendenfunktion*

$$z \mapsto \varphi_X(z) := \mathbf{P}[z^X] = \sum_{k=0}^{\infty} z^k \mathbf{P}[X = k].$$

Wir bemerken, dass diese für $z \in [0, 1]$ mit der Laplace-Transformierten von X eng verwandt ist, weil ja (mit $z = e^{-t}$)

$$\mathcal{L}_X(t) = \mathbf{P}[e^{-tX}] = \mathbf{P}[z^X] = \varphi_X(z).$$

Insbesondere übertragen sich zwei Eigenschaften von Laplace-Transformierten zu Erzeugendenfunktionen.

1. *Erzeugendenfunktionen verteilungsbestimmend, siehe Proposition 10.25:* Die Verteilung von X ist eindeutig bestimmt durch $z \mapsto \varphi_X(z)$ für $z \in [0, 1]$.
2. *Schwache Konvergenz äquivalent zur Konvergenz der Erzeugendenfunktionen, siehe Theorem 10.33:* Sei X_1, X_2, \dots eine Folge von Zufallsvariablen mit Werten in \mathbb{Z}_+ , so dass $\varphi_{X_n}(z) \xrightarrow{n \rightarrow \infty} \varphi(z)$ für $z \in [0, 1]$ für eine Funktion φ , die in 1 stetig von unten ist. Dann ist $X_n \xrightarrow{n \rightarrow \infty} X$ für eine Zufallsvariable X mit Erzeugendenfunktion φ .

Manchmal sind Erzeugendenfunktionen praktische Werkzeuge. Durch ihre Definition sind sie Potenzreihen mit Konvergenzradius $r \geq 1$. Man weiß, dass im Inneren des Konvergenzradius Ableitung und Summe vertauschen. Ist also etwa $r > 1$, schreiben wir

$$\varphi'_X(1) = \sum_{k=0}^{\infty} k z^{k-1} \mathbf{P}(X = k) \Big|_{z=1} = \sum_{k=0}^{\infty} k \mathbf{P}(X = k) = \mathbf{P}[X].$$

Analoge Rechnungen für höhere Ableitungen sind ebenfalls möglich.

Definition 11.3 (Asymptotische Vernachlässigbarkeit). Eine triangonale Familie von Zufallsvariablen $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ mit $m_1, m_2, \dots \in \mathbb{N}$ ist asymptotisch vernachlässigbar falls für $n = 1, 2, \dots$ die Zufallsvariablen X_{n1}, \dots, X_{n,m_n} unabhängig sind und

$$\sup_{j=1,\dots,m_n} \mathbf{P}(|X_{nj}| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0 \quad (11.2)$$

für alle $\varepsilon > 0$. Falls $X_{ij} \geq 0$ für alle i, j , so ist auch $m_n = \infty$ zugelassen.

Bemerkung 11.4 (Äquivalente Formulierung). 1. Für eine triangonale Familie von Zufallsvariablen $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ gilt (11.2) genau dann, wenn

$$\sup_{j=1,\dots,m_n} \mathbf{E}[|X_{nj}| \wedge 1] \xrightarrow{n \rightarrow \infty} 0.$$

2. Sei $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ eine triangonale von \mathbb{Z}_+ -wertigen Zufallsvariablen. Dann ist (11.2) genau dann, wenn

$$\inf_{z \in [0,1]} \inf_{j=1,\dots,m_n} \varphi_{X_{nj}}(z) = \inf_{j=1,\dots,m_n} \varphi_{X_{nj}}(0) = \inf_{j=1,\dots,m_n} \mathbf{P}(|X_{nj}| = 0) \xrightarrow{n \rightarrow \infty} 1. \quad (11.3)$$

Theorem 11.5 (Poisson Konvergenz). Sei $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ eine Familie asymptotisch vernachlässigbarer Zufallsvariablen mit Werten in \mathbb{Z}_+ und $X \sim \text{Poi}(\lambda)$. Dann gilt

$$\sum_{j=1}^{m_n} X_{nj} \xrightarrow{n \rightarrow \infty} X$$

genau dann, wenn gilt:

1. $\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 1) \xrightarrow{n \rightarrow \infty} 0$
2. $\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} = 1) \xrightarrow{n \rightarrow \infty} \lambda.$

Wir bereiten den Beweis mit einem Lemma vor.

Lemma 11.6. Sei $(\lambda_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ eine triangonale Familie asymptotisch vernachlässigbarer, nicht-negativer Konstanten und $\lambda \in [0; \infty]$. Dann gilt

$$\prod_{j=1}^{m_n} (1 - \lambda_{nj}) \xrightarrow{n \rightarrow \infty} e^{-\lambda} \iff \sum_{j=1}^{m_n} \lambda_{nj} \xrightarrow{n \rightarrow \infty} \lambda.$$

Beweis. Zunächst sei bemerkt, dass $\log(1 - x) = -x + \varepsilon(x)$ für $x > 0$ mit $\varepsilon(x)/x \xrightarrow{x \rightarrow 0} 0$. Da $\sup_{j=1,\dots,m_n} \lambda_{nj} < 1$ für große n , ist die linke Aussage äquivalent zu

$$-\lambda = \lim_{n \rightarrow \infty} \sum_{j=1}^{m_n} \log(1 - \lambda_{nj}) = - \lim_{n \rightarrow \infty} \sum_{j=1}^{m_n} \lambda_{nj} \left(1 - \frac{\varepsilon(\lambda_{nj})}{\lambda_{nj}}\right) = - \lim_{n \rightarrow \infty} \sum_{j=1}^{m_n} \lambda_{nj},$$

da

$$\sup_{j=1,\dots,m_n} \frac{\varepsilon(\lambda_{nj})}{\lambda_{nj}} \xrightarrow{n \rightarrow \infty} 0.$$

Daraus folgt die Aussage. \square

Beweis von Theorem 11.5. Wir bezeichnen mit $\varphi_{n,j}$ die Erzeugendenfunktion von $X_{n,j}$. Nach Bemerkung 11.2.2 ist die Konvergenzaussage der schwachen Konvergenz im Theorem äquivalent zur punktweisen Konvergenz von $\prod_{j=1}^{m_n} \varphi_{n,j}(z) \xrightarrow{n \rightarrow \infty} e^{-\lambda(1-z)}$, da

$$\varphi_X(z) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} z^k = e^{-\lambda(1-z)}.$$

Wegen Lemma 11.6 gilt dies genau dann, wenn

$$A_n(z) := \sum_{j=1}^{m_n} (1 - \varphi_{nj}(z)) \xrightarrow{n \rightarrow \infty} \lambda(1 - z), \quad (11.4)$$

da die Familie $(1 - \varphi_{nj}(z))_{n=1,2,\dots,j=1,\dots,m_n}$ für jedes $z \in [0, 1]$ nach (11.3) asymptotisch vernachlässigbar ist. Wir zerlegen $A(z) = A_n^1(z) + A_n^2(z)$ mit

$$A_n^1(z) = \sum_{k=1}^{\infty} (1 - z) \sum_{j=1}^{m_n} \mathbf{P}(X_{nj} = k) = (1 - z) \sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 0),$$

$$A_n^2(z) = \sum_{k=2}^{\infty} (z - z^k) \sum_{j=1}^{m_n} \mathbf{P}(X_{nj} = k).$$

Zunächst sei festgestellt, dass $z(1 - z) \leq z - z^k \leq z$ für alle $k = 2, 3, \dots$. Damit ist

$$z(1 - z) \sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 1) \leq A_n^2(z) \leq z \sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 1). \quad (11.5)$$

Wenden wir uns nun dem Beweis der Behauptung zu.

' \Rightarrow ': Es gelte also nun (11.4). Für $z = 0$ bedeutet das, da $\varphi_{nj}(0) = \mathbf{P}(X_{nj} = 0)$, dass

$$\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 0) = \sum_{j=1}^{m_n} (1 - \varphi_{nj}(0)) \xrightarrow{n \rightarrow \infty} \lambda,$$

also $A_n^1(z) \xrightarrow{n \rightarrow \infty} \lambda(1 - z)$ für $z \in [0, 1]$. Dann muss aber auch $A_n^2(z) \xrightarrow{n \rightarrow \infty} 0$ für $z \in [0, 1]$ gelten. Wegen (11.5) bedeutet dies, dass 1. gilt. Die Aussage 2. folgt daraus durch Subtraktion.

' \Leftarrow ': Es gelte also 1. und 2. Klar ist, dass $A_n^2(z) \xrightarrow{n \rightarrow \infty} 0$ wegen (11.5). Dann gilt aber auch $A_n^1(z) \xrightarrow{n \rightarrow \infty} (1 - z)\lambda$ wegen der Voraussetzung, d.h. (11.4) ist gezeigt. \square

Beispiel 11.7 (Konvergenz geometrischer Verteilungen gegen Poisson). Sei X_{nj} , $j = 1, \dots, n$, $n = 1, 2, \dots$ geometrisch verteilt mit Parameter p_n (d.h. $\mathbf{P}(X_{nj} = k) = (1 - p_n)^{k-1} p_n$, siehe Beispiel 3.2.4) und $Y_{nj} = X_{nj} - 1$. (Somit gibt Y_{nj} die Anzahl der *Misserfolge* vor dem ersten *Erfolg* an.) Wir setzen $Y_n := \sum_{j=1}^n Y_{nj}$, was so verteilt ist wie die Anzahl der Misserfolge vor dem n -ten Erfolg. Falls $Y \sim \text{Poi}(\lambda)$ und $(1 - p_n) \cdot n \xrightarrow{n \rightarrow \infty} \lambda$, so ist $Y_n \xrightarrow{n \rightarrow \infty} Y$. Denn es gilt

$$\begin{aligned} \sum_{j=1}^n \mathbf{P}(Y_{nj} = 1) &= n(1 - p_n)p_n \xrightarrow{n \rightarrow \infty} \lambda, \\ \sum_{j=1}^n \mathbf{P}(Y_{nj} > 1) &= n(1 - p_n)^2 \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

und Theorem 11.5 liefert das Ergebnis.

11.2 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz, Theorem 11.8, verallgemeinert den Satz von deMoivre Laplace. Die Verallgemeinerung besteht darin, dass beliebige Summen unabhängiger (nicht notwendig identisch verteilter) Zufallsgrößen schwach gegen eine normalverteilte Zufallsgröße konvergieren, falls sie die *Lindeberg-Bedingung* (siehe 2. in Theorem 11.8) erfüllen.

Theorem 11.8 (Zentraler Grenzwertsatz von Lindeberg-Feller). Sei $(X_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ eine Familie von Zufallsvariablen, so dass für $n = 1, 2, \dots$ die Zufallsvariablen X_{n1}, \dots, X_{nm_n} unabhängig sind. Sei außerdem

$$\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}] \xrightarrow{n \rightarrow \infty} \mu, \quad \sum_{j=1}^{m_n} \mathbf{V}[X_{nj}] \xrightarrow{n \rightarrow \infty} \sigma^2$$

und $X \sim N(\mu, \sigma^2)$. Dann sind folgende Aussagen äquivalent:

1. $\sum_{j=1}^{m_n} X_{nj} \xrightarrow{n \rightarrow \infty} X$ und $\sup_{j=1,\dots,m_n} \mathbf{V}[X_{nj}] \xrightarrow{n \rightarrow \infty} 0$,
2. $\sum_{j=1}^{m_n} \mathbf{E}[(X_{nj} - \mathbf{E}[X_{nj}])^2; |X_{nj} - \mathbf{E}[X_{nj}]| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0$ für alle $\varepsilon > 0$.

Bevor wir den zentralen Grenzwertsatz beweisen, verweisen wir auf den Spezialfall von identisch verteilten Zufallsvariablen, der bereits in der Vorlesung Stochastik behandelt wurde.

Korollar 11.9 (Zentraler Grenzwertsatz für identische verteilte Zufallsvariable). Seien X_1, X_2, \dots unabhängig und identisch verteilt mit $\mathbf{E}[X_1] = \mu, \mathbf{V}[X_1] = \sigma^2 > 0$. Sei $S_n := \sum_{k=1}^n X_k$ und $X \sim N(0, 1)$. Dann gilt

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{n \rightarrow \infty} X.$$

Beweis. Sei $m_n = n$ und $X_{nj} = \frac{X_n - \mu}{\sqrt{n\sigma^2}}$. Dann erfüllt die Familie $(X_{nj})_{n=1,2,\dots,j=1,\dots,n}$ die Voraussetzungen von Theorem 11.8 mit $\mu = 0, \sigma^2 = 1$. Außerdem gilt

$$\sum_{j=1}^n \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] = \frac{1}{\sigma^2} \mathbf{E}[(X_1 - \mu)^2; |X_1 - \mu| > \varepsilon\sqrt{n\sigma^2}] \xrightarrow{n \rightarrow \infty} 0$$

wegen majorisierter Konvergenz. \square

Oftmals ist die Lindeberg-Bedingung nicht einfach nachzuprüfen. Einfacher ist oft die stärkere Lyapunoff-Bedingung.

Bemerkung 11.10 (Lyapunoff-Bedingung). Die Familie $(X_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ aus Theorem 11.8 genügt der Lyapunoff-Bedingung, falls für ein $\delta > 0$

$$\sum_{j=1}^{m_n} \mathbf{E}[|X_{nj} - \mathbf{E}[X_{nj}]|^{2+\delta}] \xrightarrow{n \rightarrow \infty} 0.$$

Unter den Voraussetzungen von Theorem 11.8 impliziert die Lyapunoff-Bedingung die Lindeberg-Bedingung. Um dies zu sehen, sei ohne Einschränkung $\mathbf{E}[X_{nj}] = 0$. Es gilt für alle $\varepsilon > 0$

$$x^2 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta} 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta}.$$

Gilt nun die Lyapunoff-Bedingung, so folgt die Lindeberg-Bedingung aus

$$\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \leq \frac{1}{\varepsilon^\delta} \sum_{j=1}^{m_n} \mathbf{E}[|X_{nj}|^{2+\delta}] \xrightarrow{n \rightarrow \infty} 0.$$

Der Beweis von Theorem 11.8 basiert auf der geschickten Verwendung der charakteristischen Funktionen der Zufallsvariable X_{nj} und Taylor-Approximationen. Wir bereiten den Beweis des Theorems mit zwei Lemmata vor.

Lemma 11.11 (Abschätzung). Für komplexe Zahlen $z_1, \dots, z_n, z'_1, \dots, z'_n$ mit $|z_i| \leq 1, |z'_i| \leq 1$ für $i = 1, \dots, n$ gilt

$$\left| \prod_{k=1}^n z_k - \prod_{k=1}^n z'_k \right| \leq \sum_{k=1}^n |z_k - z'_k|. \quad (11.6)$$

Beweis. Für $n = 1$ ist die Gleichung offensichtlich richtig. Gilt (11.6) für ein n , so ist

$$\begin{aligned} \left| \prod_{k=1}^{n+1} z_k - \prod_{k=1}^{n+1} z'_k \right| &\leq \left| z_{n+1} \left(\prod_{k=1}^n z_k - \prod_{k=1}^n z'_k \right) \right| + \left| (z_{n+1} - z'_{n+1}) \prod_{k=1}^n z'_k \right| \\ &\leq \sum_{k=1}^n |z_k - z'_k| + |z_{n+1} - z'_{n+1}|. \end{aligned}$$

Daraus folgt die Behauptung. \square

Lemma 11.12 (Taylor-Approximation der Exponentialfunktion). Sei $t \in \mathbb{C}$ und $n \in \mathbb{Z}_+$. Dann gilt

$$\left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| \leq \frac{2|t|^n}{n!} \wedge \frac{|t|^{n+1}}{(n+1)!}. \quad (11.7)$$

Beweis. Bezeichne $h_n(t)$ die Differenz auf der linken Seite. Für $n = 0$ folgt (11.7) aus

$$|h_0(t)| = \left| \int_0^t e^{is} ds \right| \leq \int_0^t |e^{is}| ds = |t|$$

und

$$|h_0(t)| \leq |e^{it}| + 1 = 2.$$

Allgemein gilt für $t \in \mathbb{R}$, $n \in \mathbb{N}$

$$\left| \int_0^t h_n(s) ds \right| = \left| -i(e^{it} - 1) + i \sum_{k=0}^n \frac{(it)^{k+1}}{(k+1)!} \right| = \left| ie^{it} - i \sum_{k=0}^{n+1} \frac{(it)^k}{k!} \right| = |h_{n+1}(t)|,$$

woraus (11.7) mittels Induktion folgt. \square

Bemerkung 11.13 (Notation). Im folgenden Beweis werden wir für Funktionen a und b genau dann $a \lesssim b$ schreiben, falls es eine Konstante C gibt mit $a \leq Cb$.

Beweis von Theorem 11.8. O.E. sei $\mathbf{E}[X_{nj}] = \mu = 0$ und $\sigma^2 = 1$; ansonsten ersetzen wir X_{nj} durch $\frac{X_{nj} - \mathbf{E}[X_{nj}]}{\sqrt{\sigma^2}}$. Sei $\sigma_{nj}^2 := \mathbf{V}[X_{nj}]$ sowie $\sigma_n^2 := \sum_{j=1}^{m_n} \sigma_{nj}^2 \xrightarrow{n \rightarrow \infty} 1$. Bezeichne außerdem ψ_{nj} die charakteristische Funktion von X_{nj} .

2. \Rightarrow 1. Da für jedes $\varepsilon > 0$

$$\sup_{j=1, \dots, m_n} \sigma_{nj}^2 \leq \varepsilon^2 + \sup_{j=1, \dots, m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \leq \varepsilon^2 + \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \xrightarrow{n \rightarrow \infty} \varepsilon^2, \quad (11.8)$$

ist der zweite Teil von 1. bereits gezeigt.

Seien $(Z_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ unabhängige Zufallsvariablen mit $Z_{nj} \sim N(0, \sigma_{nj}^2)$. Damit ist $Z_n := \sum_{j=1}^{m_n} Z_{nj} \sim N(0, \sigma_n^2)$. Insbesondere gilt also $Z_n \xrightarrow{n \rightarrow \infty} X$, was man etwa direkt aus der Form der charakteristischen Funktionen der Normalverteilung, Beispiel 7.13.3 abliest. Sei $\tilde{\psi}_{nj}$ die charakteristische Funktion von Z_{nj} . Dann genügt es zu zeigen, siehe Theorem 10.33, dass

$$\prod_{j=1}^{n_j} \psi_{nj}(t) - \prod_{j=1}^{m_n} \tilde{\psi}_{nj}(t) \xrightarrow{n \rightarrow \infty} 0 \quad (11.9)$$

für alle t . Mittels Lemma 11.11 und Lemma 11.12 schreiben wir

$$\begin{aligned} \left| \prod_{j=1}^{m_n} \psi_{nj}(t) - \prod_{j=1}^{m_n} \tilde{\psi}_{nj}(t) \right| &\leq \sum_{j=1}^{m_n} |\psi_{nj}(t) - \tilde{\psi}_{nj}(t)| \\ &\leq \sum_{j=1}^{m_n} |\psi_{nj}(t) - 1 + \frac{1}{2}t^2\sigma_{nj}^2| + \sum_{j=1}^{m_n} |\tilde{\psi}_{nj}(t) - 1 + \frac{1}{2}t^2\sigma_{nj}^2| \\ &\lesssim 2 \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; (1 \wedge |X_{nj}|)] + \sum_{j=1}^{m_n} |e^{-\frac{1}{2}\sigma_{nj}^2 t^2} - 1 + \frac{1}{2}t^2\sigma_{nj}^2|. \end{aligned}$$

Weiter ist

$$\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2(1 \wedge |X_{nj}|)] \leq \varepsilon \sum_{j=1}^{m_n} \sigma_{nj}^2 + \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \xrightarrow{n \rightarrow \infty} \varepsilon$$

und

$$\sum_{j=1}^{m_n} |e^{-\frac{1}{2}\sigma_{nj}^2 t^2} - 1 + \frac{1}{2}t^2 \sigma_{nj}^2| \lesssim \sum_{j=1}^{m_n} \sigma_{nj}^4 \leq \sigma_n^2 \sup_{j=1, \dots, m_n} \sigma_{nj}^2 \xrightarrow{n \rightarrow \infty} 0$$

wegen (11.8). Damit ist (11.9) bereits bewiesen.

1. \Rightarrow 2. Nach dem zweiten Teil von 1. ist für jedes $\varepsilon > 0$ mit der Chebyshev-Ungleichung

$$\sup_{j=1, \dots, m_n} \mathbf{P}[|X_{nj}| > \varepsilon] \leq \sup_{j=1, \dots, m_n} \frac{\sigma_{nj}^2}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \quad (11.10)$$

Mit Lemma 11.12 gilt

$$\sup_{j=1, \dots, m_n} |\psi_{nj}(t) - 1| \leq \sup_{j=1, \dots, m_n} \mathbf{E}[2 \wedge |t \cdot X_{nj}|] \leq 2 \sup_{j=1, \dots, m_n} \mathbf{P}[|X_{nj}| > \varepsilon] + \varepsilon |t| \xrightarrow{n \rightarrow \infty} \varepsilon |t|.$$

Insbesondere ist $\sum_{j=1}^{m_n} \log \psi_{nj}(t)$ für jedes t definiert, falls n groß genug ist. Aus der Gültigkeit von 1. folgt

$$\sum_{j=1}^{m_n} \log \psi_{nj}(t) \xrightarrow{n \rightarrow \infty} -\frac{t^2}{2}. \quad (11.11)$$

Außerdem gilt wegen $\psi'_{nj}(0) = i\mathbf{E}[X_{nj}] = 0$, $\psi''_{nj}(0) = -\mathbf{V}[X_{nj}] = -\sigma_{nj}^2$ mit Hilfe einer Taylorentwicklung von ψ_{nj} um 0

$$|\psi_{nj}(t) - 1| \lesssim \sigma_{nj}^2 |t|^2$$

und

$$\begin{aligned} \left| \sum_{j=1}^{m_n} \log \psi_{nj}(t) - \sum_{j=1}^{m_n} (\psi_{nj}(t) - 1) \right| &\lesssim \sum_{j=1}^{m_n} |\psi_{nj}(t) - 1|^2 \\ &\lesssim \sum_{j=1}^{m_n} (\sigma_{nj}^2)^2 |t|^4 \lesssim |t|^4 \sup_{j=1, \dots, m_n} \sigma_{nj}^2 \xrightarrow{n \rightarrow \infty} 0. \end{aligned} \quad (11.12)$$

Da aus der Konvergenz einer imaginären Reihe die Konvergenz ihres Realteils folgt, folgern wir aus (11.11) und (11.12) wegen $\operatorname{Re}(\psi_{nj}(t)) = \mathbf{E}[\cos(tX_{nj})]$

$$\sum_{j=1}^{m_n} \mathbf{E}[\cos(tX_{nj}) - 1] \xrightarrow{n \rightarrow \infty} -\frac{t^2}{2}$$

Für $\varepsilon > 0$ ist nun wegen $0 \leq 1 - \cos(\theta) \leq \frac{\theta^2}{2}$

$$\begin{aligned}
0 \leq \limsup_{n \rightarrow \infty} \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] &= \limsup_{n \rightarrow \infty} 1 - \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| \leq \varepsilon] \\
&\leq \limsup_{n \rightarrow \infty} 1 - \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{E}[1 - \cos(tX_{nj}); |X_{nj}| \leq \varepsilon] \\
&= \limsup_{n \rightarrow \infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{E}[1 - \cos(tX_{nj}); |X_{nj}| > \varepsilon] \quad (11.13) \\
&\leq \limsup_{n \rightarrow \infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{P}[|X_{nj}| > \varepsilon] \\
&\leq \frac{2}{\varepsilon^2 t^2} \limsup_{n \rightarrow \infty} \sum_{j=1}^{m_n} \sigma_{nj} = \frac{2}{\varepsilon^2 t^2}.
\end{aligned}$$

Da $t, \varepsilon > 0$ willkürlich waren, ist 2. gezeigt, wenn man in der letzten Ungleichungskette $t \infty$ betrachtet. \square

11.3 Mehrdimensionale Grenzwertsätze

Bisher haben wir schwache Grenzwertsätze nur für den Fall \mathbb{R} -wertiger Zufallsvariablen betrachtet. Wir verallgemeinern dies nun zu \mathbb{R}^d -wertigen Zufallsgrößen. Insbesondere geben wir eine Variante des mehrdimensionalen zentralen Grenzwertsatzes an.

Definition 11.14 (Mehrdimensionale Normalverteilung). Seien $\mu \in \mathbb{R}^d$ und $C \in \mathbb{R}^{d \times d}$ eine strikt positiv definite symmetrische Matrix.^{16,17} Die d -dimensionale Normalverteilung mit Erwartungswert μ und Kovarianzmatrix C ist das Wahrscheinlichkeitsmaß $N_{\mu, C}$ auf \mathbb{R}^d mit Dichte

$$f_{\mu, C}(x) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\left(-\frac{1}{2}(x - \mu)C^{-1}(x - \mu)^\top\right).$$

Proposition 11.15 (Eigenschaften der mehrdimensionalen Normalverteilung). Seien $\mu \in \mathbb{R}^d$, $C = AA^\top \in \mathbb{R}^{d \times d}$ eine strikt positiv definite symmetrische Matrix und I die d -dimensionale Einheitsmatrix. Es sind äquivalent:

1. $X \sim N_{\mu, C}$
2. $tX^\top \sim N_{t\mu^\top, tCt^\top}$ für jedes $t \in \mathbb{R}^d$
3. $\psi_X(t) = e^{it\mu^\top} e^{-\frac{1}{2}tCt^\top}$ für jedes $t \in \mathbb{R}^d$.

In jedem dieser Fälle gilt

4. $X \stackrel{d}{=} AY + \mu$ für $Y \sim N_{0, I}$
5. $\mathbf{E}[X_i] = \mu_i$ für $i = 1, \dots, d$

¹⁶Wir bezeichnen hier Zeilenvektoren mit x und Spaltenvektoren mit x^\top .

¹⁷Strikt positiv definit bedeutet $xCx^\top > 0$ für alle $x \in \mathbb{R}^d$. Aus der linearen Algebra ist bekannt, dass es für eine strikt positiv definite Matrix C immer eine invertierbare Matrix A gibt mit $C = AA^\top$.

6. $\mathbf{COV}[X_i, X_j] = C_{ij}$ für $i, j = 1, \dots, d$

Beweis. Sei zunächst $X \sim N_{\mu, C}$. Wir zeigen zunächst 4.-6. Die Eigenschaft 4. ist eine Anwendung des Transformationsatzes. Für $B \in \mathcal{B}(\mathbb{R}^d)$ und $T : y \mapsto Ay^\top + \mu^\top$ ist

$$\begin{aligned} N_{0, I}(T^{-1}(B)) &= \frac{1}{\sqrt{(2\pi)^d}} \int_{T^{-1}(B)} e^{-\frac{1}{2}yy^\top} dy \\ &\stackrel{y=A^{-1}(x-\mu)}{=} \frac{1}{\sqrt{(2\pi)^d}} \frac{1}{\det A} \int_B \exp\left(-\frac{1}{2}(x-\mu)(A^\top)^{-1}A^{-1}(x-\mu)^\top\right) dx \\ &= \frac{1}{\sqrt{(2\pi)^d \det C}} \int_B \exp\left(-\frac{1}{2}(x-\mu)C^{-1}(x-\mu)^\top\right) dx \\ &= N_{\mu, C}(B). \end{aligned}$$

5. folgt aus 4. mit

$$\mathbf{E}[X_i] = \mathbf{E}[\pi_i(AY + \mu)] = \pi_i\mu = \mu_i,$$

wobei π_i die Projektion auf die i -te Koordinate ist.

6. folgt ebenso aus 4. mit

$$\begin{aligned} \mathbf{COV}[X_i, X_j] &= \mathbf{E}[(\pi_i AY^\top)(\pi_j AY^\top)] = \mathbf{E}[(A_i Y^\top)(A_j Y^\top)] = \mathbf{E}[A_i Y^\top Y A_j^\top] \\ &= A_i \cdot A_j^\top = (AA^\top)_{ij} = C_{ij}. \end{aligned}$$

Wir kommen nun zur Äquivalenz von 1.-3.: '1. \Rightarrow 2.': Da $X \stackrel{d}{=} AY^\top + \mu^\top$ wie in 4., ist $tX^\top = tAY^\top + t\mu^\top$ als Linearkombination von (eindimensionalen) Normalverteilungen wieder normalverteilt. Der Erwartungswert ist offenbar $t\mu^\top$ und die Varianz

$$\mathbf{V}[tX^\top] = \mathbf{E}[(tAY^\top)^2] = \mathbf{E}[tAY^\top Y A^\top t^\top] = tAA^\top t^\top = tCt^\top.$$

'2. \Rightarrow 3.': Da $tX^\top \sim N_{t\mu^\top, tCt^\top}$, folgt die Aussage aus Beispiel 7.13.3.

'3. \Rightarrow 1.': Dies folgt aus Proposition 10.25. \square

Bemerkung 11.16 (Spezialfälle). 1. Falls C in Definition 11.14 zwar positiv, aber nicht strikt positiv definit ist (d.h. es gibt $x \in \mathbb{R}^d$, $x \neq 0$ mit $xCx = 0$), kann man $N_{\mu, C}$ nicht durch Angabe der Dichte wie in obiger Definition bestimmen. In diesem Fall definiert man $N_{\mu, C}$ durch Angabe der charakteristischen Funktion, d.h. $N_{\mu, C}$ ist die eindeutig bestimmte Verteilung auf \mathbb{R}^d mit $\psi_{N_{\mu, C}}(t) = e^{it\mu} e^{-\frac{1}{2}tCt^\top}$.

2. Ist $Y \sim N_{0, I}$ und A eine orthogonale Matrix, so ist auch $X := AY \sim N_{0, I}$. Dies folgt aus Proposition 11.15, wenn man $I = AA^\top$ schreibt und 4. benutzt.

Proposition 11.17 (Cramér-Wold Device). Sind X, X_1, X_2, \dots Zufallsvariablen mit Werten in \mathbb{R}^d . Dann gilt $X_n \xrightarrow{n \rightarrow \infty} X$ genau dann, falls $tX_n \xrightarrow{n \rightarrow \infty} tX$ für alle $t \in \mathbb{R}^d$ (wobei $(t, x) \mapsto tx$ das Skalarprodukt im \mathbb{R}^d ist).

Beweis. ' \Rightarrow ': Sei $t \in \mathbb{R}^d$ und $f \in \mathcal{C}_b(\mathbb{R})$. Dann ist $f(\cdot) \in \mathcal{C}_b(\mathbb{R}^d)$. Damit gilt $\mathbf{E}[f(tX_n)] \xrightarrow{n \rightarrow \infty} \mathbf{E}[f(tX)]$, d.h. $tX_n \xrightarrow{n \rightarrow \infty} tX$.

' \Leftarrow ': Sei π_i die Projektion auf die i -te Koordinate. Da $(\pi_i X_n)_{n=1, 2, \dots}$ nach Korollar 10.18 straff für alle i ist, sieht man, dass $(X_n)_{n=1, 2, \dots}$ straff ist. Da $\{x \mapsto e^{itx} : t \in \mathbb{R}^d\}$ eine separierende Funktionenklasse ist, folgt die Behauptung aus $\mathbf{E}[e^{itX_n}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[e^{itX}]$ für alle $t \in \mathbb{R}^d$ und Proposition 10.27. \square

Theorem 11.18 (Mehrdimensionaler zentraler Grenzwertsatz). *Seien X_1, X_2, \dots unabhängige, identische verteilte Zufallsvariablen mit Werten in \mathbb{R}^d mit $\mathbf{E}[X_n] = \mu \in \mathbb{R}^d$ und $\mathbf{COV}[X_{n,i}, X_{n,j}] = C_{ij}$ für $i, j = 1, \dots, d$ und $S_n = \sum_{i=1}^n X_i$. Ist $X \sim N_{0,C}$, so gilt*

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} X.$$

Beweis. Wir wenden den eindimensionalen zentralen Grenzwertsatz, Korollar 11.9, auf die unabhängigen, identisch verteilten Zufallsvariablen tX_1, tX_2, \dots an. Dieser liefert

$$t \frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} tX.$$

Da t beliebig war, folgt die Aussage aus Proposition 11.17. \square

12 Die bedingte Erwartung

Sei $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Wir schreiben $\mathcal{L}^1 := \mathcal{L}^1(\mathbf{P})$ für die Menge aller reellen Zufallsvariablen, deren Erwartungswert existiert. In diesem Kapitel verwenden wir wieder die Notation $\mathbb{E}[\cdot]$ für das Integral bezüglich des Wahrscheinlichkeitsmaßes \mathbf{P} , sowie $\mathcal{L}^p := \mathcal{L}^p(\mathbf{P})$.

12.1 Motivation

Definiere wie in der elementaren Stochastik für $A, G \in \mathcal{A}$ und $\mathbf{P}(G) > 0$

$$\mathbf{P}(A|G) := \frac{\mathbf{P}(A \cap G)}{\mathbf{P}(G)}$$

und analog die *bedingte Erwartung*

$$\mathbf{E}[X|G] := \frac{\mathbf{E}[X; G]}{\mathbf{P}(G)}.$$

Dann gilt $\mathbf{P}(A|G) = \mathbf{E}[1_A|G]$. Dieser Zusammenhang bedeutet, dass man bedingte Erwartungen dazu verwenden kann, bedingte Wahrscheinlichkeiten zu berechnen. Insbesondere ist der Begriff der bedingten Erwartung allgemeiner als der Begriff der bedingten Wahrscheinlichkeit.

Wir werden in diesem Kapitel die bedingte Erwartung $\mathbf{E}[X|\mathcal{G}]$ für eine Zufallsvariable X und eine σ -Algebra $\mathcal{G} \subseteq \mathcal{F}$ kennen lernen. Hierbei wird $\mathbf{E}[X|\mathcal{G}]$ eine \mathcal{G} -messbare Zufallsvariable sein. Als einfaches Beispiel sei $\{G_1, G_2, \dots\} \subseteq \mathcal{F}$ eine Partition von Ω mit $\mathbf{P}(G_i) > 0$ für $i = 1, 2, \dots$ und \mathcal{G} die erzeugte σ -Algebra. Dann setzen wir für $X \in \mathcal{L}^1$

$$\mathbf{E}[X|\mathcal{G}](\omega) := \sum_{i=1}^{\infty} \mathbf{E}[X|G_i] \cdot 1_{G_i}(\omega). \quad (12.1)$$

Es gilt also: für $\omega \in G_i$ ist die Zufallsvariable $\mathbf{E}[X|\mathcal{G}]$ gegeben durch $\mathbf{E}[X|\mathcal{G}](\omega) = \mathbf{E}[X|G_i] = \mathbf{E}[X; G_i]/\mathbf{P}(G_i)$. Insbesondere ist sie auf G_i konstant, $i = 1, 2, \dots$. Mit anderen Worten ist

$\mathbf{E}[X|\mathcal{G}]$ messbar bezüglich \mathcal{G} . Weiter gilt etwa für $J \subseteq \mathbb{N}$ und $A = \bigcup_{j \in J} G_j \in \mathcal{G}$

$$\begin{aligned} \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] &= \mathbf{E}\left[\sum_{i=1}^{\infty} \mathbf{E}[X|G_i] 1_{G_i} 1_A\right] \\ &= \sum_{j \in J} \mathbf{E}[\mathbf{E}[X|G_j] 1_{G_j}] \\ &= \sum_{j \in J} \mathbf{E}[X|G_j] \cdot \mathbf{P}(G_j) \\ &= \mathbf{E}[X; A]. \end{aligned} \tag{12.2}$$

Insbesondere gilt mit $J = \mathbb{N}$ also $\mathbf{E}[\mathbf{E}[X|\mathcal{F}]] = \mathbf{E}[X]$. Die Definition der bedingten Erwartung (12.1) lässt sich mit Hilfe der Eigenschaft (12.2) auf beliebige σ -Algebren $\mathcal{G} \subseteq \mathcal{F}$ verallgemeinern.

Beispiel 12.1 (Binomialverteilung mit zufälliger Erfolgswahrscheinlichkeit). Sei X gleichverteilt auf $[0, 1]$, d.h. die Verteilung von X hat Dichte $1_{[0,1]}$. Gegeben $X = x$ sei Y_1, \dots, Y_n eine Folge von Bernoulli-verteilten Zufallsvariablen mit Erfolgswahrscheinlichkeit x . Also ist $Y = Y_1 + \dots + Y_n$ binomialverteilt mit n und x , d.h. Y zählt die Anzahl der Erfolge in n unabhängigen Experimenten mit Erfolgswahrscheinlichkeit x . Intuitiv ist klar, dass das

$$\mathbf{P}(Y = k|X) = \binom{n}{k} X^k (1 - X)^{n-k}$$

bedeuten sollte. Dies ist allerdings bisher nicht definiert, da $\mathbf{P}(X = x) = 0$ gilt. Bemerkenswert ist jedoch, dass die rechte Seite eine $\sigma(X)$ -messbare Zufallsvariable ist.

12.2 Definition und Eigenschaften

Wir definieren nun formal die bedingte Erwartung $\mathbf{E}[X|\mathcal{G}]$ für $\mathcal{G} \subseteq \mathcal{F}$. Wie oben schon erwähnt, ist dies eine \mathcal{G} -messbare Zufallsvariable, deren Erwartungen wie in (12.2) mit denen von X übereinstimmen.

Theorem 12.2 (Existenz und Eigenschaften der bedingten Erwartung). Sei $\mathcal{G} \subseteq \mathcal{F}$ eine σ -Algebra. Dann gibt es einen fast sicher eindeutigen, linearen Operator $\mathbf{E}[\cdot|\mathcal{G}] : \mathcal{L}^1 \rightarrow \mathcal{L}^1$, so dass $\mathbf{E}[X|\mathcal{G}]$ für alle $X \in \mathcal{L}^1$ eine \mathcal{G} -messbare Zufallsvariable ist mit

1. $\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A]$ für alle $A \in \mathcal{G}$.

Weiter gilt

2. $\mathbf{E}[X|\mathcal{G}] \geq 0$, falls $X \geq 0$.
3. $\mathbf{E}[|\mathbf{E}[X|\mathcal{G}]|] \leq \mathbf{E}[|X|]$.
4. Falls $0 \leq X_n \uparrow X$ für $n \rightarrow \infty$, so ist auch $\mathbf{E}[X_n|\mathcal{G}] \uparrow \mathbf{E}[X|\mathcal{G}]$ in \mathcal{L}^1 , falls alle Erwartungen existieren.
5. Falls X eine \mathcal{G} -messbare Funktion ist, so gilt $\mathbf{E}[XY|\mathcal{G}] = X\mathbf{E}[Y|\mathcal{G}]$, falls alle Erwartungen existieren.

6. $\mathbf{E}[X\mathbf{E}[Y|\mathcal{G}]] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]Y] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]\mathbf{E}[Y|\mathcal{G}]]$, falls alle Erwartungen existieren.

7. Ist $\mathcal{H} \subseteq \mathcal{G}$, so ist $\mathbf{E}[\mathbf{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbf{E}[X|\mathcal{H}]$.

8. Ist X unabhängig von \mathcal{G} , so ist $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.

Beweis. 1. im Fall $X \in \mathcal{L}^2$: Sei M der abgeschlossene lineare Teilraum von \mathcal{L}^2 , der aus allen Funktionen besteht, die bis auf eine Nullmenge mit einer \mathcal{G} -messbaren Funktion übereinstimmen. Nach Proposition 5.9 gibt es nun fast sicher eindeutige Funktionen $Y \in M, Z \perp M$ mit $X = Y + Z$. Wir definieren $\mathbf{E}[X|\mathcal{G}] := Y$. Damit gilt $X - \mathbf{E}[X|\mathcal{G}] \perp M$, also $\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]; A] = 0$ für $A \in \mathcal{G}$, woraus 1. für $X \in \mathcal{L}^2$ folgt.

3. im Fall $X \in \mathcal{L}^2$: Wähle $A := \{\mathbf{E}[X|\mathcal{G}] \geq 0\}$. Nach 1. gilt dann

$$\mathbf{E}[|\mathbf{E}[X|\mathcal{G}]|] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] - \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A^c] = \mathbf{E}[X; A] - \mathbf{E}[X; A^c] \leq \mathbf{E}[|X|].$$

1. im Fall $X \in \mathcal{L}^1$: Ist $X \in \mathcal{L}^1 \supset \mathcal{L}^2$, so wähle $X_1, X_2, \dots \in \mathcal{L}^2$ mit $\|X_n - X\|_1 \xrightarrow{n \rightarrow \infty} 0$ (etwa so, dass $|X_n| := |X| \wedge n$), und definiere $\mathbf{E}[X|\mathcal{G}] := \lim_{n \rightarrow \infty} \mathbf{E}[X_n|\mathcal{G}]$. Dieser Grenzwert existiert in \mathcal{L}^1 , da wegen 3.

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X_m|\mathcal{G}]|] = \mathbf{E}[|\mathbf{E}[X_n - X_m|\mathcal{G}]|] \leq \mathbf{E}[|X_n - X_m|] \xrightarrow{n, m \rightarrow \infty} 0$$

die Folge $(\mathbf{E}[X_n|\mathcal{G}])_{n=1,2,\dots}$ eine Cauchy-Folge ist und \mathcal{L}^1 vollständig ist. Außerdem gilt damit $\|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]\|_1 \xrightarrow{n \rightarrow \infty} 0$. Weiter ist für $A \in \mathcal{G}$

$$\begin{aligned} |\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]; A]| &\leq \mathbf{E}[|X1_A - X_n1_A|] \\ &\quad + |\mathbf{E}[X_n - \mathbf{E}[X_n|\mathcal{G}]; A]| \\ &\quad + \mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}]1_A - \mathbf{E}[X|\mathcal{G}]1_A|] \\ &\xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

wegen majorisierter Konvergenz und 1. folgt im Fall $X \in \mathcal{L}^1$.

3. im Fall $X \in \mathcal{L}^1$. Auch hier sieht man durch ein Approximationsargument, falls $X_1, X_2, \dots \in \mathcal{L}^2$ mit $X_n \xrightarrow{n \rightarrow \infty} \mathcal{L}^1 X$

$$\mathbf{E}[|\mathbf{E}[X|\mathcal{G}]|] = \lim_{n \rightarrow \infty} \mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}]|] \leq \lim_{n \rightarrow \infty} \mathbf{E}[|X_n|] = \mathbf{E}[|X|],$$

da wegen der umgekehrten Dreiecksungleichung etwa

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}]| - |\mathbf{E}[X|\mathcal{G}]|] \leq \mathbf{E}[|\mathbf{E}[X|\mathcal{G}] - \mathbf{E}[X_n|\mathcal{G}]|] \xrightarrow{n \rightarrow \infty} 0.$$

2. Setze $A = \{\mathbf{E}[X|\mathcal{G}] \leq 0\}$ und damit

$$0 \geq \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = 0,$$

also wegen $\mathbf{E}[X|\mathcal{G}]1_A \leq 0$ auch $\mathbf{E}[X|\mathcal{G}]1_A = 0$ fast sicher.

4. Wegen monotoner Konvergenz ist $\|X_n - X\|_1 \xrightarrow{n \rightarrow \infty} 0$, also nach 3.

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]|] = \mathbf{E}[|\mathbf{E}[X_n - X|\mathcal{G}]|] \leq \mathbf{E}[|X_n - X|] \xrightarrow{n \rightarrow \infty} 0.$$

6. im Fall $X, Y \in \mathcal{L}^2$. Nach der Definition der bedingten Erwartung ist $\mathbf{E}[X|\mathcal{G}], \mathbf{E}[Y|\mathcal{G}] \in M$, wenn M der lineare Teilraum von \mathcal{L}^2 besteht, der Funktionen beinhaltet, die bis auf eine

Nullmenge mit einer \mathcal{G} -messbaren Funktion übereinstimmen. Außerdem ist $X - \mathbf{E}[X|\mathcal{G}] \perp M$.
Damit ist

$$\mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}])\mathbf{E}[Y|\mathcal{G}]] = 0.$$

6. im Fall $X, Y \in \mathcal{L}^1$. Wähle $X_1, Y_1, X_2, Y_2, \dots \in \mathcal{L}^2$ mit $X_n \uparrow X, Y_n \uparrow Y$. Wegen 4. und majorisierter Konvergenz gilt dann, falls alle Erwartungen existieren,

$$\mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}])\mathbf{E}[Y|\mathcal{G}]] = \lim_{n \rightarrow \infty} \mathbf{E}[(X_n - \mathbf{E}[X_n|\mathcal{G}])\mathbf{E}[Y_n|\mathcal{G}]] = 0.$$

5. Wegen 1. ist $\mathbf{E}[X|\mathcal{G}]1_A = X1_A$ für $A \in \mathcal{G}$ fast sicher. Damit ist auch

$$\mathbf{E}[XY; A] = \mathbf{E}[X\mathbf{E}[Y|\mathcal{G}]; A]$$

nach 6. Hieraus folgt nach 1. bereits $\mathbf{E}[XY|\mathcal{G}] = X\mathbf{E}[Y|\mathcal{G}]$.

7. Da $\mathcal{H} \subseteq \mathcal{G}$, ist für $A \in \mathcal{H}$

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = \mathbf{E}[\mathbf{E}[X|\mathcal{H}]; A]$$

nach 1. Hier aus folgt aber $\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; \mathcal{H}] = \mathbf{E}[X|\mathcal{H}]$.

8. Sicherlich ist $\mathbf{E}[X]$ messbar bezüglich \mathcal{G} . Für $A \in \mathcal{G}$ ist außerdem

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = \mathbf{E}[X]\mathbf{E}[1_A] = \mathbf{E}[\mathbf{E}[X]; A]$$

und damit $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$. □

Bemerkung 12.3 (Interpretation und alternativer Beweis). 1. Sei $X \in \mathcal{L}^2$. Wie der Beweis von 1. in Theorem 12.2 zeigt, ist $X - \mathbf{E}[X|\mathcal{G}]$ senkrecht auf dem linearen Teilraum aller \mathcal{G} -messbaren Funktionen. Insbesondere ist $\mathbf{E}[X|\mathcal{G}]$ diejenige \mathcal{G} -messbare Zufallsvariable, die (im Sinne der \mathcal{L}^2 -Norm) der Zufallsvariable X am nächsten kommt. Deswegen kann man sagen, dass $\mathbf{E}[X|\mathcal{G}]$ die beste Schätzung von X ist, wenn Informationen aus der σ -Algebra \mathcal{G} zur Verfügung stehen.

2. Die fast sicher eindeutige Existenz der bedingten Erwartung mit der Eigenschaft 1. in Theorem 12.2 kann man anders als oben mit Hilfe des Satzes von Radon-Nikodým (Korollar 5.16) beweisen:

Sei zunächst $X \geq 0$. Setze $\tilde{\mathbf{P}} := \mathbf{P}|_{\mathcal{G}}$, die Einschränkung von \mathbf{P} auf \mathcal{G} , und $\mu(\cdot) := \tilde{\mathbf{E}}[X; \cdot]$ ein endliches Maß. Dann gilt offenbar $\mu \ll \tilde{\mathbf{P}}$. Der Satz von Radon-Nikodým stellt sicher, dass μ eine Dichte bzgl. $\tilde{\mathbf{P}}$ hat, d.h. es eine \mathcal{G} -messbare Zufallsvariable Z gibt mit

$$\mathbf{E}[X; A] = \tilde{\mathbf{E}}[X; A] = \mu(A) = \tilde{\mathbf{E}}[Z; A] = \mathbf{E}[Z; A]$$

für alle $A \in \mathcal{G}$. Damit erfüllt Z die Eigenschaften von 1. aus Theorem 12.2. Der allgemeinen Fall (d.h. X kann auch negative Werte annehmen) folgt dann mit der Zerlegung $X = X^+ - X^-$.

Zum Beweis der (fast sicheren) Eindeutigkeit der bedingten Erwartung sei Z' eine weitere, \mathcal{G} -messbare Zufallsvariable mit $\mathbf{E}[Z'; A] = \mathbf{E}[X; A]$ für alle $A \in \mathcal{G}$. Dann ist $B := \{Z' - \mathbf{E}[X|\mathcal{G}] > 0\} \in \mathcal{G}$ und $\mathbf{E}[\mathbf{E}[X|\mathcal{G}] - Z'; B] = \mathbf{E}[X - X; B] = 0$ und ebenso $\mathbf{E}[\mathbf{E}[X|\mathcal{G}] - Z'; B^c] = 0$. Das heißt also $Z' = \mathbf{E}[X|\mathcal{G}]$ fast sicher.

Proposition 12.4 (Jensen'sche Ungleichung für bedingte Erwartungen). Sei I ein offenes Intervall, $\mathcal{G} \subseteq \mathcal{A}$ und $X \in \mathcal{L}^1$ mit Werten in I und $\varphi : I \rightarrow \mathbb{R}$ konvex. Dann gilt

$$\mathbf{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbf{E}[X|\mathcal{G}]).$$

Beweis. Der Beweis verläuft analog zu dem der Jensen'schen Ungleichung im unbedingten Fall, Proposition 7.6: Da I offen ist, liegt $\mathbf{E}[X|\mathcal{G}] \in I$ fast sicher. Wir erinnern an die Definition von λ in (7.2). Weiter ist, wie in (7.3) für $x \in I$

$$\varphi(x) \geq \varphi(\mathbf{E}[X|\mathcal{G}]) + \lambda(\mathbf{E}[X|\mathcal{G}])(x - \mathbf{E}[X|\mathcal{G}])$$

und damit

$$\begin{aligned} \mathbf{E}[\varphi(X)|\mathcal{G}] &\geq \mathbf{E}[\varphi(\mathbf{E}[X|\mathcal{G}])|\mathcal{G}] + \mathbf{E}[\lambda(\mathbf{E}[X|\mathcal{G}]) \cdot (X - \mathbf{E}[X|\mathcal{G}])|\mathcal{G}] \\ &= \varphi(\mathbf{E}[X|\mathcal{G}]). \end{aligned} \quad \square$$

Lemma 12.5 (Gleichgradige Integrierbarkeit und bedingte Erwartung). Sei $X \in \mathcal{L}^1$. Dann ist die Familie $(\mathbf{E}[X|\mathcal{G}])_{\mathcal{G} \subseteq \mathcal{A}}$ gleichgradig integrierbar.

Beweis. Da $\{X\}$ gleichgradig integrierbar ist, gibt es nach Lemma 8.9 eine monoton wachsende konvexe Funktion $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ mit $\frac{\varphi(x)}{x} \xrightarrow{x \rightarrow \infty} \infty$ und $\mathbf{E}[\varphi(|X|)] < \infty$. Mit Theorem 12.2.3 folgt

$$\sup_{\mathcal{F} \subseteq \mathcal{A}} \mathbf{E}[\varphi(|\mathbf{E}[X|\mathcal{F}]|)] \leq \mathbf{E}[\varphi(|X|)] < \infty.$$

Damit ist $\{\mathbf{E}[X|\mathcal{F}] : \mathcal{F} \subseteq \mathcal{A} \text{ } \sigma\text{-Algebra}\}$ gleichgradig integrierbar, wieder nach Lemma 8.9. \square

Theorem 12.6 (Majorisierte und monotone Konvergenz für bedingte Erwartungen). Sei $\mathcal{G} \subseteq \mathcal{F}$ und $X_1, X_2, \dots \in \mathcal{L}^1$. Es gelte eine der Bedingungen:

1. Sei $X \in \mathcal{L}^1$, so dass $X_n \uparrow X$ fast sicher.
2. Ist $Y \in \mathcal{L}^1$, so dass $|X_n| \leq |Y|$ für alle n , und $X_n \xrightarrow{n \rightarrow \infty} X$ fast sicher.

Dann gilt

$$\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[X|\mathcal{G}]$$

fast sicher und in \mathcal{L}^1 .

Beweis. Für die \mathcal{L}^1 -Konvergenz hat man in beiden Fällen mit Theorem 12.2.3

$$\begin{aligned} \mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]|] &= \mathbf{E}[|\mathbf{E}[X_n - X|\mathcal{G}]|] \\ &\leq \mathbf{E}[|X_n - X|] \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Die fast sichere Konvergenz teilen wir in die beiden Fälle auf: im Fall 1. ist nach Theorem 12.2.2 klar, dass $\mathbf{E}[X_n|\mathcal{G}]$ monoton wächst. Außerdem ist für $A \in \mathcal{F}$ mit dem Satz der monotonen Konvergenz

$$\mathbf{E}\left[\sup_n \mathbf{E}[X_n|\mathcal{G}]; A\right] = \sup_n \mathbf{E}[\mathbf{E}[X_n|\mathcal{G}]; A] = \sup_n \mathbf{E}[X_n; A] = \mathbf{E}[\sup_n X_n; A] = \mathbf{E}[X; A].$$

Damit ist aber gezeigt, dass fast sicher $\sup_n \mathbf{E}[X_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$ gilt.

Im Fall 2. setzen wir

$$Y_n := \sup_{k \geq n} X_k \downarrow \limsup_n X_n = X \text{ fast sicher,}$$

$$Z_n := \inf_{k \geq n} X_k \uparrow \liminf_n X_n = X \text{ fast sicher.}$$

Damit ist $-Y \leq Z_n \leq X_n \leq Y_n \leq Y$, also insbesondere $Y_1, Z_1, Y_2, Z_2, \dots \in \mathcal{L}^1$, also ist nach 1.

$$\mathbf{E}[X|\mathcal{G}] = \lim_{n \rightarrow \infty} \mathbf{E}[Z_n|\mathcal{G}] \leq \lim_{n \rightarrow \infty} \mathbf{E}[X_n|\mathcal{G}] \leq \lim_{n \rightarrow \infty} \mathbf{E}[Y_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$$

fast sicher. Insbesondere ist also $\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[X|\mathcal{G}]$ fast sicher. \square

12.3 Der Fall $\mathcal{G} = \sigma(X)$

Im Falle $\mathcal{G} = \sigma(X)$ bedeutet $\mathbf{E}[Y|X] := \mathbf{E}[Y|\sigma(X)]$ die Erwartung von Y , gegeben, dass die Zufallsvariable X festgelegt ist. Dies ist eine Funktion von X , wie Proposition 12.7 zeigt.

Proposition 12.7 (Bedingung auf eine Zufallsvariable). *Sei (Ω', \mathcal{F}') ein Messraum, X eine Ω' -wertige Zufallsvariable mit Werten in Ω' und $Y \in \mathcal{L}^1$. Dann existiert eine $\mathcal{F}'/\mathcal{B}(\mathbb{R})$ -messbare Abbildung $\varphi : \Omega' \rightarrow \mathbb{R}$ mit $\mathbf{E}[Y|X] = \varphi(X)$.*

Beweis. Klar nach Lemma 7.2 \square

Beispiel 12.8 (Zufällige Erfolgswahrscheinlichkeit). Betrachten wir die in Beispiel 12.1 gestellte Frage nach der Existenz der bedingten Wahrscheinlichkeit $\mathbf{P}(Y = k|X)$, wobei X uniform auf $[0, 1]$ ist und X unabhängig binomial verteilt mit n und X . Wir zeigen nun, dass

$$\mathbf{P}(Y = k|X) = \binom{n}{k} X^k (1 - X)^{n-k}. \quad (12.3)$$

Sei $A = \{X \in I\}$ für $I \in \mathcal{B}([0, 1])$, d.h. A ist eine $\sigma(X)$ -messbare Menge. Dann gilt

$$\mathbf{E}[1_{Y=k}; A] = \mathbf{P}(Y = k, X \in I) = \int_I \binom{n}{k} x^k (1 - x)^{n-k} dx = \mathbf{E}\left[\binom{n}{k} X^k (1 - X)^{n-k}; A\right]$$

Dies bedeutet aber, dass (12.3) stimmt.

Beispiel 12.9 (Summen unabhängiger identisch verteilter Zufallsvariable). Seien X_1, X_2, \dots eine Folge unabhängiger, identisch verteilter Zufallsvariablen, $\mu = \mathbf{E}[X_1]$ und $S_n := X_1 + \dots + X_n$. Dann ist

$$\mathbf{E}[S_n|X_1] = \mathbf{E}[X_1|X_1] + \mathbf{E}[X_2 + \dots + X_n|X_1] = X_1 + (n - 1)\mu,$$

$$\mathbf{E}[X_1|S_n] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i|S_n] = \frac{1}{n} \mathbf{E}[S_n|S_n] = \frac{1}{n} S_n.$$

In der zweiten Rechnung ist also beispielsweise für $X = S_n$ und $Y = X_1$ die Funktion φ aus Proposition 12.7 gegeben durch $\varphi(x) = \frac{1}{n}x$.

Beispiel 12.10 (Buffon's Nadelproblem). Auf einer Ebene liegen Geraden im horizontalen Abstand 1. Es werden Nadeln der Länge 1 auf die Ebene geworfen; siehe Abbildung 12.1. Betrachten wir eine Nadel. Wir setzen

$$Z := \begin{cases} 1, & \text{falls die Nadel eine Gerade schneidet} \\ 0, & \text{sonst} \end{cases}.$$

Dabei ist der Mittelpunkt der Nadel X von der linken Geraden entfernt und die Verlängerung

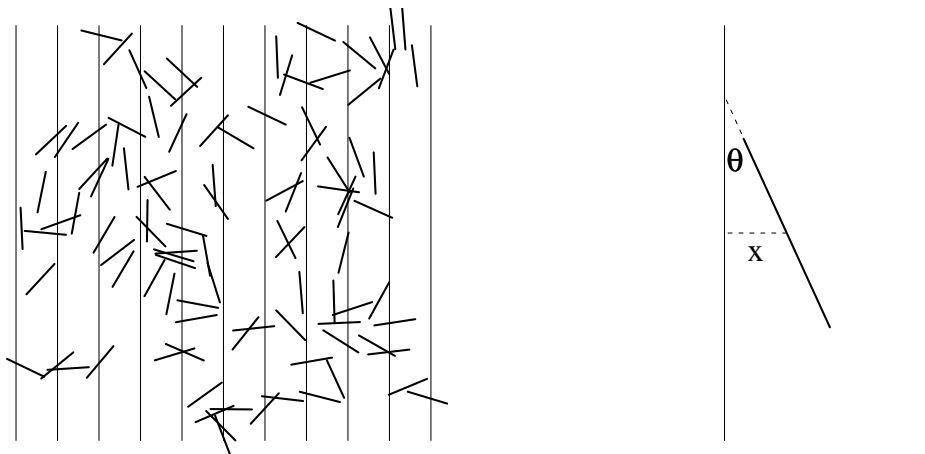


Abbildung 12.1: Skizze zu Buffon's Nadelproblem.

der Nadel geht einen spitzen Winkel Θ mit der linken Gerade ein. Damit ist X uniform auf $[0; 1]$, Θ uniform auf $[0; \frac{\pi}{2}]$ unabhängig und es gilt

$$\mathbf{P}(Z = 1|\Theta) = \mathbf{P}(X \leq \frac{1}{2} \sin(\Theta) \text{ oder } X \geq 1 - \frac{1}{2} \sin(\Theta)|\Theta) = \sin(\Theta).$$

Damit ist

$$\mathbf{P}(Z = 1) = \mathbf{E}[\mathbf{P}(Z = 1|\Theta)] = \mathbf{E}[\sin(\Theta)] = \frac{2}{\pi} \int_0^{\pi/2} (\sin(\theta)d\theta) = \frac{2}{\pi}.$$

Dies kann man so interpretieren: will man durch Simulation (d.h. also durch ein Monte-Carlo Verfahren) den numerischen Wert von π herausfinden, kann man Buffon's Nadeln simulieren. Da jede einzelne Nadel die Wahrscheinlichkeit $\frac{2}{\pi}$ hat, eine vertikale Linien zu treffen, ist etwa

$$\pi \approx \frac{2}{\text{Anteil der Nadeln, die eine Vertikale treffen}}$$

nach dem Gesetz der großen Zahlen.

Beispiel 12.11 (Suchen in Listen). Gegeben seien n Namen von Personen, die aus r verschiedenen Städten kommen. Jede Person kommt (unabhängig von jeder anderen) mit Wahrscheinlichkeit p_j aus Stadt j , $j = 1, \dots, r$. Die Namen (zusammen mit anderen persönlichen Daten) werden in r verschiedene (ungeordnete) Listen eingetragen. Will man nun eine (zufällige, nach den Wahrscheinlichkeiten p_1, \dots, p_r verteilte) Person in der Liste suchen, bestimmt man zunächst die Stadt j , aus der die Person kommt und sucht anschließend in Liste

j nach dem Personennamen. Bis man feststellt, dass der Name nicht auf der Liste auftaucht, muss man die zu findende Person mit Namen auf der Liste vergleichen. Die Frage ist nun: Wie oft hat man im Mittel ohne Erfolg den Namen der zu findenden Person mit Namen auf der Liste verglichen, bis man endgültig weiß, dass die Person nicht auf der Liste steht?

Wir definieren zunächst ein paar Zufallsvariablen:

J : Nummer der Stadt, aus der die zu suchende Person kommt

L : Anzahl der unerfolgreichen Vergleiche, bis man den Namen der zu findenden Person findet

Z_j : Anzahl der Personen aus Stadt j

sowie $Z = (Z_1, \dots, Z_r)$. Um $\mathbf{E}[L]$ zu ermitteln, bestimmen wir zunächst

$$\mathbf{P}(L = a | J, Z) = 1_{Z_J = a}$$

und damit

$$\mathbf{P}(L = a | Z) = \sum_{j=1}^r p_j 1_{Z_j = a}.$$

Daraus folgern wir

$$\mathbf{E}[L | Z] = \sum_{a=1}^{\infty} \sum_{j=1}^r a \cdot p_j \cdot 1_{Z_j = a} = \sum_{j=1}^r p_j Z_j$$

und deshalb

$$\mathbf{E}[L] = \mathbf{E}[\mathbf{E}[L | Z]] = \sum_{j=1}^r p_j \mathbf{E}[Z_j] = n \cdot \sum_{j=1}^r p_j^2.$$

12.4 Bedingte Unabhängigkeit

In Abschnitt 9 haben wir bereits die Unabhängigkeit von σ -Algebren (oder von Zufallsvariablen) kennen gelernt. Bedingte Erwartungen und Unabhängigkeit sind eng verwandt, wie das erste Lemma zeigt.

Lemma 12.12 (Bedingte Wahrscheinlichkeit und Unabhängigkeit). *Die σ -Algebren $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ sind genau dann unabhängig, wenn $\mathbf{P}(G | \mathcal{H}) = \mathbf{P}(G)$ für alle $G \in \mathcal{G}$.*

Beweis. '⇒': Wenn \mathcal{G} und \mathcal{H} unabhängig sind, so ist für $G \in \mathcal{G}, H \in \mathcal{H}$

$$\mathbf{E}[\mathbf{P}(G), H] = \mathbf{P}(G \cap H) = \mathbf{E}[\mathbf{P}(G | \mathcal{H}), H].$$

Damit ist $\mathbf{P}(G | \mathcal{H}) = \mathbf{P}(G)$ nach der Definition der bedingten Erwartung.

'⇐': Gilt also $\mathbf{P}(G | \mathcal{H}) = \mathbf{P}(G)$, so folgt für $H \in \mathcal{H}$

$$\mathbf{P}(G \cap H) = \mathbf{E}[1_G, H] = \mathbf{E}[\mathbf{P}(G | \mathcal{H}), H] = \mathbf{E}[\mathbf{P}(G), H] = \mathbf{P}(G) \cdot \mathbf{P}(H). \quad \square$$

Oft benötigt man das Konzept der Unabhängigkeit auch noch in einer bedingten Form. Dafür geben wir zunächst ein wichtiges Beispiel an.

Beispiel 12.13 (Markov-Ketten). Sei E eine abzählbare Menge. Eine Markov-Kette $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ ist eine Familie von E -wertigen Zufallsvariablen, so dass für alle $A \subseteq E$

$$\mathbf{P}(X_{t+1} \in A | X_0, \dots, X_t) = \mathbf{P}(X_{t+1} \in A | X_t). \quad (12.4)$$

Dies bedeutet: wenn man die Verteilung von X_{t+1} wissen will, und dabei schon die Informationen der Zufallsvariable X_t zur Verfügung hat, bringt die Information über die Zufallsvariablen X_0, \dots, X_{t-1} keine zusätzliche Information. Man sagt auch:

Gegeben X_t ist X_{t+1} unabhängig von X_0, \dots, X_{t-1} .

Oder in Termen von σ -Algebren:

Gegeben $\sigma(X_t)$ ist $\sigma(X_{t+1})$ unabhängig von $\sigma(X_0, \dots, X_{t-1})$.

Etwas umgangssprachlich sagt man auch: gegeben die Gegenwart (das ist der Zustand zur Zeit t , X_t) ist die Zukunft (d.h. X_{t+1}) unabhängig von der Vergangenheit (das sind die Zustände X_0, \dots, X_{t-1}).

Ein einfaches Beispiel für eine Markov-Kette ist die ein-dimensionale Irrfahrt: seien Y_1, Y_2, \dots unabhängig und identisch verteilt, so dass $\mathbf{P}(Y_1 = 1) = p$ und $\mathbf{P}(Y_1 = -1) = q$ für ein $p \in [0, 1]$. Weiter sei $X_0 = 0$ und $X_t = Y_1 + \dots + Y_t$. Dann ist $(X_t)_{t \geq 0}$ eine Markov-Kette, denn

$$\mathbf{P}(X_{t+1} = k | X_0, \dots, X_t) = \begin{cases} p, & k = X_t + 1, \\ q, & k = X_t - 1. \end{cases}$$

Insbesondere definiert die rechte Seite eine X_t -messbare Zufallsvariable und ist damit gleich $\mathbf{P}(X_{t+1} = k | X_t)$.

Definition 12.14 (Bedingte Unabhängigkeit). Sei $\mathcal{G} \subseteq \mathcal{F}$. Eine Familie $(\mathcal{C}_i)_{i \in I}$ von Mengensystemen mit $\mathcal{C}_i \subseteq \mathcal{F}$ heißt unabhängig gegeben \mathcal{G} , falls

$$\mathbf{P}\left(\bigcap_{j \in J} A_j | \mathcal{G}\right) = \prod_{j \in J} \mathbf{P}(A_j | \mathcal{G}) \quad (12.5)$$

für alle $J \in I$ und $A_j \in \mathcal{C}_j, j \in J$, gilt.

Analog definiert man die bedingte Unabhängigkeit für Zufallsvariablen. Sei Y eine Zufallsvariable. Eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen ist unabhängig gegeben \mathcal{G} (bzw. Y) falls $(\sigma(X_i))_{i \in I}$ unabhängig gegeben \mathcal{G} (bzw. $\sigma(Y)$) ist.

Beispiel 12.15 (Einfache Fälle). Seien $\mathcal{G} \subseteq \mathcal{F}$ eine σ -Algebra und $(\mathcal{C}_i)_{i \in I}$ eine Familie von Mengensystemen.

1. Ist $\mathcal{G} = \mathcal{F}$, so ist $(\mathcal{C}_i)_{i \in I}$ immer unabhängig gegeben \mathcal{G} .
2. Ist $\mathcal{G} = \{\emptyset, \Omega\}$, so ist $(\mathcal{C}_i)_{i \in I}$ genau dann unabhängig gegeben \mathcal{G} , wenn $(\mathcal{C}_i)_{i \in I}$ unabhängig sind.

Beispiel 12.16 (Binomialverteilung mit zufälliger Erfolgswahrscheinlichkeit). Wir betrachten nochmal den Münzwurf mit zufälliger Erfolgswahrscheinlichkeit aus Beispiel 12.1 und 12.8. Hier war X uniform auf $[0, 1]$ verteilt und, gegeben X sind Y_1, \dots, Y_n Bernoulli-verteilt. Nun sollte ja gelten, dass (Y_1, \dots, Y_n) unabhängig gegeben X sind. Genau wie in Beispiel 12.8 berechnen wir für $A = \{X \in I\}$ für ein $I \in \mathcal{B}([0, 1])$ und $y_1, \dots, y_n \in \{0, 1\}$ und $k := y_1 + \dots + y_n$

$$\begin{aligned} \mathbf{E}[1_{Y_1=y_1, \dots, Y_n=y_n}, A] &= \mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n, X \in I) \\ &= \int_I x^{y_1 + \dots + y_n} (1-x)^{n-y_1 - \dots - y_n} dx = \mathbf{E}[X^k (1-X)^{n-k}, A], \end{aligned}$$

also

$$\mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n | X) = X^k (1-X)^{n-k}.$$

Analog zeigt man für $i = 1, \dots, n$

$$\mathbf{P}(Y_i = y_i | X) = X^{y_i} (1 - X)^{1 - y_i}.$$

Daraus folgt

$$\mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n | X) = \prod_{i=1}^n \mathbf{P}(Y_i = y_i | X),$$

also sind (Y_1, \dots, Y_n) unabhängig gegeben X .

Lemma 12.12 gibt es auch in folgender Version, in der die Unabhängigkeit durch die bedingte Unabhängigkeit ausgetauscht ist.

Proposition 12.17 (Bedingte Wahrscheinlichkeit und bedingte Unabhängigkeit). Sei $\mathcal{K} \subseteq \mathcal{F}$ eine σ -Algebra. Die σ -Algebren $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ sind genau dann unabhängig gegeben \mathcal{K} , wenn $\mathbf{P}(G | \sigma(\mathcal{H}, \mathcal{K})) = \mathbf{P}(G | \mathcal{K})$ für alle $G \in \mathcal{G}$.

Beweis. '⇒': Wenn \mathcal{G} und \mathcal{H} unabhängig gegeben \mathcal{K} sind, so ist für $G \in \mathcal{G}, H \in \mathcal{H}, K \in \mathcal{K}$

$$\mathbf{E}[\mathbf{P}(G | \mathcal{K}), H \cap K] = \mathbf{E}[\mathbf{P}(G | \mathcal{K}) \mathbf{P}(H | \mathcal{K}), K] = \mathbf{E}[\mathbf{P}(G \cap H | \mathcal{K}), K] = \mathbf{P}(G \cap H \cap K).$$

Nun kann man zeigen, dass das Mengensystem

$$\mathcal{D} := \{A \in \sigma(\mathcal{H}, \mathcal{K}) : \mathbf{E}[\mathbf{P}(G | \mathcal{K}), A] = \mathbf{P}(G \cap A)\}$$

ein schnittstabiles Dynkin-System ist mit $\mathcal{D} \supseteq \mathcal{H}, \mathcal{K}$. Nun folgt mit Theorem 2.13, dass $\mathcal{D} = \sigma(\mathcal{H}, \mathcal{K})$, woraus $\mathbf{P}(G | \sigma(\mathcal{H}, \mathcal{K})) = \mathbf{P}(G | \mathcal{K})$ folgt.

'⇐': Gilt also $\mathbf{P}(G | \sigma(\mathcal{H}, \mathcal{K})) = \mathbf{P}(G | \mathcal{K})$, so folgt für $H \in \mathcal{H}$

$$\mathbf{P}(G \cap H | \mathcal{K}) = \mathbf{E}[\mathbf{P}(G | \sigma(\mathcal{H}, \mathcal{K})), H | \mathcal{K}] = \mathbf{E}[\mathbf{P}(G | \mathcal{K}), H | \mathcal{K}] = \mathbf{P}(G | \mathcal{K}) \cdot \mathbf{P}(H | \mathcal{K}). \quad \square$$

Beispiel 12.18 (Markov-Ketten). Betrachten wir nochmal die Markov-Kette $(X_t)_{t=0,1,2,\dots}$ aus Beispiel 12.13. Für festes t setzen wir $\mathcal{G} = \sigma(X_{t+1}), \mathcal{H} = \sigma(X_0, \dots, X_{t-1}), \mathcal{K} = \sigma(X_t)$. Die Markov-Eigenschaft (12.4) sagt nun für $G \in \mathcal{G}, H \in \mathcal{H}, K \in \mathcal{K}$, dass $\mathbf{P}(G | \sigma(\mathcal{H}, \mathcal{K})) = \mathbf{P}(G | \mathcal{K})$. Nach Proposition 12.17 bedeutet dies, dass X_{t+1} und (X_0, \dots, X_{t-1}) unabhängig gegeben X_t sind.

12.5 Reguläre Version der bedingten Verteilung

Wir haben im Abschnitt 12.1 gesehen, wie die bedingte Wahrscheinlichkeit $\mathbf{P}(A | \mathcal{G}) := \mathbf{E}[1_A | \mathcal{G}]$ für eine σ -Algebra $\mathcal{G} \subseteq \mathcal{F}$ definiert ist. Dies bedeutet jedoch noch *nicht*, dass wir ein Wahrscheinlichkeitsmaß $A \mapsto \mathbf{P}(A | \mathcal{G})$ definiert haben; siehe hierzu die nächste Bemerkung. In den meisten Fällen kann man jedoch ein solches (zufälliges, \mathcal{G} -messbares) Maß definieren, die (oder besser: eine) reguläre Version der bedingten Verteilung.

Bemerkung 12.19 (Bedingte Wahrscheinlichkeiten und bedingte Verteilungen).

Sei $\mathcal{G} \subseteq \mathcal{F}$ eine σ -Algebra und $A_1, A_2, \dots \in \mathcal{F}$ mit $A_i \cap A_j = \emptyset$. Dann ist für $B \in \mathcal{G}$

$$\begin{aligned} \mathbf{E}\left[\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n \mid \mathcal{G}\right); B\right] &= \mathbf{E}\left[\mathbf{E}\left[1_{\bigcup_{n=1}^{\infty} A_n} \mid \mathcal{G}\right]; B\right] = \mathbf{E}\left[1_{\bigcup_{n=1}^{\infty} A_n}; B\right] \\ &= \mathbf{E}\left[\sum_{n=1}^{\infty} 1_{A_n}; B\right] = \sum_{n=1}^{\infty} \mathbf{E}\left[1_{A_n}; B\right] \\ &= \sum_{n=1}^{\infty} \mathbf{E}\left[\mathbf{P}(A_n \mid \mathcal{G}); B\right] = \mathbf{E}\left[\sum_{n=1}^{\infty} \mathbf{P}(A_n \mid \mathcal{G}); B\right] \end{aligned}$$

und damit

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n \mid \mathcal{G}\right) = \sum_{n=1}^{\infty} \mathbf{P}(A_n \mid \mathcal{G}) \quad (12.6)$$

\mathbf{P} -fast überall. Das bedeutet, dass es eine von A_1, A_2, \dots abhängige Nullmenge gibt, so dass (12.6) für alle ω außerhalb dieser Nullmenge gilt. Da es aber überabzählbar viele Folgen $A_1, A_2, \dots \in \mathcal{F}$ gibt, muss es damit nicht notwendigerweise eine Nullmenge N geben, so dass (12.6) für jede Wahl von $A_1, A_2, \dots \in \mathcal{F}$ und $\omega \notin N$ gilt. Falls es jedoch ein solches N gibt, werden wir sagen, dass eine *reguläre Version der bedingten Verteilung* von \mathbf{P} gegeben \mathcal{G} existiert. Bedingungen hierfür werden wir in Theorem 12.22 kennenlernen.

Wir erinnern an den Begriff des stochastischen Kernes; siehe Definition 6.9.

Definition 12.20 (Reguläre Version der bedingten Verteilung). Sei (Ω', \mathcal{F}') ein Messraum, Y eine Ω' -wertige messbare Zufallsvariable und $\mathcal{G} \subseteq \mathcal{F}$. Ein stochastischer Kern $\kappa_{Y, \mathcal{G}}$ von (Ω, \mathcal{G}) nach (Ω', \mathcal{F}') heißt reguläre Version der bedingten Verteilung von Y , gegeben \mathcal{G} , falls

$$\kappa_{Y, \mathcal{G}}(\omega, B) = \mathbf{P}(Y \in B \mid \mathcal{G})(\omega)$$

für \mathbf{P} -fast alle ω und jedes $B \in \mathcal{F}'$.

Bemerkung 12.21 (Auf eine Zufallsvariable bedingte Verteilung). 1. Für den stochastischen Kern aus Definition 12.20 reicht es, die Eigenschaft (ii) aus Definition 6.9 nur für einen schnittstabilen Erzeuger \mathcal{C} von \mathcal{F} zu fordern. Es ist nämlich stets

$$\mathcal{D} := \{A' \in \mathcal{F}' : \omega \mapsto \kappa(\omega, A') \text{ ist } \mathcal{A}\text{-messbar}\}$$

ein Dynkin-System. Damit ist nach Theorem 2.13 auch $\mathcal{D} = \sigma(\mathcal{C})$.

2. Sei $\mathcal{G} = \sigma(X)$ für eine Zufallsvariable X in Definition 12.20.2. Ist dann $\kappa_{Y, \sigma(X)}$ eine reguläre Version der bedingten Erwartung von Y gegeben $\sigma(X)$, so ist $\omega \mapsto \kappa_{Y, \sigma(X)}(\omega, A')$ $\sigma(X)$ -messbar für alle $A' \in \mathcal{A}'$. Damit gibt es nach Proposition 12.7 eine nach $\sigma(X)/\mathcal{B}([0; 1])$ -messbare Abbildung $\varphi_{A'} : \Omega \rightarrow [0; 1]$ mit $\varphi_{A'} \circ X = \kappa_{Y, \sigma(X)}(\cdot, A')$. Wir setzen dann

$$\kappa_{Y, X}(x, A') := \varphi_{A'}(x)$$

und sagen $\kappa_{Y, X}$ ist die reguläre Version der bedingten Verteilung von Y gegeben X .

Theorem 12.22 (Existenz der regulären Version der bedingten Verteilung). Sei (E, r) ein vollständiger und separabler metrischer Raum, ausgestattet mit der Borel'schen σ -Algebra, $\mathcal{G} \subseteq \mathcal{F}$ eine σ -Algebra und Y eine (nach \mathcal{F} messbare) Zufallsvariable mit Werten in E . Dann existiert eine reguläre Version der bedingten Verteilung von Y gegeben \mathcal{G} .

Bevor wir das Theorem beweisen können, benötigen wir eine Eigenschaft (Proposition 12.24) über vollständige, separable metrische Räume.

Definition 12.23 (Borel'scher Raum). 1. Zwei Messräume (Ω, \mathcal{F}) und (Ω', \mathcal{F}') heißen isomorph, falls es eine bijektive, nach \mathcal{F}/\mathcal{F}' -messbare Abbildung $\varphi : \Omega \rightarrow \Omega'$ gibt, so dass φ^{-1} nach \mathcal{F}'/\mathcal{F} -messbar ist.

2. Ein Messraum (Ω, \mathcal{F}) heißt Borel'scher Raum, falls es eine Borel'sche Menge $A \in \mathcal{B}(\mathbb{R})$ gibt, so dass (Ω, \mathcal{F}) und $(A, \mathcal{B}(A))$ isomorph sind.

Proposition 12.24 (Polnische und Borel'sche Räume). Jeder vollständige und separable metrische Raum (E, r) , ausgestattet mit der Borel'schen σ -Algebra, ist ein Borel'scher Raum.

Beweis. Siehe etwa Dudley, Real analysis and probability, Theorem 13.1.1. \square

Beweis von Theorem 12.22. Wir beweisen das Theorem unter der schwächeren Voraussetzung, dass E , ausgestattet mit der Borel'schen σ -Algebra, ein Borel'scher Raum ist. O.E. können wir also annehmen, dass $E \in \mathcal{B}(\mathbb{R})$ ist. Die Strategie unseres Beweises besteht darin, eine Verteilungsfunktion der bedingten Verteilung zu finden, indem diese erst für rationale Werte festgelegt wird, bevor sie auf alle reellen Zahlen fortgesetzt wird.

Für $r \in \mathbb{Q}$ sei F_r eine Version von $\mathbf{P}(Y \leq r|\mathcal{G})$ (d.h. $F_r = \mathbf{P}(Y \leq r|\mathcal{G})$ fast sicher. Sei $A \in \mathcal{F}$ so, dass für $\omega \in A$ die Abbildung $r \mapsto F_r(\omega)$ nicht-fallend ist mit Grenzwerten 1 und 0 bei $\pm\infty$. Da A durch abzählbar viele Bedingungen gegeben ist, die alle fast sicher erfüllt sind, folgt $\mathbf{P}(A) = 1$. Definiere nun für $x \in \mathbb{R}$

$$F_x(\omega) := 1_A(\omega) \cdot \inf_{r>x} F_r(\omega) + 1_{A^c}(\omega) \cdot 1_{x \geq 0}.$$

Damit ist $x \mapsto F_x(\omega)$ für alle ω eine Verteilungsfunktion. Definiere

$$\kappa(\omega, \cdot) := \text{Ma\ss}, \text{ das durch } x \mapsto F_x(\omega) \text{ definiert ist.}$$

Für $r \in \mathbb{Q}$ und $B = (-\infty; r]$ ist

$$\omega \mapsto \kappa(\omega, B) = 1_A(\omega) \cdot \mathbf{P}(Y \leq r|\mathcal{G})(\omega) + 1_{A^c}(\omega) \cdot 1_{r \geq 0} \quad (12.7)$$

(nach \mathcal{F}) messbar. Da $\{(-\infty; r] : r \in \mathbb{Q}\}$ ein schnittstabiler Erzeuger von $\mathcal{B}(\mathbb{R})$ ist, ist nach Bemerkung 12.21 die Abbildung $\omega \mapsto \kappa(\omega, B)$ für alle $B \in \mathcal{F}$ messbar. Also ist κ ein stochastischer Kern.

Es bleibt zu zeigen, dass κ eine reguläre Version der bedingten Verteilung ist. Da (12.7) auf einem schnittstabilen Erzeuger von \mathcal{E} gilt, gilt für $\omega \in A$

$$\kappa(\omega, B) = \mathbf{P}(Y \in B|\mathcal{G})(\omega).$$

Mit anderen Worten, κ ist eine reguläre Version der bedingten Verteilung. \square

13 Martingale

Wir beginnen nun, stochastische Prozesse zu behandeln. Insbesondere kommen wir zu einer wichtigen Klasse stochastischer Prozesse, den Martingalen. Oftmals werden sie als ein *faïres Spiel* bezeichnet. Einfach gesagt ist ein Martingal ein reellwertiger stochastischer Prozess, dessen Inkremente im Mittel verschwinden.

Im ganzen Abschnitt sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, (E, r) ein vollständiger und separabler metrischer Raum und $I \subseteq \mathbb{R}$ eine geordnete Indexmenge (meist $I = \{0, 1, 2, \dots\}$ oder $I = \mathbb{R}_+$).

13.1 Definition und ein paar Beispiele

Zunächst kümmern wir uns um die elementare Frage danach, was ein stochastische Prozess überhaupt ist.

Definition 13.1 (Stochastischer Prozess, Filtration). 1. Sei $\mathcal{X} = (X_t)_{t \in I}$ so, dass $X_t : \Omega \rightarrow E$ $\mathcal{F}/\mathcal{B}(E)$ -messbar ist. Dann heißt \mathcal{X} ein E -wertiger (stochastischer) Prozess. Für $\omega \in \Omega$ heißt die Abbildung $I \rightarrow E$, gegeben durch $X(\omega) : t \mapsto X_t(\omega)$ ein Pfad von X .

2. Sei $0 < p < \infty$. Ein reellwertiger Prozess $\mathcal{X} = (X_t)_{t \in I}$ heißt p -fach integrierbar, falls $\mathbf{E}[|X_t|^p] < \infty$ für alle $t \in I$. Er heißt L^p -beschränkt, falls $\sup_{t \in I} \mathbf{E}[|X_t|^p] < \infty$.

3. Eine Familie $(\mathcal{F}_t)_{t \in I}$ von σ -Algebren mit $\mathcal{F}_t \subseteq \mathcal{F}, t \in I$, heißt Filtration, falls $\mathcal{F}_s \subseteq \mathcal{F}_t$ für alle $s \leq t$.

4. Der stochastische Prozess $(X_t)_{t \in I}$ heißt an eine Filtration $(\mathcal{F}_t)_{t \in I}$ adaptiert, falls X_t für alle $t \in I$ eine $\mathcal{F}_t/\mathcal{B}(E)$ -messbare Zufallsvariable ist.

5. Für einen stochastischen Prozess $\mathcal{X} = (X_t)_{t \in I}$ heißt die Filtration $\mathcal{F} = (\mathcal{F}_t)_{t \in I}$ mit $\mathcal{F}_t = \sigma(X_s : s \leq t)$ die von \mathcal{X} erzeugte Filtration.

Im Folgenden sei immer eine Filtration $(\mathcal{F}_t)_{t \in I}$ gegeben sein. Adaptiertheit eines stochastischen Prozesses wird immer bezüglich $(\mathcal{F}_t)_{t \in I}$ gemeint sein.

Beispiel 13.2 (Markov-Kette, ein einfaches Martingal). 1. Seien $(X_t)_{t \in I}$ unabhängig. Dann ist $\mathcal{X} = (X_t)_{t \in I}$ ein (sehr einfacher) stochastischer Prozess.

2. Eine recht einfache Abhängigkeitsstruktur zwischen Zufallsvariablen haben wir bereits bei (zeit-diskreten) Markov-Ketten kennen gelernt; siehe etwa Beispiel 12.13. Sei hierzu $I = \{0, 1, 2, \dots\}$, E höchstens abzählbar und $P = (p_{ij})_{i,j \in E}$ eine stochastische Matrix, d.h.

$$p_{ij} \geq 0, \quad \sum_{j \in E} p_{ij} = 1$$

für alle $i, j \in E$. Wir definieren rekursiv die Verteilung von $\mathcal{X} = (X_t)_{t \in I}$ mittels

$$\mathbf{P}(X_{t+1} = j | X_0, \dots, X_t) = \mathbf{P}(X_{t+1} = j | X_t) = p_{X_t, j}.$$

Es gilt mit $\mu_i = \mathbf{P}(X_0 = i)$

$$\mathbf{P}(X_0 = x_0, \dots, X_t = x_t) = \mu_{x_0} p_{x_0, x_1} \cdots p_{x_{t-1}, x_t},$$

wodurch nicht nur eine Verteilung auf E^I eindeutig bestimmt ist, sondern auch der stochastische Prozess $\mathcal{X} = (X_t)_{t \in I}$.

3. Für eine \mathcal{F} -messbare Zufallsvariable X kann man einen stochastischen Prozess definieren, nämlich $\mathcal{X} = (X_t)_{t \in I}$ mit

$$X_t := \mathbf{E}[X | \mathcal{F}_t]. \quad (*)$$

Natürlich ist dann, wegen Theorem 12.2.7

$$\mathbf{E}[X_t | \mathcal{F}_s] = \mathbf{E}[\mathbf{E}[X | \mathcal{F}_t] | \mathcal{F}_s] = \mathbf{E}[X | \mathcal{F}_s] = X_s.$$

Stochastische Prozesse \mathcal{X} mit der Eigenschaft werden wir Martingale nennen. In Abschnitt 13.4 wird es dann (unter anderem) darum gehen, wann es zu einem Martingal $(X_t)_{t \in I}$ eine Zufallsvariable X gibt, so dass (*) gilt.

Definition 13.3 ((Sub-, Super-)Martingal). Sei $\mathcal{X} = (X_t)_{t \in I}$ ein adaptierter, reellwertiger stochastischer Prozess mit $\mathbf{E}[|X_t|] < \infty$, $t \in I$. Dann heißt \mathcal{X}

$$\begin{aligned} & \text{Martingal, falls } \mathbf{E}[X_t | \mathcal{F}_s] = X_s \text{ für } s, t \in I, s < t, \\ & \text{Sub-Martingal, falls } \mathbf{E}[X_t | \mathcal{F}_s] \geq X_s \text{ für } s, t \in I, s < t, \\ & \text{Super-Martingal, falls } \mathbf{E}[X_t | \mathcal{F}_s] \leq X_s \text{ für } s, t \in I, s < t. \end{aligned}$$

Genauer sagen wir dass \mathcal{X} ein $(\mathcal{F}_t)_{t \in I}$ -*(Sub, Super)-Martingal* ist.

Bemerkung 13.4 (Martingaleigenschaft bei diskreter Indexmenge). Ist I diskret, etwa $I = \{0, 1, 2, \dots\}$, so ist ein reellwertiger stochastischer Prozess $\mathcal{X} = (X_t)_{t \in I}$ genau dann ein Martingal, wenn $\mathbf{E}[|X_t|] < \infty$, $t \in I$ und $\mathbf{E}[X_t | \mathcal{F}_{t-1}] = X_{t-1}$ für alle $t = 1, 2, \dots$ gilt. Es gilt dann nämlich für $s, t \in I, s \leq t$,

$$\mathbf{E}[X_t | \mathcal{F}_s] = \mathbf{E}[\dots \mathbf{E}[\mathbf{E}[X_t | \mathcal{F}_{t-1}] | \mathcal{F}_{t-2}] \dots | \mathcal{F}_s] = X_s$$

nach Theorem 12.2.7. Analoges gilt für Sub- und Super-Martingale.

Beispiel 13.5 (Summen und Produkte integrierbarer Zufallsvariablen).

1. Sei X_1, X_2, \dots eine Folge unabhängiger, integrierbarer Zufallsvariablen mit $\mathbf{E}[X_i] = 0$, $i = 1, 2, \dots$ und $\mathcal{F}_t := \sigma(X_1, \dots, X_t)$. Weiter sei $S_0 := 0$ und für $t = 1, 2, \dots$

$$S_t := \sum_{i=1}^t X_i.$$

Dann ist

$$\mathbf{E}[S_t | \mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1} + X_t | \mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t | \mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t] = S_{t-1},$$

d.h. $(S_t)_{t=0,1,2,\dots}$ ist ein Martingal.

Falls $\mathbf{E}[X_i] \geq 0$ für alle $i = 1, 2, \dots$, so ist $(S_t)_{t \geq 0}$ ein Sub-Martingal.

2. Sei $I = \{-1, -2, \dots\}$ und X_1, X_2, \dots seien integrierbare, unabhängige, identisch verteilte Zufallsvariablen. Weiter setzen wir für $t \in I$

$$S_t := \frac{1}{|t|} \sum_{i=1}^{|t|} X_i$$

und $\mathcal{F}_t := \sigma(\dots, S_{t-1}, S_t)$. Dann ist für $t \in I$

$$\begin{aligned} \mathbf{E}[S_t | \mathcal{F}_{t-1}] &= \mathbf{E}\left[\frac{1}{|t|} \sum_{i=1}^{|t|} X_i \mid S_{t-1}, S_{t-2}, \dots\right] \\ &= \frac{1}{|t|} \sum_{i=1}^{|t|} \mathbf{E}\left[X_i \mid \sum_{i=1}^{|t|+1} X_i\right] \\ &= \mathbf{E}\left[X_1 \mid \sum_{i=1}^{|t|+1} X_i\right] \\ &= \frac{1}{|t-1|} \sum_{i=1}^{|t-1|} X_i \\ &= S_{t-1} \end{aligned}$$

nach Beispiel 12.9. Speziell ist

$$\begin{aligned} \mathbf{E}[X_1 | \mathcal{F}_t] &= \mathbf{E}\left[X_1 \mid \sum_{i=1}^{|t|} X_i\right] \\ &= \frac{1}{|t|} \sum_{i=1}^{|t|} X_i \\ &= S_t. \end{aligned}$$

3. Sei $I = \{0, 1, 2, \dots\}$ und X_1, X_2, \dots eine Folge unabhängiger, integrierbarer Zufallsvariablen mit $\mathbf{E}[X_i] = 1, i = 1, 2, \dots$ und $\mathcal{F}_t := \sigma(X_1, \dots, X_t)$. Weiter ist $S_0 := 1$ und für $t = 1, 2, \dots$

$$S_t := \prod_{i=1}^t X_i.$$

Dann ist, falls S_1, S_2, \dots integrierbar sind,

$$\mathbf{E}[S_t | \mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1} X_t | \mathcal{F}_{t-1}] = S_{t-1} \cdot \mathbf{E}[X_t | \mathcal{F}_{t-1}] = S_{t-1} \cdot \mathbf{E}[X_t] = S_{t-1},$$

d.h. $(S_t)_{t \in I}$ ist ein Martingal.

Falls $\mathbf{E}[X_i] \geq 1$ für alle $i = 1, 2, \dots$, so ist $(S_t)_{t \in I}$ ein Sub-Martingal.

Beispiel 13.6 (Verzweigungsprozesse in diskreter Zeit). Wir betrachten ein einfaches Modell für eine sich zufällige entwickelnde Population. Seien $X_i^{(t)}$ unabhängige, $\{0, 1, 2, \dots\}$ -wertige Zufallsvariable und $\mu = \mathbf{E}[X_i^{(t)}]$. Hier steht $X_i^{(t)}$ für die Anzahl der Nachkommen des

iten Individuums der Generation t mit $i, t = 0, 1, \dots, t = 1, 2, \dots$. Startend mit $Z_0 = k$ setzen wir

$$Z_{t+1} = \sum_{i=1}^{Z_t} X_i^{(t)},$$

also ist $\mathcal{Z} = (Z_t)_{t=0,1,2,\dots}$ der stochastische Prozess der Gesamtzahl an Individuen. Die Verteilung von $X_i^{(t)}$ heißt auch die *Nachkommensverteilung*.

Der Prozess \mathcal{Z} ist genau dann ein (nicht-negatives) Martingal (bezüglich der von \mathcal{Z} erzeugten Filtration), wenn $\mathbf{E}[X_i^{(t)}] = 1$, jedes Individuum im Mittel also einen Nachkommen hat. Es gilt nämlich für $t = 1, 2, \dots$

$$\mathbf{E}[Z_{t+1} - Z_t | \mathcal{F}_t] = \mathbf{E}\left[\sum_{i=1}^{Z_t} X_i^{(t)} - Z_t | \mathcal{F}_t\right] = (\mu - 1)Z_t.$$

Im Fall $\mu > 1$ ist \mathcal{Z} ein Sub-Martingal, und im Fall $\mu < 1$ ein Super-Martingal. Außerdem heißt¹⁸

- \mathcal{Z} ein kritischer Verzweigungsprozess, falls $\mu = 1$,
- \mathcal{Z} ein super-kritischer Verzweigungsprozess, falls $\mu > 1$,
- \mathcal{Z} ein sub-kritischer Verzweigungsprozess, falls $\mu < 1$.

Allgemein ist $(Z_t/\mu^t)_{t=0,1,2,\dots}$ ein (nicht-negatives) Martingal, weil genau wie in der letzten Rechnung

$$\mathbf{E}[Z_{t+1} - \mu Z_t | \mathcal{F}_t] = \mu Z_t - \mu Z_t = 0.$$

Bemerkenswert ist außerdem, dass $\mathbf{E}[Z_{t+1} | \mathcal{F}_t] = \mu Z_t$ gilt, woraus man rekursiv folgern kann, dass

$$\mathbf{E}[Z_t] = \mu^t.$$

□

Wir schließen diesen Abschnitt mit einer einfachen Aussage, wie man aus bekannten (Sub)-Martingalen weitere Sub-Martingale erhält.

Proposition 13.7 (Konvexe Funktionen von Martingalen sind Sub-Martingale).

Sei $\mathcal{X} = (X_t)_{t \in I}$ ein stochastischer Prozess und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ konvex. Falls $\varphi(X) = (\varphi(X_t))_{t \in I}$ integrierbar ist und eine der beiden Bedingungen

1. \mathcal{X} ist ein Martingal
2. \mathcal{X} ist ein Sub-Martingal und φ ist nicht-fallend

erfüllt ist, so ist $\varphi(\mathcal{X}) = (\varphi(X_t))_{t \in I}$ ein Sub-Martingal.

Beweis. Ist \mathcal{X} ein Martingal so ist $\varphi(X_s) = \varphi(\mathbf{E}[X_t | \mathcal{F}_s])$. Ist \mathcal{X} ein Sub-Martingal und φ nicht-fallend, gilt $\varphi(X_s) \leq \varphi(\mathbf{E}[X_t | \mathcal{F}_s])$. In beiden Fällen gilt damit für $s \leq t$ fast sicher wegen der Jensen'schen Ungleichung für bedingte Erwartungen, Proposition 12.4, dass

$$\varphi(X_s) \leq \varphi(\mathbf{E}[X_t | \mathcal{F}_s]) \leq \mathbf{E}[\varphi(X_t) | \mathcal{F}_s],$$

d.h. $\varphi(\mathcal{X})$ ist ein Sub-Martingal. □

¹⁸Es mag irritierend erscheinen, dass ein *superkritischer Verzweigungsprozess* ein Sub-Martingal und ein *subkritischer Verzweigungsprozess* ein Super-Martingal ist.

13.2 Das stochastische Integral als Martingal

In diesem Abschnitt sei immer $I = \{0, 1, 2, \dots\}$ (wobei sich alle Ergebnisse auch auf eine diskrete Indexmenge $I = \{t_0, t_1, \dots\}$ mit $t_0 < t_1 < \dots$ übertragen lassen). Alle hier eingeführten Konzepte haben ein Analogon für Prozesse in stetiger Zeit. Allerdings sind die entsprechenden Aussagen dann deutlich aufwändiger zu formulieren und zu beweisen. Einige dieser analogen Aussagen werden erst in der Vorlesung *Stochastische Integration* formuliert.

Definition 13.8 (Previsibler Prozess). Ein stochastischen Prozess \mathcal{X} heißt $(\mathcal{F}_t)_{t \in I}$ -previsibel, falls $X_0 = 0$ und X_t \mathcal{F}_{t-1} -messbar ist, $t = 1, 2, \dots$

Proposition 13.9 (Doob-Zerlegung). Sei $I = \{0, 1, 2, \dots\}$. Jeder adaptierte Prozess $\mathcal{X} = (X_t)_{t \in I}$ hat eine fast sicher eindeutige Zerlegung $\mathcal{X} = \mathcal{M} + \mathcal{A}$, wobei \mathcal{M} ein Martingal und \mathcal{A} previsibel ist. Insbesondere ist \mathcal{X} genau dann ein Sub-Martingal falls \mathcal{A} fast sicher nicht fällt.

Beweis. Definiere den previsiblen Prozess $\mathcal{A} = (A_t)_{t \in I}$ durch

$$A_t = \sum_{s=1}^t \mathbf{E}[X_s - X_{s-1} | \mathcal{F}_{s-1}]. \quad (13.1)$$

Dann ist $\mathcal{M} = \mathcal{X} - \mathcal{A}$ ein Martingal, denn

$$\mathbf{E}[M_t - M_{t-1} | \mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}] - (A_t - A_{t-1}) = 0.$$

Kommen wir zur Eindeutigkeit der Darstellung. Falls $\mathcal{X} = \mathcal{M} + \mathcal{A}$ für ein Martingal \mathcal{M} und einen previsiblen Prozess \mathcal{A} , so ist $A_t - A_{t-1} = \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}]$ für alle $t = 1, 2, \dots$, d.h. (13.1) gilt fast sicher. \square

Definition 13.10 (Quadratische Variation, wachsender Prozess). Sei $I = \{0, 1, 2, \dots\}$ und $\mathcal{X} = (X_t)_{t \in I}$ ein quadratisch integrierbares Martingal. Der fast sicher eindeutig bestimmte, previsible Prozess $(\langle \mathcal{X} \rangle_t)_{t \in I}$, für den $(X_t^2 - \langle \mathcal{X} \rangle_t)_{t \in I}$ ein Martingal ist, heißt der quadratische Variationsprozess (oder auch der wachsende Prozess) von \mathcal{X} .

Proposition 13.11 (Wachsender Prozess und Varianz). Sei $I = \{0, 1, 2, \dots\}$, $\mathcal{X} = (X_t)_{t \in I}$ ein Martingal mit quadratischem Variationsprozess $\langle \mathcal{X} \rangle = (\langle \mathcal{X} \rangle_t)_{t \in I}$. Dann ist

$$\langle \mathcal{X} \rangle_t = \sum_{s=1}^t \mathbf{E}[X_s^2 - X_{s-1}^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t \mathbf{E}[(X_s - X_{s-1})^2 | \mathcal{F}_{s-1}]$$

und

$$\mathbf{E}[\langle \mathcal{X} \rangle_t] = \mathbf{V}[X_t - X_0].$$

Beweis. Wie im Beweis von Proposition 13.9 kann man den Prozess $\langle \mathcal{X} \rangle$ mittels (13.1) schreiben. Daraus folgt sofort das erste Gleichheitszeichen. Das zweite folgt, da $\mathbf{E}[X_s X_{s-1} | \mathcal{F}_{s-1}] = X_{s-1}^2$. Weiter ist

$$\mathbf{E}[\langle \mathcal{X} \rangle_t] = \sum_{s=1}^t \mathbf{E}[X_s^2 - X_{s-1}^2] = \mathbf{E}[X_t^2 - X_0^2] = \mathbf{E}[(X_t - X_0)^2] = \mathbf{V}[X_t - X_0].$$

\square

Beispiel 13.12 (Wachsende Prozesse). 1. Sei $\mathcal{S} = (S_t)_{t \in I}$ mit $S_t = \sum_{s=1}^t X_s$ wie in Beispiel 13.5.1 mit quadratische integrierbaren Zufallsvariablen X_1, X_2, \dots . Dann ist nach Proposition 13.11

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[X_s^2].$$

Insbesondere ist der quadratische Variationsprozess von \mathcal{S} deterministisch.

2. Sei $\mathcal{S} = (S_t)_{t \in I}$ mit $S_t = \prod_{s=1}^t X_s$ wie in Beispiel 13.5.3 mit quadratisch integrierbaren Zufallsvariablen X_1, X_2, \dots . Dann ist

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[(S_s - S_{s-1})^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{E}[(X_s - 1)^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{V}[X_s].$$

Insbesondere ist in diesem Beispiel der Prozess $\langle \mathcal{S} \rangle$ echt stochastisch.

Definition 13.13 (Diskretes stochastisches Integral). Sei $I = \{0, 1, 2, \dots\}$ und $\mathcal{H} = (H_t)_{t \in I}, \mathcal{X} = (X_t)_{t \in I}$ stochastische Prozesse mit Werten in \mathbb{R} . Ist \mathcal{X} adaptiert und \mathcal{H} previsibel, so definieren wir das stochastische Integral $\mathcal{H} \cdot \mathcal{X} = ((\mathcal{H} \cdot \mathcal{X})_t)_{t \in I}$ durch

$$(\mathcal{H} \cdot \mathcal{X})_t = \sum_{s=1}^t H_s (X_s - X_{s-1})$$

für alle $t \in I$. Ist \mathcal{X} ein Martingal, so nennen wir $\mathcal{H} \cdot \mathcal{X}$ eine Martingaltransformierte von \mathcal{X} .

Proposition 13.14 (Stabilität der stochastischen Integrale). Sei $I = \{0, 1, 2, \dots\}$ und $\mathcal{X} = (X_t)_{t \in I}$ ein adaptierter, reellwertiger Prozess mit $\mathbf{E}[|X_0|] < \infty$.

1. \mathcal{X} ist genau dann ein Martingal, wenn für jeden previsiblen Prozess $\mathcal{H} = (H_t)_{t \in I}$ das stochastische Integral $\mathcal{H} \cdot \mathcal{X}$ ein Martingal ist.
2. \mathcal{X} ist genau dann ein Sub-Martingal (Super-Martingal), wenn für jeden previsiblen, nicht-negativen Prozess $\mathcal{H} = (H_t)_{t \in I}$ das stochastische Integral $\mathcal{H} \cdot \mathcal{X}$ ein Sub-Martingal (Super-Martingal) ist.

Beweis. 1. '⇒': Man schreibt sofort

$$\begin{aligned} \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_{t+1} - (\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_t] &= \mathbf{E}[H_{t+1}(X_{t+1} - X_t) | \mathcal{F}_t] \\ &= H_{t+1} \mathbf{E}[X_{t+1} - X_t | \mathcal{F}_t] \\ &= 0. \end{aligned}$$

'⇐': Sei $t \in I$ und $H_s := 1_{\{s=t\}}$. Dann ist $\mathcal{H} = (H_s)_{s \in I}$ deterministisch, insbesondere previsibel. Da $(\mathcal{H} \cdot \mathcal{X})_{t-1} = 0$ gilt, folgt

$$0 = \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}] = \mathbf{E}[X_t | \mathcal{F}_{t-1}] - X_{t-1}$$

Daraus folgt die Behauptung.

2. folgt analog. □

Beispiel 13.15 (Quadratische Variation für stochastische Integrale). Sei $I = \{0, 1, 2, \dots\}$, $\mathcal{X} = (X_t)_{t \in I}$ ein Martingal und $\mathcal{H} = (H_t)_{t \in I}$ previsibel. Dann ist wegen Proposition 13.11

$$\begin{aligned} \langle \mathcal{H} \cdot \mathcal{X} \rangle_t &= \sum_{s=1}^t \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_s - (\mathcal{H} \cdot \mathcal{X})_{s-1}]^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t \mathbf{E}[H_s^2 (X_s - X_{s-1})^2 | \mathcal{F}_{s-1}] \\ &= \sum_{s=1}^t H_s^2 \cdot \mathbf{E}[(X_s - X_{s-1})^2 | \mathcal{F}_{s-1}]. \end{aligned}$$

Inbesondere gilt also

$$\mathbf{V}[(\mathcal{H} \cdot \mathcal{X})_t] = \sum_{s=1}^t \mathbf{E}[H_s^2 \cdot (X_s - X_{s-1})^2].$$

Beispiel 13.16 (Auszahlung bei Spielsystemen). Martingaltransformierte kann man auch als Auszahlungen von Spielsystemen interpretieren. Gegeben, eine zufällige Größe entwickelt sich gemäß des adaptierten Prozesses $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$. Wettet man vor Zeit t mit einem Einsatz H_t (basierend auf den Erfahrungen, die aus X_0, \dots, X_{t-1} gewonnen wurden) auf die Änderung der zufälligen Größe $X_t - X_{t-1}$, so ist $(\mathcal{H} \cdot \mathcal{X})_t$ der bis zur Zeit t realisierte Gewinn. Gegeben der zugrunde liegende Prozess \mathcal{X} ist ein Martingal, zeigt Proposition 13.14, dass der erzielte Gewinn $\mathcal{H} \cdot \mathcal{X}$ für jede Strategie \mathcal{H} ein Martingal ist. Insbesondere ist der erwartete Gewinn 0.

Als Beispiel betrachten wir das Petersburger Paradoxon: eine faire Münze wird unendlich oft geworfen. In jeder Runde setzt ein Spieler einen Einsatz in beliebiger Höhe. Kommt Kopf, verliert er ihn, kommt Zahl, so wird der Einsatz verdoppelt wieder ausbezahlt. Das Paradox besteht aus folgender Strategie: startend mit einem Einsatz von 1 beim ersten Münzwurf, verdoppelt der Spieler bei jedem Misserfolg seinen Einsatz. Kommt der erste Erfolg im t -ten Wurf, so beträgt sein bisheriger Einsatz $\sum_{i=1}^t 2^{i-1} = 2^t - 1$. Da der letzte Einsatz 2^{t-1} war, bekommt der Spieler also 2^t zurück, hat also sicher einen Gewinn von 1 gemacht, obwohl das Spiel fair war.

Um dieses Spiel mittels Martingalen zu analysieren, sei X_1, X_2, \dots eine unabhängig, identisch verteilte Folge mit $\mathbf{P}(X_1 = -1) = \mathbf{P}(X_1 = 1) = \frac{1}{2}$, und $S_0 = 0, S_t = \sum_{i=1}^t X_i$. Dann ist $\mathcal{S} = (S_t)_{t=0,1,2,\dots}$ ein Martingal. Weiter sei H_t der Einsatz im t -ten Spiel. Also ist

$$(\mathcal{H} \cdot \mathcal{S})_t = \sum_{i=1}^t H_i (S_i - S_{i-1}) = \sum_{i=1}^t H_i X_i$$

der Gewinn nach dem t -ten Spiel. Da mit \mathcal{S} auch $\mathcal{H} \cdot \mathcal{S}$ ein Martingal ist, gilt

$$\lim_{t \rightarrow \infty} \mathbf{E}[(\mathcal{H} \cdot \mathcal{S})_t] = \mathbf{E}[(\mathcal{H} \cdot \mathcal{S})_1] = \mathbf{E}[X_1] = 0,$$

d.h. der mittlere Gewinn nach langer Zeit ist 0, unabhängig von der Strategie \mathcal{H} . Oben haben wir den Einsatz

$$H_t := 2^{t-1} 1_{\{S_{t-1} = -(t-1)\}} \tag{13.2}$$

betrachtet und gezeigt, dass für den Gewinn $(\mathcal{H} \cdot \mathcal{S})_t \xrightarrow[t \rightarrow \infty]{f.s.} 1$ gilt.

Wie bewerten wir nun die Strategie (13.2)? Sei T die zufällige Zeit des Gewinns, d.h. T ist geometrisch verteilt mit Parameter $\frac{1}{2}$. Insbesondere ist T fast sicher endlich. Dann ist

$$\mathbf{E}\left[\sum_{t=1}^{\infty} H_t\right] = \sum_{k=1}^{\infty} \frac{1}{2^k} (2^k - 1) = \infty,$$

d.h. für die obige Strategie benötigt man unter Umständen sehr viel Kapital.

13.3 Gestoppte Martingale

Sei $\mathcal{X} = (X_t)_{t \in I}$ ein stochastischer Prozess. Unter einem gestoppten stochastischen Prozess versteht man den Prozess $\mathcal{X}^T := (X_{T \wedge t})_{t \in I}$, wobei T eine I -wertige Zufallsvariable ist. Der Prozess \mathcal{X}^T hält also an, wenn T erreicht ist. Spezielle Zufallsvariablen T heißen Stoppzeiten, deren Eintreten zur Zeit t mittels der σ -Algebra \mathcal{F}_t entschieden werden kann. (Man denke etwa an einen Spieler, der ein faires Spiel spielt, und zu einer zufälligen Zeit, z.B. wenn er genug gewonnen oder verloren hat, aufhört.) Wir werden in diesem Abschnitt das Optional Stopping Theorem kennenlernen, das besagt, dass (zu einer Stoppzeit) gestoppte Martingale wieder Martingale sind; siehe Theorem 13.19. Das Optional Sampling Theorem gibt Bedingungen an, für die die Martingale-Eigenschaft nicht nur zu festen, sondern auch zu zufälligen Stoppzeiten gilt; siehe Theorem 13.23.

Definition 13.17 (Stoppzeit). 1. Eine zufällige Zeit ist eine Zufallsvariable mit Werten in \bar{I} (dem Abschluss von I). Eine zufällige Zeit T heißt $((\mathcal{F}_t)_{t \in I})$ -Stoppzeit falls $\{T \leq t\} \in \mathcal{F}_t$ für alle $t \in I$.

2. Jede Stoppzeit T definiert die σ -Algebra

$$\mathcal{F}_T := \{A \in \mathcal{A} : A \cap \{T \leq t\} \in \mathcal{F}_t, t \in I\}$$

der T -Vergangenheit.

3. Sei $B \in \mathcal{B}(E)$. Dann ist die Treffzeit von B definiert als

$$T_B := \inf\{t : X_t \in B\}.$$

4. Für eine zufällige Zeit T ist X_T definiert durch $\omega \mapsto X_{T(\omega)}(\omega)$. Weiter ist $\mathcal{X}^T := (X_{T \wedge t})_{t \in I}$ der bei T gestoppte Prozess.

Bemerkung 13.18 (Interpretation und Treffzeiten). 1. Sei $\mathcal{X} = (X_t)_{t \in I}$ ein stochastischer Prozess und $(\mathcal{F}_t)_{t \in I}$ die kanonische Filtration. Man kann \mathcal{F}_t als die Information auffassen, die zum Zeitpunkt t durch Kenntnis von $(X_s)_{0 \leq s \leq t}$ verfügbar ist. Ist nun T eine Stoppzeit, so gilt $\{T \leq t\} \in \mathcal{F}_t$. Also lässt sich das Eintreten des Ereignisses $\{T \leq t\}$ durch Kenntnis von $(X_s)_{s \leq t}$ entscheiden. Mit anderen Worten kann durch Kenntnis des stochastischen Prozesses bis zur Zeit t entschieden werden, ob die Stoppzeit T spätestens jetzt gerade abgelaufen ist.

2. Ist I höchstens abzählbar und $B \in \mathcal{B}(E)$, so ist T_B eine Stoppzeit. Es ist nämlich

$$\{T_B \leq t\} = \bigcup_{s \leq t} \underbrace{\{X_s \in B\}}_{\in \mathcal{F}_s \subseteq \mathcal{F}_t} \in \mathcal{F}_t.$$

Proposition 13.19 (Optional Stopping). Sei $I = \{0, 1, 2, \dots\}$ und $\mathcal{X} = (X_t)_{t \in I}$ ein (Sub, Super)-Martingal und T eine Stoppzeit. Dann ist $\mathcal{X}^T = (X_{T \wedge t})_{t \in I}$ ein (Sub, Super)-Martingal.

Beweis. Wir zeigen die Aussage nur für den Fall, dass \mathcal{X} ein Sub-Martingal ist. Die anderen Aussagen ergeben sich aus Symmetriegründen. Für ein Sub-Martingal \mathcal{X} und $\{T > t-1\} \in \mathcal{F}_t$ ist

$$\begin{aligned} \mathbf{E}[X_{T \wedge t} - X_{T \wedge (t-1)} | \mathcal{F}_{t-1}] &= \mathbf{E}[(X_t - X_{t-1})1_{\{T > t-1\}} | \mathcal{F}_{t-1}] \\ &= 1_{\{T > t-1\}} \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}] \geq 0, \end{aligned}$$

d.h. \mathcal{X}^T ist ein Sub-Martingal. □

Proposition 13.20 (Messbarkeit von X_T). Sei I höchstens abzählbar, $\mathcal{X} = (X_t)_{t \in I}$ ein an die Filtration $(\mathcal{F}_t)_{t \in I}$ adaptierter, stochastischer Prozess und T eine $(\mathcal{F}_t)_{t \in I}$ -Stoppzeit. Dann ist

$$X_T : \begin{cases} \{T < \infty\} & \rightarrow E \\ \omega & \mapsto X_{T(\omega)}(\omega) \end{cases}$$

messbar bezüglich $\{T < \infty\} \cap \mathcal{F}_T / \mathcal{B}(E)$.

Beweis. Wir müssen zeigen, dass $\{X_T \in B, T \leq t\} \in \mathcal{F}_t$ für $B \in \mathcal{B}(E)$ gilt, $t \in I$. Per Definition von \mathcal{F}_T gilt dann nämlich, dass $\{X_T \in B\} \in \mathcal{F}_T$, woraus die Behauptung folgt. Da aber $\{X_T \in B, T \leq t\} = \{X_{T \wedge t} \in B, T \leq t\}$, genügt es zu zeigen, dass $X_{T \wedge t}$ messbar ist bezüglich \mathcal{F}_t , $t \in I$. Wir können also oBdA annehmen, dass $T \leq t$ gilt. Wir schreiben $X_T = Y_t \circ \psi$, wobei $\psi(\omega) := (T(\omega), \omega)$ messbar ist bezüglich $\mathcal{F}_t / (I \cap \mathcal{B}([0, t]) \otimes \mathcal{F}_t)$ und $Y_t(s, \omega) = X_s(\omega)$ nach Voraussetzung $I \cap \mathcal{B}([0, t]) \otimes \mathcal{F}_t / \mathcal{B}(E)$ -messbar ist. Die Behauptung folgt nun mit Lemma 4.5.2. □

Lemma 13.21 (Bedingen auf \mathcal{F}_T). Sei $I = \{0, 1, 2, \dots\}$, $\mathcal{X} = (X_t)_{t \in I}$ ein Martingal und T eine durch t beschränkte Stoppzeit. Dann gilt $X_T = \mathbf{E}[X_t | \mathcal{F}_T]$.

Beweis. Nach der Definition der bedingten Erwartung und da X_T \mathcal{F}_T -messbar ist (siehe Proposition 13.20), müssen wir zeigen, dass $\mathbf{E}[X_t; A] = \mathbf{E}[X_T; A]$ für $A \in \mathcal{F}_T$ gilt. Es ist $\{T = s\} \cap A \in \mathcal{F}_s$ für $s \in I$, also

$$\begin{aligned} \mathbf{E}[X_T; A] &= \sum_{s=1}^t \mathbf{E}[X_s; \{T = s\} \cap A] \\ &= \sum_{s=1}^t \mathbf{E}[\mathbf{E}[X_t | \mathcal{F}_s]; \{T = s\} \cap A] \\ &= \sum_{s=1}^t \mathbf{E}[X_t; \{T = s\} \cap A] \\ &= \mathbf{E}[X_t; A]. \end{aligned}$$

□

Lemma 13.22 (Gleichgradige Integrierbarkeit und Stoppzeiten).

Sei $I = \{0, 1, 2, \dots\}$. Ein Martingal $\mathcal{X} = (X_t)_{t \in I}$ ist genau dann gleichgradig integrierbar, wenn die Familie $\{X_T : T \text{ fast sicher endliche Stoppzeit}\}$ gleichgradig integrierbar ist.

Beweis. '⇐': klar.

'⇒': Nach Lemma 8.9 gibt es eine konvexe Funktion $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ mit $\frac{f(x)}{x} \xrightarrow{x \rightarrow \infty} \infty$ und $\sup_{t \in I} \mathbf{E}[f(|X_t|)] =: L < \infty$. Sei T eine fast sicher endliche Stoppzeit, dann ist nach Lemma 13.21 (angewendet auf die fast sicher endliche Stoppzeit $T \wedge t$) $\mathbf{E}[X_t | \mathcal{F}_{T \wedge t}] = X_{T \wedge t}$. Da $\{T \leq t\} \in \mathcal{F}_{T \wedge t}$ folgt mit der Jensen'schen Ungleichung

$$\begin{aligned} \mathbf{E}[f(|X_T|), \{T \leq t\}] &= \mathbf{E}[f(|X_{T \wedge t}|), \{T \leq t\}] \\ &= \mathbf{E}[f(|\mathbf{E}[X_t | \mathcal{F}_{T \wedge t}]|), \{T \leq t\}] \\ &\leq \mathbf{E}[\mathbf{E}[f(|X_t|) | \mathcal{F}_{T \wedge t}], \{T \leq t\}] \\ &= \mathbf{E}[f(|X_t|), \{T \leq t\}] \leq L. \end{aligned}$$

Damit ist $\mathbf{E}[f(|X_T|)] \leq L$, d.h. die Behauptung folgt mit Lemma 8.9. \square

In Beispiel 13.16 war $\mathcal{H} \cdot \mathcal{S}$ ein Martingal, T eine Stoppzeit und $\mathbf{E}[(\mathcal{H} \cdot \mathcal{S})_t] = 0 \neq 1 = (\mathcal{H} \cdot \mathcal{S})_T$. Falls T beschränkt gewesen wäre, wäre diese Ungleichung nicht möglich gewesen, wie wir nun zeigen.

Theorem 13.23 (Optional Sampling Theorem). *Sei $I = \{0, 1, 2, \dots\}$, $S \leq T$ fast sicher endliche Stoppzeiten und $\mathcal{X} = (X_t)_{t \in I}$ ein Sub-Martingal. Ist entweder T beschränkt oder \mathcal{X} gleichgradig integrierbar, so ist X_T integrierbar und $X_S \leq \mathbf{E}[X_T | \mathcal{F}_S]$.*

Beweis. Wir führen den Beweis zunächst im Fall einer beschränkten Stoppzeit T . Sei also $T \leq t$ für ein $t \in I$. Wir verwenden die Doob-Zerlegung $\mathcal{X} = \mathcal{M} + \mathcal{A}$ von \mathcal{X} in das Martingal \mathcal{M} und den monoton nicht-fallenden Prozess \mathcal{A} . Dann ist mit Lemma 13.21 und $\mathcal{F}_S \subseteq \mathcal{F}_T$ nach Theorem 12.2.7

$$\begin{aligned} X_S &= M_S + A_S = \mathbf{E}[M_t + A_S | \mathcal{F}_S] \\ &\leq \mathbf{E}[M_t + A_T | \mathcal{F}_S] \\ &= \mathbf{E}[\mathbf{E}[M_t | \mathcal{F}_T] + A_T | \mathcal{F}_S] \\ &= \mathbf{E}[M_T + A_T | \mathcal{F}_S] \\ &= \mathbf{E}[X_T | \mathcal{F}_S]. \end{aligned}$$

Sei nun T unbeschränkt und \mathcal{X} gleichgradig integrierbar. Sei $\mathcal{X} = \mathcal{M} + \mathcal{A}$ die Doob-Zerlegung von \mathcal{X} in das Martingal \mathcal{M} und den nicht-fallenden previsiblen Prozess $\mathcal{A} \geq 0$ mit $A_0 = 0$. Da

$$\mathbf{E}[|A_t|] = \mathbf{E}[A_t] = \mathbf{E}[X_t - X_0] \leq \mathbf{E}[|X_0|] + \sup_{s \in I} \mathbf{E}[|X_s|]$$

gilt $A_t \uparrow A_\infty$ für ein $A_\infty \geq 0$ mit $\mathbf{E}[A_\infty] < \infty$. Mit Lemma 8.9 kann man folgern, dass auch \mathcal{M} gleichgradig integrierbar ist. Wir wenden nun das Optional Sampling Theorem auf die beschränkten Stoppzeiten $S \wedge t$, $T \wedge t$ und \mathcal{M} an. Für $A \in \mathcal{F}_S$ ist $\{S \leq t\} \cap A \in \mathcal{F}_{S \wedge t}$, also

$$\mathbf{E}[M_{T \wedge t}, \{S \leq t\} \cap A] = \mathbf{E}[\mathbf{E}[M_{T \wedge t} | \mathcal{F}_{S \wedge t}], \{S \leq t\} \cap A] = \mathbf{E}[M_{S \wedge t}, \{S \leq t\} \cap A].$$

Da nach Lemma 13.22 die Menge $\{M_{S \wedge t}, M_{T \wedge t} : t \in I\}$ gleichgradig integrierbar ist, gilt mit Theorem 8.11

$$\mathbf{E}[M_T, A] = \lim_{t \rightarrow \infty} \mathbf{E}[M_{T \wedge t}, \{S \leq t\} \cap A] = \lim_{t \rightarrow \infty} \mathbf{E}[M_{S \wedge t}, \{S \leq t\} \cap A] = \mathbf{E}[M_S, A],$$

d.h. $\mathbf{E}[M_T|\mathcal{F}_S] = M_S$. Ferner folgt

$$\mathbf{E}[X_T|\mathcal{F}_S] = \mathbf{E}[M_T|\mathcal{F}_S] + A_S + \mathbf{E}[A_T - A_S|\mathcal{F}_S] \geq M_S + A_S = X_S.$$

□

Das Optional Sampling Theorem bietet eine einfache Möglichkeit der Charakterisierung von Martingalen.

Lemma 13.24 (Charakterisierung von Martingalen). *Sei $I = \{0, 1, 2, \dots\}$, und $\mathcal{X} = (X_t)_{t \in I}$ ein adaptierter, stochastischer Prozess. Dann ist \mathcal{X} genau dann ein Martingal, falls $\mathbf{E}[X_S] = \mathbf{E}[X_T]$ für Stoppzeiten S, T gilt, die nur zwei Werte annehmen.*

Beweis. '⇒': Klar nach dem Optional Sampling Theorem.

'⇐': Sei $s \leq t$, $A \in \mathcal{F}_s$ und $T = s1_A + t1_{A^c}$. Dann ist T eine Stoppzeit und es gilt

$$0 = \mathbf{E}[X_t - X_T] = \mathbf{E}[X_t] - \mathbf{E}[X_s, A] - \mathbf{E}[X_t, A^c] = \mathbf{E}[X_t - X_s, A].$$

Da A beliebig war, folgt $\mathbf{E}[X_t|\mathcal{F}_s] = X_s$, also ist \mathcal{X} ein Martingal. □

Beispiel 13.25 (Wald'sche Identitäten, Ruin-Problem). 1. Seien $X_1, X_2, \dots \in \mathcal{L}^1$ unabhängig mit $\mu := \mathbf{E}[X_1] = \mathbf{E}[X_2] = \dots$, und $S_t := \sum_{s=1}^t X_s$. Weiter sei T eine fast sicher beschränkte Stoppzeit. Dann gilt die erste Wald'sche Identität

$$\mathbf{E}[S_T] = \mathbf{E}[T]\mu.$$

Denn: der Prozess $\mathcal{M} = (M_t)_{t=0,1,2,\dots}$ mit $M_0 = 0$, $M_t = S_t - t\mu$ für $t = 1, 2, \dots$ ist ein Martingal, und nach dem Optional Sampling Theorem

$$0 = \mathbf{E}[M_T] = \mathbf{E}[S_T] - \mathbf{E}[T]\mu.$$

Ist außerdem $X_1, X_2, \dots \in L^2$ mit $\sigma^2 = \mathbf{V}[X_1] = \mathbf{V}[X_2] = \dots$ und T unabhängig von X_1, X_2, \dots , so gilt die zweite Wald'sche Identität

$$\mathbf{V}[S_T] = \mathbf{E}[T]\sigma^2 + \mathbf{V}[T]\mu^2.$$

Denn: $(M_t^2 - \langle M \rangle_t)_{t=0,1,2,\dots}$ ist ein Martingal, und $\langle M \rangle_t = t\sigma^2$ nach Beispiel 13.12, also

$$0 = \mathbf{E}[M_T^2 - \langle M \rangle_T] = \mathbf{E}[M_T^2] - \mathbf{E}[T]\sigma^2.$$

Außerdem ist wegen der Unabhängigkeit von T und X_1, X_2, \dots

$$\mathbf{COV}[S_T, T] = \mathbf{E}[\mathbf{E}[X_1 + \dots + X_T|T]T] - \mu\mathbf{E}[T]^2 = \mu\mathbf{V}[T],$$

also

$$\mathbf{E}[M_T^2] = \mathbf{V}[S_T - T\mu] = \mathbf{V}[S_T] + \mu^2\mathbf{V}[T] - 2\mu\mathbf{COV}[S_T, T] = \mathbf{V}[S_T] - \mu^2\mathbf{V}[T].$$

In allen beiden Wald'schen Identitäten kann man die Voraussetzung, dass T beschränkt ist, abschwächen.

2. Seien $k \in \mathbb{N}$ und X_1, X_2, \dots unabhängige und identisch verteilte Zufallsvariablen mit $\mathbf{P}(X_1 = 1) = 1 - \mathbf{P}(X_1 = -1) = p := 1 - q$. Für $N \in \mathbb{N}$ mit $0 < k < N$ sei $S_0 = k$ und $S_t = S_0 + \sum_{i=1}^t X_i$. Weiter sei $T := \inf\{t : S_t \in \{0, N\}\}$ und $p_k := \mathbf{P}(S_T = 0)$. Das bedeutet: man spielt ein Spiel, startend mit Kapital k , bis man entweder ruiniert ist oder Kapital N besitzt. In jedem Schritt gewinnt man mit Wahrscheinlichkeit p eine Kapitaleinheit und verliert mit Wahrscheinlichkeit $q = 1 - p$ eine Kapitaleinheit. Dann ist die Wahrscheinlichkeit, ruiniert zu werden (0 Kapitaleinheiten zu besitzen), gegeben durch p_k .

Im Fall $p = \frac{1}{2}$ ist $(S_t)_{t=0,1,2,\dots}$ ein Martingal, und damit nach dem Optional Sampling Theorem

$$k = \mathbf{E}[S_T] = N(1 - \mathbf{P}(S_T = 0)),$$

also

$$\mathbf{P}(S_T = 0) = \frac{N - k}{N}.$$

Eine ähnliche Rechnung erlaubt die Bestimmung von p_k für den Fall $p \neq \frac{1}{2}$.

Weiter berechnen wir nun mittels des Optional Sampling Theorems für $p \neq \frac{1}{2}$

$$p_k := \mathbf{P}(S_T = 0) = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}. \quad (13.3)$$

Denn: es gilt

$$\mathbf{E}\left[\left(\frac{q}{p}\right)^{X_1}\right] = \frac{q}{p}p + \frac{p}{q}q = 1$$

und damit ist $\mathcal{Y} = (Y_t)_{t=0,1,2,\dots}$, definiert durch $Y_t := \left(\frac{q}{p}\right)^{S_t}$ nach Beispiel 13.5.3 ein Martingal. Da T fast sicher endlich ist, ist $Y_{T \wedge t}$ wegen Proposition 13.19 ein Martingal, das durch 1 und $\left(\frac{q}{p}\right)^N$ beschränkt ist. Wegen Theorem 13.23 ist

$$\left(\frac{q}{p}\right)^k = \mathbf{E}[Y_0] = \mathbf{E}[Y_T] = p_k + (1 - p_k)\left(\frac{q}{p}\right)^N,$$

woraus (13.3) folgt.

3. Betrachten wir einen fairen Münzwurf. Wie lange dauert es, bis zum ersten Mal das Muster ZKZK auftritt? (K und Z stehen hier für Kopf und Zahl.)

Um dies zu berechnen, betrachten wir das folgende Spiel: vor dem ersten Münzwurf wettet eine Spielerin einen Euro auf Z. Falls sie verliert, hört sie auf, falls sie gewinnt, setzt sie vor dem nächsten Wurf zwei Euro auf K. Verliert sie im zweiten Wurf, hört sie auf, im Fall eines Gewinns wettet sie vier Euro auf Z. Verliert sie im dritten Wurf, hört sie auf, gewinnt sie, wettet sie acht Euro auf K. Falls sie also beim vierten Wurf gewinnt, hat sie insgesamt 15 Euro gewonnen. In allen anderen Fällen verliert sie einen Euro.

Nehmen wir nun an, dass vor jedem Münzwurf eine neue Spielerin nach obiger Strategie spielt. Das Spiel wird beendet, wenn das erste Mal eine Spielerin 15 Euro gewinnt.

Sei X_t der Gesamtgewinn aller Spielerinnen bis zur Zeit t und T die Zeit, zu der das Spiel gestoppt wird, weil zum ersten Mal das Muster ZKZK aufgetreten ist. Sicher ist

$$|X_t| \leq 15 \cdot t, \quad \mathbf{P}[T > 4t] \leq \frac{15^t}{16}.$$

Damit hat $(X_{t \wedge T} : t = 1, 2, \dots)$ eine integrierbare Majorante, ist also nach Beispiel 8.8.2 gleichgradig integrierbar. Damit können wir das Optional Stopping-Theorem anwenden, d.h. $(X_{T \wedge t})_{t=1,2,\dots}$ ist ein Martingal.

Sicher ist

$$X_T = 15 - 1 + 3 - 1 - (T - 4)$$

da die ersten $T - 4$ Spielerinnen, sowie Spielerinnen $T - 3$ und $T - 1$ einen Verlust von einem Euro hinnehmen mussten. Spielerin $T - 2$ hat zur Zeit T einen Gewinn von drei Euro und Spielerin $T - 4$ hat 15 Euro gewonnen. Also

$$0 = \mathbf{E}[X_T] = \mathbf{E}[15 - 1 + 3 - 1 - (T - 4)] = -\mathbf{E}[T] - 20,$$

also $\mathbf{E}[T] = 20$. Interessant ist, dass erwartet werden kann, dass beispielsweise das Muster ZKZK schon nach 16 Münzwürfen auftritt.

13.4 Martingalkonvergenzsätze

Wieder ist $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, I abzählbar (hier ist auch erlaubt, dass I dicht in $[0, \infty)$ ist) und $(\mathcal{F}_t)_{t \in I}$ eine Filtration. Wir kennen Konvergenzsätze, etwa das starke Gesetz der großen Zahlen. Martingale konvergieren unter relativ schwachen Voraussetzungen.

Wir beginnen in Proposition 13.27 mit den Doob'schen Ungleichungen. Diese machen Aussagen über die Verteilung von $\sup_{s \leq t} X_s$, falls $\mathcal{X} = (X_t)_{t \in I}$ ein (Sub, Super)-Martingal ist.

Lemma 13.26 (Maximal-Ungleichung). *Ist I höchstens abzählbar und $\mathcal{X} = (X_t)_{t \in I}$ ein Sub-Martingal, so ist für $\lambda > 0$*

$$\lambda \mathbf{P}[\sup_{s \leq t} X_s \geq \lambda] \leq \mathbf{E}[X_t, \sup_{s \leq t} X_s \geq \lambda] \leq \mathbf{E}[|X_t|, \sup_{s \leq t} X_s \geq \lambda].$$

Beweis. Die zweite Ungleichung ist trivial. Für die erste bemerken wir, dass es wegen monotoner Konvergenz (durch Wahl von immer feineren Indexmengen in $[0, t]$) genügt, den diskreten Fall, also etwa $I = \{0, 1, 2, \dots\}$, zu betrachten. Wir erinnern an die Definition von T_B aus Definition 13.17, die nach Bemerkung 13.18.2 eine Stoppzeit ist und setzen

$$T = t \wedge T_{[\lambda; \infty)}.$$

Nach dem Optional Sampling Theorem 13.23 ist

$$\begin{aligned} \mathbf{E}[X_t] &\geq \mathbf{E}[X_T] = \mathbf{E}[X_T; \sup_{s \leq t} X_s \geq \lambda] + \mathbf{E}[X_T; \sup_{s \leq t} X_s < \lambda] \\ &\geq \lambda \mathbf{P}[\sup_{s \leq t} X_s \geq \lambda] + \mathbf{E}[X_t; \sup_{s \leq t} X_s < \lambda]. \end{aligned}$$

Subtrahieren des letzten Terms ergibt die Ungleichung. □

Proposition 13.27 (Doob'sche L^p -Ungleichung). Sei I höchstens abzählbar und $\mathcal{X} = (X_t)_{t \in I}$ ein Martingal oder ein positives Sub-Martingal.

1. Für $p \geq 1$ und $\lambda > 0$ ist

$$\lambda^p \mathbf{P}[\sup_{s \leq t} |X_s| \geq \lambda] \leq \mathbf{E}[|X_t|^p].$$

2. Für $p > 1$ ist

$$\mathbf{E}[|X_t|^p] \leq \mathbf{E}[\sup_{s \leq t} |X_s|^p] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_t|^p].$$

Beweis. Wieder genügt es – wegen monotoner Konvergenz – den Fall $I = \{0, 1, 2, \dots\}$ zu betrachten.

1. Nach Proposition 13.7 ist $(|X_t|^p)_{t \in I}$ ein Sub-Martingal und die Behauptung folgt nach Lemma 13.26.

2. Die erste Ungleichung ist klar. Für die zweite Ungleichung beachte, dass nach Lemma 13.26 gilt, dass

$$\lambda \mathbf{P}\{\sup_{s \leq t} |X_s| \geq \lambda\} \leq \mathbf{E}[|X_s|; \sup_{s \leq t} |X_s| \geq \lambda].$$

Also ist für $K > 0$

$$\begin{aligned} \mathbf{E}[\sup_{s \leq t} (|X_s| \wedge K)^p] &= \mathbf{E}\left[\int_0^{\sup_{s \leq t} |X_s| \wedge K} p\lambda^{p-1} d\lambda\right] \\ &= \mathbf{E}\left[\int_0^K p\lambda^{p-1} 1_{\{\lambda < \sup_{s \leq t} |X_s|\}} d\lambda\right] \\ &= \int_0^K p\lambda^{p-1} \mathbf{P}(\sup_{s \leq t} |X_s| \geq \lambda) d\lambda \\ &\leq \int_0^K p\lambda^{p-2} \mathbf{E}[|X_t|, \sup_{s \leq t} |X_s| \geq \lambda] d\lambda \\ &= p\mathbf{E}\left[|X_t| \int_0^{\sup_{s \leq t} |X_s| \wedge K} \lambda^{p-2} d\lambda\right] \\ &= \frac{p}{p-1} \mathbf{E}[|X_t| (\sup_{s \leq t} |X_s| \wedge K)^{p-1}] \\ &\leq \frac{p}{p-1} \mathbf{E}[\sup_{s \leq t} (|X_s| \wedge K)^p]^{(p-1)/p} \cdot \mathbf{E}[|X_t|^p]^{1/p}, \end{aligned}$$

wobei wir im letzten Schritt die Hölder-Ungleichung verwendet haben. Potenziert man beide Seiten mit p und teilt anschließend durch $\mathbf{E}[\sup_{s \leq t} (|X_s| \wedge K)^p]^{p-1}$, folgt

$$\mathbf{E}[\sup_{s \leq t} (|X_s|)^p] = \lim_{K \rightarrow \infty} \mathbf{E}[\sup_{s \leq t} (|X_s| \wedge K)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_t|^p].$$

□

Für die Martingalkonvergenzsätze ist das Aufkreuzungslemma 13.29 zentral. Bild 13.1 verdeutlicht die Definition einer Aufkreuzung.

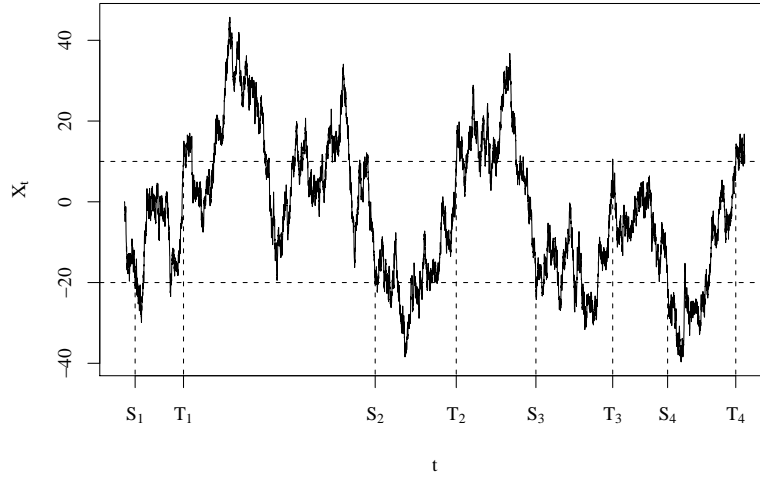


Abbildung 13.1

Eine Illustration der Stoppzeiten $S_1, T_1, S_2, T_2, \dots$ aus Definition 13.28

Definition 13.28. Sei I höchstens abzählbar und $\mathcal{X} = (X_t)_{t \in I}$ ein reellwertiger stochastischer Prozess. Für $a < b$ ist eine Aufkreuzung ein Stück Pfad $(X_r)_{s \leq r \leq s'}$ mit $X_s \leq a$ und $X_{s'} \geq b$. Um die Anzahl solcher Aufkreuzungen zu zählen, führen wir Stoppzeiten $0 =: T_0 < S_1 < T_1 < S_2 < T_2 < \dots$ durch

$$\begin{aligned} S_k &:= \inf\{t \geq T_{k-1} : X_t \leq a\}, \\ T_k &:= \inf\{t \geq S_k : X_t \geq b\} \end{aligned}$$

mit $\inf \emptyset = \infty$ ein. Die k -te Aufkreuzung zwischen a und b ist hier zwischen S_k und T_k . Weiter ist

$$U_{a,b}^t := \sup\{k : T_k \leq t\}$$

die Anzahl der Aufkreuzungen zwischen a und b bis Zeit t .

Lemma 13.29 (Aufkreuzungslemma). Sei I höchstens abzählbar und $\mathcal{X} = (X_t)_{t \in I}$ ein Sub-Martingal. Dann ist

$$\mathbf{E}[U_{a,b}^t] \leq \frac{\mathbf{E}[(X_t - a)^+]}{b - a}.$$

Beweis. Wieder können wir – wegen monotoner Konvergenz – annehmen, dass $I = \{0, 1, 2, \dots\}$. Da nach Proposition 13.7 mit \mathcal{X} auch $((X_t - a)^+)_{t \in I}$ ein Sub-Martingal ist und die Aufkreuzungen zwischen a und b von \mathcal{X} dieselben sind wie die Aufkreuzungen von $((X_t - a)^+)_{t \in I}$ zwischen 0 und $b - a$, können wir \mathbb{E} annehmen, dass $\mathcal{X} \geq 0$ und $a = 0$ gilt. Wir definieren den Prozess $\mathcal{H} = (H_t)_{t \in I}$ durch

$$H_t := \sum_{k \geq 1} 1_{\{S_k < t \leq T_k\}},$$

d.h. $H_t = 1$ genau dann, wenn t in einer Aufkreuzung liegt. Da

$$\{H_t = 1\} = \bigcup_{k \geq 1} \{S_k \leq t - 1\} \cap \{T_k > t - 1\},$$

ist H previsible.

Gegeben $T_k < \infty$ ist offenbar $X_{T_k} - X_{S_k} \geq b$. Weiter ist in diesem Fall

$$(\mathcal{H} \cdot \mathcal{X})_{T_k} = \sum_{i=1}^k \sum_{s=S_{i-1}}^{T_i} (X_s - X_{s-1}) = \sum_{i=1}^k (X_{T_i} - X_{S_i}) \geq kb.$$

Für $t \in \{T_k, \dots, S_{k+1}\}$ ist $(\mathcal{H} \cdot \mathcal{X})_t = (\mathcal{H} \cdot \mathcal{X})_{T_k}$ und für $t \in \{S_k + 1, \dots, T_k\}$ ist $(\mathcal{H} \cdot \mathcal{X})_t \geq (\mathcal{H} \cdot \mathcal{X})_{S_k} = (\mathcal{H} \cdot \mathcal{X})_{T_{k-1}}$. Deswegen ist $(\mathcal{H} \cdot \mathcal{X})_t \geq bU_{0,b}^t$. Aus Proposition 13.14 folgt, dass $((1 - \mathcal{H}) \cdot \mathcal{X})$ ein Sub-Martingal ist, insbesondere $\mathbf{E}[(1 - \mathcal{H}) \cdot \mathcal{X}]_t \geq 0$. Mit $X_t - X_0 = (1 \cdot \mathcal{X})_t = (\mathcal{H} \cdot \mathcal{X})_t + ((1 - \mathcal{H}) \cdot \mathcal{X})_t$ gilt

$$\mathbf{E}[X_t] \geq \mathbf{E}[X_t - X_0] \geq \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t] \geq b\mathbf{E}[U_{0,b}^t].$$

□

Theorem 13.30 (Martingalkonvergenzsatz für Sub-Martingale). Sei $I \subseteq [0, \infty)$ abzählbar, $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t \in I} \mathcal{F}_t)$ und $\mathcal{X} = (X_t)_{t \in I}$ ein Sub-Martingal mit $\sup_{t \in I} \mathbf{E}[X_t^+] < \infty$. Dann gibt es eine Nullmenge N , so dass \mathcal{X} außerhalb von N entlang jeder auf- oder absteigenden Folge in I konvergiert.

Ist insbesondere $I = \{0, 1, 2, \dots\}$, \mathcal{X} ein Sub-Martingal mit $\sup_{t \in I} \mathbf{E}[X_t^+] < \infty$, dann existiert eine \mathcal{F}_∞ -messbare, integrierbare Zufallsvariable X_∞ und $X_t \xrightarrow{t \rightarrow \infty}_{f.s.} X_\infty$.

Beweis. Wegen Lemma 13.29 ist $\mathbf{P}(U_{a,b}^t < \infty) = 1$ für alle a, b, t . Deshalb ist

$$N := \bigcup_{\substack{a < b \\ a, b \in \mathbb{Q}}} \{\sup_{t \in I} U_{a,b}^t = \infty\}$$

eine Nullmenge. Angenommen es existiert eine auf- oder absteigende Folge $t_1, t_2, \dots \in I$, so dass $\mathbf{P}(\liminf_{n \rightarrow \infty} X_{t_n} < \limsup_{n \rightarrow \infty} X_{t_n}) > 0$. Für $a, b \in \mathbb{Q}$ sei

$$B(a, b) := \{\liminf_{n \rightarrow \infty} X_{t_n} < a < b < \limsup_{n \rightarrow \infty} X_{t_n}\}.$$

Da $\{\liminf_{n \rightarrow \infty} X_{t_n} < \limsup_{n \rightarrow \infty} X_{t_n}\} = \bigcup_{a, b \in \mathbb{Q}} B(a, b)$, existieren $a, b \in \mathbb{Q}$ mit $\mathbf{P}(B(a, b)) > 0$. Allerdings gilt $\sup_t U_{a,b}^t = \infty$ auf $B(a, b)$ im Widerspruch dazu, dass N eine Nullmenge ist. Also folgt die fast sichere Konvergenz entlang jeder auf- oder absteigenden Folge.

Sei nun $I = \{0, 1, 2, \dots\}$. Da alle X_t \mathcal{F}_∞ -messbar sind, ist auch X_∞ \mathcal{F}_∞ -messbar. Es bleibt zu zeigen, dass X_∞ integrierbar ist. Nach Fatou's Lemma ist

$$\mathbf{E}[X_\infty^+] \leq \sup_{t \in I} \mathbf{E}[X_t^+] < \infty.$$

Außerdem ist, da \mathcal{X} ein Sub-Martingal ist, wieder mit Fatou's Lemma

$$\mathbf{E}[X_\infty^-] \leq \liminf_{t \rightarrow \infty} \mathbf{E}[X_t^-] = \liminf_{t \rightarrow \infty} (\mathbf{E}[X_t^+] - \mathbf{E}[X_t]) \leq \sup_{t \in I} \mathbf{E}[X_t^+] - \mathbf{E}[X_0] < \infty.$$

□

Korollar 13.31 (Martingalkonvergenzsatz für positive Super-Martingale). Sei $I \subseteq [0, \infty)$ höchstens abzählbar, $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t \in I} \mathcal{F}_t)$ und $\mathcal{X} = (X_t)_{t \in I}$ ein nicht-negatives Super-Martingal. Dann existiert eine \mathcal{F}_u -messbare, integrierbare Zufallsvariable X_u mit $\mathbf{E}[X_u] \leq \mathbf{E}[X_0]$ und $X_t \xrightarrow[t \rightarrow u]{f_s} X_u$.

Beweis. Theorem 13.30, angewandt auf das Sub-Martingal $-\mathcal{X}$ liefert den fast sicheren Limes. Mit dem Lemma von Fatou ist außerdem

$$\mathbf{E}[X_u] \leq \liminf_{t \rightarrow u} \mathbf{E}[X_t] \leq \mathbf{E}[X_0].$$

□

Beispiel 13.32 (Konvergenz von Verzweigungsprozessen). Betrachten wir einen kritischen oder sub-kritischen Verzweigungsprozess $\mathcal{Z} = (Z_t)_{t=0,1,2,\dots}$ aus Beispiel 13.6 (wobei die Nachkommensverteilung nicht degeneriert ist, also nicht $X_i^{(t)} = 1$ fast sicher gilt). Diese sind nicht-negative Super-Martingale, also müssen diese nach Korollar 13.31 fast sicher gegen eine Zufallsvariable Z_∞ konvergieren. Es muss hierbei $\mathbf{P}(Z_\infty > 0) = 0$ gelten, da sonst die fast sichere Konvergenz verletzt ist. (Eine Population mit einer positiven Anzahl von Individuen hat eine positive Wahrscheinlichkeit, in einer Generation ihre Größe zu verändern.) Also ist

$$Z_t \xrightarrow[t \rightarrow \infty]{} Z_\infty := 0$$

fast sicher.

Im Fall des kritischen Verzweigungsprozesses ist es wichtig einzusehen, dass $(Z_t)_{t=0,1,2,\dots,\infty}$ kein Martingal ist, weil $\mathbf{E}[Z_\infty | \mathcal{F}_t] = \mathbf{E}[0 | \mathcal{F}_t] \neq Z_t$ mit positiver Wahrscheinlichkeit gilt.

Ist \mathcal{Z} superkritisch, so ist $(Z_t/\mu^t)_{t=0,1,2,\dots}$ ein nicht-negatives Martingal, das ebenfalls nach obigem Korollar fast sicher konvergiert.

Theorem 13.33 (Konvergenzsatz für gleichgradig integrierbare Martingale). Sei I abzählbar mit $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t \in I} \mathcal{F}_t)$ und $\mathcal{X} = (X_t)_{t \in I}$ ein (Super, Sub)-Martingal. Dann sind folgende Aussagen äquivalent:

1. \mathcal{X} ist gleichgradig integrierbar.
2. Es existiert eine \mathcal{F}_u -messbare Zufallsvariable X_u , so dass $(X_t)_{t \in I \cup u}$ ein (Super, Sub)-Martingal ist.
3. Es existiert eine \mathcal{F}_u -messbare Zufallsvariable X_u mit $X_t \xrightarrow[t \rightarrow u]{f_s, L^1} X_u$.

Beweis. 2. \Rightarrow 1. folgt direkt aus Lemma 12.5. 1. \Rightarrow 3. Wegen Lemma 8.9 ist $\sup_{t \in I} \mathbf{E}[|X_t|] < \infty$. Die fast sichere Konvergenz folgt damit aus Theorem 13.30 und die L^1 -Konvergenz damit aus Theorem 8.11.

3. \Rightarrow 2.: Den Beweis, dass $(X_t)_{t \in I \cup \{u\}}$ ein (Super, Sub)-Martingal ist, führen wir exemplarisch für Sub-Martingale, d.h. $\mathbf{E}[\mathbf{E}[X_u | \mathcal{F}_s]; A] \geq \mathbf{E}[X_s; A]$ für $A \in \mathcal{F}_s$ und $s \in I$. Wegen der L^1 -Konvergenz nach Theorem 12.2.3 ist $\mathbf{E}[|\mathbf{E}[X_t | \mathcal{F}_s] - \mathbf{E}[X_u | \mathcal{F}_s]|] \xrightarrow[t \rightarrow u]{} 0$ und damit gilt für $A \in \mathcal{F}_s$

$$\mathbf{E}[\mathbf{E}[X_u | \mathcal{F}_s]; A] = \lim_{t \rightarrow \infty} \mathbf{E}[\mathbf{E}[X_t | \mathcal{F}_s]; A] \geq \mathbf{E}[X_s; A],$$

d.h. $\mathbf{E}[X_u | \mathcal{F}_s] \geq X_s$ fast sicher. □

Theorem 13.34 (Martingalkonvergenzsatz für L^p -beschränkte Martingale). Sei I abzählbar mit $\sup I = u \in [0, \infty)$, $\mathcal{F}_u = \sigma(\bigcup_{t \in I} \mathcal{F}_t)$, $p > 1$ und $\mathcal{X} = (X_t)_{t \in I}$ ein L^p -beschränktes Martingal. Dann gibt es eine \mathcal{F}_u -messbare Zufallsvariable X_u mit $\mathbf{E}[|X_u|^p] < \infty$, $X_t \xrightarrow{t \uparrow u}_{f_s, L^1} X_u$. Außerdem ist $(|X_t|^p)_{t \in I}$ gleichgradig integrierbar.

Beweis. Wegen Lemma 8.9 ist \mathcal{X} gleichgradig integrierbar. Nach Theorem 13.33 gibt es damit den Limes X_u mit $X_t \xrightarrow{t \uparrow u}_{f_s, L^1} X_u$. Nach der Doob'schen Ungleichung aus Proposition 13.27 ist für $t \in I$

$$\mathbf{E}[\sup_{t \in I} |X_t|^p] = \lim_{t \uparrow u} \mathbf{E}[\sup_{s \leq t} |X_s|^p] \leq \lim_{t \uparrow u} \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_t|^p] < \infty.$$

Damit ist $(|X_t|^p)_{t \in I}$ gleichgradig integrierbar nach Beispiel 8.8.3. Nach dem Lemma von Fatou und Lemma 8.9 ist $\mathbf{E}[|X_u|^p] \leq \sup_{t \in I} \mathbf{E}[|X_t|^p] < \infty$ und Theorem 8.11 liefert die Konvergenz in L^p . \square

Beispiel 13.35 (Verzweigungsprozess). Sei \mathcal{Z} ein Verzweigungsprozess wie in Beispiel 13.6 und Beispiel 13.32 mit $Z_0 = k$. Die quadratische Variation von $\mathcal{Y} = (Y_t)_{t=0,1,2,\dots}$, gegeben $Y_t = Z_t/\mu^t$ ist nach Proposition 13.11 gegeben als

$$\begin{aligned} \langle \mathcal{Y} \rangle_t &= \sum_{s=1}^t \frac{1}{\mu^{2s}} \mathbf{E} \left[\left(\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} - \mu Z_{s-1} \right)^2 \middle| \mathcal{F}_{s-1} \right] \\ &= \sum_{s=1}^t \frac{1}{\mu^{2s}} \mathbf{V} \left[\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} \middle| Z_{s-1} \right] \\ &= \sum_{s=1}^t \frac{1}{\mu^{2s}} Z_{s-1} \cdot \mathbf{V}[X_1^{(1)}]. \end{aligned}$$

Hat also insbesondere die Nachkommensverteilung ein zweites Moment, gilt also $\mathbf{V}[X_1^{(1)}] =: \sigma^2 < \infty$, folgt

$$\mathbf{V}[Y_t] = \sum_{s=1}^t \frac{1}{\mu^{2s}} \mathbf{E}[Z_s] \cdot \sigma^2 = k\sigma^2 \sum_{s=1}^t \frac{1}{\mu^s}.$$

Ist $\mu \leq 1$, ist \mathcal{Y} also nicht L^2 -beschränkt, aber für $\mu > 1$ ist $\sup_{t=0,1,2,\dots} \mathbf{V}[Y_t] < \infty$. Damit gibt es also eine \mathcal{F}_∞ -messbare, quadratisch integrierbare Zufallsvariable Y_∞ , so dass $(Y_t)_{t=0,1,2,\dots,\infty}$ ein Martingal ist.

Beispiel 13.36 (Produkt von Zufallsvariablen). Seien $I = \{1, 2, \dots\}$, X_1, X_2, \dots nicht-negative, unabhängige, integrierbare Zufallsvariable mit $\mathbf{E}[X_t] = 1, t \in I$ und $S_t := \prod_{s=1}^t X_s$ nach Beispiel 13.5.2 ein Martingal. Nach Korollar 13.31 gibt es damit ein S_∞ , so dass $S_t \xrightarrow{t \rightarrow \infty}_{f_s} S_\infty$. Definiere

$$a_t := \mathbf{E}[\sqrt{X_t}].$$

Wir zeigen nun:

$$\{S_t : t \in I\} \text{ gleichgradig integrierbar} \iff \prod_{t=1}^{\infty} a_t > 0.$$

Insbesondere gilt dann auch $S_t \xrightarrow{t \rightarrow \infty}_{L^1} S_\infty$. Im Beweis setzen wir für $t = 1, 2, \dots$

$$W_t := \prod_{s=1}^t \frac{\sqrt{X_s}}{a_s}.$$

Damit ist $(W_t)_{t=1,2,\dots}$ ein Martingal. Auch hier folgt, dass es ein W_∞ gibt mit $W_t \xrightarrow{t \rightarrow \infty}_{f.s.} W_\infty$.

' \Leftarrow ': Wegen der Jensen'schen Ungleichung ist $a_t^2 = (\mathbf{E}[\sqrt{X_t}])^2 \leq \mathbf{E}[X_t] = 1$, also $a_t \leq 1$. Es gilt

$$\sup_{t \in I} \mathbf{E}[W_t^2] = \sup_{t \in I} \mathbf{E}\left[\prod_{s=1}^t \frac{X_s}{a_s^2}\right] = \sup_{t \in I} \prod_{s=1}^t \frac{\mathbf{E}[X_s]}{a_s^2} \leq \frac{1}{\left(\prod_{s=1}^\infty a_s\right)^2} < \infty.$$

Damit ist $(W_t)_{t \in I}$ ein L^2 -beschränktes Martingal, Nach Theorem 13.34 ist $\{W_t^2 : t \in I\}$ gleichgradig integrierbar. Daraus folgt auch die gleichgradige Integrierbarkeit von $\{S_t : t \in I\}$. ' \Rightarrow ': Nehmen wir an, dass $\prod_{s=1}^\infty a_s = 0$. Da W_t einen fast sicheren, endlichen Limes hat, muss $S_t = \prod_{s=1}^t X_s \xrightarrow{t \rightarrow \infty}_{f.s.} 0$ gelten. Falls $\{S_t : t \in I\}$ gleichgradig integrierbar wäre, wäre $0 = \mathbf{E}[S_\infty] = \lim_{t \rightarrow \infty} \mathbf{E}[S_t] = 1$, also ein Widerspruch.

Theorem 13.37 (Konvergenz von bedingten Erwartungswerten).

1. Sei $I \subseteq [0, \infty)$ abzählbar mit $\sup I = u \in (0, \infty]$, $(\mathcal{F}_t)_{t \in I}$ eine Filtration und $\mathcal{F}_u = \sigma(\bigcup_{t \in I} \mathcal{F}_t)$. Dann gilt für $X \in \mathcal{L}^1$, dass

$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t \uparrow u}_{f.s., L^1} \mathbf{E}[X|\mathcal{F}_u].$$

2. Sei $I \subseteq (-\infty, \infty)$ abzählbar mit $\inf I = u \in [-\infty, \infty)$, $(\mathcal{F}_t)_{t \in I}$ eine Filtration und $\mathcal{F}_u = \bigcap_{t \in I} \mathcal{F}_t$. Dann gilt für $X \in \mathcal{L}^1$, dass

$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t \downarrow u}_{f.s., L^1} \mathbf{E}[X|\mathcal{F}_u].$$

Beweis. Wir zeigen nur 1., da der Beweis von 2. analog verläuft. Mit $\mathbf{E}[|\mathbf{E}[X|\mathcal{F}_t]|] \leq \mathbf{E}[|X|] < \infty$ konvergiert nach Theorem 13.30 das Martingal $(\mathbf{E}[X|\mathcal{F}_t])_{t \in I}$ fast sicher. Die L^1 -Konvergenz folgt mit Theorem 13.33 und Lemma 12.5. Der Grenzwert X_u kann hierbei \mathcal{F}_u -messbar gewählt werden. Wir werden nun zeigen, dass $X_u = \mathbf{E}[X|\mathcal{F}_u]$, woraus die Behauptung folgt.

Klar ist, dass $\mathbf{E}[\mathbf{E}[X|\mathcal{F}_t], A] = \mathbf{E}[X, A]$ für alle $A \in \mathcal{F}_s$ und $s \leq t$ gilt. Mit $t \uparrow u$ ist damit $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ für alle $A \in \mathcal{F}_s$ und mit $s \uparrow u$ gilt auch $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ für alle $A \in \mathcal{F}_u$. Da X_u nach \mathcal{F}_u -messbar ist, gilt damit $X_u = \mathbf{E}[X|\mathcal{F}_u]$. \square

Wir kommen nun zu Rückwärtsmartingalen, das sind Martingale mit nach unten unbeschränkter Indexmenge $I \subseteq (-\infty, 0]$. Diese konvergieren unter sehr schwachen Voraussetzungen.

Theorem 13.38 (Martingalkonvergenzsatz für Rückwärtsmartingale). Sei $I \subseteq (-\infty, 0]$ diskret, $\inf I = u \in (-\infty, 0]$, $\mathcal{F}_u = \bigcap_{t \in I} \mathcal{F}_t$ und $\mathcal{X} = (X_t)_{t \in I}$ ein Sub-Martingal. Dann sind äquivalent

1. Es gibt eine \mathcal{F}_u -messbare, integrierbare Zufallsvariable X_u mit $X_t \xrightarrow{t \downarrow u}_{f.s., L^1} X_u$

2. $\inf_{t \in I} \mathbf{E}[X_t] > -\infty$.

Dann ist auch $(X_t)_{t \in I \cup \{u\}}$ ein Sub-Martingal. Insbesondere konvergiert jedes Rückwärtsmartingal fast sicher und in L^1 .

Beweis. Ohne Einschränkung sei $I = \{\dots, -2, -1, 0\}$ und $u = -\infty$.

'1. \Rightarrow 2.': Aus der Konvergenz im Mittel folgt

$$\inf_{t \in I} \mathbf{E}[X_t] = \lim_{t \rightarrow -\infty} \mathbf{E}[X_t] = \mathbf{E}[X_{-\infty}] > -\infty.$$

'2. \Rightarrow 1.': Die fast sichere Konvergenz folgt wie im Beweis von Theorem 13.30, wobei die Bedingung $\sup_{t \in I} \mathbf{E}[X_t^+] < \infty$ wegen $I \subseteq (-\infty, 0]$ durch $\inf_{t \in I} \mathbf{E}[X_t^-] < \infty$ ersetzt werden muss. Weiter definieren wir für $t = \dots, -2, -1, 0$

$$Y_t := \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}] \geq 0.$$

Dann gilt

$$\mathbf{E}\left[\sum_{t=0}^{-\infty} Y_t\right] = \mathbf{E}[X_0] - \inf_{t \in I} \mathbf{E}[X_t] < \infty.$$

Damit ist $\sum_{t=0}^{-\infty} Y_t < \infty$ fast sicher, und wir definieren

$$A_t = \sum_{s \leq t} Y_s, \quad M_t = X_t - A_t$$

Nun ist $(A_t)_{t \in I}$ wegen $\mathbf{E}[A_0] < \infty$ gleichgradig integrierbar, und $(M_t)_{t \in I}$ ist wegen Lemma 12.5 gleichgradig integrierbar. Damit ist \mathcal{X} gleichgradig integrierbar, und die L^1 -Konvergenz folgt. Der Beweis, dass $(X_t)_{t \in I \cup \{-\infty\}}$ ein Sub-Martingal ist, verläuft analog zum Beweis in 13.33. \square

Beispiel 13.39 (Das starke Gesetz der großen Zahlen). Seien $X_1, X_2, \dots \in L^1$ unabhängig identisch verteilt. Für $t \in \{\dots, -2, -1\}$ setzen wir wie im Beispiel 13.5.2

$$S_t := \frac{1}{|t|} \sum_{s=1}^{|t|} X_s$$

und $\mathcal{F}_t = \sigma(\dots, S_{t-1}, S_t) = \sigma(S_t, X_{t+1}, X_{t+2}, \dots)$. Dann ist $(S_t)_{t \in I}$ ein Rückwärtsmartingal mit $S_t = \mathbf{E}[X_1 | \mathcal{F}_t]$. Nach Theorem 13.38 konvergiert S_t fast sicher und in L^1 gegen eine Zufallsgröße $S_{-\infty}$. Diese ist messbar bzgl. $\mathcal{F}_{-\infty}$, jedoch auch bzgl. $\mathcal{T}(X_1, X_2, \dots)$, der terminalen σ -Algebra der Familie $\{X_1, X_2, \dots\}$. Da diese σ -Algebra nach dem Kolmogoroff'schen 0-1-Gesetz trivial ist, ist $S_{-\infty}$ fast sicher konstant. Da $(S_t)_{t \in I \cup \{-\infty\}}$ ein Martingal ist, folgt also

$$\frac{1}{|t|} \sum_{s=1}^{|t|} X_s = S_t \xrightarrow[t \rightarrow -\infty]{f.s., L^1} S_{-\infty} = \mathbf{E}[S_{-\infty}] = \mathbf{E}[S_{-1}] = \mathbf{E}[X_1].$$

Die fast sichere Konvergenz ist jedoch genau die Aussage des Gesetzes der großen Zahlen.

Wir kommen nun noch zu einer Anwendung der Martingalkonvergenzsätze, einer Verbesserung des Lemmas von Borel-Cantelli, Theorem 9.8. Hierzu benötigen Wir ein Lemma.

Lemma 13.40 (Konvergenz und wachsender Prozess). Sei $\mathcal{M} = (M_t)_{t=0,1,2,\dots}$ ein L^2 -integrierbares Martingal, wobei $|M_t - M_{t-1}| \leq K$ für ein K und alle $t = 1, 2, \dots$ gilt. Dann gibt es eine Nullmenge N , so dass

$$\begin{aligned} \{\langle \mathcal{M} \rangle_\infty < \infty\} &\subseteq \{\lim_{t \rightarrow \infty} M_t \text{ existiert}\} \cup N, \\ \{\langle \mathcal{M} \rangle_\infty = \infty\} &\subseteq \{\lim_{t \rightarrow \infty} M_t / \langle \mathcal{M} \rangle_t = 0\} \cup N. \end{aligned}$$

Beweis. Wir beginnen mit der ersten Aussage. Zunächst ist für jedes $k = 1, 2, 3, \dots$ die zufällige Zeit

$$T_k := \inf\{t : \langle \mathcal{M} \rangle_t > k\}$$

eine Stoppzeit. Daraus folgt bereits

$$\{\langle \mathcal{M} \rangle_\infty < \infty\} = \bigcup_{k=1}^{\infty} \{T_k = \infty\}. \quad (13.4)$$

Weiter ist damit der gestoppte Prozess $(\langle \mathcal{M} \rangle_{t \wedge T_k})_{t=0,1,2,\dots}$ prävisibel, denn für $A \in \mathcal{B}(\mathbb{R})$ gilt

$$\{\langle \mathcal{M} \rangle_{t \wedge T_k} \in A\} = (\{T_k > t-1\} \cap \{\langle \mathcal{M} \rangle_t \in A\}) \cup \bigcup_{s=0}^{t-1} \{T_k = s, \langle \mathcal{M} \rangle_s \in A\} \in \mathcal{F}_{t-1}.$$

Betrachten wir nun die Martingale $(\mathcal{M}^{T_k})^2 - \langle \mathcal{M} \rangle^{T_k} = (\mathcal{M}^2 - \langle \mathcal{M} \rangle)^{T_k}$ für $k = 1, 2, \dots$. Es ist $\langle \mathcal{M}^{T_k} \rangle = \langle \mathcal{M} \rangle^{T_k}$ und $\langle \mathcal{M} \rangle^{T_k}$ ist durch $k + K^2$ beschränkt. Also ist \mathcal{M}^{T_k} in L^2 beschränkt und konvergiert damit fast sicher. Auf einer Menge $\{T_k = \infty\}$ konvergiert jedoch \mathcal{M}^{T_k} genau dann, wenn \mathcal{M} konvergiert. Zusammen mit (13.4) folgt die Aussage.

Für die zweite Aussage betrachten wir das Martingal $\mathcal{X} := (1 + \langle \mathcal{M} \rangle)^{-1} \cdot \mathcal{M}$. Da $(1 + \langle \mathcal{M} \rangle)^{-1}$ beschränkt ist und \mathcal{M} ein L^2 -integrierbares Martingal ist, ist \mathcal{X} ein L^2 -integrierbares Martingal. Außerdem ist nach Beispiel 15.16

$$\begin{aligned} \langle \mathcal{X} \rangle_t &= \left(\frac{1}{(1 + \langle \mathcal{M} \rangle)^2} \cdot \langle \mathcal{M} \rangle \right)_t = \sum_{s=1}^t \frac{1}{(1 + \langle \mathcal{M} \rangle_s)^2} (\langle \mathcal{M} \rangle_s - \langle \mathcal{M} \rangle_{s-1}) \\ &\leq \sum_{s=1}^t \frac{1}{(1 + \langle \mathcal{M} \rangle_s)(1 + \langle \mathcal{M} \rangle_{s-1})} (\langle \mathcal{M} \rangle_s - \langle \mathcal{M} \rangle_{s-1}) = \sum_{s=1}^t \frac{1}{1 + \langle \mathcal{M} \rangle_{s-1}} - \frac{1}{1 + \langle \mathcal{M} \rangle_s} \\ &= 1 - \frac{1}{1 + \langle \mathcal{M} \rangle_t}. \end{aligned}$$

Damit konvergiert nach 1. das Martingal \mathcal{X} , d.h. insbesondere

$$\sum_{s=1}^{\infty} \frac{M_s - M_{s-1}}{1 + \langle \mathcal{M} \rangle_s} < \infty.$$

Nun liefert das Kronecker-Lemma 9.24, dass

$$\frac{\sum_{s=1}^t M_s - M_{s-1}}{\langle \mathcal{M} \rangle_t} \xrightarrow{t \rightarrow \infty} 0$$

auf $\{\langle \mathcal{M} \rangle_\infty = \infty\}$. □

Theorem 13.41 (Erweiterung des Borel-Cantelli Lemmas). Sei $A_t \in \mathcal{F}_t$, $t = 0, 1, 2, \dots$ sowie

$$X_s := \mathbf{P}(A_s | \mathcal{F}_{s-1}).$$

1. Auf $\sum_{t=1}^{\infty} X_t < \infty$ treten nur endlich viele der A_t ein, d.h.

$$\left\{ \sum_{t=1}^{\infty} X_t < \infty \right\} \subseteq \left\{ \sum_{t=1}^{\infty} 1_{A_t} < \infty \right\}.$$

2. Auf $\sum_{t=1}^{\infty} X_t = \infty$ gilt $\sum_{t=1}^{\infty} 1_{A_t} / \sum_{t=1}^{\infty} X_t = 1$, also

$$\left\{ \sum_{t=1}^{\infty} X_t = \infty \right\} \subseteq \left\{ \sum_{t=1}^{\infty} 1_{A_t} / \sum_{t=1}^{\infty} X_t = 1 \right\} \subseteq \left\{ \sum_{t=1}^{\infty} 1_{A_t} = \infty \right\}.$$

Bemerkung 13.42 (Erweiterung). Das Borel-Cantelli-Lemma aus Theorem 9.8 kann nun einfach hergeleitet werden. Ist nämlich

$$\mathbf{E} \left[\sum_{t=1}^{\infty} X_t \right] = \sum_{t=1}^{\infty} \mathbf{P}(A_t) < \infty,$$

so gilt $\sum_{t=1}^{\infty} X_t < \infty$ fast sicher. Die Aussage liefert nun, dass höchstens endlich viele der A_n eintreten. Falls weiter A_1, A_2, \dots unabhängig sind, so setzen wir $\mathcal{F}_t = \sigma(A_1, \dots, A_t)$ und damit $X_s = \mathbf{E}[1_{A_s} | \mathcal{F}_{s-1}] = \mathbf{P}(A_s)$. Gilt nun $\sum_{t=1}^{\infty} \mathbf{P}(A_t) = \infty$, treten also unendlich viele der A_n ein.

Beweis. Wir betrachten das Martingal \mathcal{M} mit

$$M_t = \sum_{s=1}^t 1_{A_s} - X_s.$$

Es gilt

$$\langle \mathcal{M} \rangle_t = \sum_{s=1}^t \mathbf{E}[1_{A_s}^2 - X_s^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t X_s(1 - X_s) \leq \sum_{s=1}^t X_s.$$

Ist nun $\sum_{t=1}^{\infty} X_t < \infty$, so konvergiert \mathcal{M} nach Lemma 13.40.1. Also gilt auch $\sum_{t=1}^{\infty} 1_{A_t} < \infty$. Ist nun $\sum_{t=1}^{\infty} X_t = \infty$ und $\langle \mathcal{M} \rangle_{\infty} < \infty$, so konvergiert \mathcal{M} und die Behauptung ist klar.

Ist nun $\sum_{t=1}^{\infty} X_t = \infty$ und $\langle \mathcal{M} \rangle_{\infty} = \infty$, so gilt $M_t / \langle \mathcal{M} \rangle_t \xrightarrow{t \rightarrow \infty} 0$ nach Lemma 13.40.2. Daraus folgt dann

$$\left| \frac{\sum_{s=1}^t 1_{A_s}}{\sum_{s=1}^t X_s} - 1 \right| = \left| \frac{M_t}{\sum_{s=1}^t X_s} \right| \leq \left| \frac{M_t}{\langle \mathcal{M} \rangle_t} \right| \xrightarrow{t \rightarrow \infty} 0.$$

□

13.5 Der zentrale Grenzwertsatz für Martingale

Der zentrale Grenzwertsatz aus Abschnitt 11.2 gibt die Konvergenz einer Summe von *unabhängigen* Zufallsvariablen – geeignet transformiert – gegen eine Normalverteilung. Nun behandeln wir den Fall einer Folge von Martingalen $\mathcal{M}^1 = (M_t^1)_{t=0,1,2,\dots}, \mathcal{M}^2 = (M_t^2)_{t=0,1,2,\dots}, \dots$, jeweils gestartet in 0, die wir mittels $X_t^n := M_t^n - M_{t-1}^n, t = 1, 2, \dots$ wieder als Summe schreiben können, da ja nun $M_t^n = X_1^n + \dots + X_t^n$ gilt. Man beachte nun, dass die Familie X_1^n, X_2^n, \dots nicht unabhängig sein müssen. Dennoch können wir – unter geeigneten Voraussetzungen die Konvergenz in Verteilung gegen eine normalverteilte Zufallsvariable zeigen.

Theorem 13.43 (Zentraler Grenzwertsatz für Martingale). *Sei $I^n = \{0, 1, 2, \dots, t_n\}$ und $\mathcal{M}^n = (M_t^n)_{t \in I^n}$ ein Martingal mit $M_0^n = 0$ bezüglich einer Filtration $\mathcal{F}^n = (\mathcal{F}_t^n)_{t \in I^n}$, $n = 1, 2, \dots$. Für $X_t^n := M_t^n - M_{t-1}^n$ (mit $t = 1, \dots, t_n$) gelte*

$$\mathbf{E}[\max_{1 \leq s \leq t_n} |X_s^n|] \xrightarrow{n \rightarrow \infty} 0, \quad (13.5)$$

$$\sum_{s=1}^{t_n} (X_s^n)^2 \xrightarrow{n \rightarrow \infty}_p \sigma^2 > 0. \quad (13.6)$$

Dann ist $M_{t_n}^n \xrightarrow{n \rightarrow \infty} X$ mit $X \sim N(0, \sigma^2)$.

Wir benötigen im Beweis des Theorems zwei Lemmas.

Lemma 13.44 (Konvergenz von Produkten von Zufallsvariablen). *Seien $U_1, U_2, \dots, T_1, T_2, \dots$ Zufallsvariablen, die den folgenden Bedingungen genügen:*

1. $U_n \xrightarrow{n \rightarrow \infty}_p u$,
2. $(T_n)_{n=1,2,\dots}$ und $(T_n U_n)_{n=1,2,\dots}$ sind uniform integrierbar,
3. $\mathbf{E}[T_n] \xrightarrow{n \rightarrow \infty} 1$,

Dann gilt $\mathbf{E}[T_n U_n] \xrightarrow{n \rightarrow \infty} u$.

Beweis. Wegen 3. genügt es zu zeigen, dass $\mathbf{E}[T_n(U_n - u)] \xrightarrow{n \rightarrow \infty} 0$. Hierzu zeigen wir lediglich $T_n(U_n - u) \xrightarrow{n \rightarrow \infty}_p 0$, was die L^1 -Konvergenz wegen 2. impliziert. Insbesondere gilt dann $\mathbf{E}[T_n(U_n - u)] \xrightarrow{n \rightarrow \infty} 0$. Sei hierzu $\varepsilon > 0$ und K groß genug, so dass $\sup_n \mathbf{P}(|T_n| > K) \leq \varepsilon$. (Ein solches K existiert wegen der uniformen Integrierbarkeit von T_1, T_2, \dots) Nun schreiben wir (man beachte, dass für $x, y \geq 0$ und $\delta, \varepsilon > 0$ immer gilt, dass $xy > \delta\varepsilon \Rightarrow x > \delta$ oder $y > \varepsilon$)

$$\limsup_{n \rightarrow \infty} \mathbf{P}(|T_n(U_n - u)| > \varepsilon) \leq \limsup_{n \rightarrow \infty} \mathbf{P}(|U_n - u| > \varepsilon/K) + \mathbf{P}(|T_n| > K) \leq \varepsilon.$$

Daraus folgt bereits die Behauptung. □

Lemma 13.45 (Abschätzung der Exponentialfunktion). *1. Es gibt ein $C > 0$ und eine Funktion r mit $|r(x)| \leq C|x^3|$, so dass*

$$\exp(ix) = (1 + ix) \exp(-x^2/2 + r(x))$$

für alle $x \in \mathbb{R}$ gilt.

2. Es gilt $|1 + ix| \leq e^{x^2/2}$ für alle $x \in \mathbb{R}$.

Beweis. 1. Es genügt, die Behauptung für kleine $|x|$ zu zeigen, da sie für große $|x|$ trivial ist. Mit Hilfe von Lemma 11.12 schreiben wir

$$\begin{aligned} & \left| \exp(ix) - (1 + ix) \exp(-x^2/2) \right| \\ &= \left| \exp(ix) - 1 - ix + x^2/2 - (1 + ix)(\exp(-x^2/2) - 1 + x^2/2) + ix^3/3 \right| \\ &\leq \left| \exp(ix) - 1 - ix + x^2/2 \right| + |1 + ix| \cdot \left| \exp(-x^2/2) - 1 + x^2/2 \right| + |x^3/3| \\ &\leq \frac{|x^3|}{6} + |1 + ix| \cdot \left(\frac{|x^2|}{2} \wedge \frac{|x^4|}{8} \right) + \frac{|x^3|}{3} \leq |x^3| \end{aligned}$$

für alle x . Daraus folgt die Behauptung für kleine $|x|$, und damit ist 1. bewiesen. Für 2. genügt es, $|1 + ix|^2 = 1 + x^2 \leq e^{x^2}$ zu schreiben und die Wurzel zu ziehen. \square

Beweis von Theorem 13.43. Zunächst definieren wir

$$Z_s^n := X_s^n 1_{\sum_{r=1}^{s-1} (X_r^n)^2 \leq 2\sigma^2}$$

sowie $N_t^n := \sum_{s=1}^t Z_s^n$. Dann ist $(N_t^n)_{t=1,2,\dots}$ ein $(\mathcal{F}_t^n)_{t \in I^n}$ -Martingal, denn

$$\mathbf{E}[N_t^n - N_{t-1}^n | \mathcal{F}_{t-1}^n] = \mathbf{E}[Z_t^n | \mathcal{F}_{t-1}^n] = 1_{\sum_{r=1}^{s-1} (X_r^n)^2 \leq 2\sigma^2} \cdot \mathbf{E}[X_t^n | \mathcal{F}_{t-1}^n] = 0,$$

da $M_t^n = X_1^n + \dots + X_t^n$. Nun ist

$$\begin{aligned} \mathbf{P}\left(\max_{t=1,\dots,t_n} |M_t^n - N_t^n| > 0\right) &= \mathbf{P}(M_t^n \neq N_t^n \text{ für ein } t \in I^n) \\ &= \mathbf{P}(X_t^n \neq Z_t^n \text{ für ein } t \in I^n) \\ &= \mathbf{P}\left(\sum_{s=1}^{t_n} (X_s^n)^2 > 2\sigma^2\right) \xrightarrow{n \rightarrow \infty} 0, \end{aligned} \tag{13.7}$$

wobei die Konvergenz aus (13.6) folgt. Nun gilt also $M_{t_n}^n - N_{t_n}^n \xrightarrow{n \rightarrow \infty} 0$, also genügt es nach dem Satz von Slutsky, Korollar 10.9, $N_{t_n}^n \xrightarrow{n \rightarrow \infty} X \sim N(0, \sigma^2)$ zu zeigen. Hierfür werden wir für beliebiges $\lambda \in \mathbb{R}$

$$\mathbf{E}[e^{i\lambda N_{t_n}^n}] \xrightarrow{n \rightarrow \infty} e^{-i\lambda^2 \sigma^2 / 2}$$

zeigen. Mit der Funktion r aus Lemma 13.45 gilt nun

$$\mathbf{E}[e^{i\lambda N_{t_n}^n}] = \prod_{s=1}^{t_n} (1 + i\lambda Z_s^n) \cdot \exp\left(-\frac{\lambda^2}{2} \sum_{s=1}^{t_n} (Z_s^n)^2 + \sum_{s=1}^{t_n} r(\lambda Z_s^n)\right).$$

Wir setzen nun

$$T_n := \prod_{s=1}^{t_n} (1 + i\lambda Z_s^n), \quad U_n := \exp\left(-\frac{\lambda^2}{2} \sum_{s=1}^{t_n} (Z_s^n)^2 + \sum_{s=1}^{t_n} r(\lambda Z_s^n)\right)$$

und zeigen, dass für diese Zufallsvariablen die Voraussetzungen von Lemma 13.44 gelten (mit $u = e^{-\lambda^2 \sigma^2 / 2}$). Für 1. ist zunächst wegen (13.7)

$$\lim_{n \rightarrow \infty} \sum_{s=1}^{t_n} (Z_s^n)^2 = \lim_{n \rightarrow \infty} \sum_{s=1}^{t_n} (X_s^n)^2 = \sigma^2.$$

Weiter ist mit C aus Lemma 13.45

$$\begin{aligned} \left| \sum_{s=1}^{t_n} r(\lambda Z_s^n) \right| &\leq C \cdot |\lambda^3| \cdot \sum_{s=1}^{t_n} |Z_s^n|^3 \leq C \cdot |\lambda^3| \cdot \sum_{s=1}^{t_n} |X_s^n|^3 \\ &\leq C \cdot |\lambda^3| \cdot \max_{1 \leq s \leq t_n} |X_s^n| \cdot \sum_{s=1}^{t_n} |X_s^n|^2 \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

wobei die Konvergenz aus (13.5) und (13.6) folgt.

Für 2. gilt zunächst $|T_n U_n| = |e^{i\lambda N_{t_n}^n}| = 1$, woraus bereits die uniforme Integrierbarkeit von $(T_n U_n)_{n \in I^n}$ folgt. Für die uniforme Integrierbarkeit von $(T_n)_{n \in I^n}$ definieren wir

$$J_n := \inf \left\{ s \leq t_n : \sum_{r=1}^s (X_r^n)^2 > 2\sigma^2 \right\} \wedge t_n$$

und schreiben

$$\begin{aligned} |T_n| &= \prod_{s=1}^{J_n-1} |1 + i\lambda Z_s^n| \cdot |1 + i\lambda Z_{J_n}^n| \leq \exp\left(\frac{\lambda^2}{2} \sum_{s=1}^{J_n-1} (X_s^n)^2\right) (1 + \lambda |X_{J_n}^n|) \\ &\leq \exp(\lambda^2 \sigma^2) \cdot (1 + |\lambda| \cdot \max_{1 \leq s \leq t_n} |X_s^n|). \end{aligned}$$

Da $\max_{1 \leq s \leq t_n} |X_s^n| \xrightarrow{n \rightarrow \infty} L^1 0$, ist insbesondere die Familie $(\max_{1 \leq s \leq t_n} |X_s^n|)_{n=1,2,\dots}$ uniform integrierbar, woraus die uniforme Integrierbarkeit von $(T_n)_{n=1,2,\dots}$ folgt.

Wir kommen nun zu 3., indem wir $\mathbf{E}[T_n] = 1$ zeigen. Wegen $\mathbf{E}[Z_s^n | \mathcal{F}_{s-1}^n] = 0$ für alle $s = 1, \dots, t_n$ ist nämlich

$$\begin{aligned} \mathbf{E}[T_n] &= \mathbf{E}\left[\prod_{s=1}^{t_n} (1 + i\lambda Z_s^n)\right] \\ &= \mathbf{E}\left[(1 + i\lambda Z_1^n) \cdot \mathbf{E}[(1 + i\lambda Z_2^n) \cdots \mathbf{E}[1 + \lambda Z_{t_n}^n | \mathcal{F}_{t_n-1}^n] \cdots | \mathcal{F}_1^n]]\right] = 1. \end{aligned}$$

Nun folgt die behauptung direkt mit Lemma 13.45. \square

Beispiel 13.46. 1. Seien X_1, X_2, \dots unabhängige, identisch verteilte, reellwertige Zufallsvariable mit $\mathbf{E}[X_1] = 0$ und endlicher Varianz $\mathbf{V}[X_1] = \sigma^2$. Bekannterweise ist dann $\mathcal{M}^n = (M_t^n)_{t=0,1,2,\dots}$ mit

$$M_t^n = \frac{1}{\sqrt{n}} \sum_{s=1}^t X_s$$

ein Martingal und

$$M_n^n \xrightarrow{n \rightarrow \infty} X \sim N(0, \sigma^2).$$

Dies lässt sich auch mittels Theorem 13.43 zeigen: zunächst stellen wir fest, dass $\int_0^\infty t \mathbf{P}(|X_1| > t) dt < \infty$ wegen des endlichen zweiten Moments ist. Damit ist $\mathbf{P}(|X_1| > t) = o(1/t^2)$ für $t \rightarrow \infty$, lässt sich also schreiben als $\mathbf{P}(|X_1| > t) = a(t)/t^2$ mit $a(t) \xrightarrow{t \rightarrow \infty} 0$. Daraus folgt

$$\begin{aligned} \mathbf{E}\left[\max_{1 \leq s \leq n} |X_s| / \sqrt{n}\right] &= \int_0^\infty \mathbf{P}(\max_{1 \leq s \leq n} |X_s| > t\sqrt{n}) dt = \int_0^\infty 1 - (1 - \mathbf{P}(|X_1| > t\sqrt{n}))^n dt \\ &= \int_0^\infty 1 - \left(1 - \frac{a(t\sqrt{n})}{t^2 n}\right)^n dt \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

wegen majorisierter Konvergenz. Weiter gilt mit dem Gesetz großer Zahlen, dass

$$\frac{1}{n} \sum_{s=1}^n X_s^2 \xrightarrow[n \rightarrow \infty]{f.s.} \sigma^2.$$

Also sind die Voraussetzungen von Theorem 13.43 erfüllt.

2. Wir bringen noch ein Beispiel einer Folge von Martingalen, die auf Summen von abhängigen Zufallsvariablen führen. Hierfür erinnern wir an das stochastische Integral aus Definition 13.13. Seien Y_1, Y_2, \dots unabhängige, identisch verteilte, beschränkte Zufallsvariable mit $\mathbf{E}[Y_1] = 0$ und $\mathbf{V}[Y_1] = 1$ sowie $\mathcal{H} = (H_t)_{t=0,1,2,\dots}$ und $\mathcal{M}^n = (M_t^n)_{t=0,1,2,\dots}$ gegeben als

$$H_s = \frac{1}{s-1} (Y_1^2 + \dots + Y_{s-1}^2), \quad M_t^n = \frac{1}{\sqrt{n}} \sum_{s=1}^t Y_s.$$

Dann ist

$$(\mathcal{H} \cdot \mathcal{M}^n)_t = \frac{1}{\sqrt{n}} \sum_{s=1}^t Y_s \frac{1}{s-1} \sum_{r=1}^{s-1} Y_r^2$$

ein Martingal mit

$$X_t^n := (\mathcal{H} \cdot \mathcal{M}^n)_t - (\mathcal{H} \cdot \mathcal{M}^n)_{t-1} = \frac{1}{\sqrt{n}} Y_t \frac{1}{t-1} \sum_{r=1}^{t-1} Y_r^2.$$

(Man beachte, dass (X_1^n, X_2^n, \dots) keine unabhängige Familie ist.) Nun gilt (13.5) wegen der Beschränktheit von Y_1, Y_2, \dots . Weiter berechnen wir

$$\sum_{s=1}^n (X_s^n)^2 = \frac{1}{n} \sum_{s=1}^n Y_s^2 \left(\frac{1}{s-1} \sum_{r=1}^{s-1} Y_r^2 \right)^2 \xrightarrow[n \rightarrow \infty]{} 1,$$

woraus nun $(\mathcal{H} \cdot \mathcal{M}^n)_n \xrightarrow[n \rightarrow \infty]{} X \sim N(0, 1)$ folgt.

14 Ausblicke

Heutzutage werden viele Dinge des realen Lebens mit Hilfe stochastischer Methoden analysiert. Wir geben hier noch beispielhaft ein paar Anmerkungen aus drei Gebieten.

14.1 Zufällige Graphen

Mittels Graphen werden oftmals (mögliche) Interaktionen modelliert. Insgesamt steht dann der Graph für ein soziales Netzwerk, oder aber für zulluläre Prozesse, oder eben irgendeinen Lebensbereich, der sich mit einem Graph darstellen lässt. Basierend auf der Arbeit

Hermann, F., Pfaffelhuber, P. Large-scale behavior of the partial duplication random graph. ALEA, Lat. Am. J. Probab. Math. Stat. 13, 687–710, 2016.

soll nun ein Modell für einen zufälligen Graphen vorgestellt werden. Ein paar Resultate illustrieren außerdem, wie wichtig Martingale sein können.

Definition 14.1 (Graph, Grad, Clique).

1. Ein (ungerichteter, schleifenfreier) Graph ist ein Tupel $G = (V, E)$, wobei V die Menge der Knoten und $E \subseteq \{\{v, w\} : v, w \in V, v \neq w\}$ die Menge der Kanten ist.
2. Eine k -Clique in $G = (V, E)$ ist eine Teilmenge $V' \subseteq V$ mit $|V'| = k$ und $\{\{v, w\} : v, w \in V', v \neq w\} \subseteq E$ (also ein vollständiger Teilgraph mit k Knoten). Wir bezeichnen mit $C_k(G)$ die Anzahl der k -Cliquen in G .
3. Für den Graphen $G = (V, E)$ und $v \in V$, definieren wir den Grad von v durch

$$D_v := D_v(G) := |\{w : \{v, w\} \in E\}|.$$

Außerdem ist die (absolute und relative) Gradverteilung gegeben durch $(F_k(G))_{k=0,1,2,\dots}$ bzw. $(F_k^\circ(G))_{k=0,1,2,\dots}$ mittels

$$F_k(G) := |\{v : D_v(G) = k\}|, \quad F_k^\circ(G) := \frac{1}{|V|} F_k(G).$$

Die erzeugende Funktion der Gradverteilung ist

$$H_q(G) := \sum_{k=0}^{\infty} F_k(G) q^k, \quad H_q^\circ(G) := \sum_{k=0}^{\infty} F_k^\circ(G) q^k \quad \text{für } q \in [0, 1].$$

In Abbildung 14.1 ist eine Illustration der folgenden Definition.

Definition 14.2 (Partielles Duplikations-Modell). Sei $p \in [0, 1]$. Wir definieren eine Markov-Kette $\mathcal{G} = (G_n)_{n=n_0, n_0+1, \dots}$ mit Werten in Graphen und nennen sie das Partielle Duplikations-Modell, wobei $G_n = (V_n, E_n)$ der Graph zur Zeit n mit Knotenmenge V_n und Kantenmenge $E_n \subseteq \{\{v, w\} : v, w \in V_n, v \neq w\}$ ist, $n = n_0, n_0 + 1, \dots$. Startend in $G_{n_0} = (V_{n_0}, E_{n_0})$ mit $|V_{n_0}| = n_0$, ist die Dynamik wiefolgt: In Zeitschritt $n + 1$ wird ein Knoten $v \in V_n$ zufällig ausgewählt. Daraufhin wird ein neuer Knoten v' zum Graph hinzugefügt und jede Verbindung zu v (d.h. jede Kante $e = \{v, w\}$ für ein $w \in V_n$) unabhängig mit Wahrscheinlichkeit p kopiert, d.h. $\{v', w\} \in E_{n+1}$ mit Wahrscheinlichkeit p .

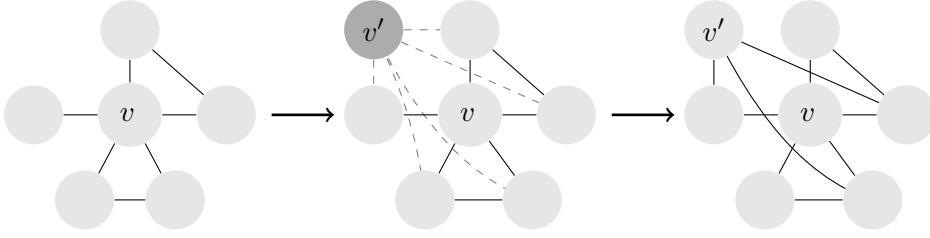


Abbildung 14.1: Illustration eines Schrittes im partiellen Duplikations-Modell; siehe auch Definition 14.2. Zur Zeit $n = 6$ (da im linken Bild 6 Knoten zu sehen sind), wird Knoten v ausgewählt, und v' erzeugt. Anschließend wird jede gestrichelte Linie mit Wahrscheinlichkeit p beibehalten. Das Ergebnis ist ein Graph mit 7 Knoten.

Wir wollen nun ein Resultat zur Anzahl der Cliques beweisen, sowie einen Ausblick auf weitere Eigenschaften des Modells geben.

Theorem 14.3 (Cliques). Sei $k \geq 2$, $C_k(G_{n_0}) > 0$ and $\mathcal{F}_\infty := \sigma(G_n; n \geq n_0)$. Dann gibt es eine \mathcal{F}_∞ -messbare, integrierbare Zufallsvariable $C_k(\infty)$ mit $\mathbf{P}(C_k(\infty) > 0) > 0$, und

$$n^{-kp^{k-1}} C_k(n) \xrightarrow[n \rightarrow \infty]{f.s., L^2} C_k(\infty). \quad (14.1)$$

Außerdem gilt

$$\mathbf{E}[C_k(n)] = C_k(n_0) \cdot \prod_{m=n_0}^{n-1} \frac{m + kp^{k-1}}{m} \quad (14.2)$$

Bevor wir dieses Resultat beweisen können, benötigen wir eine wichtige Aussage, wie man mit bestimmten Martingalen umgehen kann.

Lemma 14.4 (Martingal-Abschätzungen). Sei $I = \{t_0, t_0 + 1, \dots\}$, $\mathcal{X} = (X_t)_{t \in I}$ ein nicht-negativer, integrierbarer, an eine Filtration $(\mathcal{F}_t)_{t \in I}$ adaptierter stochastischer Prozess, $\mathcal{F}_\infty := \sigma\left(\bigcup_{t=t_0}^{\infty} \mathcal{F}_t\right)$ sowie $x_0 := \mathbf{E}[X_{t_0}] > 0$. Sei außerdem $a > -n_0$ und

$$\mathbf{E}[X_{t+1} | \mathcal{F}_t] = \left(1 + \frac{a}{n}\right) X_t, \quad t = t_0, t_0 + 1, \dots$$

Dann gilt:

1. Der Prozess $\mathcal{M} = (M_t)_{t \geq n_0}$, definiert durch $M_{t_0} = X_{t_0}$ und

$$M_t = X_t \cdot \prod_{s=t_0}^{t-1} \frac{s}{s+a}$$

ist ein Martingal und

$$\mathbf{E}[X_t] = x_0 \cdot \prod_{s=t_0}^{t-1} \frac{s+a}{s}. \quad (14.3)$$

2. Es gibt eine nicht-negative Zufallsvariable $X_\infty \in \mathcal{L}^1(\mathcal{F}_\infty)$ mit $\mathbf{E}[X_\infty] \leq x_0 \Gamma(t_0)/\Gamma(t_0 + a)$, so dass

$$t^{-a} X_t \xrightarrow[t \rightarrow \infty]{f.s.} X_\infty.$$

Gilt zusätzlich $\limsup_{t \rightarrow \infty} \mathbf{E}[X_t^r]/t^{ar} < \infty$ für ein $r > 1$, dann gilt die Konvergenz auch in \mathcal{L}^r und es gilt $\mathbf{P}(X_\infty > 0) > 0$.

Bemerkung 14.5. Im Beweis benötigen wir eine Standard-Abschätzung der Gamma-Funktion,

$$\prod_{s=t_0}^{t-1} \frac{s+a}{s} = \frac{\Gamma(t+a)}{\Gamma(t)} \cdot \frac{\Gamma(t_0)}{\Gamma(t_0+a)} \xrightarrow[t \rightarrow \infty]{\sim} \frac{t^a \Gamma(t_0)}{\Gamma(t_0+a)},$$

wobei $a_t \xrightarrow[t \rightarrow \infty]{\sim} b_t$ genau dann, wenn $a_t/b_t \xrightarrow[t \rightarrow \infty]{} 1$.

Beweis. Die Martingal-Eigenschaft von \mathcal{M} ist einfach nachzurechnen. Deshalb gilt (14.3) mit Induktion. Mit dem Martingal-Konvergenzsatz, Korollar 13.31, folgt die fast sichere Konvergenz für das nicht-negative Martingal \mathcal{M} mit $\mathbf{E}[M_\infty] \leq \mathbf{E}[M_{t_0}]$. Deshalb gilt

$$t^{-a} X_t = t^{-a} M_t \cdot \prod_{s=t_0}^{t-1} \frac{s+a}{s} \xrightarrow[t \rightarrow \infty]{} M_\infty \cdot \frac{\Gamma(t_0)}{\Gamma(t_0+a)} =: X_\infty$$

fast sicher. Nach Theorem 13.34 gilt die Konvergenz in \mathcal{L}^r , falls \mathcal{M} auch \mathcal{L}^r -beschränkt ist. Wir berechnen, wobei $c > 0$ so gewählt ist, dass $\mathbf{E}[X_t^r] \leq ct^{ar}$, sowie ein weiteres $c' > 0$,

$$\sup_t \mathbf{E}[M_t^r] = \sup_t \mathbf{E}[X_t^r] \left(\prod_{s=t_0}^{t-1} \frac{s}{s+a} \right)^r \leq \sup_t \frac{\Gamma(t_0+a)^r c t^{ar}}{\Gamma(t_0)^r c' t^{ar}} < \infty,$$

und die Behauptung ist gezeigt. Insbesondere ist die Konvergenz in \mathcal{L}^1 , so dass $\mathbf{E}[M_\infty] > 0$ und $\mathbf{P}(X_\infty > 0) > 0$ impliziert. \square

Beweis von Theorem 14.3. Wir zeigen zunächst

$$\mathbf{E}[C_k(n+1)|\mathcal{F}_n] = C_k(n) \left(1 + \frac{k}{n} p^{k-1} \right). \quad (14.4)$$

Eine neue k -Clique tritt nämlich genau dann auf, wenn ein Knoten v kopiert wird, der Teil einer k -Clique ist, sowie alle $k-1$ Kanten zu seinen Nachbarn innerhalb der k -Clique. Darum gilt

$$\mathbf{E}[C_k(n+1) - C_k(n)|\mathcal{F}_n] = \sum_{i=1}^{C_k(n)} \mathbf{P}(\text{Clique } i \text{ wird kopiert} | \mathcal{F}_n) = C_k(n) \frac{k}{n} p^{k-1}.$$

Wir verwenden nun (14.4) zusammen mit Lemma 14.4. Die fast sichere Konvergenz folgt aus Lemma 14.4.2 mit $a = kp^{k-1}$, und für (14.2) verwenden wir (14.3). Wir skizzieren noch die \mathcal{L}^2 -Konvergenz in (14.1). Hierfür betrachten wir die Anzahl der Paare von k -Cliquen zur Zeit, $\binom{C_k(n)}{2}$. Sei $C_{k,\ell}(n)$ die Anzahl dieser Paare, die genau ℓ Knoten gemeinsam haben. (Also ist die Anzahl der disjunkten Paare die 0-Paare, und ein $(k-1)$ -Paare von k -Cliquen ist eine $(k+1)$ -Clique, bei der eventuell eine Kante fehlt. Es gilt also

$$\binom{C_k(n)}{2} = \sum_{\ell=0}^{k-1} C_{k,\ell}(n). \quad (14.5)$$

Für jedes ℓ -Paar von k -Cliques gibt es vier Möglichkeiten, wie ein neues ℓ -Paar im nächsten Schritt entstehen kann:

1. Jede neue Clique bildet ein $(k-1)$ -Paar mit der Clique, von der sie kopiert wurde. Dies passiert im Mittel $\frac{kp^{k-1}}{n}C_k(n)$ mal im nächsten Schritt. In den nächsten drei Fällen werden wir dies ignorieren.
2. Einer der $2(k-\ell)$ nicht-gemeinsamen Knoten wird zur Duplikation ausgewählt und die Clique, zu der er gehört wird kopiert. Da die neue Clique die ℓ Knoten der nicht-duplizierten Clique behält, wird ein neues ℓ -Paar gebildet.

$$\text{Wahrscheinlichkeit: } \frac{2(k-\ell)}{n}p^{k-1}$$

3. Einer der ℓ gemeinsamen Knoten wird zur Duplikation ausgewählt und beide Cliques werden dupliziert. Dies ergibt offenbar ein neues ℓ -Paar. Zusätzlich gibt es zwei neue Paare, die durch die ursprüngliche und die Kopie der neuen Clique gebildet werden. Dies sind dann $(\ell-1)$ -Paare, da diese Cliques den neuen Knoten nicht teilen.

$$\text{Wahrscheinlichkeit: } \frac{\ell}{n}p^{2k-\ell-1}$$

4. Einer der ℓ gemeinsamen Knoten wird zur Duplikation ausgewählt und nur eine der beiden Cliques wird dupliziert. Ähnlich wie in 3. wird ein neues $(\ell-1)$ -Paar gebildet.

$$\text{Wahrscheinlichkeit: } \frac{\ell}{n} \cdot 2p^{k-1}(1-p^{k-\ell}) = \frac{\ell}{n}2p^{k-1} - \frac{\ell}{n}2p^{2k-\ell-1}$$

Daraus folgt nun für $\ell \leq k-2$, dass

$$\begin{aligned} & \mathbf{E}[C_{k,\ell}(n+1) - C_{k,\ell}(n) \mid \mathcal{F}_n] \\ &= \frac{2(k-\ell)p^{k-1} + \ell p^{2k-\ell-1}}{n}C_{k,\ell}(n) + \frac{2(\ell+1)p^{k-1}}{n}C_{k,\ell+1}(n) \end{aligned} \quad (14.6)$$

und für $\ell = k-1$

$$\begin{aligned} & \mathbf{E}[C_{k,k-1}(n+1) - C_{k,k-1}(n) \mid \mathcal{F}_n] \\ &= \frac{2p^{k-1} + (k-1)p^k}{n}C_{k,k-1}(n) + \frac{kp^{k-1}}{n}C_k(n). \end{aligned} \quad (14.7)$$

Nun lassen wir einige Abschätzungen aus. Aus der obigen Rekursion ergibt sich iterativ, dass $\limsup_{n \rightarrow \infty} \mathbf{E}[C_k(n)^2]/n^{2kp^{k-1}} < \infty$. Aus Lemma 14.4.2 folgt nun die \mathcal{L}^2 -Konvergenz. \square

Ohne Beweis, aber mit ein paar Vorüberlegungen, geben wir nun noch ein Resultat über das allgemeine Verhalten des Duplikations-Modells.

Theorem 14.6 (Häufigkeit isolierter Knoten). *Sei p^* die eindeutige Lösung von $pe^p = 1$ (oder $p + \log p = 0$). Dann gilt:*

1. Für $p \leq p^*$ ist $\sup_{q \in [0,1]} |H_q^\circ(n) - 1| \xrightarrow{n \rightarrow \infty}_{f.s.} 0$. Für $q = 0$ folgt insbesondere $F_0^\circ(n) \xrightarrow{n \rightarrow \infty} 1$, d.h. nach langer Zeit besteht der gesamte Graph nur noch aus isolierten Knoten.

2. Für $p^* < p < 1$ gilt

$$\mathbf{E}[H_q^\circ(n)] \xrightarrow{n \rightarrow \infty} 1 - \left(1 - \frac{1}{p} \log\left(\frac{1}{p}\right)\right) \cdot \sum_{k=1}^{\infty} \frac{S_k^\circ(n_0)}{k!} (-1)^{k-1} \prod_{\ell=1}^{k-1} \left(1 - \frac{p^\ell}{p^\ell}\right).$$

Bemerkung 14.7. Obwohl wir das Theorem nicht beweisen, wollen wir doch eine Formel herleiten, die der Startpunkt der Analyse der Gradverteilung darstellt. Es gilt nämlich

$$\begin{aligned} \mathbf{E}[F_k(n+1)|\mathcal{F}_n] &= F_k(n) + p(k-1)F_{k-1}^\circ(n) - pkF_k^\circ(n) \\ &\quad + \sum_{\ell \geq k} F_\ell^\circ(n) \binom{\ell}{k} p^k (1-p)^{\ell-k}. \end{aligned}$$

Denn: Die Anzahl der Knoten mit Grad k , also F_k , wächst genau in zwei Fällen. Entweder wird ein Knoten mit Grad $\ell \geq k$ dupliziert, zusammen mit genau k Kanten (Wahrscheinlichkeit $\binom{\ell}{k} p^k (1-p)^{\ell-k}$), oder ein Nachbar eines Knotens mit Grad $k-1$ wird dupliziert, zusammen mit der verbindenden Kante. Weiter verringert sich F_k um 1, falls ein Nachbar eines Knotens mit Grad k zusammen mit der verbindenden Kante dupliziert wird. Genau diese drei Fälle finden sich in obiger Gleichung wieder. Weiter gilt nun, durch Multiplikation mit q^k und Summation,

$$\begin{aligned} \mathbf{E}[H_q(n+1) - H_q(n)|\mathcal{F}_n] &= pq^2 \sum_{k=1}^{\infty} (k-1)F_{k-1}^\circ(n)q^{k-2} - pq \sum_{k=0}^{\infty} kF_k^\circ(n)q^{k-1} \\ &\quad + \sum_{k=0}^{\infty} \sum_{\ell=k}^{\infty} F_\ell^\circ(n) \binom{\ell}{k} p^k (1-p)^{\ell-k} q^k \\ &= -pq(1-q) \frac{d}{ds} \left(\sum_{k=0}^{\infty} F_k^\circ(n) s^k \right) \Big|_{s=q} + \sum_{\ell=0}^{\infty} F_\ell^\circ(n) (1-p+pq)^\ell \\ &= -pq(1-q) \frac{d}{ds} H_s^\circ(n) \Big|_{s=q} + H_{1-p+pq}^\circ(n). \end{aligned}$$

Aus dieser Gleichung sieht man, dass man die Entwicklung der Funktion $q \mapsto H_q^\circ$ studieren kann. Dies führt dann letztlich auf einen Beweis von Theorem 14.6.

14.2 Populationsgenetik

Als Teil der Evolutionstheorie beschäftigt sich die Populationsgenetik mit der Ausbreitung von (genetischen) Typen in Populationen. Man modelliert meistens (zumindest werden wir hier nichts anderes tun) eine Population konstanter Größe N . Jedes Individuum kann Nachkommen bekommen und gibt dabei das eigene genetische Material weiter. In seltenen Fällen kann es zu Mutationen kommen, d.h. Fehlern beim Kopieren des genetischen Materials bei der Vererbung. Wir führen zunächst das Wright-Fisher-Modell als Standard-Modell der Populationsgenetik ein.

Definition 14.8 (Das Wright-Fisher-Modell). Sei $N \in \mathbb{N}$. Das Wright-Fisher-Modell der Größe N ist ein zeit-diskretes Populationsmodell, bei dem in jeder Generation die gesamte Population ausgetauscht wird. Jedes Individuum der Generation n wählt dabei einen Vorfahren unabhängig und gleichverteilt aus der Menge aller Individuen der Generation $n-1$.

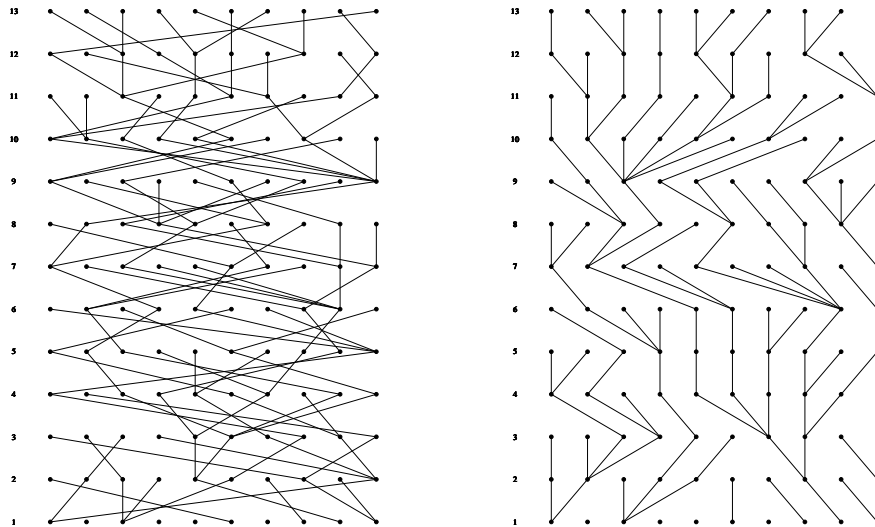


Abbildung 14.2

Darstellung eines Wright-Fisher-Modells mit $N = 10$ Individuen. Das rechte Bild ist nur eine umgeordnete Version des linken.

In Abbildung 14.2 ist ein Beispiel einer Realisierung eines Wright-Fisher-Modells mit $N = 10$ und $t = 13$ Generationen dargestellt.

Lemma 14.9 (Allelhäufigkeiten im Wright-Fisher-Modell). *Angenommen, ein Allel (d.h. ein genetischer Typ) tritt zur Zeit 0 in k der N Individuen auf. Dann ist der stochastische Prozess $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$, der die Entwicklung der Allelhäufigkeit im Wright-Fisher-Modell beschreibt, eine zeitlich homogene Markov-Kette mit Übergangswahrscheinlichkeit*

$$\mathbf{P}(X_{t+1} = k | X_t) = \binom{N}{k} \left(\frac{X_t}{N}\right)^k \left(1 - \frac{X_t}{N}\right)^{N-k}.$$

Weiter ist $(X_t)_{t=0,1,2,\dots}$ ein Martingal mit

$$\langle \mathcal{X} \rangle_t = \sum_{s=1}^{t-1} X_s \left(1 - \frac{X_s}{N}\right).$$

Beweis. Gegeben $X_t = x$ hat jedes Individuum in Generation $t + 1$ dieselbe Chance X_t/N , einen der X_t Allel-Träger aus Generation t als Vorfahren zu wählen. Daraus folgt $X_{t+1} \sim B(N, X_t/N)$, also die Formel für die Übergangswahrscheinlichkeit. Weiter ist damit $\mathbf{E}[X_{t+1} | X_t] = X_t$ und die Martingaleigenschaft folgt. Nach Definition der quadratischen Variation ist

$$\langle \mathcal{X} \rangle_t = \sum_{s=1}^t \mathbf{E}[(X_s - X_{s-1})^2 | X_{s-1}] = \sum_{s=1}^t X_{s-1} \left(1 - \frac{X_{s-1}}{N}\right).$$

□

Bemerkung 14.10 (Genetische Drift). Die Form der quadratischen Variation von \mathcal{X} lässt darauf schließen, dass sich \mathcal{X} nach $x \approx N/2$ am meisten Fluktuationen aufweist. Betrachtet man außerdem $\mathcal{P} = (P_t)_{t=0,1,2,\dots}$ mit $P_t = X_t/N$, also die relative Häufigkeit des Allels, so stellt man fest, dass $\langle \mathcal{P} \rangle_t = \frac{1}{N} \sum_{s=1}^{t-1} P_s(1-P_s)$, also fluktuiert P_t umso langsamer je größer die Population ist. Die Größe der Fluktuationen in \mathcal{P} wird auch als genetische Drift bezeichnet.

Bemerkung 14.11 (Ahnenlinien). Üblicherweise hat man es mit großen Populationen zu tun, von denen eine kleine Stichprobe (der Größe n) gezogen und untersucht wird. Aufgrund der Struktur des Wright-Fisher-Modells lassen sich auch einzelne Ahnenlinien verfolgen. Folgende Vorüberlegung ist dabei hilfreich: In einer Stichprobe der Größe n wollen wir die Wahrscheinlichkeit berechnen, dass diese Individuen vor einer Generation n verschiedene Vorfahren hatten. Dies ist¹⁹

$$\begin{aligned} & \mathbb{P}[n \text{ unterschiedliche Vorfahren vor einer Generation}] \\ &= \left(1 - \frac{1}{N}\right) \cdot \dots \cdot \left(1 - \frac{n-1}{N}\right) = 1 - \frac{\binom{n}{2}}{N} + \mathcal{O}\left(\frac{1}{N^2}\right). \end{aligned}$$

Weiter ist

$$\mathbb{P}[\text{weniger als } n-1 \text{ Vorfahren vor einer Generation}] \leq \frac{\binom{N}{n-2}(n-2)^n}{N^n} = \mathcal{O}\left(\frac{1}{N^2}\right)$$

sowie

$$\mathbb{P}[n-1 \text{ unterschiedliche Vorfahren vor einer Generation}] = \frac{\binom{n}{2}}{N} + \mathcal{O}\left(\frac{1}{N^2}\right).$$

Da Vorfahren in verschiedenen Generationen unabhängig gewählt werden, wiederholen sich diese Rechnungen für die Zeit, die weiter in der Vergangenheit liegt. Abbildung 14.2 liefert ein Beispiel. Wir sehen also, dass die Wartezeit T_n^N , bis zum ersten Mal zwei der n Individuen einen gemeinsamen Vorfahren haben, in etwa $\text{geo}\left(\frac{\binom{n}{2}}{N}\right)$ -verteilt ist. Da nach einer Übungsaufgabe $T_n^N/N \xrightarrow{N \rightarrow \infty} T_n \sim \exp\left(\frac{\binom{n}{2}}{N}\right)$, haben wir die folgende Definition motiviert.

Definition 14.12 (n -Coaleszent). Der n -Coaleszent \mathcal{T}_n ist ein zufälliger (ultrametrischer²⁰) Baum, der folgendermaßen zustande kommt: Ausgehend von n Linien verschmilzt jedes Paar von Linien mit Rate 1. Sind noch $2 \leq k \leq n$ Linien übrig, ist also die Gesamtrate, mit der eine Verschmelzung stattfindet, $\binom{k}{2}$. Jedes der $\binom{k}{2}$ Paare hat dabei die gleiche Wahrscheinlichkeit, zu verschmelzen. Der Prozess endet, wenn nur noch eine Linie übrig ist.

Die Zeit, bei der der Coaleszent zum ersten mal k Linien hat, bezeichnen wir mit S_k und $T_k = S_k - S_{k+1} \sim \exp\left(\frac{k}{2}\right)$ ist die Zeitdauer, die der Coaleszent mit genau k Linien verbringt.

Den Prozess, der die Linien im Coaleszenten zählt, nennen wir $(K_t)_{t \geq 0}$. Er ist ein reiner Todesprozess mit $\mu_k = \frac{k}{2}$.

Bemerkung 14.13. 1. Ein paar Realisierungen des Coaleszenten für $n = 4$ und $n = 20$ sind in Abbildung 14.3 abgedruckt.

¹⁹Wir verwenden Landau-Symbole, d.h. wir schreiben $a_n = O(b_n)$ falls $\limsup_{n \rightarrow \infty} a_n/b_n < \infty$.

²⁰Bekanntlich ist eine Ultrametrik eine Metrik r , bei der die Dreiecksungleichung zu $r(x, z) \leq r(x, y) \vee r(y, z)$ verschärft werden kann. Ein ultrametrischer Baum ist durch die Metrik auf den Blättern gegeben, d.h. der Abstand zwischen zwei Blättern im Baum ist die Weglänge im Baum von einem zum anderen Blatt.

2. Unsere Konstruktion des Coaleszenten geht von den Blättern des Baumes aus und endet bei der Wurzel. Stochastisch äquivalent ist die Beschreibung von der Wurzel zu den Blättern: startend mit der Wurzel und zwei daranhängenden Linien vergeht im Zustand von k Linien eine $\exp\left(\frac{k}{2}\right)$ -verteilte Zeit. Ist diese zu Ende, wird eine zufällig ausgewählte Linie verzweigt. Dies geschieht bis zu der Zeit, bis kurz bevor die $n + 1$ -te Linie erzeugt wird.
3. An dieser Vorwärts-Konstruktion sieht man, dass es auch den Coaleszent mit unendlich vielen Blättern gibt. Es ist nämlich

$$\mathbf{E}\left[\sum_{k=2}^{\infty} T_k\right] = \sum_{k=2}^{\infty} \frac{1}{\binom{k}{2}} < \infty,$$

also insbesondere $\mathbf{P}\left(\sum_{k=2}^{\infty} T_k < \infty\right) = 1$.

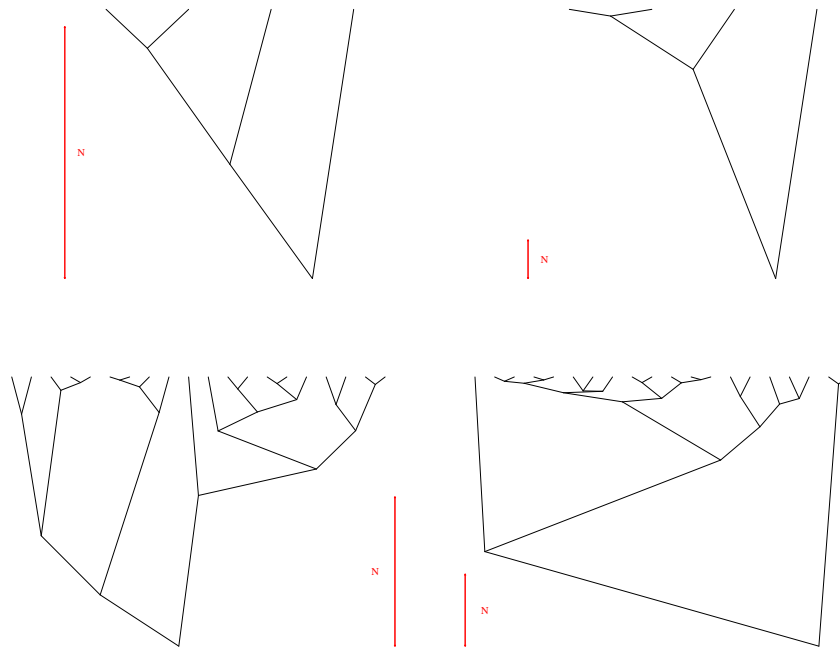


Abbildung 14.3: Coaleszenten für $n = 4$ und $n = 20$ Blätter.

Proposition 14.14 (Eigenschaften des n -Coaleszenten). Für einen n -Coaleszenten sei $S_1 = \sum_{k=2}^n T_k$ die Zeit des jüngsten gemeinsamen Stichproben-Vorfahrens, und

$$L_n := \sum_{k=2}^n kT_k$$

die gesamte Baumlänge.

1. Es gilt

$$\begin{aligned}\mathbf{E}[S_1] &= 2\left(1 - \frac{1}{n}\right) \xrightarrow{n \rightarrow \infty} 2, \\ \mathbf{V}[S_1] &= 8\left(\sum_{k=2}^n \frac{1}{k^2}\right) - 4\left(1 - \frac{1}{n}\right)^2 \xrightarrow{n \rightarrow \infty} \frac{4}{3}\pi^2.\end{aligned}$$

2. Seien $X_1, \dots, X_{n-1} \sim \exp(1)$ unabhängig. Dann ist $L_n \sim 2 \max_{i=1, \dots, n-1} X_i$. Weiter gilt

$$\begin{aligned}\mathbf{E}[L_n] &= 2 \sum_{k=1}^{n-1} \frac{1}{k}, & \mathbf{V}[L_n] &= 4 \sum_{k=1}^{n-1} \frac{1}{k^2}, \\ \frac{1}{2}L_n - \log n &\xrightarrow{N \rightarrow \infty} G,\end{aligned}$$

wobei G Gumbel-verteilt ist, d.h. die Verteilungsfunktion von G ist $F(x) = e^{-e^{-x}}$.

Beweis. Wir erinnern daran, dass $S_1 = T_n + \dots + T_2$ und $T_k \sim \exp\left(\frac{k}{2}\right)$. 1. Daraus folgt

$$\begin{aligned}\mathbf{E}[S_1] &= \sum_{k=2}^n \frac{2}{k(k-1)} = 2 \sum_{k=2}^n \frac{1}{k-1} - \frac{1}{k} = 2\left(1 - \frac{1}{n}\right), \\ \mathbf{V}[S_1] &= \sum_{k=2}^n \frac{4}{k^2(k-1)^2} = 4 \sum_{k=2}^n \left(\frac{1}{(k-1)} - \frac{1}{k}\right)^2 \\ &= 4 \left[2 \left(\sum_{k=2}^n \frac{1}{k^2}\right) + 1 - \frac{1}{n^2} - \sum_{k=2}^{n-1} \frac{2}{k(k-1)} \right] \\ &= 8 \left(\sum_{k=2}^n \frac{1}{k^2}\right) - 4\left(1 - \frac{1}{n}\right)^2.\end{aligned}$$

Es gilt $kT_k \sim \exp((k-1)/2)$. Daraus folgt bereits

$$\begin{aligned}\mathbf{E}[L_n] &= \sum_{k=2}^n \frac{2}{k-1} = 2 \sum_{k=1}^{n-1} \frac{1}{k}, \\ \mathbf{V}[L_n] &= 4 \sum_{k=1}^{n-1} \frac{1}{k^2}.\end{aligned}$$

Für die Darstellung $L_n = 2 \max_{i=1, \dots, n-1} X_i$ bemerken wir folgendes. Die Zufallsgrößen X_1, \dots, X_{n-1} sind die Zeiten, zu denen $n-1$ unabhängige, exponentialverteilte Wecker klingeln. Die Gesamtrate, mit der ein Wecker klingelt, ist zunächst $n-1$. Nach Ablauf des ersten Weckers sind noch $n-2$ übrig, so dass die Gesamtrate $n-2$ ist. Der letzte Wecker klingelt also zu einer Zeit, die nach $Y_{n-1} + \dots + Y_1$ verteilt ist, wobei $Y_k \sim \exp(k)$. Also ist

$$L_n = \sum_{k=2}^n kT_k \sim 2 \sum_{k=1}^{n-1} Y_k = 2 \max_{i=1, \dots, n-1} X_i.$$

Nun ist

$$\begin{aligned} \mathbf{P}\left(\frac{1}{2}L_n - \log n \leq x\right) &= \mathbf{P}\left(\max_{i=1, \dots, n-1} X_i - \log n \leq x\right) = (1 - e^{-(x+\log n)})^{n-1} \\ &= \left(1 - \frac{e^{-x}}{n}\right)^{n-1} \xrightarrow{n \rightarrow \infty} e^{-e^{-x}}. \end{aligned}$$

Daraus folgt die Konvergenz gegen die Gumbel-Verteilung. \square

Aus der Biologie ist bekannt, dass sich bei der Vererbung von Genmaterial von einer auf die nächste Generation Fehler einschleichen, wobei man auch von Mutationen spricht. Im Wright-Fisher-Modell wird Mutation dadurch realisiert, dass ein Nachkomme mit Wahrscheinlichkeit $1 - \mu$ dasselbe genetische Material wie sein Vorfahre (in der vorigen Generation) hat, und mit Wahrscheinlichkeit μ etwas anderes. Nimmt man nun an, dass $\mu = \vartheta/(2N)$, und betrachtet die Anzahl M^N der Mutationen in N Generationen, so ist diese $B(N, \vartheta/(2N))$ -verteilt. Mit Beispiel 11.1 folgt, dass $M^N \xrightarrow{M} \sim \text{Poi}(\theta/2)$. Deshalb definieren wir folgendes Mutationsmodell im Coaleszenten.

Definition 14.15 (Der Coaleszent im unendlich-viele-Sites-Modell). Sei \mathcal{T}_n ein Coaleszent wie in Definition 14.12 und $\theta > 0$. Entlang eines Astes der Länge t wird der Baum mit einer $\text{Poi}(t\theta/2)$ -verteilten Anzahl von Mutationen markiert. Diese Markierungen heißen auch SNPs (Single Nucleotide Polymorphisms). Ein SNP hat Größe i , falls er so auf \mathcal{T}_n liegt, dass er zu genau i Blättern führt. Es bezeichne $(S_i)_{i=1, \dots, n-1}$ das Frequenzspektrum und $S = S_1 + \dots + S_{n-1}$ die Gesamtzahl an SNPs.

Bemerkung 14.16. Für die Gesamtzahl S der SNPs auf \mathcal{T}_n lassen sich die ersten beiden Momente recht einfach berechnen. Es gilt, da S gegeben \mathcal{T}_n gerade eine $\text{Poi}(\theta L(\mathcal{T}_n)/2)$ -verteilt ist, nach einer Übungsaufgabe und Proposition 14.14

$$\begin{aligned} \mathbf{E}[S] &= \mathbf{E}[\mathbf{E}[S|L(\mathcal{T}_n)]] = \mathbf{E}[\theta L(\mathcal{T}_n)/2] = \theta \sum_{k=1}^{n-1} \frac{1}{k}, \\ \mathbf{V}[S] &= \mathbf{E}[\mathbf{V}[S|L(\mathcal{T}_n)]] + \mathbf{V}[\mathbf{E}[S|L(\mathcal{T}_n)]] \\ &= \mathbf{E}[\theta L(\mathcal{T}_n)/2] + \mathbf{V}[\theta L(\mathcal{T}_n)/2] = \vartheta \sum_{k=1}^{n-1} \frac{1}{k} + \theta^2 \sum_{k=1}^{n-1} \frac{1}{k^2}. \end{aligned}$$

Wir berechnen nun im unendlich-viele-Sites-Modell das erwartete Frequenzspektrum.

Proposition 14.17. Im Coaleszent im unendlich-viele-Sites-Modell ist das erwartete Frequenzspektrum gegeben durch

$$\mathbf{E}[S_i] = \frac{\theta}{i}.$$

Bemerkung 14.18 (Polya-Urne). Im Beweis wird die aus der Stochastik bekannte Polya-Urne auftauchen, die wir kurz wiederholen wollen. Wir stellen uns eine Urne mit k Kugeln vor, die alle eine unterschiedliche Farbe besitzen. Wir ziehen eine Kugel, notieren die Farbe, und legen die Kugel zusammen mit einer Kugel derselben Farbe zurück. Dies wiederholen wir, bis n Kugeln in der Urne sind. Wenn wir uns nun fragen, wie groß die Wahrscheinlichkeit ist, dass i blaue Kugeln in der Urne am Ende (wenn n Kugeln in der Urne sind) sind, wenn am Anfang (wenn k Kugeln in der Urne sind) eine blaue Kugel enthalten ist, so lässt sich

dies folgendermaßen berechnen: Insgesamt müssen in den $n - k$ Ziehungen von Kugeln aus der Urne genau $i - 1$ -mal eine blaue gezogen worden sein, damit am Ende i blaue Kugeln in der Urne sind. Betrachten wir etwa die Möglichkeit, dass alle $i - 1$ blauen Kugeln in die Urne gelegt werden, bevor irgendeine andersfarbige Kugel gezogen wird, so erhalten wir

$$\frac{1 \cdots (i - 1) \cdot (k - 1) \cdots (n - i - 1)}{k \cdots (n - 1)}.$$

Insgesamt gibt es aber $\binom{n-k}{i-1}$ Möglichkeiten, in welchen Schritten die blauen Kugeln in die Urne gelegt werden können. Deshalb ist die oben gesuchte Wahrscheinlichkeit gleich

$$\binom{n-k}{i-1} \frac{(i-1)!(k-1) \cdots (n-i-1)}{k \cdots (n-1)} = \frac{k-1}{i} \binom{n-k}{i-1} \frac{i!}{(n-i) \cdots (n-1)} = \frac{\binom{n-k}{i-1} k-1}{\binom{n-1}{i} i}.$$

Beweis von Proposition 14.17. Im unendlich-viele-Sites-Modell sagen wir, ein SNP $x \in I$ habe Größe i , wenn er in genau i der n Blätter von \mathcal{T}_n vorkommt. Außerdem sagen wir, der Coaleszent ist in Zustand k , wenn er gerade k Linien hat. Dabei muss man bedenken, dass jeder SNP in einer Zeit im Coaleszenten vorkommt, in der dieser einen Zustand k hat und der SNP auf einer Linie sitzen muss, die insgesamt zu i Blättern führt. Nun ist es so, dass man den Coaleszenten auch von $k \leq n$ nach n Linien generieren kann, indem man bei k Linien eine $\exp\left(\frac{k}{2}\right)$ -verteilte Zeit wartet und dann eine zufällig ausgewählte Linie nimmt, und diese verzweigt. Das Verfahren endet dann kurz bevor die Verzweigung in $n + 1$ Linien ientreten würde. Mit anderen Worten ist die Wahrscheinlichkeit, dass eine von k Linien genau zu i Blättern führt, dieselbe wie die oben berechnete Wahrscheinlichkeit, dass eine Farbe in der Polya-Urne zu der Zeit, wenn n Kugeln in der Urne sind, genau i -mal vorkommt, wenn zu der Zeit, als in der Urne k Linien waren, genau einmal vorkam, also

$$\mathbf{P}[l\text{-te Linie im Zustand } k \text{ hat Größe } i] = \frac{\binom{n-k}{i-1} k-1}{\binom{n-1}{i} i}.$$

Insgesamt folgt

$$\begin{aligned} \mathbb{E}[S_i] &= \sum_{k=2}^n \sum_{l=1}^k \mathbb{P}[l\text{-te Linie im Zustand } k \text{ hat Größe } i] \cdot \\ &\quad \mathbb{E}[\text{Anzahl der Mutation auf der } i\text{-ten Linie im Zustand } k] \\ &= \sum_{k=2}^n \sum_{l=1}^k \frac{\binom{n-k}{i-1} k-1}{\binom{n-1}{i} i} \frac{\theta}{k(k-1)} \\ &= \frac{\theta}{i} \frac{1}{\binom{n-1}{i}} \sum_{k=2}^n \binom{n-k}{i-1} \\ &= \frac{\theta}{i} \frac{1}{\binom{n-1}{i}} \sum_{k=2}^n \left(\binom{n-(k-1)}{i} - \binom{n-k}{i} \right) \\ &= \frac{\theta}{i} \frac{1}{\binom{n-1}{i}} \left(\sum_{k=1}^{n-1} \binom{n-k}{i} - \sum_{k=2}^n \binom{n-k}{i} \right) = \frac{\theta}{i}. \end{aligned}$$

□

14.3 Finanzmathematik

Als Anwendung der Wahrscheinlichkeitstheorie etabliert sich seit ein paar Jahrzehnten die Finanzmathematik. Aus diesem Bereich werden wir in diesem abschließenden Abschnitt das Binomialmodell der Optionspreisbewertung kennen lernen.

Bemerkung 14.19 (Finanzderivate). Auf Finanzmärkten gibt es verschiedene Arten von Wertpapieren. Die einfachsten hiervon sind etwa festverzinsliche Wertpapiere oder Aktien. Derivate sind Wertpapiere, die direkt an den Preis eines anderen, zu Grunde liegenden Wertpapiere, gekoppelt sind. Betrachten wir als Beispiel eine *europäische Call-Option*. Das zugrunde liegende Wertpapier ist eine bestimmte Aktie. Der Besitzer der europäischen Option hat das Recht, aber nicht die Pflicht, die Aktie zu einer Zeit τ zu einem Preis K zu erwerben. Die Frage, die hier beantwortet werden soll ist:

Was ist der faire Preis einer europäischen Call-Option mit Parametern τ und K ?

Klar ist, dass der faire Preis vom heutigen Aktienkurs sowie von der vermutlichen Wertentwicklung der Aktie abhängt. Die Wahrscheinlichkeitstheorie stellt für den Preis der Aktie einen geeigneten Rahmen zur Verfügung, da man den Aktienpreis der Zukunft als Zufallsvariable betrachten kann.

Die Anzahl verschiedener Finanzderivate wächst ständig. Bei Put-Optionen besteht das Recht des Besitzers nicht darin, die Aktie zu kaufen, sondern sie zu verkaufen. Bei einem Future besteht nicht die Recht auf Erwerb/Verkauf der Aktie, sondern die Pflicht. Amerikanische Optionen erlauben dem Besitzer, schon vor der Zeit τ das Recht auf Erwerb/Verkauf der Aktie auszuüben. Bei all diesen Fragen stellt sich die Frage nach dem fairen Preis.

Bemerkung 14.20 (Finanzmarktmodell). Neben der Aktie und dem Finanzderivat wird hier angenommen, dass es auch ein festverzinsliches Wertpapier gibt. Für 1 Geldeinheit (GE) in Zeit $t - 1$ bekommt man $1 + r$ GEen zur Zeit $1 + r$, d.h. der Zinssatz ist r . Es wird stets angenommen, dass jederzeit zu diesem Zinssatz Geld geliehen oder angelegt werden kann.

Geht man von einem Händler in einem solchen Finanzmarkt aus, so hat er jederzeit die Wahl zwischen folgenden Möglichkeiten:

- Geld zum Zinssatz von r leihen,
- Geld zum Zinssatz r anlegen,
- in die Aktie investieren,
- Aktien verkaufen.

Zur letzten Möglichkeit sei bemerkt, dass auch Leerverkäufe getätigt werden können. Das heißt, die Aktie wird zum heutigen Preis verkauft, obwohl man sie nicht besitzt, und wird erst zu einem späteren Zeitpunkt geliefert.

Folgende Definition stellt die Grundlage für ein einfaches Aktienpreismodell dar. Natürlich macht das Modell viele Vereinfachungen, aber das haben mathematische Modelle so an sich.

Definition 14.21 (Binomialmodell des Aktienpreisprozesses). Seien M_1, M_2, \dots unabhängig und identisch verteilt mit $\mathbf{P}(M_1 = u) = p, \mathbf{P}(M_1 = d) = q$ mit $0 < d < u$ und $p + q = 1$. Für $S_0 \in \mathbb{R}_+$ definieren wir den Aktienpreisprozess S_0, S_1, \dots durch

$$S_t = M_t \cdots M_1 \cdot S_0.$$

Weiter sei $0 < r < 1$ der Zinssatz eines festverzinslichen Wertpapiers.

Definition 14.22 (Portfolio und Handelsstrategien). 1. Ein Portfolio besteht aus einer Anzahl von Aktien und einer Anzahl von festverzinslichen Wertpapieren.

2. Eine Handelsstrategie ist gegeben durch den Wert X_0 eines Anfangsportfolios und Funktionen $\Delta_0 = \Delta_0(S_0), \Delta_1 = \Delta_1(S_0, S_1), \dots$, die die Anzahl der Aktien im Portfolio zu den Zeitpunkten $t = 0, 1, 2, \dots$ beschreiben. Wir fordern, dass Δ_t messbar bezüglich S_0, \dots, S_t ist, $t = 1, 2, \dots$

Das bedeutet: Zum Zeitpunkt $t = 0$ bezeichnet $\Delta_0 = \Delta_0(S_0)$ die Anzahl der Aktien des Startportfolios, außerdem ist $X_0 - \Delta_0 S_0$ die Anzahl des festverzinslichen Wertpapiers im Portfolio. Der Wert des Portfolios zur Zeit $t = 1$ ist dann

$$X_1 = \Delta_0 S_1 + (X_0 - \Delta_0 S_0)(1 + r).$$

Auf Grundlage des Aktienpreises S_1 wird festgelegt, wie viel des Wertes X_1 in Aktien investiert werden sollen (das sind dann $\Delta_1(S_0, S_1)$). Die Anzahl an festverzinslichen Wertpapieren ist dann $X_1 - \Delta_1 S_1$.

Bemerkung 14.23 (Annahmen). Im Binomialmodell werden einige Annahmen getroffen, die mehr oder weniger realistisch sind:

1. Sowohl die Aktie als auch das festverzinsliche Wertpapier sind beliebig teilbar. (Da es sich üblicherweise um eine Großzahl an gehandelter Papiere handelt, ist diese Annahme oft gerechtfertigt.)
2. Der Zinssatz des festverzinslichen Wertpapiers hängt nicht davon ab, ob der Händler sich Geld leiht oder Geld in das Wertpapier investiert. (Dies gilt für Privatanleger sicher nicht, für Banken annähernd schon.)
3. Der Kauf- und Verkaufspreis der Aktie sind gleich. (Dies ist in der Praxis nicht erfüllt, und kann zu großen Abweichungen des Modells führen.)
4. Gegeben den Aktienpreis S_t zur Zeit t , kann der Aktienpreis S_{t+1} zur Zeit $t + 1$ nur zwei Werte annehmen. (Dies ist in der Praxis sicher verletzt, macht die Sache aber mathematisch handhabbarer.)

Bemerkung 14.24 (Arbitrage). Grundlegend bei der Theorie der Finanzmärkte ist die Annahme, dass man nie Geld risikolos gewinnen kann. Die Begründung ist, dass es in echten Märkten zwar manchmal solche Situationen gibt, diese aber sofort durch geeigneten Handel wieder verschwinden. Für das Binomialmodell heißt das: Arbitragefreiheit herrscht genau dann, wenn für alle Werte von Δ_0 gilt, dass

$$\mathbf{P}(X_1 > 0 | X_0 = 0) < 1.$$

Lemma 14.25 (Keine Arbitrage im Binomialmodell). *Das Binomialmodell ist genau dann arbitrage-frei, wenn $d < 1 + r < u$.*

Beweis. Wir berechnen direkt

$$\mathbf{P}(X_1 > 0 | X_0 = 0) = \mathbf{P}(\Delta_0 S_1 / S_0 > (1 + r)\Delta_0) = \mathbf{P}(\Delta_0 M_1 > (1 + r)\Delta_0).$$

Da Δ_0 beliebig ist, ist die Arbitragefreiheit äquivalent zu

$$\mathbf{P}(M_1 > 1 + r) < 1,$$

$$\mathbf{P}(M_1 < 1 + r) < 1.$$

Und dies bedeutet wiederum $d < 1 + r < u$. □

Wir kommen nun zum nicht-trivialen Teil dieses Abschnittes. Ziel ist es, den fairen Preis der europäischen Call-Option für das arbitragefreie Binomial-Modell zu berechnen.

Definition 14.26 (Europäisches Finanzderivat). *Ein (europäisches) Finanzderivat wird bestimmt durch die Ausübungszeit τ und eine Funktion φ , so dass der Wert des Finanzderivats zur Zeit τ gerade $\varphi(S_\tau)$ beträgt. Im Fall der europäischen Call-Option ist $\varphi(S_\tau) = (S_\tau - K)^+$, denn: (i) Ist $S_\tau < K$, so ist die Option zum Zeitpunkt τ wertlos, weil man ja die Aktie billiger als zu K GEen erwerben kann. (ii) Ist $S_\tau > K$, so kann man die Option ausüben, die Aktie also zu K GEen erwerben und gleich wieder für S_τ GEen verkaufen. Der Erlös ist gerade $S_\tau - K$.*

Die Strategie, um den fairen Preis der Call-Option zu ermitteln, veranschaulichen wir an einem Beispiel im Fall $\tau = 1$.

Beispiel 14.27 (Hedging). Sei $u = 2, d = \frac{1}{2}, r = \frac{1}{4}$, sowie $S_0 = 4$ und $K = 5$. Das bedeutet, dass die europäische Option zur Zeit 1 entweder 3 GE wert ist, falls $M_1 = u$ (Wahrscheinlichkeit p), oder 0 GEen wert ist, falls $M_1 = l$ (Wahrscheinlichkeit q), also

$$\varphi(X_1) = \begin{cases} 3, & \text{falls } M_1 = u, \\ 0, & \text{falls } M_1 = d. \end{cases}$$

Um den fairen Preis der Option zu ermitteln, versuchen wir, mit Hilfe eines Portfolios aus Aktien und festverzinslichem Wertpapier dieselbe Auszahlungsfunktion herzustellen. Wir beginnen mit einem Portfolio aus 0.5 Aktien und -0.8 Einheiten des festverzinslichen Wertpapiers. Dieses Portfolio hat einen Wert von $X_0 = 0.5 \cdot 4 - 0.8 = 1.2$. Zur Zeit 1 hat dieses Portfolio einen Wert von

$$X_1 = 0.5 \cdot M_1 \cdot 4 - 0.8 \cdot (1 + r) = \begin{cases} 3, & \text{falls } M_1 = u, \\ 0, & \text{falls } M_1 = d. \end{cases}$$

Dies ist dieselbe Auszahlungsfunktion wie die der europäischen Option. Da der Wert des Portfolios 1.2 GEen war, muss der faire Preis der europäischen Option also 1.2 GEen sein. Wäre der Preis etwa höher, könnte man eine solche Option verkaufen, und gleichzeitig die oben beschriebene Handelsstrategie wählen, wobei der Differenzbetrag noch in das festverzinsliche Wertpapier angelegt wird. Damit ist $X_0 = 0$, aber $X_1 > 0$ fast sicher, es würde sich also eine Arbitrage-Möglichkeit ergeben.

Man sagt, dass die angegebene Handelsstrategie die Option repliziert. Außerdem wird das Replizieren auch als Hedging bezeichnet.

Proposition 14.28 (Fairer Preis der europäischen Call-Option für $\tau = 1$). *Der faire Preis einer europäischen Call-Option, die zur Zeit 1 einen Wert von $V_1 = \varphi(S_1) = (S_1 - K)^+$ besitzt, ist im Binomialmodell*

$$V_0 = \frac{1}{1+r} (\tilde{p} \cdot \varphi(uS_0) + \tilde{q} \cdot \varphi(dS_0)),$$

wobei

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}.$$

Ist außerdem $X_0 = V_0$ mit

$$\Delta_0 = \frac{\varphi(uS_0) - \varphi(dS_0)}{uS_0 - dS_0},$$

so ist $X_1 = \varphi(S_1)$ fast sicher.

Bemerkung 14.29 (Interpretation). Die Formel für den Optionspreis interpretiert man am besten so: Sei $\tilde{\mathbf{P}}$ ein Wahrscheinlichkeitsmaß mit

$$\tilde{\mathbf{P}}(M_1 = u) = \tilde{p}, \quad \tilde{\mathbf{P}}(M_1 = d) = \tilde{q}.$$

Dann gilt

$$(1+r)V_0 = \tilde{\mathbf{E}}[\varphi(S_1)|S_0],$$

wobei $\tilde{\mathbf{E}}$ der Erwartungswert bezüglich $\tilde{\mathbf{P}}$ ist.

Beweis von Proposition 14.28. Wir gehen genau wie im Beispiel 14.27 vor. Wir nehmen ein Portfolio an, das aus Δ_0 Aktien besteht, und einen Wert von X_0 hat. Damit sind $X_0 - \Delta_0 S_0$ in das festverzinsliche Wertpapier investiert. Der Wert des Portfolios zur Zeit 1 ist nun

$$X_1 = \Delta_0 S_1 + (X_0 - \Delta_0 S_0)(1+r)$$

Damit $X_1 = \varphi(S_1)$ (fast sicher) gelten soll, muss also

$$\begin{aligned} \varphi(uS_0) &= \Delta_0 uS_0 + (X_0 - \Delta_0 S_0)(1+r), \\ \varphi(dS_0) &= \Delta_0 dS_0 + (X_0 - \Delta_0 S_0)(1+r). \end{aligned}$$

Löst man diese zwei Gleichungen nach X_0, Δ_0 auf, so erhält man

$$\begin{aligned} \Delta_0 &= \frac{\varphi(uS_0) - \varphi(dS_0)}{uS_0 - dS_0}, \\ X_0 &= \frac{1}{1+r} (\tilde{p} \cdot \varphi(uS_0) + \tilde{q} \cdot \varphi(dS_0)). \end{aligned} \quad \square$$

Iteriert man die letzte Proposition, bekommt man den fairen Preis der europäischen Call-Option für beliebiges τ .

Theorem 14.30 (Fairer Preis der europäischen Call-Option). *Der faire Preis zur Zeit t einer europäischen Call-Option, die zur Zeit τ einen Wert von $V_\tau = \varphi(S_\tau) = (S_\tau - K)^+$ besitzt, ist im Binomialmodell*

$$V_t = V_t(S_t) = \frac{1}{(1+r)^{\tau-t}} \sum_{s=0}^{\tau-t} \binom{\tau-t}{s} \tilde{p}^s \tilde{q}^{\tau-t-s} \varphi(u^s d^{\tau-t-s} S_t), \quad (14.8)$$

wobei

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}.$$

Ist außerdem $X_t = V_t$

$$\Delta_s = \frac{V_{s+1}(uS_s) - V_{s+1}(dS_s)}{uS_s - dS_s},$$

so ist $X_s = V_s$ fast sicher für $s = t, \dots, \tau - 1$.

Bemerkung 14.31 (Preisberechnung mit bedingten Erwartungen). Der Preis der Option berechnet sich am einfachsten mittels bedingter Erwartungen. Wir haben ja schon in Bemerkung 14.29 gesehen, dass $V_0 = \tilde{\mathbf{E}}[\varphi(S_1)|S_0]/(1+r)$ gilt. Damit ist wohl auch für den Fall $\tau > 1$

$$\begin{aligned} V_t &= \tilde{\mathbf{E}}[\varphi(S_\tau)|S_t]/(1+r)^{\tau-t} \\ &= \tilde{\mathbf{E}}[\tilde{\mathbf{E}}[\dots \tilde{\mathbf{E}}[\varphi(S_\tau)|S_{\tau-1}]|S_{\tau-2}] \dots |S_t]/(1+r)^{\tau-t} \\ &= \tilde{\mathbf{E}}[\tilde{\mathbf{E}}[\dots \tilde{\mathbf{E}}[\tilde{p}\varphi(uS_{\tau-1}) + \tilde{q}\varphi(dS_{\tau-1})|S_{\tau-2}] \dots |S_t]/(1+r)^{\tau-t-1} \\ &= \dots = \sum_{s=0}^{\tau-t} \binom{\tau-t}{s} \tilde{p}^s \tilde{q}^{\tau-t-s} \varphi(u^s d^{\tau-t-s} S_t). \end{aligned}$$

Beweis von Theorem 14.30. Am einfachsten zeigt man das Theorem durch Induktion über $t = \tau - 1, \tau - 2, \dots, 0$. Für $t = \tau - 1$ ist die Aussage gerade die von Proposition 14.28. Gilt also das Theorem schon für ein t . Gegeben S_{t-1} , weiß man also schon, dass $V_t(uS_{t-1})$ und $V_t(dS_{t-1})$ die fairen Preise der Option zur Zeit t sind, je nachdem ob $M_t = u$ oder $M_t = d$. Damit befindet man sich wieder in der Situation von Proposition 14.28, also

$$V_{t-1} = \frac{1}{1+r} (\tilde{p} \cdot V_t(uS_{t-1}) + \tilde{q} \cdot V_t(dS_{t-1})).$$

Diese Rekursion wird aber gerade von (14.8) gelöst. Genauso sieht man die Form für Δ_s ein. \square