Einführung in die Stochastik

Vorlesungsskript

Benedikt Jahnel

Technische Universität Braunschweig

Stand: 27. Oktober 2026

Inhaltsverzeichnis

Zur Literatur **4**

1 Mathematische Beschreibung von Zufallssituationen **5**
1.1 Wahrscheinlichkeitsräume . . . 5
1.1.1 Ergebnisräume . . . 5
1.1.2 Ereignisräume . . . 6
1.1.3 Wahrscheinlichkeitsräume . . . 9
1.2 Eigenschaften und Konstruktion von Wahrscheinlichkeitsmaßen . . . 9
1.3 Zufallsvariablen . . . 14

2 Stochastische Standardmodelle **18**
2.1 Urnenmodelle . . . 18
2.1.1 Ziehen mit Zurücklegen und mit Reihenfolge . . . 18
2.1.2 Ziehen ohne Zurücklegen, aber mit Reihenfolge . . . 18
2.1.3 Ziehen ohne Zurücklegen und ohne Reihenfolge . . . 18
2.1.4 Ziehen mit Zurücklegen und ohne Reihenfolge . . . 19
2.2 Die Poisson-Verteilung . . . 21
2.3 Die Gamma-Verteilung . . . 22
2.4 Die Beta-Verteilung . . . 23
2.5 Die Normalverteilung . . . 24

3 Bedingte Wahrscheinlichkeit und Unabhängigkeit **26**
3.1 Bedingte Wahrscheinlichkeit . . . 26
3.2 Mehrstufige Modelle . . . 28
3.3 Unabhängigkeit . . . 31
3.4 Existenz unabhängiger Zufallsvariablen, Produktmaße . . . 35
3.5 Der Poisson-Prozess . . . 39
3.6 Asymptotische Ereignisse . . . 42

4 Erwartungswert und Varianz **44**
4.1 Erwartungswert . . . 44
4.1.1 Der diskrete Fall . . . 44
4.1.2 Der allgemeine Fall . . . 47
4.2 Das Wartezeitparadoxon . . . 50
4.3 Varianz und Kovarianz . . . 51
4.4 Erzeugende Funktionen . . . 54

5 Gesetz der großen Zahl und zentraler Grenzwertsatz **57**
5.1 Das Gesetz der großen Zahl . . . 57
5.1.1 Das schwache Gesetz der großen Zahl . . . 57
5.1.2 Das starke Gesetz der großen Zahl . . . 63
5.2 Die Normalapproximation der Binomialverteilung . . . 66
5.3 Der zentrale Grenzwertsatz . . . 70
5.4 Normal- versus Poisson-Approximation . . . 75

5.5 Große Abweichungen 77

6 Markov-Ketten **81**
6.1 Die Markov-Eigenschaft 81
6.2 Absorptionswahrscheinlichkeiten 84
6.3 Asymptotische Stationarität 88
6.4 Wiederkehrzeiten 91

7 Maßtheoretische Grundlagen der Wahrscheinlichkeitstheorie **96**
7.1 Abstrakte Maßräume 96
7.2 Lebesgue-Integration auf abstrakten Maßräumen 103
7.3 L^p-Räume und Satz von Radon-Nikodym 113
7.4 Hauptsatz der Differential- und Integralrechnung 119

Wiederholung und Ausblick **123**

Zur Literatur

Das Skript folgt über weite Strecken dem Lehrbuch von Georgii [Geo15], geht aber an einigen Stellen darüber hinaus. Ergänzend werden die Lehrbücher von Klenke [Kle20], Bauer [Bau02] und Durrett [Dur19] empfohlen; vollständige Angaben stehen im Literaturverzeichnis am Ende.

Zum Weiterlesen eignen sich je nach Thema:

- *Einstieg und Kapitel 1–5:* Krengel [Kre05] auf dem Niveau dieser Vorlesung, Henze [Hen23] als besonders zugänglicher Einstieg sowie Kersting und Wakolbinger [KW10].
- *Maß- und Integrationstheorie (Kapitel 1 und 7):* Elstrodt [Els18], Schilling [Sch17] und Cohn [Coh13].
- *Markov-Ketten (Kapitel 6):* Norris [Nor97] und Häggström [Häg02].
- *Übungsmaterial:* Grimmett und Stirzaker [GS20b] mit dem zugehörigen Aufgabenband [GS20a].

1 Mathematische Beschreibung von Zufallssituationen

1.1 Wahrscheinlichkeitsräume

Ziel dieses Abschnitts ist es, Zufallsexperimente mathematisch zu modellieren. Ein solches Modell besteht aus drei Bestandteilen: der Menge der möglichen Ausgänge (Ergebnisraum), dem System der Ereignisse, denen eine Wahrscheinlichkeit zugeordnet werden soll (Ereignissystem), und schließlich dieser Zuordnung selbst (Wahrscheinlichkeitsmaß).

1.1.1 Ergebnisräume

Beispiel 1.1. (1) *Einmaliges Würfeln.* Die möglichen Augenzahlen bilden den Ergebnisraum

$$\Omega = \{1, \dots, 6\}.$$

(2) *n-maliges Würfeln*, $n \in \mathbb{N}$. Hier ist

$$\Omega = \{1, \dots, 6\}^n = \{\omega = (\omega_1, \dots, \omega_n) \colon \omega_i \in \{1, \dots, 6\}\},$$

wobei ω_i die Augenzahl im i-ten Wurf bezeichnet, $1 \le i \le n$. Interessiert man sich nicht für die Reihenfolge der Würfe, sondern nur dafür, wie oft jede Augenzahl gefallen ist, so ist auch

$$\widehat{\Omega} = \Big\{(k_1, \dots, k_6) \in \mathbb{N}_0^6 \colon \sum_{a=1}^{6} k_a = n\Big\}$$

ein geeigneter Ergebnisraum. Dabei ist k_a die Anzahl der Würfe mit Augenzahl a; zum Beispiel ist k_1 die Anzahl der Einsen beim n-maligen Würfeln.

(3) *Unendliche Münzwurffolge.* Codieren wir „Kopf“ mit 0 und „Zahl“ mit 1, so ist

$$\Omega = \{0, 1\}^{\mathbb{N}} = \{\omega = (\omega_i)_{i \in \mathbb{N}} \colon \omega_i \in \{0, 1\}\},$$

wobei ω_i die Münzseite im i-ten Wurf angibt.

Allgemein besteht der *erste Schritt* der mathematischen Modellierung einer Zufallssituation darin, einen geeigneten *Ergebnisraum* $\Omega \neq \emptyset$ festzulegen. Wie Beispiel 1.1(2) zeigt, ist diese Wahl nicht eindeutig; sie hängt davon ab, welche Aspekte des Experiments man beschreiben möchte.

1.1.2 Ereignisräume

Ereignisse werden durch Teilmengen von Ω beschrieben, nämlich durch die Menge aller Ausgänge, bei denen das Ereignis eintritt.

Beispiel 1.2. (1) *Gerade Augenzahl beim einmaligen Würfeln.* Für $\Omega = \{1, \dots, 6\}$ ist

$$A = \{2, 4, 6\} = \{\omega \in \Omega \colon \omega \in 2\mathbb{N}\}.$$

(2) *Mindestens k-mal Zahl beim n-maligen Münzwurf.* Für $\Omega = \{0,1\}^n$ und $0 \le k \le n$ ist

$$A = \Big\{\omega = (\omega_1, \dots, \omega_n) \in \Omega \colon \sum_{i=1}^{n} \omega_i \ge k\Big\}.$$

Allgemein besteht der *zweite Schritt* der Modellbildung in der Festlegung eines geeigneten *Ereignissystems* $\mathcal{F} \subset \mathcal{P}(\Omega)$, wobei

$$\mathcal{P}(\Omega) = \{A \colon A \subset \Omega\}$$

die *Potenzmenge* von Ω bezeichnet.

Achtung: Ist Ω überabzählbar, so ist $\mathcal{P}(\Omega)$ im Allgemeinen *zu groß* für ein sinnvolles Ereignissystem. Das zeigt der folgende Satz am Beispiel des unendlich oft wiederholten fairen Münzwurfs.

Für $n \in \mathbb{N}$ bezeichne

$$T_n \colon \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}, \qquad (\omega_1, \omega_2, \dots) \mapsto (\omega_1, \dots, \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, \dots)$$

die Abbildung, die das Ergebnis des n-ten Wurfs umdreht („Flip“). Bei einer fairen Münze sollte ein solcher Flip die Wahrscheinlichkeiten nicht verändern.

Satz 1.3 (Vitali)**.** *Sei $\Omega = \{0,1\}^{\mathbb{N}}$. Dann gibt es keine Abbildung $\mathbb{P} \colon \mathcal{P}(\Omega) \to [0,1]$ mit den folgenden Eigenschaften:*

(N) Normierung: $\mathbb{P}(\Omega) = 1$.

(A) σ-Additivität: *Sind $A_1, A_2, \dots \subset \Omega$ paarweise disjunkt, so gilt*

$$\mathbb{P}\Big(\bigcup_{i \ge 1} A_i\Big) = \sum_{i \ge 1} \mathbb{P}(A_i).$$

(I) Flip-Invarianz: *Für alle $A \subset \Omega$ und $n \ge 1$ gilt $\mathbb{P}(T_n A) = \mathbb{P}(A)$.*

Beweis. Angenommen, eine solche Abbildung $\mathbb{P}$ existiert. Wir definieren eine Äquivalenzrelation auf Ω durch

$$\omega \sim \omega' \quad :\Longleftrightarrow \quad \exists N \in \mathbb{N}\ \forall n \ge N \colon\ \omega_n = \omega'_n,$$

d. h. ω und ω' unterscheiden sich nur in endlich vielen Koordinaten. Nach dem Auswahlaxiom gibt es eine Menge $A \subset \Omega$, die aus jeder Äquivalenzklasse genau ein Element enthält.

Sei $\mathcal{S} = \{S \subset \mathbb{N} \colon |S| < \infty\}$ das System der endlichen Teilmengen von $\mathbb{N}$. Wegen

$$\mathcal{S} = \{\emptyset\} \cup \bigcup_{m \ge 1} \mathcal{S}_m, \qquad \mathcal{S}_m \coloneqq \{S \subset \mathbb{N} \colon S \ne \emptyset,\ \max S = m\},$$

und $|\mathcal{S}_m| = 2^{m-1} < \infty$ für alle m ist $\mathcal{S}$ abzählbar. Für $S = \{n_1, \dots, n_k\} \in \mathcal{S}$ sei

$$T_S \coloneqq \prod_{n \in S} T_n = T_{n_1} \circ \cdots \circ T_{n_k}$$

der Gesamtflip aller Koordinaten in S (die T_n kommutieren; $T_\emptyset = \mathrm{id}$). Dann gilt:

(1) $\Omega = \bigcup_{S \in \mathcal{S}} T_S A$. Denn zu jedem $\omega \in \Omega$ gibt es ein $\omega' \in A$ mit $\omega \sim \omega'$. Die Menge S der Koordinaten, in denen sich ω und ω' unterscheiden, ist endlich, und es gilt $\omega = T_S \omega' \in T_S A$.

(2) Die Mengen $T_S A$, $S \in \mathcal{S}$, sind paarweise disjunkt. Denn ist $T_S \omega = T_{S'} \omega'$ mit $\omega, \omega' \in A$, so folgt $\omega \sim T_S \omega = T_{S'} \omega' \sim \omega'$, also $\omega = \omega'$, da A aus jeder Klasse nur ein Element enthält. Aus $T_S \omega = T_{S'} \omega$ folgt dann $S = S'$.

Mit (N), (A) und (wiederholt angewandt) (I) folgt

$$1 = \mathbb{P}(\Omega) = \sum_{S \in \mathcal{S}} \mathbb{P}(T_S A) = \sum_{S \in \mathcal{S}} \mathbb{P}(A) \in \{0, \infty\},$$

da über abzählbar unendlich viele gleiche Summanden summiert wird. Das ist ein Widerspruch. □

Als Ausweg verzichtet man darauf, *allen* Teilmengen von Ω eine Wahrscheinlichkeit zuzuordnen, und beschränkt sich auf ein geeignetes System von Ereignissen. Dieses soll unter den üblichen logischen Operationen abgeschlossen sein.

Definition 1.4 (σ-Algebra und Ereignisraum)**.** Sei $\Omega \neq \emptyset$. Ein Mengensystem $\mathcal{F} \subset \mathcal{P}(\Omega)$ mit

(a) $\Omega \in \mathcal{F}$,

(b) $A \in \mathcal{F} \;\Rightarrow\; A^{\mathrm{c}} \coloneqq \Omega \setminus A \in \mathcal{F}$ (logisches Verneinen),

(c) $A_1, A_2, \dots \in \mathcal{F} \;\Rightarrow\; \bigcup_{i \geq 1} A_i \in \mathcal{F}$ (logisches Oder)

heißt *σ-Algebra* in Ω. Das Paar $(\Omega, \mathcal{F})$ heißt dann *Ereignisraum* oder *messbarer Raum*, die Elemente von $\mathcal{F}$ heißen *Ereignisse*.

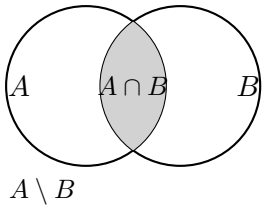
Bemerkung 1.5. Jede σ-Algebra $\mathcal{F}$ hat auch die folgenden Eigenschaften:

(1) $\emptyset \in \mathcal{F}$, denn $\emptyset = \Omega^{\mathrm{c}}$ nach Definition 1.4(a) und (b).

(2) Für $A, B \in \mathcal{F}$ gilt

$$\begin{aligned} A \cup B &\in \mathcal{F}, && \text{denn } A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \dots, \\ A \cap B &\in \mathcal{F}, && \text{denn } A \cap B = (A^{\mathrm{c}} \cup B^{\mathrm{c}})^{\mathrm{c}}, \\ A \setminus B &\in \mathcal{F}, && \text{denn } A \setminus B = A \cap B^{\mathrm{c}}. \end{aligned}$$

(3) Abzählbare Durchschnitte von Mengen in $\mathcal{F}$ gehören zu $\mathcal{F}$, denn $\bigcap_{i \geq 1} A_i = \big(\bigcup_{i \geq 1} A_i^{\mathrm{c}}\big)^{\mathrm{c}}$.



Bemerkung und Definition 1.6 (Erzeugte σ-Algebra)**.** Ist $\Omega \neq \emptyset$ und $\mathcal{G} \subset \mathcal{P}(\Omega)$, so gibt es genau eine kleinste σ-Algebra $\mathcal{F} = \sigma(\mathcal{G})$ in Ω mit $\mathcal{F} \supset \mathcal{G}$. Sie heißt die *von $\mathcal{G}$ erzeugte σ-Algebra*, und $\mathcal{G}$ heißt ein *Erzeuger* von $\mathcal{F}$.

Beweis. Sei $\mathfrak{A}$ das System aller σ-Algebren $\mathcal{A}$ in Ω mit $\mathcal{A} \supset \mathcal{G}$. Es ist $\mathfrak{A} \neq \emptyset$, da $\mathcal{P}(\Omega) \in \mathfrak{A}$. Setze

$$\mathcal{F} := \bigcap_{\mathcal{A} \in \mathfrak{A}} \mathcal{A}.$$

Dann gilt (a) $\Omega \in \mathcal{F}$, da $\Omega \in \mathcal{A}$ für alle $\mathcal{A} \in \mathfrak{A}$; (b) aus $A \in \mathcal{F}$ folgt $A \in \mathcal{A}$, also $A^c \in \mathcal{A}$ für alle $\mathcal{A} \in \mathfrak{A}$ und damit $A^c \in \mathcal{F}$; (c) analog folgt aus $A_1, A_2, \ldots \in \mathcal{F}$, dass $\bigcup_{i \geq 1} A_i \in \mathcal{F}$. Also ist $\mathcal{F}$ eine σ-Algebra mit $\mathcal{F} \supset \mathcal{G}$, und nach Konstruktion gilt $\mathcal{F} \subset \mathcal{A}$ für jedes $\mathcal{A} \in \mathfrak{A}$. Damit ist $\mathcal{F}$ die kleinste solche σ-Algebra; insbesondere ist sie eindeutig. □

Beispiel und Definition 1.7. (1) *Potenzmenge.* Sei Ω abzählbar und $\mathcal{G} = \{\{\omega\} \colon \omega \in \Omega\}$ das System der einelementigen Mengen. Dann gilt $\sigma(\mathcal{G}) = \mathcal{P}(\Omega)$, denn jedes $A \subset \Omega$ ist die abzählbare Vereinigung $A = \bigcup_{\omega \in A} \{\omega\}$.

(2) *Borel'sche σ-Algebra.* Sei $\Omega = \mathbb{R}^n$ und

$$\mathcal{G} = \Big\{ \prod_{i=1}^{n} [a_i, b_i] \colon a_i < b_i,\ a_i, b_i \in \mathbb{Q} \Big\}$$

das System der kompakten Quader mit rationalen Eckpunkten. Dann heißt $\mathcal{B}^n := \sigma(\mathcal{G})$ die *Borel'sche σ-Algebra* auf $\mathbb{R}^n$, und jedes $A \in \mathcal{B}^n$ heißt eine *Borel-Menge*. Man schreibt $\mathcal{B} := \mathcal{B}^1$.

Beachte:

(i) Jede offene Menge $A \subset \mathbb{R}^n$ ist Borel'sch. Denn zu jedem $\omega \in A$ gibt es ein $Q \in \mathcal{G}$ mit $\omega \in Q \subset A$, und daher ist

$$A = \bigcup_{Q \in \mathcal{G} \colon\ Q \subset A} Q$$

eine Vereinigung von abzählbar vielen Mengen aus $\mathcal{G}$ (da $\mathcal{G}$ abzählbar ist).

(ii) Jede abgeschlossene Menge $A \subset \mathbb{R}^n$ ist Borel'sch, da A^c offen ist.

(iii) $\mathcal{B}^n$ enthält „praktisch alle" Teilmengen des $\mathbb{R}^n$, die in Anwendungen auftreten, aber nicht alle: Mit einem ähnlichen Auswahlargument wie im Beweis von Satz 1.3 kann man nicht-Borel'sche Mengen konstruieren.

(iv) Es gilt $\mathcal{B} = \sigma(\mathcal{G}')$ mit $\mathcal{G}' = \{(-\infty, c] \colon c \in \mathbb{R}\}$ (Übung).

(v) Für $\emptyset \neq \Omega \subset \mathbb{R}^n$ ist

$$\mathcal{B}^n_\Omega := \{A \cap \Omega \colon A \in \mathcal{B}^n\}$$

eine σ-Algebra in Ω, die *Borel'sche σ-Algebra auf Ω* (auch *Spur-σ-Algebra*).

Beispiel und Definition 1.8 (Produkt-σ-Algebra)**.** Sei $I \neq \emptyset$ eine beliebige Indexmenge, und für jedes $i \in I$ sei $E_i \neq \emptyset$ eine Menge mit einer σ-Algebra $\mathcal{E}_i$. Sei

$$\Omega = \prod_{i \in I} E_i$$

das kartesische Produkt und $\pi_i \colon \Omega \to E_i$ die Projektion auf die i-te Koordinate. Mit

$$\mathcal{G} = \{\pi_i^{-1} A_i \colon i \in I,\ A_i \in \mathcal{E}_i\}$$

heißt

$$\bigotimes_{i \in I} \mathcal{E}_i := \sigma(\mathcal{G})$$

die *Produkt-σ-Algebra* der $\mathcal{E}_i$ auf Ω. Im Fall $E_i = E$ und $\mathcal{E}_i = \mathcal{E}$ für alle $i \in I$ schreiben wir $\mathcal{E}^{\otimes I}$ statt $\bigotimes_{i \in I} \mathcal{E}_i$. Beispielsweise gilt $\mathcal{B}^n = \mathcal{B}^{\otimes n}$ (Übung).

1.1.3 Wahrscheinlichkeitsräume

Der *dritte Schritt* der Modellbildung besteht darin, jedem Ereignis eine Wahrscheinlichkeit zuzuordnen.

Definition 1.9 (Wahrscheinlichkeitsmaß)**.** Sei $(\Omega, \mathcal{F})$ ein Ereignisraum und $\mathbb{P} \colon \mathcal{F} \to [0,1]$ eine Funktion mit den Eigenschaften

(N) *Normierung:* $\mathbb{P}(\Omega) = 1$,

(A) *σ-Additivität:* Für alle paarweise disjunkten $A_1, A_2, \ldots \in \mathcal{F}$ gilt

$$\mathbb{P}\Big(\bigcup_{i \geq 1} A_i\Big) = \sum_{i \geq 1} \mathbb{P}(A_i),$$

Dann heißt $\mathbb{P}$ ein *Wahrscheinlichkeitsmaß* (kurz: W-Maß) oder *Wahrscheinlichkeitsverteilung* auf $(\Omega, \mathcal{F})$. Das Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ heißt *Wahrscheinlichkeitsraum* (kurz: W-Raum).

Beispiel und Definition 1.10 (Dirac-Maß)**.** Sei $(\Omega, \mathcal{F})$ ein Ereignisraum und $\xi \in \Omega$. Dann ist

$$\delta_\xi(A) = \begin{cases} 1, & \text{falls } \xi \in A, \\ 0, & \text{sonst,} \end{cases} \qquad A \in \mathcal{F},$$

ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$, das *Dirac-Maß* in ξ. Es beschreibt den deterministischen Spezialfall, in dem mit Sicherheit der Ausgang ξ eintritt.

1.2 Eigenschaften und Konstruktion von Wahrscheinlichkeitsmaßen

Satz 1.11 (Rechenregeln für Wahrscheinlichkeitsmaße)**.** *Jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf einem Ereignisraum $(\Omega, \mathcal{F})$ hat für beliebige Ereignisse $A, B, A_1, A_2, \ldots \in \mathcal{F}$ die folgenden Eigenschaften:*

(a) $\mathbb{P}(\emptyset) = 0$.

(b) $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$ *(endliche Additivität)*

(c) $A \subset B \;\Rightarrow\; \mathbb{P}(A) \le \mathbb{P}(B)$ *(Monotonie)*

(d) $\mathbb{P}\big(\bigcup_{i\ge1} A_i\big) \le \sum_{i\ge1} \mathbb{P}(A_i)$ *(σ-Subadditivität)*

(e) *Gilt $A_n \uparrow A$ (d. h. $A_1 \subset A_2 \subset \cdots$ und $A = \bigcup_{n\ge1} A_n$) oder $A_n \downarrow A$ (d. h. $A_1 \supset A_2 \supset \cdots$ und $A = \bigcap_{n\ge1} A_n$), so gilt*

$$\mathbb{P}(A_n) \xrightarrow[n\to\infty]{} \mathbb{P}(A). \qquad (\sigma\text{-Stetigkeit})$$

Beweis. (a) Da $\Omega, \emptyset, \emptyset, \dots$ paarweise disjunkt sind, gilt nach (A)

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset \cup \emptyset \cup \dots) = \mathbb{P}(\Omega) + \sum_{i\ge1} \mathbb{P}(\emptyset),$$

und da $\mathbb{P}(\Omega) = 1 < \infty$, folgt $\mathbb{P}(\emptyset) = 0$.

(b) Ist $A \cap B = \emptyset$, so folgt aus (A) und (a)

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \cup B \cup \emptyset \cup \emptyset \cup \dots) = \mathbb{P}(A) + \mathbb{P}(B) + 0 + 0 + \dots,$$

also die Behauptung. Im allgemeinen Fall zerlegen wir in paarweise disjunkte Mengen,

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B), \qquad A = (A \setminus B) \cup (A \cap B), \qquad B = (B \setminus A) \cup (A \cap B),$$

und erhalten mit dem bereits gezeigten Spezialfall

$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + 2\,\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B).$$

(c) Für $A \subset B$ ist $B = A \cup (B \setminus A)$ eine disjunkte Vereinigung, also nach (b)

$$\mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \setminus A)}_{\ge 0} \ge \mathbb{P}(A).$$

(d) Die Mengen $A_i \setminus \bigcup_{j<i} A_j$, $i \ge 1$, sind paarweise disjunkt mit Vereinigung $\bigcup_{i\ge1} A_i$. Mit (A) und (c) folgt

$$\mathbb{P}\Big(\bigcup_{i\ge1} A_i\Big) = \sum_{i\ge1} \mathbb{P}\Big(A_i \setminus \bigcup_{j<i} A_j\Big) \le \sum_{i\ge1} \mathbb{P}(A_i).$$

(e) Sei $A_n \uparrow A$ und $A_0 := \emptyset$. Dann ist $A = \bigcup_{i\ge1}(A_i \setminus A_{i-1})$ eine disjunkte Vereinigung, und $A_n = \bigcup_{i=1}^{n}(A_i \setminus A_{i-1})$. Nach (A) und (b) gilt daher

$$\mathbb{P}(A) = \sum_{i\ge1} \mathbb{P}(A_i \setminus A_{i-1}) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_i \setminus A_{i-1}) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Der Fall $A_n \downarrow A$ folgt daraus durch Übergang zu den Komplementen: Es gilt $A_n^{\mathrm{c}} \uparrow A^{\mathrm{c}}$, und mit (b) ist $\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^{\mathrm{c}}) \to 1 - \mathbb{P}(A^{\mathrm{c}}) = \mathbb{P}(A)$. □

Eine weitere wichtige Folgerung aus der σ-Additivität ist, dass ein Wahrscheinlichkeitsmaß bereits durch seine Werte auf einem geeigneten Erzeuger der σ-Algebra festgelegt ist.

Satz 1.12 (Eindeutigkeitssatz)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathcal{F} = \sigma(\mathcal{G})$ für ein $\mathcal{G} \subset \mathcal{P}(\Omega)$. Ist $\mathcal{G}$ $\cap$-stabil, d. h. gilt $A \cap B \in \mathcal{G}$ für alle $A, B \in \mathcal{G}$, so ist $\mathbb{P}$ bereits durch seine Einschränkung $\mathbb{P}|_{\mathcal{G}} \colon \mathcal{G} \to [0,1]$ eindeutig bestimmt.*

Beweis. Sei Q ein weiteres Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$ mit $\mathbb{P}(A) = Q(A)$ für alle $A \in \mathcal{G}$. Wir betrachten

$$\mathcal{D} := \{A \in \mathcal{F} \colon \mathbb{P}(A) = Q(A)\} \supset \mathcal{G}.$$

Dann gilt:

(a) $\Omega \in \mathcal{D}$,

(b) für alle $A, B \in \mathcal{D}$ mit $A \subset B$ gilt $B \setminus A \in \mathcal{D}$,

(c) für alle paarweise disjunkten $A_1, A_2, \ldots \in \mathcal{D}$ gilt $\bigcup_{i \geq 1} A_i \in \mathcal{D}$.

In der Tat folgt (a) aus (N); (b) folgt aus Satz 1.11(b), denn

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A) = Q(B) - Q(A) = Q(B \setminus A);$$

und (c) folgt aus (A).

Ein Mengensystem $\mathcal{D} \subset \mathcal{P}(\Omega)$ mit den Eigenschaften (a)–(c) heißt *Dynkin-System*. Wie im Beweis von Bemerkung und Definition 1.6 sieht man, dass es zu jedem $\mathcal{G} \subset \mathcal{P}(\Omega)$ ein kleinstes Dynkin-System $d(\mathcal{G}) \supset \mathcal{G}$ gibt, das *von $\mathcal{G}$ erzeugte Dynkin-System*. Wegen $\mathcal{D} \supset \mathcal{G}$ gilt also $\mathcal{D} \supset d(\mathcal{G})$. Nach dem folgenden Lemma ist $d(\mathcal{G}) = \sigma(\mathcal{G}) = \mathcal{F}$, und somit $\mathcal{D} = \mathcal{F}$, d. h. $\mathbb{P} = Q$. □

Lemma 1.13 (Dynkin-Argument)**.** *Für jedes $\cap$-stabile Mengensystem $\mathcal{G} \subset \mathcal{P}(\Omega)$ gilt $d(\mathcal{G}) = \sigma(\mathcal{G})$.*

Beweis. Zu „$\subset$“: Jede σ-Algebra ist auch ein Dynkin-System (für $A \subset B$ ist $B \setminus A = B \cap A^c$, vgl. Bemerkung 1.5). Also ist $\sigma(\mathcal{G})$ ein Dynkin-System, das $\mathcal{G}$ enthält, und wegen der Minimalität von $d(\mathcal{G})$ gilt $d(\mathcal{G}) \subset \sigma(\mathcal{G})$.

Zu „$\supset$“: Es genügt zu zeigen, dass $d(\mathcal{G})$ eine σ-Algebra ist; dann folgt $\sigma(\mathcal{G}) \subset d(\mathcal{G})$ aus der Minimalität von $\sigma(\mathcal{G})$.

Schritt 1: $d(\mathcal{G})$ ist $\cap$-stabil. Sei

$$\mathcal{D}_1 := \{A \subset \Omega \colon A \cap B \in d(\mathcal{G}) \text{ für alle } B \in \mathcal{G}\}.$$

Das System $\mathcal{D}_1$ ist ein Dynkin-System: Es ist $\Omega \in \mathcal{D}_1$ wegen $\mathcal{G} \subset d(\mathcal{G})$; für $A \subset A'$ aus $\mathcal{D}_1$ und $B \in \mathcal{G}$ ist $(A' \setminus A) \cap B = (A' \cap B) \setminus (A \cap B) \in d(\mathcal{G})$; und für paarweise disjunkte $A_1, A_2, \ldots \in \mathcal{D}_1$ ist $\big(\bigcup_i A_i\big) \cap B = \bigcup_i (A_i \cap B)$ eine disjunkte Vereinigung von Mengen aus $d(\mathcal{G})$. Da $\mathcal{G}$ $\cap$-stabil ist, gilt $\mathcal{G} \subset \mathcal{D}_1$, also $d(\mathcal{G}) \subset \mathcal{D}_1$. Das heißt: $A \cap B \in d(\mathcal{G})$ für alle $A \in d(\mathcal{G})$ und $B \in \mathcal{G}$.

Ebenso ist

$$\mathcal{D}_2 := \{A \subset \Omega \colon A \cap B \in d(\mathcal{G}) \text{ für alle } B \in d(\mathcal{G})\}$$

ein Dynkin-System, und nach dem eben Gezeigten gilt $\mathcal{G} \subset \mathcal{D}_2$. Also folgt $d(\mathcal{G}) \subset \mathcal{D}_2$, d. h. $d(\mathcal{G})$ ist $\cap$-stabil.

Schritt 2: $d(\mathcal{G})$ ist eine σ-Algebra. Eigenschaft (a) aus Definition 1.4 gilt, da jedes Dynkin-System Ω enthält, (b) folgt aus $A^c = \Omega \setminus A$. Zu (c): Für $A_1, A_2, \ldots \in d(\mathcal{G})$ und $i \geq 1$ ist

$$B_i := A_i \setminus \bigcup_{j<i} A_j = A_i \cap \bigcap_{j<i} (\Omega \setminus A_j) \in d(\mathcal{G})$$

nach Schritt 1 (endlicher Durchschnitt). Die B_i sind paarweise disjunkt, also gilt

$$\bigcup_{i \geq 1} A_i = \bigcup_{i \geq 1} B_i \in d(\mathcal{G}).$$ □

Wie steht es um die *Existenz* von Wahrscheinlichkeitsmaßen mit vorgegebenen Werten auf einem Erzeuger? Hierauf gibt der *Fortsetzungssatz von Carathéodory* aus der Maßtheorie eine Antwort, der in den Übungen angedeutet wird. Für unsere Zwecke genügt die Existenz des Lebesgue-Integrals, die wir als bekannt voraussetzen.

Tatsache 1.14 (Lebesgue-Integral)**.** *Für jede Funktion* $f\colon \mathbb{R}^n \to [0,\infty]$*, die* Borel-messbar *ist, d. h.*

$$\{x \in \mathbb{R}^n \colon f(x) \le c\} = f^{-1}([0,c]) \in \mathcal{B}^n \qquad \text{für alle } c > 0,$$

kann das Lebesgue-Integral $\int f(x)\,\mathrm{d}x \in [0,\infty]$ *so erklärt werden, dass gilt:*

(a) *Für jede (auf einem kompakten Quader) Riemann-integrierbare Funktion* $f \ge 0$*, außerhalb des Quaders durch* 0 *fortgesetzt, stimmt* $\int f(x)\,\mathrm{d}x$ *mit dem Riemann-Integral überein.*

(b) *Für alle messbaren* $f_1, f_2, \ldots \ge 0$ *gilt*

$$\int \sum_{n\ge 1} f_n(x)\,\mathrm{d}x = \sum_{n\ge 1} \int f_n(x)\,\mathrm{d}x.$$

(c) Lemma von Fatou: *Für alle messbaren* $f_1, f_2, \ldots \ge 0$ *gilt*

$$\int \liminf_{n\to\infty} f_n(x)\,\mathrm{d}x \le \liminf_{n\to\infty} \int f_n(x)\,\mathrm{d}x.$$

(d) Satz von der majorisierten Konvergenz (Lebesgue): *Seien* $f_n, f, g\colon \mathbb{R}^n \to \mathbb{R}$ *messbar mit* $f_n(x) \to f(x)$ *und* $|f_n(x)| \le g(x)$ *für alle* $x \in \mathbb{R}^n$ *und* $n \ge 1$*, sowie* $\int g(x)\,\mathrm{d}x < \infty$*. Dann gilt*

$$\lim_{n\to\infty} \int f_n(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x.$$

Dabei heißt eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ *messbar, wenn* $\{f \le c\} \in \mathcal{B}^n$ *für alle* $c \in \mathbb{R}$ *gilt, und für solche* f *mit* $\int |f(x)|\,\mathrm{d}x < \infty$ *ist* $\int f(x)\,\mathrm{d}x := \int f^+(x)\,\mathrm{d}x - \int f^-(x)\,\mathrm{d}x$ *mit* $f^\pm := \max\{\pm f, 0\}$.

Bemerkung und Definition 1.15. (1) Die Lebesgue-Integration wird in den Übungen weiter vertieft.

(2) Für $A \subset \mathbb{R}^n$ schreiben wir

$$\mathbb{1}_A(x) = \mathbb{1}\{x \in A\} = \begin{cases} 1, & \text{falls } x \in A, \\ 0, & \text{sonst,} \end{cases}$$

für die *Indikatorfunktion* von A. Für $A \in \mathcal{B}^n$ ist $\mathbb{1}_A$ messbar.

(3) Die Abbildung $\lambda^n\colon \mathcal{B}^n \to [0,\infty]$,

$$\lambda^n(A) := \int \mathbb{1}\{x \in A\}\,\mathrm{d}x,$$

ist ein *Maß* auf $(\mathbb{R}^n, \mathcal{B}^n)$, d. h. es gilt $\lambda^n(\emptyset) = 0$, und es erfüllt die σ-Additivitätseigenschaft (A) (nach Tatsache 1.14(b)), im Allgemeinen aber nicht die Normierung (N). Die Abbildung λ^n heißt das (n-dimensionale) *Lebesgue-Maß* auf $\mathbb{R}^n$. Für $\Omega \in \mathcal{B}^n$ heißt die Einschränkung λ^n_Ω von λ^n auf $\mathcal{B}^n_\Omega$ das *Lebesgue-Maß auf* Ω.

Satz 1.16 (Konstruktion von Wahrscheinlichkeitsmaßen durch Dichten)**.**

(a) Diskreter Fall: *Sei Ω abzählbar und $\rho = (\rho(\omega))_{\omega\in\Omega}$ eine Folge in $[0,1]$ mit $\sum_{\omega\in\Omega}\rho(\omega) = 1$. Dann bestimmt ρ genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{P}(\Omega))$ mit*

$$\mathbb{P}(A) = \sum_{\omega\in A}\rho(\omega) \qquad \text{für alle } A \in \mathcal{P}(\Omega).$$

Umgekehrt ist jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{P}(\Omega))$ von dieser Form mit $\rho(\omega) = \mathbb{P}(\{\omega\})$. Die Folge ρ heißt Zähldichte *von $\mathbb{P}$.*

(b) Stetiger Fall: *Ist $\Omega \in \mathcal{B}^n$ und $\rho\colon \Omega \to [0,\infty)$ messbar mit $\int_\Omega \rho(x)\,\mathrm{d}x = 1$, so gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{B}^n_\Omega)$ mit*

$$\mathbb{P}(A) = \int_A \rho(x)\,\mathrm{d}x \qquad \text{für alle } A \in \mathcal{B}^n_\Omega.$$

Die Funktion ρ heißt Dichtefunktion *(oder* Lebesgue-Dichte*) von $\mathbb{P}$.*

Dabei ist $\int_A \rho(x)\,\mathrm{d}x := \int \rho(x)\mathbb{1}_A(x)\,\mathrm{d}x$, wobei ρ außerhalb von Ω durch 0 fortgesetzt wird.

Beweis. (a) Durch $\mathbb{P}(A) := \sum_{\omega\in A}\rho(\omega)$ wird eine Abbildung $\mathbb{P}\colon \mathcal{P}(\Omega) \to [0,1]$ definiert (die Reihe hat nichtnegative Glieder, ihr Wert hängt also nicht von der Summationsreihenfolge ab). Es gilt $\mathbb{P}(\Omega) = 1$, und für paarweise disjunkte $A_1, A_2, \ldots$ folgt die σ-Additivität aus dem Umordnungssatz für Reihen mit nichtnegativen Gliedern:

$$\mathbb{P}\Big(\bigcup_{i\ge1} A_i\Big) = \sum_{\omega\in\bigcup_i A_i}\rho(\omega) = \sum_{i\ge1}\sum_{\omega\in A_i}\rho(\omega) = \sum_{i\ge1}\mathbb{P}(A_i).$$

Die Eindeutigkeit und die Umkehrung folgen aus der σ-Additivität: Für jedes W-Maß $\mathbb{P}$ auf $(\Omega, \mathcal{P}(\Omega))$ und jedes $A \subset \Omega$ ist $A = \bigcup_{\omega\in A}\{\omega\}$ eine abzählbare disjunkte Vereinigung, also $\mathbb{P}(A) = \sum_{\omega\in A}\mathbb{P}(\{\omega\})$.

(b) Die Eindeutigkeit gilt, da $\mathbb{P}$ durch die Formel festgelegt ist. Die Normierung (N) gilt nach Voraussetzung. Für paarweise disjunkte $A_1, A_2, \ldots \in \mathcal{B}^n_\Omega$ ist $\mathbb{1}_{\bigcup_i A_i} = \sum_{i\ge1}\mathbb{1}_{A_i}$, und Tatsache 1.14(b), angewandt auf $f_i = \rho\,\mathbb{1}_{A_i}$, liefert die σ-Additivität (A). □

Beispiel und Definition 1.17 (Gleichverteilung)**.** (1) Sei Ω endlich. Das Wahrscheinlichkeitsmaß zur Zähldichte $\rho(\omega) = 1/|\Omega|$, $\omega \in \Omega$, heißt die (*diskrete*) *Gleichverteilung* auf Ω. Sie ist gegeben durch

$$\mathcal{U}_\Omega(A) = \sum_{\omega\in A}\rho(\omega) = \frac{|A|}{|\Omega|}, \qquad A \subset \Omega.$$

(2) Ist $\Omega \in \mathcal{B}^n$ mit $0 < \lambda^n(\Omega) < \infty$, so heißt das Wahrscheinlichkeitsmaß zur Dichtefunktion $\rho(x) = 1/\lambda^n(\Omega)$, $x \in \Omega$, die (*stetige*) *Gleichverteilung* auf Ω. Sie ist gegeben durch

$$\mathcal{U}_\Omega(A) = \int_A \rho(x)\,\mathrm{d}x = \frac{\lambda^n(A)}{\lambda^n(\Omega)}, \qquad A \in \mathcal{B}^n_\Omega.$$

1.3 Zufallsvariablen

Häufig interessiert man sich nicht für den genauen Ausgang eines Zufallsexperiments, sondern nur für gewisse Aspekte davon. Der Übergang zu einer solchen gröberen Beschreibung geschieht durch Abbildungen zwischen Ergebnisräumen.

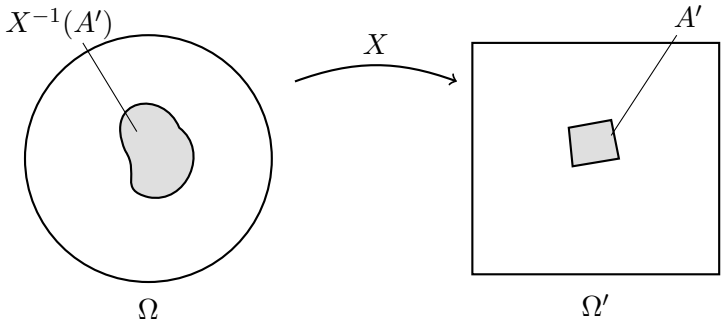
Beispiel 1.18. Wir betrachten den n-maligen Münzwurf. Der Ergebnisraum $\Omega = \{0,1\}^n$ beschreibt jeden einzelnen Ausgang vollständig. Interessiert man sich nur für die Anzahl der Einsen (also nicht für die Reihenfolge), so genügt der Ergebnisraum $\Omega' = \{0,1,\dots,n\}$; er entspricht einer geringeren Beobachtungstiefe. Der Übergang wird beschrieben durch die Abbildung

$$X\colon \Omega \to \Omega', \qquad \omega = (\omega_1,\dots,\omega_n) \mapsto \sum_{i=1}^{n} \omega_i.$$

Sind $\mathcal{F}$ und $\mathcal{F}'$ σ-Algebren in Ω bzw. Ω' und ist auf $(\Omega,\mathcal{F})$ ein Wahrscheinlichkeitsmaß gegeben, so fordern wir, damit jedem Ereignis $A' \in \mathcal{F}'$ eine Wahrscheinlichkeit zugeordnet werden kann,

$$X^{-1}(A') \in \mathcal{F} \qquad \text{für alle } A' \in \mathcal{F}',$$

d. h. die Menge der Ergebnisse, die zu dem Ereignis A' führen, muss selbst ein Ereignis (also eine messbare Menge) sein.



Definition 1.19 (Zufallsvariable)**.** Seien $(\Omega,\mathcal{F})$ und $(\Omega',\mathcal{F}')$ zwei Ereignisräume. Dann heißt jede Abbildung $X\colon \Omega \to \Omega'$ mit

$$X^{-1}(A') \in \mathcal{F} \qquad \text{für alle } A' \in \mathcal{F}'$$

eine *Zufallsvariable* (kurz: ZV) von $(\Omega,\mathcal{F})$ nach $(\Omega',\mathcal{F}')$, oder auch ein *Zufallselement* von Ω'. Man sagt auch, X sei *messbar*.

Bemerkung 1.20. (1) Wir schreiben oft

$$\{X \in A'\} \coloneqq \{\omega \in \Omega\colon X(\omega) \in A'\} = X^{-1}(A').$$

(2) Ist $\mathcal{F}' = \sigma(\mathcal{G}')$ für ein $\mathcal{G}' \subset \mathcal{P}(\Omega')$, so gilt:

$$X \text{ ist eine Zufallsvariable} \quad\Longleftrightarrow\quad X^{-1}(A') \in \mathcal{F} \text{ für alle } A' \in \mathcal{G}'.$$

(Beweis als Übung.)

Beispiel 1.21. (1) Ist $\mathcal{F} = \mathcal{P}(\Omega)$, so ist jede Abbildung $X\colon \Omega \to \Omega'$ eine Zufallsvariable.

(2) Ist $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B})$, so ist X genau dann eine Zufallsvariable, wenn

$$\{X \le c\} = X^{-1}((-\infty, c]) \in \mathcal{F} \qquad \text{für alle } c \in \mathbb{R}$$

gilt, oder äquivalent, wenn

$$\{X < c\} = X^{-1}((-\infty, c)) \in \mathcal{F} \qquad \text{für alle } c \in \mathbb{R}.$$

Dies folgt aus Bemerkung 1.20(2), da sowohl die Intervalle $(-\infty, c]$ als auch die Intervalle $(-\infty, c)$, $c \in \mathbb{R}$, die Borel'sche σ-Algebra erzeugen: Für die ersten ist das Beispiel und Definition 1.7(iv), und wegen

$$(-\infty, c) = \bigcup_{n\ge 1} (-\infty, c - \tfrac{1}{n}], \qquad (-\infty, c] = \bigcap_{n\ge 1} (-\infty, c + \tfrac{1}{n})$$

erzeugen beide Systeme dieselbe σ-Algebra.

(3) Ist $\emptyset \ne \Omega \subset \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}^n_\Omega$ und $X\colon \Omega \to \mathbb{R}$ *stetig*, so ist X eine Zufallsvariable, denn $\{X \le c\}$ ist abgeschlossen in Ω, also von der Form $A \cap \Omega$ mit einer abgeschlossenen Menge $A \subset \mathbb{R}^n$, und liegt damit in $\mathcal{B}^n_\Omega$.

Satz 1.22 (Verteilung einer Zufallsvariablen)**.** *Ist X eine Zufallsvariable von $(\Omega, \mathcal{F}, \mathbb{P})$ nach $(\Omega', \mathcal{F}')$, so wird durch*

$$\mathbb{P}'(A') := \mathbb{P}(X^{-1}(A')) = \mathbb{P}(\{X \in A'\}), \qquad A' \in \mathcal{F}',$$

ein Wahrscheinlichkeitsmaß $\mathbb{P}'$ auf $(\Omega', \mathcal{F}')$ definiert. Das Maß $\mathbb{P}'$ heißt die Verteilung von X unter $\mathbb{P}$ *oder das* Bild von $\mathbb{P}$ unter X *und wird mit $\mathbb{P} \circ X^{-1}$ bezeichnet. Zwei Zufallsvariablen heißen* identisch verteilt*, falls sie die gleiche Verteilung haben.*

Beweis. $\mathbb{P}'$ ist wohldefiniert, da $X^{-1}(A') \in \mathcal{F}$ für alle $A' \in \mathcal{F}'$. Die Abbildung $\mathbb{P}'$ erfüllt (N), denn $\mathbb{P}'(\Omega') = \mathbb{P}(X \in \Omega') = \mathbb{P}(\Omega) = 1$. Zu (A): Sind $A'_1, A'_2, \ldots \in \mathcal{F}'$ paarweise disjunkt, so sind auch die Urbilder $X^{-1}(A'_i)$ paarweise disjunkt, und es gilt

$$\mathbb{P}'\Big(\bigcup_{i\ge1} A'_i\Big) = \mathbb{P}\Big(X^{-1}\Big(\bigcup_{i\ge1} A'_i\Big)\Big) = \mathbb{P}\Big(\bigcup_{i\ge1} X^{-1}(A'_i)\Big) = \sum_{i\ge1} \mathbb{P}(X^{-1}(A'_i)) = \sum_{i\ge1} \mathbb{P}'(A'_i). \qquad \square$$

Für eine reellwertige Zufallsvariable X auf $(\Omega, \mathcal{F}, \mathbb{P})$ ist die Verteilung $\mathbb{P} \circ X^{-1}$ bereits durch die Werte

$$\mathbb{P}(X \le c) = \mathbb{P} \circ X^{-1}((-\infty, c]), \qquad c \in \mathbb{R},$$

eindeutig festgelegt. Denn die Familie der Mengen $(-\infty, c]$, $c \in \mathbb{R}$, ist ein $\cap$-stabiler Erzeuger von $\mathcal{B}$, so dass der Eindeutigkeitssatz 1.12 greift.

Definition 1.23 (Verteilungsfunktion)**.** Sei X eine reellwertige Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann heißt die Funktion

$$F_X\colon \mathbb{R} \to [0,1], \qquad F_X(c) := \mathbb{P}(X \le c),$$

die *Verteilungsfunktion* von X. Ist $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$, so heißt

$$F_{\mathbb{P}} := F_{\mathrm{id}}\colon c \mapsto \mathbb{P}((-\infty, c])$$

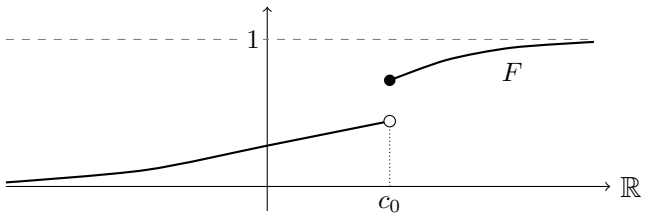
die *Verteilungsfunktion von* $\mathbb{P}$; dabei ist $\mathrm{id}\colon x \mapsto x$ die identische Abbildung.

Bemerkung 1.24. (1) Es gilt $F_X = F_{\mathbb{P}\circ X^{-1}}$, denn $\mathbb{P}(X \le c) = \mathbb{P} \circ X^{-1}\big((-\infty, c]\big)$.

(2) Jede Verteilungsfunktion F ist monoton wachsend, rechtsstetig, und es gilt

$$\lim_{c\to-\infty} F(c) = 0, \qquad \lim_{c\to\infty} F(c) = 1.$$

Dies folgt aus den Rechenregeln in Satz 1.11: Die Monotonie folgt aus Satz 1.11(c); für $c_n \downarrow c$ gilt $\{X \le c_n\} \downarrow \{X \le c\}$, für $c_n \downarrow -\infty$ gilt $\{X \le c_n\} \downarrow \emptyset$ und für $c_n \uparrow \infty$ gilt $\{X \le c_n\} \uparrow \Omega$, so dass die übrigen Aussagen aus der σ-Stetigkeit (Satz 1.11(e)) folgen.



Umgekehrt ist jede Funktion mit den Eigenschaften aus Bemerkung 1.24(2) die Verteilungsfunktion einer Zufallsvariablen. Dabei bezeichnet $\mathcal{U}_{(0,1)}$ die Gleichverteilung auf $(0,1)$ aus Beispiel und Definition 1.17.

Proposition 1.25 (Quantil-Transformation)**.** *Zu jeder monoton wachsenden, rechtsstetigen Funktion $F\colon \mathbb{R} \to [0,1]$ mit $\lim_{c\to-\infty} F(c) = 0$ und $\lim_{c\to\infty} F(c) = 1$ existiert eine reelle Zufallsvariable X auf dem Wahrscheinlichkeitsraum $\big((0,1), \mathcal{B}_{(0,1)}, \mathcal{U}_{(0,1)}\big)$ mit $F_X = F$, nämlich*

$$X(u) = \inf\{c \in \mathbb{R}\colon F(c) \ge u\}, \qquad u \in (0,1).$$

Beweis. Sei $u \in (0,1)$. Wegen $\lim_{c\to\infty} F(c) = 1 > u$ ist die Menge $\{c \in \mathbb{R}\colon F(c) \ge u\}$ nicht leer, und wegen $\lim_{c\to-\infty} F(c) = 0 < u$ und der Monotonie von F ist sie nach unten beschränkt. Also gilt $X(u) \in \mathbb{R}$.

Wir zeigen nun für alle $c \in \mathbb{R}$ und $u \in (0,1)$:

$$X(u) \le c \quad \Longleftrightarrow \quad u \le F(c).$$

(X ist also eine Art „linksseitige Umkehrfunktion“ von F.) Gilt $u \le F(c)$, so liegt c in der Menge, über die das Infimum gebildet wird, also $X(u) \le c$. Gilt umgekehrt $X(u) \le c$, so gibt es zu jedem $\varepsilon > 0$ ein $c' < X(u)+\varepsilon$ mit $F(c') \ge u$, wegen der Monotonie also $F(X(u)+\varepsilon) \ge u$. Mit $\varepsilon \downarrow 0$ folgt aus der Rechtsstetigkeit $F(X(u)) \ge u$ und damit $F(c) \ge F(X(u)) \ge u$.

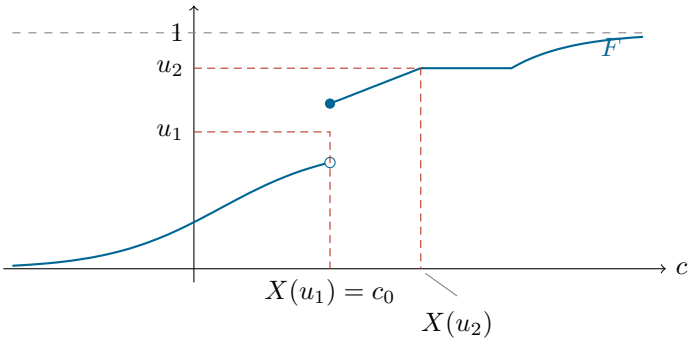
Folglich ist

$$\{X \le c\} = (0, F(c)] \cap (0,1) \in \mathcal{B}_{(0,1)},$$

so dass X nach Beispiel 1.21(2) eine Zufallsvariable ist, und es gilt

$$F_X(c) = \mathbb{P}(X \le c) = \mathcal{U}_{(0,1)}\big((0, F(c)] \cap (0,1)\big) = F(c). \qquad \square$$

Anschaulich entsteht X durch „Umkehren“ des Graphen von F, wobei Sprünge und Plateaus von F zu Plateaus bzw. Sprüngen von X werden:



Schematisch: $X(u)$ ist die kleinste Stelle c mit $F(c) \geq u$. Liegt u in einem Sprung von F (hier u_1), so ist $X(u)$ die Sprungstelle c_0; liegt u auf der Höhe eines Plateaus (hier u_2), so ist $X(u)$ dessen linker Rand.

Bemerkung und Definition 1.26 (Existenz einer Verteilungsdichte)**.** Sei X eine reelle Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$. Man sagt, F_X (bzw. X) besitze die *Dichtefunktion* $\rho\colon \mathbb{R} \to [0, \infty)$, wenn ρ messbar ist und

$$F_X(c) = \int_{-\infty}^{c} \rho(x)\,\mathrm{d}x \qquad \text{für alle } c \in \mathbb{R}$$

gilt. Die Funktion ρ wird dann *Verteilungsdichte* von X genannt. Nach dem Eindeutigkeitssatz 1.12 ist ρ dann eine Dichtefunktion von $\mathbb{P} \circ X^{-1}$ im Sinne von Satz 1.16(b); die Normierung $\int \rho(x)\,\mathrm{d}x = 1$ folgt dabei aus $F_X(c) \to 1$ für $c \to \infty$ mit monotoner Konvergenz (vgl. Tatsache 1.14(b)).

Insbesondere gilt: F_X ist genau dann stetig differenzierbar, wenn X eine stetige Verteilungsdichte ρ besitzt, und dann ist $F_X' = \rho$. Dies folgt aus dem Hauptsatz der Differential- und Integralrechnung: Ist ρ stetig, so ist F_X eine Stammfunktion von ρ. Ist umgekehrt F_X stetig differenzierbar, so ist $\rho \coloneqq F_X' \geq 0$ stetig, und es gilt $F_X(c) - F_X(a) = \int_a^c \rho(x)\,\mathrm{d}x$ für $a < c$; mit $a \to -\infty$ folgt $F_X(c) = \int_{-\infty}^c \rho(x)\,\mathrm{d}x$ (monotone Konvergenz, Tatsache 1.14(b)).

2 Stochastische Standardmodelle

2.1 Urnenmodelle

Wir betrachten eine Urne mit N unterscheidbaren Kugeln, die wir mit den Zahlen der Menge

$$M = \{1, 2, \ldots, N\}$$

bezeichnen. Aus der Urne ziehen wir nacheinander n Kugeln. Ein Ergebnis wird zunächst durch ein Tupel

$$(k_1, \ldots, k_n) \in \{1, \ldots, N\}^n$$

beschrieben, wobei $k_i \in M$ die im i-ten Zug gezogene Kugel ist. Wir betrachten verschiedene Varianten dieses Experiments und die jeweils zugehörigen Ergebnisräume.

2.1.1 Ziehen mit Zurücklegen und mit Reihenfolge

Nach jedem Zug wird die gezogene Kugel in die Urne zurückgelegt, und wir notieren die Reihenfolge der Entnahme. Der Ergebnisraum ist

$$\Omega_{\mathrm{I}} = \{(k_1, \ldots, k_n) \colon k_1, \ldots, k_n \in M\} = M^n, \qquad \text{wobei } |\Omega_{\mathrm{I}}| = N^n.$$

2.1.2 Ziehen ohne Zurücklegen, aber mit Reihenfolge

Nun sei $n \leq N$, und gezogene Kugeln werden nicht zurückgelegt. Dann ist

$$\Omega_{\mathrm{II}} = \{(k_1, \ldots, k_n) \colon k_1, \ldots, k_n \in M \text{ paarweise verschieden}\},$$

und hier gilt

$$|\Omega_{\mathrm{II}}| = N(N-1)\cdots(N-n+1) = \frac{N!}{(N-n)!}.$$

2.1.3 Ziehen ohne Zurücklegen und ohne Reihenfolge

Kommt es auf die Reihenfolge nicht an, so ist nur die Menge der gezogenen Kugeln relevant:

$$\Omega_{\mathrm{III}} = \{\{k_1, \ldots, k_n\} \subset M \colon k_1, \ldots, k_n \text{ paarweise verschieden}\} = \{A \subset M \colon |A| = n\}.$$

Da jede n-elementige Menge $\{k_1, \ldots, k_n\}$ auf genau $n!$ Arten angeordnet werden kann, gilt

$$|\Omega_{\mathrm{III}}| = \frac{|\Omega_{\mathrm{II}}|}{n!} = \frac{N!}{n!\,(N-n)!} = \binom{N}{n}$$

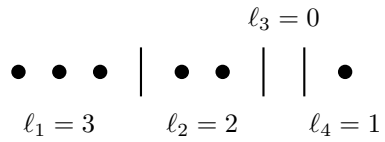
(Binomialkoeffizient).

2.1.4 Ziehen mit Zurücklegen und ohne Reihenfolge

Hier notieren wir nur, wie oft jede Kugel gezogen wurde:

$$\Omega_{\mathrm{IV}} \coloneqq \Big\{(\ell_1, \dots, \ell_N) \in \mathbb{N}_0^N \colon \sum_{i=1}^{N} \ell_i = n\Big\},$$

wobei ℓ_i die Anzahl der Ziehungen der Kugel i ist. Zum Abzählen stellen wir ein Ergebnis als Folge von n Punkten (den Ziehungen) und $N-1$ Trennstrichen dar: Vor dem ersten Strich stehen ℓ_1 Punkte, zwischen dem ersten und zweiten Strich ℓ_2 Punkte usw. Zum Beispiel entspricht für $N = 4$ und $n = 6$ die Folge



dem Ergebnis $(3, 2, 0, 1)$. Ein Ergebnis ist also eindeutig durch die Positionen der $N-1$ Striche unter den insgesamt $n + N - 1$ Plätzen bestimmt. Daher gilt

$$|\Omega_{\mathrm{IV}}| = \big|\{A \subset \{1, \dots, n+N-1\} \colon |A| = N-1\}\big| = \binom{n+N-1}{N-1} = \binom{n+N-1}{n},$$

wobei wir das Ergebnis aus Abschnitt 2.1.3 (mit $n + N - 1$ Kugeln, von denen $N - 1$ ohne Zurücklegen und ohne Reihenfolge gezogen werden) verwendet haben.

Interessiert man sich nur dafür, wie viele Kugeln jeder „Farbe“ gezogen werden, gelangt man zu den folgenden beiden Verteilungen.

Definition 2.1 (Hypergeometrische Verteilung und Multinomialverteilung)**.** Sei E eine Menge mit $2 \le |E| < \infty$ und $n \ge 1$. Wir setzen

$$\Omega_n \coloneqq \Big\{\vec{k} = (k_i)_{i \in E} \in \mathbb{N}_0^E \colon \sum_{i \in E} k_i = n\Big\}.$$

(1) Sei $\vec{N} = (N_i)_{i \in E} \in \mathbb{N}_0^E$ mit $N \coloneqq \sum_{i \in E} N_i \ge n$. Dann heißt das Wahrscheinlichkeitsmaß $\mathcal{H}_{n,\vec{N}}$ auf $(\Omega_n, \mathcal{P}(\Omega_n))$ mit der Zähldichte

$$\mathcal{H}_{n,\vec{N}}(\{\vec{k}\}) = \frac{1}{\binom{N}{n}} \prod_{i \in E} \binom{N_i}{k_i}, \qquad \vec{k} \in \Omega_n,$$

die *hypergeometrische Verteilung* (mit der Konvention $\binom{N_i}{k_i} = 0$ für $k_i > N_i$).

(2) Sei $\rho = (\rho(i))_{i \in E}$ eine Zähldichte auf E. Dann heißt das Wahrscheinlichkeitsmaß $\mathcal{M}_{n,\rho}$ auf $(\Omega_n, \mathcal{P}(\Omega_n))$ mit der Zähldichte

$$\mathcal{M}_{n,\rho}(\{\vec{k}\}) = \binom{n}{\vec{k}} \prod_{i \in E} \rho(i)^{k_i}, \qquad \text{wobei} \quad \binom{n}{\vec{k}} \coloneqq \begin{cases} \dfrac{n!}{\prod_{i \in E} k_i!}, & \text{falls } \sum_{i \in E} k_i = n, \\ 0, & \text{sonst,} \end{cases}$$

die *Multinomialverteilung*.

Interpretation. Eine Urne enthalte N_i Kugeln der Farbe $i \in E$. Zieht man n Kugeln ohne Zurücklegen und ohne Beachtung der Reihenfolge (Abschnitt 2.1.3), so ist $\mathcal{H}_{n,\vec{N}}(\{\vec{k}\})$ die Wahrscheinlichkeit, genau k_i Kugeln der Farbe i (für alle $i \in E$) zu erhalten: Unter den $\binom{N}{n}$ gleich wahrscheinlichen Teilmengen gibt es $\prod_{i\in E}\binom{N_i}{k_i}$ günstige. Insbesondere ist $\mathcal{H}_{n,\vec{N}}$ tatsächlich eine Zähldichte. Zieht man dagegen mit Zurücklegen, wobei die Farbe i jeweils mit Wahrscheinlichkeit $\rho(i)$ gezogen wird, und notiert nur die Anzahlen der Farben, so erhält man die Multinomialverteilung; dass sich die Gewichte zu 1 summieren, folgt aus dem Multinomialsatz $\sum_{\vec{k}\in\Omega_n}\binom{n}{\vec{k}}\prod_{i\in E}\rho(i)^{k_i} = \bigl(\sum_{i\in E}\rho(i)\bigr)^n = 1$.

Im Spezialfall $E = \{0,1\}$ mit $\rho(1) = p \in [0,1]$, $\rho(0) = 1-p$, hängt $\vec{k} = (n-k,k)$ nur von der Anzahl $k = k_1$ der Einsen ab, und man erhält die *Binomialverteilung*

$$\mathcal{M}_{n,\rho}(\{\vec{k}\}) = \mathcal{B}_{n,p}(\{k\}) = \binom{n}{k}p^k(1-p)^{n-k}, \qquad k \in \{0,\dots,n\}.$$

Ist die Anzahl der Kugeln in der Urne sehr groß im Vergleich zur Anzahl der Ziehungen, so sollte es kaum eine Rolle spielen, ob mit oder ohne Zurücklegen gezogen wird.

Satz 2.2 (Multinomialapproximation der hypergeometrischen Verteilung)**.** *Sei ρ eine Zähldichte auf E, und für jedes N sei $\vec{N} = (N_i)_{i\in E}$ mit $\sum_{i\in E} N_i = N$ so gewählt, dass*

$$\frac{N_i}{N} \xrightarrow[N\to\infty]{} \rho(i) \qquad \textit{für alle } i \in E.$$

Dann gilt

$$\mathcal{H}_{n,\vec{N}}(\{\vec{k}\}) \xrightarrow[N\to\infty]{} \mathcal{M}_{n,\rho}(\{\vec{k}\}) \qquad \textit{für alle } \vec{k} \in \Omega_n.$$

Beweis. Sei $\vec{k} \in \Omega_n$ fest. Für jedes $i \in E$ gilt

$$\binom{N_i}{k_i} = \frac{N_i(N_i-1)\cdots(N_i-k_i+1)}{k_i!} = \frac{N^{k_i}}{k_i!}\prod_{j=0}^{k_i-1}\Bigl(\frac{N_i}{N} - \frac{j}{N}\Bigr),$$

also

$$\frac{\binom{N_i}{k_i}}{N^{k_i}} \xrightarrow[N\to\infty]{} \frac{\rho(i)^{k_i}}{k_i!} \qquad \text{und ebenso} \qquad \frac{\binom{N}{n}}{N^n} \xrightarrow[N\to\infty]{} \frac{1}{n!}.$$

Wegen $\sum_{i\in E} k_i = n$ folgt

$$\mathcal{H}_{n,\vec{N}}(\{\vec{k}\}) = \frac{\prod_{i\in E}\binom{N_i}{k_i}/N^{k_i}}{\binom{N}{n}/N^n} \xrightarrow[N\to\infty]{} \frac{\prod_{i\in E}\rho(i)^{k_i}/k_i!}{1/n!} = \binom{n}{\vec{k}}\prod_{i\in E}\rho(i)^{k_i} = \mathcal{M}_{n,\rho}(\{\vec{k}\}).$$

Mit der Notation $a(N) \sim b(N)$ für $a(N)/b(N) \to 1$ besagt die Rechnung im Fall $\rho(i) > 0$ gerade $\binom{N_i}{k_i} \sim N_i^{k_i}/k_i!$, d. h. für große N_i spielt es keine Rolle, ob mit oder ohne Zurücklegen gezogen wird. □

2.2 Die Poisson-Verteilung

Wir betrachten die Anzahl der Versicherungsfälle (z. B. Schadensmeldungen bei einer Versicherung) in einem Zeitintervall $[0,t]$; als Ergebnisraum wählen wir also $\Omega = \mathbb{N}_0$. Zur Herleitung eines Modells zerlegen wir $[0,t]$ in n gleich lange Teilintervalle und nehmen an, dass in jedem Teilintervall (unabhängig von den anderen) mit Wahrscheinlichkeit

$$p_n = \alpha \frac{t}{n}$$

ein Schaden passiert; dabei ist $\alpha > 0$ ein freier Parameter (die „Schadensrate“). Die Anzahl der Teilintervalle mit Schaden ist dann binomialverteilt. Für große n sollte dies die Anzahl der Schäden gut beschreiben, und wir setzen

$$\mathbb{P}(\{k\}) := \lim_{n\to\infty} \mathcal{B}_{n,p_n}(\{k\}) \qquad \text{(Wahrscheinlichkeit für } k \text{ Schäden)},$$

sofern der Grenzwert existiert. Dies klärt der folgende Satz.

Satz 2.3 (Poisson-Approximation)**.** *Sei $\lambda > 0$ und $(p_n)_{n\geq 1}$ eine Folge in $[0,1]$ mit $np_n \to \lambda$. Dann gilt für alle $k \in \mathbb{N}_0$*

$$\lim_{n\to\infty} \mathcal{B}_{n,p_n}(\{k\}) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} =: \mathrm{Pois}_\lambda(\{k\}),$$

und Pois_λ ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$. Es heißt die Poisson-Verteilung *zum Parameter λ.*

Beweis. Sei $k \in \mathbb{N}_0$ fest. Wir schreiben wieder $a_n \sim b_n$, falls $a_n/b_n \to 1$. Für $n \geq k$ ist

$$\mathcal{B}_{n,p_n}(\{k\}) = \binom{n}{k} p_n^k (1-p_n)^{n-k}.$$

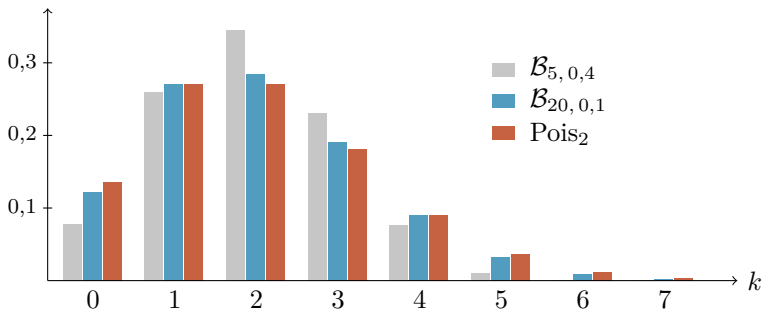
Wegen $\binom{n}{k} = n(n-1)\cdots(n-k+1)/k! \sim n^k/k!$, $np_n \to \lambda$ und $(1-p_n)^{-k} \to 1$ (denn $p_n \to 0$) folgt

$$\mathcal{B}_{n,p_n}(\{k\}) \sim \frac{n^k}{k!}\, p_n^k (1-p_n)^{n-k} \sim \frac{\lambda^k}{k!}(1-p_n)^n = \frac{\lambda^k}{k!}\Big(1 - \frac{np_n}{n}\Big)^n \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda},$$

wobei wir im letzten Schritt $(1 - x_n/n)^n \to \mathrm{e}^{-\lambda}$ für $x_n \to \lambda$ verwendet haben. Ferner ist $\mathrm{Pois}_\lambda(\{k\}) \geq 0$ und

$$\sum_{k\in\mathbb{N}_0} \mathrm{Pois}_\lambda(\{k\}) = \mathrm{e}^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1,$$

d. h. Pois_λ ist eine Zähldichte und definiert nach Satz 1.16(a) ein Wahrscheinlichkeitsmaß. □



Zähldichten von $\mathcal{B}_{n,p_n}$ mit $np_n = 2$ für $n = 5$ und $n = 20$ im Vergleich zur Poisson-Verteilung Pois_2 (exakt berechnet). Schon für $n = 20$ ist der Unterschied gering.

2.3 Die Gamma-Verteilung

Anknüpfend an die Herleitung der Poisson-Verteilung mit $\lambda = \alpha t$ betrachten wir nun den Zeitpunkt der r-ten Schadensmeldung, $r \in \mathbb{N}$, im Intervall $[0, \infty)$. Die r-te Meldung erfolgt genau dann bis zum Zeitpunkt t, wenn in $[0, t]$ mindestens r Schäden auftreten. Die Idee ist daher, die Verteilung $\mathbb{P}$ dieses Zeitpunkts durch

$$\mathbb{P}((0,t]) = 1 - \mathrm{Pois}_{\alpha t}(\{0,\dots,r-1\}) = 1 - \mathrm{e}^{-\alpha t} \sum_{k=0}^{r-1} \frac{(\alpha t)^k}{k!} = \int_0^t \frac{\alpha^r}{(r-1)!}\, x^{r-1} \mathrm{e}^{-\alpha x}\, \mathrm{d}x$$

festzulegen. Die letzte Gleichung sieht man durch Ableiten nach t: Die Ableitung der linken Seite ist eine Teleskopsumme,

$$\alpha \mathrm{e}^{-\alpha t} \sum_{k=0}^{r-1} \frac{(\alpha t)^k}{k!} - \alpha \mathrm{e}^{-\alpha t} \sum_{k=1}^{r-1} \frac{(\alpha t)^{k-1}}{(k-1)!} = \frac{\alpha^r}{(r-1)!}\, t^{r-1} \mathrm{e}^{-\alpha t},$$

und beide Seiten verschwinden für $t = 0$. Die Verteilung hat also eine Dichte, was die folgende Definition motiviert.

Definition 2.4 (Gamma-Verteilung und Exponentialverteilung)**.** Für $\alpha, r > 0$ heißt das Wahrscheinlichkeitsmaß $\Gamma_{\alpha,r}$ auf $([0,\infty), \mathcal{B}_{[0,\infty)})$ mit der Dichtefunktion

$$\gamma_{\alpha,r}(x) = \frac{\alpha^r}{\Gamma(r)}\, x^{r-1} \mathrm{e}^{-\alpha x}, \qquad x > 0$$

(und $\gamma_{\alpha,r}(0) \coloneqq 0$) die *Gamma-Verteilung* mit Ratenparameter (inversem Skalenparameter) α und Formparameter r. Dabei ist

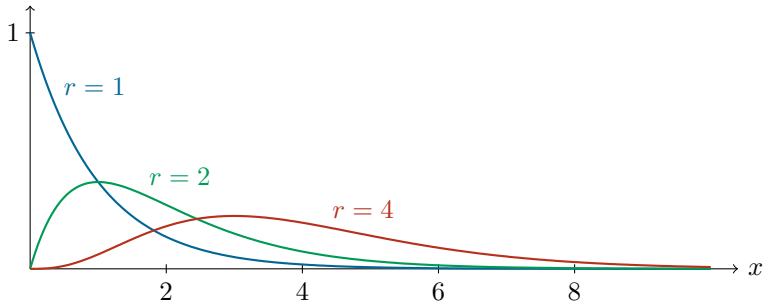
$$\Gamma(r) = \int_0^\infty x^{r-1} \mathrm{e}^{-x}\, \mathrm{d}x$$

die *Gamma-Funktion*; für $r \in \mathbb{N}$ gilt $\Gamma(r) = (r-1)!$. Die Dichte

$$\gamma_{\alpha,1}(x) = \alpha \mathrm{e}^{-\alpha x}, \qquad x > 0,$$

ist die Dichte der *Exponentialverteilung* zum Parameter α.

Mit der Substitution $y = \alpha x$ sieht man $\int_0^\infty \gamma_{\alpha,r}(x)\, \mathrm{d}x = 1$, so dass $\gamma_{\alpha,r}$ tatsächlich eine Dichtefunktion im Sinne von Satz 1.16(b) ist.



Dichten $\gamma_{1,r}$ der Gamma-Verteilung für $r = 1$ (Exponentialverteilung), $r = 2$ und $r = 4$: die Wartezeit auf den r-ten Schaden bei Intensität $\alpha = 1$ verschiebt sich mit r nach rechts und streut stärker.

2.4 Die Beta-Verteilung

In Analogie zur Herleitung der Gamma-Verteilung fragen wir nun nach der Verteilung des Zeitpunkts des r-ten von insgesamt n Schäden im Zeitintervall $(0,1)$, wobei jeder Schadenszeitpunkt unabhängig von den anderen gemäß $\mathcal{U}_{(0,1)}$ verteilt sein soll ($1 \le r \le n$). Seien $T_1, \dots, T_n \in (0,1)$ die (ungeordneten) Schadenszeitpunkte und $T_1' \le \cdots \le T_n'$ dieselben Zeitpunkte in aufsteigender Reihenfolge. Da jede der $n!$ möglichen Anordnungen dieselbe Wahrscheinlichkeit hat, erwarten wir

$$\mathbb{P}(T_r' \le c) = n! \int_0^1 \mathrm{d}t_1 \cdots \int_0^1 \mathrm{d}t_n \, \mathbb{1}\{t_1 < \cdots < t_n\} \, \mathbb{1}\{t_r \le c\}.$$

Halten wir $t_r = s$ fest, so liefert die Integration über die $r-1$ geordneten Zeitpunkte in $(0,s)$ den Faktor $s^{r-1}/(r-1)!$ und die über die $n-r$ geordneten Zeitpunkte in $(s,1)$ den Faktor $(1-s)^{n-r}/(n-r)!$. Also ist

$$\mathbb{P}(T_r' \le c) = \frac{n!}{(r-1)!\,(n-r)!} \int_0^c s^{r-1}(1-s)^{n-r} \, \mathrm{d}s,$$

wobei

$$\frac{n!}{(r-1)!\,(n-r)!} = B(r, n-r+1)^{-1}$$

mit der *Euler'schen Beta-Funktion*

$$B(a,b) \coloneqq \int_0^1 s^{a-1}(1-s)^{b-1} \, \mathrm{d}s, \qquad a,b > 0.$$

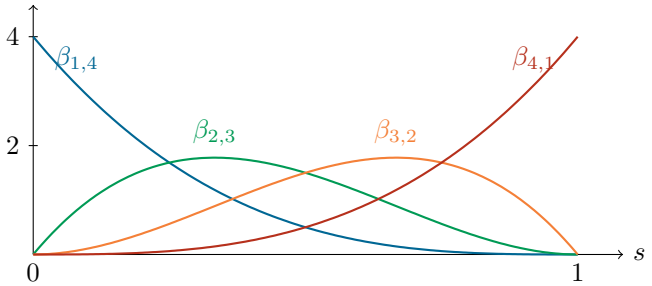
(Dass $B(r, n-r+1)$ diesen Wert hat, folgt gerade aus der Rechnung für $c = 1$.)

Definition 2.5 (Beta-Verteilung)**.** Für $a, b > 0$ heißt das Wahrscheinlichkeitsmaß $\beta_{a,b}$ auf $\big((0,1), \mathcal{B}_{(0,1)}\big)$ mit der Dichtefunktion

$$\beta_{a,b}(s) = B(a,b)^{-1} \, s^{a-1}(1-s)^{b-1}, \qquad 0 < s < 1,$$

die *Beta-Verteilung* zu den Parametern a, b. (Wir bezeichnen hier das Maß und seine Dichte mit demselben Symbol $\beta_{a,b}$; die Dichte wird stets mit Argument geschrieben.)

Nach der obigen Herleitung hat der Zeitpunkt T_r' des r-ten von n Schäden die Verteilung $\beta_{r,n-r+1}$.



Die Zeitpunkte $T_1' \le \cdots \le T_4'$ von $n = 4$ unabhängig und uniform in $(0,1)$ verteilten Schäden: T_r' hat die Dichte $\beta_{r,n-r+1}$, $r = 1, \dots, 4$.

2.5 Die Normalverteilung

Sei $v > 0$. Wir betrachten die N-dimensionale Kugel mit Radius $\sqrt{vN}$, also

$$\Omega_N := \{x \in \mathbb{R}^N : |x|^2 \le vN\},$$

mit der Gleichverteilung $\mathbb{P}_N = \mathcal{U}_{\Omega_N}$ auf Ω_N. Ferner sei

$$\pi_1 : \Omega_N \to \mathbb{R}, \qquad x \mapsto x_1,$$

die Projektion auf die erste Koordinate. Wir untersuchen die Verteilung von π_1 unter $\mathbb{P}_N$ für $N \to \infty$.

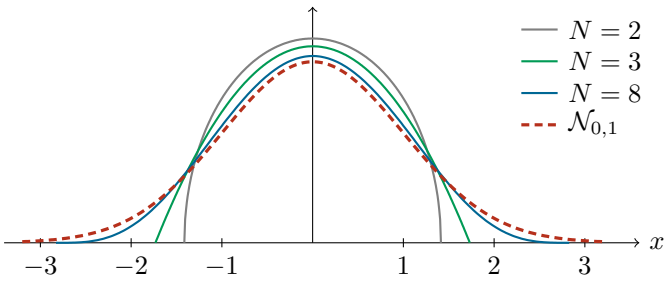
Satz 2.6 (Die Normalverteilung)**.** *Für alle $a < b$ gilt*

$$\lim_{N\to\infty} \mathbb{P}_N(a \le \pi_1 \le b) = \int_a^b \frac{1}{\sqrt{2\pi v}} \mathrm{e}^{-\frac{x^2}{2v}} \,\mathrm{d}x.$$

Für $m \in \mathbb{R}$ und $v > 0$ heißt das Wahrscheinlichkeitsmaß $\mathcal{N}_{m,v}$ auf $(\mathbb{R}, \mathcal{B})$ mit der Dichtefunktion

$$\frac{1}{\sqrt{2\pi v}} \mathrm{e}^{-\frac{(x-m)^2}{2v}}, \qquad x \in \mathbb{R},$$

die Normalverteilung *zu den Parametern m und v. Im Fall $m = 0$, $v = 1$ heißt $\mathcal{N}_{0,1}$* Standardnormalverteilung.



Dichte der ersten Koordinate π_1 eines in der Kugel Ω_N gleichverteilten Punktes ($v = 1$) für $N = 2, 3, 8$ und die Dichte von $\mathcal{N}_{0,1}$ (exakt berechnet). Schon für moderate N ist der Unterschied kaum sichtbar.

Beweis. Sei $N \ge 2$ so groß, dass $vN > a^2 \vee b^2$, wobei $a \vee b := \max\{a, b\}$; dann gilt $[a, b] \subset (-\sqrt{vN}, \sqrt{vN})$. Für $r \ge 0$ bezeichne $B_{N-1}(r)$ die $(N-1)$-dimensionale Kugel mit Radius r um 0; es ist $\lambda^{N-1}(B_{N-1}(r)) = c_{N-1}\, r^{N-1}$ mit $c_{N-1} := \lambda^{N-1}(B_{N-1}(1))$. Integrieren wir zuerst über $x_2, \dots, x_N$ und dann über x_1 (Satz von Fubini, siehe Satz 7.20; ihn sowie die Transformationsformel für Lebesgue-Integrale, hier für Polarkoordinaten, setzen wir aus der Analysis als bekannt voraus), so erhalten wir

$$\begin{aligned}
\mathbb{P}_N(a \le \pi_1 \le b) &= \frac{1}{\lambda^N(\Omega_N)} \int \cdots \int \mathbb{1}\Big\{a \le x_1 \le b,\ \sum_{i=1}^N x_i^2 \le vN\Big\} \,\mathrm{d}x_1 \cdots \mathrm{d}x_N \\
&= \frac{1}{\lambda^N(\Omega_N)} \int_a^b \lambda^{N-1}\Big(B_{N-1}\big(\sqrt{vN - x_1^2}\big)\Big) \,\mathrm{d}x_1 \\
&= \frac{1}{\lambda^N(\Omega_N)} \int_a^b (vN - x_1^2)^{\frac{N-1}{2}} c_{N-1} \,\mathrm{d}x_1.
\end{aligned}$$

Ebenso ist $\lambda^N(\Omega_N) = \int_{-\sqrt{vN}}^{\sqrt{vN}} (vN - x_1^2)^{(N-1)/2} c_{N-1}\,\mathrm{d}x_1$. Kürzen von $c_{N-1}(vN)^{(N-1)/2}$ liefert

$$\mathbb{P}_N(a \le \pi_1 \le b) = \frac{\int_a^b f_N(x)\,\mathrm{d}x}{\int_{-\infty}^{\infty} f_N(x)\,\mathrm{d}x}, \qquad f_N(x) := \Big(1 - \frac{x^2}{vN}\Big)^{\frac{N-1}{2}} \mathbb{1}\{|x| \le \sqrt{vN}\}.$$

Für jedes $x \in \mathbb{R}$ gilt $f_N(x) \to \mathrm{e}^{-x^2/(2v)}$ für $N \to \infty$, und wegen $1-y \le \mathrm{e}^{-y}$ sowie $(N-1)/N \ge 1/2$ ist

$$0 \le f_N(x) \le \Big(\mathrm{e}^{-\frac{x^2}{vN}}\Big)^{\frac{N-1}{2}} \le \mathrm{e}^{-\frac{x^2}{4v}},$$

wobei die rechte Seite integrierbar ist. Mit dem Satz von der majorisierten Konvergenz (Tatsache 1.14(d)) folgt daher

$$\mathbb{P}_N(a \le \pi_1 \le b) \xrightarrow[N\to\infty]{} \frac{\int_a^b \mathrm{e}^{-\frac{x^2}{2v}}\,\mathrm{d}x}{\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2v}}\,\mathrm{d}x}.$$

Den Nenner berechnen wir mit Polarkoordinaten (*Gauß'sches Integral*):

$$\Big[\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2v}}\,\mathrm{d}x\Big]^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2+y^2}{2v}}\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\infty} r\,\mathrm{e}^{-\frac{r^2}{2v}}\,\mathrm{d}r = 2\pi\Big[-v\,\mathrm{e}^{-\frac{r^2}{2v}}\Big]_0^{\infty} = 2\pi v.$$

Also ist der Nenner gleich $\sqrt{2\pi v}$, und die Behauptung folgt. Insbesondere ist nach der Verschiebung $x \mapsto x - m$ auch die Funktion $x \mapsto (2\pi v)^{-1/2}\mathrm{e}^{-(x-m)^2/(2v)}$ eine Dichtefunktion. □

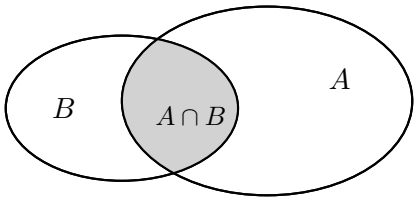
3 Bedingte Wahrscheinlichkeit und Unabhängigkeit

3.1 Bedingte Wahrscheinlichkeit

Wir betrachten zwei Ereignisse A, B, wobei $\mathbb{P}(A) > 0$. Angenommen, wir erfahren, dass A eingetreten ist. Wie sollten wir dann die Wahrscheinlichkeit von B neu bewerten? Wir suchen eine Verteilung $\mathbb{P}_A$ mit den folgenden Eigenschaften:

(a) $\mathbb{P}_A(A) = 1$, d. h. A tritt unter $\mathbb{P}_A$ sicher ein;

(b) es gibt eine Konstante $c_A > 0$ mit $\mathbb{P}_A(B) = c_A\,\mathbb{P}(B)$ für alle Ereignisse $B \subset A$, d. h. innerhalb von A bleiben die Wahrscheinlichkeitsverhältnisse erhalten.



Proposition 3.1 (Bedingte Wahrscheinlichkeit)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{F}$ mit $\mathbb{P}(A) > 0$. Dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}_A$ auf $(\Omega, \mathcal{F})$ mit den Eigenschaften (a) und (b), nämlich*

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \qquad B \in \mathcal{F}.$$

Die Größe $\mathbb{P}(B \mid A) \coloneqq \mathbb{P}_A(B)$ heißt die bedingte Wahrscheinlichkeit von B gegeben A *(oder* bezüglich A*). Im Fall $\mathbb{P}(A) = 0$ setzen wir $\mathbb{P}(B \mid A) \coloneqq 0$ für alle $B \in \mathcal{F}$.*

Beweis von Proposition 3.1. Erfülle $\mathbb{P}_A$ die Eigenschaften (a) und (b). Wegen (a) ist $\mathbb{P}_A(B \setminus A) \le \mathbb{P}_A(A^{\mathrm{c}}) = 0$, und aus (b) mit $B = A$ folgt $1 = \mathbb{P}_A(A) = c_A\mathbb{P}(A)$, also $c_A = 1/\mathbb{P}(A)$. Damit gilt für alle $B \in \mathcal{F}$

$$\mathbb{P}_A(B) = \mathbb{P}_A(A \cap B) + \mathbb{P}_A(B \setminus A) = c_A\,\mathbb{P}(A \cap B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Das zeigt die Eindeutigkeit. Umgekehrt ist $B \mapsto \mathbb{P}(A \cap B)/\mathbb{P}(A)$ ein Wahrscheinlichkeitsmaß (die Normierung folgt aus $\mathbb{P}(A \cap \Omega) = \mathbb{P}(A)$, die σ-Additivität folgt aus der von $\mathbb{P}$), und es erfüllt nach Konstruktion (a) und (b). □

Interpretation. $\mathbb{P}(B \mid A)$ beschreibt die Wahrscheinlichkeit für das Eintreten von B, wenn das Eintreten von A bereits feststeht.

(a) *Frequentistisch:* $\mathbb{P}(B \mid A)$ beschreibt (näherungsweise) den Anteil der Versuche, in denen B eintritt, unter all denjenigen Versuchen, in denen A eintritt.

(b) *Subjektiv:* Ist $\mathbb{P}$ meine Einschätzung der Lage, so ist $\mathbb{P}(\,\cdot \mid A)$ meine Einschätzung, nachdem ich über das Eintreten von A informiert wurde.

Bedingte Wahrscheinlichkeiten drücken keine Kausalität aus: Aus $\mathbb{P}(B \mid A) > \mathbb{P}(B)$ folgt nicht, dass A das Ereignis B „bewirkt".

Satz 3.2 (Fallunterscheidungs- und Bayes-Formel)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\Omega = \bigcup_{i\in I} B_i$ eine Zerlegung von Ω in abzählbar viele paarweise disjunkte Ereignisse $B_i \in \mathcal{F}$. Dann gilt für alle $A \in \mathcal{F}$:*

(a) $\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(B_i)\,\mathbb{P}(A \mid B_i)$ *(Fallunterscheidungsformel)*

(b) *Ist $\mathbb{P}(A) > 0$, so gilt für alle $k \in I$*

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(B_k)\,\mathbb{P}(A \mid B_k)}{\sum_{i\in I} \mathbb{P}(B_i)\,\mathbb{P}(A \mid B_i)}. \qquad \text{(Bayes-Formel)}$$

Beweis. (a) Für alle $i \in I$ gilt nach Definition $\mathbb{P}(B_i)\mathbb{P}(A \mid B_i) = \mathbb{P}(A \cap B_i)$, im Fall $\mathbb{P}(B_i) = 0$ wegen der Konvention in Proposition 3.1. Da A die disjunkte Vereinigung der Mengen $A \cap B_i$, $i \in I$, ist, folgt (a) aus der σ-Additivität.

(b) folgt direkt aus (a), denn $\mathbb{P}(B_k \mid A) = \mathbb{P}(A \cap B_k)/\mathbb{P}(A) = \mathbb{P}(B_k)\mathbb{P}(A \mid B_k)/\mathbb{P}(A)$. □

Beispiel 3.3 (Corona-Test)**.** Wir nehmen an:

(1) 4 % der Bevölkerung sind aktuell an Covid erkrankt (*Prävalenz*).

(2) Der Test spricht bei 90 % der Erkrankten an (*Sensitivität* 90 %).

(3) Der Test spricht bei 20 % der Gesunden an, d. h. er fällt bei 80 % der Gesunden negativ aus (*Spezifität* 80 %).

Mit welcher Wahrscheinlichkeit ist eine zufällig getestete Person

(a) tatsächlich krank, wenn sie einen positiven Test hat?

(b) tatsächlich krank, obwohl ihr Test negativ ist?

Zur Modellierung sei

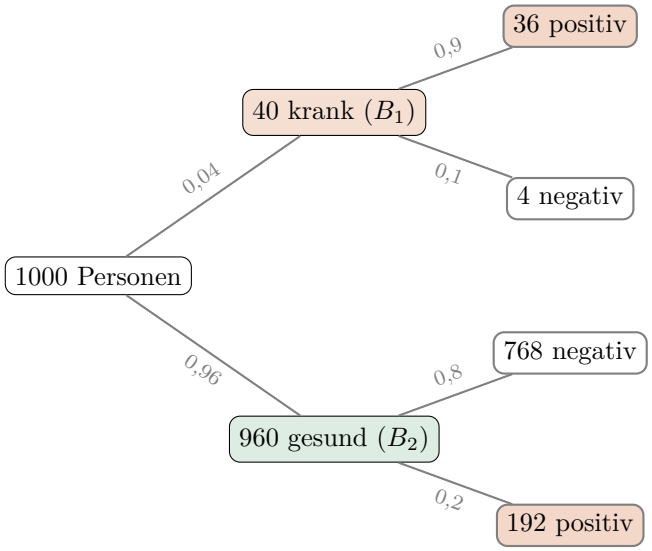
(1) $(\Omega, \mathcal{P}(\Omega), \mathcal{U}_\Omega)$, wobei Ω die (endliche) Bevölkerung ist und $\mathcal{U}_\Omega$ die Gleichverteilung (zufällige Auswahl einer Person),

(2) $B_1 \subset \Omega$ die Menge der Kranken und $B_2 = \Omega \setminus B_1 = B_1^{\mathrm{c}}$ die Menge der Gesunden,

(3) $A \subset \Omega$ die Menge der Personen mit positivem Test.

Nach Annahme ist

$$\mathbb{P}(B_1) = 0{,}04, \quad \mathbb{P}(B_2) = 0{,}96, \quad \mathbb{P}(A \mid B_1) = 0{,}9, \quad \mathbb{P}(A \mid B_2) = 0{,}2.$$



Das zweistufige Modell als Baum, mit den erwarteten Anzahlen unter 1000 getesteten Personen. Von den $36 + 192 = 228$ positiv Getesteten sind nur 36 krank, also $\mathbb{P}(B_1 \mid A) = 36/228 = 3/19 \approx 0{,}16$: Die vielen falsch-positiven Tests der großen Gruppe der Gesunden überwiegen.

Mit der Bayes-Formel (Satz 3.2(b)) folgt:

(a) $\mathbb{P}(B_1 \mid A) = (0{,}04 \cdot 0{,}9)/(0{,}04 \cdot 0{,}9 + 0{,}96 \cdot 0{,}2) = 0{,}036/0{,}228 = 3/19 \approx 0{,}16$. Der *positive Vorhersagewert* (die „positive Korrektheit“) des Tests ist also gering.

(b) Mit $\mathbb{P}(A^c \mid B_1) = 0{,}1$ und $\mathbb{P}(A^c \mid B_2) = 0{,}8$ ist

$$\mathbb{P}(B_1 \mid A^c) = \frac{0{,}04 \cdot 0{,}1}{0{,}04 \cdot 0{,}1 + 0{,}96 \cdot 0{,}8} = \frac{0{,}004}{0{,}772} \approx 0{,}005.$$

Der *negative Vorhersagewert* $\mathbb{P}(B_2 \mid A^c) \approx 0{,}995$ ist also hoch.

Der Test ist somit gut geeignet, eine vorliegende Erkrankung auszuschließen, aber wenig aussagekräftig, um sie nachzuweisen.

3.2 Mehrstufige Modelle

Viele Zufallsexperimente laufen in mehreren Stufen ab, wobei die Wahrscheinlichkeiten in jeder Stufe vom bisherigen Verlauf abhängen. Ein solches Experiment lässt sich durch die bedingten Wahrscheinlichkeiten in den einzelnen Stufen beschreiben.

Satz 3.4 (Konstruktion von Wahrscheinlichkeitsmaßen durch bedingte Wahrscheinlichkeiten)**.** *Seien $\Omega_1, \dots, \Omega_n \neq \emptyset$, $n \geq 2$, abzählbare Ergebnisräume. Sei ρ_1 eine Zähldichte auf Ω_1, und für jedes $k = 2, \dots, n$ und jedes $(\omega_1, \dots, \omega_{k-1}) \in \Omega_1 \times \cdots \times \Omega_{k-1}$ sei $\rho_{k;\omega_1 \dots \omega_{k-1}}$ eine Zähldichte auf Ω_k. Sei $\Omega = \prod_{i=1}^n \Omega_i$ und $X_i \colon \Omega \to \Omega_i$ die i-te Projektion. Dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{P}(\Omega))$ mit*

(a) $\mathbb{P}(X_1 = \omega_1) = \rho_1(\omega_1)$ *für alle* $\omega_1 \in \Omega_1$,

(b) $\mathbb{P}(X_k = \omega_k \mid X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1}) = \rho_{k;\omega_1 \dots \omega_{k-1}}(\omega_k)$ *für alle* $k = 2, \dots, n$ *und* $\omega_i \in \Omega_i$, *falls* $\mathbb{P}(X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1}) > 0$.

$\mathbb{P}$ *ist gegeben durch*

$$\mathbb{P}(\{\omega\}) = \rho_1(\omega_1)\,\rho_{2;\omega_1}(\omega_2)\,\rho_{3;\omega_1\omega_2}(\omega_3)\cdots\rho_{n;\omega_1\ldots\omega_{n-1}}(\omega_n) \qquad \textit{für alle } \omega = (\omega_1,\ldots,\omega_n) \in \Omega. \tag{$*$}$$

Beweis. Eindeutigkeit. Da Ω abzählbar ist, ist $\mathbb{P}$ nach Satz 1.16(a) durch die Werte $\mathbb{P}(\{\omega\})$ festgelegt. Sei $\omega \in \Omega$ und $A_i := \{X_i = \omega_i\}$, also $\{\omega\} = \bigcap_{i=1}^n \{X_i = \omega_i\} = \bigcap_{i=1}^n A_i$. Für jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ gilt die *Multiplikationsformel* („Treppenprodukt“)

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\cdots\mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}),$$

sofern $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$; dies sieht man durch Einsetzen der Definition, da sich das Produkt teleskopartig kürzt. Erfüllt $\mathbb{P}$ (a) und (b), so ergibt sich daraus ($*$). Ist dagegen $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) = 0$, so ist $\mathbb{P}(\{\omega\}) = 0$; sei dann k minimal mit $\mathbb{P}(A_1 \cap \cdots \cap A_k) = 0$. Nach (a) bzw. (b) gilt $0 = \mathbb{P}(A_1 \cap \cdots \cap A_k) = \mathbb{P}(A_1 \cap \cdots \cap A_{k-1})\,\rho_{k;\omega_1\ldots\omega_{k-1}}(\omega_k)$ (bzw. $0 = \rho_1(\omega_1)$ für $k = 1$), also verschwindet auch die rechte Seite von ($*$). In jedem Fall gilt also ($*$).

Existenz. Wir definieren $\mathbb{P}$ durch ($*$). Für $1 \le k \le n$ und $\omega_1,\ldots,\omega_k$ gilt dann

$$\begin{aligned}
\mathbb{P}(X_1 = \omega_1,\ldots,X_k = \omega_k) &= \sum_{\omega_{k+1}\in\Omega_{k+1},\ldots,\omega_n\in\Omega_n} \mathbb{P}(\{(\omega_1,\ldots,\omega_n)\}) \\
&= \rho_1(\omega_1)\rho_{2;\omega_1}(\omega_2)\cdots\rho_{k;\omega_1\ldots\omega_{k-1}}(\omega_k) \\
&\quad\cdot \sum_{\omega_{k+1}\in\Omega_{k+1}} \rho_{k+1;\omega_1\ldots\omega_k}(\omega_{k+1})\cdots \underbrace{\sum_{\omega_n\in\Omega_n} \rho_{n;\omega_1\ldots\omega_{n-1}}(\omega_n)}_{=1} \\
&= \rho_1(\omega_1)\rho_{2;\omega_1}(\omega_2)\cdots\rho_{k;\omega_1\ldots\omega_{k-1}}(\omega_k),
\end{aligned}$$

indem man die Summen von innen nach außen auswertet. Für $k = 1$ folgt (a), und Summation über ω_1 ergibt $\mathbb{P}(\Omega) = 1$, d. h. ($*$) ist eine Zähldichte. Außerdem folgt

$$\mathbb{P}(X_1 = \omega_1,\ldots,X_k = \omega_k) = \mathbb{P}(X_1 = \omega_1,\ldots,X_{k-1} = \omega_{k-1})\,\rho_{k;\omega_1\ldots\omega_{k-1}}(\omega_k),$$

und nach Division durch $\mathbb{P}(X_1 = \omega_1,\ldots,X_{k-1} = \omega_{k-1}) > 0$ erhalten wir (b). □

Beispiel 3.5 (Populationsgenetik: Gesetz von Hardy–Weinberg)**.** Wir betrachten ein Gen mit zwei Allelen A und a, also die Genotypen AA, Aa, aa, die von zwei Eltern zufällig an ein Kind weitergegeben werden: Jeder Elternteil gibt jeweils eines seiner beiden Allele mit Wahrscheinlichkeit $1/2$ weiter. Für die Genotypen der Eltern gilt

$$\omega_1 \in \{AA, Aa, aa\}, \qquad \omega_2 \in \{AA, Aa, aa\},$$

und für das Kind ist

$$\rho_{3;\omega_1\omega_2}(AA) = \begin{cases} 1, & \text{für } \omega_1 = \omega_2 = AA, \\ \frac{1}{2}, & \text{für } \omega_1 = AA,\ \omega_2 = Aa \ \text{ oder } \ \omega_1 = Aa,\ \omega_2 = AA, \\ \frac{1}{4}, & \text{für } \omega_1 = \omega_2 = Aa, \\ 0, & \text{sonst;} \end{cases}$$

analog (durch Vertauschen von A und a) ist $\rho_{3;\omega_1\omega_2}(aa)$ gegeben, und $\rho_{3;\omega_1\omega_2}(Aa) = 1 - \rho_{3;\omega_1\omega_2}(AA) - \rho_{3;\omega_1\omega_2}(aa)$. Die Eltern werden unabhängig voneinander aus der Population gewählt, in der die Genotypen AA, Aa, aa mit den Anteilen u, $2v$, w auftreten:

$$\rho_1 = \rho_{2;\omega_1} = (u, 2v, w) \qquad \text{mit } u + 2v + w = 1.$$

Sei $\mathbb{P}$ das zugehörige Wahrscheinlichkeitsmaß aus Satz 3.4 und X_3 der Genotyp des Kindes. Mit der Fallunterscheidungsformel erhalten wir für die Anteile $u_1, 2v_1, w_1$ in der Kindergeneration

$$\begin{aligned}
u_1 &= \mathbb{P}(X_3 = AA) = u^2 + 2 \cdot u \cdot 2v \cdot \frac{1}{2} + (2v)^2 \cdot \frac{1}{4} = u^2 + 2uv + v^2 = (u+v)^2, \\
w_1 &= \mathbb{P}(X_3 = aa) = (w+v)^2 \qquad \text{(aus Symmetriegründen)}, \\
2v_1 &= \mathbb{P}(X_3 = Aa) = 1 - u_1 - w_1 = (u + 2v + w)^2 - (u+v)^2 - (w+v)^2 = 2(u+v)(w+v).
\end{aligned}$$

Wie sieht es in der nächsten Generation aus? Wenden wir dieselbe Rechnung auf $(u_1, 2v_1, w_1)$ an, so folgt

$$u_2 = (u_1 + v_1)^2 = \big((u+v)^2 + (u+v)(w+v)\big)^2 = (u+v)^2 \underbrace{\big((u+v) + (w+v)\big)^2}_{=1} = (u+v)^2 = u_1,$$

und ebenso $w_2 = w_1$ und $v_2 = v_1$. Ab der ersten Nachkommengeneration ändert sich die Verteilung der Genotypen also nicht mehr!

Wir wollen nun auch unendlich viele Stufen zulassen. Dann ist der Produktraum im Allgemeinen überabzählbar, und wir müssen mit der Produkt-σ-Algebra arbeiten.

Satz 3.6 (Wahrscheinlichkeitsmaße auf unendlichen Produkträumen)**.** *Für jedes $i \in \mathbb{N}$ sei $\Omega_i \neq \emptyset$ eine abzählbare Menge, und ρ_1 sei eine Zähldichte auf Ω_1. Ferner sei für alle $k \geq 2$ und $\omega_i \in \Omega_i$, $i < k$, $\rho_{k;\omega_1 \dots \omega_{k-1}}$ eine Zähldichte auf Ω_k. Sei $\Omega = \prod_{i \geq 1} \Omega_i$, $X_i \colon \Omega \to \Omega_i$ die Projektion auf die i-te Koordinate und*

$$\mathcal{F} = \bigotimes_{i \geq 1} \mathcal{P}(\Omega_i)$$

die Produkt-σ-Algebra (Beispiel und Definition 1.8). Dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{F})$ mit

$$\mathbb{P}(X_1 = \omega_1, \dots, X_k = \omega_k) = \rho_1(\omega_1)\, \rho_{2;\omega_1}(\omega_2) \cdots \rho_{k;\omega_1 \dots \omega_{k-1}}(\omega_k)$$

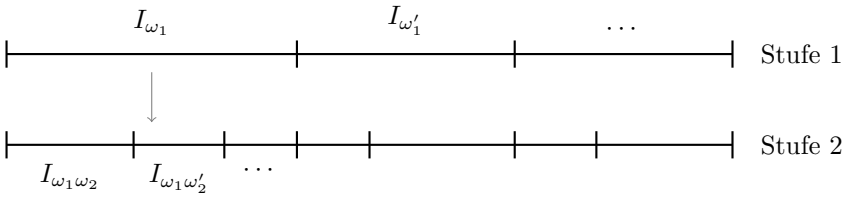
für alle $k \geq 1$ und $\omega_i \in \Omega_i$.

Beweis. Eindeutigkeit. Sie folgt aus dem Eindeutigkeitssatz 1.12, da

$$\mathcal{G} = \{\{X_1 = \omega_1, \dots, X_k = \omega_k\} \colon k \geq 1,\ \omega_i \in \Omega_i\} \cup \{\emptyset\}$$

ein $\cap$-stabiler Erzeuger von $\mathcal{F}$ ist. In der Tat ist der Durchschnitt zweier solcher „Zylindermengen“ entweder leer oder gleich der Zylindermenge mit der längeren Bedingung. Ferner gilt $\mathcal{G} \subset \mathcal{F}$, und jeder Erzeuger $\{X_i \in A_i\}$, $A_i \subset \Omega_i$, von $\mathcal{F}$ ist die abzählbare Vereinigung der Mengen $\{X_1 = \omega_1, \dots, X_i = \omega_i\}$ mit $\omega_1 \in \Omega_1, \dots, \omega_{i-1} \in \Omega_{i-1}$, $\omega_i \in A_i$; also ist $\sigma(\mathcal{G}) = \mathcal{F}$.

Existenz. Wir führen die Existenz auf die des Lebesgue-Maßes zurück. Betrachte den Wahrscheinlichkeitsraum $([0,1), \mathcal{B}_{[0,1)}, \lambda)$ mit $\lambda = \mathcal{U}_{[0,1)}$. Sei $(I_{\omega_1})_{\omega_1 \in \Omega_1}$ eine Zerlegung von $[0,1)$ in (eventuell leere) halboffene, aneinandergrenzende Intervalle der Länge $\rho_1(\omega_1)$. Weiter zerlegen wir jedes Intervall I_{ω_1} in halboffene Intervalle $(I_{\omega_1\omega_2})_{\omega_2 \in \Omega_2}$ der Länge $\rho_1(\omega_1)\rho_{2;\omega_1}(\omega_2)$, jedes $I_{\omega_1\omega_2}$ in Intervalle $(I_{\omega_1\omega_2\omega_3})_{\omega_3 \in \Omega_3}$ der Länge $\rho_1(\omega_1)\rho_{2;\omega_1}(\omega_2)\rho_{3;\omega_1\omega_2}(\omega_3)$ usw. Das ist möglich, da sich die Längen jeweils zur Länge des zerlegten Intervalls aufsummieren.



Für jedes $x \in [0,1)$ und jedes $k \geq 1$ gibt es dann genau ein Intervall der k-ten Stufe, das x enthält. Also gibt es genau eine Folge $Z(x) = (Z_1(x), Z_2(x), \dots) \in \Omega$ mit

$$x \in I_{Z_1(x)\dots Z_k(x)} \qquad \text{für alle } k \geq 1.$$

Die Abbildung $Z\colon [0,1) \to \Omega$ ist eine Zufallsvariable von $([0,1), \mathcal{B}_{[0,1)})$ nach $(\Omega, \mathcal{F})$: Für jedes $A = \{X_1 = \omega_1, \dots, X_k = \omega_k\} \in \mathcal{G} \setminus \{\emptyset\}$ ist

$$Z^{-1}(A) = \{x \in [0,1)\colon Z_1(x) = \omega_1, \dots, Z_k(x) = \omega_k\} = I_{\omega_1\dots\omega_k} \in \mathcal{B}_{[0,1)},$$

und die Behauptung folgt mit Bemerkung 1.20(2). Nach Satz 1.22 ist daher $\mathbb{P} := \lambda \circ Z^{-1}$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$, und es gilt

$$\mathbb{P}(X_1 = \omega_1, \dots, X_k = \omega_k) = \lambda(I_{\omega_1\dots\omega_k}) = \rho_1(\omega_1)\, \rho_{2;\omega_1}(\omega_2) \cdots \rho_{k;\omega_1\dots\omega_{k-1}}(\omega_k),$$

d. h. $\mathbb{P}$ hat die gewünschte Eigenschaft. □

3.3 Unabhängigkeit

Intuitiv sollten zwei Ereignisse A und B unabhängig heißen, wenn die Information über das Eintreten des einen die Wahrscheinlichkeit des anderen nicht verändert, d. h. wenn

$$\mathbb{P}(A \,|\, B) = \mathbb{P}(A) \qquad \text{und} \qquad \mathbb{P}(B \,|\, A) = \mathbb{P}(B), \qquad \text{falls } \mathbb{P}(A), \mathbb{P}(B) > 0.$$

Beide Gleichungen sind äquivalent zu $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$; diese symmetrische Formulierung ist auch im Fall $\mathbb{P}(A) = 0$ oder $\mathbb{P}(B) = 0$ sinnvoll.

Definition 3.7 (Unabhängigkeit)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \in \mathcal{F}$ heißen (*stochastisch*) *unabhängig* bezüglich $\mathbb{P}$, falls

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\, \mathbb{P}(B).$$

Beispiel 3.8. (1) Eine Urne enthalte $s \geq 1$ schwarze und $w \geq 1$ weiße Kugeln. Wir ziehen zweimal *mit* Zurücklegen (Modell: Gleichverteilung auf Ω_{I} mit $N = s + w$). Seien $A =$ {erster Zug weiß} und $B =$ {zweiter Zug weiß}. Dann gilt

$$\mathbb{P}(A \cap B) = \frac{w^2}{(s+w)^2} = \frac{w}{s+w} \cdot \frac{w}{s+w} = \mathbb{P}(A)\, \mathbb{P}(B),$$

d. h. A und B sind unabhängig.

(2) Ziehen wir dagegen *ohne* Zurücklegen (Modell: Gleichverteilung $\mathbb{P}'$ auf Ω_{II}), so ist weiterhin $\mathbb{P}'(A) = w/(s+w)$ und auch

$$\mathbb{P}'(B) = \frac{w(w-1) + s\,w}{(s+w)(s+w-1)} = \frac{w}{s+w},$$

aber

$$\mathbb{P}'(A \cap B) = \frac{w}{s+w} \cdot \frac{w-1}{s+w-1} < \frac{w}{s+w} \cdot \frac{w}{s+w} = \mathbb{P}'(A)\,\mathbb{P}'(B),$$

denn $(w-1)(s+w) < w(s+w-1)$ ist äquivalent zu $s > 0$. Also sind A und B hier *nicht* unabhängig.

(3) Zwei faire Würfel werden geworfen, also $\Omega = \{1,\dots,6\}^2$ mit der Gleichverteilung $\mathbb{P} = \mathcal{U}_\Omega$. Seien

$$A = \{\text{Augensumme ist } 7\} = \{(k,\ell) \in \Omega \colon k+\ell = 7\},$$
$$B = \{\text{erster Würfel zeigt } 6\} = \{(k,\ell) \in \Omega \colon k = 6\}.$$

Dann ist $|A| = |B| = 6$ und $|A \cap B| = 1$, also

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A)\,\mathbb{P}(B).$$

A und B sind also stochastisch unabhängig, obwohl die Augensumme vom ersten Wurf abhängt: Stochastische Unabhängigkeit bedeutet *keine* kausale Unabhängigkeit!

Wir übertragen den Begriff der Unabhängigkeit nun auf beliebig viele Ereignisse.

Definition 3.9 (Unabhängigkeit beliebig vieler Ereignisse)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $I \neq \emptyset$ eine beliebige Indexmenge. Eine Familie $(A_i)_{i\in I}$ von Ereignissen in $\mathcal{F}$ heißt *unabhängig* bezüglich $\mathbb{P}$, falls für alle $J \subset I$ mit $J \neq \emptyset$ und $|J| < \infty$ gilt

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

Es genügt im Allgemeinen nicht, diese Produktformel nur für Paare von Ereignissen zu fordern:

Beispiel 3.10 (Unabhängigkeit versus paarweise Unabhängigkeit)**.** Wir betrachten den zweifachen fairen Münzwurf, also $\Omega = \{0,1\}^2$ mit der Gleichverteilung $\mathbb{P} = \mathcal{U}_\Omega$, und die Ereignisse

$$A = \{1\} \times \{0,1\}, \qquad B = \{0,1\} \times \{1\}, \qquad C = \{(0,0),(1,1)\}$$

(erster Wurf 1, zweiter Wurf 1, beide Würfe gleich). Dann ist $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$ und

$$\mathbb{P}(A \cap B) = \tfrac{1}{4} = \mathbb{P}(A)\mathbb{P}(B), \qquad \mathbb{P}(A \cap C) = \tfrac{1}{4} = \mathbb{P}(A)\mathbb{P}(C), \qquad \mathbb{P}(B \cap C) = \tfrac{1}{4} = \mathbb{P}(B)\mathbb{P}(C),$$

d. h. A, B, C sind paarweise unabhängig. Aber wegen $A \cap B \cap C = \{(1,1)\}$ gilt

$$\mathbb{P}(A \cap B \cap C) = \tfrac{1}{4} \neq \tfrac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C),$$

also ist die Familie (A, B, C) nicht unabhängig.

Definition 3.11 (Unabhängigkeit von Zufallsvariablen)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $I \neq \emptyset$ eine beliebige Indexmenge, und für jedes $i \in I$ sei $Y_i \colon \Omega \to \Omega_i$ eine Zufallsvariable auf $(\Omega, \mathcal{F})$ mit Werten in einem Ereignisraum $(\Omega_i, \mathcal{F}_i)$. Die Familie $(Y_i)_{i\in I}$ heißt *unabhängig* bezüglich $\mathbb{P}$, falls für jede Wahl von Mengen $B_i \in \mathcal{F}_i$, $i \in I$, die Familie von Ereignissen $\big(\{Y_i \in B_i\}\big)_{i\in I}$ unabhängig bezüglich $\mathbb{P}$ ist, also

$$\mathbb{P}\Big(\bigcap_{i\in J} \{Y_i \in B_i\}\Big) = \prod_{i\in J} \mathbb{P}(Y_i \in B_i) \qquad \text{für alle } \emptyset \neq J \subset I \text{ mit } |J| < \infty. \tag{$**$}$$

Wie beim Eindeutigkeitssatz genügt es, die Produktformel auf $\cap$-stabilen Erzeugern nachzuprüfen.

Satz 3.12 (Charakterisierung der Unabhängigkeit)**.** *In der Situation von Definition 3.11 sei $\mathcal{F}_i = \sigma(\mathcal{G}_i)$ mit einem $\cap$-stabilen Mengensystem $\mathcal{G}_i \subset \mathcal{P}(\Omega_i)$, $i \in I$. Dann gilt:*

$$(Y_i)_{i\in I} \text{ ist unabhängig} \quad \Longleftrightarrow \quad \big(\{Y_i \in B_i\}\big)_{i\in I} \text{ ist unabhängig für alle } B_i \in \mathcal{G}_i,\ i \in I.$$

Beweis. „$\Rightarrow$“ gilt, da $\mathcal{G}_i \subset \mathcal{F}_i$.

„$\Leftarrow$“: Wir zeigen durch Induktion über $n \in \mathbb{N}_0$: $(**)$ gilt für alle endlichen $\emptyset \neq J \subset I$ und alle $B_i \in \mathcal{F}_i$, $i \in J$, mit

$$\big|\{i \in J \colon B_i \notin \mathcal{G}_i\}\big| \leq n.$$

Induktionsanfang $n = 0$: Dann gilt $B_i \in \mathcal{G}_i$ für alle $i \in J$, und $(**)$ gilt nach Voraussetzung.

Induktionsschritt $n \to n+1$: Sei $J \subset I$ endlich und seien $B_i \in \mathcal{F}_i$, $i \in J$, gegeben mit $\big|\{i \in J \colon B_i \notin \mathcal{G}_i\}\big| = n+1$. Ist $|J| = 1$, so besteht das Produkt in $(**)$ aus einem Faktor, und es ist nichts zu zeigen; sei also $|J| \geq 2$. Wähle ein $j \in J$ mit $B_j \notin \mathcal{G}_j$ und setze $J' = J \setminus \{j\}$ sowie

$$A = \bigcap_{i\in J'} \{Y_i \in B_i\}.$$

Unter den B_i, $i \in J'$, liegen höchstens n nicht in $\mathcal{G}_i$, so dass nach Induktionsannahme

$$\mathbb{P}(A) = \prod_{i\in J'} \mathbb{P}(Y_i \in B_i)$$

gilt. Ist $\mathbb{P}(A) = 0$, so sind beide Seiten von $(**)$ gleich 0, und es ist nichts zu zeigen. Sei also $\mathbb{P}(A) > 0$. Wir betrachten die beiden Wahrscheinlichkeitsmaße

$$B' \mapsto \mathbb{P}(Y_j \in B' \,|\, A) \qquad \text{und} \qquad B' \mapsto \mathbb{P}(Y_j \in B')$$

auf $(\Omega_j, \mathcal{F}_j)$ (Bilder von $\mathbb{P}_A$ bzw. $\mathbb{P}$ unter Y_j, vgl. Satz 1.22). Für $B' \in \mathcal{G}_j$ liegen unter den Mengen B_i, $i \in J'$, und B' höchstens n nicht im jeweiligen Erzeuger, also gilt nach Induktionsannahme

$$\mathbb{P}(A \cap \{Y_j \in B'\}) = \mathbb{P}(A)\,\mathbb{P}(Y_j \in B'), \quad \text{d. h.} \quad \mathbb{P}(Y_j \in B' \,|\, A) = \mathbb{P}(Y_j \in B').$$

Die beiden Maße stimmen also auf dem $\cap$-stabilen Erzeuger $\mathcal{G}_j$ von $\mathcal{F}_j$ überein und nach dem Eindeutigkeitssatz 1.12 daher auf ganz $\mathcal{F}_j$. Insbesondere folgt für $B' = B_j$

$$\mathbb{P}\Big(\bigcap_{i\in J} \{Y_i \in B_i\}\Big) = \mathbb{P}(A)\,\mathbb{P}(Y_j \in B_j \,|\, A) = \mathbb{P}(A)\,\mathbb{P}(Y_j \in B_j) = \prod_{i\in J} \mathbb{P}(Y_i \in B_i),$$

also $(**)$. □

Korollar 3.13 (Unabhängigkeit von Indikatorfunktionen)**.** *Eine Familie $(A_i)_{i\in I}$ von Ereignissen ist genau dann unabhängig, wenn die Familie $(\mathbb{1}_{A_i})_{i\in I}$ der Indikatorfunktionen unabhängig ist. Insbesondere gilt: Ist $(A_i)_{i\in I}$ unabhängig und $C_i \in \{A_i, A_i^c, \Omega, \emptyset\}$ für alle $i \in I$, so ist auch $(C_i)_{i\in I}$ unabhängig.*

Beweis. Jede Indikatorfunktion $\mathbb{1}_A$, $A \in \mathcal{F}$, ist eine Zufallsvariable mit Werten in $\{0,1\}$ (versehen mit $\mathcal{P}(\{0,1\})$), und $\mathcal{P}(\{0,1\})$ hat den $\cap$-stabilen Erzeuger $\mathcal{G} = \{\{1\}\}$ mit $\{\mathbb{1}_A \in \{1\}\} = \{\mathbb{1}_A = 1\} = A$. Nach Satz 3.12 ist $(\mathbb{1}_{A_i})_{i\in I}$ also genau dann unabhängig, wenn $(A_i)_{i\in I}$ unabhängig ist.

Ist nun $(A_i)_{i\in I}$ unabhängig, so ist $\big(\{\mathbb{1}_{A_i} \in B_i\}\big)_{i\in I}$ für alle $B_i \subset \{0,1\}$ unabhängig. Wegen $\{\mathbb{1}_{A_i} \in B_i\} \in \{A_i, A_i^c, \Omega, \emptyset\}$, wobei jede dieser vier Mengen bei geeigneter Wahl von B_i auftritt, folgt die zweite Aussage. □

Korollar 3.14 (Unabhängigkeit endlich vieler Zufallsvariablen)**.** *Sei $(Y_i)_{1\le i\le n}$ eine Familie von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt:*

(1) Diskreter Fall: *Hat jedes Y_i einen abzählbaren Wertebereich Ω_i (mit $\mathcal{F}_i = \mathcal{P}(\Omega_i)$), so ist $(Y_i)_{1\le i\le n}$ genau dann unabhängig, wenn*

$$\mathbb{P}(Y_1 = \omega_1, \dots, Y_n = \omega_n) = \prod_{i=1}^n \mathbb{P}(Y_i = \omega_i) \qquad \text{für alle } \omega_i \in \Omega_i.$$

(2) Reeller Fall: *Ist jedes Y_i reellwertig, so ist $(Y_i)_{1\le i\le n}$ genau dann unabhängig, wenn*

$$\mathbb{P}(Y_1 \le c_1, \dots, Y_n \le c_n) = \prod_{i=1}^n \mathbb{P}(Y_i \le c_i) \qquad \text{für alle } c_i \in \mathbb{R}.$$

Beweis. In beiden Fällen ist „⇒“ ein Spezialfall der Definition der Unabhängigkeit, angewandt auf die Ereignisse $\{Y_i = \omega_i\}$ bzw. $\{Y_i \le c_i\}$. Für „⇐“ wenden wir Satz 3.12 an. Dabei ist zu beachten, dass die rechten Seiten nur die Produktformel für $J = \{1, \dots, n\}$ liefern; wir nehmen daher jeweils die ganze Menge Ω_i in den Erzeuger auf und setzen in der Produktformel $B_i = \Omega_i$ für $i \notin J$.

(1) $\mathcal{G}_i = \{\{\omega_i\} \colon \omega_i \in \Omega_i\} \cup \{\emptyset, \Omega_i\}$ ist ein $\cap$-stabiler Erzeuger von $\mathcal{F}_i = \mathcal{P}(\Omega_i)$. Für $B_i = \{\omega_i\}$ bzw. $B_i = \emptyset$ $(i = 1, \dots, n)$ ist die Produktformel die Voraussetzung (bzw. beide Seiten sind 0). Ist $B_i = \Omega_i$ für i aus einer Teilmenge der Indizes, so summiert man die Voraussetzung über die entsprechenden $\omega_i \in \Omega_i$ und erhält wegen der σ-Additivität wieder die Produktformel.

(2) Ebenso ist $\mathcal{G}_i = \{(-\infty, c] \colon c \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$ ein $\cap$-stabiler Erzeuger der Borel'schen σ-Algebra. Die Produktformel für $B_i = \mathbb{R}$ erhält man aus der Voraussetzung, indem man die entsprechenden $c_i \uparrow \infty$ laufen lässt und die σ-Stetigkeit (Satz 1.11(e)) verwendet. □

Beispiel 3.15 (Produktmaße)**.** Sei E eine endliche Menge, ρ eine Zähldichte auf E und $n \ge 2$. Sei $\mathbb{P} = \rho^{\otimes n}$ das *n-fache Produktmaß* auf $\Omega = E^n$, d. h. das Wahrscheinlichkeitsmaß mit der Zähldichte

$$\rho^{\otimes n}(\omega) = \prod_{i=1}^n \rho(\omega_i), \qquad \omega = (\omega_1, \dots, \omega_n) \in E^n$$

(vgl. die geordnete Stichprobe mit Zurücklegen). Sei $X_i\colon \Omega \to E$ die Projektion auf die i-te Koordinate. Dann ist

$$\mathbb{P}(X_1 = \omega_1, \dots, X_n = \omega_n) = \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \prod_{i=1}^{n} \rho(\omega_i),$$

und durch Summation über alle ω_j, $j \neq i$, folgt $\mathbb{P}(X_i = \omega_i) = \rho(\omega_i)$. Nach Korollar 3.14(1) sind die X_i, $1 \leq i \leq n$, also unabhängig bezüglich $\mathbb{P}$.

Satz 3.16 (Kombination unabhängiger Zufallsvariablen)**.** *Sei $(Y_i)_{i\in I}$ eine unabhängige Familie von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in den Ereignisräumen $(\Omega_i, \mathcal{F}_i)$, $i \in I$. Sei $(I_k)_{k\in K}$ eine Familie paarweise disjunkter Teilmengen von I, und für $k \in K$ sei $(\tilde{\Omega}_k, \tilde{\mathcal{F}}_k)$ ein Ereignisraum und*

$$\tilde{Y}_k := \varphi_k((Y_i)_{i\in I_k})$$

für Zufallsvariablen $\varphi_k\colon (\prod_{i\in I_k} \Omega_i, \bigotimes_{i\in I_k} \mathcal{F}_i) \to (\tilde{\Omega}_k, \tilde{\mathcal{F}}_k)$. Dann ist die Familie $(\tilde{Y}_k)_{k\in K}$ unabhängig.

Beweis. Übung. □

3.4 Existenz unabhängiger Zufallsvariablen, Produktmaße

Satz 3.17 (Konstruktion unabhängiger Zufallsvariablen)**.** *Sei I eine abzählbare Indexmenge, und für jedes $i \in I$ sei $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ ein Wahrscheinlichkeitsraum. Dann existieren ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und unabhängige Zufallsvariablen $Y_i\colon \Omega \to \Omega_i$ mit $\mathbb{P} \circ Y_i^{-1} = \mathbb{P}_i$ für alle $i \in I$.*

Beweis. Es genügt, den Fall $I = \mathbb{N}$ zu betrachten (ist I endlich, so ergänzt man die Familie etwa durch triviale Wahrscheinlichkeitsräume $(\{0\}, \mathcal{P}(\{0\}), \delta_0)$). Wir beweisen die Aussage nur unter zusätzlichen Voraussetzungen an die Ω_i.

1. Fall: Alle Ω_i, $i \in \mathbb{N}$, sind abzählbar (mit $\mathcal{F}_i = \mathcal{P}(\Omega_i)$). Dann folgt die Aussage aus Satz 3.6 mit

$$\rho_1(\omega_1) = \mathbb{P}_1(\{\omega_1\}), \qquad \rho_{k;\omega_1\dots\omega_{k-1}}(\omega_k) = \mathbb{P}_k(\{\omega_k\}),$$

$\Omega = \prod_{i\geq 1} \Omega_i$, $\mathcal{F} = \bigotimes_{i\geq 1} \mathcal{P}(\Omega_i)$ und $Y_i = X_i$ (Projektionen): Es gilt $\mathbb{P}(X_1 = \omega_1, \dots, X_k = \omega_k) = \prod_{i=1}^{k} \mathbb{P}_i(\{\omega_i\})$ für alle k; durch Summation folgt $\mathbb{P}(X_i = \omega_i) = \mathbb{P}_i(\{\omega_i\})$, also $\mathbb{P} \circ X_i^{-1} = \mathbb{P}_i$, und nach Korollar 3.14(1) sind $X_1, \dots, X_k$ für jedes k unabhängig. Da die Unabhängigkeit nach Definition 3.11 nur endliche Teilfamilien betrifft, ist damit $(X_i)_{i\geq 1}$ unabhängig.

2. Fall: $\Omega_i = [0, 1]$ und $\mathbb{P}_i = \mathcal{U}_{[0,1]}$ für alle $i \in \mathbb{N}$. Nach Fall 1 (angewandt auf die abzählbare Indexmenge $\mathbb{N}\times\mathbb{N}$) gibt es auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ unabhängige $\{0, 1\}$-wertige Zufallsvariablen $(Y_{ij})_{i,j\geq 1}$ mit $\mathbb{P}(Y_{ij} = 1) = \mathbb{P}(Y_{ij} = 0) = 1/2$. Setze

$$Y_i := \sum_{j\geq 1} Y_{ij}\, 2^{-j} = \varphi((Y_{ij})_{j\geq 1}), \qquad \varphi\colon \{0, 1\}^{\mathbb{N}} \to [0, 1], \quad \varphi(y) = \sum_{j\geq 1} y_j 2^{-j},$$

d. h. Y_i ist die Zahl mit der Binärentwicklung $(Y_{ij})_{j\geq 1}$.

φ ist eine Zufallsvariable bezüglich $\mathcal{P}(\{0, 1\})^{\otimes\mathbb{N}}$ und $\mathcal{B}_{[0,1]}$. Seien $X_k\colon \{0, 1\}^{\mathbb{N}} \to \{0, 1\}$ die Projektionen. Für $n \geq 1$ und $0 \leq m < 1$ mit Binärdarstellung $m = \sum_{k=1}^{n} y_k 2^{-k}$, $y_k \in \{0, 1\}$, gilt

$$\varphi^{-1}([m, m + 2^{-n}]) = \{X_1 = y_1, \dots, X_n = y_n\} \cup R,$$

wobei R aus höchstens zwei Folgen besteht (den Binärentwicklungen von m bzw. $m+2^{-n}$, die nicht mit $y_1, \dots, y_n$ beginnen; jede Zahl hat höchstens zwei Binärentwicklungen). Einpunktige Mengen $\{y\} = \bigcap_{n\geq 1}\{X_1 = y_1, \dots, X_n = y_n\}$ liegen in $\mathcal{P}(\{0,1\})^{\otimes\mathbb{N}}$, also auch $\varphi^{-1}([m, m+2^{-n}])$. Da die dyadischen Intervalle $[m, m+2^{-n}]$ die σ-Algebra $\mathcal{B}_{[0,1]}$ erzeugen, ist φ nach Bemerkung 1.20(2) messbar. Damit ist auch $Y_i = \varphi \circ (Y_{ij})_{j\geq 1}$ eine Zufallsvariable.

Verteilung von Y_i. Jede feste Folge $y \in \{0,1\}^{\mathbb{N}}$ hat die Wahrscheinlichkeit $\mathbb{P}((Y_{ij})_{j\geq 1} = y) \leq \mathbb{P}(Y_{i1} = y_1, \dots, Y_{in} = y_n) = 2^{-n}$ für alle n, also 0. Mit der obigen Zerlegung folgt

$$\mathbb{P}(Y_i \in [m, m+2^{-n}]) = \mathbb{P}(Y_{i1} = y_1, \dots, Y_{in} = y_n) = 2^{-n}.$$

Insbesondere gilt $\mathbb{P}(Y_i = m) = 0$ für alle i und alle dyadischen Zahlen m, denn $\{Y_i = m\} = \{(Y_{ij})_{j\geq 1} \in \varphi^{-1}(\{m\})\}$, und $\varphi^{-1}(\{m\})$ besteht aus höchstens zwei Folgen, die jeweils Wahrscheinlichkeit 0 haben. Daher ist

$$\mathbb{P}(Y_i \in (m, m+2^{-n}]) = 2^{-n} = \mathcal{U}_{[0,1]}((m, m+2^{-n}]).$$

Die halboffenen dyadischen Intervalle $(m, m+2^{-n}]$ bilden zusammen mit $\emptyset$ und $[0,1]$ einen $\cap$-stabilen Erzeuger von $\mathcal{B}_{[0,1]}$ (zwei solche Intervalle sind entweder disjunkt oder ineinander enthalten). Nach dem Eindeutigkeitssatz 1.12 folgt $\mathbb{P} \circ Y_i^{-1} = \mathcal{U}_{[0,1]}$. Schließlich impliziert Satz 3.16 (mit den disjunkten Indexmengen $I_i = \{i\} \times \mathbb{N}$) die Unabhängigkeit der Y_i, $i \geq 1$, als „Zusammenfassungen“ von unabhängigen Zufallsvariablen.

3. Fall: $(\Omega_i, \mathcal{F}_i) = (\mathbb{R}, \mathcal{B})$ für alle $i \in \mathbb{N}$. Nach Fall 2 gibt es unabhängige Zufallsvariablen $(Y_i)_{i\geq 1}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathbb{P} \circ Y_i^{-1} = \mathcal{U}_{[0,1]}$. Wegen $\mathcal{U}_{[0,1]}((0,1)) = 1$ gilt $\mathbb{P}(0 < Y_i < 1 \text{ für alle } i \geq 1) = 1$. Mit der Quantil-Transformation (Proposition 1.25, angewandt auf $F = F_{\mathbb{P}_i}$) gibt es zu jedem $\mathbb{P}_i$ auf $(\mathbb{R}, \mathcal{B})$ eine Zufallsvariable $\varphi_i \colon (0,1) \to \mathbb{R}$ mit Verteilung $\mathbb{P}_i = \mathcal{U}_{(0,1)} \circ \varphi_i^{-1}$; wir setzen φ_i beliebig messbar auf $[0,1]$ fort (z. B. $\varphi_i(0) = \varphi_i(1) = 0$), was wegen $\mathbb{P}(Y_i \in \{0,1\}) = 0$ die Verteilung von $\varphi_i \circ Y_i$ nicht ändert. Dann ist

$$\mathbb{P} \circ (\varphi_i \circ Y_i)^{-1} = \mathcal{U}_{[0,1]} \circ \varphi_i^{-1} = \mathbb{P}_i,$$

und wieder nach Satz 3.16 hat die Familie $(\varphi_i \circ Y_i)_{i\geq 1}$ die gewünschten Eigenschaften.

4. Fall: allgemeine Ereignisräume. Dieser Fall ist maßtheoretisch aufwendiger; siehe z. B. [Kle20] (ohne Beweis). □

Korollar 3.18 (Existenz von Produktmaßen)**.** *Sei $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)_{i\in I}$ eine abzählbare Familie von Wahrscheinlichkeitsräumen und*

$$\Omega = \prod_{i\in I} \Omega_i, \qquad \mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i.$$

Dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{F})$, bezüglich dessen die Projektionen $X_i \colon \Omega \to \Omega_i$ unabhängig sind mit Verteilung $\mathbb{P}_i$, d. h.

$$\mathbb{P}(X_i \in A_i \text{ für alle } i \in J) = \prod_{i\in J} \mathbb{P}_i(A_i) \qquad \text{für alle } \emptyset \neq J \subset I \text{ mit } |J| < \infty \text{ und } A_i \in \mathcal{F}_i.$$

Das Maß $\mathbb{P}$ heißt das Produkt *der $\mathbb{P}_i$ und wird mit $\bigotimes_{i\in I} \mathbb{P}_i$ bezeichnet, bzw. mit $Q^{\otimes I}$, falls $\mathbb{P}_i = Q$ für alle $i \in I$.*

Beweis. Eindeutigkeit. Sie folgt aus dem Eindeutigkeitssatz 1.12, da die Mengen

$$\{X_i \in A_i \text{ für alle } i \in J\}$$

mit $\emptyset \neq J \subset I$ endlich und $A_i \in \mathcal{F}_i$ einen $\cap$-stabilen Erzeuger von $\mathcal{F}$ bilden.

Existenz. Seien $(Y_i)_{i\in I}$ gemäß Satz 3.17 unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega', \mathcal{F}', \mathbb{P}')$ mit $\mathbb{P}' \circ Y_i^{-1} = \mathbb{P}_i$. Dann ist $Y = (Y_i)_{i\in I} \colon \Omega' \to \Omega$ eine Zufallsvariable, denn für alle endlichen $\emptyset \neq J \subset I$ und $A_i \in \mathcal{F}_i$ gilt

$$Y^{-1}\{X_i \in A_i \text{ für alle } i \in J\} = \{Y_i \in A_i \text{ für alle } i \in J\} \in \mathcal{F}'.$$

Also ist $\mathbb{P} := \mathbb{P}' \circ Y^{-1}$ wohldefiniert, und

$$\mathbb{P}(X_i \in A_i \text{ für alle } i \in J) = \mathbb{P}'(Y_i \in A_i \text{ für alle } i \in J) = \prod_{i\in J} \mathbb{P}'(Y_i \in A_i) = \prod_{i\in J} \mathbb{P}_i(A_i). \qquad \square$$

Bemerkung 3.19 (Gemeinsame Verteilung unabhängiger Zufallsvariablen). Für eine abzählbare Familie $(Y_i)_{i\in I}$ von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(\Omega_i, \mathcal{F}_i)$ gilt

$$(Y_i)_{i\in I} \text{ unabhängig bzgl. } \mathbb{P} \quad \Longleftrightarrow \quad \mathbb{P} \circ \big((Y_i)_{i\in I}\big)^{-1} = \bigotimes_{i\in I} \mathbb{P} \circ Y_i^{-1}.$$

Die gemeinsame Verteilung unabhängiger Zufallsvariablen ist also das Produkt ihrer Einzelverteilungen. Denn für $\emptyset \neq J \subset I$ endlich und $A_i \in \mathcal{F}_i$ ist

$$\mathbb{P} \circ \big((Y_i)_{i\in I}\big)^{-1}\big(\{\omega \colon \omega_i \in A_i \ \forall i \in J\}\big) = \mathbb{P}(Y_i \in A_i \ \forall i \in J),$$

und die Unabhängigkeit bedeutet gerade, dass dies gleich $\prod_{i\in J} \mathbb{P}(Y_i \in A_i)$ ist. Das ist genau die charakterisierende Eigenschaft des Produktmaßes aus Korollar 3.18; da die Mengen dieser Form einen $\cap$-stabilen Erzeuger der Produkt-σ-Algebra bilden, folgt die Äquivalenz aus dem Eindeutigkeitssatz 1.12.

Beispiel 3.20. (1) *Unabhängige Wiederholung.* Die unabhängige Wiederholung desselben Experiments $(\Omega, \mathcal{F}, \mathbb{P})$ wird realisiert durch das Experiment

$$(\Omega^{\mathbb{N}}, \ \mathcal{F}^{\otimes\mathbb{N}}, \ \mathbb{P}^{\otimes\mathbb{N}})$$

mit dem Produktraum $\Omega^{\mathbb{N}}$, der Produkt-σ-Algebra $\mathcal{F}^{\otimes\mathbb{N}}$ und dem Produktmaß $\mathbb{P}^{\otimes\mathbb{N}}$.

(2) *Bernoulli-Prozess.* Im Spezialfall $\Omega = \{0,1\}$, $\mathbb{P}(\{1\}) = p \in [0,1]$, heißt $\mathbb{P}^{\otimes\mathbb{N}}$ das (unendliche) *Bernoulli-Maß* mit Erfolgswahrscheinlichkeit p. Eine Folge $(Y_i)_{i\in\mathbb{N}}$ von $\{0,1\}$-wertigen Zufallsvariablen mit gemeinsamer Verteilung $\mathbb{P}^{\otimes\mathbb{N}}$ heißt *Bernoulli-Prozess* (oder *Bernoulli-Folge*) zu p. Entsprechend heißt eine Familie $(Y_i)_{1\leq i\leq n}$ von $\{0,1\}$-wertigen Zufallsvariablen mit gemeinsamer Verteilung $\mathbb{P}^{\otimes n}$ eine *endliche Bernoulli-Folge* (der Länge n) zu p.

(3) *Produkte von Dichten.* Seien $\Omega_i = \mathbb{R}$ und $\mathbb{P}_i$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ mit Dichtefunktion ρ_i, $1 \leq i \leq n$. Dann ist $\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i$ das Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit der Dichte

$$\rho(x) = \prod_{i=1}^n \rho_i(x_i), \qquad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Bezeichnet nämlich X_i die i-te Projektion, so gilt für alle $c_1, \dots, c_n \in \mathbb{R}$

$$\begin{aligned}\mathbb{P}(X_1 \le c_1, \dots, X_n \le c_n) &= \prod_{i=1}^n \mathbb{P}_i((-\infty, c_i]) = \prod_{i=1}^n \int_{-\infty}^{c_i} \rho_i(x_i)\,\mathrm{d}x_i \\ &= \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_n} \rho_1(x_1) \cdots \rho_n(x_n)\,\mathrm{d}x_n \cdots \mathrm{d}x_1 \\ &= \int \rho(x)\, \mathbb{1}\{x \in (-\infty, c_1] \times \cdots \times (-\infty, c_n]\}\,\mathrm{d}x\end{aligned}$$

(Satz von Fubini für nichtnegative Funktionen). Da die Mengen $(-\infty, c_1] \times \cdots \times (-\infty, c_n]$ einen $\cap$-stabilen Erzeuger von $\mathcal{B}^n$ bilden, folgt mit dem Eindeutigkeitssatz 1.12

$$\mathbb{P}(A) = \int_A \rho(x)\,\mathrm{d}x \qquad \text{für alle } A \in \mathcal{B}^n.$$

Achtung: Unendliche Produktmaße (etwa auf $\mathbb{R}^{\mathbb{N}}$) haben keine Dichte!

Wir betrachten nun die Summe zweier unabhängiger reellwertiger Zufallsvariablen Y_1, Y_2 mit Verteilungen Q_1 bzw. Q_2. Nach Bemerkung 3.19 hat (Y_1, Y_2) die Verteilung $Q_1 \otimes Q_2$ auf $\mathbb{R}^2$. Mit der (stetigen, also messbaren) Abbildung

$$A\colon \mathbb{R}^2 \to \mathbb{R}, \qquad (x_1, x_2) \mapsto x_1 + x_2,$$

ist $Y_1 + Y_2 = A(Y_1, Y_2)$, und $Y_1 + Y_2$ hat folglich die Verteilung $(Q_1 \otimes Q_2) \circ A^{-1}$.

Definition 3.21 (Faltung)**.** Seien Q_1, Q_2 zwei Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$. Dann heißt das Wahrscheinlichkeitsmaß

$$Q_1 * Q_2 := (Q_1 \otimes Q_2) \circ A^{-1}$$

auf $(\mathbb{R}, \mathcal{B})$ die *Faltung* von Q_1 und Q_2.

Bemerkung 3.22 (Faltung von Dichten). (a) *Diskreter Fall:* Sind Q_1 und Q_2 auf $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$ konzentriert mit Zähldichten ρ_1 und ρ_2, so ist auch $Q_1 * Q_2$ auf $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$ konzentriert und hat die Zähldichte

$$\rho_1 * \rho_2(k) := \sum_{l \in \mathbb{Z}} \rho_1(l)\, \rho_2(k - l), \qquad k \in \mathbb{Z}.$$

(b) *Stetiger Fall:* Haben Q_1 und Q_2 Dichtefunktionen ρ_1 und ρ_2, so hat $Q_1 * Q_2$ die Dichtefunktion

$$\rho_1 * \rho_2(x) := \int \rho_1(y)\, \rho_2(x - y)\,\mathrm{d}y, \qquad x \in \mathbb{R}.$$

Beweis. (a) Da $Q_1 \otimes Q_2$ auf $\mathbb{Z}^2$ konzentriert ist, ist $(Q_1 \otimes Q_2) \circ A^{-1}$ auf $\mathbb{Z}$ konzentriert, und für $k \in \mathbb{Z}$ gilt

$$Q_1 \otimes Q_2(\{A = k\}) = \sum_{l_1, l_2 \in \mathbb{Z}} \rho_1(l_1)\rho_2(l_2)\, \mathbb{1}\{l_1 + l_2 = k\} = \sum_{l \in \mathbb{Z}} \rho_1(l)\rho_2(k - l) = \rho_1 * \rho_2(k).$$

(b) Für $c \in \mathbb{R}$ gilt nach Beispiel 3.20(3) und mit der Substitution $x = z + y$ im inneren Integral

$$\begin{aligned}Q_1 \otimes Q_2(\{A \le c\}) &= \int \mathrm{d}y\, \rho_1(y) \int \mathrm{d}z\, \rho_2(z)\, \mathbb{1}\{z + y \le c\} \\ &= \int \mathrm{d}y\, \rho_1(y) \int \mathrm{d}x\, \rho_2(x - y)\, \mathbb{1}\{x \le c\} \\ &= \int_{-\infty}^{c} \rho_1 * \rho_2(x)\,\mathrm{d}x,\end{aligned}$$

wobei wir im letzten Schritt die Integrationsreihenfolge vertauscht haben (Fubini). Mit $c \to \infty$ folgt insbesondere $\int \rho_1 * \rho_2(x)\,\mathrm{d}x = 1$, d. h. $\rho_1 * \rho_2$ ist eine Dichtefunktion. Da $\{(-\infty, c] : c \in \mathbb{R}\}$ ein $\cap$-stabiler Erzeuger von $\mathcal{B}$ ist, stimmen $Q_1 * Q_2$ und das Wahrscheinlichkeitsmaß mit Dichte $\rho_1 * \rho_2$ nach dem Eindeutigkeitssatz 1.12 überein. □

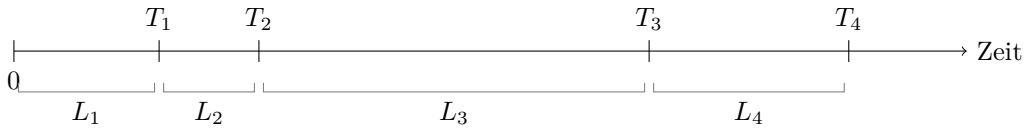
Beispiel 3.23 (Anwendung der Faltung)**.** Für die Normalverteilungen $\mathcal{N}_{m_1,v_1}$ und $\mathcal{N}_{m_2,v_2}$ gilt

$$\mathcal{N}_{m_1,v_1} * \mathcal{N}_{m_2,v_2} = \mathcal{N}_{m_1+m_2,\, v_1+v_2}.$$

Dies rechnet man mit Bemerkung 3.22(b) nach (Übung). Die Summe zweier unabhängiger normalverteilter Zufallsvariablen ist also wieder normalverteilt.

3.5 Der Poisson-Prozess

Wir betrachten eine Folge $(L_i)_{i\geq 1}$ von unabhängigen, exponentialverteilten Zufallsvariablen mit Parameter $\alpha > 0$ (d. h. mit Verteilung $\Gamma_{\alpha,1}$) auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$; ein solcher existiert nach Satz 3.17. Wir deuten L_i als zufällige Wartezeit zwischen dem $(i-1)$-ten und dem i-ten Eintreten eines Ereignisses (etwa eines Schadensfalls bei einer Versicherung).



Sei

$$T_k := \sum_{i=1}^{k} L_i \quad (k \geq 0,\ T_0 = 0)$$

der k-te Zeitpunkt und

$$N_t := \sum_{k\geq 1} \mathbb{1}\{T_k \leq t\}, \qquad t \geq 0,$$

die Anzahl der Zeitpunkte bis zur Zeit t.

Satz 3.24 (Der Poisson-Prozess)**.** *Die N_t sind Zufallsvariablen, und für $0 = t_0 < t_1 < \cdots < t_n$ sind die Zuwächse $N_{t_i} - N_{t_{i-1}}$, $1 \leq i \leq n$, unabhängig und Poisson-verteilt mit Parameter $\alpha(t_i - t_{i-1})$. Die Familie $(N_t)_{t\geq 0}$ heißt* Poisson-Prozess mit Intensität $\alpha > 0$.

Beweis. Als Summen von Zufallsvariablen sind die T_k Zufallsvariablen. Wegen

$$\{N_t \geq k\} = \{T_k \leq t\} \in \mathcal{F} \qquad \text{bzw.} \qquad \{N_t = k\} = \{T_k \leq t < T_{k+1}\} \in \mathcal{F}$$

ist daher auch N_t für jedes $t \geq 0$ eine Zufallsvariable (mit Werten in $\mathbb{N}_0 \cup \{\infty\}$; dass $N_t < \infty$ fast sicher gilt, ergibt sich unten).

Wir führen den Rest des Beweises für $n = 2$ aus, der allgemeine Fall geht analog. Sei also $t_1 = s$, $t_2 = t$ mit $0 < s < t$, und seien $k, l \in \mathbb{N}_0$. Nach Bemerkung 3.19 und Beispiel 3.20(3)

hat $(L_1, \dots, L_m)$ für jedes m die Dichte $\prod_{i=1}^m \alpha \mathrm{e}^{-\alpha x_i}$ auf $(0,\infty)^m$. Sei zunächst $l \geq 1$ und $m := k + l + 1$. Dann gilt

$$\begin{aligned}\mathbb{P}(N_s = k,\ N_t - N_s = l) &= \mathbb{P}\big(T_k \leq s < T_{k+1} \leq T_{k+l} \leq t < T_{k+l+1}\big) \\ &= \int_0^\infty \cdots \int_0^\infty \mathrm{d}x_1 \cdots \mathrm{d}x_m\, \alpha^m \mathrm{e}^{-\alpha \sum_{i=1}^m x_i} \\ &\qquad \cdot \mathbb{1}\Big\{\sum_{i=1}^k x_i \leq s < \sum_{i=1}^{k+1} x_i,\ \sum_{i=1}^{k+l} x_i \leq t < \sum_{i=1}^m x_i\Big\}.\end{aligned}$$

Wir integrieren zuerst über x_m. Für festes $u := \sum_{i=1}^{k+l} x_i \leq t$ ist

$$\int_0^\infty \mathrm{d}x_m\, \alpha \mathrm{e}^{-\alpha x_m}\, \mathbb{1}\{t < u + x_m\} = \mathrm{e}^{-\alpha(t-u)},$$

und zusammen mit dem Faktor $\mathrm{e}^{-\alpha u}$ ergibt sich $\mathrm{e}^{-\alpha t}$. Also ist

$$\begin{aligned}\mathbb{P}(N_s = k,\ N_t - N_s = l) = \mathrm{e}^{-\alpha t}\alpha^{k+l} \int \mathrm{d}x_1 \cdots \mathrm{d}x_k\, \mathbb{1}\Big\{\sum_{i=1}^k x_i \leq s\Big\} \\ \cdot \Big(\int \mathrm{d}x_{k+1} \cdots \mathrm{d}x_{k+l}\, \mathbb{1}\Big\{s < \sum_{i=1}^{k+1} x_i,\ \sum_{i=1}^{k+l} x_i \leq t\Big\}\Big),\end{aligned}$$

wobei alle Integrale über $(0,\infty)$ laufen. Das innere Integral (in der Klammer) hängt zunächst über $\sum_{i=1}^k x_i$ noch von den äußeren Variablen ab. Im inneren Integral substituieren wir $y_1 = \sum_{i=1}^{k+1} x_i - s$ und $y_j = x_{k+j}$ für $2 \leq j \leq l$; es wird zu

$$\int_0^\infty \cdots \int_0^\infty \mathrm{d}y_1 \cdots \mathrm{d}y_l\, \mathbb{1}\{y_1 + \cdots + y_l \leq t - s\} = \frac{(t-s)^l}{l!},$$

dem Volumen eines l-dimensionalen Simplex der Kantenlänge $t - s$ (Induktion über l); insbesondere hängt es nicht mehr von $x_1, \dots, x_k$ ab. Ebenso ist das äußere Integral gleich $s^k/k!$. Insgesamt erhalten wir

$$\begin{aligned}\mathbb{P}(N_s = k,\ N_t - N_s = l) &= \mathrm{e}^{-\alpha t}\alpha^{k+l}\frac{s^k}{k!}\frac{(t-s)^l}{l!} = \Big(\mathrm{e}^{-\alpha s}\frac{(\alpha s)^k}{k!}\Big)\Big(\mathrm{e}^{-\alpha(t-s)}\frac{(\alpha(t-s))^l}{l!}\Big) \\ &= \mathrm{Pois}_{\alpha s}(\{k\})\,\mathrm{Pois}_{\alpha(t-s)}(\{l\}).\end{aligned}$$

Für $l = 0$ ist $\{N_s = k,\ N_t - N_s = 0\} = \{T_k \leq s,\ T_{k+1} > t\}$, und dieselbe Rechnung (ohne das innere Integral) liefert das analoge Ergebnis.

Die rechte Seite ist das Produkt der Zähldichten zweier Poisson-Verteilungen. Summation über alle $k, l \in \mathbb{N}_0$ ergibt 1; insbesondere sind N_s und $N_t - N_s$ fast sicher endlich. Summation über l bzw. k zeigt, dass N_s die Verteilung $\mathrm{Pois}_{\alpha s}$ und $N_t - N_s$ die Verteilung $\mathrm{Pois}_{\alpha(t-s)}$ hat, und da die gemeinsame Zähldichte das Produkt dieser Zähldichten ist, sind $N_s = N_s - N_0$ und $N_t - N_s$ nach Korollar 3.14 (diskreter Fall) unabhängig. □

Korollar 3.25 (Faltung der Poisson- und Gammaverteilung)**.** *Für alle $\lambda, \mu > 0$ gilt*

$$\mathrm{Pois}_\lambda * \mathrm{Pois}_\mu = \mathrm{Pois}_{\lambda+\mu},$$

und für alle $\alpha > 0$ und $r, s \in \mathbb{N}$ gilt

$$\Gamma_{\alpha,r} * \Gamma_{\alpha,s} = \Gamma_{\alpha,r+s}.$$

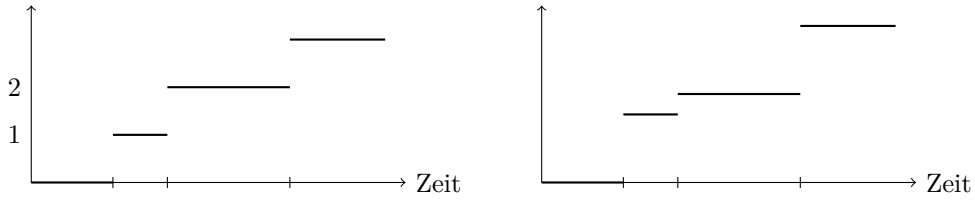
Ohne Beweis. (Beide Aussagen lassen sich mit Bemerkung 3.22 direkt nachrechnen.)

Beispiel 3.26 (Erweiterungen des Poisson-Prozesses)**.** Wir skizzieren zwei Erweiterungen des Poisson-Prozesses.

(1) *Zusammengesetzter Poisson-Prozess.* Sei $(N_t)_{t\geq 0}$ ein Poisson-Prozess und $(Z_i)_{i\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter, reellwertiger Zufallsvariablen mit $Z_i \geq 0$, die auch von $(N_t)_{t\geq 0}$ unabhängig sind. Dann heißt

$$S_t \coloneqq \sum_{i=1}^{N_t} Z_i, \qquad t \geq 0,$$

ein *zusammengesetzter Poisson-Prozess* (engl. *compound Poisson process*). Die Z_i beschreiben die Sprunghöhen, etwa die Schadenshöhen bei einer Versicherung.



Poisson-Prozess N_t zusammengesetzter Poisson-Prozess S_t

Bezeichnet $c > 0$ die konstante Prämienrate, so ist

$$V_t \coloneqq S_t - ct$$

der Verlust der Versicherung bis zur Zeit t. Für $a > 0$ ist die *Ruinwahrscheinlichkeit*

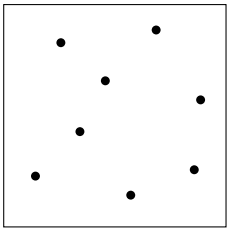
$$r(a) \coloneqq \mathbb{P}\Big(\sup_{t\geq 0} V_t > a\Big) = \mathbb{P}\Big(\sup_{k\geq 1} \sum_{i=1}^{k} (Z_i - cL_i) > a\Big);$$

hierbei haben wir benutzt, dass V_t zwischen zwei Sprungzeiten fällt, also $\sup_{t\geq 0} V_t = \max\{V_0, \sup_{k\geq 1} V_{T_k}\}$ mit $V_0 = 0 < a$ gilt, und dass $V_{T_k} = \sum_{i=1}^{k} Z_i - cT_k = \sum_{i=1}^{k}(Z_i - cL_i)$.

(2) *Poisson'scher Punktprozess* (räumliche Erweiterung des Poisson-Prozesses) auf $\Lambda = [0, L]^2$. Sei $(N_t)_{t\geq 0}$ ein Poisson-Prozess mit Intensität αL und $(Z_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen, auf $[0, L]$ gleichverteilten Zufallsvariablen, die auch von $(N_t)_{t\geq 0}$ unabhängig ist. Dann heißt die zufällige Punktmenge

$$X^\alpha \coloneqq \{(T_k, Z_k) \colon 1 \leq k \leq N_L\} \subset \Lambda$$

der *Poisson'sche Punktprozess* auf Λ mit Intensität α. Die Anzahl $\#X^\alpha = N_L$ der Punkte in Λ ist $\mathrm{Pois}_{\alpha L^2}$-verteilt, im Mittel liegen also α Punkte pro Flächeneinheit in Λ (die Poisson-Verteilung Pois_λ hat den Erwartungswert λ, vgl. später Beispiel 4.23(2)).



$\Lambda = [0, L]^2$

3.6 Asymptotische Ereignisse

Definition 3.27 (Asymptotische Ereignisse)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(Y_k)_{k\geq 1}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{F})$ mit Werten in $(\Omega_k, \mathcal{F}_k)$. Ein Ereignis $A \in \mathcal{F}$ heißt *asymptotisch* für $(Y_k)_{k\geq 1}$, falls für alle $n \geq 1$ gilt: A hängt nur von $(Y_k)_{k\geq n}$ ab, d. h. es gibt ein

$$B_n \in \bigotimes_{k\geq n} \mathcal{F}_k \qquad \text{mit} \qquad A = \{(Y_k)_{k\geq n} \in B_n\}.$$

Sei $\mathcal{A} = \mathcal{A}(Y_k \colon k \geq 1)$ das System der asymptotischen Ereignisse.

Bemerkung 3.28. $\mathcal{A}$ ist eine Sub-σ-Algebra von $\mathcal{F}$ und heißt die *asymptotische σ-Algebra* der Folge $(Y_k)_{k\geq 1}$. In der Tat ist $\Omega = \{(Y_k)_{k\geq n} \in \prod_{k\geq n} \Omega_k\}$, mit $A = \{(Y_k)_{k\geq n} \in B_n\}$ ist $A^{\mathrm{c}} = \{(Y_k)_{k\geq n} \in B_n^{\mathrm{c}}\}$, und für $A^{(1)}, A^{(2)}, \ldots \in \mathcal{A}$ mit zugehörigen Mengen $B_n^{(j)}$ ist $\bigcup_j A^{(j)} = \{(Y_k)_{k\geq n} \in \bigcup_j B_n^{(j)}\}$.

Die σ-Algebra $\mathcal{A}$ enthält durchaus interessante, nicht-triviale Ereignisse, wie die folgenden Beispiele zeigen.

(1) Für $A_k \in \mathcal{F}_k$ ist

$$A = \{Y_k \in A_k \text{ für unendlich viele } k\} = \bigcap_{m\geq 1} \bigcup_{k\geq m} \{Y_k \in A_k\} = \limsup_{k\to\infty} \{Y_k \in A_k\}$$

liegt in $\mathcal{A}(Y_k \colon k \geq 1)$, denn für jedes n gilt auch $A = \bigcap_{m\geq n} \bigcup_{k\geq m} \{Y_k \in A_k\}$, und dieses Ereignis hängt nur von $(Y_k)_{k\geq n}$ ab.

(2) Ist $(\Omega_k, \mathcal{F}_k) = (\mathbb{R}, \mathcal{B})$ für alle k und sind $a \leq b$, so ist

$$A = \Big\{ \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} Y_k \text{ existiert und liegt in } [a,b] \Big\} \in \mathcal{A}(Y_k \colon k \geq 1),$$

denn für jedes feste n gilt $\frac{1}{N}\sum_{k=1}^{n-1} Y_k \to 0$ für $N \to \infty$; die ersten $n-1$ Summanden beeinflussen also weder die Existenz noch den Wert des Grenzwerts. Mit $M_N := \frac{1}{N}\sum_{k=n}^{N} Y_k$, $N \geq n$, ist daher

$$A = \bigcap_{j\geq 1} \bigcup_{N_0\geq n} \bigcap_{N,N'\geq N_0} \Big\{ |M_N - M_{N'}| \leq \tfrac{1}{j},\ a - \tfrac{1}{j} \leq M_N \leq b + \tfrac{1}{j} \Big\}$$

(Cauchy-Kriterium). Jedes Ereignis in den geschweiften Klammern ist von der Form $\{(Y_n, \ldots, Y_{\max\{N,N'\}}) \in C\}$ mit einer Borel-Menge C, da stetige Abbildungen messbar sind; also ist $A = \{(Y_k)_{k\geq n} \in B_n\}$ mit einem $B_n \in \bigotimes_{k\geq n} \mathcal{B}$.

Für unabhängige Zufallsvariablen sind asymptotische Ereignisse in einem sehr starken Sinne deterministisch.

Satz 3.29 (0-1-Gesetz von Kolmogorov)**.** *Sei $(Y_k)_{k\geq 1}$ eine Folge unabhängiger Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(\Omega_k, \mathcal{F}_k)$. Dann gilt für jedes $A \in \mathcal{A}(Y_k \colon k \geq 1)$ entweder $\mathbb{P}(A) = 0$ oder $\mathbb{P}(A) = 1$.*

Beweis. Sei $A \in \mathcal{A}(Y_k \colon k \geq 1)$, und sei $\mathcal{G}$ das Mengensystem bestehend aus den Ereignissen

$$C = \{Y_1 \in C_1, \dots, Y_n \in C_n\}, \qquad n \geq 1,\ C_k \in \mathcal{F}_k.$$

Es ist $\mathcal{G} = \{\{(Y_k)_{k\geq 1} \in D\} \colon D \in \mathcal{Z}\}$, wobei $\mathcal{Z}$ das System der Zylindermengen $C_1 \times \cdots \times C_n \times \prod_{k>n} \Omega_k$ bezeichnet; $\mathcal{Z}$ ist ein $\cap$-stabiler Erzeuger von $\bigotimes_{k\geq 1} \mathcal{F}_k$.

Sei $C = \{Y_1 \in C_1, \dots, Y_n \in C_n\} \in \mathcal{G}$. Da A asymptotisch ist, gilt $A = \{(Y_k)_{k>n} \in B_{n+1}\}$ für ein $B_{n+1} \in \bigotimes_{k>n} \mathcal{F}_k$. Somit ist

$$\mathbb{1}_C = \prod_{k=1}^{n} \mathbb{1}\{Y_k \in C_k\} \ \text{ eine Funktion von } (Y_1, \dots, Y_n)$$

und

$$\mathbb{1}_A = \mathbb{1}\{(Y_k)_{k>n} \in B_{n+1}\} \ \text{ eine Funktion von } (Y_k)_{k>n}.$$

Nach dem Satz über die Kombination unabhängiger Zufallsvariablen (Satz 3.16), angewandt auf die disjunkten Indexmengen $\{1, \dots, n\}$ und $\{n+1, n+2, \dots\}$, sind $\mathbb{1}_C$ und $\mathbb{1}_A$ unabhängig, d. h. $\mathbb{P}(C \cap A) = \mathbb{P}(C)\mathbb{P}(A)$ für alle $C \in \mathcal{G}$.

Also sind die Zufallsvariablen $(Y_k)_{k\geq 1}$ (mit Werten in $\prod_k \Omega_k$) und $\mathbb{1}_A$ unabhängig: Nach dem Satz über die Charakterisierung der Unabhängigkeit durch $\cap$-stabile Erzeuger (Satz 3.12) genügt es, dies für die Ereignisse $\{(Y_k)_{k\geq 1} \in D\}$, $D \in \mathcal{Z}$, und $\{\mathbb{1}_A = 1\} = A$ nachzuprüfen. Insbesondere ist das Ereignis $A = \{(Y_k)_{k\geq 1} \in B_1\}$ unabhängig von $A = \{\mathbb{1}_A = 1\}$, also

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A) \cdot \mathbb{P}(A),$$

und daher $\mathbb{P}(A) \in \{0, 1\}$. □

Satz 3.30 (Lemma von Borel–Cantelli)**.** *Sei $(A_k)_{k\geq 1}$ eine Folge von Ereignissen in $(\Omega, \mathcal{F}, \mathbb{P})$ und*

$$A \coloneqq \{\omega \in \Omega \colon \omega \in A_k \text{ für unendlich viele } k\} = \limsup_{k\to\infty} A_k.$$

(a) *Ist $\sum_{k\geq 1} \mathbb{P}(A_k) < \infty$, so gilt $\mathbb{P}(A) = 0$.*

(b) *Ist $\sum_{k\geq 1} \mathbb{P}(A_k) = \infty$ und ist $(A_k)_{k\geq 1}$ unabhängig, so gilt $\mathbb{P}(A) = 1$.*

Beweis. (a) Für jedes $m \geq 1$ gilt $A \subset \bigcup_{k\geq m} A_k$, also nach Satz 1.11

$$\mathbb{P}(A) \leq \sum_{k\geq m} \mathbb{P}(A_k) \xrightarrow[m\to\infty]{} 0,$$

da $\sum_{k\geq 1} \mathbb{P}(A_k) < \infty$ (Restglieder einer konvergenten Reihe).

(b) Es gilt $A^{\mathrm{c}} = \bigcup_{m\geq 1} \bigcap_{k\geq m} A_k^{\mathrm{c}}$. Mit der σ-Subadditivität und der σ-Stetigkeit (Satz 1.11) sowie der Unabhängigkeit der Komplemente $(A_k^{\mathrm{c}})_{k\geq 1}$ (Korollar 3.13) folgt

$$\begin{aligned}
\mathbb{P}(A^{\mathrm{c}}) &\leq \sum_{m\geq 1} \mathbb{P}\Big(\bigcap_{k\geq m} A_k^{\mathrm{c}}\Big) = \sum_{m\geq 1} \lim_{n\to\infty} \mathbb{P}\Big(\bigcap_{k=m}^{n} A_k^{\mathrm{c}}\Big) = \sum_{m\geq 1} \lim_{n\to\infty} \prod_{k=m}^{n} (1 - \mathbb{P}(A_k)) \\
&\leq \sum_{m\geq 1} \lim_{n\to\infty} \exp\Big(-\sum_{k=m}^{n} \mathbb{P}(A_k)\Big) = \sum_{m\geq 1} 0 = 0,
\end{aligned}$$

wobei wir $1 - x \leq \mathrm{e}^{-x}$ für $x \in \mathbb{R}$ benutzt haben und im letzten Schritt $\sum_{k\geq m} \mathbb{P}(A_k) = \infty$ für jedes m. Also ist $\mathbb{P}(A) = 1$. □

4 Erwartungswert und Varianz

Erwartungswert und Varianz sind zwei fundamentale Kenngrößen einer Verteilung:

Erwartungswert	$\widehat{=}$	typischer, mittlerer Wert,
Varianz	$\widehat{=}$	Streuung um den typischen Wert.

4.1 Erwartungswert

4.1.1 Der diskrete Fall

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X\colon \Omega \to \mathbb{R}$ eine Zufallsvariable. Die Zufallsvariable X heißt *diskret*, falls $X(\Omega) \coloneqq \{X(\omega)\colon \omega \in \Omega\}$ höchstens abzählbar ist.

Definition 4.1 (Erwartungswert diskreter Zufallsvariablen)**.** Sei X eine diskrete Zufallsvariable. Man sagt, X *besitzt einen Erwartungswert*, falls

$$\sum_{x \in X(\Omega)} |x|\, \mathbb{P}(X = x) < \infty,$$

und nennt in diesem Fall

$$\mathbb{E}[X] = \mathbb{E}_{\mathbb{P}}[X] \coloneqq \sum_{x \in X(\Omega)} x\, \mathbb{P}(X = x)$$

den *Erwartungswert* von X. Man schreibt dann $X \in \mathcal{L}^1(\mathbb{P})$ oder kurz $X \in \mathcal{L}^1$.

Bemerkung 4.2. (1) Der Erwartungswert hängt nur von der Verteilung $\mathbb{P} \circ X^{-1}$ von X ab: Es gilt

$$X \in \mathcal{L}^1(\mathbb{P}) \quad \Longleftrightarrow \quad \mathrm{id}_{X(\Omega)} \in \mathcal{L}^1(\mathbb{P} \circ X^{-1}),$$

und in diesem Fall ist $\mathbb{E}_{\mathbb{P}}[X] = \mathbb{E}_{\mathbb{P}\circ X^{-1}}[\mathrm{id}_{X(\Omega)}]$. Dabei fassen wir X als Zufallsvariable mit Werten in der abzählbaren Menge $X(\Omega)$, versehen mit $\mathcal{P}(X(\Omega))$, auf, d. h. $\mathbb{P} \circ X^{-1}$ als Wahrscheinlichkeitsmaß auf $\big(X(\Omega), \mathcal{P}(X(\Omega))\big)$.

(2) Ist X nichtnegativ, so ist $\sum_{x \in X(\Omega)} x\, \mathbb{P}(X = x)$ stets wohldefiniert in $[0, \infty]$ (sogar wenn man den Wert $X = \infty$ zulässt), und wir bezeichnen diese Summe wieder mit $\mathbb{E}[X]$. Insbesondere gilt für beliebige diskrete X

$$X \in \mathcal{L}^1(\mathbb{P}) \quad \Longleftrightarrow \quad \mathbb{E}\big[|X|\big] < \infty.$$

Beispiel 4.3. (1) *Indikatorfunktion.* Für $A \in \mathcal{F}$ ist $\mathbb{1}_A \in \mathcal{L}^1(\mathbb{P})$ und

$$\mathbb{E}[\mathbb{1}_A] = 0 \cdot \mathbb{P}(\mathbb{1}_A = 0) + 1 \cdot \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A).$$

(2) *Abzählbarer Ergebnisraum.* Ist Ω abzählbar, so ist jede Zufallsvariable X diskret, und durch Zusammenfassen der ω mit gleichem Wert $X(\omega)$ erhält man

$$\sum_{x\in X(\Omega)} |x|\,\mathbb{P}(X=x) = \sum_{\omega\in\Omega} |X(\omega)|\,\mathbb{P}(\{\omega\}).$$

Also gilt $X \in \mathcal{L}^1(\mathbb{P}) \iff \sum_{\omega\in\Omega} |X(\omega)|\,\mathbb{P}(\{\omega\}) < \infty$, und in diesem Fall ist

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\{\omega\}).$$

(3) *Binomialverteilung.* Sei $X_1, \dots, X_n$ eine endliche Bernoulli-Folge mit Erfolgswahrscheinlichkeit $p \in [0,1]$ und $S := \sum_{i=1}^n X_i$ die Anzahl der Erfolge, also $\mathbb{P}\circ S^{-1} = \mathcal{B}_{n,p}$. Dann ist

$$\mathbb{E}[S] = \sum_{k=0}^n k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^n np\underbrace{\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}}_{=\mathcal{B}_{n-1,p}(\{k-1\})} = np.$$

(4) *Geometrische Verteilung.* Sei $(X_n)_{n\in\mathbb{N}}$ eine Bernoulli-Folge zu $p \in (0,1]$ und $T := \inf\{n \geq 1\colon X_n = 1\}$ der Zeitpunkt des ersten Erfolgs. Dann ist T geometrisch verteilt, d. h. für $k \geq 1$ gilt

$$\mathbb{P}(T=k) = \mathbb{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = (1-p)^{k-1}p,$$

und mit der Ableitung der geometrischen Reihe $\sum_{k\geq 0} s^k = 1/(1-s)$ (gliedweise Differentiation der Potenzreihe für $|s| < 1$; für $p = 1$ ist $T = 1$ fast sicher) folgt

$$\mathbb{E}[T] = \sum_{k\geq 1} k\,p(1-p)^{k-1} = p\,\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=1-p}\sum_{k\geq 0} s^k = p\,\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=1-p}\frac{1}{1-s} = p\,\frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

Satz 4.4 (Rechenregeln für den Erwartungswert)**.** *Seien $X, Y, X_n, Y_n\colon \Omega \to \mathbb{R}$, $n \geq 1$, diskrete Zufallsvariablen; in (a) und (c) dürfen X und Y auch den Wert $+\infty$ annehmen. Dann gilt:*

(a) Monotonie: *Sind $X, Y \in \mathcal{L}^1$ und ist $X \leq Y$, so ist $\mathbb{E}[X] \leq \mathbb{E}[Y]$. Dasselbe gilt (in $[0,\infty]$) für nichtnegative $X \leq Y$.*

(b) Linearität: *Sind $X, Y \in \mathcal{L}^1$, so ist für alle $c \in \mathbb{R}$ auch $cX \in \mathcal{L}^1$ mit $\mathbb{E}[cX] = c\,\mathbb{E}[X]$, und es ist $X + Y \in \mathcal{L}^1$ mit $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.*

(c) σ-Additivität bzw. monotone Konvergenz: *Ist $X_n \geq 0$ für alle n und $X = \sum_{n\geq 1} X_n$, so gilt*

$$\mathbb{E}[X] = \sum_{n\geq 1}\mathbb{E}[X_n] \qquad (\text{in } [0,\infty]).$$

Ist $Y_n \in \mathcal{L}^1$ für alle n und gilt $Y_n \uparrow Y$ für $n \to \infty$, so gilt $\mathbb{E}[Y] = \lim_{n\to\infty}\mathbb{E}[Y_n]$. Dabei ist $\mathbb{E}[Y] := \mathbb{E}[Y_1] + \mathbb{E}[Y - Y_1] \in (-\infty, \infty]$ mit $\mathbb{E}[Y - Y_1] \in [0,\infty]$ wie in Bemerkung 4.2(2); für $Y \in \mathcal{L}^1$ stimmt dies nach (b) mit Definition 4.1 überein. Für $Y \notin \mathcal{L}^1$ ist $\mathbb{E}[Y] = \infty$, denn aus $\mathbb{E}[Y - Y_1] < \infty$ folgte $Y - Y_1 \in \mathcal{L}^1$ und damit $Y = Y_1 + (Y - Y_1) \in \mathcal{L}^1$; der Wert hängt also nicht von der Wahl der Folge ab.

(d) Produktregel bei Unabhängigkeit: *Sind $X, Y \in \mathcal{L}^1$ unabhängig, so ist $XY \in \mathcal{L}^1$ und*

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

Beweis. (a) Da $X \le Y$, ist $\mathbb{P}(X = x, Y = y) = 0$ für $y < x$. Da alle auftretenden Reihen absolut konvergieren, dürfen wir umordnen und erhalten

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\substack{x\in X(\Omega)\\ y\in Y(\Omega)}} x\,\mathbb{P}(X = x, Y = y) = \sum_{x\in X(\Omega)} \sum_{\substack{y\in Y(\Omega)\\ y\ge x}} x\,\mathbb{P}(X = x, Y = y)\\ &\le \sum_x \sum_y y\,\mathbb{P}(X = x, Y = y) = \mathbb{E}[Y].\end{aligned}$$

Dieselbe Rechnung zeigt die Monotonie auch für nichtnegative diskrete $X \le Y$ mit Werten in $[0, \infty]$.

(b) Die Aussage $\mathbb{E}[cX] = c\,\mathbb{E}[X]$ folgt durch Umindizieren (für $c \neq 0$ ist $\mathbb{P}(cX = z) = \mathbb{P}(X = z/c)$). Die Summe $X + Y$ ist eine diskrete Zufallsvariable mit Werten in der abzählbaren Menge $X(\Omega) + Y(\Omega)$, und

$$\begin{aligned}\sum_z |z|\,\mathbb{P}(X + Y = z) &= \sum_{z,x} |z|\,\mathbb{P}(X = x,\, Y = z - x)\\ &\le \sum_{z,x} \big(|x| + |z - x|\big)\,\mathbb{P}(X = x,\, Y = z - x)\\ &= \sum_x |x|\,\mathbb{P}(X = x) + \sum_y |y|\,\mathbb{P}(Y = y) < \infty.\end{aligned}$$

Die gleiche Rechnung ohne Betragsstriche (mit Gleichheit in der zweiten Zeile) liefert $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. Für nichtnegative X, Y gilt dies ebenso in $[0, \infty]$.

(c) *Zu „$\ge$“:* Für alle N ist $X \ge S_N := \sum_{n=1}^N X_n$, also nach (a) und (b)

$$\mathbb{E}[X] \ge \sum_{n=1}^N \mathbb{E}[X_n] \qquad \Longrightarrow \qquad \mathbb{E}[X] \ge \sum_{n=1}^\infty \mathbb{E}[X_n].$$

Zu „$\le$“: Wir zeigen etwas mehr: Ist $X \ge 0$ diskret und gilt nur $X \le \sum_{n\ge1} X_n$, so ist $\mathbb{E}[X] \le \sum_{n\ge1} \mathbb{E}[X_n]$. Seien dazu $0 < c < 1$, $K > 0$ und

$$\tau := \inf\{N \ge 1 \colon S_N \ge c \min\{X, K\}\}.$$

Da $S_N \uparrow \sum_n X_n \ge X$ und $c\min\{X, K\} < X$ auf $\{X > 0\}$ (sowie $\tau = 1$ auf $\{X = 0\}$), ist $\tau < \infty$ überall. Dann ist τ eine Zufallsvariable, und $S_\tau = \sum_{n=1}^\tau X_n$ ist eine diskrete Zufallsvariable (mit Werten in $\bigcup_N S_N(\Omega)$) mit $c\min\{X, K\} \le S_\tau$. Mit (a), (b) und Vertauschen von Summen mit nichtnegativen Gliedern folgt

$$\begin{aligned}c\,\mathbb{E}\big[\min\{X, K\}\big] \le \mathbb{E}[S_\tau] &= \sum_{x\in S_\tau(\Omega)} x \sum_{N\ge1} \mathbb{P}(\tau = N,\, S_N = x) = \sum_{N\ge1} \mathbb{E}\big[\mathbb{1}\{\tau = N\} S_N\big]\\ &= \sum_{N\ge1} \sum_{n=1}^N \mathbb{E}\big[\mathbb{1}\{\tau = N\} X_n\big]\\ &= \sum_{n\ge1} \sum_{N\ge n} \sum_{x\in X_n(\Omega)} x\,\mathbb{P}(\tau = N,\, X_n = x)\\ &= \sum_{n\ge1} \sum_{x\in X_n(\Omega)} x\,\mathbb{P}(\tau \ge n,\, X_n = x) \le \sum_{n\ge1} \mathbb{E}[X_n].\end{aligned}$$

Mit $c \uparrow 1$ und anschließend $K \uparrow \infty$ folgt die Behauptung, denn (monotone Konvergenz für Reihen mit nichtnegativen Gliedern)

$$\mathbb{E}\big[\min\{X, K\}\big] = \sum_x \min\{x, K\}\,\mathbb{P}(X = x) \xrightarrow[K\uparrow\infty]{} \sum_x x\,\mathbb{P}(X = x) = \mathbb{E}[X].$$

Die zweite Aussage folgt aus der ersten mit $X_n := Y_{n+1} - Y_n \geq 0$ und $X := Y - Y_1 = \sum_{n\geq 1} X_n$: Nach Definition von $\mathbb{E}[Y]$ ist

$$\mathbb{E}[Y] - \mathbb{E}[Y_1] = \mathbb{E}[Y - Y_1] = \sum_{n\geq 1} \big(\mathbb{E}[Y_{n+1}] - \mathbb{E}[Y_n]\big) = \lim_{n\to\infty} \mathbb{E}[Y_n] - \mathbb{E}[Y_1].$$

Dabei ist $\mathbb{E}[Y] = \infty$ genau dann, wenn $\mathbb{E}[Y - Y_1] = \infty$ ist (etwa wenn $\mathbb{P}(Y = \infty) > 0$).

(d) XY ist eine diskrete Zufallsvariable, und mit der Unabhängigkeit von X und Y gilt

$$\begin{aligned}\sum_z |z|\,\mathbb{P}(XY = z) &= \sum_{z\neq 0} |z| \sum_{x\neq 0} \mathbb{P}(X = x,\, Y = \tfrac{z}{x}) \\ &= \sum_{x\neq 0}\sum_{y\neq 0} |x||y|\,\mathbb{P}(X = x)\mathbb{P}(Y = y) = \mathbb{E}\big[|X|\big]\,\mathbb{E}\big[|Y|\big] < \infty.\end{aligned}$$

Ebenso (ohne Betragsstriche) folgt die Behauptung. □

4.1.2 Der allgemeine Fall

Für eine beliebige reelle Zufallsvariable $X\colon \Omega \to \mathbb{R}$ betrachten wir die *Diskretisierungen*

$$X_{(n)} := \frac{\lfloor nX \rfloor}{n}, \qquad n \geq 1,$$

von X, wobei $\lfloor x \rfloor := \max\{k \in \mathbb{Z}\colon k \leq x\}$ die Gauß-Klammer bezeichnet. Es ist also

$$X_{(n)}(\omega) = \frac{k}{n}, \qquad \text{falls } \frac{k}{n} \leq X(\omega) < \frac{k+1}{n},\ k \in \mathbb{Z}.$$

Wegen $\{X_{(n)} = k/n\} = \{k/n \leq X < (k+1)/n\} \in \mathcal{F}$ ist $X_{(n)}$ eine diskrete Zufallsvariable.

Lemma 4.5. (a) *Für alle $n \geq 1$ gilt $X_{(n)} \leq X < X_{(n)} + 1/n$.*

(b) *Ist $X_{(n)} \in \mathcal{L}^1$ für ein n, so ist $X_{(m)} \in \mathcal{L}^1$ für alle m, und $\big(\mathbb{E}[X_{(n)}]\big)_{n\geq 1}$ ist eine Cauchy-Folge.*

Beweis. (a) folgt direkt aus der Definition. Daraus folgt für alle $m, n \geq 1$

$$X_{(m)} \leq X < X_{(n)} + \frac{1}{n} \qquad \text{und} \qquad X_{(n)} \leq X < X_{(m)} + \frac{1}{m},$$

also $|X_{(m)}| \leq |X_{(n)}| + \max\{1/m, 1/n\}$.

(b) Ist $X_{(n)} \in \mathcal{L}^1$ für ein n, so folgt aus der letzten Ungleichung mit Satz 4.4(a) (für nichtnegative Zufallsvariablen) und (b), dass $X_{(m)} \in \mathcal{L}^1$ für alle m. Ferner liefert Satz 4.4(a), (b)

$$\mathbb{E}[X_{(m)}] \leq \mathbb{E}[X_{(n)}] + \frac{1}{n} \qquad \text{und} \qquad \mathbb{E}[X_{(n)}] \leq \mathbb{E}[X_{(m)}] + \frac{1}{m},$$

also $\big|\mathbb{E}[X_{(n)}] - \mathbb{E}[X_{(m)}]\big| \leq \max\{1/m, 1/n\}$. □

Nach Lemma 4.5 können wir den Erwartungswert einer beliebigen reellen Zufallsvariablen als Grenzwert der Erwartungswerte ihrer Diskretisierungen erklären.

Definition 4.6 (Erwartungswert)**.** Sei $X\colon \Omega \to \mathbb{R}$ eine beliebige reellwertige Zufallsvariable. Man sagt, dass X einen *Erwartungswert besitzt*, kurz $X \in \mathcal{L}^1(\mathbb{P})$, falls $X_{(n)} \in \mathcal{L}^1(\mathbb{P})$ für ein (oder, äquivalent, alle) $n \geq 1$. In diesem Fall heißt

$$\mathbb{E}[X] \coloneqq \lim_{n\to\infty} \mathbb{E}[X_{(n)}]$$

der *Erwartungswert* von X.

Ist X diskret, so stimmt diese Definition mit Definition 4.1 überein: Nach Lemma 4.5(a) und Satz 4.4(a), (b) ist $X \in \mathcal{L}^1$ genau dann, wenn $X_{(n)} \in \mathcal{L}^1$, und es gilt $|\mathbb{E}[X_{(n)}] - \mathbb{E}[X]| \leq 1/n$. Ist $X \geq 0$, so definiert man ebenso $\mathbb{E}[X] \coloneqq \lim_{n\to\infty} \mathbb{E}[X_{(n)}] \in [0,\infty]$ (vgl. Bemerkung 4.2(2)).

Bemerkung 4.7. (1) In der Integrationstheorie schreibt man manchmal

$$\int X \,\mathrm{d}\mathbb{P} = \int X(\omega)\,\mathbb{P}(\mathrm{d}\omega) \qquad \text{für } \mathbb{E}_{\mathbb{P}}[X].$$

(2) Wieder gilt $X \in \mathcal{L}^1(\mathbb{P}) \iff \mathrm{id}_{\mathbb{R}} \in \mathcal{L}^1(\mathbb{P}\circ X^{-1})$, und dann $\mathbb{E}_{\mathbb{P}}[X] = \mathbb{E}_{\mathbb{P}\circ X^{-1}}[\mathrm{id}_{\mathbb{R}}]$; der Erwartungswert hängt also nur von der Verteilung von X ab. (Denn $X_{(n)}$ hat unter $\mathbb{P}$ dieselbe Verteilung wie $(\mathrm{id}_{\mathbb{R}})_{(n)}$ unter $\mathbb{P}\circ X^{-1}$.)

Satz 4.8 (Rechenregeln)**.** *Die Rechenregeln (a)–(d) aus Satz 4.4 gelten auch für nicht diskrete Zufallsvariablen.*

Beweis. (a) Klar, denn aus $X \leq Y$ folgt $X_{(n)} \leq Y_{(n)}$ für alle n.

(b) Sei $c \in \mathbb{R}$. Nach Lemma 4.5(a) gilt

$$\big|(cX)_{(n)} - cX\big| \leq \frac{1}{n} \qquad \text{und} \qquad \big|cX_{(n)} - cX\big| \leq \frac{|c|}{n},$$

also $|(cX)_{(n)} - cX_{(n)}| \leq (1+|c|)/n$. Insbesondere ist $(cX)_{(n)} \in \mathcal{L}^1$, und mit Satz 4.4(a), (b) folgt

$$\Big|\mathbb{E}\big[(cX)_{(n)}\big] - c\,\mathbb{E}\big[X_{(n)}\big]\Big| \leq \frac{1+|c|}{n} \xrightarrow[n\to\infty]{} 0.$$

Ebenso kann die Additivität durch diskrete Approximation gezeigt werden, denn $|(X+Y)_{(n)} - X_{(n)} - Y_{(n)}| \leq 2/n$.

(c) Die Ungleichung „$\mathbb{E}[X] \geq \sum_{n\geq 1} \mathbb{E}[X_n]$“ zeigt man genau wie im diskreten Fall mit (a) und (b). Für die Ungleichung „$\leq$“ sei für $n, k \geq 1$

$$Y_{n,k} \coloneqq (X_n)_{(2^{n+k})}, \qquad \text{also} \qquad Y_{n,k} \leq X_n < Y_{n,k} + 2^{-n-k}.$$

Dann sind $X_{(m)}$ und $Y_{n,k} + 2^{-n-k}$ nichtnegative diskrete Zufallsvariablen mit

$$X_{(m)} \leq X = \sum_{n\geq 1} X_n \leq \sum_{n\geq 1} \big(Y_{n,k} + 2^{-n-k}\big)$$

für alle $m \geq 1$. Nach der im Beweis von Satz 4.4(c) gezeigten verschärften Ungleichung und mit (a), (b) folgt

$$\mathbb{E}[X_{(m)}] \leq \sum_{n\geq 1} \mathbb{E}\big[Y_{n,k} + 2^{-n-k}\big] = \sum_{n\geq 1} \underbrace{\mathbb{E}[Y_{n,k}]}_{\leq \mathbb{E}[X_n]} + 2^{-k} \leq \sum_{n\geq 1} \mathbb{E}[X_n] + 2^{-k}.$$

Mit $m \to \infty$ und dann $k \to \infty$ folgt $\mathbb{E}[X] \le \sum_{n\ge 1} \mathbb{E}[X_n]$. Die Aussage über $Y_n \uparrow Y$ folgt daraus wie im diskreten Fall.

(d) Es gilt

$$\begin{aligned}|(XY)_{(n)} - X_{(n)}Y_{(n)}| &\le |(XY)_{(n)} - XY| + |X(Y - Y_{(n)})| + |Y_{(n)}(X - X_{(n)})| \\ &\le \frac{1}{n} + \frac{1}{n}\Big(|X| + |Y| + \frac{1}{n}\Big).\end{aligned}$$

Mit X, Y sind auch die diskreten Zufallsvariablen $X_{(n)}$ und $Y_{(n)}$ unabhängig (Satz 3.16), also ist nach Satz 4.4(d) $X_{(n)}Y_{(n)} \in \mathcal{L}^1$ mit $\mathbb{E}[X_{(n)}Y_{(n)}] = \mathbb{E}[X_{(n)}]\mathbb{E}[Y_{(n)}]$. Aus der obigen Abschätzung folgt $XY \in \mathcal{L}^1$ und

$$\Big|\mathbb{E}\big[(XY)_{(n)}\big] - \mathbb{E}[X_{(n)}]\,\mathbb{E}[Y_{(n)}]\Big| \le \frac{1}{n}\Big(1 + \mathbb{E}|X| + \mathbb{E}|Y| + \frac{1}{n}\Big) \xrightarrow[n\to\infty]{} 0,$$

also $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. □

Satz 4.9 (Erwartungswert bei existierender Dichte)**.** *Sei $\Omega \in \mathcal{B}^d$ und $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B}^d_\Omega)$ mit Dichtefunktion ρ, und sei X eine reellwertige Zufallsvariable auf Ω. Dann gilt*

$$X \in \mathcal{L}^1(\mathbb{P}) \quad \Longleftrightarrow \quad \int_\Omega |X(\omega)|\,\rho(\omega)\,\mathrm{d}\omega < \infty,$$

und in diesem Fall ist

$$\mathbb{E}[X] = \int_\Omega X(\omega)\,\rho(\omega)\,\mathrm{d}\omega.$$

Beweis. Für jedes n gilt nach Tatsache 1.14(b)

$$\sum_{k\in\mathbb{Z}} \Big|\frac{k}{n}\Big|\,\mathbb{P}\Big(X_{(n)} = \frac{k}{n}\Big) = \sum_{k\in\mathbb{Z}} \Big|\frac{k}{n}\Big| \int_{\{k/n \le X < (k+1)/n\}} \rho(\omega)\,\mathrm{d}\omega = \int_\Omega |X_{(n)}(\omega)|\,\rho(\omega)\,\mathrm{d}\omega.$$

Also ist $X_{(n)} \in \mathcal{L}^1(\mathbb{P})$ genau dann, wenn $\int_\Omega |X_{(n)}(\omega)|\rho(\omega)\,\mathrm{d}\omega < \infty$, und wegen $|X - X_{(n)}| \le 1/n$ und $\int_\Omega \rho(\omega)\,\mathrm{d}\omega = 1$ ist dies äquivalent zu $\int_\Omega |X(\omega)|\rho(\omega)\,\mathrm{d}\omega < \infty$. Ebenso (Zerlegung in Positiv- und Negativteil) ist

$$\mathbb{E}[X_{(n)}] = \int_\Omega X_{(n)}(\omega)\,\rho(\omega)\,\mathrm{d}\omega \xrightarrow[n\to\infty]{} \int_\Omega X(\omega)\,\rho(\omega)\,\mathrm{d}\omega,$$

denn

$$\int_\Omega X_{(n)}(\omega)\rho(\omega)\,\mathrm{d}\omega \le \int_\Omega X(\omega)\rho(\omega)\,\mathrm{d}\omega \le \int_\Omega X_{(n)}(\omega)\rho(\omega)\,\mathrm{d}\omega + \frac{1}{n}.$$ □

Bemerkung 4.10. Besitzt die Zufallsvariable $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}^d)$ eine Dichte, ist also ρ eine Dichtefunktion von $\mathbb{P} \circ X^{-1}$ auf $\mathbb{R}^d$, und ist $f\colon \mathbb{R}^d \to \mathbb{R}$ messbar, so gilt $f \circ X \in \mathcal{L}^1(\mathbb{P})$ genau dann, wenn $\int_{\mathbb{R}^d} |f(x)|\rho(x)\,\mathrm{d}x < \infty$, und dann ist

$$\mathbb{E}[f \circ X] = \int_{\mathbb{R}^d} f(x)\,\rho(x)\,\mathrm{d}x.$$

Denn $f \circ X$ hat unter $\mathbb{P}$ dieselbe Verteilung wie f unter $\mathbb{P} \circ X^{-1}$; die Behauptung folgt also aus Bemerkung 4.7(2) und Satz 4.9.

Neben dem Erwartungswert gibt es weitere Möglichkeiten, einen „mittleren Wert“ einer reellen Zufallsvariablen X anzugeben; manchmal ist dies erhellender.

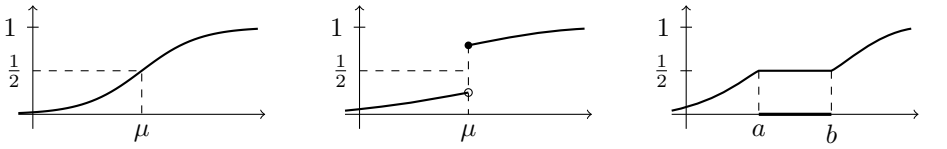
Definition 4.11 (Median)**.** Sei X eine reellwertige Zufallsvariable mit Verteilung Q auf $(\mathbb{R}, \mathcal{B})$. Eine Zahl $\mu \in \mathbb{R}$ heißt *Median* von X (bzw. von Q), falls

$$\mathbb{P}(X \geq \mu) \geq \frac{1}{2} \qquad \text{und} \qquad \mathbb{P}(X \leq \mu) \geq \frac{1}{2}.$$

Bemerkung 4.12. (1) Der Median ist im Allgemeinen nicht eindeutig.

(2) Ein Median existiert immer: Ist F_X die Verteilungsfunktion von X, so ist zum Beispiel $\mu := \inf\{c \in \mathbb{R} \colon F_X(c) \geq 1/2\}$ ein Median. Denn wegen der rechtsseitigen Stetigkeit ist $\mathbb{P}(X \leq \mu) = F_X(\mu) \geq 1/2$, und für $c < \mu$ ist $F_X(c) < 1/2$, also $\mathbb{P}(X < \mu) = \lim_{c \uparrow \mu} F_X(c) \leq 1/2$ und somit $\mathbb{P}(X \geq \mu) \geq 1/2$.

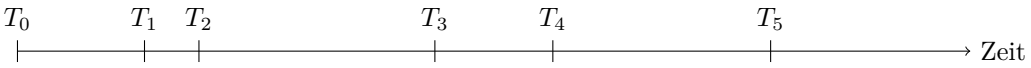
Man kann einen Median am Graphen der Verteilungsfunktion F_X ablesen:



eindeutiger Median | Sprung über $\frac{1}{2}$ | jedes $\mu \in [a, b]$ ist Median

4.2 Das Wartezeitparadoxon

Wir betrachten einen Poisson-Prozess mit Intensität $\alpha > 0$ und Sprungzeiten $0 = T_0 < T_1 < T_2 < \dots$, die wir als Ankunftszeiten von Bussen an einer Haltestelle deuten.



Nach Konstruktion (Abschnitt 3.5) sind die Zwischenankunftszeiten $L_k := T_k - T_{k-1}$, $k \geq 1$,

(1) unabhängig,

(2) exponentialverteilt mit Parameter α, also $\mathbb{E}[L_k] = \int_0^\infty x\, \alpha \mathrm{e}^{-\alpha x}\, \mathrm{d}x = 1/\alpha$, und

(3) *gedächtnislos:* Für alle $s, t \geq 0$ und alle k gilt

$$\mathbb{P}(L_k > s + t \,|\, L_k > s) = \mathbb{P}(L_k > t), \qquad \text{denn} \qquad \frac{\mathrm{e}^{-\alpha(s+t)}}{\mathrm{e}^{-\alpha s}} = \mathrm{e}^{-\alpha t}.$$

Frage: Wie lange muss ein Fahrgast, der zur Zeit $t > 0$ an der Haltestelle ankommt, im Mittel auf den nächsten Bus warten?

Naive Idee: Die erwartete Zeit zwischen zwei Bussen ist $1/\alpha$, und der Ankunftszeitpunkt ist innerhalb eines solchen Intervalls gleichverteilt. Die Wartezeit wäre dann gleichverteilt auf $[0, 1/\alpha]$, mit Erwartungswert

$$\alpha \int_0^{1/\alpha} x \, \mathrm{d}x = \frac{\alpha}{2} x^2 \Big|_0^{1/\alpha} = \frac{\alpha}{2} \cdot \frac{1}{\alpha^2} = \frac{1}{2\alpha}.$$

Rechnung: Die zufällige Wartezeit ist

$$W_t := \min\{T_k - t \colon k \geq 1,\ T_k \geq t\}.$$

Für $s \geq 0$ ist $W_t > s$ genau dann, wenn im Zeitintervall $[t, t+s]$ kein Bus ankommt. Für $c > 0$, $k \geq 1$ und $0 < \varepsilon < c$ ist $\mathbb{P}(T_k = c) \leq \mathbb{P}(N_c - N_{c-\varepsilon} \geq 1) = 1 - \mathrm{e}^{-\alpha\varepsilon}$, also $\mathbb{P}(T_k = c) = 0$. Daher unterscheidet sich $\{W_t > s\}$ von $\{N_{t+s} - N_t = 0\}$ nur um eine Nullmenge, und nach Satz 3.24 gilt

$$\mathbb{P}(W_t > s) = \mathbb{P}(N_{t+s} - N_t = 0) = \mathrm{e}^{-\alpha s}.$$

Folglich ist W_t exponentialverteilt mit Parameter α, und

$$\mathbb{E}[W_t] = \frac{1}{\alpha} \neq \frac{1}{2\alpha}.$$

Interpretation: Egal, wann man ankommt: Die erwartete Wartezeit entspricht genau dem erwarteten Abstand zwischen zwei Ankünften. Der Grund: Mein Ankunftszeitpunkt t liegt mit höherer Wahrscheinlichkeit in einem *großen* Zeitintervall zwischen zwei Bussen! Um dies zu sehen, betrachten wir auch die seit der letzten Busankunft vergangene Zeit

$$V_t := \min\{t - T_k \colon k \geq 0,\ T_k \leq t\}.$$

Für $0 \leq s < t$ ist $V_t > s$ genau dann, wenn kein T_k in $[t-s, t]$ liegt; bis auf eine Nullmenge (s. o.) ist dies das Ereignis $\{N_t - N_{t-s} = 0\}$, also $\mathbb{P}(V_t > s) = \mathbb{P}(N_t - N_{t-s} = 0) = \mathrm{e}^{-\alpha s}$, und $\mathbb{P}(V_t > s) = 0$ für $s \geq t$ (wegen $T_0 = 0$). Also hat V_t dieselbe Verteilung wie $\min\{L_1, t\}$, und nach Bemerkung 4.10 ist

$$\mathbb{E}[V_t] = \int_0^\infty (x \wedge t)\, \alpha \mathrm{e}^{-\alpha x}\, \mathrm{d}x = \int_0^t x\, \alpha \mathrm{e}^{-\alpha x}\, \mathrm{d}x + t\, \mathrm{e}^{-\alpha t} = \frac{1}{\alpha} - \frac{1}{\alpha}\mathrm{e}^{-\alpha t}$$

(partielle Integration). Damit folgt

$$\mathbb{E}[V_t + W_t] = \frac{2}{\alpha} - \frac{1}{\alpha}\mathrm{e}^{-\alpha t} \approx \frac{2}{\alpha} \qquad \text{für große } t.$$

Das Zeitintervall zwischen zwei Bussen, in das der Zeitpunkt t fällt, ist also im Mittel tatsächlich ungefähr doppelt so lang wie ein typisches Intervall. In diesem Sinne stimmt die naive Idee doch: Der Fahrgast wartet im Mittel die Hälfte des Intervalls, in dem er ankommt.

4.3 Varianz und Kovarianz

Definition 4.13 (Momente)**.** Sei $X \colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ eine Zufallsvariable und $r \in \mathbb{N}$. Falls $X^r \in \mathcal{L}^1(\mathbb{P})$, so nennt man $\mathbb{E}[X^r]$ das *r-te Moment* von X und schreibt $X \in \mathcal{L}^r(\mathbb{P}) = \mathcal{L}^r$. Da für $r < s$

$$|X|^r \leq 1 + |X|^s,$$

gilt $\mathcal{L}^s \subset \mathcal{L}^r$.

Definition 4.14 (Varianz und Kovarianz)**.** Für $X, Y \in \mathcal{L}^2$ heißt

(1)
$$\mathrm{Var}[X] = \mathrm{Var}_{\mathbb{P}}[X] := \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

die *Varianz* von X und $\sqrt{\mathrm{Var}[X]}$ die *Streuung* oder *Standardabweichung* von X bzgl. $\mathbb{P}$;

(2)
$$\operatorname{Cov}(X,Y) := \mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] = \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$$
die *Kovarianz* von X und Y;

(3) X und Y *unkorreliert*, falls $\operatorname{Cov}(X,Y)=0$.

Dabei ist $XY \in \mathcal{L}^1$ wegen $|XY| \le (1/2)(X^2+Y^2)$, und die jeweils zweite Darstellung folgt durch Ausmultiplizieren mit der Linearität des Erwartungswerts.

Bemerkung 4.15. (1) Die Varianz sagt etwas über die Abweichung vom Durchschnitt aus, die Kovarianz etwas über die Abweichung vom unabhängigen Verhalten von X und Y (vgl. Satz 4.16(4)).

(2) Erwartungswert und Varianz hängen nur von der Verteilung ab.

Satz 4.16 (Rechenregeln)**.** *Seien $X, Y, X_1, \dots, X_n \in \mathcal{L}^2$ und $a, b, c, d \in \mathbb{R}$. Dann gilt:*

(1) $aX+b,\ cY+d \in \mathcal{L}^2$ und $\operatorname{Cov}(aX+b, cY+d) = ac\operatorname{Cov}(X,Y)$. Insbesondere gilt
$$\operatorname{Var}[aX+b] = a^2 \operatorname{Var}[X].$$

(2) $\operatorname{Cov}(X,Y)^2 \le \operatorname{Var}[X]\operatorname{Var}[Y]$ *(Cauchy–Schwarz-Ungleichung)*

(3) $\sum_{i=1}^n X_i \in \mathcal{L}^2$ und
$$\operatorname{Var}\Big[\sum_{i=1}^n X_i\Big] = \sum_{i=1}^n \operatorname{Var}[X_i] + \sum_{i\ne j} \operatorname{Cov}(X_i, X_j).$$
Insbesondere gilt, falls $X_1, \dots, X_n$ paarweise unkorreliert sind,
$$\operatorname{Var}\Big[\sum_{i=1}^n X_i\Big] = \sum_{i=1}^n \operatorname{Var}[X_i].$$

(4) Sind X, Y unabhängig, so sind X, Y unkorreliert.

Beweis. (1) Wegen $(aX+b)^2 \le 2a^2X^2 + 2b^2$ ist $aX+b \in \mathcal{L}^2$, und die Formel folgt durch direktes Nachrechnen, da $aX+b-\mathbb{E}[aX+b] = a(X-\mathbb{E}[X])$.

(2) Wegen (1) können wir $\mathbb{E}[X]=\mathbb{E}[Y]=0$ annehmen. Für alle $t \in \mathbb{R}$ gilt
$$0 \le \mathbb{E}\big[(X-tY)^2\big] = \mathbb{E}[X^2] - 2t\,\mathbb{E}[XY] + t^2\,\mathbb{E}[Y^2].$$
Ist $\mathbb{E}[Y^2] > 0$, so liefert die Wahl $t = \mathbb{E}[XY]/\mathbb{E}[Y^2]$
$$0 \le \mathbb{E}[X^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]}, \qquad \text{also} \qquad \operatorname{Cov}(X,Y)^2 = \mathbb{E}[XY]^2 \le \mathbb{E}[X^2]\mathbb{E}[Y^2] = \operatorname{Var}[X]\operatorname{Var}[Y].$$
Ist $\mathbb{E}[Y^2]=0$, so ist $0 \le \mathbb{E}[X^2] - 2t\,\mathbb{E}[XY]$ für alle $t \in \mathbb{R}$, also $\mathbb{E}[XY]=0$, und beide Seiten der Ungleichung sind 0.

(3) Wegen $\big(\sum_{i=1}^n X_i\big)^2 \le n\sum_{i=1}^n X_i^2$ ist $\sum_i X_i \in \mathcal{L}^2$. Wegen (1) können wir $\mathbb{E}[X_i]=0$ für alle i annehmen. Dann ist
$$\operatorname{Var}\Big[\sum_{i=1}^n X_i\Big] = \mathbb{E}\Big[\Big(\sum_{i=1}^n X_i\Big)^2\Big] = \sum_{i,j=1}^n \mathbb{E}[X_iX_j] = \sum_{i=1}^n \operatorname{Var}[X_i] + \sum_{i\ne j} \operatorname{Cov}(X_i, X_j).$$

(4) Nach der Produktregel (Satz 4.8 mit Satz 4.4(d)) ist $\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$. □

Bemerkung 4.17. (1) Ist $X \in \mathcal{L}^2$ mit $\operatorname{Var}[X] > 0$, so gilt für

$$X^* := \frac{X - \mathbb{E}[X]}{\sqrt{\operatorname{Var}[X]}}$$

$\mathbb{E}[X^*] = 0$ und $\operatorname{Var}[X^*] = 1$. Man nennt X^* die *Standardisierung* von X.

(2) *Achtung:* Unkorreliertheit bezieht sich nur auf den gemeinsamen Erwartungswert $\mathbb{E}[XY]$, Unabhängigkeit dagegen auf die ganze gemeinsame Verteilung. Aus Unkorreliertheit folgt im Allgemeinen *nicht* Unabhängigkeit:

Beispiel 4.18 (Unabhängigkeit vs. Unkorreliertheit)**.** Sei $\Omega = \{1,2,3\}$ und $\mathbb{P}$ die Gleichverteilung auf Ω. Wir definieren

$$X\colon\ 1 \mapsto 1,\ 2 \mapsto 0,\ 3 \mapsto -1, \qquad\qquad Y\colon\ 1 \mapsto 0,\ 2 \mapsto 1,\ 3 \mapsto 0.$$

Dann ist $XY = 0$ und $\mathbb{E}[X] = 0$, also $\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$. Aber

$$\mathbb{P}(X = 1, Y = 1) = 0 \qquad \text{und} \qquad \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \neq 0,$$

d. h. X und Y sind unkorreliert, aber nicht unabhängig.

Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathcal{B})$ schreiben wir $\mathbb{E}[\mathbb{P}] := \mathbb{E}_{\mathbb{P}}[\mathrm{id}]$ und $\operatorname{Var}[\mathbb{P}] := \operatorname{Var}_{\mathbb{P}}[\mathrm{id}]$ (sofern existent); nach Bemerkung 4.15(2) sind dies Erwartungswert und Varianz jeder Zufallsvariablen mit Verteilung $\mathbb{P}$.

Beispiel 4.19. (1) *Varianz der Binomialverteilung.* Sei $X_1, \dots, X_n$ eine Bernoulli-Folge zu p, also $\sum_{i=1}^n X_i$ binomialverteilt (Beispiel 4.3(3)). Die X_i sind unabhängig, also paarweise unkorreliert (Satz 4.16(4)), und es ist $\operatorname{Var}[X_i] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = p - p^2 = p(1-p)$. Mit Satz 4.16(3) folgt

$$\operatorname{Var}[\mathcal{B}_{n,p}] = \sum_{i=1}^n \operatorname{Var}[X_i] = np(1-p).$$

(2) *Normalverteilung.* Mit Satz 4.9 und der Substitution $y = (x-m)/(\sqrt{v})$ gilt

$$\begin{aligned}\mathbb{E}[\mathcal{N}_{m,v}] &= \frac{1}{\sqrt{2\pi v}} \int x\, \mathrm{e}^{-\frac{(x-m)^2}{2v}}\, \mathrm{d}x = \frac{1}{\sqrt{2\pi v}} \int \mathrm{e}^{-\frac{y^2}{2}} (m + y\sqrt{v})\sqrt{v}\, \mathrm{d}y \\ &= m + \frac{\sqrt{v}}{\sqrt{2\pi}} \int y\, \mathrm{e}^{-\frac{y^2}{2}}\, \mathrm{d}y = m,\end{aligned}$$

da der Integrand im letzten Integral ungerade (und integrierbar) ist und $\frac{1}{\sqrt{2\pi}} \int \mathrm{e}^{-y^2/2}\, \mathrm{d}y = 1$. Ebenso folgt mit partieller Integration

$$\begin{aligned}\operatorname{Var}[\mathcal{N}_{m,v}] &= \frac{1}{\sqrt{2\pi v}} \int (x-m)^2 \mathrm{e}^{-\frac{(x-m)^2}{2v}}\, \mathrm{d}x = \frac{v}{\sqrt{2\pi}} \int y^2 \mathrm{e}^{-\frac{y^2}{2}}\, \mathrm{d}y \\ &= \frac{v}{\sqrt{2\pi}} \Big(\underbrace{\Big[-y\, \mathrm{e}^{-\frac{y^2}{2}}\Big]_{-\infty}^{\infty}}_{=0} + \int \mathrm{e}^{-\frac{y^2}{2}}\, \mathrm{d}y \Big) = v.\end{aligned}$$

4.4 Erzeugende Funktionen

Wir betrachten einen Wahrscheinlichkeitsraum $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0), \mathbb{P})$ mit Zähldichte $\rho(k) = \mathbb{P}(\{k\})$, $k \in \mathbb{N}_0$. Erinnerung: Eine Potenzreihe legt ihre Koeffizienten eindeutig fest; die Zähldichte ρ ist also durch die Potenzreihe mit den Koeffizienten $\rho(k)$ eindeutig bestimmt.

Definition 4.20 (Erzeugende Funktion)**.** Sei ρ eine Zähldichte auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ und $\mathbb{P}$ das zugehörige Wahrscheinlichkeitsmaß. Dann nennt man

$$\varphi_{\mathbb{P}}(s) \coloneqq \sum_{k\geq 0} \rho(k)\, s^k, \qquad 0 \leq s \leq 1,$$

die *erzeugende Funktion* von $\mathbb{P}$ bzw. ρ.

Wegen $\sum_{k\geq 0} \rho(k) = 1$ konvergiert die Reihe für alle $|s| \leq 1$ absolut; ihr Konvergenzradius ist also mindestens 1.

Beispiel 4.21. (1) *Binomialverteilung:* Nach dem binomischen Lehrsatz ist

$$\varphi_{\mathcal{B}_{n,p}}(s) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k = \big(sp + (1-p)\big)^n.$$

(2) *Poisson-Verteilung:*

$$\varphi_{\mathrm{Pois}_\lambda}(s) = \sum_{k\geq 0} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} s^k = \mathrm{e}^{-\lambda(1-s)}.$$

Satz 4.22 (Momentenberechnung mit erzeugenden Funktionen)**.** *Sei $\varphi_{\mathbb{P}}$ die erzeugende Funktion von $\mathbb{P}$. Dann gilt:*

(1) Für alle $k \geq 0$ ist $\rho(k) = \varphi_{\mathbb{P}}^{(k)}(0)/k!$. Insbesondere ist die Zuordnung $\mathbb{P} \mapsto \varphi_{\mathbb{P}}$ injektiv.

(2) $\mathbb{E}[\mathbb{P}] < \infty$ (d. h. $\mathrm{id} \in \mathcal{L}^1(\mathbb{P})$) genau dann, wenn $\varphi'_{\mathbb{P}}(1) \coloneqq \lim_{s\uparrow 1} \varphi'_{\mathbb{P}}(s)$ endlich ist. In diesem Fall ist

$$\mathbb{E}[\mathbb{P}] = \varphi'_{\mathbb{P}}(1).$$

(3) $\mathrm{Var}[\mathbb{P}] < \infty$ (d. h. $\mathrm{id} \in \mathcal{L}^2(\mathbb{P})$) genau dann, wenn $\varphi''_{\mathbb{P}}(1) \coloneqq \lim_{s\uparrow 1} \varphi''_{\mathbb{P}}(s)$ endlich ist. In diesem Fall ist

$$\mathrm{Var}[\mathbb{P}] = \varphi''_{\mathbb{P}}(1) - \mathbb{E}[\mathbb{P}]^2 + \mathbb{E}[\mathbb{P}].$$

Beweis. Wir schreiben kurz $\varphi = \varphi_{\mathbb{P}}$.

(1) folgt aus dem Satz von Taylor für Potenzreihen (gliedweises Differenzieren im Innern des Konvergenzbereichs).

(2) Für $0 \leq s < 1$ gilt mit der geometrischen Summenformel

$$\frac{\varphi(1) - \varphi(s)}{1-s} = \sum_{k\geq 0} \rho(k) \frac{1-s^k}{1-s} = \sum_{k\geq 0} \rho(k) \sum_{j=0}^{k-1} s^j.$$

Alle Summanden sind nichtnegativ und monoton wachsend in s und in der Anzahl der Summanden. Daher existiert der Grenzwert $s \uparrow 1$ in $[0,\infty]$, und man darf die Suprema vertauschen:

$$\lim_{s\uparrow 1} \frac{\varphi(1)-\varphi(s)}{1-s} = \sup_{s<1}\sup_{n\geq 1} \sum_{k=0}^{n} \rho(k) \sum_{j=0}^{k-1} s^j = \sup_{n\geq 1} \sum_{k=0}^{n} \rho(k)\, k = \sum_{k\geq 0} \rho(k)\, k = \mathbb{E}[\mathbb{P}].$$

Genauso erhält man mit $\varphi'(s) = \sum_{k\geq 1} \rho(k)\, k\, s^{k-1}$ für $s < 1$

$$\lim_{s\uparrow 1} \varphi'(s) = \sum_{k\geq 0} \rho(k)\, k = \mathbb{E}[\mathbb{P}] \in [0,\infty].$$

Alle Ausdrücke sind also genau dann endlich, wenn $\mathbb{E}[\mathbb{P}] < \infty$, und stimmen dann überein.

(3) Wie in (2) gilt $\varphi''(s) = \sum_{k\geq 2} \rho(k)\, k(k-1)\, s^{k-2} \uparrow \sum_{k\geq 0} \rho(k)\, k(k-1)$ für $s \uparrow 1$. Dieser Grenzwert ist genau dann endlich, wenn $\sum_k \rho(k) k^2 < \infty$, und dann ist

$$\operatorname{Var}[\mathbb{P}] = \sum_k \rho(k)k^2 - \mathbb{E}[\mathbb{P}]^2 = \sum_k \rho(k)k(k-1) + \sum_k \rho(k)k - \mathbb{E}[\mathbb{P}]^2 = \varphi''(1) - \mathbb{E}[\mathbb{P}]^2 + \mathbb{E}[\mathbb{P}]. \quad \square$$

Beispiel 4.23. (1) *Binomialverteilung:* Mit Beispiel 4.21(1) und Satz 4.22 ist

$$\mathbb{E}[\mathcal{B}_{n,p}] = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=1} \big(sp+(1-p)\big)^n = n\big(sp+(1-p)\big)^{n-1} p\,\Big|_{s=1} = np$$

und

$$\begin{aligned}\operatorname{Var}[\mathcal{B}_{n,p}] &= \frac{\mathrm{d}^2}{\mathrm{d}s^2}\Big|_{s=1} \big(sp+(1-p)\big)^n - (np)^2 + np \\ &= \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=1} \Big(np\big(sp+1-p\big)^{n-1}\Big) - (np)^2 + np \\ &= n(n-1)p^2 \underbrace{\big(sp+1-p\big)^{n-2}\Big|_{s=1}}_{=1} - (np)^2 + np = -np^2 + np = np(1-p),\end{aligned}$$

in Übereinstimmung mit Beispiel 4.3(3) und Beispiel 4.19(1).

(2) *Poisson-Verteilung:* Mit Beispiel 4.21(2) ist

$$\mathbb{E}[\mathrm{Pois}_\lambda] = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=1} \mathrm{e}^{-\lambda(1-s)} = \mathrm{e}^{-\lambda}\lambda\mathrm{e}^{\lambda} = \lambda$$

und

$$\begin{aligned}\operatorname{Var}[\mathrm{Pois}_\lambda] = \frac{\mathrm{d}^2}{\mathrm{d}s^2}\Big|_{s=1} \mathrm{e}^{-\lambda(1-s)} - \lambda^2 + \lambda &= \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=1} \mathrm{e}^{-\lambda}\lambda\mathrm{e}^{\lambda s} - \lambda^2 + \lambda \\ = \mathrm{e}^{-\lambda}\lambda^2\mathrm{e}^{\lambda} - \lambda^2 + \lambda = \lambda &= \mathbb{E}[\mathrm{Pois}_\lambda].\end{aligned}$$

(Da die erzeugenden Funktionen hier auf ganz $\mathbb{R}$ definiert und beliebig oft differenzierbar sind, stimmen die Grenzwerte $\lim_{s\uparrow 1} \varphi'(s)$ und $\lim_{s\uparrow 1} \varphi''(s)$ mit den gewöhnlichen Ableitungen an der Stelle $s = 1$ überein.)

Definition 4.24 (Erzeugende Funktion einer Zufallsvariablen)**.** Sei X eine $\mathbb{N}_0$-wertige Zufallsvariable. Dann heißt

$$\varphi_X(s) \coloneqq \varphi_{\mathbb{P}\circ X^{-1}}(s) = \sum_{k\geq 0} s^k\, \mathbb{P}(X=k) = \mathbb{E}\big[s^X\big], \qquad 0 \leq s \leq 1,$$

die *erzeugende Funktion* von X (mit der Konvention $0^0 \coloneqq 1$).

Satz 4.25 (Erzeugende Funktion von Summen unabhängiger Zufallsvariablen)**.** *Sind X, Y unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen, so gilt*

$$\varphi_{X+Y}(s) = \varphi_X(s)\,\varphi_Y(s), \qquad 0 \le s \le 1.$$

Beweis. Für festes s sind s^X und s^Y als messbare Funktionen der unabhängigen Zufallsvariablen X bzw. Y unabhängig (Satz 3.16) und beschränkt. Mit der Produktregel (Satz 4.4(d)) folgt

$$\varphi_{X+Y}(s) = \mathbb{E}\big[s^{X+Y}\big] = \mathbb{E}\big[s^X s^Y\big] = \mathbb{E}\big[s^X\big]\mathbb{E}\big[s^Y\big] = \varphi_X(s)\,\varphi_Y(s). \qquad \square$$

5 Gesetz der großen Zahl und zentraler Grenzwertsatz

In diesem Kapitel behandeln wir zwei fundamentale Grenzwertsätze für das Langzeitverhalten von empirischen Mitteln unabhängiger Zufallsvariablen.

5.1 Das Gesetz der großen Zahl

Wir betrachten die empirischen Mittel

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

unabhängiger Zufallsvariablen $X_1, X_2, \ldots$. Ist etwa $X_i = \mathbb{1}\{Y_i \in A\}$ für unabhängige, identisch verteilte Y_i mit Verteilung Q, so entspricht die Idee

$$Q(A) \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{Y_i \in A\}$$

der *frequentistischen* Interpretation von Wahrscheinlichkeit: Die Wahrscheinlichkeit eines Ereignisses ist seine relative Häufigkeit bei vielen unabhängigen Wiederholungen.

5.1.1 Das schwache Gesetz der großen Zahl

Beispiel 5.1 (Bernoulli-Experiment)**.** Sei $X_1, X_2, \ldots$ eine Bernoulli-Folge zu $p = 1/2$ (fairer Münzwurf). Dann ist $\sum_{i=1}^{2n} X_i$ gemäß $\mathcal{B}_{2n,1/2}$ verteilt, und

$$\mathbb{P}\Big(\frac{1}{2n}\sum_{i=1}^{2n} X_i = \frac{1}{2}\Big) = \mathcal{B}_{2n,\frac{1}{2}}(\{n\}) = \binom{2n}{n} 2^{-2n}. \qquad (*)$$

Mit der *Stirling-Formel* (ohne Beweis)

$$n! = \sqrt{2\pi n}\, n^n \mathrm{e}^{-n+\eta(n)} \qquad \text{mit} \qquad 0 < \eta(n) < \frac{1}{12n}, \text{ insbesondere } \eta(n) \xrightarrow[n\to\infty]{} 0,$$

folgt

$$(*) = \frac{(2n)!}{(n!)^2}\, 2^{-2n} \underset{n\to\infty}{\sim} \frac{\sqrt{2\pi \cdot 2n}\,(2n)^{2n}}{\big(\sqrt{2\pi n}\, n^n\big)^2}\, 2^{-2n} = \frac{1}{\sqrt{\pi n}} \xrightarrow[n\to\infty]{} 0.$$

Die Wahrscheinlichkeit, dass die relative Häufigkeit *exakt* gleich 1/2 ist, geht also gegen 0. Man kann daher nur erwarten, dass das empirische Mittel mit hoher Wahrscheinlichkeit *nahe* bei 1/2 liegt. Dies führt zu folgendem Konvergenzbegriff.

Definition 5.2 (Konvergenz in Wahrscheinlichkeit)**.** Seien $Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Man sagt, Y_n *konvergiert stochastisch* (oder *in Wahrscheinlichkeit*) gegen Y, und schreibt $Y_n \xrightarrow{\mathbb{P}} Y$, falls

$$\mathbb{P}\big(|Y_n - Y| \le \varepsilon\big) \xrightarrow[n\to\infty]{} 1 \qquad \text{für alle } \varepsilon > 0.$$

Proposition 5.3 (Markov-Ungleichung)**.** *Sei Y eine reelle Zufallsvariable und $f\colon [0,\infty) \to [0,\infty)$ monoton wachsend mit $f(x) > 0$ für alle $x > 0$. Dann gilt für alle $\varepsilon > 0$*

$$\mathbb{P}\big(|Y| \ge \varepsilon\big) \le \frac{\mathbb{E}\big[f(|Y|)\big]}{f(\varepsilon)}.$$

Beweis. Da f monoton ist, ist $\{f \le c\}$ für jedes c ein Intervall, also eine Borel-Menge; daher ist $f(|Y|) \ge 0$ eine Zufallsvariable, und ihr Erwartungswert ist in $[0,\infty]$ definiert. Wegen der Monotonie von f gilt $f(\varepsilon)\mathbb{1}\{|Y| \ge \varepsilon\} \le f(|Y|)$, und mit der Monotonie des Erwartungswerts folgt

$$f(\varepsilon)\,\mathbb{P}\big(|Y| \ge \varepsilon\big) = \mathbb{E}\big[f(\varepsilon)\mathbb{1}\{|Y| \ge \varepsilon\}\big] \le \mathbb{E}\big[f(|Y|)\big]. \qquad \square$$

Korollar 5.4 (Chebyshev-Ungleichung)**.** *Für $Y \in \mathcal{L}^2$ und $\varepsilon > 0$ gilt*

$$\mathbb{P}\big(|Y - \mathbb{E}[Y]| \ge \varepsilon\big) \le \frac{\operatorname{Var}[Y]}{\varepsilon^2}.$$

Beweis. Markov-Ungleichung für $Y' \coloneqq Y - \mathbb{E}[Y]$ und $f(x) = x^2$. $\square$

Satz 5.5 (Schwaches Gesetz der großen Zahl in $\mathcal{L}^2$)**.** *Seien $(X_i)_{i\ge1}$ paarweise unkorrelierte Zufallsvariablen in $\mathcal{L}^2$ mit*

$$v \coloneqq \sup_{i\ge1} \operatorname{Var}[X_i] < \infty.$$

Dann gilt für alle $\varepsilon > 0$

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\big(X_i - \mathbb{E}[X_i]\big)\Big| \ge \varepsilon\Big) \le \frac{v}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0,$$

also $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \xrightarrow{\mathbb{P}} 0$.

Beweis. Sei $Y_n \coloneqq \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$. Dann ist $Y_n \in \mathcal{L}^2$, $\mathbb{E}[Y_n] = 0$, und nach Satz 4.16(1), (3)

$$\operatorname{Var}[Y_n] = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}[X_i] \le \frac{v}{n}.$$

Die Behauptung folgt durch Anwenden der Chebyshev-Ungleichung (Korollar 5.4). $\square$

Satz 5.6 (Schwaches Gesetz der großen Zahl in $\mathcal{L}^1$)**.** *Seien $(X_i)_{i\ge1}$ paarweise unabhängige und identisch verteilte Zufallsvariablen in $\mathcal{L}^1$. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_1].$$

Beweis. Wir zerlegen jedes X_i in einen beschränkten Anteil und einen Rest: Sei

$$X_i^\flat := X_i \mathbb{1}\{|X_i| \le i^{1/4}\} \quad \text{(Abschneiden)}, \qquad X_i^\sharp := X_i - X_i^\flat \quad \text{(Rest)},$$

und für $\star \in \{\flat, \sharp\}$

$$Y_n^\star := \frac{1}{n}\sum_{i=1}^n (X_i^\star - \mathbb{E}[X_i^\star]).$$

Wir zeigen $Y_n^\flat \xrightarrow{\mathbb{P}} 0$ und $Y_n^\sharp \xrightarrow{\mathbb{P}} 0$.

(1) Die $X_i^\flat$ sind als Funktionen der X_i ebenfalls paarweise unabhängig (Satz 3.16), also paarweise unkorreliert (Satz 4.16(4)), und als beschränkte Zufallsvariablen in $\mathcal{L}^2$. Wegen $|X_i^\flat - \mathbb{E}[X_i^\flat]| \le 2i^{1/4}$ ist

$$\operatorname{Var}[X_i^\flat] = \mathbb{E}\big[(X_i^\flat - \mathbb{E}[X_i^\flat])^2\big] \le 4i^{1/2},$$

also nach Satz 4.16(1), (3)

$$\operatorname{Var}[Y_n^\flat] = \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}[X_i^\flat] \le \frac{1}{n^2}\cdot n \cdot 4n^{1/2} = 4n^{-1/2} \xrightarrow[n\to\infty]{} 0,$$

und daher nach der Chebyshev-Ungleichung $Y_n^\flat \xrightarrow{\mathbb{P}} 0$.

(2) Da die X_i identisch verteilt sind, gilt

$$\mathbb{E}[|X_i^\sharp|] = \mathbb{E}[|X_i|] - \mathbb{E}[|X_i|\mathbb{1}\{|X_i| \le i^{1/4}\}] = \mathbb{E}[|X_1|] - \mathbb{E}[|X_1|\mathbb{1}\{|X_1| \le i^{1/4}\}] \xrightarrow[i\to\infty]{} 0$$

nach dem Satz von der monotonen Konvergenz (Satz 4.8). Wegen $|X_i^\sharp - \mathbb{E}[X_i^\sharp]| \le |X_i^\sharp| + \mathbb{E}|X_i^\sharp|$ folgt

$$\mathbb{E}[|Y_n^\sharp|] \le \frac{2}{n}\sum_{i=1}^n \mathbb{E}[|X_i^\sharp|] \xrightarrow[n\to\infty]{} 0,$$

denn die Cesàro-Mittel einer Nullfolge bilden eine Nullfolge. Mit der Markov-Ungleichung (Proposition 5.3 mit $f(x) = x$) folgt $Y_n^\sharp \xrightarrow{\mathbb{P}} 0$.

Wegen $X_i^\flat + X_i^\sharp = X_i$ und $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ ist

$$\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_1] = Y_n^\flat + Y_n^\sharp,$$

und die Aussage folgt mit dem folgenden Lemma 5.7. □

Lemma 5.7 (Stabilität der stochastischen Konvergenz)**.** *Seien Y_n, Z_n, $n \ge 1$, reelle Zufallsvariablen und $a_n \in \mathbb{R}$. Dann gilt:*

(1) Aus $Y_n \xrightarrow{\mathbb{P}} 0$ und $Z_n \xrightarrow{\mathbb{P}} 0$ folgt $Y_n + Z_n \xrightarrow{\mathbb{P}} 0$.

(2) Aus $Y_n \xrightarrow{\mathbb{P}} 0$ und der Beschränktheit von $(a_n)_{n\ge1}$ folgt $a_n Y_n \xrightarrow{\mathbb{P}} 0$.

Beweis. Sei $\varepsilon > 0$.

(1) Ist $|Y_n + Z_n| > \varepsilon$, so ist $|Y_n| > \varepsilon/2$ oder $|Z_n| > \varepsilon/2$. Also gilt

$$\mathbb{P}(|Y_n + Z_n| > \varepsilon) \le \mathbb{P}(|Y_n| > \tfrac{\varepsilon}{2}) + \mathbb{P}(|Z_n| > \tfrac{\varepsilon}{2}) \xrightarrow[n\to\infty]{} 0.$$

(2) Sei $c > 0$ mit $|a_n| \le c$ für alle n. Dann gilt

$$\mathbb{P}(|a_n Y_n| > \varepsilon) \le \mathbb{P}(|Y_n| > \tfrac{\varepsilon}{c}) \xrightarrow[n\to\infty]{} 0.$$ □

Das schwache Gesetz der großen Zahl hat zahlreiche Anwendungen, auch außerhalb der Stochastik. Wir betrachten drei Beispiele.

Beispiel 5.8 (Approximationssatz von Weierstraß)**.** Sei $f\colon [0,1] \to \mathbb{R}$ stetig. Lässt sich f gleichmäßig durch Polynome approximieren? Wir betrachten die Funktionen

$$f_n(p) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\binom{n}{k} p^k (1-p)^{n-k}, \qquad p \in [0,1],$$

die Polynome vom Grad höchstens n sind (*Bernstein-Polynome*). Ist $(X_i)_{i\geq 1}$ ein Bernoulli-Prozess zum Parameter p (unter dem W-Maß $\mathbb{P}_p$ mit Erwartungswert $\mathbb{E}_p$), so ist $\sum_{i=1}^n X_i$ binomialverteilt zu den Parametern n, p, und daher

$$f_n(p) = \mathbb{E}_p\Big[f\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)\Big].$$

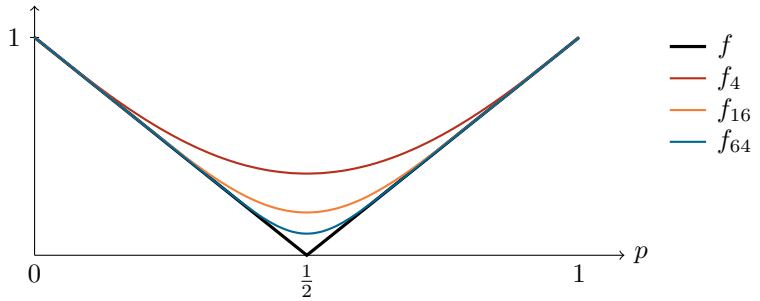
Sei $\varepsilon > 0$. Da f auf dem Kompaktum $[0,1]$ gleichmäßig stetig ist, gibt es ein $\delta > 0$ mit $|f(x) - f(y)| < \varepsilon$ für alle $x, y \in [0,1]$ mit $|x-y| < \delta$. Mit $\|f\| := \sup_{x\in[0,1]} |f(x)|$ folgt

$$\begin{aligned}
|f_n(p) - f(p)| &= \Big|\mathbb{E}_p\Big[f\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) - f(p)\Big]\Big| \leq \mathbb{E}_p\Big[\Big|f\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) - f(p)\Big|\Big] \\
&= \mathbb{E}_p\Big[\Big|f\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) - f(p)\Big|\Big(\mathbb{1}\Big\{\Big|\frac{1}{n}\sum_{i=1}^n X_i - p\Big| \geq \delta\Big\} + \mathbb{1}\Big\{\Big|\frac{1}{n}\sum_{i=1}^n X_i - p\Big| < \delta\Big\}\Big)\Big] \\
&\leq 2\|f\|\,\mathbb{P}_p\Big(\Big|\frac{1}{n}\sum_{i=1}^n X_i - p\Big| \geq \delta\Big) + \varepsilon.
\end{aligned}$$

Nach der Chebyshev-Ungleichung (Korollar 5.4) und wegen $\mathrm{Var}_p[\sum_{i=1}^n X_i] = np(1-p) \leq n/4$ ist das

$$\leq \varepsilon + \frac{2\|f\|\,np(1-p)}{n^2\delta^2} \leq \varepsilon + \frac{\|f\|}{2n\delta^2} \xrightarrow[n\to\infty]{} \varepsilon.$$

Da diese Abschätzung nicht von p abhängt und $\varepsilon > 0$ beliebig war, folgt $\sup_{p\in[0,1]} |f_n(p) - f(p)| \to 0$, d. h. $f_n \to f$ gleichmäßig (also bezüglich der Supremumsnorm). Insbesondere lässt sich jede stetige Funktion auf $[0,1]$ gleichmäßig durch Polynome approximieren.



Bernstein-Polynome f_n zu $f(p) = |2p-1|$ für $n = 4, 16, 64$ (exakt berechnet). Der Fehler ist dort am größten, wo f nicht glatt ist, und fällt bei $p = 1/2$ wie $n^{-1/2}$, im Einklang mit der Streuung $\sqrt{p(1-p)/n}$ von $\frac{1}{n}\sum_{i=1}^n X_i$.

Beispiel 5.9 (Monte-Carlo-Integration)**.** Gesucht ist eine Approximation für $\int_0^1 f(x)\,\mathrm{d}x$, wobei $f\colon [0,1] \to [0,c]$ messbar ist. Seien dazu $X_1, \dots, X_n$ unabhängig und gleichverteilt auf $[0,1]$. Die Idee ist

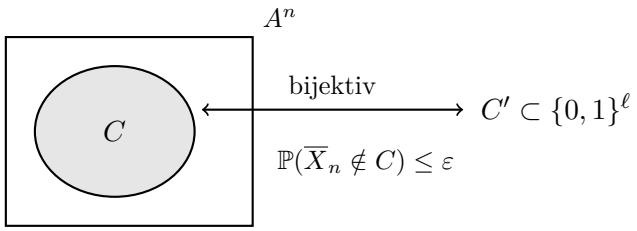
$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \approx \mathbb{E}[f(X_1)] = \int_0^1 f(x)\,\mathrm{d}x.$$

Da die $f(X_i)$ unabhängig sind und $\operatorname{Var}[f(X_1)] \le \mathbb{E}[f(X_1)^2] \le c^2$ gilt, liefert die Chebyshev-Ungleichung (Korollar 5.4) eine explizite Fehlerschranke:

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \int_0^1 f(x)\,\mathrm{d}x\Big| \ge \varepsilon\Big) \le \frac{\operatorname{Var}[f(X_1)]}{n\varepsilon^2} \le \frac{c^2}{n\varepsilon^2}.$$

Beispiel 5.10 (Quellencodierungssatz von Shannon)**.** Wir betrachten einen Nachrichtenblock $\overline{X}_n = (X_1, \dots, X_n)$, wobei die Buchstaben X_i unabhängig und identisch verteilt sind mit Werten in einem endlichen Alphabet A und Zähldichte ρ mit $\rho(a) > 0$ für alle $a \in A$. Der Nachrichtenblock soll in Binärdarstellung codiert werden. Wie viele Binärstellen müssen (asymptotisch in n) bereitgestellt werden, damit die gesamte Nachricht mit hoher Wahrscheinlichkeit fehlerfrei übertragen werden kann?

Dazu sucht man eine Menge $C \subset A^n$ von Nachrichten, die mit großer Wahrscheinlichkeit eintritt, und codiert sie mittels einer Bijektion $C \leftrightarrow C' \subset \{0,1\}^\ell$; für Nachrichten außerhalb von C nimmt man eine fehlerhafte Übertragung in Kauf.



Für $0 < \varepsilon < 1$ sei also

$$L(n,\varepsilon) := \min\{\ell \ge 0\colon\ \exists\, C \subset A^n \text{ mit } |C| \le 2^\ell \text{ und } \mathbb{P}(\overline{X}_n \in C) \ge 1-\varepsilon\}$$

die minimale Anzahl benötigter Binärstellen.

Satz. Für alle $0 < \varepsilon < 1$ gilt

$$\lim_{n\to\infty} \frac{L(n,\varepsilon)}{n} = H(\rho) := -\sum_{a\in A} \rho(a)\log_2 \rho(a).$$

Bemerkung. $H(\rho)$ heißt die *Entropie* der Zähldichte ρ. Es gilt $H(\rho) \ge 0$, da alle Summanden $-\rho(a)\log_2\rho(a) \ge 0$ sind, und (mit der Konvention $0\log_2 0 := 0$ auch für Zähldichten mit Nullstellen) $H(\rho) = 0$, falls $\rho(a) = 1$ für ein $a \in A$ gilt (keine Unsicherheit). Für welches ρ wird $H(\rho)$ maximal? Wir betrachten dazu Richtungsableitungen in Richtungen $\tilde\rho$ mit $\sum_{a\in A}\tilde\rho(a) = 0$ (damit $\rho + \varepsilon\tilde\rho$ für kleine ε wieder eine Zähldichte ist). Ist ρ eine Extremstelle mit $\rho(a) > 0$ für

alle a, so gilt (mit dem natürlichen Logarithmus, was nur einen konstanten Faktor ausmacht)

$$\begin{aligned}0 &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\sum_{a\in A}\Big[\big(\rho(a)+\varepsilon\tilde\rho(a)\big)\log\big(\rho(a)+\varepsilon\tilde\rho(a)\big)-\rho(a)\log\rho(a)\Big]\\ &= \sum_{a\in A}\rho(a)\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big[\log\big(\rho(a)+\varepsilon\tilde\rho(a)\big)-\log\rho(a)\Big]+\sum_{a\in A}\tilde\rho(a)\log\rho(a)\\ &= \underbrace{\sum_{a\in A}\rho(a)\frac{\tilde\rho(a)}{\rho(a)}}_{=0}+\sum_{a\in A}\tilde\rho(a)\log\rho(a).\end{aligned}$$

Wählt man $\tilde\rho=\delta_a-\delta_b$ für $a\neq b$ (d. h. $\tilde\rho(a)=1=-\tilde\rho(b)$, sonst 0), so folgt $\log\rho(a)=\log\rho(b)$, also $\rho(a)=\rho(b)$ für alle $a,b\in A$: Die einzige solche Extremstelle ist die Gleichverteilung, und da H konkav ist, wird dort das Maximum angenommen. Für die Zähldichte $\rho\equiv 1/|A|$ von $\mathcal{U}_A$ gilt

$$H(\rho)=\log_2|A| \qquad \text{(maximale Unsicherheit).}$$

Beweis des Satzes. Die Zufallsvariablen $Y_i\coloneqq-\log_2\rho(X_i)$, $i\geq 1$, sind unabhängig und identisch verteilt mit $\mathbb{E}[Y_i]=-\sum_{a\in A}\rho(a)\log_2\rho(a)=H(\rho)$; da A endlich ist, sind sie beschränkt, also in $\mathcal{L}^2$. Sei $0<\delta<1-\varepsilon$. Nach dem schwachen Gesetz der großen Zahl (Satz 5.5) gibt es ein n_0, sodass für alle $n\geq n_0$

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n Y_i-H(\rho)\Big|>\delta\Big)<\delta.$$

Schreiben wir $\rho^{\otimes n}(\omega)\coloneqq\prod_{i=1}^n\rho(\omega_i)$ für $\omega\in A^n$, so ist $\frac{1}{n}\sum_{i=1}^n Y_i=-(1/n)\log_2\rho^{\otimes n}(\overline{X}_n)$, und daher gilt für $n\geq n_0$

$$\mathbb{P}(\overline{X}_n\in B_n)>1-\delta \qquad \text{mit} \qquad B_n\coloneqq\big\{\omega\in A^n\colon\ 2^{-n(H(\rho)+\delta)}\leq\rho^{\otimes n}(\omega)\leq 2^{-n(H(\rho)-\delta)}\big\}.$$

Obere Abschätzung. Ist zusätzlich $\delta\leq\varepsilon$, so ist $C=B_n$ in der Definition von $L(n,\varepsilon)$ zulässig, also $L(n,\varepsilon)\leq\min\{\ell\colon|B_n|\leq 2^\ell\}$. Wegen

$$1=\sum_{\omega\in A^n}\rho^{\otimes n}(\omega)\geq\sum_{\omega\in B_n}\rho^{\otimes n}(\omega)\geq|B_n|\,2^{-n(H(\rho)+\delta)}$$

ist $|B_n|\leq 2^{n(H(\rho)+\delta)}$, also $L(n,\varepsilon)\leq n(H(\rho)+\delta)+1$. Es folgt $\limsup_{n\to\infty}L(n,\varepsilon)/n\leq H(\rho)+\delta$, und mit $\delta\downarrow 0$ erhalten wir $\limsup_{n\to\infty}L(n,\varepsilon)/n\leq H(\rho)$.

Untere Abschätzung. Sei $C\subset A^n$ eine Menge wie in der Definition von $L(n,\varepsilon)$, d. h. $|C|\leq 2^{L(n,\varepsilon)}$ und $\mathbb{P}(\overline{X}_n\in C)\geq 1-\varepsilon$. Für $\omega\in B_n$ ist $\rho^{\otimes n}(\omega)\,2^{n(H(\rho)-\delta)}\leq 1$, also

$$\begin{aligned}2^{L(n,\varepsilon)}\geq|C|\geq|C\cap B_n|&\geq\sum_{\omega\in C\cap B_n}\rho^{\otimes n}(\omega)\,2^{n(H(\rho)-\delta)}\\ &=\mathbb{P}(\overline{X}_n\in C\cap B_n)\,2^{n(H(\rho)-\delta)}\geq(1-\varepsilon-\delta)\,2^{n(H(\rho)-\delta)},\end{aligned}$$

denn $\mathbb{P}(\overline{X}_n\in C\cap B_n)=\mathbb{P}(\overline{X}_n\in C)-\mathbb{P}(\overline{X}_n\in C\setminus B_n)\geq\mathbb{P}(\overline{X}_n\in C)-\mathbb{P}(\overline{X}_n\notin B_n)\geq 1-\varepsilon-\delta$. Es folgt

$$L(n,\varepsilon)\geq\log_2(1-\varepsilon-\delta)+n\big(H(\rho)-\delta\big),$$

also $\liminf_{n\to\infty}L(n,\varepsilon)/n\geq H(\rho)-\delta$ und mit $\delta\downarrow 0$ schließlich $\liminf_{n\to\infty}L(n,\varepsilon)/n\geq H(\rho)$. □

5.1.2 Das starke Gesetz der großen Zahl

Definition 5.11 (Fast sichere Konvergenz von Zufallsvariablen)**.** Seien $Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen auf einem W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$. Man sagt, Y_n *konvergiert* $\mathbb{P}$*-fast sicher* gegen Y, falls

$$\mathbb{P}\Big(\{\omega \in \Omega\colon\ Y_n(\omega) \xrightarrow[n\to\infty]{} Y(\omega)\}\Big) = 1.$$

Bemerkung.

(1) $\{Y_n \to Y\}$ ist ein Ereignis, denn

$$\{Y_n \to Y\} = \bigcap_{k\ge 1} \bigcup_{m\ge 1} \bigcap_{n\ge m} \Big\{|Y_n - Y| < \frac{1}{k}\Big\} \in \mathcal{F}.$$

(2) Fast sichere Konvergenz impliziert stochastische Konvergenz. Denn sei $\varepsilon > 0$ und

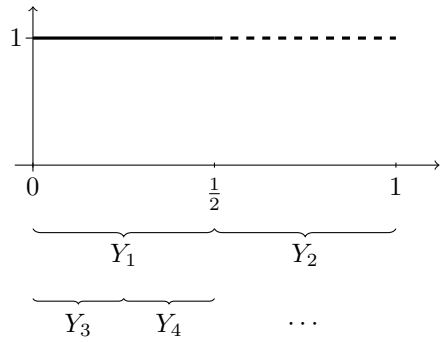
$$A_n := \Big\{\omega\colon\ \sup_{k\ge n} |Y_k(\omega) - Y(\omega)| > \varepsilon\Big\}.$$

Dann gilt $A_n \downarrow A := \{\omega\colon\ |Y_n(\omega) - Y(\omega)| > \varepsilon \text{ für unendlich viele } n\}$. Konvergiert $Y_n \to Y$ $\mathbb{P}$-fast sicher, so ist $\mathbb{P}(A) = 0$, und mit der σ-Stetigkeit (Satz 1.11(e)) folgt $\mathbb{P}(A_n) \to 0$. Wegen $\mathbb{P}(A_n) \ge \mathbb{P}(|Y_n - Y| > \varepsilon)$ folgt $\mathbb{P}(|Y_n - Y| > \varepsilon) \to 0$.

(3) Stochastische Konvergenz impliziert *nicht* fast sichere Konvergenz. Dazu betrachten wir den W-Raum $([0,1], \mathcal{B}_{[0,1]}, \mathcal{U}_{[0,1]})$ und die „Treppenfolge“

$$Y_1 = \mathbb{1}_{[0,\frac{1}{2}]}, \quad Y_2 = \mathbb{1}_{[\frac{1}{2},1]}, \quad Y_3 = \mathbb{1}_{[0,\frac{1}{4}]}, \quad Y_4 = \mathbb{1}_{[\frac{1}{4},\frac{1}{2}]}, \quad \dots,$$

allgemein $Y_n = \mathbb{1}_{[(j-1)2^{-k},\, j2^{-k}]}$ für $n = 2^k - 2 + j$ mit $k \ge 1$ und $1 \le j \le 2^k$.



Für $0 < \varepsilon < 1$ und $n = 2^k - 2 + j$ wie oben gilt $\mathbb{P}(|Y_n| > \varepsilon) = 2^{-k} \to 0$, also $Y_n \xrightarrow{\mathbb{P}} 0$. Aber für jedes $\omega \in [0,1]$ gilt $Y_n(\omega) = 1$ für unendlich viele n (in jeder „Stufe“ k mindestens einmal) und $Y_n(\omega) = 0$ für unendlich viele n, also $Y_n(\omega) \not\to 0$.

Gilt das Gesetz der großen Zahl also auch „ω-weise“, d. h. gilt $\frac{1}{n}\sum_{i=1}^n X_i \to \mathbb{E}[X_1]$ $\mathbb{P}$-fast sicher, wenn die X_i unabhängig und identisch verteilt sind? Die Antwort gibt der folgende Satz.

Satz 5.12 (Starkes Gesetz der großen Zahl)**.** *Seien $(X_i)_{i\ge 1}$ paarweise unkorrelierte Zufallsvariablen in $\mathcal{L}^2$ mit $v := \sup_{i\ge 1} \operatorname{Var}[X_i] < \infty$. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \xrightarrow[n\to\infty]{} 0 \qquad \mathbb{P}\text{-}\textit{fast sicher.}$$

Bemerkung. Auch eine $\mathcal{L}^1$-Version (für paarweise unabhängige, identisch verteilte $X_i \in \mathcal{L}^1$, vgl. Satz 5.6) ist korrekt; ohne Beweis.

Beweis. Ohne Einschränkung sei $\mathbb{E}[X_i] = 0$ für alle i (sonst ersetze X_i durch $X_i - \mathbb{E}[X_i]$). Sei $Y_n := \frac{1}{n}\sum_{i=1}^n X_i$.

1. Schritt: $Y_{n^2} \to 0$ $\mathbb{P}$-fast sicher. Wegen der paarweisen Unkorreliertheit ist $\operatorname{Var}[\sum_{i=1}^{n^2} X_i] = \sum_{i=1}^{n^2} \operatorname{Var}[X_i] \le n^2 v$. Für $\varepsilon > 0$ liefert die Chebyshev-Ungleichung (Korollar 5.4)

$$\mathbb{P}\big(|Y_{n^2}| \ge \varepsilon\big) \le \frac{n^2 v}{n^4 \varepsilon^2} = \frac{v}{n^2 \varepsilon^2},$$

und dies ist summierbar in n. Nach dem Lemma von Borel–Cantelli (Teil (a)) gilt also

$$\mathbb{P}\big(|Y_{n^2}| \ge \varepsilon \text{ für unendlich viele } n\big) = 0.$$

Damit folgt

$$\mathbb{P}\Big(\exists k \ge 1\colon\ |Y_{n^2}| \ge \tfrac{1}{k} \text{ für unendlich viele } n\Big) \le \sum_{k\ge 1} \mathbb{P}\Big(|Y_{n^2}| \ge \tfrac{1}{k} \text{ für unendlich viele } n\Big) = 0,$$

d. h. $Y_{n^2} \to 0$ $\mathbb{P}$-fast sicher.

2. Schritt: Kontrolle der Abweichungen zwischen n^2 und $(n+1)^2$. Für $m \in \mathbb{N}$ sei $n = n(m)$ diejenige Zahl mit $n^2 \le m < (n+1)^2$. Mit der Chebyshev-Ungleichung und der Unkorreliertheit folgt

$$\mathbb{P}\Big(\Big|\sum_{i=1}^m X_i - \sum_{i=1}^{n^2} X_i\Big| \ge \varepsilon n^2\Big) \le \frac{1}{\varepsilon^2 n^4} \operatorname{Var}\Big[\sum_{i=n^2+1}^m X_i\Big] \le \frac{(m-n^2)\,v}{\varepsilon^2 n^4}.$$

Daher ist

$$\begin{aligned}\sum_{m\ge 1} \mathbb{P}\Big(\Big|\sum_{i=1}^m X_i - \sum_{i=1}^{n(m)^2} X_i\Big| \ge \varepsilon\, n(m)^2\Big) &\le \frac{v}{\varepsilon^2} \sum_{m\ge 1} \frac{m - n(m)^2}{n(m)^4} = \frac{v}{\varepsilon^2} \sum_{n\ge 1} \sum_{m=n^2}^{(n+1)^2-1} \frac{m-n^2}{n^4} \\ &= \frac{v}{\varepsilon^2} \sum_{n\ge 1} \frac{1}{n^4} \sum_{k=1}^{2n} k = \frac{v}{\varepsilon^2} \sum_{n\ge 1} \frac{2n(2n+1)}{2n^4} < \infty.\end{aligned}$$

Mit Borel–Cantelli folgt wie im 1. Schritt

$$\mathbb{P}\Big(\Big|\frac{1}{n(m)^2}\sum_{i=1}^m X_i - Y_{n(m)^2}\Big| \xrightarrow[m\to\infty]{} 0\Big) = 1.$$

Zusammen mit dem 1. Schritt (beachte $n(m) \to \infty$) folgt

$$\mathbb{P}\Big(\frac{1}{n(m)^2}\sum_{i=1}^m X_i \xrightarrow[m\to\infty]{} 0\Big) = 1.$$

Wegen $n(m)^2 \le m < (n(m)+1)^2$ gilt $m/n(m)^2 \to 1$, und da $Y_m = n(m)^2/m \cdot \frac{1}{n(m)^2}\sum_{i=1}^m X_i$, folgt

$$\mathbb{P}\Big(Y_m \xrightarrow[m\to\infty]{} 0\Big) = 1. \qquad \square$$

Beispiel 5.13 (Fast alle reellen Zahlen sind normal; Borel 1909)**.** Sei $q \geq 2$ eine natürliche Zahl. Jedes $x \in [0,1]$ besitzt eine q-adische Darstellung

$$x = \sum_{i \geq 1} x_i q^{-i} \qquad \text{mit Ziffern } x_i \in \{0, \dots, q-1\}$$

(die für abzählbar viele x nicht eindeutig ist; diese Nullmenge spielt im Folgenden keine Rolle). Die Zahl x heißt *normal* (zur Basis q), falls für jedes $k \geq 1$ jede Ziffernfolge $a = (a_1, \dots, a_k) \in \{0, \dots, q-1\}^k$ in den Nachkommastellen von x mit derselben asymptotischen Häufigkeit vorkommt, d. h. falls

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{(x_i, \dots, x_{i+k-1}) = a\} = q^{-k} \qquad \text{für alle } k \geq 1,\ a \in \{0, \dots, q-1\}^k. \tag{5.1}$$

(1) Es sind nur wenige explizite Beispiele normaler Zahlen bekannt, etwa (zur Basis 10)

$$0{,}123456789101112131\dots \qquad \text{oder} \qquad 0{,}235711131719\dots$$

(Aneinanderreihung der natürlichen Zahlen bzw. der Primzahlen). Ob etwa π oder e normal sind, ist unbekannt.

(2) Überabzählbar viele Zahlen sind nicht normal, z. B. alle Elemente der Cantor-Menge (deren Ziffern zur Basis 3 nie den Wert 1 annehmen).

(3) *Satz von Borel:* Fast alle $x \in [0,1]$ (bezüglich des Lebesgue-Maßes) sind normal, und zwar zu jeder Basis q.

Beweis von (3). Wir betrachten den W-Raum $([0,1], \mathcal{B}_{[0,1]}, \mathcal{U}_{[0,1]})$ und fassen die Ziffern x_i als Zufallsvariablen auf; diese sind unabhängig und gleichverteilt auf $\{0, \dots, q-1\}$, denn für alle $n \geq 1$ und $a_1, \dots, a_n \in \{0, \dots, q-1\}$ ist $\{x_1 = a_1, \dots, x_n = a_n\}$ bis auf eine Nullmenge ein Intervall der Länge q^{-n}, also $\mathcal{U}_{[0,1]}(x_1 = a_1, \dots, x_n = a_n) = q^{-n} = \prod_{i=1}^{n} \frac{1}{q}$. Seien $k \geq 1$ und $a \in \{0, \dots, q-1\}^k$ fest. Die Blöcke

$$\overline{X}_i := (x_i, \dots, x_{i+k-1}), \qquad i \geq 1,$$

sind gleichverteilt auf $\{0, \dots, q-1\}^k$, aber *nicht* unabhängig. Jedoch ist $\overline{X}_1$ unabhängig von $\overline{X}_{k+1}$, $\overline{X}_2$ von $\overline{X}_{k+2}$ usw., denn diese Blöcke überlappen sich nicht. Genauer ist für jedes $j \in \{1, \dots, k\}$ die Folge $(\overline{X}_{ik+j})_{i \geq 0}$ unabhängig und identisch verteilt. Die Indikatoren $\mathbb{1}\{\overline{X}_{ik+j} = a\}$, $i \geq 0$, sind daher unabhängig (insbesondere paarweise unkorreliert), beschränkt und haben Erwartungswert q^{-k}. Nach dem starken Gesetz der großen Zahl (Satz 5.12) hat das Ereignis

$$C := \bigcap_{j=1}^{k} \Big\{ \lim_{m \to \infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{1}\{\overline{X}_{ik+j} = a\} = q^{-k} \Big\}$$

die Wahrscheinlichkeit 1. Auf C gilt

$$\frac{1}{mk} \sum_{i=1}^{mk} \mathbb{1}\{\overline{X}_i = a\} = \frac{1}{k} \sum_{j=1}^{k} \underbrace{\frac{1}{m} \sum_{i=0}^{m-1} \mathbb{1}\{\overline{X}_{ik+j} = a\}}_{\to\, q^{-k}} \xrightarrow[m \to \infty]{} q^{-k}.$$

Für beliebiges $n \geq k$ sei $m = m(n) \geq 1$ mit $mk \leq n < (m+1)k$. Dann gilt

$$\frac{m}{m+1} \cdot \frac{1}{mk} \sum_{i=1}^{mk} \mathbb{1}\{\overline{X}_i = a\} \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\overline{X}_i = a\} \leq \frac{m+1}{m} \cdot \frac{1}{(m+1)k} \sum_{i=1}^{(m+1)k} \mathbb{1}\{\overline{X}_i = a\},$$

und beide Schranken konvergieren auf C gegen q^{-k}, da $m/(m+1) \to 1$. Also gilt die Bedingung in (5.1) für dieses k und a mit Wahrscheinlichkeit 1. Da es nur abzählbar viele k, a (und Basen q) gibt, gilt (5.1) für alle k und a (und alle q) ebenfalls mit Wahrscheinlichkeit 1. □

5.2 Die Normalapproximation der Binomialverteilung

Sei $(X_i)_{i\geq 1}$ ein Bernoulli-Prozess zum Parameter $0 < p < 1$, und sei

$$S_n = \sum_{i=1}^{n} X_i$$

die Anzahl der Erfolge bis zur Zeit n; S_n ist also $\mathcal{B}_{n,p}$-verteilt. Wie verhalten sich die Fluktuationen $S_n - \mathbb{E}[S_n] = S_n - np$ für $n \to \infty$? Aus dem schwachen Gesetz der großen Zahl wissen wir, dass $\mathbb{P}(|S_n - np| \geq \varepsilon n) \to 0$, d. h. Abweichungen der Größenordnung n werden unwahrscheinlich. Gesucht ist eine Folge (a_n), für die $\mathbb{P}(|S_n - np| \leq a_n)$ einen nichttrivialen Limes hat.

Bemerkung 5.14 (Bernoulli-Fluktuationen). Es gilt

$$\mathbb{P}(|S_n - np| \leq a_n) \xrightarrow[n\to\infty]{} \begin{cases} 1, & \text{falls } a_n/\sqrt{n} \to \infty, \\ 0, & \text{falls } a_n/\sqrt{n} \to 0. \end{cases}$$

Denn:

(1) Nach der Chebyshev-Ungleichung gilt

$$\mathbb{P}(|S_n - np| > a_n) \leq \frac{np(1-p)}{a_n^2} \xrightarrow[n\to\infty]{} 0, \qquad \text{falls } a_n/\sqrt{n} \to \infty.$$

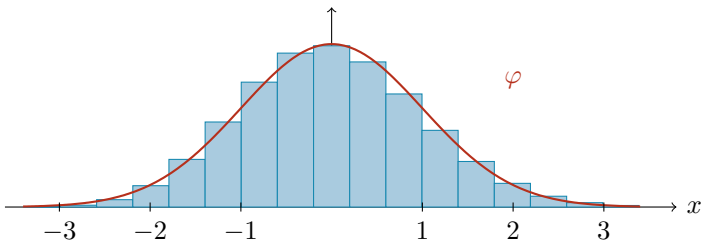
(2) Es gilt

$$\mathbb{P}(|S_n - np| \leq a_n) = \sum_{k:\ |k-np|\leq a_n} \mathcal{B}_{n,p}(\{k\}) \leq (2a_n + 1) \max_{k:\ |k-np|\leq a_n} \mathcal{B}_{n,p}(\{k\}).$$

Ist $a_n/\sqrt{n} \to 0$, so gilt für alle k mit $|k - np| \leq a_n$ und hinreichend große n, dass $k(n-k)/n \geq np(1-p)/2$ (da $k/n \to p$ gleichmäßig). Mit der Stirling-Formel (vgl. Beispiel 5.1) und der Abschätzung $\mathrm{e}^{-nh(k/n)} \leq 1$ aus dem folgenden Beweis von Satz 5.15 folgt $\mathcal{B}_{n,p}(\{k\}) \leq c/\sqrt{n}$ für eine Konstante $c > 0$ und alle diese k. (Genauer gilt sogar $\max_k \mathcal{B}_{n,p}(\{k\}) \sim 1/\sqrt{2\pi np(1-p)}$.) Also ist

$$\mathbb{P}(|S_n - np| \leq a_n) \leq (2a_n + 1)\frac{c}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0, \qquad \text{falls } a_n/\sqrt{n} \to 0.$$

Die interessante Größenordnung ist also $a_n = c\sqrt{n}$, und wir betrachten $\mathbb{P}(|S_n - np| \leq c\sqrt{n})$. Die Antwort wird durch die Gauß'sche Glockenkurve gegeben.



Für $n = 30$, $p = 0{,}3$: über jedem Punkt $x_n(k)$ ein Balken der Breite $1/\sqrt{np(1-p)}$ und der Höhe $\sqrt{np(1-p)}\,\mathcal{B}_{n,p}(\{k\})$, also mit Fläche $\mathcal{B}_{n,p}(\{k\})$, im Vergleich zur Dichte φ (exakt berechnet).

Satz 5.15 (de Moivre–Laplace)**.** *Sei $0 < p < 1$ und $c > 0$. Sei*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}, \qquad x \in \mathbb{R},$$

die Dichte der Standardnormalverteilung $\mathcal{N}_{0,1}$, und für $k \in \{0, \dots, n\}$ sei

$$x_n(k) = \frac{k - np}{\sqrt{np(1-p)}}$$

die Standardisierung von k. Dann gilt

$$\lim_{n\to\infty} \max_{k:\ |x_n(k)|\le c} \Big| \frac{\sqrt{np(1-p)}\ \mathcal{B}_{n,p}(\{k\})}{\varphi(x_n(k))} - 1 \Big| = 0.$$

Beweis. Wir zeigen

$$\mathcal{B}_{n,p}(\{k\}) \underset{n\to\infty}{\sim} \frac{1}{\sqrt{np(1-p)}}\,\varphi(x_n(k))$$

gleichmäßig in allen k mit $|x_n(k)| \le c$, d. h. für $|k - np| \le c\sqrt{np(1-p)}$. Für solche k gilt $k \to \infty$ und $n - k \to \infty$ gleichmäßig, sodass wir auf $n!$, $k!$ und $(n-k)!$ die Stirling-Formel $m! \sim \sqrt{2\pi m}\, m^m \mathrm{e}^{-m}$ anwenden können:

$$\begin{aligned}\mathcal{B}_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} &\underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \Big(\frac{np}{k}\Big)^k \Big(\frac{n(1-p)}{n-k}\Big)^{n-k} \\ &\underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi np(1-p)}}\,\mathrm{e}^{-nh(k/n)}\end{aligned}$$

mit

$$h(s) = s \log \frac{s}{p} + (1-s) \log \frac{1-s}{1-p}, \qquad 0 < s < 1.$$

Dabei haben wir benutzt:

(1) Es gilt

$$\begin{aligned}\frac{k(n-k)}{n^2 p(1-p)} &= \frac{\big((k-np)+np\big)\big(n(1-p)-(k-np)\big)}{n^2 p(1-p)} \\ &= 1 + \underbrace{\frac{k-np}{np}}_{\to 0} - \underbrace{\frac{k-np}{n(1-p)}}_{\to 0} - \underbrace{\frac{(k-np)^2}{n^2 p(1-p)}}_{\to 0},\end{aligned}$$

gleichmäßig in k, da $|k - np| \le c\sqrt{np(1-p)}$. Also ist $\sqrt{n/(k(n-k))} \sim 1/\sqrt{np(1-p)}$.

(2) Es gilt

$$\begin{aligned}\Big(\frac{np}{k}\Big)^k \Big(\frac{n(1-p)}{n-k}\Big)^{n-k} &= \exp\Big\{-k \log \frac{k}{np} - (n-k) \log \frac{n-k}{n(1-p)}\Big\} \\ &= \exp\Big\{-n\Big(\frac{k}{n} \log \frac{k/n}{p} + \Big(1 - \frac{k}{n}\Big) \log \frac{1-k/n}{1-p}\Big)\Big\} = \mathrm{e}^{-nh(k/n)}.\end{aligned}$$

Übrigens ist $h \ge 0$, denn wegen der Konkavität des Logarithmus (Jensen) ist $-h(s) = s \log(p/s) + (1-s) \log((1-p)/(1-s)) \le \log(p + 1 - p) = 0$; also ist stets $\mathrm{e}^{-nh(k/n)} \le 1$.

Wegen $|k/n-p| \le c\sqrt{p(1-p)}/\sqrt{n}$ gilt $k/n \to p$ gleichmäßig, und wir entwickeln h um p nach Taylor:

$$h\Big(\frac{k}{n}\Big) = h(p) + h'(p)\Big(\frac{k}{n}-p\Big) + \frac{h''(p)}{2}\Big(\frac{k}{n}-p\Big)^2 + O\Big(\Big|\frac{k}{n}-p\Big|^3\Big).$$

Hier ist $h(p)=0$,

$$h'(s) = \log\frac{s}{p} + 1 - \log\frac{1-s}{1-p} - 1 = \log\frac{s}{p} - \log\frac{1-s}{1-p}, \qquad \text{also } h'(p)=0,$$

$$h''(s) = \frac{1}{s} + \frac{1}{1-s} = \frac{1}{s(1-s)}, \qquad \text{also } h''(p) = \frac{1}{p(1-p)},$$

und h''' ist in einer Umgebung von p beschränkt. Wegen $|k/n-p|^3 = O(n^{-3/2})$ (gleichmäßig in k) folgt

$$h\Big(\frac{k}{n}\Big) = \frac{1}{2p(1-p)}\Big(\frac{k}{n}-p\Big)^2 + O(n^{-3/2}), \qquad \text{also} \qquad nh\Big(\frac{k}{n}\Big) = \frac{x_n(k)^2}{2} + O(n^{-1/2}).$$

Insgesamt erhalten wir gleichmäßig in k mit $|x_n(k)| \le c$

$$\sqrt{np(1-p)}\,\mathcal{B}_{n,p}(\{k\}) \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x_n(k)^2/2+O(n^{-1/2})} \underset{n\to\infty}{\sim} \varphi(x_n(k)). \qquad \square$$

Korollar 5.16. *Sei $0<p<1$. Dann gilt für alle $0 \le k \le l \le n$*

$$\mathcal{B}_{n,p}(\{k,\dots,l\}) = \Phi\Big(\frac{l+\frac12-np}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\frac{k-\frac12-np}{\sqrt{np(1-p)}}\Big) + \delta_n(k,l),$$

wobei

$$\Phi(c) = \int_{-\infty}^{c} \varphi(x)\,\mathrm{d}x, \qquad c \in \mathbb{R},$$

die Verteilungsfunktion der Standardnormalverteilung ist und

$$\max_{0\le k\le l\le n} |\delta_n(k,l)| \xrightarrow[n\to\infty]{} 0.$$

Die Terme $\pm 1/2$ heißen Stetigkeitskorrektur*; ihr Einfluss verschwindet für $n\to\infty$.*

Beweis. Wir schreiben $\sigma_n := \sqrt{np(1-p)}$ und verwenden die Standardisierung $x_n(t) = (t-np)/\sigma_n$ aus Satz 5.15 auch für nicht ganzzahlige t. Für alle j sei

$$\Delta_j\Phi := \Phi\big(x_n(j+\tfrac12)\big) - \Phi\big(x_n(j-\tfrac12)\big)$$

das Gauß-Integral über das Intervall $[x_n(j-1/2), x_n(j+1/2)]$ der Länge $1/\sigma_n$. Dann gilt für alle $c>0$

$$\max_{j:\ |x_n(j)|\le c} \Big|\frac{\Delta_j\Phi}{\sigma_n^{-1}\varphi(x_n(j))} - 1\Big| \xrightarrow[n\to\infty]{} 0. \tag{5.2}$$

Denn es ist

$$\frac{\Delta_j\Phi}{\sigma_n^{-1}\varphi(x_n(j))} = \sigma_n \int_{x_n(j)-\frac{1}{2\sigma_n}}^{x_n(j)+\frac{1}{2\sigma_n}} \exp\Big(\frac{(x_n(j)+x)(x_n(j)-x)}{2}\Big)\,\mathrm{d}x,$$

und für $|x_n(j)| \le c$ und x im Integrationsbereich ist

$$\Big|\frac{(x_n(j)+x)(x_n(j)-x)}{2}\Big| \le \frac12\Big(2c + \frac{1}{2\sigma_n}\Big)\frac{1}{2\sigma_n} \xrightarrow[n\to\infty]{} 0.$$

Der Integrand konvergiert also gleichmäßig gegen 1, und da das Integrationsintervall die Länge $1/\sigma_n$ hat, folgt (5.2).

Zusammen mit Satz 5.15 folgt: Zu $\varepsilon > 0$ und $c > 0$ gibt es ein $n(\varepsilon, c)$, sodass

$$\max_{j:\ |x_n(j)|\le c} \left| \frac{\mathcal{B}_{n,p}(\{j\})}{\Delta_j \Phi} - 1 \right| \le \varepsilon \qquad \text{für alle } n \ge n(\varepsilon, c).$$

Da sich die $\Delta_j \Phi$ teleskopisch aufsummieren, gilt

$$\delta_n(k,l) = \sum_{j=k}^{l} (\mathcal{B}_{n,p}(\{j\}) - \Delta_j \Phi).$$

Für $0 \le k \le l \le n$ mit $|x_n(j)| \le c$ für alle $j \in \{k, \dots, l\}$ folgt daher für $n \ge n(\varepsilon, c)$

$$|\delta_n(k,l)| \le \sum_{j=k}^{l} |\mathcal{B}_{n,p}(\{j\}) - \Delta_j \Phi| \le \varepsilon \underbrace{\sum_{j=k}^{l} \Delta_j \Phi}_{\le 1} \le \varepsilon. \tag{5.3}$$

Es bleiben die Schwänze zu kontrollieren. Wähle $c > 0$ so groß, dass $\Phi\big(c - 1/\sqrt{p(1-p)}\big) > 1 - \varepsilon$ (also auch $\Phi\big(-c + 1/\sqrt{p(1-p)}\big) < \varepsilon$), und setze

$$k_c := \lceil np - c\sigma_n \rceil, \qquad l_c := \lfloor np + c\sigma_n \rfloor.$$

Sei ferner $n_0 \ge n(\varepsilon, c)$ so groß, dass $np - c\sigma_n \ge 0$, $np + c\sigma_n \le n$ und $2c\sigma_n \ge 1$ für alle $n \ge n_0$ gilt (möglich, da $\sigma_n \to \infty$, $\sigma_n/n \to 0$ und $0 < p < 1$). Für $n \ge n_0$ ist dann $0 \le k_c \le l_c \le n$. Außerdem ist $x_n(k_c) \ge -c$ und $x_n(l_c) \le c$, also $x_n(j) \in [-c, c]$ für alle $j \in \{k_c, \dots, l_c\}$. Ferner gilt $l_c + 1/2 \ge np + c\sigma_n - 1/2$, also wegen $\sigma_n \ge \sqrt{p(1-p)}$

$$x_n\big(l_c + \tfrac{1}{2}\big) \ge c - \frac{1}{2\sigma_n} \ge c - \frac{1}{\sqrt{p(1-p)}}, \qquad \text{analog} \qquad x_n\big(k_c - \tfrac{1}{2}\big) \le -c + \frac{1}{\sqrt{p(1-p)}}.$$

Mit (5.3) folgt für $n \ge n_0$

$$\mathcal{B}_{n,p}(\{k_c, \dots, l_c\}) \ge \Phi\big(x_n(l_c + \tfrac{1}{2})\big) - \Phi\big(x_n(k_c - \tfrac{1}{2})\big) - \varepsilon > (1 - \varepsilon) - \varepsilon - \varepsilon = 1 - 3\varepsilon.$$

Die Binomialverteilung hat also außerhalb von $\{k_c, \dots, l_c\}$ höchstens Masse 3ε, und die Normalverteilung hat außerhalb von $[x_n(k_c - 1/2), x_n(l_c + 1/2)]$ höchstens Masse 2ε. Für beliebige $0 \le k \le l \le n$ zerlegen wir die Summe für $\delta_n(k,l)$ in die Indizes $j \in \{k_c, \dots, l_c\}$ und die übrigen und erhalten mit (5.3)

$$|\delta_n(k,l)| \le \varepsilon + \sum_{j \notin \{k_c, \dots, l_c\}} \mathcal{B}_{n,p}(\{j\}) + \sum_{j \notin \{k_c, \dots, l_c\}} \Delta_j \Phi < \varepsilon + 3\varepsilon + 2\varepsilon = 6\varepsilon$$

für alle $n \ge n_0$. Da $\varepsilon > 0$ beliebig war, folgt die Behauptung. □

Korollar 5.17 (Zentraler Grenzwertsatz für Bernoulli-Folgen)**.** *Sei $(X_i)_{i \ge 1}$ ein Bernoulli-Prozess zum Parameter $p \in (0,1)$, und seien*

$$S_n = \sum_{i=1}^{n} X_i \qquad \textit{und} \qquad S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Dann gilt für alle $a < b$

$$\lim_{n \to \infty} \mathbb{P}(a \le S_n^* \le b) = \int_a^b \varphi(x)\, \mathrm{d}x = \Phi(b) - \Phi(a).$$

Die Konvergenz ist sogar gleichmäßig in $a < b$; insbesondere gilt $\lim_{n\to\infty} \sup_{b \in \mathbb{R}} |\mathbb{P}(S_n^* \le b) - \Phi(b)| = 0$, d. h. die Verteilungsfunktionen konvergieren gleichmäßig.

Beweis. Mit $\sigma_n = \sqrt{np(1-p)}$ gilt, da S_n nur Werte in $\{0,\dots,n\}$ annimmt, $\mathbb{P}(a \le S_n^* \le b) = \mathcal{B}_{n,p}(\{k_{n,a},\dots,l_{n,b}\})$ mit

$$k_{n,a} = \max\{\lceil np + \sigma_n a\rceil, 0\}, \qquad l_{n,b} = \min\{\lfloor np + \sigma_n b\rfloor, n\}$$

(und $\{k_{n,a},\dots,l_{n,b}\} = \emptyset$ für $k_{n,a} > l_{n,b}$). Wegen $x_n(0) = -\sqrt{np/(1-p)} \to -\infty$ und $x_n(n) = \sqrt{n(1-p)/p} \to \infty$ gilt

$$\gamma_n := \frac{1}{\sqrt{2\pi}\,\sigma_n} + \Phi\big(x_n(0)\big) + 1 - \Phi\big(x_n(n)\big) \xrightarrow[n\to\infty]{} 0.$$

Wir zeigen

$$\big|\Phi(x_n(k_{n,a} - \tfrac12)) - \Phi(a)\big| \le \gamma_n \qquad \text{und} \qquad \big|\Phi(x_n(l_{n,b} + \tfrac12)) - \Phi(b)\big| \le \gamma_n. \tag{5.4}$$

Ist $k_{n,a} = \lceil np + \sigma_n a\rceil$, so gilt

$$\big|x_n(k_{n,a} - \tfrac12) - a\big| = \Big|\frac{k_{n,a} - \frac12 - np}{\sigma_n} - a\Big| = \Big|\frac{\lceil np + a\sigma_n\rceil - (np + a\sigma_n) - \frac12}{\sigma_n}\Big| \le \frac{1}{\sigma_n},$$

und nach dem Mittelwertsatz (beachte $\Phi' = \varphi \le 1/\sqrt{2\pi}$) folgt die erste Ungleichung in (5.4). Andernfalls ist $k_{n,a} = 0$ und $np + \sigma_n a < 0$, d. h. $a < x_n(0)$; da auch $x_n(-\frac12) < x_n(0)$ ist, liegen beide Werte $\Phi(x_n(-\frac12))$ und $\Phi(a)$ in $[0, \Phi(x_n(0))]$, und ihr Abstand ist höchstens $\Phi(x_n(0)) \le \gamma_n$. Die zweite Ungleichung folgt analog (mit $x_n(n)$ statt $x_n(0)$).

Ist $k_{n,a} \le l_{n,b}$, so ist $0 \le k_{n,a} \le l_{n,b} \le n$, und mit Korollar 5.16 und (5.4) folgt

$$\begin{aligned}
&\Big|\mathcal{B}_{n,p}(\{k_{n,a},\dots,l_{n,b}\}) - (\Phi(b) - \Phi(a))\Big| \\
&\quad\le \Big|\mathcal{B}_{n,p}(\{k_{n,a},\dots,l_{n,b}\}) - \Big(\Phi(x_n(l_{n,b} + \tfrac12)) - \Phi(x_n(k_{n,a} - \tfrac12))\Big)\Big| \\
&\qquad + \big|\Phi(x_n(l_{n,b} + \tfrac12)) - \Phi(b)\big| + \big|\Phi(x_n(k_{n,a} - \tfrac12)) - \Phi(a)\big| \\
&\quad\le \max_{0\le k\le l\le n} |\delta_n(k,l)| + 2\gamma_n \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

Ist $k_{n,a} > l_{n,b}$, so ist die linke Seite gleich $\Phi(b) - \Phi(a)$, und es gilt ebenfalls $\Phi(b) - \Phi(a) \le \gamma_n$: Ist $l_{n,b} < 0$, so ist $b < x_n(0)$ und $\Phi(b) - \Phi(a) \le \Phi(x_n(0))$; ist $k_{n,a} > n$, so ist $a > x_n(n)$ und $\Phi(b) - \Phi(a) \le 1 - \Phi(x_n(n))$; andernfalls ist $\lceil np + \sigma_n a\rceil > \lfloor np + \sigma_n b\rfloor$, d. h. das Intervall $[np + \sigma_n a, np + \sigma_n b]$ enthält keine ganze Zahl, also ist $b - a < 1/\sigma_n$ und $\Phi(b) - \Phi(a) \le 1/(\sqrt{2\pi}\,\sigma_n)$. Die Schranke $\max_{0\le k\le l\le n} |\delta_n(k,l)| + 2\gamma_n$ hängt nicht von a, b ab; daraus folgt die gleichmäßige Konvergenz. Lässt man schließlich $a \to -\infty$ gehen, so folgt auch die gleichmäßige Konvergenz der Verteilungsfunktionen. □

5.3 Der zentrale Grenzwertsatz

Kernaussage: Die Normalapproximation im Satz von de Moivre–Laplace gilt nicht nur für Bernoulli-Prozesse, sondern für beliebige u. i. v. Prozesse in $\mathcal{L}^2$ und sogar noch allgemeiner.

Zur Erinnerung: Die Verteilungsfunktion einer reellen Zufallsvariablen X ist $F_X(c) := \mathbb{P}(X \le c)$, $c \in \mathbb{R}$.

Definition 5.18 (Verteilungskonvergenz)**.** Sei $(Y_n)_{n\geq 1}$ eine Folge reellwertiger Zufallsvariablen, die nicht notwendigerweise auf demselben W-Raum definiert sind. Man sagt, Y_n *konvergiert in Verteilung* gegen die reelle Zufallsvariable Y bzw. gegen deren Verteilung Q auf $(\mathbb{R}, \mathcal{B})$, falls

$$F_{Y_n}(c) \xrightarrow[n\to\infty]{} F_Y(c) = F_Q(c)$$

für alle $c \in \mathbb{R}$ gilt, an denen F_Y stetig ist. Man schreibt dann

$$Y_n \xrightarrow{\mathcal{L}} Y \text{ (bzw. } Q) \qquad \text{oder} \qquad Y_n \xrightarrow{d} Y \text{ (bzw. } Q).$$

[$\mathcal{L}$ steht für „law“, d für „distribution“.]

Satz 5.19 (Charakterisierung der Verteilungskonvergenz)**.** *In der Situation von Definition 5.18 sind die folgenden Aussagen äquivalent:*

(a) $Y_n \xrightarrow{\mathcal{L}} Y$.

(b) $\mathbb{E}[f(Y_n)] \to \mathbb{E}[f(Y)]$ *für alle stetigen und beschränkten Funktionen* $f\colon \mathbb{R} \to \mathbb{R}$.

Ist F_Y stetig, so ist ebenfalls äquivalent:

(c) F_{Y_n} *konvergiert gleichmäßig gegen* F_Y*, d. h.*

$$\|F_{Y_n} - F_Y\| := \sup_{c\in\mathbb{R}} |F_{Y_n}(c) - F_Y(c)| \xrightarrow[n\to\infty]{} 0.$$

Beweis. (a)⇒(b): Sei f stetig und beschränkt, $\|f\| := \sup_x |f(x)|$, und sei $\varepsilon > 0$. Wähle $c_0 < c_1 < \cdots < c_k$ mit

(i) c_i ist eine Stetigkeitsstelle von F_Y für alle $i \in \{0, \dots, k\}$,

(ii) $F_Y(c_0) < \varepsilon$ und $F_Y(c_k) > 1 - \varepsilon$,

(iii) $|f(x) - f(c_i)| < \varepsilon$ für alle $i \in \{1, \dots, k\}$ und alle $x \in [c_{i-1}, c_i]$.

Das ist möglich, denn F_Y ist monoton und hat deshalb höchstens abzählbar viele Unstetigkeitsstellen (die Stetigkeitsstellen liegen also dicht), und f ist auf jedem Kompaktum sogar gleichmäßig stetig. Damit gilt

$$\begin{aligned}
\mathbb{E}[f(Y_n)] &= \sum_{i=1}^{k} \mathbb{E}\big[f(Y_n)\mathbb{1}\{c_{i-1} < Y_n \leq c_i\}\big] + \mathbb{E}\big[f(Y_n)\mathbb{1}\{Y_n \in \mathbb{R} \setminus (c_0, c_k]\}\big] \\
&\leq \sum_{i=1}^{k} \big(f(c_i) + \varepsilon\big)\big[F_{Y_n}(c_i) - F_{Y_n}(c_{i-1})\big] + \|f\|\big(F_{Y_n}(c_0) + 1 - F_{Y_n}(c_k)\big) \\
&\xrightarrow[n\to\infty]{} \sum_{i=1}^{k} \big(f(c_i) + \varepsilon\big)\big[F_Y(c_i) - F_Y(c_{i-1})\big] + \|f\|\big(F_Y(c_0) + 1 - F_Y(c_k)\big),
\end{aligned}$$

und der letzte Summand ist $\leq 2\varepsilon\|f\|$. Ferner gilt nach (iii)

$$\begin{aligned}
\sum_{i=1}^{k} \big(f(c_i) + \varepsilon\big)\big[F_Y(c_i) - F_Y(c_{i-1})\big] &= \sum_{i=1}^{k} \mathbb{E}\big[\big(f(c_i) + \varepsilon\big)\mathbb{1}\{c_{i-1} < Y \leq c_i\}\big] \\
&\leq \sum_{i=1}^{k} \mathbb{E}\big[\big(f(Y) + 2\varepsilon\big)\mathbb{1}\{c_{i-1} < Y \leq c_i\}\big] \\
&\leq \mathbb{E}[f(Y)] + \|f\|\,\mathbb{P}(Y \notin (c_0, c_k]) + 2\varepsilon\,\mathbb{P}(c_0 < Y \leq c_k) \\
&\leq \mathbb{E}[f(Y)] + 2\varepsilon\|f\| + 2\varepsilon.
\end{aligned}$$

Also ist

$$\limsup_{n\to\infty} \mathbb{E}[f(Y_n)] \le \mathbb{E}[f(Y)] + 2\varepsilon(1+2\|f\|) \xrightarrow[\varepsilon\downarrow 0]{} \mathbb{E}[f(Y)].$$

Mit $-f$ statt f folgt $\liminf_{n\to\infty} \mathbb{E}[f(Y_n)] \ge \mathbb{E}[f(Y)]$, also (b).

(b)⇒(a): Sei $c \in \mathbb{R}$ und $\delta > 0$. Wähle eine stetige und beschränkte Funktion f mit

$$\mathbb{1}\{x \le c\} \le f(x) \le \mathbb{1}\{x \le c+\delta\} \qquad \text{für alle } x \in \mathbb{R}$$

(etwa $f = 1$ auf $(-\infty, c]$, $f = 0$ auf $[c+\delta, \infty)$ und linear dazwischen). Dann gilt

$$\limsup_{n\to\infty} F_{Y_n}(c) = \limsup_{n\to\infty} \mathbb{P}(Y_n \le c) \le \lim_{n\to\infty} \mathbb{E}[f(Y_n)] = \mathbb{E}[f(Y)] \le F_Y(c+\delta).$$

Ist c eine Stetigkeitsstelle von F_Y, so folgt $F_Y(c+\delta) \to F_Y(c)$ für $\delta \downarrow 0$, also

$$\limsup_{n\to\infty} F_{Y_n}(c) \le F_Y(c).$$

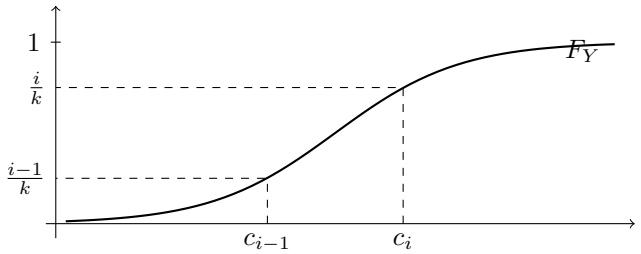
Ebenso folgt mit einer stetigen Funktion f mit $\mathbb{1}\{x \le c-\delta\} \le f(x) \le \mathbb{1}\{x \le c\}$, dass

$$\liminf_{n\to\infty} F_{Y_n}(c) \ge F_Y(c-\delta) \xrightarrow[\delta\downarrow 0]{} F_Y(c).$$

(a)⇒(c): Sei F_Y stetig, $k \in \mathbb{N}$ und $\varepsilon = 1/k$. Da F_Y stetig ist, gibt es nach dem Zwischenwertsatz $c_i \in \mathbb{R}$ mit

$$F_Y(c_i) = \frac{i}{k} \qquad \text{für alle } i \in \{1, \dots, k-1\};$$

setze $c_0 := -\infty$, $c_k := +\infty$ (mit $F_Y(-\infty) := F_{Y_n}(-\infty) := 0$ und $F_Y(+\infty) := F_{Y_n}(+\infty) := 1$). Die c_i zerlegen $\mathbb{R}$ in k Intervalle, auf denen F_Y jeweils um genau ε wächst.



Sei

$$(*) := \|F_{Y_n} - F_Y\| = \sup_{c\in\mathbb{R}} |F_{Y_n}(c) - F_Y(c)|.$$

Wir nutzen, dass F_{Y_n} und F_Y monoton wachsend sind. Sei $c \in \mathbb{R}$.

- Falls $F_{Y_n}(c) \ge F_Y(c)$, betrachte das kleinste i mit $c_i \ge c$. Ist $i = k$, so gilt $F_{Y_n}(c) - F_Y(c) \le 1 - F_Y(c_{k-1}) = \varepsilon$. Sonst ist $c_{i-1} < c \le c_i$, und es folgt

$$F_{Y_n}(c) - F_Y(c) \le F_{Y_n}(c_i) - F_Y(c_{i-1}) = F_{Y_n}(c_i) - F_Y(c_i) + \varepsilon \le |F_{Y_n}(c_i) - F_Y(c_i)| + \varepsilon.$$

- Falls $F_{Y_n}(c) < F_Y(c)$, betrachte das größte i mit $c_i < c$. Ist $i = 0$, so gilt $F_Y(c) - F_{Y_n}(c) \le F_Y(c_1) = \varepsilon$. Sonst ist $c_i < c \le c_{i+1}$, und es folgt

$$F_Y(c) - F_{Y_n}(c) \le F_Y(c_{i+1}) - F_{Y_n}(c_i) = F_Y(c_i) - F_{Y_n}(c_i) + \varepsilon \le |F_{Y_n}(c_i) - F_Y(c_i)| + \varepsilon.$$

Also ist

$$(*) \le \varepsilon + \max_{0<i<k} \left|F_{Y_n}(c_i) - F_Y(c_i)\right| \xrightarrow[n\to\infty]{} \varepsilon$$

nach (a), da alle c_i Stetigkeitsstellen von F_Y sind. Da $\varepsilon = 1/k$ beliebig klein gewählt werden kann, folgt $(*) \to 0$.

(c)⇒(a) gilt, da aus gleichmäßiger Konvergenz die punktweise Konvergenz $F_{Y_n}(c) \to F_Y(c)$ für alle $c \in \mathbb{R}$ folgt. □

Satz 5.20 (Zentraler Grenzwertsatz)**.** *Sei $(X_i)_{i\ge1}$ eine u. i. v. Folge reeller Zufallsvariablen in $\mathcal{L}^2$ mit $\mathbb{E}[X_i] = m$ und $\operatorname{Var}[X_i] = v > 0$. Dann gilt*

$$S_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sqrt{v}} \xrightarrow{\mathcal{L}} \mathcal{N}_{0,1},$$

also (nach Satz 5.19, da Φ stetig ist) $\|F_{S_n^*} - \Phi\| \to 0$ für $n \to \infty$.

Bemerkung 5.21. (1) *Warum gerade $\mathcal{N}_{0,1}$?* Die Standardnormalverteilung hat eine Fixpunkteigenschaft (Faltungsinvarianz): Sind die X_i u. i. v. mit $X_i \sim \mathcal{N}_{0,1}$, so ist auch

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \sim \mathcal{N}_{0,1} \qquad \text{für alle } n \ge 1.$$

Die Normalverteilung ist die einzige Verteilung mit Erwartungswert 0 und Varianz 1 mit dieser Eigenschaft. (Das folgt sogar aus Satz 5.20: Hat Q diese Eigenschaft, so ist $S_n^* \sim Q$ für alle n, und zugleich $S_n^* \xrightarrow{\mathcal{L}} \mathcal{N}_{0,1}$, also $Q = \mathcal{N}_{0,1}$.)

(2) Ohne die Voraussetzung $X_i \in \mathcal{L}^2$ ist die Aussage im Allgemeinen falsch. Sind z. B. die X_i u. i. v. Cauchy-verteilt zum Parameter $a > 0$, also mit Dichte

$$\rho(x) = \frac{1}{\pi} \frac{a}{x^2 + a^2}, \qquad x \in \mathbb{R},$$

und Verteilungsfunktion $F_{X_i}(c) = 1/2 + (1/\pi)\arctan(c/a)$, so ist $\sum_{i=1}^n X_i$ Cauchy-verteilt zum Parameter na (Übung), also

$$S_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \sim \text{Cauchy}(a\sqrt{n}), \qquad F_{S_n^*}(c) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{c}{a\sqrt{n}} \xrightarrow[n\to\infty]{} \frac{1}{2},$$

und S_n^* konvergiert nicht in Verteilung gegen $\mathcal{N}_{0,1}$. Hier ist sogar $X_i \notin \mathcal{L}^1$. Aber auch $X_i \in \mathcal{L}^1$ allein reicht nicht aus, und ebenso genügt es nicht, statt der Unabhängigkeit nur paarweise Unabhängigkeit oder Unkorreliertheit vorauszusetzen (ohne Beweis).

(3) Falls $X_i \in \mathcal{L}^3$, so gibt es eine Abschätzung für die Konvergenzgeschwindigkeit, den *Satz von Berry–Esséen* (ohne Beweis):

$$\|F_{S_n^*} - \Phi\| \le 0{,}8 \, \frac{\mathbb{E}\left[|X_1 - \mathbb{E}[X_1]|^3\right]}{v^{3/2}} \cdot \frac{1}{\sqrt{n}}.$$

Beweis von Satz 5.20. Ohne Einschränkung sei $m = 0$ und $v = 1$ (sonst ersetze X_i durch $(X_i - m)/\sqrt{v}$). Nach Satz 5.19 genügt es zu zeigen, dass

$$\mathbb{E}[f(S_n^*)] \xrightarrow[n\to\infty]{} \mathbb{E}_{\mathcal{N}_{0,1}}[f] \qquad \text{für alle stetigen und beschränkten } f.$$

Aus dem Beweis von (b)⇒(a) in Satz 5.19 geht hervor, dass es sogar genügt, diese Konvergenz für zweimal differenzierbare f zu zeigen, für die f, f' und f'' beschränkt sind und f'' gleichmäßig stetig ist. (Dort kann man nämlich statt der stückweise linearen auch eine entsprechende glatte Funktion f wählen.)

Sei $(Y_i)_{i\geq 1}$ eine u. i. v. Folge mit $Y_1 \sim \mathcal{N}_{0,1}$, die unabhängig von $(X_i)_{i\geq 1}$ ist (gegebenenfalls auf einem Produktraum). Nach Bemerkung 5.21(1) ist

$$T_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \sim \mathcal{N}_{0,1}.$$

Es genügt also zu zeigen:

$$\lim_{n\to\infty} \Big| \mathbb{E}\big[f(S_n^*) - f(T_n^*)\big] \Big| = 0.$$

Seien

$$X_{i,n} := \frac{X_i}{\sqrt{n}}, \qquad Y_{i,n} := \frac{Y_i}{\sqrt{n}}, \qquad W_{i,n} := \sum_{j=1}^{i-1} Y_{j,n} + \sum_{j=i+1}^{n} X_{j,n}.$$

Wegen $W_{i,n}+X_{i,n} = W_{i-1,n}+Y_{i-1,n}$ für $1 < i \leq n$ sowie $W_{1,n}+X_{1,n} = S_n^*$ und $W_{n,n}+Y_{n,n} = T_n^*$ erhalten wir die Teleskopsumme

$$f(S_n^*) - f(T_n^*) = \sum_{i=1}^n \Big[f(W_{i,n} + X_{i,n}) - f(W_{i,n} + Y_{i,n}) \Big].$$

Nach Taylor gilt

$$f(W_{i,n} + X_{i,n}) = f(W_{i,n}) + f'(W_{i,n})X_{i,n} + \frac{1}{2} f''(W_{i,n}) X_{i,n}^2 + R_{X,i,n}$$

mit dem (Lagrange-)Restglied

$$R_{X,i,n} = \frac{1}{2} X_{i,n}^2 \big[f''(W_{i,n} + \theta X_{i,n}) - f''(W_{i,n}) \big]$$

für ein (zufälliges) $\theta \in [0,1]$. Somit gilt $|R_{X,i,n}| \leq X_{i,n}^2 \|f''\|$. Da f'' gleichmäßig stetig ist, gibt es außerdem zu $\varepsilon > 0$ ein $\delta > 0$ mit $|f''(x) - f''(y)| \leq \varepsilon$ für $|x-y| \leq \delta$, also $|R_{X,i,n}| \leq (1/2) X_{i,n}^2 \varepsilon$ für $|X_{i,n}| \leq \delta$. Zusammen:

$$|R_{X,i,n}| \leq X_{i,n}^2 \Big[\frac{\varepsilon}{2} \mathbb{1}\{|X_{i,n}| \leq \delta\} + \|f''\| \, \mathbb{1}\{|X_{i,n}| > \delta\} \Big].$$

Analoges gilt für $f(W_{i,n} + Y_{i,n})$ mit Restglied $R_{Y,i,n}$. Damit folgt

$$\Big| \mathbb{E}\big[f(S_n^*) - f(T_n^*)\big] \Big| \leq \sum_{i=1}^n \mathbb{E}\big[|R_{X,i,n}| + |R_{Y,i,n}| \big], \tag{5.5}$$

denn die übrigen Terme heben sich im Erwartungswert weg:

(1) Terme mit f: $\mathbb{E}\big[f(W_{i,n}) - f(W_{i,n})\big] = 0$.

(2) Terme mit f': Da $W_{i,n}$ und $X_{i,n}$ unabhängig sind, gilt

$$\mathbb{E}\big[f'(W_{i,n}) X_{i,n}\big] = \mathbb{E}\big[f'(W_{i,n})\big] \, \mathbb{E}[X_{i,n}] = 0,$$

und ebenso für $Y_{i,n}$.

(3) Terme mit f'' (außer dem Restglied): Ebenso wegen der Unabhängigkeit

$$\mathbb{E}[f''(W_{i,n})(X_{i,n}^2 - Y_{i,n}^2)] = \mathbb{E}[f''(W_{i,n})]\Big(\underbrace{\mathbb{E}[X_{i,n}^2]}_{=1/n} - \underbrace{\mathbb{E}[Y_{i,n}^2]}_{=1/n}\Big) = 0.$$

Die rechte Seite von (5.5) ist

$$\begin{aligned}
&\leq \sum_{i=1}^{n}\Big[\frac{\varepsilon}{2}\mathbb{E}[X_{i,n}^2 + Y_{i,n}^2] + \|f''\|\,\mathbb{E}[X_{i,n}^2\mathbb{1}\{|X_{i,n}| > \delta\} + Y_{i,n}^2\mathbb{1}\{|Y_{i,n}| > \delta\}]\Big] \\
&= \varepsilon + \|f''\|\,\mathbb{E}[X_1^2\mathbb{1}\{|X_1| > \delta\sqrt{n}\} + Y_1^2\mathbb{1}\{|Y_1| > \delta\sqrt{n}\}].
\end{aligned}$$

Dabei ist

$$\mathbb{E}[X_1^2\mathbb{1}\{|X_1| > \delta\sqrt{n}\}] = 1 - \mathbb{E}[X_1^2\mathbb{1}\{|X_1| \leq \delta\sqrt{n}\}] \xrightarrow[n\to\infty]{} 0,$$

denn $X_1^2\mathbb{1}\{|X_1| \leq \delta\sqrt{n}\} \uparrow X_1^2$ und $\mathbb{E}[X_1^2] = 1$ (σ-Additivität bzw. monotone Konvergenz); ebenso für Y_1. Es folgt

$$\limsup_{n\to\infty}\Big|\mathbb{E}\big[f(S_n^*) - f(T_n^*)\big]\Big| \leq \varepsilon \qquad \text{für alle } \varepsilon > 0,$$

also die Behauptung. □

Bemerkung. Falls $\mathbb{E}[|X_1|^3] < \infty$ und f dreimal differenzierbar mit beschränkter Ableitung f''' ist, so gilt $|R_{X,i,n}| \leq \|f'''\|\,|X_{i,n}|^3/6$ (und ebenso für Y), und (5.5) ist

$$\leq \frac{C\,\|f'''\|}{\sqrt{n}} \qquad \text{mit } C = \frac{\mathbb{E}[|X_1|^3] + \mathbb{E}[|Y_1|^3]}{6}.$$

5.4 Normal- versus Poisson-Approximation

Betrachte einen Bernoulli-Prozess $(X_i)_{1\leq i\leq n}$ zum Parameter p und setze

$$Y_i^{(n)} = \frac{X_i - p}{\sqrt{np(1-p)}}.$$

Dann ist $\sum_{i=1}^n Y_i^{(n)}$ nach Korollar 5.17 näherungsweise $\mathcal{N}_{0,1}$-verteilt. Hier sind alle Summanden klein: $|Y_i^{(n)}| \leq 1/\sqrt{np(1-p)} \to 0$, aber sie sind nicht selten von 0 verschieden, denn $\mathbb{P}(Y_i^{(n)} \neq 0) = 1$.

Andererseits ist $\sum_{i=1}^n X_i^{(n)}$ näherungsweise Pois_λ-verteilt, wenn $(X_i^{(n)})_{1\leq i\leq n}$ ein Bernoulli-Prozess zum Parameter p_n mit $np_n \to \lambda$ ist (Poisson'scher Grenzwertsatz). Hier sind die Summanden nicht klein, $|X_i^{(n)}| \not\to 0$, aber selten von 0 verschieden: $\mathbb{P}(X_i^{(n)} \neq 0) = p_n \to 0$.

Satz 5.22 (Poisson-Approximation der Binomialverteilung)**.** *Für $0 < p < 1$ und $n \in \mathbb{N}$ gilt für den* Variationsabstand *von $\mathcal{B}_{n,p}$ und* Pois_{np}

$$\|\mathcal{B}_{n,p} - \mathrm{Pois}_{np}\| := \sum_{k\geq 0}\big|\mathcal{B}_{n,p}(\{k\}) - \mathrm{Pois}_{np}(\{k\})\big| \leq 2np^2.$$

Beweis mittels Kopplung. Es ist

$$S \coloneqq \sum_{i=1}^{n} X_i \sim \mathcal{B}_{n,p}, \qquad T \coloneqq \sum_{i=1}^{n} Y_i \sim \mathrm{Pois}_{np},$$

wenn die X_i u. i. v. Bernoulli-verteilt zum Parameter p und die Y_i u. i. v. Pois_p-verteilt sind. *Kopplung:* Wir wählen einen gemeinsamen W-Raum, auf dem X_i *nicht* unabhängig von Y_i ist, sondern möglichst nah bei Y_i liegt.

Seien dazu $Z_1, \dots, Z_n$ u. i. v. Zufallsvariablen mit Werten in $\{-1, 0, 1, 2, \dots\}$ und

$$\mathbb{P}(Z_i = k) = \mathrm{Pois}_p(\{k\}) \ \ \forall\, k \geq 1, \qquad \mathbb{P}(Z_i = 0) = 1 - p, \qquad \mathbb{P}(Z_i = -1) = \mathrm{e}^{-p} - (1-p),$$

wobei $\mathcal{B}_{1,p}(\{0\}) = 1 - p \leq \mathrm{e}^{-p} = \mathrm{Pois}_p(\{0\})$ benutzt wurde (die Gewichte sind also nichtnegativ und summieren sich zu 1). Wir setzen

$$X_i \coloneqq \mathbb{1}\{Z_i \neq 0\}, \qquad Y_i \coloneqq \max\{Z_i, 0\}.$$

Dann sind die X_i u. i. v. Bernoulli-verteilt zum Parameter p (denn $\mathbb{P}(Z_i \neq 0) = p$) und die Y_i u. i. v. Pois_p-verteilt (denn $\mathbb{P}(Y_i = 0) = \mathbb{P}(Z_i \in \{-1, 0\}) = \mathrm{e}^{-p}$). Betrachte

$$D \coloneqq \{\exists\, i \leq n \text{ mit } X_i \neq Y_i\} = \bigcup_{i \leq n} \{Z_i \notin \{0, 1\}\}.$$

Dann gilt $S = T$ auf D^{c}, und damit

$$\begin{aligned}
\|\mathcal{B}_{n,p} - \mathrm{Pois}_{np}\| &= \sum_{k \geq 0} \big|\mathbb{P}(S = k) - \mathbb{P}(T = k)\big| = \sum_{k \geq 0} \big|\mathbb{P}(\{S = k\} \cap D) - \mathbb{P}(\{T = k\} \cap D)\big| \\
&\leq \sum_{k \geq 0} \big(\mathbb{P}(\{S = k\} \cap D) + \mathbb{P}(\{T = k\} \cap D)\big) = 2\mathbb{P}(D) \leq 2 \sum_{i=1}^{n} \mathbb{P}(Z_i \notin \{0, 1\}) \\
&= 2n\big(1 - p\mathrm{e}^{-p} - (1-p)\big) = 2np\big(1 - \mathrm{e}^{-p}\big) \leq 2np^2,
\end{aligned}$$

denn $1 - \mathrm{e}^{-p} \leq p$. □

Bemerkung 5.23. (1) Der Beweis lässt sich leicht erweitern auf unabhängige $X_i \sim \mathcal{B}_{1,p_i}$ und $\mathrm{Pois}_{\sum_{i=1}^n p_i}$, mit der Schranke $2\sum_{i=1}^n p_i^2$ für den Variationsabstand.

(2) Für alle $\lambda, \delta > 0$ gilt $\|\mathrm{Pois}_{\lambda+\delta} - \mathrm{Pois}_\lambda\| \leq 2\delta$. Denn ist $X \sim \mathrm{Pois}_\lambda$ unabhängig von $Y \sim \mathrm{Pois}_\delta$, so ist $X + Y \sim \mathrm{Pois}_{\lambda+\delta}$, und $X + Y$ und X unterscheiden sich nur auf $\{Y \geq 1\}$. Wie im Beweis von Satz 5.22 folgt

$$\|\mathrm{Pois}_{\lambda+\delta} - \mathrm{Pois}_\lambda\| \leq 2\mathbb{P}(Y \geq 1) = 2\big(1 - \mathrm{e}^{-\delta}\big) \leq 2\delta.$$

(3) Wann sollte die Binomialverteilung durch die Normalverteilung approximiert werden, und wann durch die Poisson-Verteilung?

(i) Die Poisson-Approximation ist gut, falls np^2 klein ist.

(ii) Die Normalapproximation ist gut, falls die Schranke im Satz von Berry–Esséen (Bemerkung 5.21(3)) klein ist. Für $X_1 \sim \mathcal{B}_{1,p}$ ist

$$\frac{\mathbb{E}\big[|X_1 - \mathbb{E}[X_1]|^3\big]}{v^{3/2}} \cdot \frac{1}{\sqrt{n}} = \frac{p(1-p)^3 + (1-p)p^3}{\sqrt{n}\,\big(p(1-p)\big)^{3/2}} = \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}},$$

und da $1/2 \leq p^2 + (1-p)^2 \leq 1$, ist dies genau dann klein, wenn $np(1-p)$ groß ist.

5.5 Große Abweichungen

Unter bestimmten Voraussetzungen ist die Konvergenz im schwachen Gesetz der großen Zahl sogar exponentiell schnell.

Satz 5.24 (Große Abweichungen)**.** *Seien $(X_i)_{i\geq 1}$ u. i. v. reellwertige Zufallsvariablen mit Dichte f, deren* logarithmische momenterzeugende Funktion

$$\Lambda(t) \coloneqq \log \mathbb{E}\big[\mathrm{e}^{tX_1}\big] = \log \int_{\mathbb{R}} \mathrm{e}^{tx} f(x)\,\mathrm{d}x$$

für alle $t \in \mathbb{R}$ endlich ist. Sei

$$\Lambda^*(x) \coloneqq \sup_{t\in\mathbb{R}} \{tx - \Lambda(t)\}, \qquad x \in \mathbb{R},$$

die Legendre-Transformierte *von Λ, und sei $S_n \coloneqq \sum_{i=1}^n X_i$. Dann gilt für alle $x > \mathbb{E}[X_1]$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq xn) = -\Lambda^*(x).$$

Bemerkung. Satz 5.24 besagt insbesondere: Für jedes $x > \mathbb{E}[X_1]$ und jedes $\varepsilon > 0$ gibt es ein $N \in \mathbb{N}$, so dass

$$\mathrm{e}^{-(\Lambda^*(x)+\varepsilon)n} \leq \mathbb{P}(S_n \geq xn) \leq \mathrm{e}^{-(\Lambda^*(x)-\varepsilon)n} \qquad \text{für alle } n \geq N.$$

Wählt man in der Definition von Λ^* speziell $t = 0$, so erhält man

$$\Lambda^*(x) \geq 0 \cdot x - \Lambda(0) = -\log \mathbb{E}[\mathrm{e}^0] = 0.$$

Man kann zeigen, dass $\Lambda^*(x) = 0$ nur für $x = \mathbb{E}[X_1]$ gilt (ohne Beweis). Für $x > \mathbb{E}[X_1]$ ist also $\Lambda^*(x) > 0$, d. h. die Wahrscheinlichkeiten $\mathbb{P}(S_n \geq xn)$ fallen exponentiell schnell ab. In diesem Sinne ist die Konvergenz im schwachen Gesetz der großen Zahl sogar exponentiell schnell.

Beweis von Satz 5.24. Reduktion. Wir können ohne Einschränkung $\mathbb{E}[X_1] < 0$ und $x = 0$ annehmen. Andernfalls betrachten wir $\tilde{X}_i \coloneqq X_i - x$; dann ist $\mathbb{E}[\tilde{X}_1] = \mathbb{E}[X_1] - x < 0$, und für die zugehörigen Funktionen gilt

$$\tilde{\Lambda}(t) = \log \mathbb{E}\big[\mathrm{e}^{t(X_1-x)}\big] = \Lambda(t) - tx, \qquad \tilde{\Lambda}^*(0) = \sup_{t\in\mathbb{R}} \{-\Lambda(t) + tx\} = \Lambda^*(x),$$

sowie $\{S_n \geq xn\} = \{\tilde{S}_n \geq 0\}$ mit $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_i$.

Sei also $\mathbb{E}[X_1] < 0$ und $x = 0$. Wir setzen

$$M(t) \coloneqq \mathbb{E}\big[\mathrm{e}^{tX_1}\big] = \mathrm{e}^{\Lambda(t)} \in (0, \infty), \qquad r \coloneqq \mathrm{e}^{-\Lambda^*(0)} = \mathrm{e}^{\inf_{t\in\mathbb{R}} \Lambda(t)} = \inf_{t\in\mathbb{R}} M(t).$$

Da M auf ganz $\mathbb{R}$ endlich ist, darf man (mit majorisierter Konvergenz) unter dem Erwartungswert differenzieren und erhält

$$M'(t) = \mathbb{E}\big[X_1 \mathrm{e}^{tX_1}\big], \qquad M''(t) = \mathbb{E}\big[X_1^2 \mathrm{e}^{tX_1}\big] > 0,$$

denn wegen $\mathbb{E}[X_1] < 0$ ist X_1 nicht fast sicher gleich 0. Also ist M strikt konvex mit $M'(0) = \mathbb{E}[X_1] < 0$.

Fall 1: $\mathbb{P}(X_1 \le 0) = 1$. Dann ist $M'(t) = \mathbb{E}[X_1 \mathrm{e}^{tX_1}] < 0$ für alle $t \in \mathbb{R}$, d. h. M ist strikt fallend, und mit majorisierter Konvergenz folgt

$$r = \lim_{t\to\infty} M(t) = \lim_{t\to\infty} \mathbb{E}\big[\mathrm{e}^{tX_1}\big] = \mathbb{P}(X_1 = 0).$$

Da alle $X_i \le 0$ fast sicher sind, gilt $\{S_n \ge 0\} = \{X_1 = \cdots = X_n = 0\}$ fast sicher, also wegen der Unabhängigkeit

$$\mathbb{P}(S_n \ge 0) = \mathbb{P}(X_1 = \cdots = X_n = 0) = r^n,$$

und somit $(1/n)\log \mathbb{P}(S_n \ge 0) = \log r = -\Lambda^*(0)$ für alle n (mit $\log 0 := -\infty$).

Fall 2: $\mathbb{P}(X_1 < 0) > 0$ *und* $\mathbb{P}(X_1 > 0) > 0$. Dann gilt $\lim_{t\to\infty} M(t) = \infty = \lim_{t\to-\infty} M(t)$. Da M stetig und strikt konvex ist, besitzt M eine eindeutige Minimalstelle $\tau \in \mathbb{R}$, d. h.

$$M(\tau) = r \qquad \text{und} \qquad M'(\tau) = 0,$$

und wegen $M'(0) < 0$ und der Monotonie von M' ist $\tau > 0$.

Obere Abschätzung. Mit der Markov-Ungleichung und der Unabhängigkeit der X_i folgt

$$\mathbb{P}(S_n \ge 0) = \mathbb{P}(\mathrm{e}^{\tau S_n} \ge 1) \le \mathbb{E}\big[\mathrm{e}^{\tau S_n}\big] = M(\tau)^n = r^n,$$

also

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \ge 0) \le \log r = -\Lambda^*(0).$$

Untere Abschätzung. Es bleibt

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \ge 0) \ge -\Lambda^*(0)$$

zu zeigen. Wir nehmen dazu an, dass X_1 eine Dichte f besitzt. Die *Idee* ist, die Verteilung so abzuändern, dass der Wert $x = 0$ für die neuen Zufallsvariablen zum typischen Wert, nämlich zum Erwartungswert, wird. Genauer seien $\tilde{X}_1, \tilde{X}_2, \ldots$ u. i. v. Zufallsvariablen mit der Dichte

$$g(x) := r^{-1}\mathrm{e}^{\tau x} f(x), \qquad x \in \mathbb{R}.$$

Dann gilt:

(1) g ist eine Dichte, denn $g \ge 0$ und $\int_{-\infty}^{\infty} g(x)\,\mathrm{d}x = r^{-1}\mathbb{E}\big[\mathrm{e}^{\tau X_1}\big] = r^{-1}M(\tau) = 1$.

(2) Für $\tilde{M}(t) := \mathbb{E}\big[\mathrm{e}^{t\tilde{X}_1}\big]$ gilt $\tilde{M}(t) = r^{-1}\mathbb{E}\big[\mathrm{e}^{(t+\tau)X_1}\big] = r^{-1}M(t+\tau) < \infty$ für alle $t \in \mathbb{R}$.

(3) $\mathbb{E}[\tilde{X}_1] = \tilde{M}'(0) = r^{-1}M'(\tau) = 0$; d. h. $x = 0$ ist nun der Erwartungswert.

(4) $\tilde{v} := \mathrm{Var}[\tilde{X}_1] = \tilde{M}''(0) - \tilde{M}'(0)^2 = r^{-1}M''(\tau) \in (0, \infty)$.

Mit $\tilde{S}_n := \sum_{i=1}^n \tilde{X}_i$ erhalten wir

$$\begin{aligned}
\mathbb{P}(S_n \ge 0) &= \int \mathrm{d}x_1 \cdots \mathrm{d}x_n\, f(x_1)\cdots f(x_n)\, \mathbb{1}\{x_1 + \cdots + x_n \ge 0\} \\
&= \int \mathrm{d}x_1 \cdots \mathrm{d}x_n\, g(x_1)\cdots g(x_n)\, r^n \mathrm{e}^{-\tau\sum_{i=1}^n x_i}\, \mathbb{1}\{x_1 + \cdots + x_n \ge 0\} \\
&= r^n\, \mathbb{E}\Big[\mathrm{e}^{-\tau\tilde{S}_n} \mathbb{1}\{\tilde{S}_n \ge 0\}\Big].
\end{aligned}$$

Wegen $\log r = -\Lambda^*(0)$ genügt es also zu zeigen, dass

$$\liminf_{n\to\infty} \underbrace{\frac{1}{n}\log \mathbb{E}\Big[\mathrm{e}^{-\tau\tilde{S}_n}\mathbb{1}\{\tilde{S}_n \geq 0\}\Big]}_{=:a_n} \geq 0.$$

Sei dazu $c > 0$. Da $\tau > 0$, ist $\mathrm{e}^{-\tau\tilde{S}_n} \geq \mathrm{e}^{-\tau c\sqrt{n}}$ auf $\{0 \leq \tilde{S}_n \leq c\sqrt{n}\}$, und daher

$$\begin{aligned} a_n &\geq \frac{1}{n}\log \mathbb{E}\Big[\mathrm{e}^{-\tau\tilde{S}_n}\mathbb{1}\{0 \leq \tilde{S}_n \leq c\sqrt{n}\}\Big] \geq \frac{1}{n}\log\Big(\mathrm{e}^{-\tau c\sqrt{n}}\,\mathbb{P}\Big(\frac{\tilde{S}_n}{\sqrt{n}} \in [0,c]\Big)\Big) \\ &= -\frac{\tau c}{\sqrt{n}} + \frac{1}{n}\log \mathbb{P}\Big(\frac{\tilde{S}_n}{\sqrt{\tilde{v}n}} \in \Big[0, \frac{c}{\sqrt{\tilde{v}}}\Big]\Big). \end{aligned}$$

Der erste Summand konvergiert gegen 0. Nach dem zentralen Grenzwertsatz (Satz 5.20), angewandt auf die u. i. v. Folge $(\tilde{X}_i)$ mit Erwartungswert 0 und Varianz $\tilde{v} > 0$, gilt

$$\mathbb{P}\Big(\frac{\tilde{S}_n}{\sqrt{\tilde{v}n}} \in \Big[0, \frac{c}{\sqrt{\tilde{v}}}\Big]\Big) \xrightarrow[n\to\infty]{} \Phi\Big(\frac{c}{\sqrt{\tilde{v}}}\Big) - \frac{1}{2} > 0,$$

so dass auch der zweite Summand gegen 0 konvergiert. Also gilt $\liminf_{n\to\infty} a_n \geq 0$, und zusammen mit der oberen Abschätzung folgt die Behauptung. □

Bemerkung. Die Aussage von Satz 5.24 gilt auch ohne die Existenz von Dichten; im Beweis ersetzt man dann die Dichte g durch die entsprechend „gekippte“ Verteilung $r^{-1}\mathrm{e}^{\tau x}\,\mathbb{P}\circ X_1^{-1}(\mathrm{d}x)$.

Beispiel 5.25. (1) *Standardnormalverteilung.* Sei $X_1 \sim \mathcal{N}_{0,1}$. Dann ist

$$\Lambda(t) = \log \mathbb{E}[\mathrm{e}^{tX_1}] = \log\Big(\frac{1}{\sqrt{2\pi}}\int \mathrm{d}x\,\mathrm{e}^{tx}\mathrm{e}^{-\frac{x^2}{2}}\Big) = \log\Big(\frac{1}{\sqrt{2\pi}}\int \mathrm{d}x\,\mathrm{e}^{-\frac{(x-t)^2}{2}}\,\mathrm{e}^{\frac{t^2}{2}}\Big) = \frac{t^2}{2},$$

und somit

$$\Lambda^*(x) = \sup_{t\in\mathbb{R}}(tx - \Lambda(t)) = \sup_{t\in\mathbb{R}}\Big(tx - \frac{t^2}{2}\Big) = \frac{x^2}{2},$$

wobei das Supremum in $t = x$ angenommen wird. Also gilt für alle $x > 0$

$$\frac{1}{n}\log \mathbb{P}\Big(\sum_{i=1}^n X_i \geq xn\Big) \xrightarrow[n\to\infty]{} -\frac{x^2}{2}.$$

(2) *Bernoulli-Verteilung.* Sei X_1 Bernoulli-verteilt zum Parameter $p \in (0,1)$. Dann ist

$$\Lambda(t) = \log \mathbb{E}[\mathrm{e}^{tX_1}] = \log(1 - p + p\mathrm{e}^t) \quad \text{und} \quad \Lambda^*(x) = \sup_{t\in\mathbb{R}}\{tx - \log(1-p+p\mathrm{e}^t)\}.$$

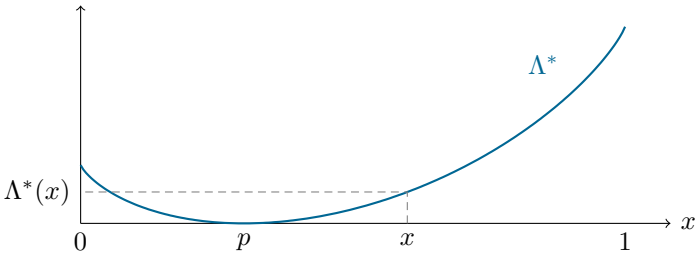
Für $x \in (0,1)$ ergibt sich die Maximalstelle aus der Nullstelle der Ableitung,

$$x = \frac{p\mathrm{e}^t}{1-p+p\mathrm{e}^t} \quad \Longleftrightarrow \quad t = \log\frac{x(1-p)}{p(1-x)},$$

und Einsetzen (mit $1 - p + p\mathrm{e}^t = (1-p)(1 + x/(1-x)) = (1-p)/(1-x)$) liefert

$$\Lambda^*(x) = x\log\frac{x(1-p)}{p(1-x)} - \log\frac{1-p}{1-x} = x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}.$$

Dies ist die *relative Entropie* der Bernoulli-Verteilung zum Parameter x bezüglich derjenigen zum Parameter p. Da X_1 keine Dichte besitzt, ist Satz 5.24 hier nicht direkt anwendbar; die Aussage $\frac{1}{n}\log \mathbb{P}(\sum_{i=1}^n X_i \geq xn) \to -\Lambda^*(x)$ für $p < x < 1$ folgt aus der Bemerkung im Anschluss an den Beweis von Satz 5.24 (die obere Abschätzung gilt ohnehin ohne Dichte). Dasselbe gilt in (3).



Die Ratenfunktion Λ^* der Bernoulli-Verteilung für $p = 0{,}3$ (exakt berechnet). Sie ist konvex und verschwindet nur in $x = p$: Für $x > p$ fällt $\mathbb{P}\big(\sum_{i=1}^n X_i \geq xn\big)$ exponentiell schnell, mit Rate $\Lambda^*(x)$.

(3) *Symmetrische* ± 1*-Variablen.* Sei $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. Dann ist

$$\Lambda(t) = \log\Big(\frac{1}{2}\mathrm{e}^{-t} + \frac{1}{2}\mathrm{e}^{t}\Big) = \log\cosh(t),$$

und eine analoge Rechnung (Maximalstelle t mit $\tanh t = x$) ergibt für $|x| < 1$

$$\Lambda^*(x) = \frac{1+x}{2}\log(1+x) + \frac{1-x}{2}\log(1-x).$$

6 Markov-Ketten

Markov-Ketten sind Folgen von Zufallsvariablen mit „kurzem Gedächtnis“. Ähnlich wie bei der Modellierung mit Differentialgleichungen soll das Verhalten zum nächsten Zeitpunkt nur vom aktuellen Zustand abhängen, nicht aber von der weiter zurückliegenden Vergangenheit.

6.1 Die Markov-Eigenschaft

Sei $E \neq \emptyset$ eine höchstens abzählbare Menge und $\Pi = (\Pi(x,y))_{x,y\in E}$ eine *stochastische Matrix*, d. h. für jedes $x \in E$ ist $\Pi(x,\cdot)$ eine Zähldichte auf E:

$$\Pi(x,y) \geq 0 \quad \text{für alle } x,y \in E, \qquad \sum_{y\in E} \Pi(x,y) = 1 \quad \text{für alle } x \in E.$$

Definition 6.1 (Markov-Kette)**.** Eine Folge $X_0, X_1, \ldots$ von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in E heißt *Markov-Kette* mit *Zustandsraum* E und *Übergangsmatrix* Π, wenn für alle $n \geq 0$ und $x_0, \ldots, x_{n+1} \in E$ gilt

$$\mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \ldots, X_n = x_n) = \Pi(x_n, x_{n+1}),$$

sofern $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) > 0$. Die Verteilung $\alpha = \mathbb{P} \circ X_0^{-1}$ von X_0 heißt *Startverteilung* der Markov-Kette.

Bemerkung 6.2 (Existenz und Veranschaulichung).

(1) Die Eigenschaft, dass die bedingte Verteilung von X_{n+1} nur von x_n abhängt, nicht aber von $x_0, \ldots, x_{n-1}$, heißt *Markov-Eigenschaft.*

(2) Die Übergangsmatrix hängt nicht von n ab. Dies ist der spezielle Fall *stationärer Übergangswahrscheinlichkeiten*, auf den wir uns hier beschränken.

(3) Markov-Ketten sind ein Spezialfall der in Satz 3.6 konstruierten Maße. Dabei nummerieren wir die Koordinaten mit $k \geq 0$ statt mit $k \geq 1$ (die Koordinate ω_k hier entspricht ω_{k+1} in Satz 3.6), setzen $\Omega_k := E$ für alle $k \geq 0$ und wählen die Startzähldichte $\rho_0(x) := \alpha(\{x\})$, $x \in E$, sowie die bedingten Zähldichten

$$\rho_{k;\omega_0\ldots\omega_{k-1}}(\omega_k) := \Pi(\omega_{k-1}, \omega_k), \qquad k \geq 1.$$

Daher gibt es zu jeder Startverteilung α auf E genau ein Wahrscheinlichkeitsmaß $\mathbb{P}^\alpha$ auf dem Produktraum

$$(\Omega, \mathcal{F}) := \Big(\prod_{k\geq 0} E, \bigotimes_{k\geq 0} \mathcal{P}(E)\Big),$$

so dass die Projektionen $X_n \colon (\omega_k)_{k\geq 0} \mapsto \omega_n \in E$ eine Markov-Kette zu Π und α bilden. Wir arbeiten im Folgenden mit diesem „kanonischen“ Modell.

(4) Ist $\alpha = \delta_x$ für ein $x \in E$, so schreiben wir $\mathbb{P}^x$ statt $\mathbb{P}^{\delta_x}$ für die Markov-Kette, die „sicher in x startet“. Dann gilt

$$\mathbb{P}^x(X_1 = x_1, \dots, X_n = x_n) = \Pi(x, x_1)\,\Pi(x_1, x_2) \cdots \Pi(x_{n-1}, x_n)$$

für alle $n \geq 1$ und $x_1, \dots, x_n \in E$. Summation über $x_1, \dots, x_{n-1} \in E$ liefert

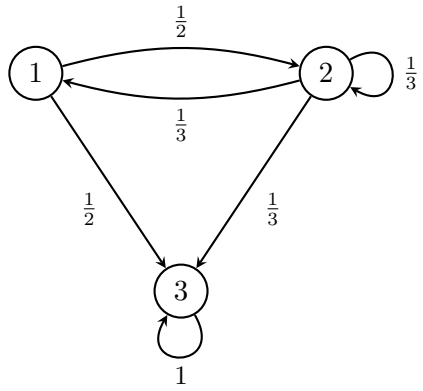
$$\mathbb{P}^x(X_n = y) = \Pi^n(x, y) \qquad \text{für alle } x, y \in E,$$

wobei Π^n das n-fache Matrixprodukt von Π mit sich selbst bezeichnet.

(5) *Darstellung durch einen Übergangsgraphen.* Zum Beispiel gehört zum Zustandsraum $E = \{1, 2, 3\}$ und

$$\Pi = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}$$

der folgende Graph; die Pfeile sind mit den Übergangswahrscheinlichkeiten beschriftet.



(6) *Simulation.* Wir fixieren eine Anordnung von E (z. B. $E \subset \mathbb{Z}$) und definieren

$$f\colon E \times (0,1) \to E, \qquad (x, u) \mapsto f(x, u) \coloneqq \min\Big\{z \in E\colon \sum_{y \leq z} \Pi(x, y) \geq u\Big\},$$

d. h. $f(x, \cdot)$ ist die Quantil-Transformation von $\Pi(x, \cdot)$. Seien nun $U_1, U_2, \dots$ unabhängig und gleichverteilt auf $(0, 1)$ und $X_0 = x_0 \in E$. Dann ist

$$X_n \coloneqq f(X_{n-1}, U_n), \qquad n \geq 1,$$

eine Markov-Kette zu Π mit Start in x_0. Denn wegen der Unabhängigkeit der U_i gilt für alle $x_1, \dots, x_n \in E$

$$\begin{aligned} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) &= \prod_{i=1}^{n} \mathbb{P}\big(f(x_{i-1}, U_i) = x_i\big) \\ &= \prod_{i=1}^{n} \mathbb{P}\Big(\sum_{y < x_i} \Pi(x_{i-1}, y) < U_i \leq \sum_{y \leq x_i} \Pi(x_{i-1}, y)\Big) \\ &= \prod_{i=1}^{n} \Big(\sum_{y \leq x_i} \Pi(x_{i-1}, y) - \sum_{y < x_i} \Pi(x_{i-1}, y)\Big) = \prod_{i=1}^{n} \Pi(x_{i-1}, x_i), \end{aligned}$$

und dies ist genau die Formel aus (4).

Beispiel 6.3. (1) *Irrfahrt auf $E = \mathbb{Z}$.* Seien $(Z_i)_{i\geq 1}$ u. i. v. $\mathbb{Z}$-wertige Zufallsvariablen mit Zähldichte ρ. Setze $X_0 = 0$ und $X_n = \sum_{i=1}^{n} Z_i$ für $n \geq 1$. Dann ist $(X_n)_{n\geq 0}$ eine Markov-Kette mit Übergangsmatrix

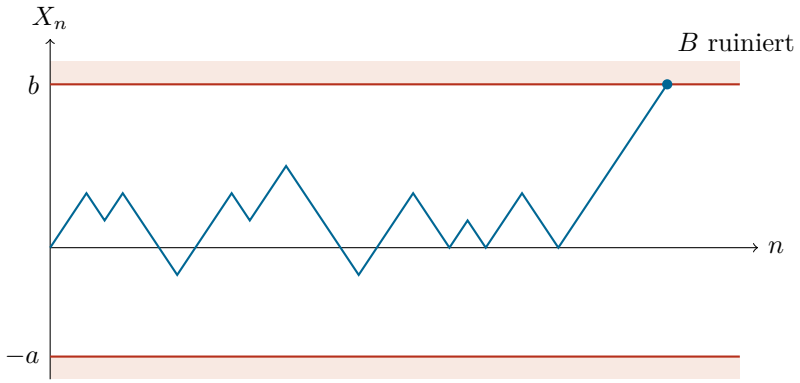
$$\Pi(x, y) = \rho(y - x), \qquad x, y \in \mathbb{Z}.$$

Denn da Z_{n+1} von $(X_0, \dots, X_n)$ unabhängig ist, gilt

$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n) &= \mathbb{P}(Z_{n+1} = x_{n+1} - x_n \,|\, X_0 = x_0, \dots, X_n = x_n) \\ &= \rho(x_{n+1} - x_n).\end{aligned}$$

(2) *Ruinproblem.* Zwei Spieler A und B besitzen a bzw. b Euro. Beide werfen wiederholt eine faire Münze, und je nach Ergebnis zahlt einer dem anderen 1 Euro. Das Spiel endet, sobald einer der Spieler ruiniert ist. Sei X_n der Gewinn von A nach Runde n; dabei ist $X_n \in \{-a, \dots, b\}$. Dann ist $(X_n)_{n\geq 0}$ eine Markov-Kette mit

$$\Pi(x, y) = \begin{cases} \frac{1}{2}, & \text{falls } -a < x < b \text{ und } |y - x| = 1, \\ 1, & \text{falls } x = y \in \{-a, b\}, \\ 0, & \text{sonst.} \end{cases}$$



Ein simulierter Pfad des Ruinproblems für $a = 4$, $b = 6$: Die Kette läuft als symmetrische Irrfahrt, bis sie einen der absorbierenden Zustände $-a$ oder b erreicht (hier nach 34 Runden den Zustand b).

Die Markov-Eigenschaft in Definition 6.1 bezieht sich auf einen einzelnen Zeitschritt. Der folgende Satz zeigt, dass sich die gesamte Zukunft einer Markov-Kette nach dem Zeitpunkt n bei gegebener Gegenwart $X_n = x$ wie eine neu in x gestartete Kette verhält, unabhängig von der Vergangenheit.

Satz 6.4 (Markov-Eigenschaft)**.** *Sei $(X_n)_{n\geq 0}$ eine Markov-Kette zur Übergangsmatrix Π und Startverteilung α im kanonischen Modell aus Bemerkung 6.2(3), und sei $\mathcal{F} = \bigotimes_{k\geq 0} \mathcal{P}(E)$. Dann gilt für alle $n \geq 0$, $A \in \mathcal{F}$, $B \subset E^n$ und $x \in E$*

$$\mathbb{P}^\alpha((X_n, X_{n+1}, \dots) \in A \,|\, (X_0, \dots, X_{n-1}) \in B,\ X_n = x) = \mathbb{P}^x(A),$$

sofern $\mathbb{P}^\alpha((X_0, \dots, X_{n-1}) \in B,\ X_n = x) > 0$. (Für $n = 0$ entfällt die Bedingung an $(X_0, \dots, X_{n-1})$.)

Beweis. Seien n, B und x fest mit $c \coloneqq \mathbb{P}^\alpha((X_0,\dots,X_{n-1}) \in B,\ X_n = x) > 0$. Beide Seiten sind als Funktionen von A Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{F})$: die rechte nach Konstruktion, die linke als bedingte Wahrscheinlichkeit (Proposition 3.1), ausgewertet auf den Urbildern $\{(X_n, X_{n+1}, \dots) \in A\}$. Nach dem Eindeutigkeitssatz 1.12 genügt es daher, die Gleichheit auf einem $\cap$-stabilen Erzeuger von $\mathcal{F}$ zu zeigen. Ein solcher ist das System der *Zylindermengen*

$$\{X_0 = x_0, \dots, X_k = x_k\}, \qquad k \geq 0,\ x_0, \dots, x_k \in E,$$

zusammen mit $\emptyset$; der Durchschnitt zweier Zylindermengen ist wieder eine Zylindermenge oder leer.

Sei also $A = \{X_i = x_i \text{ für alle } 0 \leq i \leq k\}$. Nach Bemerkung 6.2(3) und Satz 3.6 gilt die Produktformel $\mathbb{P}^\alpha(X_0 = y_0, \dots, X_m = y_m) = \alpha(y_0)\,\Pi(y_0, y_1) \cdots \Pi(y_{m-1}, y_m)$ für alle $m \geq 0$ und $y_0, \dots, y_m \in E$; damit folgt

$$\begin{aligned}
&\mathbb{P}^\alpha\big((X_0,\dots,X_{n-1}) \in B,\ X_n = x,\ X_{n+i} = x_i \text{ für alle } 0 \leq i \leq k\big) \\
&\quad = \sum_{(y_0,\dots,y_{n-1}) \in B} \alpha(y_0)\,\Pi(y_0, y_1) \cdots \Pi(y_{n-1}, x)\ \mathbb{1}\{x = x_0\}\,\Pi(x_0, x_1) \cdots \Pi(x_{k-1}, x_k) \\
&\quad = \mathbb{P}^\alpha\big((X_0,\dots,X_{n-1}) \in B,\ X_n = x\big)\,\mathbb{P}^x(X_i = x_i \text{ für alle } 0 \leq i \leq k)
\end{aligned}$$

(für $n = 0$ ist die Summe durch $\alpha(x)$ zu ersetzen). Division durch c liefert die Behauptung für Zylindermengen und damit für alle $A \in \mathcal{F}$. □

6.2 Absorptionswahrscheinlichkeiten

Definition 6.5 (Absorptionswahrscheinlichkeit)**.** Ein Zustand $z \in E$ heißt *absorbierend* bezüglich der Übergangsmatrix Π, falls $\Pi(z, z) = 1$. In diesem Fall heißt

$$h_z(x) \coloneqq \mathbb{P}^x(X_n = z \text{ schließlich}) = \mathbb{P}^x\Big(\bigcup_{N \geq 0} \bigcap_{n \geq N} \{X_n = z\}\Big), \qquad x \in E,$$

die *Absorptionswahrscheinlichkeit* in z bei Start in x.

Bemerkung und Definition 6.6 (Stoppzeiten)**.** Für $z \in E$ heißt

$$\tau_z \coloneqq \inf\{n \geq 1 \colon X_n = z\} \qquad (\inf \emptyset \coloneqq \infty)$$

die *Eintrittszeit* in den Zustand z. Bei Start in z spricht man von der *Rückkehrzeit* nach z. Die Zufallsvariable τ_z besitzt die *Stoppzeiteigenschaft*: Für jedes $n \geq 1$ ist

$$\{\tau_z = n\} = \{X_1 \neq z, \dots, X_{n-1} \neq z,\ X_n = z\} \in \mathcal{F}_n \coloneqq \sigma(X_0, \dots, X_n),$$

d. h. ob $\{\tau_z = n\}$ eintritt, hängt nur von $(X_0, \dots, X_n)$ ab.

Satz 6.7 (Charakterisierung der Absorptionswahrscheinlichkeiten)**.** *Sei $z \in E$ absorbierend. Dann gilt für alle $x \in E$*

$$h_z(x) = \mathbb{P}^x(\tau_z < \infty) = \lim_{n \to \infty} \mathbb{P}^x(X_n = z).$$

Außerdem ist h_z die kleinste nichtnegative Funktion $h \colon E \to [0, \infty)$ mit $h(z) = 1$ und

$$\sum_{y \in E} \Pi(x, y)\,h(y) = h(x) \qquad \text{für alle } x \in E.$$

Bemerkung. Die letzte Bedingung kann man kurz als $\Pi h = h$ schreiben; h ist also ein rechter Eigenvektor von Π zum Eigenwert 1. Eine solche Funktion heißt *harmonisch* bezüglich Π.

Beweis. Schritt 1: $h_z(x) = \lim_n \mathbb{P}^x(X_n = z)$. Mit der σ-Stetigkeit (Satz 1.11(e)) gilt zunächst

$$\mathbb{P}^z(X_i = z \text{ für alle } i \geq 0) = \lim_{k\to\infty} \mathbb{P}^z(X_i = z \text{ für alle } i \leq k) = \lim_{k\to\infty} \Pi(z,z)^k = 1.$$

Ist $\mathbb{P}^x(X_n = z) > 0$, so folgt mit der Markov-Eigenschaft (Satz 6.4, angewandt mit $B = E^n$)

$$\mathbb{P}^x(X_{n+i} = z \text{ für alle } i \geq 0) = \mathbb{P}^x(X_n = z)\,\mathbb{P}^z(X_i = z \text{ für alle } i \geq 0) = \mathbb{P}^x(X_n = z);$$

ist $\mathbb{P}^x(X_n = z) = 0$, so sind beide Seiten gleich 0. Die Ereignisse $\{X_{n+i} = z \text{ für alle } i \geq 0\}$ wachsen in n und haben die Vereinigung $\{X_n = z \text{ schließlich}\}$. Mit der σ-Stetigkeit folgt $\mathbb{P}^x(X_n = z) \to h_z(x)$ für $n \to \infty$.

Schritt 2: $h_z(x) = \mathbb{P}^x(\tau_z < \infty)$. Für $1 \leq k \leq n$ ist $\{\tau_z = k\} = \{(X_0, \dots, X_{k-1}) \in B_k,\ X_k = z\}$ mit $B_k = E \times (E \setminus \{z\})^{k-1}$. Ist $\mathbb{P}^x(\tau_z = k) > 0$, so liefert Satz 6.4

$$\mathbb{P}^x(X_n = z \,|\, \tau_z = k) = \mathbb{P}^z(X_{n-k} = z) = \Pi(z,z)^{n-k} = 1.$$

Da aus $X_n = z$ für $n \geq 1$ stets $\tau_z \leq n$ folgt, erhalten wir mit der Fallunterscheidungsformel

$$\mathbb{P}^x(X_n = z) = \sum_{k=1}^{n} \mathbb{P}^x(X_n = z,\ \tau_z = k) = \sum_{k=1}^{n} \mathbb{P}^x(\tau_z = k)\,\mathbb{P}^x(X_n = z \,|\, \tau_z = k) = \mathbb{P}^x(\tau_z \leq n),$$

wobei Summanden mit $\mathbb{P}^x(\tau_z = k) = 0$ verschwinden. Für $n \to \infty$ konvergiert die rechte Seite gegen $\mathbb{P}^x(\tau_z < \infty)$, die linke nach Schritt 1 gegen $h_z(x)$.

Schritt 3: h_z *ist harmonisch.* Nach Definition ist $h_z \geq 0$, und nach Schritt 1 ist $h_z(z) = \lim_n \mathbb{P}^z(X_n = z) = 1$. Sei $A := \{X_m = z \text{ schließlich}\} \in \mathcal{F}$; dann ist $\{(X_1, X_2, \dots) \in A\} = A$. Mit Satz 6.4 (für $n = 1$, $B = E$) folgt für alle $x \in E$

$$\begin{aligned}\sum_{y\in E} \Pi(x,y)\,h_z(y) &= \sum_{y\in E} \mathbb{P}^x(X_1 = y)\,\mathbb{P}^y(A) = \sum_{y\in E} \mathbb{P}^x(X_1 = y,\ (X_1, X_2, \dots) \in A) \\ &= \mathbb{P}^x(A) = h_z(x),\end{aligned}$$

wobei Summanden mit $\mathbb{P}^x(X_1 = y) = 0$ verschwinden.

Schritt 4: Minimalität. Sei $h \geq 0$ eine weitere Funktion mit $\Pi h = h$ und $h(z) = 1$. Durch Iteration folgt $h = \Pi^n h$ für alle $n \geq 1$, also

$$h(x) = \Pi^n h(x) = \sum_{y\in E} \Pi^n(x,y)\,h(y) \geq \Pi^n(x,z)\,h(z) = \mathbb{P}^x(X_n = z) \xrightarrow[n\to\infty]{} h_z(x)$$

nach Bemerkung 6.2(4) und Schritt 1. □

Beispiel 6.8 (Ruinproblem, Fortsetzung)**.** Wir betrachten wieder das Ruinproblem aus Beispiel 6.3(2). Die Zustände $-a$ und b sind absorbierend. Sei

$$r_A := h_{-a}(0)$$

die Ruinwahrscheinlichkeit von Spieler A. Nach Satz 6.7 gilt für $-a < x < b$

$$h_{-a}(x) = \Pi h_{-a}(x) = \tfrac{1}{2} h_{-a}(x-1) + \tfrac{1}{2} h_{-a}(x+1),$$

d. h. $h_{-a}(x+1) - h_{-a}(x) = h_{-a}(x) - h_{-a}(x-1)$. Die Zuwächse $c := h_{-a}(x+1) - h_{-a}(x)$ hängen also nicht von $x \in \{-a, \dots, b-1\}$ ab, und mit einer Teleskopsumme folgt

$$h_{-a}(x) - h_{-a}(-a) = \sum_{k=0}^{x+a-1} \big(h_{-a}(-a+k+1) - h_{-a}(-a+k)\big) = (x+a)\,c, \qquad -a \le x \le b.$$

Wegen $h_{-a}(-a) = 1$ und $h_{-a}(b) = 0$ (aus dem absorbierenden Zustand b wird $-a$ nie erreicht) liefert $x = b$

$$-1 = (a+b)\,c \quad \Longleftrightarrow \quad c = -\frac{1}{a+b},$$

und daher

$$r_A = h_{-a}(0) = 1 + ac = 1 - \frac{a}{a+b} = \frac{b}{a+b}.$$

Ebenso erhält man für die Ruinwahrscheinlichkeit von Spieler B den Wert $r_B := h_b(0) = a/(a+b)$, also $r_A + r_B = 1$. Da die Ereignisse „Absorption in $-a$“ und „Absorption in b“ disjunkt sind, endet das Spiel also mit Wahrscheinlichkeit 1 nach einer endlichen (zufälligen) Zeit mit dem Ruin eines der beiden Spieler.

Beispiel 6.9 (Verzweigungsprozesse)**.** Wir betrachten eine Population, in der jedes Individuum der n-ten Generation unabhängig von allen anderen mit Wahrscheinlichkeit $\rho(k)$ durch $k \in \mathbb{N}_0$ Nachkommen ersetzt wird; dabei ist ρ eine Zähldichte auf $\mathbb{N}_0$. Sei X_n die Anzahl der Individuen der n-ten Generation. Dann ist $(X_n)_{n\ge0}$ eine Markov-Kette auf $E = \mathbb{N}_0$ mit Übergangsmatrix

$$\Pi(k,l) = \rho^{*k}(l) = \sum_{l_1+\cdots+l_k=l} \rho(l_1)\cdots\rho(l_k), \qquad k,l \in \mathbb{N}_0,$$

d. h. $\Pi(k,\cdot)$ ist die k-fache Faltung von ρ, mit $\rho^{*0} := \delta_0$; insbesondere ist $\Pi(1,\cdot) = \rho$, und der Zustand 0 ist absorbierend. Interessant ist nur der Fall $0 < \rho(0) < 1$: Ist $\rho(0) = 0$, so hat jedes Individuum mindestens einen Nachkommen und die Population kann nicht aussterben; ist $\rho(0) = 1$, so stirbt sie sofort aus. Wir betrachten die *Aussterbewahrscheinlichkeit*

$$q := h_0(1)$$

bei einem Stammvater. Es gilt:

(1) Für alle $k, n \ge 0$ ist $\mathbb{P}^k(X_n = 0) = \mathbb{P}^1(X_n = 0)^k$. Für $n = 0$ ist nämlich

$$\mathbb{P}^k(X_0 = 0) = \delta_{0,k} \qquad \text{und} \qquad \mathbb{P}^1(X_0 = 0)^k = 0^k = \delta_{0,k}$$

(mit $0^0 = 1$). Gilt die Aussage für ein n, so folgt mit der Markov-Eigenschaft und der Fallunterscheidung nach X_1

$$\begin{aligned}\mathbb{P}^k(X_{n+1} = 0) &= \sum_{l\ge0} \Pi(k,l)\,\mathbb{P}^l(X_n = 0) = \sum_{l_1,\dots,l_k\ge0} \rho(l_1)\cdots\rho(l_k)\,\mathbb{P}^1(X_n = 0)^{l_1+\cdots+l_k} \\ &= \Big(\sum_{l\ge0} \Pi(1,l)\,\mathbb{P}^1(X_n = 0)^l\Big)^k = \mathbb{P}^1(X_{n+1} = 0)^k.\end{aligned}$$

Im Grenzwert $n \to \infty$ folgt daraus (vgl. Satz 6.7) $h_0(k) = q^k$ für alle $k \ge 0$.

(2) $q = h_0(1)$ ist der kleinste Fixpunkt in $[0,1]$ der erzeugenden Funktion

$$\varphi_\rho(s) = \sum_{k\ge0} \rho(k) s^k, \qquad s \in [0,1].$$

Denn einerseits gilt nach (1)

$$q = h_0(1) = \Pi h_0(1) = \sum_{k\geq 0} \rho(k) h_0(k) = \sum_{k\geq 0} \rho(k) q^k = \varphi_\rho(q).$$

Ist andererseits $s \in [0,1]$ ein weiterer Fixpunkt von φ_ρ, so gilt für $h(k) := s^k$ und alle $l \geq 0$

$$\Pi h(l) = \sum_{k\geq 0} \rho^{*l}(k) s^k = \Big(\sum_{k\geq 0} \rho(k) s^k\Big)^l = \varphi_\rho(s)^l = s^l = h(l),$$

d. h. $h \geq 0$ ist ein rechter Eigenvektor von Π zum Eigenwert 1 mit $h(0) = 1$. Da h_0 nach Satz 6.7 die kleinste solche Funktion ist, folgt $s = h(1) \geq h_0(1) = q$.

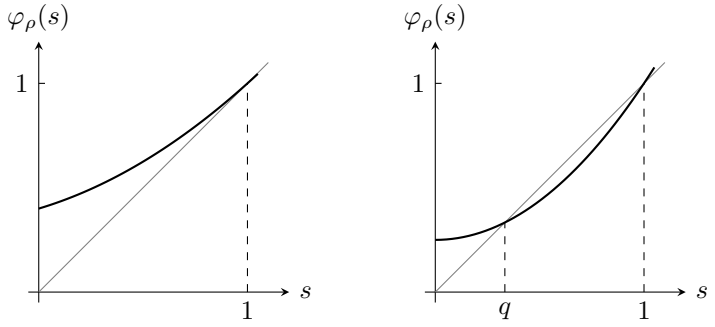
(3) φ_ρ ist entweder linear (falls $\rho(0) + \rho(1) = 1$) oder strikt konvex auf $[0,1]$, und es gilt

$$\varphi_\rho(0) = \rho(0) < 1, \qquad \varphi_\rho(1) = 1, \qquad \varphi_\rho'(1) = \mathbb{E}(\rho) := \sum_{k\geq 1} k\rho(k) \in [0,\infty].$$

Denn im Fall $\rho(0) + \rho(1) = 1$ ist $\varphi_\rho(s) = \rho(0) + s\rho(1)$; andernfalls ist

$$\varphi_\rho'(s) = \sum_{k\geq 1} k\rho(k) s^{k-1}, \qquad \varphi_\rho''(s) = \sum_{k\geq 2} k(k-1)\rho(k) s^{k-2} > 0 \quad \text{für } s \in (0,1),$$

und $\varphi_\rho'(1) = \mathbb{E}(\rho)$ (als linksseitige Ableitung). Im (interessanten) Fall $0 < \rho(0) < 1$ folgt: Ist $\mathbb{E}(\rho) \leq 1$, so ist $\varphi_\rho(s) > s$ für alle $s \in [0,1)$, d. h. φ_ρ besitzt nur den Fixpunkt 1. Ist dagegen $\mathbb{E}(\rho) > 1$, so ist $\varphi_\rho(s) < s$ für $s < 1$ nahe bei 1, und wegen $\varphi_\rho(0) = \rho(0) > 0$ besitzt φ_ρ nach dem Zwischenwertsatz noch einen weiteren Fixpunkt in $(0,1)$.



$\mathbb{E}(\rho) \leq 1$: $q = 1$ $\qquad$ $\mathbb{E}(\rho) > 1$: $q < 1$

Als Ergebnis erhalten wir im (interessanten) Fall $0 < \rho(0) < 1$:

(i) Ist $\mathbb{E}(\rho) \leq 1$, so ist $q = 1$, d. h. die Population stirbt mit Sicherheit aus.

(ii) Ist $\mathbb{E}(\rho) > 1$, so ist $q \in (0,1)$, d. h. die Population hat eine positive Überlebenswahrscheinlichkeit.

Bemerkung und Definition 6.10 (Eintrittswahrscheinlichkeiten)**.** Sei $z \in E$ ein beliebiger (nicht notwendig absorbierender) Zustand mit Eintrittszeit $\tau_z = \inf\{n \geq 1 \colon X_n = z\}$. Wir betrachten die „zur Zeit τ_z gestoppte Markov-Kette“

$$\tilde{X}_n := X_{\min(n,\tau_z)}, \qquad n \geq 0.$$

Bei Start in $x \neq z$ (allgemeiner unter jeder Startverteilung α mit $\alpha(z) = 0$) ist diese eine Markov-Kette zur Übergangsmatrix

$$\tilde{\Pi}(x,y) = \begin{cases} \Pi(x,y), & \text{falls } x \neq z, \\ 1, & \text{falls } x = y = z, \\ 0, & \text{sonst,} \end{cases}$$

bei der der Zustand z absorbierend gemacht wurde. Für $x \neq z$ ist dann die Eintrittswahrscheinlichkeit $h_z(x) \coloneqq \mathbb{P}^x(\tau_z < \infty)$ gerade die Wahrscheinlichkeit, dass $(\tilde{X}_n)$ bei Start in x in z absorbiert wird. Damit lassen sich die Resultate über Absorptionswahrscheinlichkeiten auf Eintrittswahrscheinlichkeiten übertragen.

6.3 Asymptotische Stationarität

Sei nun $|E| < \infty$. Wie sieht das Langzeitverhalten einer Markov-Kette aus? *Idee:* Wenn die Markov-Kette nicht in einzelne Bereiche zerfällt, die nicht miteinander kommunizieren, so sollte sie sich nach langer Zeit in einem Gleichgewichtszustand einpendeln.

Satz 6.11 (Ergodensatz für Markov-Ketten)**.** *Sei $|E| < \infty$, und es gebe ein $k \geq 1$ mit $\Pi^k(x,y) > 0$ für alle $x, y \in E$. Dann existiert für alle $y \in E$ der Grenzwert*

$$\lim_{n\to\infty} \Pi^n(x,y) = \alpha(y) > 0$$

und hängt nicht von $x \in E$ ab. Dabei ist α die eindeutige Zähldichte auf E mit

$$\sum_{x\in E} \alpha(x)\Pi(x,y) = \alpha(y) \qquad \text{für alle } y \in E. \tag{$*$}$$

Bemerkung 6.12. (1) Statt ($*$) kann man auch kurz $\alpha\Pi = \alpha$ schreiben und α als *linken Eigenvektor* von Π zum Eigenwert 1 auffassen. Unter der Startverteilung α ist die Markov-Kette *zeitlich invariant* (*stationär*) in dem Sinne, dass

$$\mathbb{P}^\alpha\big((X_n, X_{n+1}, \ldots) \in A\big) = \mathbb{P}^\alpha(A) \qquad \text{für alle } A \in \mathcal{F},\ n \geq 0,$$

und α heißt die *stationäre (Start-)Verteilung*. Denn mit der Markov-Eigenschaft und $\alpha\Pi^n = \alpha$ gilt

$$\begin{aligned} \mathbb{P}^\alpha\big((X_n, X_{n+1}, \ldots) \in A\big) &= \sum_{x\in E} \mathbb{P}^\alpha(X_n = x)\, \mathbb{P}^x(A) = \sum_{x\in E}\sum_{y\in E} \alpha(y)\Pi^n(y,x)\, \mathbb{P}^x(A) \\ &= \sum_{x\in E} \alpha(x)\mathbb{P}^x(A) = \mathbb{P}^\alpha(A), \end{aligned}$$

unabhängig von n.

(2) Unter den Voraussetzungen von Satz 6.11 gilt sogar

$$\sup_{A\in\mathcal{F}} \Big|\mathbb{P}^x\big((X_n, X_{n+1}, \ldots) \in A\big) - \mathbb{P}^\alpha(A)\Big| \xrightarrow[n\to\infty]{} 0.$$

Denn für jedes $A \in \mathcal{F}$ ist

$$\begin{aligned} \Big|\mathbb{P}^x\big((X_n, X_{n+1}, \ldots) \in A\big) - \mathbb{P}^\alpha(A)\Big| &= \Big|\sum_{y\in E} \big(\mathbb{P}^x(X_n = y) - \alpha(y)\big)\mathbb{P}^y(A)\Big| \\ &\leq \sum_{y\in E} \big|\Pi^n(x,y) - \alpha(y)\big| \xrightarrow[n\to\infty]{} 0, \end{aligned}$$

da E endlich ist.

(3) *Null-Eins-Gesetz von Orey.* Ähnlich wie im 0-1-Gesetz von Kolmogorov (Satz 3.29) gilt unter den Voraussetzungen von Satz 6.11

$$\mathbb{P}^x(A) = \mathbb{P}^\alpha(A) \in \{0,1\} \qquad \text{für alle } A \in \mathcal{A} \text{ und } x \in E,$$

wobei $\mathcal{A} = \mathcal{A}(X_n \colon n \geq 0)$ die σ-Algebra der asymptotischen Ereignisse bezeichnet (Definition 3.27). Denn ist $A \in \mathcal{A}$, so gibt es zu jedem $n \geq 0$ ein $B_n \in \mathcal{F}$ mit

$$A = \big\{(X_n, X_{n+1}, \ldots) \in B_n\big\} = E^n \times B_n \qquad \text{(im kanonischen Modell)},$$

und es gilt $B_n = E^k \times B_{n+k}$ für alle $k \geq 0$, also ist auch $B_n \in \mathcal{A}$. Mit (1) folgt $\mathbb{P}^\alpha(A) = \mathbb{P}^\alpha(B_n)$ und daher mit (2)

$$\big|\mathbb{P}^x(A) - \mathbb{P}^\alpha(A)\big| = \big|\mathbb{P}^x\big((X_n, X_{n+1}, \ldots) \in B_n\big) - \mathbb{P}^\alpha(B_n)\big| \xrightarrow[n\to\infty]{} 0.$$

Da die linke Seite nicht von n abhängt, ist $\mathbb{P}^x(A) = \mathbb{P}^\alpha(A)$ für alle x; ebenso (angewandt auf $B_n \in \mathcal{A}$) $\mathbb{P}^x(B_n) = \mathbb{P}^\alpha(B_n)$ für alle x und n. Für alle $x_0, \ldots, x_n \in E$ mit $\mathbb{P}^\alpha(X_0 = x_0, \ldots, X_n = x_n) > 0$ folgt mit der Markov-Eigenschaft

$$\mathbb{P}^\alpha\big(A \,\big|\, X_0 = x_0, \ldots, X_n = x_n\big) = \mathbb{P}^{x_n}(B_n) = \mathbb{P}^\alpha(B_n) = \mathbb{P}^\alpha(A).$$

Also ist A bezüglich $\mathbb{P}^\alpha$ unabhängig von $(X_0, \ldots, X_n)$ für jedes n, d. h. von jedem Ereignis in dem $\cap$-stabilen Erzeuger $\bigcup_{n\geq 0} \sigma(X_0, \ldots, X_n)$ von $\mathcal{F}$, und damit (Dynkin-Argument) von jedem Ereignis in $\mathcal{F}$ – insbesondere von sich selbst. Somit ist $\mathbb{P}^\alpha(A) = \mathbb{P}^\alpha(A)^2$, also $\mathbb{P}^\alpha(A) \in \{0,1\}$.

Beweis von Satz 6.11. Schritt 1: Kontraktion. Für Zähldichten ρ_1, ρ_2 auf E betrachten wir den *Variationsabstand*

$$\|\rho_1 - \rho_2\| = \sum_{z\in E} |\rho_1(z) - \rho_2(z)|.$$

Für jede stochastische Matrix Π gilt

$$\begin{aligned}\|\rho_1\Pi - \rho_2\Pi\| &= \sum_{x\in E} \Big|\sum_{y\in E} \rho_1(y)\Pi(y,x) - \sum_{y\in E} \rho_2(y)\Pi(y,x)\Big| \\ &\leq \sum_{x,y\in E} |\rho_1(y) - \rho_2(y)|\,\Pi(y,x) = \|\rho_1 - \rho_2\|.\end{aligned}$$

Für Π^k gilt sogar eine strikte Kontraktion: Da E endlich ist und $\Pi^k > 0$, gibt es ein $\delta \in (0,1)$ mit

$$\Pi^k(x,y) \geq \frac{\delta}{|E|} \qquad \text{für alle } x, y \in E.$$

Mit $U(x,y) := 1/|E|$ gilt also $\Pi^k \geq \delta U$ (elementweise), und $V := (1-\delta)^{-1}(\Pi^k - \delta U)$ ist ebenfalls eine stochastische Matrix mit $\Pi^k = \delta U + (1-\delta)V$. Da $\rho_1 U = \rho_2 U$ (beides ist die Gleichverteilung auf E), folgt

$$\|\rho_1\Pi^k - \rho_2\Pi^k\| \leq \delta \underbrace{\|\rho_1 U - \rho_2 U\|}_{=0} + (1-\delta)\|\rho_1 V - \rho_2 V\| \leq (1-\delta)\|\rho_1 - \rho_2\|.$$

Für $n \geq 0$ und $m := \lfloor n/k \rfloor$ ergibt sich daraus mit der ersten Abschätzung (angewandt auf Π^{n-km}) und wegen $\|\rho_1 - \rho_2\| \leq 2$

$$\begin{aligned}\|\rho_1\Pi^n - \rho_2\Pi^n\| &\leq \|\rho_1\Pi^{km} - \rho_2\Pi^{km}\| \\ &\leq (1-\delta)^m\|\rho_1 - \rho_2\| \leq 2(1-\delta)^m \xrightarrow[n\to\infty]{} 0.\end{aligned} \tag{$\circledast$}$$

Schritt 2: Existenz und Eindeutigkeit von α. Sei ρ eine Zähldichte auf E. Die Folge $(\rho\Pi^n)_{n\geq 0}$ liegt in der kompakten Menge $[0,1]^E$; also gibt es eine Teilfolge (n_k) mit $\rho\Pi^{n_k} \to \alpha$ für $k \to \infty$, und α ist (da E endlich ist) wieder eine Zähldichte. Mit $\rho_1 = \rho$ und $\rho_2 = \rho\Pi$ in (⊛) folgt $\|\rho\Pi^{n_k} - \rho\Pi^{n_k+1}\| \to 0$ und daher

$$\alpha = \lim_{k\to\infty} \rho\Pi^{n_k} = \lim_{k\to\infty} \rho\Pi^{n_k+1} = \alpha\Pi,$$

d. h. α ist stationär. Außerdem gilt

$$\alpha(y) = \alpha\Pi^k(y) = \sum_{x\in E} \alpha(x)\Pi^k(x,y) \geq \frac{\delta}{|E|} > 0 \qquad \text{für alle } y \in E.$$

Mit (⊛) und $\rho_2 = \alpha = \alpha\Pi^n$ folgt $\|\rho\Pi^n - \alpha\| \to 0$ für jede Zähldichte ρ. Ist ρ selbst stationär, so ist $\rho = \rho\Pi^n \to \alpha$, also $\rho = \alpha$; das zeigt die Eindeutigkeit. Für $\rho = \delta_x$, $x \in E$, erhalten wir schließlich

$$\delta_x\Pi^n = \Pi^n(x,\cdot) \xrightarrow[n\to\infty]{} \alpha. \qquad \square$$

Bemerkung. Die Abschätzung (⊛) liefert sogar eine exponentielle Konvergenzgeschwindigkeit: Wegen $\lfloor n/k \rfloor \geq n/k - 1$ und $\log(1-\delta) \leq -\delta$ ist

$$2(1-\delta)^m = 2\mathrm{e}^{\lfloor n/k\rfloor \log(1-\delta)} \leq 2\mathrm{e}^{\delta}\mathrm{e}^{-cn} \qquad \text{mit } c := \frac{\delta}{k} > 0.$$

Beispiel 6.13. (1) *Kartenmischen.* Wir betrachten einen Stapel von N Spielkarten. Es gibt $N!$ mögliche Reihenfolgen, die wir mit den Permutationen $\pi \in S_N =: E$ identifizieren. Ein Mischverfahren beschreiben wir durch

$$\pi_0 \;\longrightarrow\; X_1 = \xi_1 \circ \pi_0 \;\longrightarrow\; X_2 = \xi_2 \circ X_1 \;\longrightarrow\; \cdots$$

mit u. i. v. zufälligen Permutationen $\xi_1, \xi_2, \ldots$. Gutes Mischen bedeutet, dass die Verteilung der Reihenfolge nahezu die Gleichverteilung ist. Wie schnell mischt das Verfahren?

Sei $\rho(\pi) = \mathbb{P}(\xi_1 = \pi)$ die Zähldichte von ξ_1. Dann ist

$$\mathbb{P}(X_1 = \pi_1, \ldots, X_n = \pi_n) = \rho(\pi_1 \circ \pi_0^{-1}) \cdots \rho(\pi_n \circ \pi_{n-1}^{-1}),$$

also ist (X_n) eine Markov-Kette mit Übergangsmatrix $\Pi(\pi,\pi') = \rho(\pi' \circ \pi^{-1})$. Diese ist sogar *doppelt stochastisch*: Es gilt

$$\sum_{\pi'\in E} \Pi(\pi,\pi') = \sum_{\pi'\in E} \rho(\pi' \circ \pi^{-1}) = 1 \qquad \text{und} \qquad \sum_{\pi\in E} \Pi(\pi,\pi') = \sum_{\pi\in E} \rho(\pi' \circ \pi^{-1}) = 1,$$

da sowohl $\pi' \mapsto \pi' \circ \pi^{-1}$ (bei festem π) als auch $\pi \mapsto \pi' \circ \pi^{-1}$ (bei festem π') Bijektionen von E auf sich sind. Folglich ist die Gleichverteilung $\alpha = \mathcal{U}_{S_N}$ stationär, denn

$$\sum_{\pi\in E} \alpha(\pi)\Pi(\pi,\pi') = \frac{1}{N!} = \alpha(\pi') \qquad \text{für alle } \pi' \in E.$$

Nach Satz 6.11 gilt daher

$$\lim_{n\to\infty} \mathbb{P}^{\pi_0}(X_n = \pi) = \frac{1}{N!} \qquad \text{für alle } \pi, \pi_0 \in E,$$

falls $\Pi^k > 0$ für ein $k \geq 1$ ist, und zwar mit exponentieller Geschwindigkeit.

(2) *Markov Chain Monte Carlo (MCMC) für das Ising-Modell.* Sei α eine Zähldichte auf einer endlichen, aber sehr großen Menge E. Wie kann α simuliert werden?

Ising-Modell. Sei $\Lambda \subset \mathbb{Z}^2$ endlich und $E = \{-1, 1\}^\Lambda$; dabei beschreibt $x_i \in \{\pm 1\}$ den „Spin“ am Ort $i \in \Lambda$. Betrachte

$$\alpha(x) = Z^{-1} \exp\Big(\sum_{i,j \in \Lambda:\ i \neq j} J(i,j)\, x_i x_j \Big), \qquad x \in E,$$

mit einer *Kopplungskonstanten* J, z. B. $J(i,j) = \mathbb{1}\{i \text{ und } j \text{ sind Nachbarn}\}$, und der Normierungskonstanten $Z = \sum_{x \in E} \exp\big(\sum_{i \neq j} J(i,j) x_i x_j\big)$. Dies ist das paradigmatische Modell für spontane Magnetisierung in einem Eisenblock.

Die Konstante Z ist eine Summe über $2^{|\Lambda|}$ Terme und daher schwer auszurechnen. *Idee:* Betrachte eine Markov-Kette, für die α stationär ist. Für $x \in E$ und $j \in \Lambda$ sei $T_j x$ die Konfiguration, die aus x durch Umdrehen („Flip“) des Spins am Ort j entsteht. Ein Beispiel ist der *Metropolis-Algorithmus*: Mit einer Konstanten $0 < c < 1/|\Lambda|$ setze

$$\Pi(x,y) = \begin{cases} c \min(\alpha(y)/\alpha(x), 1), & \text{falls } y = T_j x \text{ für ein } j \in \Lambda, \\ 1 - \sum_{j \in \Lambda} \Pi(x, T_j x), & \text{falls } y = x, \\ 0, & \text{sonst.} \end{cases}$$

Der Quotient $\alpha(y)/\alpha(x)$ ist leicht zu berechnen, da sich Z herauskürzt und sich x und $y = T_j x$ nur an einer Stelle unterscheiden. Es gilt:

(i) Π ist stochastisch: $\Pi(x,y) \geq 0$ wegen $c|\Lambda| < 1$, und $\sum_y \Pi(x,y) = 1$ nach Definition.

(ii) *Detailed Balance:* $\alpha(x)\Pi(x,y) = \alpha(y)\Pi(y,x)$ für alle $x, y \in E$, denn für $y = T_j x$ sind beide Seiten gleich $c \min(\alpha(x), \alpha(y))$. Summation über x liefert $\alpha\Pi = \alpha$.

(iii) $\Pi(x,x) \geq 1 - c|\Lambda| > 0$ für alle x, und jedes y ist von jedem x durch höchstens $|\Lambda|$ Spinflips erreichbar, die jeweils positive Wahrscheinlichkeit haben. Daher gilt $\Pi^{|\Lambda|}(x,y) > 0$ für alle $x, y \in E$.

Nach Satz 6.11 konvergiert also die Verteilung der Kette gegen α, und man erhält nach genügend vielen Schritten eine für α typische Konfiguration.

6.4 Wiederkehrzeiten

Für abzählbares E betrachten wir nun Übergangsmatrizen, die keine isolierten Bereiche in E zulassen, und zwar im folgenden Sinne. (Eine Version des Ergodensatzes aus Satz 6.11 gilt für abzählbares E erst unter zusätzlichen Voraussetzungen wie Aperiodizität und positiver Rekurrenz; Irreduzibilität allein genügt nicht, denn für $E = \{0,1\}$ und $\Pi(0,1) = \Pi(1,0) = 1$ konvergiert $\Pi^n(0,0) = \mathbb{1}\{n \text{ gerade}\}$ nicht.)

Definition 6.14 (Irreduzibilität)**.** Eine Übergangsmatrix Π heißt *irreduzibel*, falls es zu allen $x, y \in E$ ein $k = k(x,y) \geq 0$ gibt mit $\Pi^k(x,y) > 0$.

Unter Irreduzibilität kann eine stationäre Verteilung, sofern sie existiert, eindeutig über die Rückkehrzeiten

$$\tau_x = \inf\{n \geq 1 \colon X_n = x\}, \qquad x \in E,$$

charakterisiert werden. Wir schreiben $\mathbb{E}^x$ bzw. $\mathbb{E}^\alpha$ für den Erwartungswert bezüglich $\mathbb{P}^x$ bzw. $\mathbb{P}^\alpha$.

Satz 6.15 (Wiederkehrsatz)**.** *Sei E höchstens abzählbar und Π irreduzibel. Ist α stationär bezüglich Π, so gilt*

$$0 < \alpha(x) = \frac{1}{\mathbb{E}^x[\tau_x]} \qquad \text{für alle } x \in E,$$

und α ist insbesondere eindeutig.

Beweis. Schritt 1. Wir zeigen zunächst, dass sogar ohne Irreduzibilität für jede stationäre Verteilung α gilt:

$$\alpha(x)\,\mathbb{E}^x[\tau_x] = \mathbb{P}^\alpha(\tau_x < \infty) \qquad \text{für alle } x \in E. \tag{$\dagger$}$$

Es ist

$$\begin{aligned}
\alpha(x)\,\mathbb{E}^x[\tau_x] &= \mathbb{E}^\alpha\big[\mathbb{1}\{X_0 = x\}\,\tau_x\big] = \mathbb{E}^\alpha\Big[\mathbb{1}\{X_0 = x\}\sum_{n\geq 0}\mathbb{1}\{\tau_x > n\}\Big] \\
&= \sum_{n\geq 0}\mathbb{P}^\alpha(X_0 = x,\ \tau_x > n) \\
&= \lim_{k\to\infty}\sum_{n=0}^{k-1}\mathbb{P}^\alpha(X_0 = x,\ \tau_x > n) = \lim_{k\to\infty}\sum_{n=0}^{k-1}\mathbb{P}^\alpha(X_0 = x,\ X_i \neq x \text{ für } 1 \leq i \leq n) \\
&= \lim_{k\to\infty}\sum_{n=0}^{k-1}\mathbb{P}^\alpha(X_{k-n} = x,\ X_i \neq x \text{ für } k-n < i \leq k).
\end{aligned}$$

Dabei haben wir in der ersten Gleichung $\mathbb{P}^\alpha(\,\cdot\,|X_0 = x) = \mathbb{P}^x$ benutzt, in der zweiten $\tau_x = \sum_{n\geq 0}\mathbb{1}\{\tau_x > n\}$ (auch für $\tau_x = \infty$), in der dritten die σ-Additivität des Erwartungswerts und in der letzten Gleichung die Stationarität der Kette unter $\mathbb{P}^\alpha$ (Bemerkung 6.12(1)) für eine Zeitverschiebung um $k-n$. Das Ereignis in der letzten Summe besagt:

(1) $k-n$ ist der letzte Zeitpunkt in $\{1,\dots,k\}$, zu dem sich die Kette in x befindet;

(2) insbesondere sind diese Ereignisse für verschiedene n disjunkt.

Ihre Vereinigung über $n \in \{0,\dots,k-1\}$ ist also das Ereignis $\{\tau_x \leq k\}$, und wir erhalten

$$\alpha(x)\,\mathbb{E}^x[\tau_x] = \lim_{k\to\infty}\mathbb{P}^\alpha(\tau_x \leq k) = \mathbb{P}^\alpha(\tau_x < \infty).$$

Schritt 2. Es bleibt zu zeigen, dass $\alpha(x) > 0$ und $\mathbb{P}^\alpha(\tau_x < \infty) = 1$ für alle $x \in E$; dann folgt die Behauptung aus ($\dagger$). Da α eine Zähldichte ist, gibt es ein x_0 mit $\alpha(x_0) > 0$, und wegen der Irreduzibilität gibt es zu $x \in E$ ein k mit $\Pi^k(x_0, x) > 0$. Also

$$\alpha(x) = \alpha\Pi^k(x) \geq \alpha(x_0)\Pi^k(x_0, x) > 0.$$

Mit ($\dagger$) folgt $\mathbb{E}^x[\tau_x] \leq 1/\alpha(x) < \infty$ für alle $x \in E$, also insbesondere $\mathbb{P}^x(\tau_x < \infty) = 1$. Es gilt sogar $\mathbb{P}^x(X_n = x \text{ unendlich oft}) = 1$ (der Beweis folgt in Satz 6.16(a)). Mit der Markov-Eigenschaft folgt für jedes $k \geq 1$

$$0 = \mathbb{P}^x(X_n \neq x \text{ für alle } n > k) = \sum_{y\in E}\Pi^k(x,y)\,\mathbb{P}^y(\tau_x = \infty),$$

also $\mathbb{P}^y(\tau_x = \infty) = 0$, falls $\Pi^k(x,y) > 0$. Da Π irreduzibel ist, gilt dies für alle $y \in E$, und somit

$$\mathbb{P}^\alpha(\tau_x < \infty) = \sum_{y\in E}\alpha(y)\underbrace{\mathbb{P}^y(\tau_x < \infty)}_{=1} = 1. \qquad \square$$

Wir zeigen nun noch eine feinere Aussage über das Wiederkehrverhalten. Für $x \in E$ betrachten wir

- die Wahrscheinlichkeit einer Rückkehr nach x, $F_1(x,x) = \mathbb{P}^x(\tau_x < \infty)$,
- die Wahrscheinlichkeit unendlich vieler Rückkehren,

$$F_\infty(x,x) = \mathbb{P}^x(X_n = x \text{ für unendlich viele } n),$$

- die erwartete Anzahl der Besuche in x,

$$G(x,x) = \mathbb{E}^x\Big[\sum_{n\geq 0} \mathbb{1}\{X_n = x\}\Big] = \sum_{n\geq 0} \Pi^n(x,x).$$

Satz 6.16 (Rekurrenz und Transienz)**.** *Sei $x \in E$. Dann gilt:*

(a) *Ist $F_1(x,x) = 1$, so ist $F_\infty(x,x) = 1$ und $G(x,x) = \infty$. In diesem Fall heißt x* rekurrent*.*

(b) *Ist $F_1(x,x) < 1$, so ist $F_\infty(x,x) = 0$ und*

$$G(x,x) = \frac{1}{1 - F_1(x,x)} < \infty.$$

In diesem Fall heißt x transient*.*

Bemerkung. Teil (a) liefert das noch fehlende Argument im Beweis von Satz 6.15.

Beweis. Sei $\sigma = \sup\{n \geq 0 \colon X_n = x\}$ der Zeitpunkt des letzten Aufenthalts in x (mit $\sigma = \infty$, falls x unendlich oft besucht wird). Dann gilt:

(1) $1 - F_\infty(x,x) = \mathbb{P}^x(\sigma < \infty)$.

(2) Für alle $n \geq 0$ ist nach der Markov-Eigenschaft

$$\mathbb{P}^x(\sigma = n) = \mathbb{P}^x(X_n = x)\,\mathbb{P}^x(X_i \neq x \text{ für alle } i \geq 1) = \Pi^n(x,x)\big(1 - F_1(x,x)\big).$$

(3) Summation über n liefert $1 - F_\infty(x,x) = G(x,x)\big(1 - F_1(x,x)\big)$.

Zu (a): Ist $F_1(x,x) = 1$, so ist nach (2) $\mathbb{P}^x(\sigma = n) = 0$ für alle n, also nach (1) $F_\infty(x,x) = 1$. Mit der Markov-Ungleichung folgt für jedes $k \in \mathbb{N}$

$$1 = \mathbb{P}^x\big(|\{n \colon X_n = x\}| \geq k\big) \leq \frac{1}{k}\, G(x,x),$$

also $G(x,x) \geq k$ für alle k, d. h. $G(x,x) = \infty$.

Zu (b): Ist $F_1(x,x) < 1$, so folgt aus (3) $G(x,x) \leq 1/(1 - F_1(x,x)) < \infty$. Mit dem Lemma von Borel–Cantelli folgt aus

$$\sum_{n\geq 0} \mathbb{P}^x(X_n = x) = G(x,x) < \infty,$$

dass $F_\infty(x,x) = 0$, und dann liefert (3) $G(x,x) = 1/(1 - F_1(x,x))$. □

Bemerkung. Wir betrachten die einfache symmetrische Irrfahrt auf $\mathbb{Z}^d$ mit

$$\Pi(x,y) = \begin{cases} \frac{1}{2d}, & \text{falls } |x-y| = 1, \\ 0, & \text{sonst.} \end{cases}$$

Dann gilt (ohne Beweis): Für $d = 1, 2$ ist jedes $x \in \mathbb{Z}^d$ rekurrent, für $d \geq 3$ ist jedes $x \in \mathbb{Z}^d$ transient.

Wie steht es um die mittlere Rückkehrzeit?

Definition 6.17 (Positive Rekurrenz)**.** Ein Zustand $x \in E$ heißt *positiv rekurrent*, falls $\mathbb{E}^x[\tau_x] < \infty$. Ist x rekurrent, aber $\mathbb{E}^x[\tau_x] = \infty$, so heißt x *nullrekurrent.*

Wegen $\mathbb{E}^x[\tau_x] < \infty \Rightarrow \mathbb{P}^x(\tau_x < \infty) = 1$ ist jeder positiv rekurrente Zustand rekurrent.

Satz 6.18 (Positive Rekurrenz und Existenz einer stationären Verteilung)**.** *Ist $x \in E$ positiv rekurrent, so existiert eine stationäre Verteilung α mit*

$$\alpha(x) = \frac{1}{\mathbb{E}^x[\tau_x]} > 0.$$

Ist umgekehrt α eine stationäre Verteilung mit $\alpha(x) > 0$, so ist x positiv rekurrent.

Beweis. Zur Umkehrung: Ist $\alpha(x) > 0$ und $\alpha\Pi = \alpha$, so folgt aus (†)

$$\mathbb{E}^x[\tau_x] = \frac{\mathbb{P}^\alpha(\tau_x < \infty)}{\alpha(x)} \leq \frac{1}{\alpha(x)} < \infty.$$

Zur Existenz: Sei $x \in E$ positiv rekurrent; insbesondere ist $\mathbb{P}^x(\tau_x < \infty) = 1$. Betrachte die erwartete Anzahl der Besuche in y vor der Rückkehr nach x,

$$\beta(y) \coloneqq \mathbb{E}^x\Big[\sum_{n=1}^{\tau_x} \mathbb{1}\{X_n = y\}\Big], \qquad y \in E.$$

Wegen $\mathbb{P}^x(X_{\tau_x} = y) = \delta_{x,y} = \mathbb{P}^x(X_0 = y)$ können wir die Summe ebenso gut von $n = 0$ bis $\tau_x - 1$ laufen lassen:

$$\beta(y) = \mathbb{E}^x\Big[\sum_{n=0}^{\tau_x - 1} \mathbb{1}\{X_n = y\}\Big] = \sum_{n \geq 0} \mathbb{P}^x(\tau_x > n,\ X_n = y).$$

Da das Ereignis $\{\tau_x > n\}$ nur von $X_0, \ldots, X_n$ abhängt, folgt mit der Markov-Eigenschaft für alle $z \in E$

$$\begin{aligned} \beta\Pi(z) &= \sum_{y \in E} \sum_{n \geq 0} \mathbb{P}^x(\tau_x > n,\ X_n = y)\, \Pi(y,z) = \sum_{n \geq 0} \sum_{y \in E} \mathbb{P}^x(\tau_x > n,\ X_n = y,\ X_{n+1} = z) \\ &= \sum_{n \geq 0} \mathbb{P}^x(\tau_x > n,\ X_{n+1} = z) = \mathbb{E}^x\Big[\sum_{n=1}^{\tau_x} \mathbb{1}\{X_n = z\}\Big] = \beta(z). \end{aligned}$$

Ferner ist $\sum_{y \in E} \beta(y) = \mathbb{E}^x[\tau_x] < \infty$ und $\beta(x) = 1$. Daher ist

$$\alpha \coloneqq \frac{1}{\mathbb{E}^x[\tau_x]}\, \beta$$

eine stationäre Verteilung, denn α ist eine Zähldichte und

$$\alpha\Pi(y) = \sum_{z \in E} \frac{1}{\mathbb{E}^x[\tau_x]}\, \beta(z)\Pi(z,y) = \frac{1}{\mathbb{E}^x[\tau_x]}\, \beta(y) = \alpha(y),$$

und es gilt $\alpha(x) = 1/(\mathbb{E}^x[\tau_x])$. □

Bemerkung.

(1) Die einfache symmetrische Irrfahrt auf $\mathbb{Z}^d$ besitzt in keiner Dimension einen positiv rekurrenten Zustand: Für $d = 1, 2$ sind alle Zustände nullrekurrent, für $d \geq 3$ sogar transient.

(2) Nach den Sätzen 6.15 und 6.18 besitzt jede irreduzible Markov-Kette mit mindestens einem positiv rekurrenten Zustand eine eindeutige stationäre Verteilung.

7 Maßtheoretische Grundlagen der Wahrscheinlichkeitstheorie

Zur Erinnerung: Ein wahrscheinlichkeitstheoretisches Modell besteht aus

(1) einem Ergebnisraum Ω, z. B. $[0,1]$, $\mathbb{N}$, $\mathbb{R}$, ...;

(2) einer σ-Algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$, z. B. $\mathcal{P}(\Omega)$, $\mathcal{B}_{[0,1]}$, $\mathcal{B}$;

(3) einem Wahrscheinlichkeitsmaß $\mathbb{P}\colon \mathcal{F} \to [0,1]$, z. B. dem Lebesgue-Maß auf $\mathcal{B}_{[0,1]}$.

Offen geblieben ist bisher die Frage: Wie konstruiert man ein Wahrscheinlichkeitsmaß auf einer σ-Algebra $\mathcal{F}$, insbesondere das Lebesgue-Maß auf $(\mathbb{R}^d, \mathcal{B}^d)$? (Bisher haben wir die Existenz des Lebesgue-Integrals in Tatsache 1.14 vorausgesetzt.)

7.1 Abstrakte Maßräume

Sei Ω eine nichtleere Menge.

Definition 7.1 (Algebra)**.** Ein Mengensystem $\mathcal{A}_0 \subset \mathcal{P}(\Omega)$ heißt *Algebra* über Ω, falls gilt:

(a) $\Omega \in \mathcal{A}_0$,

(b) $A \in \mathcal{A}_0 \;\Rightarrow\; A^{\mathrm{c}} \in \mathcal{A}_0$,

(c) $A, B \in \mathcal{A}_0 \;\Rightarrow\; A \cup B \in \mathcal{A}_0$.

Bemerkung.

(1) Für $A, B \in \mathcal{A}_0$ gilt auch

$$A \cap B = (A^{\mathrm{c}} \cup B^{\mathrm{c}})^{\mathrm{c}} \in \mathcal{A}_0 \qquad \text{und} \qquad A \triangle B = (A \cup B) \setminus (A \cap B) \in \mathcal{A}_0;$$

insbesondere sind Algebren $\cap$-stabil.

(2) Algebren unterscheiden sich von σ-Algebren dadurch, dass nur die Abgeschlossenheit unter *endlichen* Vereinigungen gefordert wird. Algebren sind oft viel einfacher zu beschreiben.

Definition 7.2 (Mengenfunktion)**.** Sei $\mathcal{A}_0$ eine Algebra über Ω und $\mu_0\colon \mathcal{A}_0 \to [0,\infty]$ eine *Mengenfunktion.*

(a) μ_0 heißt *additiv*, falls $\mu_0(\emptyset) = 0$ und für alle $A, B \in \mathcal{A}_0$ mit $A \cap B = \emptyset$ gilt

$$\mu_0(A \cup B) = \mu_0(A) + \mu_0(B).$$

(b) μ_0 heißt *σ-additiv*, falls $\mu_0(\emptyset) = 0$ und für alle paarweise disjunkten $(A_n)_{n\geq 0} \subset \mathcal{A}_0$ mit $\bigcup_{n\geq 0} A_n \in \mathcal{A}_0$ gilt

$$\mu_0\Big(\bigcup_{n\geq 0} A_n\Big) = \sum_{n\geq 0} \mu_0(A_n).$$

Definition 7.3 (Maß, Maßraum)**.** Ist $\mathcal{A}$ eine σ-Algebra und $\mu\colon \mathcal{A} \to [0,\infty]$ σ-additiv, so heißt μ ein *Maß* auf $(\Omega, \mathcal{A})$ und $(\Omega, \mathcal{A}, \mu)$ ein *Maßraum*. Ist $\mu(\Omega) = 1$, so ist μ ein Wahrscheinlichkeitsmaß. Ist $\mu(\Omega) < \infty$, so heißt μ ein *endliches Maß*. Gibt es $(A_n)_{n\geq 0} \subset \mathcal{A}$ mit $\mu(A_n) < \infty$ für alle n und $\bigcup_{n\geq 0} A_n = \Omega$, so heißt μ *σ-endlich*. Ist $A \in \mathcal{A}$ mit $\mu(A) = 0$, so heißt A eine *μ-Nullmenge*.

Wir kommen nun zur zentralen Konstruktionsaussage für Maße.

Satz 7.4 (Fortsetzungssatz von Carathéodory)**.** *Sei $\mathcal{A}_0$ eine Algebra über Ω und $\mathcal{A} = \sigma(\mathcal{A}_0)$. Ist $\mu_0\colon \mathcal{A}_0 \to [0,\infty]$ eine σ-additive Mengenfunktion, so gibt es ein Maß μ auf $(\Omega, \mathcal{A})$ mit*

$$\mu(A) = \mu_0(A) \qquad \text{für alle } A \in \mathcal{A}_0.$$

*Ist μ_0 σ-*endlich*, d. h. gibt es $\Omega_1, \Omega_2, \ldots \in \mathcal{A}_0$ mit $\bigcup_{n\geq 1} \Omega_n = \Omega$ und $\mu_0(\Omega_n) < \infty$ für alle n, so ist μ eindeutig. Dies gilt insbesondere, wenn $\mu_0(\Omega) < \infty$.*

Bemerkung. Ohne σ-Endlichkeit ist die Fortsetzung im Allgemeinen nicht eindeutig. Sei etwa $\Omega = \mathbb{Q} \cap (0,1]$, $\mathcal{A}_0$ das System der endlichen Vereinigungen von Mengen der Form $(a,b] \cap \Omega$ mit $0 \leq a \leq b \leq 1$, und $\mu_0(\emptyset) = 0$, $\mu_0(A) = \infty$ für $A \neq \emptyset$. Dann ist $\mathcal{A}_0$ eine Algebra, μ_0 ist σ-additiv (eine nichtleere abzählbare disjunkte Vereinigung enthält eine nichtleere Menge), und es gilt $\sigma(\mathcal{A}_0) = \mathcal{P}(\Omega)$, da $\{q\} = \bigcap_{n\geq 1} (\max\{0, q - 1/n\}, q] \cap \Omega$ und Ω abzählbar ist. Sowohl das Zählmaß $\mu(A) = |A|$ als auch 2μ setzen μ_0 fort.

Beispiel 7.5 (Konstruktion des Lebesgue-Maßes)**.** Sei $\Omega = (0,1]$ und $\mathcal{A}_0$ die Menge aller endlichen Vereinigungen

$$A = (a_1, b_1] \cup \cdots \cup (a_r, b_r]$$

mit $r \in \mathbb{N}_0$ und $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_r < b_r \leq 1$ (für $r = 0$ ist $A = \emptyset$). Dann ist $\mathcal{A}_0$ eine Algebra über Ω, und es gilt $\sigma(\mathcal{A}_0) = \mathcal{B}_{(0,1]}$. Wir definieren

$$\mu_0(A) \coloneqq \sum_{k=1}^{r} (b_k - a_k);$$

dieser Wert hängt nicht von der gewählten Darstellung von A ab, und μ_0 ist additiv, da sich die Längen disjunkter Intervalle addieren (und damit monoton und endlich subadditiv auf $\mathcal{A}_0$).

μ_0 ist σ-additiv. Es ist $\mu_0(\emptyset) = 0$. Seien $(A_n)_{n\geq 0} \subset \mathcal{A}_0$ paarweise disjunkt mit $A \coloneqq \bigcup_{n\geq 0} A_n \in \mathcal{A}_0$, und sei $G_n \coloneqq \bigcup_{k\leq n} A_k$, also $G_n \uparrow A$. Wegen der Additivität von μ_0 gilt

$$\mu_0(G_n) = \sum_{k\leq n} \mu_0(A_k) \xrightarrow[n\to\infty]{} \sum_{n\geq 0} \mu_0(A_n).$$

Es genügt also zu zeigen, dass $\mu_0(G_n) \uparrow \mu_0(A)$. Sei $H_n \coloneqq A \setminus G_n \in \mathcal{A}_0$; dann gilt $H_n \downarrow \emptyset$ und $\mu_0(A) = \mu_0(G_n) + \mu_0(H_n)$. Es genügt also zu zeigen, dass $\mu_0(H_n) \downarrow 0$. Angenommen, dies ist falsch. Dann gibt es ein $\varepsilon > 0$ mit $\mu_0(H_n) \geq 2\varepsilon$ für alle n, und wir zeigen, dass dann $\bigcap_n H_n \neq \emptyset$ ist, im Widerspruch zu $H_n \downarrow \emptyset$.

Zu jedem H_k gibt es (durch geringfügiges Vergrößern der linken Intervallendpunkte) ein $J_k \in \mathcal{A}_0$ mit $\overline{J_k} \subset H_k$ und $\mu_0(H_k \setminus J_k) \le \varepsilon 2^{-(k+1)}$. Wegen $H_n \subset H_k$ für $k \le n$ folgt

$$\mu_0\Big(H_n \setminus \bigcap_{k\le n} J_k\Big) \le \mu_0\Big(\bigcup_{k\le n} (H_k \setminus J_k)\Big) \le \sum_{k=0}^{n} \varepsilon 2^{-(k+1)} < \varepsilon,$$

und daher, da $\mu_0(H_n) \ge 2\varepsilon$ für alle n,

$$\mu_0\Big(\bigcap_{k\le n} J_k\Big) = \mu_0(H_n) - \mu_0\Big(H_n \setminus \bigcap_{k\le n} J_k\Big) > \varepsilon.$$

Insbesondere ist

$$K_n := \bigcap_{k\le n} \overline{J_k} \supset \bigcap_{k\le n} J_k \neq \emptyset \qquad \text{für alle } n.$$

Wähle $x_n \in K_n$. Dann ist $(x_n)_{n\ge 0}$ eine Folge in der kompakten Menge $\overline{J_0}$, besitzt also eine Teilfolge $x_{n_r} \to x \in \overline{J_0}$. Für jedes k liegen alle bis auf endlich viele x_{n_r} in $\overline{J_k}$ (nämlich alle mit $n_r \ge k$), und da $\overline{J_k}$ abgeschlossen ist, folgt $x \in \overline{J_k}$ für alle k. Also

$$x \in \bigcap_{k\ge 0} \overline{J_k} \subset \bigcap_{n\ge 0} H_n,$$

d. h. $\bigcap_n H_n \neq \emptyset$, ein Widerspruch.

Nach dem Fortsetzungssatz 7.4 gibt es daher ein Maß μ auf $\big((0,1], \mathcal{B}_{(0,1]}\big)$ mit $\mu = \mu_0$ auf $\mathcal{A}_0$, das sogenannte *Lebesgue-Maß* λ auf $(0,1]$; es ist eindeutig, da $\mu_0(\Omega) = 1 < \infty$.

Bemerkung.

(1) Durch Zusammenkleben (der Lebesgue-Maße auf den Intervallen $(k, k+1]$, $k \in \mathbb{Z}$) kann das Lebesgue-Maß auch auf $\mathbb{R}$ erklärt werden.

(2) Durch Hinzunahme von Nullmengen kann das Lebesgue-Maß sogar auf eine noch größere σ-Algebra als die Borel'sche σ-Algebra fortgesetzt werden, die sogenannte *Lebesgue'sche σ-Algebra* (siehe Bemerkung 7.10).

Strategie für den Beweis von Satz 7.4.

(1) Wir konstruieren eine Kandidaten-Mengenfunktion μ^* auf $(\Omega, \mathcal{P}(\Omega))$.

(2) Wir zeigen, dass $\mu^* = \mu_0$ auf $\mathcal{A}_0$.

(3) Wir finden eine σ-Algebra $\mathcal{L}$ auf Ω, auf der μ^* ein Maß ist.

(4) Wir zeigen $\mathcal{A}_0 \subset \mathcal{L}$; dann ist auch $\mathcal{A} = \sigma(\mathcal{A}_0) \subset \mathcal{L}$, und die Einschränkung von μ^* auf $\mathcal{A}$ ist das gesuchte Maß.

Dazu benötigen wir einige Vorbereitungen.

Lemma 7.6 (μ^*-Mengen)**.** *Sei $\mathcal{G}_0$ eine Algebra über Ω und $\mu^*\colon \mathcal{G}_0 \to [0,\infty]$ eine Mengenfunktion mit $\mu^*(\emptyset) = 0$. Wir nennen $L \in \mathcal{G}_0$ eine* μ^*-Menge*, falls*

$$\mu^*(L \cap G) + \mu^*(L^c \cap G) = \mu^*(G) \qquad \textit{für alle } G \in \mathcal{G}_0.$$

Dann ist das System $\mathcal{L}_0$ der μ^*-Mengen eine Algebra, und μ^* ist additiv auf $\mathcal{L}_0$. Weiterhin gilt für paarweise disjunkte $L_1, \dots, L_n \in \mathcal{L}_0$ und $G \in \mathcal{G}_0$

$$\mu^*\Big(\bigcup_{k=1}^{n} (L_k \cap G)\Big) = \sum_{k=1}^{n} \mu^*(L_k \cap G). \tag{$*$}$$

Anschaulich: μ^*-Mengen zerlegen jede Menge in $\mathcal{G}_0$ „richtig“.

Beweis. (1) $\mathcal{L}_0$ ist eine Algebra.

(a) $\Omega \in \mathcal{L}_0$, denn $\mu^*(\Omega \cap G) + \mu^*(\emptyset \cap G) = \mu^*(G) + 0$.

(b) $L \in \mathcal{L}_0 \Rightarrow L^c \in \mathcal{L}_0$, da die Definition symmetrisch in L und L^c ist.

(c) Seien $L_1, L_2 \in \mathcal{L}_0$ und $L \coloneqq L_1 \cap L_2$. Dann gilt

$$L^c \cap L_2 = L_2 \cap L_1^c \qquad \text{und} \qquad L^c \cap L_2^c = L_2^c.$$

Für $G \in \mathcal{G}_0$ folgt, da L_2 eine μ^*-Menge ist (angewandt auf $L^c \cap G \in \mathcal{G}_0$),

$$\mu^*(L^c \cap G) = \mu^*\big(L_2 \cap (L^c \cap G)\big) + \mu^*\big(L_2^c \cap (L^c \cap G)\big) = \mu^*(L_2 \cap L_1^c \cap G) + \mu^*(L_2^c \cap G),$$

und daher, da L_1 eine μ^*-Menge ist (angewandt auf $L_2 \cap G$) und ebenso L_2,

$$\begin{aligned}\mu^*(L^c \cap G) + \mu^*(L \cap G) &= \mu^*(L_1^c \cap L_2 \cap G) + \mu^*(L_1 \cap L_2 \cap G) + \mu^*(L_2^c \cap G) \\ &= \mu^*(L_2 \cap G) + \mu^*(L_2^c \cap G) = \mu^*(G).\end{aligned}$$

Also ist $L_1 \cap L_2 \in \mathcal{L}_0$, und mit (b) und $L_1 \cup L_2 = (L_1^c \cap L_2^c)^c$ auch $L_1 \cup L_2 \in \mathcal{L}_0$.

(2) Additivität. Für $L_1, L_2 \in \mathcal{L}_0$ mit $L_1 \cap L_2 = \emptyset$ gilt

$$(L_1 \cup L_2) \cap L_1 = L_1 \qquad \text{und} \qquad (L_1 \cup L_2) \cap L_1^c = L_2.$$

Für $G \in \mathcal{G}_0$ folgt daher, da L_1 eine μ^*-Menge ist,

$$\begin{aligned}\mu^*\big((L_1 \cap G) \cup (L_2 \cap G)\big) &= \mu^*\big((L_1 \cup L_2) \cap G\big) \\ &= \mu^*\big(((L_1 \cup L_2) \cap G) \cap L_1\big) + \mu^*\big(((L_1 \cup L_2) \cap G) \cap L_1^c\big) \\ &= \mu^*(L_1 \cap G) + \mu^*(L_2 \cap G).\end{aligned}$$

Für $G = \Omega$ folgt die Additivität von μ^* auf $\mathcal{L}_0$.

(3) Endliche Iteration des Arguments in (2) liefert (∗). □

Definition 7.7 (Äußeres Maß)**.** Sei $(\Omega, \mathcal{A})$ ein Ereignisraum. Eine Mengenfunktion $\mu^*\colon \mathcal{A} \to [0, \infty]$ heißt *äußeres Maß* auf $(\Omega, \mathcal{A})$, falls gilt:

(1) $\mu^*(\emptyset) = 0$;

(2) μ^* ist *monoton* (wachsend), d. h. für alle $A_1, A_2 \in \mathcal{A}$ gilt $A_1 \subset A_2 \Rightarrow \mu^*(A_1) \leq \mu^*(A_2)$;

(3) μ^* ist *σ-subadditiv*, d. h. für alle $(A_i)_{i \geq 0} \subset \mathcal{A}$ gilt

$$\mu^*\Big(\bigcup_{i \geq 0} A_i\Big) \leq \sum_{i \geq 0} \mu^*(A_i).$$

Bemerkung. Jedes Maß ist ein äußeres Maß (vgl. Satz 1.11, dessen Beweis sinngemäß auch für Maße gilt).

Im folgenden Lemma zeigen wir, dass die μ^*-Mengen eines äußeren Maßes nicht nur eine Algebra, sondern sogar eine σ-Algebra bilden, auf der μ^* σ-additiv ist.

Lemma 7.8 (Lemma von Carathéodory)**.** *Sei μ^* ein äußeres Maß auf $(\Omega, \mathcal{A})$. Dann bildet das System $\mathcal{L}$ der μ^*-Mengen in $\mathcal{A}$ eine σ-Algebra, und μ^* ist auf $\mathcal{L}$ sogar σ-additiv, d. h. ein Maß. Insbesondere ist $(\Omega, \mathcal{L}, \mu^*)$ ein Maßraum.*

Beweis. Nach Lemma 7.6 (mit $\mathcal{G}_0 = \mathcal{A}$) wissen wir bereits, dass $\mathcal{L}$ eine Algebra ist. Da sich jede Folge (L_i) in $\mathcal{L}$ durch $L_i' := L_i \setminus \bigcup_{j<i} L_j \in \mathcal{L}$ in eine disjunkte Folge mit derselben Vereinigung verwandeln lässt, genügt es zu zeigen: Ist $(L_i)_{i\geq 0}$ eine Folge paarweise disjunkter Mengen in $\mathcal{L}$, so ist $L := \bigcup_{i\geq 0} L_i$ eine μ^*-Menge, und es gilt

$$\mu^*(L) = \sum_{i\geq 0} \mu^*(L_i).$$

Es gilt $L \in \mathcal{A}$, da $\mathcal{A}$ eine σ-Algebra ist. Sei nun $A \in \mathcal{A}$. Da μ^* σ-subadditiv ist und $\mu^*(\emptyset) = 0$ gilt, ist μ^* auch endlich subadditiv, also

$$\mu^*(A) \leq \mu^*(L \cap A) + \mu^*(L^{\mathrm{c}} \cap A). \tag{7.1}$$

Es bleibt die umgekehrte Ungleichung zu zeigen. Für $n \geq 0$ ist $M_n := \bigcup_{k\leq n} L_k \in \mathcal{L}$, da $\mathcal{L}$ eine Algebra ist. Also gilt

$$\mu^*(A) = \mu^*(M_n \cap A) + \mu^*(M_n^{\mathrm{c}} \cap A).$$

Wegen $L^{\mathrm{c}} \subset M_n^{\mathrm{c}}$ und der Monotonie von μ^* folgt

$$\mu^*(A) \geq \mu^*(M_n \cap A) + \mu^*(L^{\mathrm{c}} \cap A) = \sum_{k\leq n} \mu^*(L_k \cap A) + \mu^*(L^{\mathrm{c}} \cap A) \qquad \text{für alle } n \geq 0,$$

wobei wir im letzten Schritt die Eigenschaft $(*)$ aus Lemma 7.6 (mit $G = A$) verwendet haben. Im Grenzübergang $n \to \infty$ erhalten wir mit der σ-Subadditivität von μ^*

$$\mu^*(A) \geq \sum_{k\geq 0} \mu^*(L_k \cap A) + \mu^*(L^{\mathrm{c}} \cap A) \geq \mu^*(L \cap A) + \mu^*(L^{\mathrm{c}} \cap A).$$

Zusammen mit (7.1) gilt hier überall Gleichheit, d. h.

$$\mu^*(A) = \sum_{k\geq 0} \mu^*(L_k \cap A) + \mu^*(L^{\mathrm{c}} \cap A) = \mu^*(L \cap A) + \mu^*(L^{\mathrm{c}} \cap A) \qquad \text{für alle } A \in \mathcal{A}.$$

Die zweite Gleichung besagt $L \in \mathcal{L}$; also ist $\mathcal{L}$ eine σ-Algebra. Wählt man in der ersten Gleichung $A = L$, so folgt $\mu^*(L) = \sum_{k\geq 0} \mu^*(L_k)$, d. h. μ^* ist σ-additiv auf $\mathcal{L}$. □

Damit können wir den Beweis des Fortsetzungssatzes abschließen, indem wir die zuvor skizzierte Strategie ausführen.

Beweis von Satz 7.4. Wir betrachten den Ereignisraum $(\Omega, \mathcal{P}(\Omega))$ und setzen für $A \subset \Omega$

$$\mu^*(A) := \inf\Big\{\sum_{n\geq 0} \mu_0(A_n) \colon (A_n)_{n\geq 0} \subset \mathcal{A}_0 \text{ mit } A \subset \bigcup_{n\geq 0} A_n\Big\}.$$

Die Menge, über die das Infimum gebildet wird, ist nicht leer, da $\Omega \in \mathcal{A}_0$ (wähle $A_0 = \Omega$, $A_n = \emptyset$ für $n \geq 1$).

Schritt 1: μ^* ist ein äußeres Maß auf $(\Omega, \mathcal{P}(\Omega))$.

(a) Es gilt $\mu^*(\emptyset) = 0$, da μ_0 als σ-additive Mengenfunktion $\mu_0(\emptyset) = 0$ erfüllt und $\emptyset$ durch $A_n = \emptyset$, $n \geq 0$, überdeckt wird.

(b) μ^* ist wachsend: Ist $A \subset B$, so ist jede Überdeckung von B durch Mengen aus $\mathcal{A}_0$ auch eine Überdeckung von A, also $\mu^*(A) \le \mu^*(B)$.

(c) μ^* ist σ-subadditiv: Sei $(G_n)_{n\ge1} \subset \mathcal{P}(\Omega)$ und $G \coloneqq \bigcup_{n\ge1} G_n$. Ist $\mu^*(G_n) = \infty$ für ein n, so ist nichts zu zeigen; sei also $\mu^*(G_n) < \infty$ für alle n. Zu $\varepsilon > 0$ und $n \ge 1$ gibt es nach Definition von μ^* eine Folge $(F_{n,k})_{k\ge0} \subset \mathcal{A}_0$ mit

$$G_n \subset \bigcup_{k\ge0} F_{n,k} \qquad \text{und} \qquad \sum_{k\ge0} \mu_0(F_{n,k}) \le \mu^*(G_n) + \varepsilon 2^{-n}.$$

Wegen $G \subset \bigcup_{n\ge1}\bigcup_{k\ge0} F_{n,k}$ (abzählbare Überdeckung durch Mengen aus $\mathcal{A}_0$) folgt

$$\mu^*(G) \le \sum_{n\ge1}\sum_{k\ge0} \mu_0(F_{n,k}) \le \sum_{n\ge1} \mu^*(G_n) + \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, gilt $\mu^*(G) \le \sum_{n\ge1} \mu^*(G_n)$.

Schritt 2: $\mu^* = \mu_0$ auf $\mathcal{A}_0$. Sei $A \in \mathcal{A}_0$.

(a) Es gilt $\mu^*(A) \le \mu_0(A)$ nach Definition, da $A, \emptyset, \emptyset, \dots$ eine zulässige Überdeckung ist.

(b) Sei umgekehrt $(A_n)_{n\ge1} \subset \mathcal{A}_0$ mit $A \subset \bigcup_{n\ge1} A_n$. Wir definieren

$$E_1 \coloneqq A_1, \qquad E_n \coloneqq A_n \cap \Big(\bigcup_{k<n} A_k\Big)^{\mathrm{c}} \quad (n \ge 2).$$

Da $\mathcal{A}_0$ eine Algebra ist, gilt $E_n \in \mathcal{A}_0$; ferner ist (E_n) eine disjunkte Folge mit $E_n \subset A_n$ und $\bigcup_n E_n = \bigcup_n A_n \supset A$. Somit ist $A = \bigcup_n (A \cap E_n)$ eine disjunkte Vereinigung von Mengen aus $\mathcal{A}_0$, deren Vereinigung wieder in $\mathcal{A}_0$ liegt. Da μ_0 nach Voraussetzung σ-additiv auf $\mathcal{A}_0$ ist (und damit insbesondere monoton, denn für $B \subset C$ in $\mathcal{A}_0$ ist $\mu_0(C) = \mu_0(B) + \mu_0(C \setminus B)$), folgt

$$\mu_0(A) = \mu_0\Big(\bigcup_{n\ge1} (A \cap E_n)\Big) = \sum_{n\ge1} \mu_0(A \cap E_n) \le \sum_{n\ge1} \mu_0(E_n) \le \sum_{n\ge1} \mu_0(A_n).$$

Bilden wir das Infimum über alle solchen Überdeckungen, so folgt $\mu_0(A) \le \mu^*(A)$.

Nach Schritt 1 und Lemma 7.8 ist μ^* insbesondere ein Maß auf $(\Omega, \mathcal{L})$, wobei $\mathcal{L}$ die σ-Algebra der μ^*-Mengen in $\mathcal{P}(\Omega)$ bezeichnet.

Schritt 3: $\mathcal{A}_0 \subset \mathcal{L}$. Sei $E \in \mathcal{A}_0$ und $G \subset \Omega$. Wegen der Subadditivität von μ^* gilt $\mu^*(G) \le \mu^*(E \cap G) + \mu^*(E^{\mathrm{c}} \cap G)$; zu zeigen ist die umgekehrte Ungleichung, und dafür dürfen wir $\mu^*(G) < \infty$ annehmen. Zu $\varepsilon > 0$ gibt es $(F_n)_{n\ge0} \subset \mathcal{A}_0$ mit

$$G \subset \bigcup_{n\ge0} F_n \qquad \text{und} \qquad \sum_{n\ge0} \mu_0(F_n) \le \mu^*(G) + \varepsilon.$$

Da μ_0 auf $\mathcal{A}_0$ additiv ist und $E \cap F_n$, $E^{\mathrm{c}} \cap F_n \in \mathcal{A}_0$, folgt

$$\sum_{n\ge0} \mu_0(F_n) = \sum_{n\ge0} \mu_0(E \cap F_n) + \sum_{n\ge0} \mu_0(E^{\mathrm{c}} \cap F_n) \ge \mu^*(E \cap G) + \mu^*(E^{\mathrm{c}} \cap G),$$

wobei wir im letzten Schritt die Definition von μ^* verwendet haben: Es gilt $E \cap G \subset \bigcup_n (E \cap F_n)$ und $E^{\mathrm{c}} \cap G \subset \bigcup_n (E^{\mathrm{c}} \cap F_n)$. Da $\varepsilon > 0$ beliebig war, folgt

$$\mu^*(G) \ge \mu^*(E \cap G) + \mu^*(E^{\mathrm{c}} \cap G),$$

also Gleichheit. Somit ist E eine μ^*-Menge, d. h. $\mathcal{A}_0 \subset \mathcal{L}$.

Abschluss. Da $\mathcal{L}$ eine σ-Algebra ist, folgt $\mathcal{A} = \sigma(\mathcal{A}_0) \subset \mathcal{L}$. Die Einschränkung $\mu := \mu^*|_{\mathcal{A}}$ ist daher ein Maß auf $(\Omega, \mathcal{A})$, das nach Schritt 2 auf $\mathcal{A}_0$ mit μ_0 übereinstimmt.

Eindeutigkeit. Seien μ, ν zwei Maße auf $(\Omega, \mathcal{A})$ mit $\mu = \nu = \mu_0$ auf $\mathcal{A}_0$.

Fall $\mu_0(\Omega) < \infty$. Das System $\mathcal{D} := \{A \in \mathcal{A} \colon \mu(A) = \nu(A)\}$ enthält $\mathcal{A}_0$ und ist ein Dynkin-System: Es ist $\Omega \in \mathcal{A}_0 \subset \mathcal{D}$; für $A \subset B$ aus $\mathcal{D}$ gilt $\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$, wobei die Differenzen wegen $\mu(\Omega) = \nu(\Omega) = \mu_0(\Omega) < \infty$ wohldefiniert sind; und die Abgeschlossenheit unter disjunkten Vereinigungen folgt aus der σ-Additivität. Da die Algebra $\mathcal{A}_0$ $\cap$-stabil ist, folgt mit Lemma 1.13

$$\mathcal{D} \supset d(\mathcal{A}_0) = \sigma(\mathcal{A}_0) = \mathcal{A},$$

also $\mu = \nu$. (Dies ist genau das Argument aus dem Beweis von Satz 1.12, das dort für Wahrscheinlichkeitsmaße formuliert wurde.)

σ-endlicher Fall. Seien $\Omega_n \in \mathcal{A}_0$ wie in Satz 7.4. Indem wir Ω_n durch $\Omega_1 \cup \cdots \cup \Omega_n \in \mathcal{A}_0$ ersetzen, dürfen wir $\Omega_n \uparrow \Omega$ annehmen; wegen der Subadditivität von μ_0 bleibt $\mu_0(\Omega_n) < \infty$. Für festes n sind $\mu_n := \mu(\,\cdot \cap \Omega_n)$ und $\nu_n := \nu(\,\cdot \cap \Omega_n)$ endliche Maße auf $(\Omega, \mathcal{A})$, die auf $\mathcal{A}_0$ übereinstimmen (denn $A \cap \Omega_n \in \mathcal{A}_0$ für $A \in \mathcal{A}_0$). Nach dem endlichen Fall gilt $\mu_n = \nu_n$ auf $\mathcal{A}$. Mit der σ-Stetigkeit von Maßen (vgl. Satz 1.11(e)) folgt für alle $A \in \mathcal{A}$

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

Beispiel 7.9. (1) *Lebesgue-Stieltjes-Maße.* Die Funktion $F \colon \mathbb{R} \to \mathbb{R}$ sei nicht fallend und rechtsstetig. Durch

$$\mu_F((a,b]) := F(b) - F(a), \qquad a < b,$$

und Additivität wird eine Mengenfunktion auf der Algebra der endlichen disjunkten Vereinigungen von Intervallen $(a, b]$ und (a, ∞) mit $-\infty \le a < b < \infty$ erklärt (dabei sei $F(\pm\infty) := \lim_{x\to\pm\infty} F(x) \in \overline{\mathbb{R}}$, und für $a = -\infty$ bzw. das Intervall (a, ∞) ist die Formel mit diesen Grenzwerten zu lesen). Analog zu Beispiel 7.5 zeigt man mit Hilfe der Rechtsstetigkeit von F, dass diese σ-additiv ist (Übung). Nach Satz 7.4 gibt es also ein Maß μ_F auf $(\mathbb{R}, \mathcal{B})$ mit $\mu_F((a,b]) = F(b) - F(a)$, und es ist eindeutig, da $\mu_F((-n,n]) = F(n) - F(-n) < \infty$ und $(-n,n] \uparrow \mathbb{R}$. Es heißt das *Lebesgue-Stieltjes-Maß* zu F. Gilt zusätzlich

$$\lim_{x\to-\infty} F(x) = 0 \qquad \text{und} \qquad \lim_{x\to\infty} F(x) = 1,$$

so ist μ_F ein Wahrscheinlichkeitsmaß.

(2) *Lebesgue-Maß auf $\mathbb{R}^d$.* Für $a = (a_1, \dots, a_d) \in [-\infty, \infty)^d$, $b = (b_1, \dots, b_d) \in (-\infty, \infty]^d$ mit $a_i < b_i$ für alle $i \le d$ setzt man (mit $(a_i, b_i]$ gelesen als $(a_i, b_i] \cap \mathbb{R}$, wie in (1))

$$\lambda^d((a,b]) := \lambda^d((a_1,b_1] \times \cdots \times (a_d,b_d]) := \prod_{i=1}^d (b_i - a_i)$$

und erhält auf analoge Weise das Lebesgue-Maß λ^d auf $(\mathbb{R}^d, \mathcal{B}^d)$: Die endlichen disjunkten Vereinigungen solcher (auch unbeschränkter) Quader bilden eine Algebra, die σ-Additivität zeigt man wie in Beispiel 7.5, und dann wendet man Satz 7.4 an. Es ist eindeutig, da λ^d auf den Würfeln $(-n,n]^d \uparrow \mathbb{R}^d$ endlich ist.

Bemerkung 7.10. Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und

$$\mathcal{N} \coloneqq \{N \subset \Omega\colon \text{es gibt } A \in \mathcal{A} \text{ mit } \mu(A) = 0 \text{ und } N \subset A\}$$

das System der Teilmengen von μ-Nullmengen. Man beachte, dass die Mengen $N \in \mathcal{N}$ *nicht* notwendig in $\mathcal{A}$ liegen! Man definiert den *Abschluss* (die *Vervollständigung*) $\mathcal{A}^*$ von $\mathcal{A}$ als das System aller $B \in \mathcal{P}(\Omega)$, zu denen es $B_1, B_2 \in \mathcal{A}$ gibt mit

$$B_1 \subset B \subset B_2 \qquad \text{und} \qquad \mu(B_2 \setminus B_1) = 0.$$

Dann ist $\mathcal{A}^*$ eine σ-Algebra, und durch $\mu(B) \coloneqq \mu(B_1)$ wird μ zu einem Maß auf $\mathcal{A}^*$ fortgesetzt (ohne Beweis). Insbesondere enthält $\mathcal{A}^*$ alle Mengen aus $\mathcal{N}$. Die Vervollständigung $\mathcal{B}^*$ der Borel'schen σ-Algebra $\mathcal{B}$ auf $\mathbb{R}$ bezüglich des Lebesgue-Maßes λ (Beispiel 7.9) heißt die *Lebesgue'sche σ-Algebra*. Für ein Intervall $I \subset \mathbb{R}$ setzen wir $\mathcal{B}_I^* \coloneqq \{A \cap I\colon A \in \mathcal{B}^*\}$; insbesondere ist $\mathcal{B}_{[0,1]}^*$ die Lebesgue'sche σ-Algebra auf $[0,1]$.

7.2 Lebesgue-Integration auf abstrakten Maßräumen

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Wir geben den folgenden Ausdrücken einen Sinn:

$$\mu(f) = \langle f, \mu\rangle = \int_\Omega f \,\mathrm{d}\mu = \int_\Omega f(\omega)\,\mu(\mathrm{d}\omega).$$

Diese Schreibweisen bezeichnen alle das *Lebesgue-Integral* einer messbaren Abbildung $f\colon \Omega \to \overline{\mathbb{R}}$ (die also auch die Werte $\pm\infty$ annehmen darf) bezüglich μ. Ist $\mu = \lambda$ das Lebesgue-Maß, so erhalten wir das Lebesgue-Integral, das wir im ersten Teil der Vorlesung (Tatsache 1.14) benutzt haben.

Wir bauen das Integral gemäß der „maßtheoretischen Induktion“ schrittweise auf.

Schritt 1 (Indikatorfunktionen). Für $A \in \mathcal{A}$ setzen wir

$$\mu(\mathbb{1}_A) \coloneqq \int_\Omega \mathbb{1}_A \,\mathrm{d}\mu = \int_A \mathrm{d}\mu \coloneqq \mu(A),$$

wobei $\mathbb{1}_A(\omega) = 1$ für $\omega \in A$ und $\mathbb{1}_A(\omega) = 0$ sonst.

Schritt 2 (einfache Funktionen). Eine messbare Funktion $f\colon \Omega \to \mathbb{R}$ heißt *einfach*, falls

$$f = \sum_{k=1}^n a_k \mathbb{1}_{A_k} \qquad \text{mit } n \in \mathbb{N},\ a_k \in \mathbb{R} \text{ und paarweise disjunkten } A_k \in \mathcal{A}.$$

Ist f einfach und $f \geq 0$ (wir dürfen dann $a_k \geq 0$ annehmen), so setzen wir

$$\mu(f) \coloneqq \sum_{k=1}^n a_k\,\mu(A_k) \in [0, \infty]$$

mit der Konvention $0 \cdot \infty = 0$. Die Einschränkung auf $f \geq 0$ vermeidet unbestimmte Ausdrücke der Form $\infty - \infty$.

Diese Definition hängt nicht von der Darstellung von f ab: Sei

$$f = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i} = \sum_{j=1}^n \beta_j \mathbb{1}_{B_j}$$

mit paarweise disjunkten A_i bzw. B_j. Ist $A_i \cap B_j \neq \emptyset$, so gilt $\alpha_i = f(\omega) = \beta_j$ für jedes $\omega \in A_i \cap B_j$. Außerdem gilt $A_i \subset \bigcup_j B_j$, falls $\alpha_i \neq 0$, und $B_j \subset \bigcup_i A_i$, falls $\beta_j \neq 0$. Mit der Additivität von μ folgt

$$\sum_{i=1}^{m} \alpha_i \, \mu(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \, \mu(A_i \cap B_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_j \, \mu(A_i \cap B_j) = \sum_{j=1}^{n} \beta_j \, \mu(B_j).$$

Indem man zwei einfache Funktionen bezüglich der gemeinsamen Verfeinerung $(A_i \cap B_j)_{i,j}$ darstellt, erhält man ebenso für einfache $f, g \geq 0$:

(1) $\mu(f \neq g) = 0 \;\Rightarrow\; \mu(f) = \mu(g)$;

(2) $\mu(\alpha f + \beta g) = \alpha \, \mu(f) + \beta \, \mu(g)$ für alle $\alpha, \beta \geq 0$;

(3) $f \leq g \;\Rightarrow\; \mu(f) \leq \mu(g)$ (denn $g - f \geq 0$ ist einfach und $\mu(g) = \mu(f) + \mu(g - f)$ nach (2)).

Schritt 3 (nichtnegative messbare Funktionen). Für messbares $f \colon \Omega \to [0, \infty]$ setzen wir

$$\mu(f) \coloneqq \sup\{\mu(h) \colon h \text{ einfach und } 0 \leq h \leq f\}.$$

Insbesondere gilt:

(1) Für einfaches $f \geq 0$ ist diese Definition wegen Schritt 2 (3) konsistent mit Schritt 2.

(2) $f \leq g \;\Rightarrow\; \mu(f) \leq \mu(g)$; das folgt direkt aus der Definition als Supremum.

(3) *Monotone Konvergenz:* Aus $f_n \uparrow f$ (punktweise, $f_n \geq 0$ messbar) folgt $\mu(f_n) \uparrow \mu(f)$. Denn:

   a) Nach (2) ist $\mu(f_n)$ wachsend und $\lim_{n\to\infty} \mu(f_n) = \sup_n \mu(f_n) \leq \mu(f)$.

   b) Sei $h = \sum_{i=1}^{m} \alpha_i \mathbb{1}_{A_i}$ einfach mit $0 \leq h \leq f$. Es genügt zu zeigen, dass $\sup_n \mu(f_n) \geq \mu(h)$. Sei dazu $\varepsilon \in (0, 1)$ und

$$B_n^\varepsilon \coloneqq \{f_n \geq (1 - \varepsilon)h\} \in \mathcal{A}.$$

   Da $f_n \uparrow f \geq h$ und $f(\omega) > (1 - \varepsilon)h(\omega)$ für alle ω mit $h(\omega) > 0$, gilt $B_n^\varepsilon \uparrow \Omega$ für jedes ε. Mit (2) folgt

$$\mu(f_n) \geq \int (1 - \varepsilon) \, h \, \mathbb{1}_{B_n^\varepsilon} \, \mathrm{d}\mu = \sum_{i=1}^{m} (1 - \varepsilon) \, \alpha_i \, \mu(A_i \cap B_n^\varepsilon) \xrightarrow[n\to\infty]{} (1 - \varepsilon) \, \mu(h),$$

   wobei wir die σ-Stetigkeit von μ (vgl. Satz 1.11(e); der Beweis für $A_n \uparrow A$ überträgt sich wörtlich auf Maße) benutzt haben. Mit $\varepsilon \downarrow 0$ folgt die Behauptung.

(4) *Linearität:* $\mu(\alpha f + \beta g) = \alpha \, \mu(f) + \beta \, \mu(g)$ für messbare $f, g \geq 0$ und $\alpha, \beta \geq 0$. Denn setzen wir

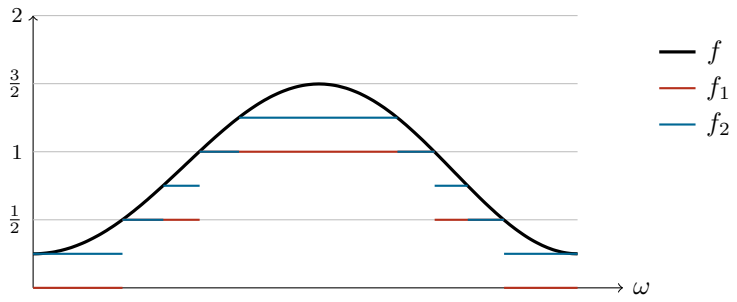
$$f_n \coloneqq \left(2^{-n} \lfloor 2^n f \rfloor\right) \wedge n, \qquad g_n \coloneqq \left(2^{-n} \lfloor 2^n g \rfloor\right) \wedge n,$$

   so nehmen f_n, g_n nur Werte in $\{k 2^{-n} \colon k = 0, 1, \ldots, n 2^n\}$ an, d. h.

$$f_n = \sum_{k=0}^{n 2^n} k 2^{-n} \, \mathbb{1}_{\{f_n = k 2^{-n}\}}$$

   ist einfach (und analog g_n), und es gilt $f_n \uparrow f$, $g_n \uparrow g$. Also gilt $\alpha f_n + \beta g_n \uparrow \alpha f + \beta g$, und mit (3) sowie Schritt 2 (2) folgt

$$\mu(\alpha f + \beta g) = \lim_{n\to\infty} \mu(\alpha f_n + \beta g_n) = \lim_{n\to\infty} \left(\alpha \, \mu(f_n) + \beta \, \mu(g_n)\right) = \alpha \, \mu(f) + \beta \, \mu(g).$$



Die einfachen Funktionen $f_n = (2^{-n}\lfloor 2^n f\rfloor) \wedge n$ aus (4) für $n = 1, 2$ (schematisch, Ω als Intervall gezeichnet): f_n rundet f auf das Raster $2^{-n}\mathbb{Z}$ ab und kappt bei n; f_1 ist durch 1 beschränkt, $f_2 \geq f_1$ ist feiner.

Schritt 4 (messbare Funktionen). Sei nun $f\colon \Omega \to [-\infty, \infty] = \overline{\mathbb{R}}$ messbar, und seien

$$f^+ := \max\{f, 0\}, \qquad f^- := \max\{-f, 0\}$$

der Positiv- und der Negativteil von f. Dann gilt

$$f = f^+ - f^-, \qquad |f| = f^+ + f^-, \qquad f^+ \leq |f|, \quad f^- \leq |f|,$$

und nach Schritt 3 (2) folgt

$$\int f^+ \,\mathrm{d}\mu \leq \int |f| \,\mathrm{d}\mu, \qquad \int f^- \,\mathrm{d}\mu \leq \int |f| \,\mathrm{d}\mu.$$

Insbesondere impliziert $\int |f| \,\mathrm{d}\mu < \infty$, dass $\int f^+ \,\mathrm{d}\mu < \infty$ und $\int f^- \,\mathrm{d}\mu < \infty$.

Definition 7.11 (Integral für messbare Funktionen)**.** Eine messbare Funktion $f\colon \Omega \to \overline{\mathbb{R}}$ heißt *μ-integrierbar*, falls $\int |f| \,\mathrm{d}\mu < \infty$. Wir setzen

$$\mathcal{L}^1(\mu) := \mathcal{L}^1(\Omega, \mathcal{A}, \mu) := \{f\colon \Omega \to \overline{\mathbb{R}}\colon f \text{ messbar und } \int |f| \,\mathrm{d}\mu < \infty\}.$$

Für $f \in \mathcal{L}^1(\mu)$ definieren wir das *Integral von f bezüglich μ* als

$$\int f \,\mathrm{d}\mu := \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu.$$

Ist lediglich $\int f^- \,\mathrm{d}\mu < \infty$, aber $\int f^+ \,\mathrm{d}\mu = \infty$, so setzen wir $\int f \,\mathrm{d}\mu := \infty$; ist umgekehrt $\int f^+ \,\mathrm{d}\mu < \infty$ und $\int f^- \,\mathrm{d}\mu = \infty$, so setzen wir $\int f \,\mathrm{d}\mu := -\infty$. Für $A \in \mathcal{A}$ schreiben wir

$$\int_A f \,\mathrm{d}\mu := \int f \,\mathbb{1}_A \,\mathrm{d}\mu.$$

Eine Eigenschaft gilt *μ-fast überall* (kurz *μ-f. ü.*; bei Wahrscheinlichkeitsmaßen auch *μ-fast sicher*, kurz *μ-f. s.*), wenn sie außerhalb einer μ-Nullmenge gilt.

Satz 7.12 (Punktweise Aussagen aus dem Integral)**.** *Sei $f\colon \Omega \to [0, \infty]$ messbar. Dann gilt:*

(a) $\int f \,\mathrm{d}\mu = 0 \iff f = 0$ *μ-f. ü.* $(d.\,h.\ \mu(\{f \neq 0\}) = 0)$.

(b) $\int f \,\mathrm{d}\mu < \infty \implies f < \infty$ *μ-f. ü.*

Beweis. (a) „⇐“: Sei $f = 0$ μ-f. ü., d. h. $N := \{f > 0\}$ erfüllt $\mu(N) = 0$. Dann gilt $f \le \infty \cdot \mathbb{1}_N$ und $n\mathbb{1}_N \uparrow \infty \cdot \mathbb{1}_N$ für $n \to \infty$. Mit Schritt 3 (2) und der monotonen Konvergenz aus Schritt 3 (3) folgt

$$0 \le \int f \,\mathrm{d}\mu \le \int \infty \cdot \mathbb{1}_N \,\mathrm{d}\mu = \lim_{n\to\infty} \int n\mathbb{1}_N \,\mathrm{d}\mu = \lim_{n\to\infty} n\,\mu(N) = 0.$$

„⇒“: Sei $N_n := \{f \ge 1/n\}$. Dann gilt $N_n \uparrow N = \{f > 0\}$ und

$$0 = \int f \,\mathrm{d}\mu \ge \int \tfrac{1}{n}\mathbb{1}_{N_n} \,\mathrm{d}\mu = \tfrac{1}{n}\,\mu(N_n).$$

Also ist $\mu(N_n) = 0$ für alle n, und mit der σ-Stetigkeit folgt $\mu(N) = \lim_n \mu(N_n) = 0$.

(b) Sei $A := \{f = \infty\}$. Für alle $n \in \mathbb{N}$ gilt $A \subset \{f \ge n\}$ und $(1/n) f \mathbb{1}_{\{f\ge n\}} \ge \mathbb{1}_{\{f\ge n\}}$, also

$$\mu(A) \le \mu(\{f \ge n\}) \le \int \tfrac{1}{n} f\, \mathbb{1}_{\{f\ge n\}} \,\mathrm{d}\mu \le \frac{1}{n} \int f \,\mathrm{d}\mu \xrightarrow[n\to\infty]{} 0. \qquad \square$$

Ist $f \in \mathcal{L}^1(\mu)$, so ist nach Satz 7.12(b) (angewandt auf $|f|$) f μ-f. ü. endlich. Für $f, g \in \mathcal{L}^1(\mu)$ sind daher Ausdrücke wie $f+g$ außerhalb einer Nullmenge wohldefiniert; auf dieser Nullmenge setzen wir sie beliebig (etwa $= 0$) fest, was nach Satz 7.13(a) die Integrale nicht beeinflusst.

Satz 7.13 (Eigenschaften des Integrals)**.** *Seien $f, g \in \mathcal{L}^1(\mu)$. Dann gilt:*

(a) Monotonie: *Ist $f \le g$ μ-f. ü., so gilt $\int f \,\mathrm{d}\mu \le \int g \,\mathrm{d}\mu$. Insbesondere folgt aus $f = g$ μ-f. ü., dass $\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$.*

(b) Dreiecksungleichung: $\left|\int f \,\mathrm{d}\mu\right| \le \int |f| \,\mathrm{d}\mu$.

(c) Linearität: *Für alle $\alpha, \beta \in \mathbb{R}$ ist $\alpha f + \beta g \in \mathcal{L}^1(\mu)$ mit*

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

Beweis. Vorab: Sind $u, v \colon \Omega \to [0, \infty]$ messbar mit $u \le v$ μ-f. ü., so gilt $\int u \,\mathrm{d}\mu \le \int v \,\mathrm{d}\mu$. Denn mit $N := \{u > v\}$ ist $\mu(N) = 0$, also $\int u\mathbb{1}_N \,\mathrm{d}\mu = 0$ nach Satz 7.12(a), und mit Schritt 3 (4) und (2) folgt

$$\int u \,\mathrm{d}\mu = \int u\mathbb{1}_{N^c} \,\mathrm{d}\mu + \int u\mathbb{1}_N \,\mathrm{d}\mu = \int u\mathbb{1}_{N^c} \,\mathrm{d}\mu \le \int v \,\mathrm{d}\mu.$$

(a) Aus $f \le g$ μ-f. ü. folgt $f^+ \le g^+$ und $f^- \ge g^-$ μ-f. ü., also nach der Vorbemerkung

$$\int f^+ \,\mathrm{d}\mu \le \int g^+ \,\mathrm{d}\mu \qquad \text{und} \qquad \int f^- \,\mathrm{d}\mu \ge \int g^- \,\mathrm{d}\mu.$$

Damit gilt

$$\int f \,\mathrm{d}\mu = \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu \le \int g^+ \,\mathrm{d}\mu - \int g^- \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu.$$

(b) Es gilt

$$\left|\int f \,\mathrm{d}\mu\right| = \left|\int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu\right| \le \int f^+ \,\mathrm{d}\mu + \int f^- \,\mathrm{d}\mu = \int |f| \,\mathrm{d}\mu,$$

wobei wir im letzten Schritt Schritt 3 (4) verwendet haben.

(c) Wegen $|\alpha f + \beta g| \le |\alpha|\,|f| + |\beta|\,|g|$ und Schritt 3 (2), (4) gilt $\alpha f + \beta g \in \mathcal{L}^1(\mu)$. Für die Linearität genügt es zu zeigen:

(i) $\int (f+g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$. Dies gilt, denn

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

also $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$ μ-f. ü., nämlich außerhalb der Nullmenge $\{|f| = \infty\} \cup \{|g| = \infty\}$ (alle Summanden nichtnegativ). Mit Schritt 3 (4) und der Vorbemerkung folgt

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

und da alle auftretenden Integrale endlich sind, ergibt Umstellen

$$\int (f+g)\,\mathrm{d}\mu = \int f^+\,\mathrm{d}\mu - \int f^-\,\mathrm{d}\mu + \int g^+\,\mathrm{d}\mu - \int g^-\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu.$$

(ii) Für $\alpha \geq 0$ gilt $(\alpha f)^\pm = \alpha f^\pm$ und daher

$$\int \alpha f\,\mathrm{d}\mu = \int \alpha f^+\,\mathrm{d}\mu - \int \alpha f^-\,\mathrm{d}\mu = \alpha \int f\,\mathrm{d}\mu.$$

(iii) Wegen $(-f)^+ = f^-$ und $(-f)^- = f^+$ gilt

$$\int (-f)\,\mathrm{d}\mu = \int (-f)^+\,\mathrm{d}\mu - \int (-f)^-\,\mathrm{d}\mu = \int f^-\,\mathrm{d}\mu - \int f^+\,\mathrm{d}\mu = -\int f\,\mathrm{d}\mu. \qquad \square$$

Bemerkung. Wie bereits früher bemerkt, können wir Bildmaße unter Zufallsvariablen (d. h. messbaren Abbildungen) $X\colon (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}')$ betrachten, also $\mu' := \mu \circ X^{-1}$. Ist nun $f \in \mathcal{L}^1(\mu')$, so gilt $f \circ X \in \mathcal{L}^1(\mu)$ und

$$\int f \circ X\,\mathrm{d}\mu = \int f\,\mathrm{d}(\mu \circ X^{-1}).$$

Dies zeigt man mit maßtheoretischer Induktion, d. h. indem man die Schritte 1–4 durchläuft (Übung).

Definition 7.14 (Dichten)**.** Sei μ ein Maß auf $(\Omega, \mathcal{A})$ und $f\colon \Omega \to [0, \infty)$ messbar. Wir sagen, dass das durch

$$\nu(A) := \int_A f\,\mathrm{d}\mu, \qquad A \in \mathcal{A},$$

definierte Maß $f\mu := \nu$ die *Dichte* f bezüglich μ hat.

Bemerkung.

(1) ν ist tatsächlich ein Maß, denn $\nu(\emptyset) = 0$, und für $A = \bigcup_{i\geq 1} A_i$ mit paarweise disjunkten $A_i \in \mathcal{A}$ gilt nach dem Satz von der monotonen Konvergenz (Schritt 3 (3)) und der Linearität (Schritt 3 (4))

$$\nu(A) = \int f\,\mathbb{1}_{\bigcup_{i\geq 1} A_i}\,\mathrm{d}\mu = \lim_{n\to\infty} \int f\,\mathbb{1}_{\bigcup_{i\leq n} A_i}\,\mathrm{d}\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \int f\,\mathbb{1}_{A_i}\,\mathrm{d}\mu = \sum_{i\geq 1} \nu(A_i).$$

(2) Dieser Dichtebegriff umfasst insbesondere die diskreten Dichten (Zähldichten) und die Lebesgue-Dichten aus dem ersten Teil der Vorlesung (vgl. Satz 1.16): Man wähle für μ das Zählmaß bzw. das Lebesgue-Maß.

(3) Es gilt $g \in \mathcal{L}^1(f\mu) \iff gf \in \mathcal{L}^1(\mu)$, und mit Hilfe der maßtheoretischen Induktion lässt sich zeigen, dass dann

$$\int g \,\mathrm{d}(f\mu) = \int gf \,\mathrm{d}\mu$$

gilt (Übung).

(4) Sind f_1 und f_2 Dichten von ν bezüglich μ und ist ν σ-endlich, so gilt $f_1 = f_2$ μ-f. ü. (Beweis zur Übung).

Satz 7.15 (Satz von der monotonen Konvergenz)**.** *Seien $f_1, f_2, \ldots \in \mathcal{L}^1(\mu)$ und $f\colon \Omega \to \overline{\mathbb{R}}$ messbar mit $f_n \uparrow f$ μ-f. ü. Dann gilt*

$$\lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu \quad (\leq \infty).$$

Beweis. Sei N eine μ-Nullmenge, sodass $f_n(\omega) \uparrow f(\omega)$ und $|f_1(\omega)| < \infty$ für alle $\omega \in N^{\mathrm{c}}$ gilt; eine solche existiert nach Voraussetzung und Satz 7.12(b). Dann sind

$$f_n' \coloneqq (f_n - f_1)\,\mathbb{1}_{N^{\mathrm{c}}} \geq 0 \qquad \text{und} \qquad f' \coloneqq (f - f_1)\,\mathbb{1}_{N^{\mathrm{c}}} \geq 0$$

messbar, und es gilt $f_n' \uparrow f'$ sogar punktweise. Nach Schritt 3 (3) folgt

$$\int f_n' \,\mathrm{d}\mu \xrightarrow[n\to\infty]{} \int f' \,\mathrm{d}\mu.$$

Da $f_n = f_n' + f_1$ und $f = f' + f_1$ μ-f. ü. gilt, folgt mit Satz 7.13

$$\int f_n \,\mathrm{d}\mu = \int f_1 \,\mathrm{d}\mu + \int f_n' \,\mathrm{d}\mu \xrightarrow[n\to\infty]{} \int f_1 \,\mathrm{d}\mu + \int f' \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

(Im Fall $\int f' \,\mathrm{d}\mu = \infty$ ist $f \notin \mathcal{L}^1(\mu)$; die letzte Gleichung folgt dann aus $f^- \leq f_1^-$ und $f^+ \geq f' - f_1^-$, sodass beide Seiten gleich ∞ sind.) □

Satz 7.16 (Lemma von Fatou)**.** *Sei $f \in \mathcal{L}^1(\mu)$, und seien $f_1, f_2, \ldots$ messbar mit $f_n \geq f$ μ-f. ü. für alle n. Dann gilt*

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

Beweis. Nach Abänderung auf einer μ-Nullmenge, die nach der Vorbemerkung im Beweis von Satz 7.13 keines der Integrale ändert, dürfen wir annehmen, dass f reellwertig ist und $f_n \geq f$ überall gilt. Für messbares $h\colon \Omega \to [0,\infty]$ gilt dann

$$\int (h + f) \,\mathrm{d}\mu = \int h \,\mathrm{d}\mu + \int f \,\mathrm{d}\mu;$$

denn aus $(h+f)^+ + f^- = (h+f)^- + h + f^+$ folgt mit Schritt 3 (4) die entsprechende Gleichung der Integrale, und wegen $(h+f)^- \leq f^-$ sind $\int (h+f)^- \,\mathrm{d}\mu$ und $\int f^{\pm} \,\mathrm{d}\mu$ endlich und dürfen abgezogen werden. Angewandt auf $h = f_n - f$ und $h = \liminf_{n\to\infty} f_n - f$ zeigt dies, dass wir f_n durch $f_n - f$ ersetzen und ohne Beschränkung der Allgemeinheit $f_n \geq 0$ annehmen können. Setze

$$g_n \coloneqq \inf_{m \geq n} f_m \leq f_n.$$

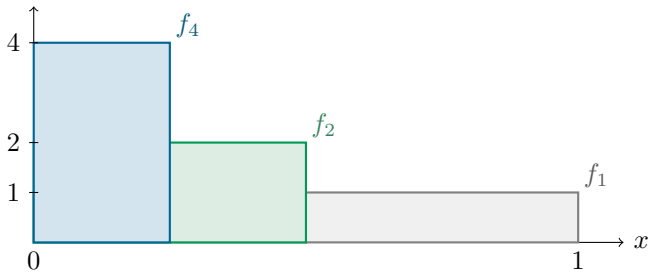
Dann gilt $0 \leq g_n \uparrow \liminf_{m\to\infty} f_m$ für $n \to \infty$, und mit Satz 7.15 (bzw. Schritt 3 (3)) und der Monotonie folgt

$$\int \liminf_{n\to\infty} f_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int g_n \,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu.$$ □

Beispiele.

(1) Die Ungleichung im Lemma von Fatou kann strikt sein: Betrachte auf $[0,1]$ mit dem Lebesgue-Maß die Funktionen $f_n := n\mathbb{1}_{[0,1/n]}$. Dann gilt $f_n \to 0$ Lebesgue-f. ü. (nämlich auf $(0,1]$), also

$$\int_0^1 \liminf_{n\to\infty} f_n(x)\,\mathrm{d}x = 0, \qquad \text{aber} \qquad \int_0^1 f_n(x)\,\mathrm{d}x = 1 \quad \text{für alle } n.$$



$f_n = n\mathbb{1}_{[0,1/n]}$ für $n = 1, 2, 4$: Jedes f_n hat Integral 1, doch die Masse konzentriert sich auf immer kleinere Intervalle und verschwindet im punktweisen Grenzwert.

(2) Wendet man das Lemma von Fatou auf $-f_n$ an, so erhält man: Sind $f_n \le f$ μ-f. ü. mit $f \in \mathcal{L}^1(\mu)$, so gilt

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

Satz 7.17 (Satz von Lebesgue bzw. Satz von der dominierten Konvergenz)**.** *Seien $(f_n)_{n\ge1}$ und f messbare Funktionen $\Omega \to \mathbb{R}$ mit $f_n(\omega) \to f(\omega)$ für alle $\omega \in \Omega$. Es gebe ein $g \in \mathcal{L}^1(\mu)$ mit $|f_n(\omega)| \le g(\omega)$ für alle n und alle $\omega \in \Omega$. Dann gilt $f \in \mathcal{L}^1(\mu)$ und*

$$\int |f_n - f|\,\mathrm{d}\mu \xrightarrow[n\to\infty]{} 0,$$

insbesondere also

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

Beweis von Satz 7.17. Aus $|f_n| \le g$ folgt im Grenzwert $|f| \le g$, also $f \in \mathcal{L}^1(\mu)$. Ferner gilt

$$0 \le |f_n - f| \le |f_n| + |f| \le 2g \qquad \text{und} \qquad 2g \in \mathcal{L}^1(\mu).$$

Nach dem Lemma von Fatou in der Form von Beispiel (2) im Anschluss an Satz 7.16 folgt

$$\limsup_{n\to\infty} \int |f_n - f|\,\mathrm{d}\mu \le \int \limsup_{n\to\infty} |f_n - f|\,\mathrm{d}\mu = 0,$$

also $\int |f_n - f|\,\mathrm{d}\mu \to 0$. Da nach Satz 7.13

$$\left|\int f_n\,\mathrm{d}\mu - \int f\,\mathrm{d}\mu\right| \le \int |f_n - f|\,\mathrm{d}\mu$$

gilt, folgt auch die zweite Aussage. □

Wendet man Satz 7.17 auf $f_n\mathbb{1}_{N^c}$ und $f\mathbb{1}_{N^c}$ für eine geeignete Nullmenge N an, so sieht man, dass es genügt, die Konvergenz $f_n \to f$ und die Abschätzung $|f_n| \le g$ nur μ-f. ü. zu fordern.

Eine wichtige Anwendung ist die Vertauschung von Differentiation und Integration.

Satz 7.18 (Vertauschung von Differentiation und Integration)**.** *Sei $I \subset \mathbb{R}$ offen, $I \neq \emptyset$, und $f\colon \Omega \times I \to \mathbb{R}$ eine Funktion mit*

(1) für alle $x \in I$ ist $\omega \mapsto f(\omega, x)$ in $\mathcal{L}^1(\mu)$;

(2) für μ-fast alle ω ist $x \mapsto f(\omega, x)$ differenzierbar auf I mit Ableitung $f'(\omega, x)$;

(3) $h \coloneqq \sup_{x \in I} |f'(\cdot, x)| \in \mathcal{L}^1(\mu)$.

Dann ist $f'(\cdot, x) \in \mathcal{L}^1(\mu)$ für alle $x \in I$, und die Funktion

$$F\colon I \to \mathbb{R}, \qquad F(x) \coloneqq \int f(\omega, x)\, \mu(\mathrm{d}\omega),$$

ist differenzierbar mit

$$F'(x) = \int f'(\omega, x)\, \mu(\mathrm{d}\omega).$$

Dabei sei $f'(\omega, x) \coloneqq 0$ für die ω aus der Ausnahmenullmenge N in (2) gesetzt, und in (3) wird die Messbarkeit von h mit vorausgesetzt.

Beweis. Sei $x_0 \in I$ und $(x_n)_{n \geq 1} \subset I$ eine Folge mit $x_n \to x_0$ und $x_n \neq x_0$ für alle n. Setze

$$g_n(\omega) \coloneqq \frac{f(\omega, x_n) - f(\omega, x_0)}{x_n - x_0}, \qquad \omega \in \Omega.$$

Nach (1) ist $g_n \in \mathcal{L}^1(\mu)$, und nach (2) gilt $g_n \to f'(\cdot, x_0)$ μ-f. ü.; insbesondere ist $f'(\cdot, x_0)$ messbar. Da I offen ist, liegt für alle hinreichend großen n die Verbindungsstrecke von x_0 und x_n in I; ohne Einschränkung gelte dies für alle n. Nach dem Mittelwertsatz gibt es dann für alle n und jedes $\omega \in N^{\mathrm{c}}$ ein $y_n(\omega) \in I$ zwischen x_0 und x_n mit

$$g_n(\omega) = f'(\omega, y_n(\omega)),$$

und insbesondere gilt $|g_n| \leq h$ μ-f. ü. für alle $n \in \mathbb{N}$. Nach dem Satz von der majorisierten Konvergenz (Satz 7.17 und die anschließende Bemerkung) folgt $f'(\cdot, x_0) \in \mathcal{L}^1(\mu)$ und

$$\lim_{n \to \infty} \frac{F(x_n) - F(x_0)}{x_n - x_0} = \lim_{n \to \infty} \int g_n(\omega)\, \mu(\mathrm{d}\omega) = \int f'(\omega, x_0)\, \mu(\mathrm{d}\omega).$$

Da die Folge (x_n) beliebig war, ist F in x_0 differenzierbar mit der behaupteten Ableitung. □

Als Nächstes vergleichen wir das Lebesgue-Integral mit dem Riemann-Integral. Dabei bezeichne λ das Lebesgue-Maß, fortgesetzt auf die Lebesgue'sche σ-Algebra (vgl. Bemerkung 7.10).

Satz 7.19 (Riemann- vs. Lebesgue-Integration)**.** *Sei $I = [a, b]$ und $f\colon I \to \mathbb{R}$ Riemann-integrierbar. Dann ist f messbar bezüglich der Lebesgue'schen σ-Algebra $\mathcal{B}_I^*$ (vgl. Bemerkung 7.10) und Lebesgue-integrierbar, und es gilt*

$$\int_I f\, \mathrm{d}\lambda = \int_a^b f(x)\, \mathrm{d}x.$$

Beweis. Wir betrachten eine Folge sich verfeinernder Zerlegungen $t^n = (t_i^n)_{i=0,\dots,k_n}$ von I mit

$$a = t_0^n < t_1^n < \cdots < t_{k_n}^n = b \qquad \text{und} \qquad \max\{t_i^n - t_{i-1}^n \colon i = 1, \dots, k_n\} \xrightarrow[n\to\infty]{} 0,$$

sowie die Unter- und Obersummen

$$U_n(f) \coloneqq \sum_{i=1}^{k_n} (t_i^n - t_{i-1}^n) \inf f([t_{i-1}^n, t_i^n)),$$
$$O_n(f) \coloneqq \sum_{i=1}^{k_n} (t_i^n - t_{i-1}^n) \sup f([t_{i-1}^n, t_i^n)).$$

Ist f Riemann-integrierbar, so ist f beschränkt, d. h. $\|f\|_\infty \coloneqq \sup_{x\in I} |f(x)| < \infty$, und es gilt

$$\lim_{n\to\infty} U_n(f) = \lim_{n\to\infty} O_n(f) = \int_a^b f(x)\,\mathrm{d}x$$

(die Summen über halboffene Teilintervalle liegen zwischen den üblichen Darboux'schen Unter- und Obersummen, die beide gegen das Riemann-Integral konvergieren). Ersetzen wir f durch $f + \|f\|_\infty$, so können wir ohne Einschränkung $f \geq 0$ annehmen. Setze

$$g_n \coloneqq f(b)\,\mathbb{1}_{\{b\}} + \sum_{i=1}^{k_n} \mathbb{1}_{[t_{i-1}^n, t_i^n)} \cdot \inf f([t_{i-1}^n, t_i^n)),$$
$$h_n \coloneqq f(b)\,\mathbb{1}_{\{b\}} + \sum_{i=1}^{k_n} \mathbb{1}_{[t_{i-1}^n, t_i^n)} \cdot \sup f([t_{i-1}^n, t_i^n)).$$

Das sind einfache Borel-messbare Funktionen mit $\int_I g_n\,\mathrm{d}\lambda = U_n(f)$ und $\int_I h_n\,\mathrm{d}\lambda = O_n(f)$. Da die Zerlegungen sich verfeinern, gilt

$$g_n \leq g_{n+1} \leq f \leq h_{n+1} \leq h_n,$$

und es gibt Borel-messbare Funktionen g, h mit $g_n \uparrow g$, $h_n \downarrow h$ sowie $g \leq f \leq h$. Mit monotoner Konvergenz (Satz 7.15) bzw. majorisierter Konvergenz (Satz 7.17, mit der integrierbaren Majorante h_1) folgt

$$\int_I g\,\mathrm{d}\lambda = \lim_{n\to\infty} \int_I g_n\,\mathrm{d}\lambda = \lim_{n\to\infty} U_n(f) = \lim_{n\to\infty} O_n(f) = \lim_{n\to\infty} \int_I h_n\,\mathrm{d}\lambda = \int_I h\,\mathrm{d}\lambda.$$

Also ist $\int_I (h-g)\,\mathrm{d}\lambda = 0$ mit $h - g \geq 0$, und Satz 7.12(a) liefert $h = g$ λ-f. ü. Wegen $g \leq f \leq h$ gilt für alle $\alpha \in \mathbb{R}$

$$\{f \leq \alpha\} = \underbrace{(\{g \leq \alpha\} \cap \{g = h\})}_{\text{Borel-Menge}} \cup \underbrace{(\{f \leq \alpha\} \cap \{g \neq h\})}_{\text{Teilmenge einer Lebesgue-Nullmenge}} \in \mathcal{B}_I^*.$$

Also ist f messbar bezüglich $\mathcal{B}_I^*$, es gilt $f = g$ λ-f. ü., und mit Satz 7.13(a) folgt

$$\int_I f\,\mathrm{d}\lambda = \int_I g\,\mathrm{d}\lambda = \lim_{n\to\infty} U_n(f) = \int_a^b f(x)\,\mathrm{d}x. \qquad \square$$

Beispiele.

(1) Die Umkehrung gilt nicht: Die Funktion $f \colon [0,1] \to \mathbb{R}$, $x \mapsto \mathbb{1}_{\mathbb{Q}}(x)$, ist nicht Riemann-integrierbar, da $O_n(f) = 1$ und $U_n(f) = 0$ für alle n. Sie ist aber λ-integrierbar mit $\int f\,\mathrm{d}\lambda = 0$, da $\mathbb{Q} \cap [0,1]$ als abzählbare Menge eine λ-Nullmenge ist.

(2) Ist $I = (a, b]$ oder $I = [0, \infty)$, so gibt es uneigentlich Riemann-integrierbare Funktionen, die nicht Lebesgue-integrierbar sind. Siehe etwa

$$f\colon [0,\infty) \to \mathbb{R}, \qquad x \mapsto \frac{\sin x}{1+x}.$$

Das uneigentliche Riemann-Integral $\int_0^\infty f(x)\,\mathrm{d}x$ existiert, aber es gilt $\int_0^\infty |f(x)|\,\mathrm{d}x = \infty$.

Für zwei Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i \in \{1,2\}$, bezeichne $\mu_1 \otimes \mu_2$ das *Produktmaß* auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, d. h. das Maß mit

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\,\mu_2(A_2) \qquad \text{für alle } A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2.$$

Sind μ_1, μ_2 σ-endlich, so existiert es (z. B. mit Satz 7.4) und ist eindeutig bestimmt (ohne Beweis).

Satz 7.20 (Satz von Fubini)**.** *Seien $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i \in \{1,2\}$, σ-endliche Maßräume und $f\colon \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ messbar bezüglich $\mathcal{A}_1 \otimes \mathcal{A}_2$. Ist $f \geq 0$ oder $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, so gilt:*

$$\omega_1 \longmapsto \int f(\omega_1,\omega_2)\,\mu_2(\mathrm{d}\omega_2) \text{ ist } \mathcal{A}_1\text{-messbar}, \qquad \omega_2 \longmapsto \int f(\omega_1,\omega_2)\,\mu_1(\mathrm{d}\omega_1) \text{ ist } \mathcal{A}_2\text{-messbar}, \tag{7.2}$$

und

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f\,\mathrm{d}(\mu_1\otimes\mu_2) &= \int_{\Omega_1}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)\,\mu_2(\mathrm{d}\omega_2)\Big)\mu_1(\mathrm{d}\omega_1)\\ &= \int_{\Omega_2}\Big(\int_{\Omega_1} f(\omega_1,\omega_2)\,\mu_1(\mathrm{d}\omega_1)\Big)\mu_2(\mathrm{d}\omega_2).\end{aligned} \tag{7.3}$$

Im Fall $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$ sind die inneren Integrale dabei nur μ_1- bzw. μ_2-f. ü. definiert (und endlich); außerhalb dieser Nullmengen setze man sie $= 0$.

Beweis. Wir benutzen maßtheoretische Induktion und betrachten der Symmetrie wegen nur jeweils die erste Aussage in (7.2) und die erste Gleichung in (7.3).

Schritt 1: Indikatorfunktionen von Rechtecken. Für $f = \mathbb{1}_A$ mit $A = A_1 \times A_2$, $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, gilt

$$\int \mathbb{1}_A(\omega_1,\omega_2)\,\mu_2(\mathrm{d}\omega_2) = \mathbb{1}_{A_1}(\omega_1)\,\mu_2(A_2),$$

und dies ist $\mathcal{A}_1$-messbar mit Integral $\mu_1(A_1)\,\mu_2(A_2) = \mu_1 \otimes \mu_2(A)$. Also gilt die Aussage für solche f.

Schritt 2: Indikatorfunktionen beliebiger Mengen $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Seien zunächst μ_1, μ_2 endlich, und sei $\mathcal{D}$ das System aller $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, für die die Aussage mit $f = \mathbb{1}_A$ gilt. Nach Schritt 1 enthält $\mathcal{D}$ das $\cap$-stabile System der Rechtecke $A_1 \times A_2$, das $\mathcal{A}_1 \otimes \mathcal{A}_2$ erzeugt. Ferner ist $\mathcal{D}$ ein Dynkin-System: $\Omega_1 \times \Omega_2 \in \mathcal{D}$ nach Schritt 1; für $A \subset B$ aus $\mathcal{D}$ ist $\mathbb{1}_{B\setminus A} = \mathbb{1}_B - \mathbb{1}_A$, und da alle auftretenden Integrale endlich sind, folgt $B \setminus A \in \mathcal{D}$ aus der Linearität; für paarweise disjunkte $A^{(1)}, A^{(2)}, \ldots \in \mathcal{D}$ folgt $\bigcup_k A^{(k)} \in \mathcal{D}$ aus $\mathbb{1}_{\bigcup_k A^{(k)}} = \sum_k \mathbb{1}_{A^{(k)}}$ mit dem Satz von der monotonen Konvergenz (angewandt auf die inneren und die äußeren Integrale; Grenzwerte messbarer Funktionen sind messbar). Mit Lemma 1.13 folgt $\mathcal{D} = \mathcal{A}_1 \otimes \mathcal{A}_2$. Im σ-endlichen Fall wähle man $\Omega_i^{(k)} \uparrow \Omega_i$ mit $\mu_i(\Omega_i^{(k)}) < \infty$, wende das eben Gezeigte auf die endlichen Maße $\mu_i(\,\cdot\, \cap \Omega_i^{(k)})$ an und gehe mit monotoner Konvergenz zum Grenzwert $k \to \infty$ über. Durch Bilden endlicher Summen (Linearität) gilt die Aussage dann für alle einfachen Funktionen $f \geq 0$.

Schritt 3: allgemeine f. Ist $f \geq 0$, so gibt es einfache Funktionen $f_n \geq 0$ mit $f_n \uparrow f$ (vgl. Schritt 3 (4) der Konstruktion des Integrals), und die Aussage folgt aus Schritt 2 mit dem Satz von der monotonen Konvergenz, angewandt auf die inneren und die äußeren Integrale. Ist $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, so schreiben wir $f = f^+ - f^-$ mit $f^+, f^- \geq 0$ und wenden das Bisherige auf f^+ und f^- an. Wegen $\int f^\pm \,\mathrm{d}(\mu_1 \otimes \mu_2) < \infty$ sind die inneren Integrale von $f^\pm$ nach Satz 7.12(b) μ_1-f. ü. endlich, und die Behauptung folgt durch Differenzbildung mit Satz 7.13. □

7.3 L^p-Räume und Satz von Radon-Nikodym

Sei $(\Omega, \mathcal{A}, \mu)$ ein σ-endlicher Maßraum. Für messbares $f\colon \Omega \to \overline{\mathbb{R}}$ definieren wir

$$\|f\|_p := \Big(\int |f|^p \,\mathrm{d}\mu\Big)^{1/p}, \qquad p \in [1, \infty),$$

sowie

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

(mit $\inf \emptyset = \infty$) und setzen für $p \in [1, \infty]$

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

Für $f \in \mathcal{L}^\infty(\mu)$ gilt $|f| \leq \|f\|_\infty$ μ-f. ü., denn $\{|f| > \|f\|_\infty\} = \bigcup_{k \geq 1} \{|f| > \|f\|_\infty + 1/k\}$ ist eine abzählbare Vereinigung von Nullmengen.

Satz 7.21 (Hölder'sche Ungleichung)**.** *Seien $p, q \in [1, \infty]$ mit $1/p + 1/q = 1$ (wobei $1/\infty := 0$), und seien $f \in \mathcal{L}^p(\mu)$, $g \in \mathcal{L}^q(\mu)$. Dann ist $fg \in \mathcal{L}^1(\mu)$, und es gilt*

$$\|fg\|_1 \leq \|f\|_p \, \|g\|_q.$$

Beweis. Die Fälle $p = 1$, $q = \infty$ (und symmetrisch $p = \infty$, $q = 1$) folgen direkt, denn $|fg| \leq |f| \, \|g\|_\infty$ μ-f. ü. Sei also $p \in (1, \infty)$, also auch $q = p/(p-1) \in (1, \infty)$. Ist $\|f\|_p = 0$ oder $\|g\|_q = 0$, so gilt nach Satz 7.12(a) $f = 0$ bzw. $g = 0$ μ-f. ü., und beide Seiten verschwinden. Andernfalls betrachten wir

$$f' := \frac{f}{\|f\|_p} \qquad \text{und} \qquad g' := \frac{g}{\|g\|_q},$$

sodass $\|f'\|_p = \|g'\|_q = 1$.

Schritt 1: Young'sche Ungleichung. Wir zeigen: Für alle $x, y \geq 0$ gilt

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Für $y = 0$ ist die linke Seite 0. Sei $y > 0$ fest und $\varphi(x) := x^p/p + y^q/q - xy$ für $x \geq 0$. Dann ist

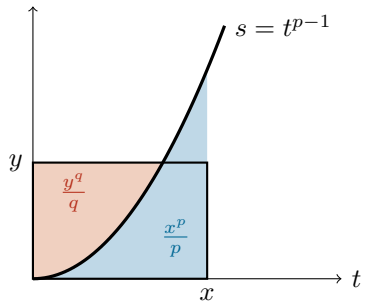
$$\varphi'(x) = x^{p-1} - y, \qquad \varphi''(x) = (p-1)\, x^{p-2} > 0 \quad (x > 0),$$

also ist φ strikt konvex mit Minimalstelle $x_0 = y^{1/(p-1)}$. Wegen $p/(p-1) = q$ gilt $x_0^p = y^q$ und $x_0 y = y^{1/(p-1)+1} = y^q$, also

$$\varphi(x_0) = \Big(\frac{1}{p} + \frac{1}{q}\Big) y^q - y^q = 0.$$

Somit ist $\varphi(x) \geq 0$ für alle $x \geq 0$.

Anschaulich ist die Young'sche Ungleichung eine Flächenaussage:



Schematisch, für $p = 3$: Das Rechteck $[0, x] \times [0, y]$ wird von den Flächen $\int_0^x t^{p-1}\,\mathrm{d}t = x^p/p$ (unter der Kurve $s = t^{p-1}$) und $\int_0^y s^{q-1}\,\mathrm{d}s = y^q/q$ (links der Kurve) überdeckt. Gleichheit gilt genau für $y = x^{p-1}$, also $y^q = x^p$.

Schritt 2. Mit der Young'schen Ungleichung folgt

$$\|f'g'\|_1 = \int |f'|\,|g'|\,\mathrm{d}\mu \le \frac{1}{p}\int |f'|^p\,\mathrm{d}\mu + \frac{1}{q}\int |g'|^q\,\mathrm{d}\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

Multiplikation mit $\|f\|_p\,\|g\|_q$ liefert die Behauptung. □

Bemerkung. Für $p = q = 2$ ist dies die *Cauchy-Schwarz-Ungleichung* $\|fg\|_1 \le \|f\|_2\,\|g\|_2$.

Satz 7.22 (Minkowski-Ungleichung)**.** *Für $p \in [1, \infty]$ und $f, g \in \mathcal{L}^p(\mu)$ gilt*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p.$$

Beweis. Für $p = \infty$ folgt die Aussage, da $|f + g| \le |f| + |g| \le \|f\|_\infty + \|g\|_\infty$ μ-f. ü., und für $p = 1$ folgt sie durch Integration von $|f + g| \le |f| + |g|$. Sei also $p \in (1, \infty)$ und $q := p/(p-1)$, d. h. $1/p + 1/q = 1$. Wegen $|f + g| \le |f| + |g|$ dürfen wir ohne Beschränkung der Allgemeinheit $f, g \ge 0$ annehmen, und ferner $\|f + g\|_p > 0$ (sonst ist nichts zu zeigen). Nun ist

$$(f + g)^p \le \big(2\max\{f, g\}\big)^p \le 2^p\,(f^p + g^p),$$

also $f + g \in \mathcal{L}^p(\mu)$. Wegen $(p-1)q = p$ ist $(f+g)^{p-1} \in \mathcal{L}^q(\mu)$ mit $\|(f+g)^{p-1}\|_q = \|f+g\|_p^{p/q} = \|f + g\|_p^{p-1}$. Die Hölder'sche Ungleichung, angewandt auf $f \cdot (f + g)^{p-1}$ und $g \cdot (f + g)^{p-1}$, liefert

$$\begin{aligned}\|f + g\|_p^p &= \int (f + g)^p\,\mathrm{d}\mu = \int f\,(f + g)^{p-1}\,\mathrm{d}\mu + \int g\,(f + g)^{p-1}\,\mathrm{d}\mu \\ &\le \|f\|_p\,\|(f + g)^{p-1}\|_q + \|g\|_p\,\|(f + g)^{p-1}\|_q = \big(\|f\|_p + \|g\|_p\big)\,\|f + g\|_p^{p-1}.\end{aligned}$$

Division durch $\|f + g\|_p^{p-1} \in (0, \infty)$ ergibt die Behauptung. □

Bemerkung.

(1) Die Minkowski-Ungleichung lässt sich als Dreiecksungleichung lesen: $\|\cdot\|_p$ ist eine Halbnorm auf dem Vektorraum $\mathcal{L}^p(\mu)$. Eine Norm ist $\|\cdot\|_p$ im Allgemeinen nicht, denn aus $\|f\|_p = 0$ folgt nur $f = 0$ μ-fast überall. (Genau genommen ist $\mathcal{L}^p(\mu)$ wegen möglicher Werte $\pm\infty$ kein Vektorraum; da diese nach Satz 7.12(b), angewandt auf $|f|^p$ (für $p = \infty$ wegen $|f| \le \|f\|_\infty$ μ-f. ü.), nur auf einer μ-Nullmenge auftreten, kann man sie dort durch 0 ersetzen und reellwertige Funktionen betrachten. Beim Übergang zu $L^p(\mu)$ in (2) spielt dies keine Rolle.)

(2) Diesen Mangel behebt man durch Übergang zum Quotientenraum: Mit

$$\mathcal{N} \coloneqq \{f\colon \Omega \to \overline{\mathbb{R}} \text{ messbar}\colon f = 0 \ \mu\text{-f. ü.}\}$$

setzt man

$$L^p(\Omega, \mathcal{A}, \mu) \coloneqq \mathcal{L}^p(\Omega, \mathcal{A}, \mu)/\mathcal{N} = \{\bar{f} \coloneqq f + \mathcal{N}\colon f \in \mathcal{L}^p(\mu)\}.$$

Dann ist $(L^p(\mu), \|\cdot\|_p)$ mit $\|\bar{f}\|_p \coloneqq \|f\|_p$ ein *Banachraum*, also ein vollständiger normierter Vektorraum (Satz von Fischer–Riesz, ohne Beweis).

(3) $L^2(\mu)$ ist sogar ein (reeller) *Hilbertraum* mit dem Skalarprodukt

$$\langle f, g\rangle \coloneqq \int fg \,\mathrm{d}\mu,$$

denn dieses induziert die Norm, $\|f\|_2 = \langle f, f\rangle^{1/2}$, und nach (2) ist $L^2(\mu)$ vollständig. Nach der Hölder-Ungleichung (mit $p = q = 2$) ist $fg \in \mathcal{L}^1(\mu)$, das Skalarprodukt also wohldefiniert.

Ist μ endlich, so sind die Räume $\mathcal{L}^p(\mu)$ ineinander enthalten:

Satz 7.23 (Kanonische Inklusion)**.** *Sei $\mu(\Omega) < \infty$ und $1 \le p' \le p \le \infty$. Dann gilt $\mathcal{L}^p(\mu) \subset \mathcal{L}^{p'}(\mu)$, und die kanonische Inklusion*

$$i\colon \mathcal{L}^p(\mu) \hookrightarrow \mathcal{L}^{p'}(\mu), \qquad f \mapsto f,$$

ist stetig (bezüglich $\|\cdot\|_p$ und $\|\cdot\|_{p'}$).

Beweis. Für $p' = p$ ist nichts zu zeigen; sei also $p' < p$.

Fall $p = \infty$. Sei $f \in \mathcal{L}^\infty(\mu)$, also $\|f\|_\infty < \infty$, und $p' \in [1, \infty)$. Dann gilt $|f|^{p'} \le \|f\|_\infty^{p'}$ μ-fast überall und daher

$$\int |f|^{p'} \,\mathrm{d}\mu \le \int \|f\|_\infty^{p'} \,\mathrm{d}\mu = \|f\|_\infty^{p'} \,\mu(\Omega) < \infty.$$

Wendet man dies auf $f - g$ an, so folgt für $f, g \in \mathcal{L}^\infty(\mu)$

$$\|f - g\|_{p'} \le \mu(\Omega)^{1/p'} \,\|f - g\|_\infty;$$

also ist i (sogar Lipschitz-)stetig.

Fall $p < \infty$. Seien $p' < p < \infty$ und $f \in \mathcal{L}^p(\mu)$. Da $|f|^{p'} \le 1$ auf $\{|f| \le 1\}$ und $|f|^{p'} \le |f|^p$ auf $\{|f| > 1\}$, gilt $|f|^{p'} \le 1 + |f|^p$ und damit

$$\int |f|^{p'} \,\mathrm{d}\mu \le \mu(\Omega) + \int |f|^p \,\mathrm{d}\mu < \infty.$$

Zur Stetigkeit seien $f, g \in \mathcal{L}^p(\mu)$ und $c > 0$. Wegen $p' - p < 0$ ist $|f - g|^{p'-p} \le c^{p'-p}$ auf $\{|f - g| > c\}$, also

$$|f - g|^{p'} = |f - g|^{p'} \mathbb{1}\{|f - g| \le c\} + |f - g|^{p'} \mathbb{1}\{|f - g| > c\} \le c^{p'} + c^{p'-p} \,|f - g|^p.$$

Integration liefert

$$\|f - g\|_{p'} \le \left(c^{p'} \mu(\Omega) + c^{p'-p} \,\|f - g\|_p^p\right)^{1/p'}.$$

Ist $\|f - g\|_p = 0$, so folgt mit $c \downarrow 0$ auch $\|f - g\|_{p'} = 0$. Andernfalls wählen wir $c = \|f - g\|_p$ und erhalten

$$\|f - g\|_{p'} \le \left(c^{p'} \mu(\Omega) + c^{p'}\right)^{1/p'} = \left(1 + \mu(\Omega)\right)^{1/p'} \,\|f - g\|_p.$$

Also ist i stetig. □

Als Nächstes vergleichen wir zwei Maße auf demselben Ereignisraum miteinander.

Definition 7.24 (Absolutstetigkeit und singuläre Maße)**.** Seien ν, μ Maße auf $(\Omega, \mathcal{A})$.

(1) ν heißt *absolutstetig* bezüglich μ (kurz: $\nu \ll \mu$), falls für alle $A \in \mathcal{A}$ gilt:

$$\mu(A) = 0 \implies \nu(A) = 0.$$

μ und ν heißen *äquivalent* (kurz: $\mu \approx \nu$), falls $\mu \ll \nu$ und $\nu \ll \mu$.

(2) μ heißt *singulär* zu ν (kurz: $\mu \perp \nu$), falls es ein $A \in \mathcal{A}$ gibt mit

$$\mu(A) = 0 \qquad \text{und} \qquad \nu(\Omega \setminus A) = 0.$$

Diese Relation ist symmetrisch.

Beispiel 7.25. (1) Sei μ ein Maß auf $(\mathbb{R}, \mathcal{B})$ mit Lebesgue-Dichte f, d. h. $\mu(A) = \int_A f \,\mathrm{d}\lambda$ für $A \in \mathcal{B}$. Ist $\lambda(A) = 0$, so folgt $\mu(A) = \int_A f \,\mathrm{d}\lambda = 0$; also gilt $\mu \ll \lambda$.

Ist zusätzlich $f > 0$ λ-fast überall, so gilt für jedes $A \in \mathcal{B}$ mit $\lambda(A) > 0$ auch $\mu(A) = \int_A f \,\mathrm{d}\lambda > 0$; also ist dann $\lambda \ll \mu$ und somit $\mu \approx \lambda$.

Ist dagegen $\lambda(\{f = 0\}) > 0$, so gilt *nicht* $\lambda \ll \mu$, denn $\mu(\{f = 0\}) = \int_{\{f=0\}} f \,\mathrm{d}\lambda = 0$.

(2) Wir betrachten die Bernoulli-Verteilungen Ber_p, Ber_q auf $\{0, 1\}$ mit $p, q \in [0, 1]$. Ist $p \in (0, 1)$, so hat nur $\emptyset$ das Ber_p-Maß 0, und es gilt $\mathrm{Ber}_q \ll \mathrm{Ber}_p$. Ist $p \in \{0, 1\}$, so ist Ber_p ein Dirac-Maß, und es gilt

$$\mathrm{Ber}_q \ll \mathrm{Ber}_p \iff p = q \qquad \text{sowie} \qquad \mathrm{Ber}_q \perp \mathrm{Ber}_p \iff q = 1 - p.$$

(3) Für die Poisson-Verteilungen Pois_α, Pois_β mit $\alpha, \beta \geq 0$ (wobei $\mathrm{Pois}_0 = \delta_0$) gilt

$$\mathrm{Pois}_\alpha \ll \mathrm{Pois}_\beta \iff \beta > 0 \text{ oder } \alpha = 0.$$

(4) Wir betrachten einen unendlichen Münzwurf, einmal mit u. i. v. $X_i \sim \mathrm{Ber}_p$, $i \geq 1$, und einmal mit u. i. v. $Y_i \sim \mathrm{Ber}_q$, $i \geq 1$. Dann gilt für $p \neq q$

$$\mathrm{Ber}_p^{\otimes \mathbb{N}} \perp \mathrm{Ber}_q^{\otimes \mathbb{N}}$$

auf $(\{0, 1\}^{\mathbb{N}}, \mathcal{P}(\{0, 1\})^{\otimes \mathbb{N}})$. Denn nach dem starken Gesetz der großen Zahlen gilt für

$$A_p := \Big\{\omega \in \{0, 1\}^{\mathbb{N}} \colon \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \omega_i = p\Big\}$$

sowohl $\mathrm{Ber}_p^{\otimes \mathbb{N}}(A_p) = 1$ als auch $\mathrm{Ber}_q^{\otimes \mathbb{N}}(A_q) = 1$. Wegen $A_p \cap A_q = \emptyset$ folgt $\mathrm{Ber}_q^{\otimes \mathbb{N}}(A_p) = 0$.

Der folgende Satz zeigt, dass sich jedes (σ-endliche) Maß in einen absolutstetigen und einen singulären Anteil zerlegen lässt.

Satz 7.26 (Zerlegungssatz von Lebesgue)**.** *Seien μ, ν σ-endliche Maße auf $(\Omega, \mathcal{A})$. Dann lässt sich ν eindeutig bezüglich μ zerlegen in einen absolutstetigen Teil $\nu_a \ll \mu$ und einen singulären Teil $\nu_s \perp \mu$, d. h.*

$$\nu = \nu_a + \nu_s.$$

Dabei hat ν_a eine Dichte $\mathrm{d}\nu_a/\mathrm{d}\mu$ bezüglich μ, die $\mathcal{A}$-messbar und μ-fast überall endlich ist, d. h. $\nu_a(A) = \int_A (\mathrm{d}\nu_a/\mathrm{d}\mu) \,\mathrm{d}\mu$ für alle $A \in \mathcal{A}$.

Den Beweis führen wir nach dem folgenden Korollar.

Korollar 7.27 (Satz von Radon–Nikodym)**.** *Seien μ, ν σ-endliche Maße auf $(\Omega, \mathcal{A})$. Dann gilt*

$$\nu \text{ hat eine Dichte bezüglich } \mu \iff \nu \ll \mu.$$

In diesem Fall ist die Dichte $\mathrm{d}\nu/\mathrm{d}\mu$ $\mathcal{A}$-messbar und μ-fast überall endlich. Sie heißt Radon–Nikodym-Ableitung *von ν nach μ.*

Beweis. „⇒“: Ist $\nu(A) = \int_A f \,\mathrm{d}\mu$ für alle $A \in \mathcal{A}$ und $\mu(A) = 0$, so ist $f\mathbb{1}_A = 0$ μ-fast überall, also $\nu(A) = 0$.

„⇐“: Sei $\nu = \nu_a + \nu_s$ die Zerlegung aus Satz 7.26, und sei $E \in \mathcal{A}$ mit $\mu(E) = 0$ und $\nu_s(\Omega \setminus E) = 0$. Wegen $\nu \ll \mu$ gilt $\nu(E) = 0$, also $\nu_s(E) \le \nu(E) = 0$ und damit $\nu_s = 0$. Folglich ist $\nu = \nu_a$, und ν hat die Dichte $\mathrm{d}\nu_a/\mathrm{d}\mu$ bezüglich μ. □

Beweis von Satz 7.26. Wir beschränken uns auf endliche Maße μ, ν; der allgemeine Fall folgt durch Ausschöpfung von Ω mit Mengen endlichen Maßes. Die Eindeutigkeit der Zerlegung beweisen wir nicht.

Schritt 1: Ein stetiges lineares Funktional. Nach Satz 7.23 ist die Inklusion

$$i \colon \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu) \hookrightarrow \mathcal{L}^1(\Omega, \mathcal{A}, \mu + \nu)$$

stetig. Da $\nu \le \mu + \nu$, gilt $\left|\int h \,\mathrm{d}\nu\right| \le \int |h| \,\mathrm{d}(\mu + \nu)$; daher ist auch die Linearform

$$h \mapsto \int h \,\mathrm{d}\nu, \qquad \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu) \to \mathbb{R},$$

stetig. Wir nutzen nun eine zentrale Aussage der Funktionalanalysis (der Erweiterung der linearen Algebra auf unendlichdimensionale Vektorräume), den

Satz von Fréchet–Riesz. Sei $(V, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und $F \colon V \to \mathbb{R}$. Dann sind äquivalent:

(i) F ist stetig und linear;

(ii) es gibt ein $f \in V$ mit $F(x) = \langle x, f \rangle$ für alle $x \in V$.

Das Element f ist eindeutig bestimmt.

Beweis. „(ii)⇒(i)“ gilt, da $x \mapsto \langle x, f \rangle$ linear und (nach der Cauchy–Schwarz-Ungleichung) stetig ist. „(i)⇒(ii)“: Ist $F \equiv 0$, so setze $f = 0$. Andernfalls ist $W := F^{-1}(\{0\})$ ein abgeschlossener linearer Unterraum von V mit $W \neq V$. Sei $W^\perp := \{v \in V \colon \langle v, w \rangle = 0 \text{ für alle } w \in W\}$. Wähle $v \in V \setminus W$ und zerlege $v = y + z$ mit $y \in W$, $z \in W^\perp$ (orthogonale Projektion auf den abgeschlossenen Unterraum W; Projektionssatz, ohne Beweis). Dann ist $z \neq 0$ und $F(z) = F(v) - F(y) = F(v) \neq 0$, also ist $u := z/F(z) \in W^\perp$ wohldefiniert mit $F(u) = 1$. Für alle $x \in V$ gilt $F(x - F(x)u) = 0$, also $x - F(x)u \in W$ und daher $\langle x - F(x)u, u \rangle = 0$. Es folgt

$$F(x) = \frac{\langle x, u \rangle}{\|u\|^2}, \qquad \text{also} \quad f = \frac{u}{\|u\|^2}$$

wie gewünscht. Eindeutigkeit: Aus $\langle x, f \rangle = \langle x, g \rangle$ für alle x folgt mit $x = f - g$, dass $\|f - g\|^2 = 0$, also $f = g$. ⌟

Streng genommen ist der Satz von Fréchet–Riesz für den Hilbertraum L^2 formuliert; wir verwenden ihn für $\mathcal{L}^2$. Der Preis dafür ist, dass das darstellende Element nur noch fast überall eindeutig ist.

Schritt 2: Die Funktion g. Nach dem Satz von Fréchet–Riesz gibt es ein $g \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu+\nu)$ mit

$$\int h \,\mathrm{d}\nu = \int hg \,\mathrm{d}(\mu+\nu) \qquad \text{für alle } h \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu+\nu).$$

Wegen $\int h \,\mathrm{d}(\mu+\nu) = \int h \,\mathrm{d}\mu + \int h \,\mathrm{d}\nu$ ist dies äquivalent zu

$$\int h(1-g) \,\mathrm{d}(\mu+\nu) = \int h \,\mathrm{d}\mu \qquad \text{für alle } h \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu+\nu). \tag{$*$}$$

Da $\mu+\nu$ endlich ist, liegen alle Indikatorfunktionen in $\mathcal{L}^2(\mu+\nu)$. Mit $h = \mathbb{1}\{g<0\}$ folgt

$$0 \le \nu(\{g<0\}) = \int_{\{g<0\}} g \,\mathrm{d}(\mu+\nu) \le 0,$$

also $\int_{\{g<0\}} g \,\mathrm{d}(\mu+\nu) = 0$; da der Integrand auf $\{g<0\}$ strikt negativ ist, folgt $(\mu+\nu)(\{g<0\}) = 0$, d. h. $g \ge 0$ $(\mu+\nu)$-fast überall. Mit $h = \mathbb{1}\{g>1\}$ in $(*)$ folgt ebenso

$$0 \le \mu(\{g>1\}) = \int_{\{g>1\}} (1-g) \,\mathrm{d}(\mu+\nu) \le 0,$$

und daraus wie eben $(\mu+\nu)(\{g>1\}) = 0$, also $g \le 1$ $(\mu+\nu)$-fast überall. Nach Abänderung auf einer $(\mu+\nu)$-Nullmenge dürfen wir $0 \le g \le 1$ überall annehmen.

Sei nun $h \ge 0$ messbar und $h_n := \min\{h, n\} \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu+\nu)$, also $h_n \uparrow h$. Der Satz von der monotonen Konvergenz, angewandt auf die Maße μ und ν sowie auf $(\mu+\nu)$ mit den nichtnegativen Integranden $h_n(1-g)$ bzw. $h_n g$, zeigt: Die Identität $(*)$ und ebenso $\int h \,\mathrm{d}\nu = \int hg \,\mathrm{d}(\mu+\nu)$ gelten sogar für alle messbaren $h \ge 0$.

Schritt 3: Die Zerlegung. Sei $E := g^{-1}(\{1\})$. Mit $h = \mathbb{1}_E$ in $(*)$ folgt $\mu(E) = \int_E (1-g) \,\mathrm{d}(\mu+\nu) = 0$. Wir definieren Maße ν_a und ν_s durch

$$\nu_a(A) := \nu(A \setminus E), \qquad \nu_s(A) := \nu(A \cap E), \qquad A \in \mathcal{A}.$$

Dann ist $\nu = \nu_a + \nu_s$, und wegen $\nu_s(\Omega \setminus E) = 0$ und $\mu(E) = 0$ gilt $\nu_s \perp \mu$.

Zur Absolutstetigkeit von ν_a: Sei zunächst $A \in \mathcal{A}$ mit $A \cap E = \emptyset$ und $\mu(A) = 0$. Nach $(*)$ mit $h = \mathbb{1}_A$ ist

$$0 = \mu(A) = \int_A (1-g) \,\mathrm{d}(\mu+\nu),$$

und da $1-g > 0$ auf A, folgt $(\mu+\nu)(A) = 0$, also $\nu(A) = 0$. Ist nun $B \in \mathcal{A}$ beliebig mit $\mu(B) = 0$, so gilt auch $\mu(B \setminus E) = 0$, und nach dem eben Gezeigten folgt

$$\nu_a(B) = \nu(B \setminus E) = 0.$$

Also ist $\nu_a \ll \mu$, und $\nu = \nu_a + \nu_s$ ist die gewünschte Zerlegung.

Schritt 4: Die Dichte. Setze

$$f := \frac{g}{1-g} \mathbb{1}_{\Omega \setminus E}$$

(mit $f := 0$ auf E); f ist messbar, nichtnegativ und überall endlich. Für $A \in \mathcal{A}$ liefert die erste Identität aus Schritt 2 mit $h = \mathbb{1}_{A \setminus E}$ und anschließend $(*)$ mit $h = f\,\mathbb{1}_{A \setminus E} \ge 0$

$$\nu_a(A) = \nu(A \setminus E) = \int_{A \setminus E} g \,\mathrm{d}(\mu+\nu) = \int_{A \setminus E} f(1-g) \,\mathrm{d}(\mu+\nu) = \int_{A \setminus E} f \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu,$$

wobei wir im letzten Schritt $\mu(E) = 0$ verwendet haben. Also ist $f = \mathrm{d}\nu_a / \mathrm{d}\mu$. □

Bemerkung. Ist $\nu \ll \mu$, so sieht man durch maßtheoretische Induktion (zuerst für Indikatorfunktionen, dann für einfache, nichtnegative und schließlich integrierbare Funktionen)

$$\int g \, \mathrm{d}\nu = \int g \, \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\mu,$$

und zwar in dem Sinne, dass, wenn eine Seite existiert, auch die andere existiert und beide übereinstimmen.

Korollar 7.28 (Produktregel für Dichten)**.** *Seien μ_1, μ_2, μ_3 σ-endliche Maße auf $(\Omega, \mathcal{A})$ mit $\mu_3 \ll \mu_2$ und $\mu_2 \ll \mu_1$. Dann gilt $\mu_3 \ll \mu_1$ und*

$$\frac{\mathrm{d}\mu_3}{\mathrm{d}\mu_1} = \frac{\mathrm{d}\mu_3}{\mathrm{d}\mu_2} \frac{\mathrm{d}\mu_2}{\mathrm{d}\mu_1} \qquad \mu_1\text{-fast überall.}$$

Beweis. (1) Aus $\mu_1(A) = 0$ folgt $\mu_2(A) = 0$ und daraus $\mu_3(A) = 0$. Also gilt $\mu_3 \ll \mu_1$.

(2) Seien $f := \mathrm{d}\mu_2/\mathrm{d}\mu_1$ und $g := \mathrm{d}\mu_3/\mathrm{d}\mu_2$ (diese existieren nach Korollar 7.27), d. h. $\mu_2(A) = \int_A f \, \mathrm{d}\mu_1$ und $\mu_3(A) = \int_A g \, \mathrm{d}\mu_2$ für $A \in \mathcal{A}$. Nach der vorangehenden Bemerkung, angewandt auf $g\mathbb{1}_A \geq 0$, gilt

$$\mu_3(A) = \int_A g \, \mathrm{d}\mu_2 = \int_A g \, \frac{\mathrm{d}\mu_2}{\mathrm{d}\mu_1} \, \mathrm{d}\mu_1 = \int_A gf \, \mathrm{d}\mu_1$$

für alle $A \in \mathcal{A}$. Also ist gf eine Dichte von μ_3 bezüglich μ_1, d. h. $\mathrm{d}\mu_3/\mathrm{d}\mu_1 = gf$ μ_1-fast überall. □

7.4 Hauptsatz der Differential- und Integralrechnung

Zur Erinnerung: Der Hauptsatz der Differential- und Integralrechnung aus der Analysis besagt:

(a) Für jede stetig differenzierbare Funktion $F\colon [a, b] \to \mathbb{R}$ gilt

$$F(x) = F(a) + \int_a^x F'(y) \, \mathrm{d}y \qquad \text{für alle } x \in [a, b].$$

(b) Für jede stetige Funktion $f\colon [a, b] \to \mathbb{R}$ ist

$$F(x) := \int_a^x f(y) \, \mathrm{d}y, \qquad x \in [a, b],$$

stetig differenzierbar mit $F' = f$.

Zum Abschluss wollen wir eine Erweiterung bzw. natürliche Formulierung dieses Hauptsatzes im Rahmen der Lebesgue-Integration andeuten.

Idee: Ist ν ein endliches Maß auf $(\mathbb{R}, \mathcal{B})$ mit $\nu \ll \lambda$, so gilt nach dem Satz von Radon–Nikodym (Korollar 7.27) mit $f = \mathrm{d}\nu/\mathrm{d}\lambda$

$$F(x) := \nu((-\infty, x]) = \int_{-\infty}^x f(y) \, \lambda(\mathrm{d}y), \qquad x \in \mathbb{R}.$$

Dies liefert einen Ableitungsbegriff für Verteilungsfunktionen: f sollte die „Ableitung“ von F sein. Als Input benötigen wir den folgenden Satz.

Satz 7.29 (Lebesgue)**.** *Sei $F\colon \mathbb{R} \to \mathbb{R}$ nichtfallend. Dann ist F Lebesgue-fast überall differenzierbar, und es gibt eine Borel-messbare Funktion $f \geq 0$ mit*

$$F'(x) = f(x) \qquad \text{für } \lambda\text{-fast alle } x \in \mathbb{R}.$$

Ohne Beweis; siehe z. B. [Coh13].

Bemerkung. Ist μ ein endliches Maß auf $(\mathbb{R}, \mathcal{B})$, so ist $F_\mu(x) \coloneqq \mu((-\infty, x])$ nichtfallend; Satz 7.29 ist also insbesondere auf Verteilungsfunktionen anwendbar.

Wann ist nun eine Funktion als Differenz zweier nichtfallender Funktionen darstellbar?

Definition 7.30 (Variation)**.** Für eine reellwertige Funktion F und $a < b$ setze

$$V_F[a,b] \coloneqq \sup\Big\{\sum_{k=1}^{r} |F(t_k) - F(t_{k-1})| \colon r \in \mathbb{N},\ a = t_0 < t_1 < \cdots < t_r = b\Big\}.$$

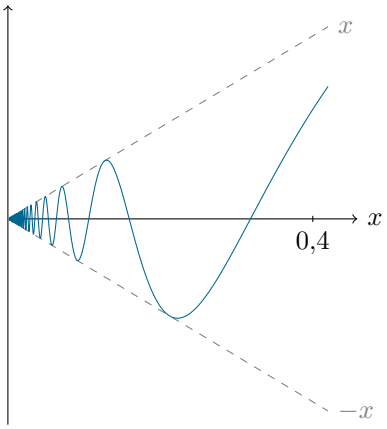
Die Größe $V_F[a,b]$ heißt die *Variation* von F auf $[a,b]$. Ist $V_F[a,b] < \infty$, so heißt F *von beschränkter Variation* auf $[a,b]$. Für $a = -\infty$ bzw. $b = \infty$ definiert man V_F analog, indem man das Supremum über alle endlichen Zerlegungspunkte $t_0 < \cdots < t_r$ in dem jeweiligen Intervall bildet; z. B. ist $V_F(-\infty, x] = \lim_{a \to -\infty} V_F[a,x]$. Die Funktion F heißt von beschränkter Variation (auf $\mathbb{R}$), wenn $V_F(-\infty, \infty) < \infty$.

Bemerkung.

(1) Die Funktion

$$F(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0, \\ 0, & x = 0, \end{cases}$$

ist stetig, aber auf $[0,1]$ *nicht* von beschränkter Variation. (An den Stellen $x_k = 1/(\pi/2 + k\pi)$ ist $|F(x_k)| = x_k$, und die Reihe $\sum_k x_k$ divergiert.)



$F(x) = x \sin(1/x)$ auf $(0; 0{,}42]$ zwischen den Geraden $\pm x$: Die Amplituden fallen nur wie x, die Anzahl der Schwingungen wächst aber so schnell, dass die Summe der Auf- und Abstiege unendlich ist.

(2) Ist F nichtfallend, so gilt

$$V_F(-\infty, \infty) = \lim_{t \to \infty} F(t) - \lim_{t \to -\infty} F(t).$$

(3) Funktionen von beschränkter Variation sind beschränkt.

(4) Ist F von beschränkter Variation, so gilt für $V_F(x) \coloneqq V_F(-\infty, x]$:

a) V_F ist nichtfallend und beschränkt;

b) $\lim_{x\to-\infty} V_F(x) = 0$;

c) ist F rechtsstetig, so ist auch V_F rechtsstetig.

(5) Es gilt $V_{\alpha F} = |\alpha|\, V_F$ für $\alpha \in \mathbb{R}$ und $V_{F+G} \le V_F + V_G$. Ist F von beschränkter Variation, so gilt für $a < b$

$$V_F(b) = V_F(a) + V_F[a, b].$$

Die Nachweise dieser Eigenschaften seien als Übung empfohlen.

Satz 7.31. *Eine Funktion $F\colon \mathbb{R} \to \mathbb{R}$ ist genau dann von beschränkter Variation, wenn sie sich als Differenz zweier beschränkter, monoton nichtfallender Funktionen darstellen lässt.*

Ohne Beweis.

Bemerkung. Insbesondere ist jede Funktion F von beschränkter Variation nach Satz 7.29 (angewandt auf die beiden monotonen Funktionen aus Satz 7.31) λ-fast überall differenzierbar.

Wann kann man also F durch Integration seiner Ableitung f zurückerhalten?

Definition 7.32 (Absolutstetigkeit von Funktionen)**.** Eine reellwertige Funktion F heißt *absolutstetig*, falls gilt: Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$, so dass für jede endliche Familie paarweise disjunkter offener Intervalle $(s_1, t_1), \dots, (s_r, t_r)$ gilt:

$$\sum_{k=1}^{r} (t_k - s_k) < \delta \implies \sum_{k=1}^{r} |F(t_k) - F(s_k)| < \varepsilon.$$

Bemerkung (Übung).

(1) Absolutstetige Funktionen sind gleichmäßig stetig (Wahl $r = 1$).

(2) Jede absolutstetige Funktion ist auf jedem beschränkten Intervall $[a, b]$ von beschränkter Variation.

(3) Jede Lipschitz-stetige Funktion ist absolutstetig.

(4) Für ein endliches Maß μ auf $(\mathbb{R}, \mathcal{B})$ mit $F_\mu(x) = \mu\big((-\infty, x]\big)$ gilt

$$\mu \ll \lambda \iff F_\mu \text{ ist absolutstetig.}$$

Satz 7.33 (Hauptsatz)**.** *(I) Für eine Funktion $F\colon [a, b] \to \mathbb{R}$ sind äquivalent:*

a) F ist absolutstetig.

b) F ist λ-fast überall auf $[a, b]$ differenzierbar, F' ist integrierbar auf $[a, b]$, und

$$F(x) = F(a) + \int_a^x F'(y)\,\mathrm{d}y \qquad \text{für alle } x \in [a, b].$$

(II) Für eine Funktion $F\colon \mathbb{R} \to \mathbb{R}$ sind äquivalent:

a) F ist absolutstetig und von beschränkter Variation mit $\lim_{x\to-\infty} F(x) = 0$.

b) F ist λ-fast überall auf $\mathbb{R}$ differenzierbar, F' ist integrierbar auf $\mathbb{R}$, und

$$F(x) = \int_{-\infty}^{x} F'(y)\,\mathrm{d}y \qquad \text{für alle } x \in \mathbb{R}.$$

Umgekehrt gilt:

(III) Ist f Lebesgue-integrierbar auf $[a,b]$, so ist

$$F(x) \coloneqq \int_{a}^{x} f(y)\,\mathrm{d}y, \qquad x \in [a,b],$$

absolutstetig mit $F'(x) = f(x)$ für λ-fast alle $x \in [a,b]$.

(IV) Ist f Lebesgue-integrierbar auf $\mathbb{R}$ und

$$F(x) \coloneqq \int_{-\infty}^{x} f(y)\,\mathrm{d}y, \qquad x \in \mathbb{R},$$

so ist F absolutstetig und von beschränkter Variation mit $\lim_{x\to-\infty} F(x) = 0$, und es gilt $F'(x) = f(x)$ für λ-fast alle $x \in \mathbb{R}$.

Ohne Beweis.

Wiederholung und Ausblick

Zum Abschluss stellen wir die wichtigsten Begriffe und Ergebnisse der Vorlesung noch einmal zusammen.

1. Wahrscheinlichkeitsräume

Grundlage sind Wahrscheinlichkeitsräume $(\Omega, \mathcal{F}, \mu)$ oder allgemeiner Maßräume. Wir haben zahlreiche Modelle kennengelernt:

(1) *Diskrete Modelle:* Ω abzählbar, $\mathcal{F} = \mathcal{P}(\Omega)$, und μ ist durch eine Zähldichte $(\rho_i)_{i\in\Omega}$ gegeben. Beispiele sind die Bernoulli-, Binomial- und Poisson-Verteilung sowie die Urnenmodelle.

(2) *Kontinuierliche Modelle* mit Hilfe des Lebesgue-Maßes: $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}^d$ und

$$\mu(A) = \int_A f \,\mathrm{d}\lambda, \qquad A \in \mathcal{B}^d,$$

mit einer Dichte f. Beispiele sind die Normal-, die Beta- und die Exponentialverteilung.

(3) *Allgemeiner Zugang:* $(\Omega, \mathcal{F}, \mu)$ beliebig. Neue Maße erhält man z. B.

- als absolutstetige Maße $\nu \ll \mu$; für σ-endliche μ, ν ist dies nach dem Satz von Radon–Nikodym (Korollar 7.27) äquivalent zu $\nu(A) = \int_A f \,\mathrm{d}\mu$ mit der Dichte $f = \mathrm{d}\nu/\mathrm{d}\mu$;
- als Bildmaße: Ist $X\colon (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ messbar, so ist $\nu = \mu \circ X^{-1}$ ein Maß auf $(\Omega', \mathcal{F}')$.

2. Integration

Gegeben seien ein Maßraum $(\Omega, \mathcal{F}, \mu)$ und messbare Funktionen $f_n, f\colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$.

(1) Definition des Integrals $\int f \,\mathrm{d}\mu$.

(2) Konvergenzsätze: Unter welchen Bedingungen folgt aus $f_n \to f$ auch $\int f_n \,\mathrm{d}\mu \to \int f \,\mathrm{d}\mu$? Hierzu dienen

a) der Satz von der monotonen Konvergenz,

b) der Satz von der dominierten (majorisierten) Konvergenz,

c) das Lemma von Fatou,

d) das Differentiationslemma (Vertauschung von Differentiation und Integration) usw.

3. Bedingte Wahrscheinlichkeit und Unabhängigkeit

(1) Bedingte Wahrscheinlichkeit $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ für $\mathbb{P}(B) > 0$; Satz von Bayes.

(2) Ereignisse A, B heißen unabhängig, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

(3) Zufallsvariablen f_1, f_2 heißen unabhängig, falls die σ-Algebren $\sigma(f_1)$ und $\sigma(f_2)$ unabhängig sind. Dies kann auf $\cap$-stabilen Erzeugern geprüft werden.

(4) Abzählbar viele unabhängige Experimente können auf einem gemeinsamen Wahrscheinlichkeitsraum realisiert werden. Feinheit: Das Produktmaß ist erst dann eindeutig bestimmt, wenn es auf dem Produktraum mit der Produkt-σ-Algebra definiert wird.

(5) Summen unabhängiger Zufallsvariablen sind gemäß der Faltung ihrer Verteilungen verteilt.

(6) Asymptotische Ereignisse:

a) *0-1-Gesetz von Kolmogorov:* Sind $Y_1, Y_2, \dots$ unabhängig und ist A ein asymptotisches Ereignis, d. h. $A \in \bigcap_{n\geq 1} \sigma(Y_k \colon k \geq n)$, so gilt $\mathbb{P}(A) \in \{0, 1\}$.

b) *Lemma von Borel–Cantelli:* Seien $A_1, A_2, \dots \in \mathcal{F}$ und

$$A \coloneqq \{\omega \in \Omega \colon \omega \in A_k \text{ für unendlich viele } k\}.$$

Dann gilt:

(i) $\sum_{k\geq 1} \mathbb{P}(A_k) < \infty \implies \mathbb{P}(A) = 0$;

(ii) $\sum_{k\geq 1} \mathbb{P}(A_k) = \infty$ und $A_1, A_2, \dots$ unabhängig $\implies \mathbb{P}(A) = 1$.

4. Erwartungswert und Varianz

Sei X eine reelle Zufallsvariable auf $(\Omega, \mathcal{F}, \mu)$.

- Erwartungswert: $\mathbb{E}[X] = \int_\Omega X \,\mathrm{d}\mu$ für $X \in \mathcal{L}^1(\mu)$.
- Momente: $\mathbb{E}[X^n] = \int_\Omega X^n \,\mathrm{d}\mu$ für $X \in \mathcal{L}^n(\mu)$.
- Varianz: $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ für $X \in \mathcal{L}^2(\mu)$.

Die Räume $\mathcal{L}^n(\mu)$ sind (nach Abänderung auf Nullmengen) Vektorräume, $L^2(\mu)$ ist sogar ein Hilbertraum, und auf $\mathcal{L}^1(\mu)$ ist das Integral linear:

$$\int (\alpha X + Y) \,\mathrm{d}\mu = \alpha \int X \,\mathrm{d}\mu + \int Y \,\mathrm{d}\mu.$$

Beispiele.

(1) Sei X eine reelle, integrierbare Zufallsvariable auf $(\mathbb{R}^d, \mathcal{B}^d, \mu)$ mit $\mu(A) = \int_A f \,\mathrm{d}\lambda$. Dann gilt (vgl. die Bemerkung nach dem Beweis von Satz 7.26)

$$\mathbb{E}_\mu[X] = \int X(\omega) \,\mu(\mathrm{d}\omega) = \int X(\omega) \,\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(\omega) \,\lambda(\mathrm{d}\omega) = \int X(\omega) f(\omega) \,\lambda(\mathrm{d}\omega).$$

(2) Erwartungswerte und Varianzen z. B. der Normal-, Exponential- und Binomialverteilung.

5. Gesetze der großen Zahlen und zentraler Grenzwertsatz

Für Zufallsvariablen $X_1, X_2, \dots$ bezeichne

$$X^{(n)} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

das *empirische Mittel*. Wir haben drei Konvergenzbegriffe kennengelernt:

(1) *Konvergenz in Wahrscheinlichkeit:* $Y_n \xrightarrow{\mathbb{P}} Y$, falls $\mathbb{P}(|Y_n - Y| > \varepsilon) \to 0$ für alle $\varepsilon > 0$.

(2) *Fast sichere Konvergenz:* $Y_n \to Y$ $\mathbb{P}$-f. s., falls $\mathbb{P}\big(\lim_{n\to\infty} Y_n = Y\big) = 1$.

(3) *Konvergenz in Verteilung:* $\mathbb{E}[f(Y_n)] \to \mathbb{E}[f(Y)]$ für alle stetigen, beschränkten Funktionen f.

Wichtige Hilfsmittel sind

(1) die *Markov-Ungleichung*: Für jede monoton wachsende Funktion $f\colon [0,\infty) \to [0,\infty)$ und $\varepsilon > 0$ mit $f(\varepsilon) > 0$ gilt

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[f(|X|)]}{f(\varepsilon)},$$

etwa mit $f(x) = x$, $f(x) = x^2$ oder $f(x) = \mathrm{e}^{tx}$, $t > 0$;

(2) das Lemma von Borel–Cantelli.

Gesetze der großen Zahlen.

(1) *Schwaches Gesetz:* Sind $X_i \in \mathcal{L}^2$ nur paarweise unkorreliert mit beschränkten Varianzen, so gilt $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \xrightarrow{\mathbb{P}} 0$, und zwar mit Rate $1/n$:

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])\Big| \geq \varepsilon\Big) \leq \frac{v}{n\varepsilon^2}, \qquad v := \sup_i \operatorname{Var}[X_i].$$

Sind die $X_i \in \mathcal{L}^1$ u. i. v., so gilt $X^{(n)} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1]$.

(2) *Starkes Gesetz:* Sind $X_i \in \mathcal{L}^2$ nur paarweise unkorreliert mit beschränkten Varianzen, so gilt $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \to 0$ fast sicher.

Zentraler Grenzwertsatz: Sind $X_1, X_2, \dots \in \mathcal{L}^2$ u. i. v. mit $\operatorname{Var}[X_1] > 0$, so konvergieren die normierten Summen

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mathbb{E}[X_i]}{\sqrt{\operatorname{Var}[X_i]}} \longrightarrow \mathcal{N}_{0,1}$$

in Verteilung gegen die Standardnormalverteilung.

6. Markov-Ketten

Eine Markov-Kette $(X_n)_{n\geq 0}$ auf einem abzählbaren Zustandsraum E mit Übergangsmatrix Π ist durch die Markov-Eigenschaft

$$\mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_n = x_n) = \Pi(x_n, x_{n+1})$$

charakterisiert. Für $z \in E$ sei $\tau_z := \inf\{n \geq 1 \colon X_n = z\}$ die Eintritts- bzw. Rückkehrzeit nach z; sie ist eine Stoppzeit. Die Bezeichnungen $\mathbb{P}^x$ und $\mathbb{E}^x$ beziehen sich auf den Start in $X_0 = x$.

(1) *Absorptionswahrscheinlichkeiten:* Ist z absorbierend, d. h. $\Pi(z, z) = 1$, so löst $h_z(x) := \mathbb{P}^x(\tau_z < \infty)$ die Gleichung

$$\sum_{y\in E} \Pi(x, y)\, h_z(y) = h_z(x), \qquad x \in E.$$

(2) a) *Ergodensatz:* Ist $|E| < \infty$ und Π irreduzibel und aperiodisch (d. h. es gibt ein $k \geq 1$ mit $\Pi^k(x, y) > 0$ für alle $x, y \in E$), so gibt es genau eine Verteilung α auf E mit

$$\Pi^n(y, x) \xrightarrow[n\to\infty]{} \alpha(x) > 0 \qquad \text{für alle } x, y \in E;$$

α ist sogar stationär, d. h. $\alpha\Pi = \alpha$.

b) *Stationäre Verteilungen und Rückkehrzeiten:* Sei E abzählbar und Π irreduzibel. Existiert eine stationäre Verteilung α, so ist sie eindeutig bestimmt, und es gilt

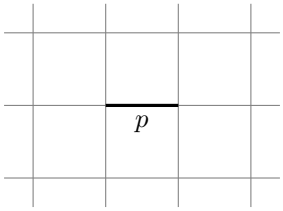
$$0 < \alpha(x) = \frac{1}{\mathbb{E}^x[\tau_x]} \qquad \text{für alle } x \in E.$$

c) *Rekurrenz und Transienz:* Entweder ist ein Zustand x rekurrent, d. h. die in x gestartete Kette kehrt mit Wahrscheinlichkeit 1 unendlich oft nach x zurück, oder er ist transient, d. h. die Kette kehrt fast sicher nur endlich oft zurück, und die erwartete Anzahl der Rückkehren ist endlich.

d) *Positive Rekurrenz:* Ist Π irreduzibel und $\mathbb{E}^x[\tau_x] < \infty$, so gibt es eine stationäre Verteilung α, d. h. $\alpha\Pi = \alpha$, und es gilt $\alpha(x) = 1/(\mathbb{E}^x[\tau_x])$.

Ausblick: Perkolation

Als Idee für weiterführende Fragestellungen sei die *Perkolation* genannt: Jede Kante eines Gitters, etwa des $\mathbb{Z}^2$, ist unabhängig von allen anderen mit Wahrscheinlichkeit p „offen“ und sonst „geschlossen“.



Literaturverzeichnis

[Bau02] H. Bauer, *Wahrscheinlichkeitstheorie*, De Gruyter Lehrbuch, De Gruyter, Berlin, 2002, 5. Auflage.

[Coh13] D. L. Cohn, *Measure Theory*, Birkhäuser Advanced Texts Basler Lehrbücher, Birkhäuser, New York, 2013, 2. Auflage.

[Dur19] R. Durrett, *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 49, Cambridge University Press, Cambridge, 2019, 5. Auflage.

[Els18] J. Elstrodt, *Maß- und Integrationstheorie*, Springer-Lehrbuch, Springer Spektrum, Berlin, 2018, 8. Auflage.

[Geo15] H.-O. Georgii, *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*, De Gruyter Studium, De Gruyter, Berlin, 2015, 5. Auflage.

[GS20a] G. R. Grimmett and D. R. Stirzaker, *One Thousand Exercises in Probability*, Oxford University Press, Oxford, 2020, 3. Auflage.

[GS20b] ———, *Probability and Random Processes*, Oxford University Press, Oxford, 2020, 4. Auflage.

[Häg02] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, London Mathematical Society Student Texts, vol. 52, Cambridge University Press, Cambridge, 2002.

[Hen23] N. Henze, *Stochastik für Einsteiger*, Springer Spektrum, Wiesbaden, 2023, 14. Auflage.

[Kle20] A. Klenke, *Wahrscheinlichkeitstheorie*, Masterclass, Springer Spektrum, Berlin, 2020, 4. Auflage.

[Kre05] U. Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, vieweg studium – Aufbaukurs Mathematik, Vieweg, Wiesbaden, 2005, 8. Auflage.

[KW10] G. Kersting and A. Wakolbinger, *Elementare Stochastik*, Mathematik Kompakt, Birkhäuser, Basel, 2010, 2. Auflage.

[Nor97] J. R. Norris, *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1997.

[Sch17] R. L. Schilling, *Measures, Integrals and Martingales*, Cambridge University Press, Cambridge, 2017, 2. Auflage.