Wahrscheinlichkeitstheorie

Lutz Dümbgen
Stephan Kötzer, Kathrin Weyermann, Niki Zumbrunnen

Universität Bern
Wintersemester 2006/2007

10. Mai 2012

Literatur

P. BILLINGSLEY, *Probability and Measure*, Wiley

H. DELING, B. HAUPT, EINFÜHRUNG IN DIE WAHRSCHEINLICHKEITSTHEORIE UND STATISTIK, Springer.

L. DÜMBGEN, STOCHASTIK FÜR INFORMATIKER, Springer.

H.-O. GEORGII, *Stochastik*, de Gruyter

C. HESSE, *Angewandte Wahrscheinlichkeitstheorie*, Vieweg.

U. KRENGEL, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg.

Inhaltsverzeichnis

Kapitel 1

Wahrscheinlichkeitsräume und Zufallsvariablen

Bisher betrachteten wir im Wesentlichen diskrete Wahrscheinlichkeitsverteilungen P auf einer Menge Ω. Das heißt, für eine Gewichtsfunktion $p : \Omega \to [0, 1]$ mit

$$\sum_{\omega \in \Omega} p(\omega) \;=\; 1$$

ist

$$P(A) \;=\; \sum_{\omega \in A} p(\omega),$$

und insbesondere $p(\omega) = P(\{\omega\})$. Eine Abbildung X von Ω in eine beliebige Menge $\mathcal{X}$ ist dann eine Zufallsvariable mit Werten in $\mathcal{X}$, und sie induziert eine diskrete Wahrscheinlichkeitsverteilung P^X auf $\mathcal{X}$: Für $B \subset \mathcal{X}$ setzen wir

$$P^X(B) \;:=\; P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\}) \;=\; \sum_{x \in B} p^X(x),$$

wobei $p^X(x) := P(X = x) = \sum_{\omega \in \Omega : X(\omega) = x} p(\omega)$.

Obwohl man in diesem Rahmen schon einige interessante Konzepte (Zufallsvariablen, bedingte Wahrscheinlichkeiten, Unabhängigkeit) und Gesetzmäßigkeiten (Markov- und Tshebyshev-Ungleichung, schwaches Gesetz der großen Zahlen) behandeln kann, ist er für viele Anwendungen zu eng.

Beispiel: Die uniforme Verteilung auf einer Zielscheibe. Ein (schlechter) Schütze zielt auf eine Zielscheibe. Unter der Bedingung, dass er die Zielscheibe überhaupt trifft, sei der getroffene Punkt uniform verteilt auf der Zielscheibe. Um dies mathematisch formulieren, betrachten wir die Zielscheibe als Teilmenge Ω von $\mathbb{R}^2$ und machen für $A \subset \Omega$ den Ansatz

$$P(A) \;:=\; \text{Fläche}(A)/\text{Fläche}(\Omega).$$

Die Frage ist nur, ob und wie man für beliebige Teilmengen A von $\mathbb{R}^2$ ihren Flächeninhalt definieren kann. Geht man davon aus, dass jede einpunktige Menge Fläche Null hat, dann wird deutlich, dass P kein diskretes Wahrscheinlichkeitsmaß sein kann.

Der unendliche Münzwurf. Angenommen man möchte ein Modell für das beliebig oftmalige Werfen einer Münze, die mit Wahrscheinlichkeit $\theta \in (0,1)$ "Zahl" und mit Wahrscheinlichkeit $1-\theta$ "Kopf" zeigt. Als Grundraum bietet sich

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

an. Der i-te Münzwurf wird durch die Abbildung

$$\Omega \ni \omega \ \mapsto \ X_i(\omega) \ := \ \omega_i$$

beschrieben. Nun suchen wir nach einem Wahrscheinlichkeitsmaß P auf Ω, so dass für beliebige $i, n \in \mathbb{N}$ gilt:

$$P(X_i = 1) \ = \ 1 - P(X_i = 0) \ = \ \theta,$$

$$X_1, X_2, \ldots, X_n \text{ sind stochastisch unabhängig.}$$

Mit anderen Worten, für beliebige $n \in \mathbb{N}$ und $y_1, y_2, \ldots, y_n \in \{0,1\}$ soll gelten:

$$P(X_1 = y_1, X_2 = y_2, \ldots, X_n = y_n) \ = \ \prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i}.$$

Wir haben also eine Idee, wie man die Wahrscheinlichkeit von Ereignissen $A \subset \Omega$ definieren sollte, die nur von *endlich vielen* Münzwürfen X_i bestimmt werden. Leider sind viele interessante Ereignisse nicht von von diesem einfachen Typ. Zum Beispiel sei

$$\widehat{\theta}_n \ := \ \frac{1}{n}\sum_{i=1}^n X_i,$$

ein Schätzwert für θ. Wenn unsere intuitive Vorstellung von Wahrscheinlichkeiten richtig ist, sollte das Ereignis

$$A^{(1)} \ := \ \Big\{\lim_{n\to\infty} \widehat{\theta}_n = \theta\Big\}$$

Wahrscheinlichkeit Eins haben. Offensichtlich wird $A^{(1)}$ nicht von endlich vielen Münzwürfen bestimmt, und wir müssen uns mit der Berechnung von $P(A^{(1)})$ noch etwas gedulden.

Ein anderes Beispiel für ein komplizierteres Ereignis betrifft eine "Irrfahrt auf $\mathbb{Z}$". Sei $W_0 := 0$ und

$$W_t \ := \ \sum_{i=1}^t (2X_i - 1) \quad \text{für } t \in \mathbb{N}.$$

Diese Folge $(W_t)_{t=0}^\infty$ beschreibt die Positionen eines Teilchens in $\mathbb{Z}$, welches zum Zeitpunkt Null in 0 startet und sich zu jedem Zeitpunkt $n \in \mathbb{N}$ einen Schritt nach links oder rechts weiterbewegt.

Zur Illustration zeigt Abbildung 1.1 eine mögliche Realisation von $(X_i)_{i=1}^n$ und die entsprechende Folge $(W_t)_{t=0}^n$ für $n = 100$. Auch das Ereignis

$$A^{(2)} := \left\{ \sup_{t \geq 0} W_t = \infty \text{ und } \inf_{t \geq 0} W_t = -\infty \right\}$$

kann man durch abzählbar viele Mengenoperationen aus einfachen Ereignissen aufbauen; siehe Übungen.

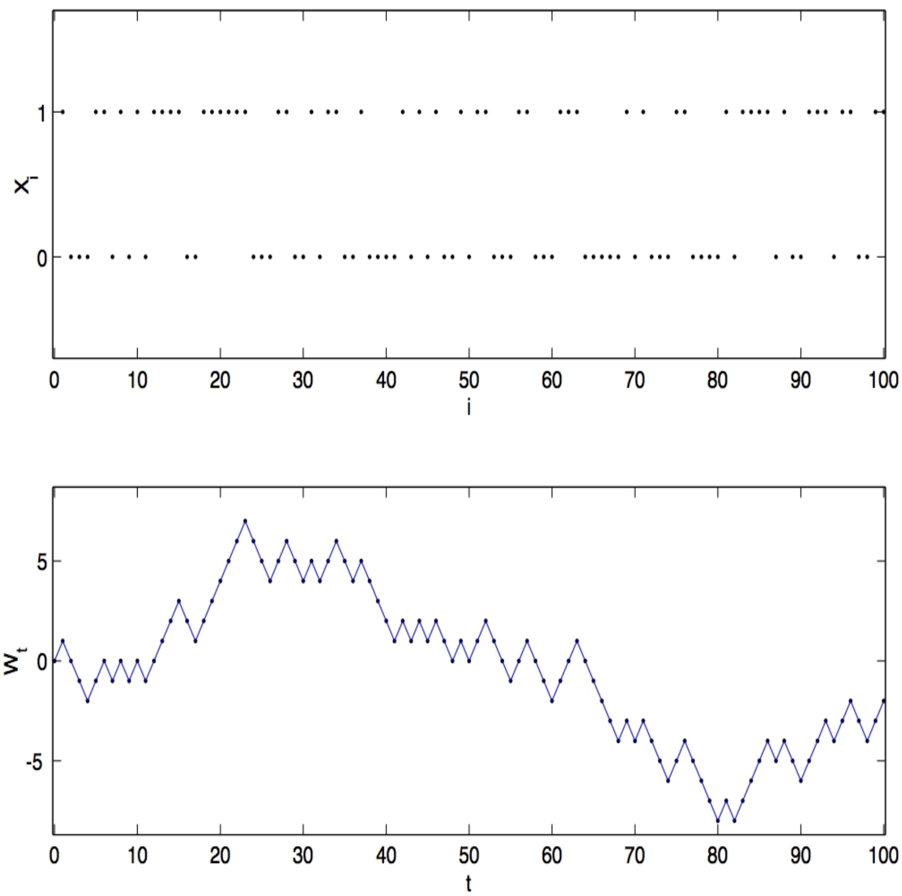

Abbildung 1.1: 100 Münzwürfe und die entsprechende Irrfahrt

Im folgenden Abschnitt werden wir sehen, wie man für die obigen beiden Beispiele Wahrscheinlichkeitsverteilungen mit den gewünschten Eigenschaften konstruieren kann.

1.1 Erste Grundlagen der Maßtheorie

1.1.1 Algebren und Inhalte

Definition 1.1 (Mengen-Algebra). Sei Ω eine Menge. Eine Mengenfamilie $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt *(Mengen-) Algebra über* Ω, wenn die folgenden drei Eigenschaften erfüllt sind:

(A.1) $\Omega \in \mathcal{A}$;

(A.2) mit $A \in \mathcal{A}$ ist auch $A^{\mathrm{c}} \in \mathcal{A}$;
(A.3) mit $A, B \in \mathcal{A}$ ist auch $A \cup B \in \mathcal{A}$.

Anmerkung 1.2 Erfüllt eine Mengenfamilie $\mathcal{A}$ Eigenschaft (A.2), so ist (A.1) äquivalent zu
(A.1') $\emptyset \in \mathcal{A}$,
und (A.3) ist äquivalent zu
(A.3') mit $A, B \in \mathcal{A}$ ist auch $A \cap B \in \mathcal{A}$.

Aus den Eigenschaften (A.1–3) folgt:
(A.3") Für $n \in \mathbb{N}$ und $A_1, A_2, \ldots, A_n \in \mathcal{A}$ sind auch $\bigcup_{i=1}^n A_i \in \mathcal{A}$ und $\bigcap_{i=1}^n A_i \in \mathcal{A}$;
(A.4) für $A, B \in \mathcal{A}$ ist auch $B \setminus A \in \mathcal{A}$.

Definition 1.3 (Inhalt und Wahrscheinlichkeitsinhalt). Sei $\mathcal{A}$ eine Algebra über einer Menge Ω. Eine Abbildung $M : \mathcal{A} \to [0, \infty]$ heißt *Inhalt auf* $\mathcal{A}$, wenn sie die folgenden zwei Eigenschaften erfüllt:
(I.1) $M(\emptyset) = 0$;
(I.2) sind $A, B \in \mathcal{A}$ disjunkt, dann ist $M(A \cup B) = M(A) + M(B)$.

Ist $M(\Omega) = 1$, dann spricht man von einem *Wahrscheinlichkeitsinhalt* auf $\mathcal{A}$.

Anmerkung 1.4 (Weitere Eigenschaften von Inhalten). Für einen Inhalt M auf einer Algebra $\mathcal{A}$ kann man (I.2) induktiv verallgemeinern:
(I.2') Für $n \in \mathbb{N}$ und paarweise disjunkte Mengen $A_1, A_2, \ldots, A_n \in \mathcal{A}$ ist

$$M\Big(\bigcup_{i=1}^n A_i\Big) \;=\; \sum_{i=1}^n M(A_i).$$

Ferner erfüllt M eine Monotonieeigenschaft:
(I.3) $M(A) \leq M(B)$ für $A, B \in \mathcal{A}$ mit $A \subset B$.

Im Falle von $M(\Omega) < \infty$ gilt die Siebformel von Sylvester-Poincaré:
(I.2") Für $n \in \mathbb{N}$ und beliebige Mengen $A_1, A_2, \ldots, A_n \in \mathcal{A}$ ist

$$M\Big(\bigcup_{i=1}^n A_i\Big) \;=\; \sum_{J \subset \{1,\ldots,n\},\, J \neq \emptyset} (-1)^{\#J-1} M\Big(\bigcap_{i \in J} A_i\Big).$$

Anmerkung 1.5 (Erste Bonferroni–Ungleichung). Für einen Inhalt M auf einer Algebra $\mathcal{A}$ gilt stets die folgende Ungleichung: Für beliebige Mengen $A_1, A_2, \ldots, A_n \in \mathcal{A}$ ist

$$M\Big(\bigcup_{i=1}^n A_i\Big) \;\leq\; \sum_{i=1}^n M(A_i);$$

mehr dazu in den Übungen.

Schlüsselbeispiel 1: Der Volumeninhalt im $\mathbb{R}^d$. Ein *Rechteck* im $\mathbb{R}^d$ ist eine Menge der Form

$$R \;=\; R_1 \times R_2 \times \cdots \times R_d$$

mit beliebigen Intervallen $R_i \subset \mathbb{R}$. Nun betrachten wir ein festes Rechteck $\Omega \subset \mathbb{R}^d$ und definieren

$$\mathcal{E} \;:=\; \Big\{ \bigcup_{i=1}^{n} R^{(i)} : n \in \mathbb{N}, R^{(i)} \subset \Omega \text{ ein Rechteck} \Big\},$$

die Menge der *Elementarmengen* in Ω. Man kann sich leicht davon überzeugen, dass $\mathcal{E}$ eine Algebra über Ω darstellt. Jede Elementarmenge lässt sich als Vereinigung von endlich vielen *paarweise disjunkten* Rechtecken darstellen. Eine solche Darstellung ist nicht eindeutig, doch liefert jede solche Darstellung ein und denselben Wert

$$\mathrm{Vol}\Big(\bigcup_{i=1}^{n} R^{(i)} \Big) \;:=\; \sum_{i=1}^{n} \mathrm{Vol}(R^{(i)})$$

mit

$$\mathrm{Vol}\big(R_1 \times R_2 \times \cdots \times R_d\big) \;:=\; \prod_{j=1}^{d} \text{Länge}(R_j).$$

Dabei verwenden wir die Konventionen, dass

$$a \cdot \infty = \infty \cdot a \;:=\; \begin{cases} \infty & \text{falls } a > 0, \\ 0 & \text{falls } a = 0, \\ -\infty & \text{falls } a < 0. \end{cases}$$

Wenn man sich davon überzeugt hat, dass dieser Volumeninhalt wohldefiniert ist, ist die Tatsache, dass $\mathrm{Vol}(\cdot)$ ein Inhalt auf $\mathcal{E}$ ist, ziemlich offensichtlich.

Schlüsselbeispiel 2: Wahrscheinlichkeitsverteilungen auf $\{0,1\}^{\mathbb{N}}$. Eine Menge $A \subset \Omega := \{0,1\}^{\mathbb{N}}$ der Form

$$A \;=\; \big\{ \omega \in \Omega : (\omega_1, \omega_2, \ldots, \omega_n) \in C \big\} \;=\; \big\{ (X_1, \ldots, X_n) \in C \big\}$$

mit beliebigem $n \in \mathbb{N}$ und $C \subset \{0,1\}^n$ nennen wir *Zylindermenge* in $\{0,1\}^n$. Es handelt sich also um eine Menge, die durch Bedingungen an *endlich viele* Komponenten von $\omega \in \Omega$ charakterisiert wird. Die obige Darstellung von A ist nicht eindeutig. Vielmehr gilt für eine natürliche Zahl $\tilde{n} > n$, $\tilde{C} \subset \{0,1\}^{\tilde{n}}$ und

$$\tilde{A} \;:=\; \big\{ (X_1, \ldots, X_{\tilde{n}}) \in \tilde{C} \big\}$$

die Beziehung:

$$A \;=\; \tilde{A} \quad \text{genau dann, wenn} \quad \tilde{C} \;=\; C \times \{0,1\}^{\tilde{n}-n}.$$

Mit dieser Vorüberlegung kann man leicht zeigen, dass die Menge $\mathcal{Z}$ aller Zylindermengen auf Ω eine Algebra ist. Jeden Wahrscheinlichkeitsinhalt auf dieser Algebra kann man wie folgt darstellen: Wir definieren

$$P\big(\big\{ (X_1, \ldots, X_n) \in C \big\}\big) \;:=\; \sum_{y \in C} p_n(y) \tag{1.1}$$

mit gewissen Gewichtsfunktionen $p_n : \{0,1\}^n \to [0,1]$, $n \in \mathbb{N}$, welche die folgenden Eigenschaften erfüllen:

$$p_1(0) + p_1(1) = 1, \tag{1.2}$$

$$p_{n+1}(y,0) + p_{n+1}(y,1) = p_n(y) \quad \text{für alle } n \in \mathbb{N}, y \in \{0,1\}^n. \tag{1.3}$$

Die zweitgenannte Eigenschaft ist wichtig, damit unsere Definition (1.1) überhaupt Sinn macht, also nicht von der speziellen Darstellung einer Zylindermenge abhängt.

Spezialfall 1: Der unendliche Münzwurf. Für $\theta \in (0,1)$ setzen wir

$$p_n(y) := \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}.$$

Diese Gewichte erfüllen die Eigenschaften (1.2–1.3).

Spezialfall 2: Ein Urnenmodell. In einer Urne befinden sich zunächst zwei Kugeln, eine mit '1' und eine mit '0' beschriftet. Zum Zeitpunkt $n \in \mathbb{N}$ zieht man rein zufällig eine Kugel und notiert die daraufstehende Zahl ($\rightsquigarrow \omega_n$). Danach legt man diese Kugel zusammen mit einer Kopie hiervon zurück. Dieses Gedankenmodell kann man durch folgende Wahrscheinlichkeitsgewichte präzisieren:

$$p_1(0) := p_1(1) := 1/2,$$

und induktiv setzen wir für $n \in \mathbb{N}$ sowie $y \in \{0,1\}^n$:

$$\begin{aligned} p_{n+1}(y,0) &:= p_n(y) \cdot \frac{n+1-\sum_{i=1}^n y_i}{n+2}, \\ p_{n+1}(y,1) &:= p_n(y) \cdot \frac{\sum_{i=1}^n y_i}{n+2}. \end{aligned}$$

Auch diese Gewichte erfüllen die Eigenschaften (1.2–1.3). In den Übungen wird dieses Urnenmodell noch genauer untersucht.

Die Eudoxos–Fortsetzung

Schon in der Antike wurden Flächen- und Volumeninhalte für andere Mengen als Rechtecke und Elementarmengen durch einen Approximationsprozess definiert. Die abstrakte Beschreibung ist wie folgt: Für einen Inhalt M auf einer Algebra $\mathcal{A}$ über Ω und eine beliebige Menge $B \subset \Omega$ definieren wir ihren inneren bzw. äußeren Inhalt als

$$\begin{aligned} \underline{M}(B) &:= \sup\bigl\{M(A) : A \in \mathcal{A}, A \subset B\bigr\}, \\ \overline{M}(B) &:= \inf\bigl\{M(A) : A \in \mathcal{A}, A \supset B\bigr\}. \end{aligned}$$

Satz 1.6 *Sei $M(\Omega) < \infty$. Mit obigen Funktionen $\underline{M}$ und $\overline{M}$ auf $\mathcal{P}(\Omega)$ ist*

$$\overline{\mathcal{A}} := \bigl\{B \subset \Omega : \underline{M}(B) = \overline{M}(B)\bigr\}$$

eine Algebra über Ω, welche $\mathcal{A}$ enthält. Ferner ist $\overline{M}$ ein Inhalt auf $\overline{\mathcal{A}}$, und $\overline{M}(A) = M(A)$ für alle $A \in \mathcal{A}$.

Übrigens lohnt es sich nicht, diesen Fortsetzungsprozess mit $(\overline{\mathcal{A}}, \overline{M})$ an Stelle von $(\mathcal{A}, M)$ zu wiederholen, denn

$$\begin{aligned} \sup\big\{\overline{M}(A) : A \in \overline{\mathcal{A}}, A \subset B\big\} &= \underline{M}(B), \\ \inf\big\{\overline{M}(A) : A \in \overline{\mathcal{A}}, A \supset B\big\} &= \overline{M}(B). \end{aligned}$$

Schlüsselbeispiel 1 (Forts.). Wendet man die Eudoxos–Fortsetzung auf die Elementarmengen in einem beschränkten Rechteck $\Omega \subset \mathbb{R}^2$ an, dann ergibt sich eine sinnvolle Definition des Flächeninhalts von Dreiecken, Ellipsen und vielen anderen Mengen. Insbesondere ergibt sich aus der Konstruktion des Riemann-Integrals, dass für eine Riemann-integrierbare Funktion $f : [a, b] \to [0, \infty)$ mit $[a, b] \subset \Omega_1$ und $\big[0, \sup(f)\big] \subset \Omega_2$ gilt:

$$B \ := \ \big\{(x, y) : a \le x \le b, 0 \le y \le f(y)\big\} \ \in \ \overline{\mathcal{E}} \quad \text{und} \quad \overline{\mathrm{Vol}}(B) \ = \ \int_a^b f(x)\, dx.$$

Dass die Eudoxos–Fortsetzung noch nicht alle Probleme löst, zeigt sich beispielsweise an der Menge

$$B \ := \ \Omega \cap \mathbb{Q}^d.$$

Da diese Menge abzählbar ist und $\mathrm{Vol}(\{\omega\}) = 0$ für jeden Punkt $\omega \in \Omega$, wäre es naheliegend, $\mathrm{Vol}(B) = 0$ zu setzen. Aber $\underline{\mathrm{Vol}}(B) = 0$ und $\overline{\mathrm{Vol}}(B) = \mathrm{Vol}(\Omega)$.

Schlüsselbeispiel 2 (Forts.). Auch hier kommen wir mit der Eudoxos–Fortsetzung noch nicht weit genug. Denn für das spezielle Ereignis

$$A^{(1)} \ := \ \Big\{\lim_{n\to\infty} \widehat{\theta}_n = \theta\Big\} \quad \text{mit} \quad \widehat{\theta}_n(\omega) \ := \ \frac{1}{n}\sum_{i=1}^n \omega_i$$

und beliebige Zylindermengen $A \subset \Omega$ gilt:

$$\begin{aligned} A \subset A^{(1)} &\quad \text{impliziert, dass} \quad A = \emptyset, \\ A \supset A^{(1)} &\quad \text{impliziert, dass} \quad A = \Omega. \end{aligned}$$

Also ist $\underline{P}(A^{(1)}) = 0$ und $\overline{P}(A^{(1)}) = 1$. Nun könnte man denken, Zylindermengen seien ganz und gar ungeeignet, um $P(A^{(1)})$ zu bestimmen, doch immerhin gilt die Darstellung

$$\begin{aligned} A^{(1)} &= \Big\{\forall \epsilon > 0,\ \exists N \in \mathbb{N},\ \forall n \ge N,\ \big|\widehat{\theta}_n - \theta\big| \le \epsilon\Big\} \\ &= \Big\{\forall k \in \mathbb{N},\ \exists N \in \mathbb{N},\ \forall n \ge N,\ \big|\widehat{\theta}_n - \theta\big| \le 2^{-k}\Big\} \\ &= \bigcap_{k\in\mathbb{N}} \left(\bigcup_{N\in\mathbb{N}} \bigcap_{n\ge N} \underbrace{\Big\{\big|\widehat{\theta}_n - \theta\big| \le 2^{-k}\Big\}}_{\in \mathcal{Z}} \right). \end{aligned}$$

Das Ereignis $A^{(1)}$ lässt sich also durch *abzählbar viele* Mengenoperationen aus Zylindermengen aufbauen.

1.1.2 Maße und σ–Algebren

Maße

Ein wesentlicher Schritt ist nun, die endliche Additivität von Inhalten (Eigenschaft (I.2)) durch eine stärkere Eigenschaft, die sogenannte *σ–Additivität* zu ersetzen:

Definition 1.7 (Maß und Wahrscheinlichkeitsmaß). Sei $\mathcal{A}$ eine Algebra über einer Menge Ω. Eine Abbildung $M : \mathcal{A} \to [0, \infty]$ heißt *Maß auf $\mathcal{A}$*, wenn sie die folgenden zwei Eigenschaften erfüllt:
(I.1) $M(\emptyset) = 0$;
($\mathbf{I}_\sigma$.2) sind $A_1, A_2, A_3, \dots$ paarweise disjunkte Mengen in $\mathcal{A}$ mit $\bigcup_n A_n \in \mathcal{A}$, dann ist

$$M\Big(\bigcup_n A_n\Big) \;=\; \sum_n M(A_n).$$

Im Falle von $M(\Omega) = 1$ nennt man M ein *Wahrscheinlichkeitsmaß auf $\mathcal{A}$*.

Zwei einfache Beispiele für Maße auf $\mathcal{P}(\Omega)$ sind das Zählmaß,

$$A \;\mapsto\; \#A,$$

und diskrete Wahrscheinlichkeitsverteilungen,

$$A \;\mapsto\; \sum_{\omega \in A} p(\omega).$$

Anmerkung 1.8 (Bonferroni–Ungleichung für Maße). Ist M ein Maß auf der Algebra $\mathcal{A}$, dann gilt für beliebige Mengen A, A_1, A_2, $A_3, \dots$ aus $\mathcal{A}$ die Ungleichung

$$M(A) \;\le\; \sum_n M(A_n) \quad \text{falls } A \subset \bigcup_n A_n.$$

Anmerkung 1.9 (σ–Superadditivität von Inhalten). Sei M ein Inhalt auf der Algebra $\mathcal{A}$. Dann gilt für *paarweise disjunkte* Mengen $A_1, A_2, A_3, \dots \in \mathcal{A}$ mit $\bigcup_n A_n \in \mathcal{A}$ stets die Ungleichung

$$M\Big(\bigcup_n A_n\Big) \;\ge\; \sum_n M(A_n);$$

siehe Übungen. Es kann aber durchaus passieren, dass diese Ungleichung strikt ist.

Beispiel 1.10 Für eine unendliche Menge Ω sei $\mathcal{A}$ die Algebra aller Teilmengen von Ω mit $\#A < \infty$ oder $\#A^c < \infty$. Dann definiert

$$M(A) \;:=\; \begin{cases} 0 & \text{wenn } \#A < \infty, \\ 1 & \text{wenn } \#A^c < \infty \end{cases}$$

einen Wahrscheinlichkeitsinhalt auf $\mathcal{A}$. Bei M handelt es sich um ein Wahrscheinlichkeitsmaß genau dann, wenn Ω überabzählbar ist; siehe Übungen.

Anmerkung 1.11 (**σ–Stetigkeit von Maßen, I**). Ist M ein Inhalt auf $\mathcal{A}$, dann ist Eigenschaft (I_σ.2) äquivalent zu folgender Stetigkeitseigenschaft: Sind $B_1 \subset B_2 \subset B_3 \subset \cdots$ Mengen in $\mathcal{A}$ mit $\bigcup_n B_n \in \mathcal{A}$, dann ist

$$M\Big(\bigcup_n B_n\Big) \;=\; \lim_{n\to\infty} M(B_n).$$

Dies ergibt sich im Wesentlichen aus den folgenden Konstruktionen:
(i) Für Mengen $B_1 \subset B_2 \subset B_3 \subset \cdots$ definieren wir $A_n := B_n \setminus B_{n-1}$, wobei $B_0 := \emptyset$. Dann sind diese Mengen A_n paarweise disjunkt, und

$$\bigcup_{i=1}^n A_i \;=\; B_n, \quad \bigcup_n A_n \;=\; \bigcup_n B_n.$$

(ii) Für paarweise disjunkte Mengen $A_1, A_2, A_3, \ldots$ definieren wir $B_n := \bigcup_{i=1}^n A_i$. Dann ist $B_1 \subset B_2 \subset B_3 \subset \cdots$ und wieder ist

$$\bigcup_n A_n \;=\; \bigcup_n B_n.$$

Sowohl in (i) als auch in (ii) ist

$$\lim_{n\to\infty} M(B_n) \;=\; \lim_{n\to\infty} \sum_{i=1}^n M(A_i) \;=\; \sum_n M(A_n).$$

Anmerkung 1.12 (**σ–Stetigkeit von Maßen, II**). Sei M ein Maß auf $\mathcal{A}$, und seien $C_1 \supset C_2 \supset C_3 \supset \cdots$ Mengen aus $\mathcal{A}$ mit $M(C_1) < \infty$ sowie $\bigcap_n C_n \in \mathcal{A}$. Dann ist

$$M\Big(\bigcap_n C_n\Big) \;=\; \lim_{n\to\infty} M(C_n);$$

siehe Übungen.

Nun beweisen wir, dass die Inhalte für unsere beiden Schlüsselbeispiele sogar Maße sind:

Satz 1.13 *Der Volumeninhalt* $\mathrm{Vol}(\cdot)$ *von Elementarmengen ist ein Maß auf* $\mathcal{E}$.

Satz 1.14 *Jeder Wahrscheinlichkeitsinhalt auf den Zylindermengen von* $\{0,1\}^{\mathbb{N}}$ *ist ein Maß auf* $\mathcal{Z}$.

Beweis von Satz 1.13. Nach Anmerkung 1.9 genügt es, folgende Ungleichung zu beweisen: Für eine beliebige (nichtleere) Elementarmenge A und paarweise disjunkte Elementarmengen $A_1, A_2, A_3, \ldots$ mit $A = \bigcup_n A_n$ ist

$$M(A) \;\le\; \sum_n M(A_n).$$

Einerseits ist

$$\mathrm{Vol}(A) \;=\; \sup\big\{\mathrm{Vol}(K) : K \in \mathcal{E}, K \subset A, K \text{ kompakt}\big\}.$$

Daher genügt es sogar zu zeigen, dass

$$M(K) \;\leq\; \sum_n M(A_n) \quad \text{für jede kompakte Elementarmenge } K \subset A. \tag{1.4}$$

Nun wählen wir für ein beliebiges $\epsilon > 0$ und zu jedem $n \in \mathbb{N}$ eine *offene* Elementarmenge $\widetilde{A}_n$ mit

$$A_n \cap K \;\subset\; \widetilde{A}_n \quad \text{und} \quad \mathrm{Vol}(\widetilde{A}_n) \;\leq\; \mathrm{Vol}(A_n) + 2^{-n}\epsilon.$$

Da K in der Menge $\bigcup_n \widetilde{A}_n$ enthalten und kompakt ist, existiert ein $N \in \mathbb{N}$ mit $K \subset \bigcup_{n=1}^N \widetilde{A}_n$. Aus Anmerkung 1.5 ergibt sich nun, dass

$$\mathrm{Vol}(K) \;\leq\; \sum_{n=1}^N \mathrm{Vol}(\widetilde{A}_n) \;\leq\; \sum_n \mathrm{Vol}(A_n) + \epsilon.$$

Da $\epsilon > 0$ beliebig klein sein kann, ergibt sich die gewünschte Ungleichung (1.4). $\square$

Beweis von Satz 1.14. Zu zeigen ist, dass P die Stetigkeitseigenschaft in Anmerkung 1.11 erfüllt. Diese ergibt sich aber aus einer rein mengentheoretischen Aussage über Zylindermengen:

Sind $B_1 \subset B_2 \subset B_3 \subset \cdots$ Zylindermengen mit $B := \bigcup_n B_n \in \mathcal{Z}$, dann ist schon $B_n = B$ für ein $n \in \mathbb{N}$.

Betrachtet man die Mengen $D_n := B \setminus B_n$, dann ergibt sich letztere Aussage aus der folgenden Behauptung: Sind $D_1 \supset D_2 \supset D_3 \supset \cdots$ Zylindermengen mit $\bigcup_n D_n = \emptyset$, dann ist $D_n = \emptyset$ für ein $n \in \mathbb{N}$.

Letztere Behauptung beweisen wir nun indirekt: Sind $D_1 \supset D_2 \supset D_3 \supset \cdots$ *nichtleere* Zylindermengen, dann gibt es eine Folge $y = (y_k)_k \in \{0,1\}^{\mathbb{N}}$, die in allen Zylindermengen D_n liegt. (Insbesondere ist dann $\bigcap_n D_n \neq \emptyset$.) Um y induktiv zu konstruieren, schreiben wir

$$Z(y_1, \ldots, y_k) \;:=\; \big\{\omega : (\omega_1, \ldots, \omega_k) = (y_1, \ldots, y_k)\big\}$$

für beliebige $k \in \mathbb{N}$ und $y_1, \ldots, y_k \in \{0,1\}$.

Induktionsanfang: Da

$$D_n \;=\; D_n \cap Z(0) \;\cup\; D_n \cap Z(1) \;\neq\; \emptyset \quad \text{für alle } n \in \mathbb{N},$$

gibt es sicher ein $y_1 \in \{0,1\}$ mit

$$D_n \cap Z(y_1) \;\neq\; \emptyset \quad \text{für alle } n \in \mathbb{N}.$$

Induktionsschritt: Für ein $k \in \mathbb{N}$ seien $y_1, \ldots, y_k \in \{0,1\}$ so gewählt, dass $D_n \cap Z(y_1, \ldots, y_k) \neq \emptyset$ für beliebige $n \in \mathbb{N}$. Dann gibt es aber auch ein $y_{k+1} \in \{0,1\}$ mit der Eigenschaft, dass

$$D_n \cap Z(y_1, \ldots, y_k, y_{k+1}) \;\neq\; \emptyset \quad \text{für alle } n \in \mathbb{N}.$$

So erhalten wir induktiv eine Folge $y = (y_k)_{k=1}^{\infty}$ mit der Eigenschaft, dass

$$D_n \cap Z(y_1, \ldots, y_k) \neq \emptyset \quad \text{für beliebige } n, k \in \mathbb{N}. \tag{1.5}$$

Bisher haben wir nur ausgenutzt, dass $D_1 \supset D_2 \supset D_3 \supset \cdots$ und $D_n \neq \emptyset$ für alle $n \in \mathbb{N}$. Schreibt man aber

$$D_n = \{\omega : (\omega_1, \ldots, \omega_{k(n)}) \in C_n\}$$

mit gewissen natürlichen Zahlen $k(n)$ und Mengen $C_n \subset \{0,1\}^{k(n)}$, dann ergibt sich aus (1.5) mit $k = k(n)$, dass $(y_1, \ldots, y_{k(n)}) \in C_n$, also $y \in D_n$ für beliebige $n \in \mathbb{N}$. □

σ–Algebren

Letztendlich möchten wir auch die bisher eingeführten Mengenalgebren zu deutlich größeren Mengensystemen erweitern.

Definition 1.15 (**σ–Algebra**). Sei Ω eine Menge. Eine Mengenfamilie $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt *σ–Algebra über* Ω, wenn die folgenden drei Eigenschaften erfüllt sind:
(A.1) $\Omega \in \mathcal{A}$;
(A.2) mit $A \in \mathcal{A}$ ist auch $A^c \in \mathcal{A}$;
(A$_\sigma$.3) mit $A_1, A_2, A_3, \ldots \in \mathcal{A}$ ist auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Anmerkung 1.16 Jede σ–Algebra ist auch eine Algebra. Umgekehrt sei $\mathcal{A}$ eine Algebra über Ω. Dann sind Eigenschaft (A$_\sigma$.3) und die beiden folgenden Eigenschaften äquivalent:

(A$_\sigma$.3') Für Mengen $B_1 \subset B_2 \subset B_3 \subset \cdots$ aus $\mathcal{A}$ ist auch $\bigcup_n B_n \in \mathcal{A}$;
(A$_\sigma$.3'') Für Mengen $C_1 \supset C_2 \supset C_3 \supset \cdots$ aus $\mathcal{A}$ ist auch $\bigcap_n C_n \in \mathcal{A}$;

Einfache Beispiele für σ–Algebren werden wir in den Übungen kennenlernen.

1.1.3 Die Maßfortsetzung

Wir betrachten nun ein Maß M auf einer Algebra $\mathcal{A}_o$ über eine Menge Ω und definieren mit dessen Hilfe eine Abbildung $M^* : \mathcal{P}(\Omega) \to [0, \infty]$ wie folgt: Für $S \subset \Omega$ sei

$$M^*(S) := \inf\Big\{\sum_n M(A_n) : A_1, A_2, A_3, \ldots \in \mathcal{A}_o \text{ mit } S \subset \bigcup_n A_n\Big\}. \tag{1.6}$$

Diese Definition ist ähnlich zur Definition von $\overline{M}(S)$ bei der Eudoxos–Fortsetzung. Der wesentliche Unterschied ist, dass wir jetzt S mit abzählbar unendlich vielen Mengen aus $\mathcal{A}_o$ überdecken. Dies führt im Allgemeinen zu einer Zahl $M^*(S) \leq \overline{M}(S)$.

Satz 1.17 ***(Carathéodory, I)***. *Sei M^* die in (1.6) definierte Funktion auf $\mathcal{P}(\Omega)$. Diese erfüllt die folgenden drei Eigenschaften:*
(M*.1) $M^*(\emptyset) = 0$;
(M*.2) $M^*(S) \leq M^*(S')$ *falls* $S \subset S' \subset \Omega$;
(M*.3) $M^*\Big(\bigcup_n S_n\Big) \leq \sum_n M^*(S_n)$ *für beliebige Mengen* $S_1, S_2, S_3, \ldots \subset \Omega$.

Ferner ist
(M*.4) $M^* = M$ *auf* $\mathcal{A}_o$, *und*
(M*.5) $M^*(S) = M(S \cap A_o) + M^*(S \setminus A_o)$ *für beliebige* $S \subset \Omega$ *und* $A_o \in \mathcal{A}_o$.

Anmerkung 1.18 (Äußere Maße). Eine beliebige Abbildung $M^* : \mathcal{P}(\Omega) \to [0, \infty]$ mit den Eigenschaften (M*.1–3) nennt man ein *äußeres Maß* auf $\mathcal{P}(\Omega)$. Die spezielle Konstruktion (1.6) ist nur eine von vielen Möglichkeiten, ein äußeres Maß zu definieren.

Schränkt man den Definitionsbereich des äußeren Maßes M^* geeignet ein, ergibt sich tatsächlich ein Maß. Dahinter steht ein allgemeiner Satz:

Satz 1.19 ***(Carathéodory, II)***. *Sei $M^* : \mathcal{P}(\Omega) \to [0, \infty]$ eine Abbildung mit den Eigenschaften (M*.1–3). Dann ist die Mengenfamilie*

$$\mathcal{A}(M^*) \;:=\; \Big\{A \subset \Omega : M^*(S) = M^*(S \cap A) + M^*(S \setminus A) \text{ für beliebige } S \subset \Omega\Big\}$$

eine σ–Algebra über Ω, und M^* ist ein Maß auf $\mathcal{A}(M^*)$.

Die Definition von $\mathcal{A}(M^*)$ in Satz 1.19 kann man wie folgt verstehen: Zerlegt man mit Hilfe von $A \subset \Omega$ eine beliebige Menge $S \subset \Omega$ in die beiden Teile $S \cap A$ und $S \setminus A$, so gilt im Allgemeinen nur die Ungleichung

$$M^*(S) \;\leq\; M^*(S \cap A) + M^*(S \setminus A).$$

Die Familie $\mathcal{A}(M^*)$ besteht also aus allen Mengen A, welche eine beliebige Menge S "glatt in zwei Teile zerlegen" in dem Sinne, dass obige Ungleichung eine Gleichung ist.

Aus den Sätzen (1.19) und (1.17) ergibt sich nun die gewünschte Maßfortsetzung für unsere beiden Schlüsselbeispiele:

Korollar 1.20 *Das Maß $\mathrm{Vol}(\cdot)$ auf den Elementarmengen im $\mathbb{R}^d$ lässt sich zu einem Maß auf einer σ–Algebra $\mathcal{A} \supset \mathcal{E}$ über $\mathbb{R}^d$ fortsetzen.*

Korollar 1.21 *Jeder Wahrscheinlichkeitsinhalt P auf den Zylindermengen in $\{0,1\}^{\mathbb{N}}$ lässt sich zu einem Wahrscheinlichkeitsmaß auf einer σ–Algebra $\mathcal{A} \supset \mathcal{Z}$ forsetzen.*

Das in Korollar 1.20 auftretende Maß nennt man das *d–dimensionale Lebesguemaß* oder das *Lebesguemaß auf $\mathbb{R}^d$*.

Beweis von Satz 1.17. Die Eigenschaften (M*.1–2) ergeben sich direkt aus der Definition (1.6). Um (M*.3) nachzuweisen, können wir uns auf den Fall, dass $M^*(S_n) < \infty$ für alle $n \in \mathbb{N}$, einschränken. Für ein beliebiges $\epsilon > 0$ und $n \in \mathbb{N}$ seien $A_{n,1}, A_{n,2}, A_{n,3}, \ldots \in \mathcal{A}_o$ mit

$$S_n \subset \bigcup_k A_{n,k} \quad \text{und} \quad \sum_k M(A_{n,k}) \leq M(S_n) + 2^{-n}\epsilon.$$

Doch dann ist

$$\bigcup_n S_n \subset \bigcup_{n,k} A_{n,k} \quad \text{und} \quad M^*\Big(\bigcup_n S_n\Big) \leq \sum_{n,k} M(A_{n,k}) \leq \sum_n M^*(S_n) + \epsilon.$$

Da $\epsilon > 0$ beliebig klein ist, folgt hieraus (M*.3).

In der Definition (1.6) von $M^*(S)$ kann man ohne Einschränkung verlangen, dass die Mengen $A_n \in \mathcal{A}_o$ paarweise disjunkt sind. Wenn außerdem $S \in \mathcal{A}_o$, darf man ohne Einschränkung verlangen, dass alle $A_n \subset S$. Doch dann ist $S = \bigcup_n A_n$ und $M(S) = \sum_n M(A_n)$, da M ein Maß auf $\mathcal{A}_o$ ist. Dies beweist (M*.4).

Zu zeigen bleibt Aussage (M*.5). Wegen (M*.1) und (M*.3) ist nur zu zeigen, dass $M^*(S) \geq M^*(S \cap A_o) + M^*(S \setminus A_o)$ für beliebige Mengen $S \subset \Omega$ und $A_o \in \mathcal{A}_o$. Doch für beliebige Mengen $A_1, A_2, A_3, \ldots \in \mathcal{A}_o$ mit $S \subset \bigcup_n A_n$ ist auch

$$S \cap A_o \subset \bigcup_n \underbrace{(A_n \cap A_o)}_{\in \mathcal{A}_o} \quad \text{und} \quad S \setminus A_o \subset \bigcup_n \underbrace{(A_n \setminus A_o)}_{\in \mathcal{A}_o},$$

so dass

$$\sum_n M(A_n) = \sum_n M(A_n \cap A_o) + \sum_n M(A_n \setminus A_o) \geq M^*(S \cap A_o) + M^*(S \setminus A_o).$$

Da die linke Seite beliebig nahe an $M^*(S)$ kommt, ergibt sich hieraus (M*.5). □

Beweis von Satz 1.19. Wir beweisen zunächst, dass $\mathcal{A}(M^*)$ eine σ–Algebra ist. Aus Eigenschaft (M*.1) und der Definition von $\mathcal{A}(M^*)$ folgt direkt, dass $\emptyset, \Omega \in \mathcal{A}(M^*)$, und dass mit $A \in \mathcal{A}(M^*)$ auch die Komplementärmenge A^c zu $\mathcal{A}(M^*)$ gehört. Zu zeigen bleibt, dass für beliebige Mengen $A_1, A_2, A_3, \ldots \in \mathcal{A}(M^*)$ auch die Vereinigungsmenge $A := \bigcup_n A_n$ zu $\mathcal{A}(M^*)$ gehört. Zu diesem Zweck betrachten wir die paarweise disjunktion Mengen $B_1 := A_1$ und $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i = \bigcap_{i=1}^{n-1} A_i^c \cap A_n$ für $n \geq 2$. Also ist $\bigcup_{n=1}^N A_n = \bigcup_{n=1}^N B_n$ für beliebige $N \in \mathbb{N}$, und $A = \bigcup_n B_n$. Für beliebige Mengen $S \subset \Omega$ ist

$$\begin{aligned}
M^*(S) &\geq M^*(S \cap A_1) + M^*(S \cap A_1^c) \\
&\geq M^*(S \cap A_1) + M^*(S \cap A_1^c \cap A_2) + M^*(S \cap A_1^c \cap A_2^c) \\
&= M^*(S \cap B_1) + M^*(S \cap B_2) + M^*(S \cap A_1^c \cap A_2^c).
\end{aligned}$$

Induktiv kann man zeigen, dass für beliebige $N \in \mathbb{N}$ gilt:

$$\begin{aligned} M^*(S) &\geq \sum_{n=1}^{N} M^*(S \cap B_n) + M^*\Big(S \cap \bigcap_{n=1}^{N} A_n^c\Big) \\ &\geq \sum_{n=1}^{N} M^*(S \cap B_n) + M^*(S \setminus A), \end{aligned}$$

denn $S \cap \bigcap_{n=1}^N A_n^c = S \setminus \bigcup_{n=1}^N A_n \supset S \setminus A$. Für $N \to \infty$ erhalten wir dann die Ungleichungen

$$\begin{aligned} M^*(S) &\geq \sum_n M^*(S \cap B_n) + M^*(S \setminus A) \\ &\geq M^*\Big(\bigcup_n S \cap B_n\Big) + M^*(S \setminus A) \;=\; M^*(S \cap A) + M^*(S \setminus A). \end{aligned}$$

Da ohnehin $M^*(S) \leq M^*(S \cap A) + M^*(S \setminus A)$, ist auch $A \in \mathcal{A}(M^*)$.

Nun zeigen wir noch, dass M^* ein Maß auf $\mathcal{A}(M^*)$ ist. Zu diesem Zweck nehmen wir an, dass die obigen Mengen $A_1, A_2, A_3, \ldots \in /AA(M^*)$ paarweise disjunkt sind, und setzen $S := A$. Dann ist $B_n = A_n$, und aus den vorangehenden Überlegungen folgt, dass

$$M^*(A) \;\geq\; \sum_n M^*(A \cap B_n) + M^*(A \setminus A) \;=\; \sum_n M^*(A_n) \;\geq\; M^*(A),$$

also $M^*(A) = \sum_n M^*(A_n)$. □

1.1.4 Eindeutigkeit von Maßen

Jetzt haben wir unser anfängliches Ziel im Wesentlichen erreicht. Die Frage ist nur, ob unsere spezielle Fortsetzung von Maßen nur eine von vielen möglichen Fortsetzungen liefert. Damit verwandt ist die Frage, inwiefern ein Maß M auf einer (σ–) Algebra bereits durch die Werte auf einer kleineren Teilfamilie festgelegt wird.

Minimale (σ–) Algebren und erzeugende Systeme. Um die vorangehenden Fragen präzise zu beantworten, benötigen wir noch eine Vorbetrachtung: Bildet man den Durchschnitt von beliebig vielen (σ–) Algebren über einer Menge Ω, dann ist auch dieser Durschschnitt eine (σ–) Algebra. Betrachten wir also eine beliebige Familie $\mathcal{D}$ von Teilmengen von Ω, dann ist

$$\sigma(\mathcal{D}) \;:=\; \bigcap \Big\{\mathcal{A} : \mathcal{A} \text{ eine } \sigma\text{–Algebra über } \Omega \text{ mit } \mathcal{D} \subset \mathcal{A}\Big\}$$

die kleinste σ–Algebra über Ω, welche $\mathcal{D}$ enthält. Man spricht auch von der *von $\mathcal{D}$ erzeugten σ–Algebra.*

Lässt sich eine beliebige σ–Algebra $\mathcal{A}$ über Ω darstellen als $\sigma(\mathcal{D})$ mit einer geeigneten Mengenfamilie $\mathcal{D} \subset \mathcal{P}(\Omega)$, so nennt man $\mathcal{D}$ ein *erzeugendes System von $\mathcal{A}$.*

Die kleinste *Algebra*, welche $\mathcal{D}$ enthält, kann man explizit darstellen durch

$$\Big\{\bigcup_{i=1}^{m}\bigcap_{j=1}^{n} D_{i,j} : m,n \in \mathbb{N}; D_{i,j} \in \mathcal{D} \cup \{\Omega\} \text{ oder } D^c \in \mathcal{D} \cup \{\Omega\}\Big\}. \tag{1.7}$$

Für $\sigma(\mathcal{D})$ gibt es in der Regel keine analoge Darstellung. Selbst relativ einfache Mengensysteme $\mathcal{D}$ führen oft zu derart komplexen Mengenfamilien $\sigma(\mathcal{D})$, dass es schwierig ist, eine Menge zu beschreiben, die *nicht* in $\sigma(\mathcal{D})$ liegt. Eine Ausnahme ist der Fall eines endlichen Mengensystems $\mathcal{D}$; siehe Übungen.

Beispiel 1.22 Sei $\Omega = \{1,2,3,4\}$, und sei $\mathcal{D} := \big\{\{1,2\},\{2,3\}\big\}$. Die kleinste Algebra, welche $\mathcal{D}$ enthält, ist gleich $\mathcal{P}(\Omega)$, und insbesondere ist $\sigma(\mathcal{D}) = \mathcal{P}(\Omega)$. Denn

$$\begin{aligned} \{1\} &= \{1,2\} \setminus \{2,3\}, & \{2\} &= \{1,2\} \cap \{2,3\}, \\ \{3\} &= \{2,3\} \setminus \{1,2\}, & \{4\} &= \Omega \setminus (\{1,2\} \cup \{2,3\}) \ \ldots \end{aligned}$$

Beispiel 1.23 Sei Ω eine beliebige Menge, und sei $\mathcal{D}$ die Menge aller einpunktigen Mengen $\{\omega\}$, $\omega \in \Omega$. Die von $\mathcal{D}$ erzeugte σ–Algebra ist gleich

$$\big\{A \subset \Omega : A \text{ oder } A^c \text{ ist abzählbar}\big\}.$$

Ist Ω selbst abzählbar, dann ist $\sigma(\mathcal{D}) = \mathcal{P}(\Omega)$.

Nun kommen wir zu dem versprochenen Eindeutigkeitssatz.

Satz 1.24 ***(Dynkin).*** *Sei $\mathcal{D} \subset \mathcal{P}(\Omega)$ ein $\cap$–stabiles Mengensystem; das heißt, mit $D, E \in \mathcal{D}$ ist auch $D \cap E \in \mathcal{D}$. Sind M_1 und M_2 zwei Maße auf $\sigma(\mathcal{D})$ mit*

$$M_1(\Omega) = M_2(\Omega) < \infty \quad \textit{und} \quad M_1(D) = M_2(D) \textit{ für alle } D \in \mathcal{D},$$

dann ist $M_1 = M_2$.

Dass man auf die Durchschnittstabilität von $\mathcal{D}$ nicht verzichten kann, zeigt folgendes Beispiel:

Beispiel 1.25 Seien Ω und $\mathcal{D}$ wie in Beispiel 1.22. Dann definieren

$$M_1(A) \ := \ \frac{\#A}{4} \quad \text{und} \quad M_2(A) \ := \ \frac{\#(A \cap \{1,3\})}{2}$$

zwei verschiedene Wahrscheinlichkeitsmaße auf $\mathcal{P}(\Omega)$, obwohl $M_1(D) = M_2(D) = 1/2$ für alle $D \in \mathcal{D}$.

Satz 1.24 wird in der Regel mit sogenannten *Dynkin–Systemen* bewiesen. Wir präsentieren in der Vorlesung einen eher konstruktiven Beweis aus zwei Hilfssätzen, von denen der zweite auch an anderer Stelle von Nutzen ist.

Hilfssatz 1.26 *Sei $\mathcal{D} \subset \mathcal{P}(\Omega)$ ein durchschnittstabiles Mengensystem, und seien M_1, M_2 Inhalte auf der kleinsten Algebra $\mathcal{A}_o$, welche $\mathcal{D}$ enthält. Dann folgt aus*

$$M_1(\Omega) = M_2(\Omega) < \infty \quad \textit{und} \quad M_1(D) = M_2(D) \textit{ für alle } D \in \mathcal{D},$$

dass $M_1 = M_2$ auf $\mathcal{A}_o$.

Hilfssatz 1.27 *Sei $\mathcal{A}_o$ eine Algebra über Ω, und sei M ein Maß auf $\sigma(\mathcal{A}_o)$ mit $M(\Omega) < \infty$. Dann existiert zu jedem $A \in \sigma(\mathcal{A}_o)$ und beliebigem $\epsilon > 0$ eine Menge $A_o \in \mathcal{A}_o$ mit*

$$M(A \triangle A_o) \ \leq \ \epsilon.$$

Beweis von Hilfssatz 1.26. Ohne Einschränkung der Allgemeinheit seien $\emptyset, \Omega \in \mathcal{D}$. Dann ergibt sich aus der $\cap$–Stabilität von $\mathcal{D}$ und (1.7) die konkrete Darstellung

$$\mathcal{A}_o \ = \ \Big\{ \bigcup_{i=1}^m D_i \cap \bigcap_{j=1}^n E_{ij}^c : m, n \in \mathbb{N}; D_i, E_{ij} \in \mathcal{D} \Big\}.$$

Für $s = 1, 2$ und eine Menge wie in dieser Beschreibung von $\mathcal{A}_o$ ergibt sich aus der Siebformel von Sylvester-Poincaré, dass

$$\begin{aligned} M_s\Big(\bigcup_{i=1}^m D_i \cap \bigcap_{j=1}^n E_{ij}^c \Big) \ &= \ \sum_{\emptyset \neq \mathcal{K} \subset \{1,\ldots,m\}} (-1)^{\#\mathcal{K}-1} M_s\Big(\bigcap_{i \in \mathcal{K}} \Big(D_i \cap \bigcap_{j=1}^n E_{ij}^c \Big) \Big) \\ &= \ \sum_{\emptyset \neq \mathcal{K} \subset \{1,\ldots,m\}} (-1)^{\#\mathcal{K}-1} M_s\Big(D_{\mathcal{K}} \cap \bigcap_{i \in \mathcal{K}, j \in \{1,\ldots,n\}} E_{ij}^c \Big) \end{aligned}$$

mit $D_{\mathcal{K}} := \bigcap_{i \in \mathcal{K}} D_i \in \mathcal{D}$. Zu zeigen bleibt also, dass $M_1(A) = M_2(A)$ für beliebige Mengen der Form

$$A \ = \ D \cap \bigcap_{j=1}^n E_j^c \quad (n \in \mathbb{N}; D, E \in \mathcal{D}).$$

Doch auch hier kann man die Siebformel einsetzen:

$$\begin{aligned} M_s\Big(D \cap \bigcap_{j=1}^n E_j^c \Big) \ &= \ M_s(D) - M_s\Big(D \setminus \Big(\bigcap_{j=1}^n E_j^c \Big) \Big) \\ &= \ M_s(D) - M_s\Big(D \cap \Big(\bigcup_{j=1}^n E_j \Big) \Big) \\ &= \ M_s(D) - M_s\Big(\bigcup_{j=1}^n D \cap E_j \Big) \\ &= \ M_s(D) - \sum_{\emptyset \mathcal{K} \subset \{1,\ldots,n\}} (-1)^{\#\mathcal{K}-1} M_s(D_{\mathcal{K}}) \end{aligned}$$

mit $D_{\mathcal{K}} := \bigcap_{j \in \mathcal{K}} D \cap E_j \in \mathcal{D}$. $\square$

Beweisidee von Hilfssatz 1.27. Sei $\mathcal{A}_*$ die Familie aller $A \in \mathcal{A}$ mit der Eigenschaft, dass zu jedem $\epsilon > 0$ ein $A_o \in \mathcal{A}_o$ existiert, so dass $M(A \triangle A_o) \leq \epsilon$. Nun zeigt man, dass $\mathcal{A}_*$ eine σ–Algebra ist, welche $\mathcal{A}_o$ enthält; siehe Übungen. □

Beweis von Satz 1.24. Aus den Voraussetzungen des Satzes und Hilfssatz 1.26 folgt, dass M_1 und M_2 auf der kleinsten Algebra $\mathcal{A}_o$m welche $\mathcal{D}$ enthält, übereinstimmen. Da $\mathcal{D} \subset \mathcal{A}_o \subset \sigma(\mathcal{D})$, ist $\sigma(\mathcal{D}) = \sigma(\mathcal{A}_o)$. Nun betrachten wir das Maß $M := M_1 + M_2$. Dieses erfüllt die Voraussetzungen von Hilfssatz 1.27, so dass zu jedem $A \in \sigma(\mathcal{D})$ und beliebigem $\epsilon > 0$ eine Menge $A_o \in \mathcal{A}_o$ existiert, so dass

$$M(A \triangle A_o) \;\leq\; \epsilon.$$

Insbesondere ist

$$\begin{aligned} |M_1(A) - M_2(A)| \;&=\; \big|(M_1(A) - M_1(A_o)) - (M_2(A) - M_2(A_o))\big| \\ &\leq\; |M_1(A) - M_1(A_o)| + |M_2(A) - M_2(A_o)| \\ &\leq\; M_1(A \triangle A_o) + M_2(A \triangle A_o) \;=\; M(A \triangle A_o) \;\leq\; \epsilon. \end{aligned}$$

Für $\epsilon \downarrow 0$ ergibt sich, dass $M_1(A) = M_2(A)$. □

Schlussbetrachtungen

Wir möchten nun die vorangehenden Überlegungen zusammenfassen und noch etwas abrunden. Um für unsere Schlüsselbeispiele Maße mit den gewünschten Eigenschaften zu konstruieren, gingen wir wie folgt vor:

Schritt 1: Wir definierten eine Algebra $\mathcal{A}_o$ über Ω und darauf einen Inhalt M.

Schritt 2: Wir zeigten, dass dieser Inhalt sogar ein Maß auf $\mathcal{A}_o$ ist.

Schritt 3: Dann definierten wir mit Hilfe dieses Maßes M auf $\mathcal{A}_o$ ein äußeres Maß M^* auf $\mathcal{P}(\Omega)$.

Schritt 4: Aus den zwei Sätzen von Carathéodory folgt nun, dass $M^* = M$ auf $\mathcal{A}_o$, und dass M^* ein Maß auf einer σ–Algebra $\mathcal{A}(M^*)$ ist, welche $\mathcal{A}_o$ enthält.

Schritt 5: Aus dem Satz von Dynkin folgt, dass die so gewonnene Fortsetzung von M zu einem Maß auf $\sigma(\mathcal{A}_o)$ eindeutig ist.

Jetzt bleiben noch zwei Fragen:

- Ist $\mathcal{A}(M^*)$ wesentlich größer als $\sigma(\mathcal{A}_o)$?
- Ist die Algebra $\overline{\mathcal{A}_o}$ aus der Eudoxos–Fortsetzung in $\sigma(\mathcal{A}_o)$ oder $\mathcal{A}(M^*)$ enthalten?

Beide Fragen können wir unter der Voraussetzung

$$M(\Omega) \;<\; \infty$$

leicht beantworten. (Es würde sogar eine schwächere Eigenschaft, die sogenannte "σ–Endlichkeit" von M genügen, aber dies sollten Ihnen die Analytiker erklären.)

Satz 1.28 *Sei M^* das aus einem Maß M auf einer Algebra $\mathcal{A}_o$ konstruierte äußere Maß, wobei $M(\Omega) < \infty$. Dann ist*

$$\mathcal{A}(M^*) \;=\; \big\{A \cup N : A \in \sigma(\mathcal{A}_o), N \subset \Omega \text{ mit } M^*(N) = 0\big\}.$$

Ferner ist die aus der Eudoxos–Fortsetzung resultierende Algebra $\overline{\mathcal{A}_o}$ in $\mathcal{A}(M^*)$ enthalten, und $M^* = \overline{M}$ auf $\overline{\mathcal{A}_o}$.

Beweis von Satz 1.28. Ein wesentlicher Punkt ist eine vereinfachte Darstellung von $M^*(S)$: Nachdem man weiss, dass M^* ein Maß auf $\mathcal{A}(M^*) \supset \sigma(\mathcal{A}_o)$ ist, kann man einfach schreiben

$$M^*(S) \;=\; \min\big\{M^*(B) : B \in \sigma(\mathcal{A}_o), B \supset S\big\}.$$

Wählt man nämlich für $k \in \mathbb{N}$ Mengen $A_{k,1}, A_{k,2}, A_{k,3}, \ldots \in \mathcal{A}_o$ mit $S \subset \bigcup_n A_{k,n}$ und $M^*(S) \leq \sum_n M(A_{k,n}) + 2^{-k}$, dann ist

$$B \;:=\; \bigcap_k \Big(\bigcup_n A_{k,n}\Big)$$

offensichtlich eine Menge aus $\sigma(\mathcal{A}_o)$, welche S enthält, und $M^*(S) = M^*(B)$.

Im Falle von $M^*(\Omega) < \infty$ kann man nun zeigen, dass zu jeder Menge $A \in \mathcal{A}(M^*)$ zwei Mengen $A', A'' \in \sigma(\mathcal{A}_o)$ existieren, so dass

$$A' \;\subset\; A \;\subset\; A'' \quad \text{und} \quad M^*(A'' \setminus A') \;=\; 0.$$

Daher ist $A = A' \cup (A \setminus A')$, und $M^*(A \setminus A') \leq M^*(A'' \setminus A') = 0$.

Umgekehrt gehört jede Menge der Form $B \cup N$ mit $B \in \sigma(\mathcal{A}_o)$ und $M^*(N) = 0$ zu $\mathcal{A}(M^*)$. Denn für jede Menge $S \subset \Omega$ ist

$$\begin{aligned} M^*(S) &= M^*(S \cap B) + M^*(S \setminus B) \\ &= M^*(S \cap B) + M^*(S \cap N) + M^*(S \setminus B) \\ &\geq M^*(S \cap (B \cup N)) + M(S \setminus (B \cup N)). \end{aligned}$$

Den Beweis der Aussagen über $\overline{\mathcal{A}_o}$ und $\overline{M}$ überlassen wir dem Leser als Übungsaufgabe. □

1.1.5 Maß- und Wahrscheinlichkeitsräume, Messbarkeit

Definition 1.29 (Messbarer Raum, Maß- und Wahrscheinlichkeitsraum).

(a) Ein Paar $(\Omega, \mathcal{A})$, bestehend aus einer Menge Ω und einer σ–Algebra $\mathcal{A}$ über Ω heißt *messbarer Raum.*

(b) Ein Tripel $(\Omega, \mathcal{A}, M)$, bestehend aus einer Menge Ω, einer σ–Algebra $\mathcal{A}$ über Ω und einem Maß M auf $\mathcal{A}$ heißt *Maßraum*.

(c) Ein Tripel $(\Omega, \mathcal{A}, P)$, bestehend aus einer Menge Ω, einer σ–Algebra $\mathcal{A}$ über Ω und einem Wahrscheinlichkeitsmaß P auf $\mathcal{A}$ heißt *Wahrscheinlichkeitsraum*. Die Elemente von $\mathcal{A}$ nennt man auch *Ereignisse*.

Nun betrachten wir Abbildungen zwischen messbaren Räumen:

Definition 1.30 (Messbare Abbildung, Zufallsvariable und ihre Verteilung). Seien $(\Omega, \mathcal{A})$ und $(\mathcal{X}, \mathcal{B})$ messbare Räume und $X : \Omega \to \mathcal{X}$ eine beliebige Abbildung.

(a) Die Abbildung X heißt *$\mathcal{A}$–$\mathcal{B}$–messbar*, wenn

$$X^{-1}(B) \in \mathcal{A} \quad \text{für beliebige } B \in \mathcal{B}.$$

Wenn aus dem Kontext klar ist, mit welchen σ–Algebren die Mengen Ω und $\mathcal{X}$ versehen sind, spricht man einfach von einer *messbaren Abbildung* $X : \Omega \to \mathcal{X}$. Insbesondere wird eine abzählbare Menge in der Regel mit der σ–Algebra aller ihrer Teilmengen versehen.

(b) Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Eine $\mathcal{A}$–$\mathcal{B}$–messbare Abbildung $X : \Omega \to \mathcal{X}$ nennt man eine *Zufallsvariable auf* $(\Omega, \mathcal{A}, P)$ *mit Werten in* $(\mathcal{X}, \mathcal{B})$. Wenn aus dem Kontext klar ist, mit welcher σ–Algebra die Menge $\mathcal{X}$ versehen ist, spricht man einfach von einer *Zufallsvariable X mit Werten in $\mathcal{X}$*. Eine solche Zufallsvariable X induziert ein neues Wahrscheinlichkeitsmaß P^X auf $\mathcal{B}$, nämlich

$$P^X(B) := P(X^{-1}(B)) \quad \text{für } B \in \mathcal{B}.$$

Man nennt P^X die *Verteilung der Zufallsvariable X*.

Das Kriterium für Messbarkeit einer Abbildung sieht auf den ersten Blick unhandlich aus. Aber mit Hilfe der folgenden Überlegung wird dies oft vereinfacht:

Anmerkung 1.31 (Zum Nachweis der Messbarkeit). Für eine belieige Abbildung $X : \Omega \to \mathcal{X}$ ist

$$\{C \subset \mathcal{X} : X^{-1}(C) \in \mathcal{A}\}$$

eine σ–Algebra über $\mathcal{X}$. Wenn nun $\mathcal{B} = \sigma(\mathcal{E})$ für eine Mengenfamilie $\mathcal{E} \subset \mathcal{P}(\mathcal{X})$, und wenn $X^{-1}(E) \in \mathcal{A}$ für alle $E \in \mathcal{E}$, dann gilt dies für beliebige Mengen aus $\mathcal{B}$, und X ist eine $\mathcal{A}$–$\mathcal{B}$–messbare Abbildung.

Beispiel 1.32 (Vom unendlichen Münzwurf zur uniformen Verteilung auf $[0,1]$). Wir gehen aus von unserem Schlüsselbeispiel 2 mit $\Omega = \{0,1\}^{\mathbb{N}}$ und einem Wahrscheinlichkeitsmaß P auf

$\sigma(\mathcal{Z})$. Nun definieren wir die Abbildung

$$U : \Omega \to [0,1], \quad U(\omega) := \sum_{n=1}^{\infty} 2^{-n}\omega_n.$$

Definiert man $\mathcal{B}$ als die kleinste σ–Algebra über $[0,1]$, welche alle Intervalle in $[0,1]$ enthält, dann ist U tatsächlich $\sigma(\mathcal{Z})$–$\mathcal{B}$–messbar. Denn man kann sich schnell davon überzeugen, dass

$$\mathcal{B} = \sigma\big(\{[0,r] : r \in [0,1)\}\big).$$

Schreibt man nun $r \in [0,1)$ als $r = \sum_{n=1}^{\infty} 2^{-n}\rho_n$ mit Ziffern $\rho_n \in \{0,1\}$, von denen unendlich viele gleich Null sind, dann ist $U^{-1}([0,r]) = \{\omega \in \Omega : U(\omega) \le r\}$ gleich

$$\bigcup_{n\in\mathbb{N}} \big\{\omega \in \Omega : \omega_i = \rho_i \text{ für alle } i < n,\, \omega_n < \rho_n\big\} \cup \big\{\omega : \omega_n = \rho_n \text{ für alle } n \in \mathbb{N}\big\}.$$

Nun betrachten wir den Spezialfall, dass P aus den Gewichten

$$p_n(y) := 2^{-n} \quad \text{für beliebige } n \in \mathbb{N} \text{ und } y \in \{0,1\}^n$$

entsteht, also das unendlich ofte Werfen einer perfekten Münze. Mit $X_i(\omega) := \omega_i$ ist dann

$$\begin{aligned}
P^U([0,r]) &= \sum_{n\in\mathbb{N}} P\big(X_i = \rho_i \text{ für alle } i < n,\, X_n < \rho_n\big) + P\big(X_n = \rho_n \text{ für alle } n \in \mathbb{N}\big) \\
&= \sum_{n\in\mathbb{N}} 2^{-n}\rho_n + 0 \\
&= r.
\end{aligned}$$

Somit erhalten wir eine Zufallsvariable U mit Werten in $([0,1],\mathcal{B})$, die auf $[0,1]$ uniform verteilt ist, das heißt,

$$P^U(J) = \text{Länge}(J) \quad \text{für beliebige Intervalle } J \subset [0,1].$$

Man kann dieses Beispiel auch in der umgekehrten Richtung betrachten. Wir betrachten den Wahrscheinlichkeitsraum $\big((0,1],\mathcal{B},P\big)$ mit der kleinsten σ–Algebra $\mathcal{B}$ über $(0,1]$, welche alle Intervalle $(r,s] \subset (0,1]$ enthält, und $P((r,s]) := s - r$. Für $u \in (0,1]$ gibt es eine dyadische Darstellung

$$u = \sum_{k=1}^{\infty} 2^{-k} X_k(u)$$

mit gewissen Ziffern $X_k(u) \in \{0,1\}$. Diese Darstellung ist eindeutig, wenn wir zusätzlich verlangen, dass $X_k(u) = 1$ für unendlich viele $k \in \mathbb{N}$. Die resultierende Abbildung $X = (X_k)_{k=1}^{\infty} : (0,1] \to \{0,1\}^{\mathbb{N}}$ ist $\mathcal{B}$–$\sigma(\mathcal{Z})$–messbar und liefert einen unendlichen Münzwurf mit perfekter Münze. Denn für jedes $n \in \mathbb{N}$ und beliebige $y_1, y_2, \ldots, y_n \in \{0,1\}$ ist

$$\big\{u \in (0,1] : X_k(u) = y_k \text{ für } k = 1,2,\ldots,n\big\} = (r, r + 2^{-n}] \quad \text{mit } r := \sum_{k=1}^{n} 2^{-k} y_k,$$

und dieses Ereignis hat Wahrscheinlichkeit 2^{-n}.

1.2 $\mathbb{R}^d$–wertige Zufallsvariablen und ihre Verteilungsfunktionen

1.2.1 Die Borelmengen im $\mathbb{R}^d$

Unser erstes Schlüsselbeispiel führt zu einem Maß auf der kleinsten σ–Algebra über $\mathbb{R}^d$, welche beliebige Rechtecke enthält. Diese σ–Algebra kann man noch auf andere Weisen charakterisieren:

Satz 1.33 *Für eine σ–Algebra $\mathcal{B}$ über $\mathbb{R}^d$ sind die folgenden Aussagen äquivalent:*

(B.1) *$\mathcal{B}$ enthält alle Mengen der Form*

$$(-\boldsymbol{\infty}, \boldsymbol{r}] \ := \ (-\infty, r_1] \times (-\infty, r_2] \times \cdots \times (-\infty, r_d] \quad \text{mit } \boldsymbol{r} = (r_i)_{i=1}^d \in \mathbb{R}^d.$$

(B.2) *$\mathcal{B}$ enthält beliebige Rechtecke im $\mathbb{R}^d$;*
(B.3) *$\mathcal{B}$ enthält alle offenen Mengen im $\mathbb{R}^d$;*
(B.4) *$\mathcal{B}$ enthält alle abgeschlossenen Mengen im $\mathbb{R}^d$.*

Definition 1.34 (Borelmengen im $\mathbb{R}^d$). Die *Borel–σ–Algebra über $\mathbb{R}^d$* ist die kleinste σ–Algebra über $\mathbb{R}^d$, welche die obigen Eigenschaften (B.1–4) erfüllt. Wir bezeichnen sie mit $\mathrm{Borel}(\mathbb{R}^d)$; ihre Elemente nennt man *Borelmengen im $\mathbb{R}^d$*.

Beweis von Satz 1.33. Dass (B.1) auch Eigenschaft (B.2) impliziert, wird in den Übungen gezeigt.

Angenommen $\mathcal{B}$ erfüllt Eigenschaft (B.2). Für eine beliebige offene Menge $U \subset \mathbb{R}^d$ und $\boldsymbol{y} \in U$ existiert ein $\delta > 0$, so dass $(\boldsymbol{y}-\delta, \boldsymbol{y}+\delta) := (y_1-\delta, y_1+\delta) \times \cdots \times (y_d-\delta, y_d+\delta)$ in U enthalten ist. Nun wählen wir einen Vektor $\widetilde{\boldsymbol{y}} \in (\boldsymbol{y}-\delta/3, \boldsymbol{y}+\delta/3) \cap \mathbb{Q}^d$ sowie ein $\widetilde{\delta} \in (\delta/3, 2\delta/3) \cap \mathbb{Q}$. Dann ist

$$\boldsymbol{y} \ \in \ (\widetilde{\boldsymbol{y}} - \widetilde{\delta}, \widetilde{\boldsymbol{y}} + \widetilde{\delta}) \ \subset \ (\boldsymbol{y}-\delta, \boldsymbol{y}+\delta) \ \subset \ U.$$

Diese Überlegung zeigt, dass U identisch ist mit der Vereinigung aller Rechtecke $(\widetilde{\boldsymbol{y}} - \widetilde{\delta}, \widetilde{\boldsymbol{y}} + \widetilde{\delta})$ derart, dass $\widetilde{\boldsymbol{y}} \in \mathbb{Q}^d$, $0 < \widetilde{\delta} \in \mathbb{Q}$ und $(\widetilde{\boldsymbol{y}} - \widetilde{\delta}, \widetilde{\boldsymbol{y}} + \widetilde{\delta}) \subset U$. Insbesondere ist U eine Vereinigung von abzählbar vielen Rechtecken, gehört also zu $\mathcal{B}$. Dies beweist Eigenschaft (B.3).

Dass die Eigenschaften (B.3–4) äquivalent sind, ergibt sich aus der Tatsache, dass Komplemente von offenen Mengen abgeschlossen und Komplemente von abgeschlossenen Mengen offen sind.

Da die speziellen Rechtecke in (B.1) abgeschlossen sind, ist (B.1) eine Folgerung aus (B.4). $\square$

Konventionen. Die Borelmengen im $\mathbb{R}^d$ bilden die Standard–σ–Algebra über $\mathbb{R}^d$. Ist irgendwann von einem "Maß auf dem $\mathbb{R}^d$" die Rede, so meint man eigentlich ein Maß auf $\mathrm{Borel}(\mathbb{R}^d)$.

Ähnliche Konventionen gelten für Abbildungen und Zufallsvariablen. Ist z.B. X eine Zufallsvariable auf einem gewissen Wahrscheinlichkeitsraum mit Werten in $(\mathbb{R}^d, \mathrm{Borel}(\mathbb{R}^d))$, dann spricht man vereinfachend von einer "Zufallsvariable X mit Werten in $\mathbb{R}^d$" oder einer "$\mathbb{R}^d$–wertigen Zufallsvariable".

Manche Autoren reservieren den Begriff "Zufallsvariable" für den Fall $d = 1$ und sprechen ansonsten von "Zufallsvektoren".

Mit Hilfe von Satz 1.33 und Anmerkung 1.31 kann man leicht zeigen, dass eine große Klasse von Abbildungen (Borel–) messbar ist:

Korollar 1.35 ***(Borel–Messbarkeit, I)***.
(a) *Jede stetige Funktion $X : \mathbb{R}^d \to \mathbb{R}^q$ ist messbar.*
(b) *Jede monotone Funktion $X : \mathbb{R} \to \mathbb{R}$ ist messbar.*

Korollar 1.36 ***(Borel–Messbarkeit, II)***. *Sei $(\Omega, \mathcal{A})$ ein messbarer Raum, und für $n \in \mathbb{N}$ sei $X_n : \Omega \to [-\infty, \infty]$ eine $\mathcal{A}$–$\mathrm{Borel}([-\infty, \infty])$–messbare Funktion. Dann sind auch $\liminf_{n\to\infty} X_n$ und $\limsup_{n\to\infty} X_n$ messbare Abbildungen.*

Den Beweis von Korollar 1.35 stellen wir als Übungsaufgabe.

Beweis von Korollar 1.36. Ein erzeugendes System für $\mathrm{Borel}([-\infty, \infty])$ sind die "Halbgeraden" $[-\infty, r]$ mit $r \in \mathbb{R}$. Mit $X := \limsup_{n\to\infty} X_n$ ist

$$X(\omega) \le r \quad \text{g.d.w.} \quad \forall k \in \mathbb{N}, \exists m \in \mathbb{N}, \forall n \ge m,\ X_n(\omega) < r + 1/k.$$

Das heißt,

$$\{X \le r\} \;=\; \bigcap_{k\in\mathbb{N}} \Big(\bigcup_{m\in\mathbb{N}} \bigcap_{n\ge m} \underbrace{\{X_m < r + 1/k\}}_{\in\mathcal{A}} \Big) \;\in\; \mathcal{A}.$$

□

1.2.2 Verteilungsfunktionen

Definition 1.37 **(Verteilungsfunktion)**.
(a) Sei P ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$. Dann nennt man

$$F_P : \mathbb{R}^d \to [0,1], \quad F_P(\boldsymbol{r}) := P\big((-\boldsymbol{\infty}, \boldsymbol{r}]\big)$$

die Verteilungsfunktion von P.
(b) Sei $\boldsymbol{X}$ eine $\mathbb{R}^d$–wertige Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann nennt man

$$F_X : \mathbb{R}^d \to [0,1], \quad F_X(\boldsymbol{r}) := P\big(X \in (-\boldsymbol{\infty}, \boldsymbol{r}]\big)$$

die Verteilungsfunktion von X.

Aus dem Eindeutigkeitssatz 1.24 von Dynkin und der Charakterisierung (B.1) in Satz 1.33 ergibt sich die Tatsache, dass ein Wahrscheinlichkeitsmaß auf dem $\mathbb{R}^d$ durch seine Verteilungsfunktion F_P eindeutig festgelegt ist. Analog wird die Verteilung einer $\mathbb{R}^d$–wertigen Zufallsvariable X durch ihre Verteilungsfunktion F_X eindeutig festgelegt:

Korollar 1.38 *Zwei Wahrscheinlichkeitsmaße P und Q auf $\mathbb{R}^d$ sind genau dann identisch, wenn ihre Verteilungsfunktionen F_P und F_Q übereinstimmen.*

Der folgende Satz nennt wesentliche Eigenschaften von Verteilungsfunktionen im Spezialfall $d = 1$.

Satz 1.39 ***(Univariate Verteilungsfunktionen).*** *Eine Funktion $F : \mathbb{R} \to [0, 1]$ ist genau dann die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes P auf $\mathbb{R}$, wenn sie die folgenden Eigenschaften hat:*
(F.1) *F ist monoton wachsend;*
(F.2) $\lim_{r\to-\infty} F(r) = 0$ *und* $\lim_{r\to\infty} F(r) = 1$;
(F.3) *F ist rechtsseitig stetig, d.h.* $\lim_{s\downarrow r} F(s) = F(r)$ *für beliebige* $r \in \mathbb{R}$.
Ferner gilt dann:
(F.4) $F(r-) := \lim_{s\uparrow r, s<r} F(s)$ *ist gleich* $P((-\infty, r))$ *für beliebige* $r \in \mathbb{R}$.

Beweis von Satz 1.39. Sei $F = F_P$ für ein Wahrscheinlichkeitsmaß P auf $\mathbb{R}$. Dann ist

$$F(r) \;=\; P((-\infty, r]) \;\leq\; P((-\infty, s]) \;=\; F(s) \quad \text{für reelle Zahlen } r < s,$$

was die Monotonie (F.1) beweist. Wegen dieser Monotonie ist

$$\begin{aligned}
\lim_{r\to\infty} F(r) \;&=\; \lim_{\mathbb{N}\ni n\to\infty} F(n) \;=\; \lim_{\mathbb{N}\ni n\to\infty} P((-\infty, n]) \\
&=\; P\Big(\bigcup_{n\in\mathbb{N}} (-\infty, n]\Big) \;=\; P(\mathbb{R}) \;=\; 1, \\
\lim_{r\to-\infty} F(r) \;&=\; \lim_{\mathbb{N}\ni n\to\infty} F(-n) \;=\; \lim_{\mathbb{N}\ni n\to\infty} P((-\infty, -n]) \\
&=\; P\Big(\bigcap_{n\in\mathbb{N}} (-\infty, -n]\Big) \;=\; P(\emptyset) \;=\; 0;
\end{aligned}$$

siehe Anmerkung 1.11 und Anmerkung 1.12. Dies zeigt (F.2). Auch die rechtsseitige Stetigkeit (F.3) und Eigenschaft (F.4) ergibt sich aus der Monotonie und den Anmerkungen 1.11 bzw. 1.12:

$$\begin{aligned}
\lim_{s\downarrow r} F(s) \;&=\; \lim_{\mathbb{N}\ni n\to\infty} F(r + 1/n) \;=\; \lim_{\mathbb{N}\ni n\to\infty} P((-\infty, r + 1/n]) \\
&=\; P\Big(\bigcap_{n\in\mathbb{N}} (-\infty, r + 1/n]\Big) \;=\; P((-\infty, r]) \;=\; F(r), \\
\lim_{s\uparrow r, s<r} F(s) \;&=\; \lim_{\mathbb{N}\ni n\to\infty} F(r - 1/n) \;=\; \lim_{\mathbb{N}\ni n\to\infty} P((-\infty, r - 1/n]) \\
&=\; P\Big(\bigcup_{n\in\mathbb{N}} (-\infty, r - 1/n]\Big) \;=\; P((-\infty, r)).
\end{aligned}$$

Dass eine Funktion $F : \mathbb{R} \to [0,1]$ mit den Eigenschaften (F.1–3) stets die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes P auf $\mathbb{R}$ ist, kann man auf zwei Arten nachweisen: Einerseits sei $\mathcal{E}_o$ die Familie aller Elementarmengen in $\mathbb{R}$, die nur aus links offenen und rechts abgeschlossenen Intervallen aufgebaut sind. Eine nichtleere solche Menge lässt sich stets schreiben als

$$A \ := \ \bigcup_{i=1}^{m} J_i$$

mit $m \in \mathbb{N}$ und paarweise disjunkten Intervallen $J_i = (a_i, b_i]$ oder $J_i = (a_i, \infty)$, wobei $-\infty \leq a_i < b_i \leq \infty$. Dann definiert

$$P(A) \ := \ \sum_{i=1}^{m} (F(b_i) - F(a_i))$$

mit $F(\infty) := 1$ einen Wahrscheinlichkeitsinhalt auf $\mathcal{E}_o$. Mit ähnlichen Argumenten wie im Beweis von Satz 1.13 kann man zeigen, dass P sogar ein Maß auf $\mathcal{E}_o$ ist, und die Maßfortsetzung liefert dann das gewünschte Maß P auf $\mathrm{Borel}(\mathbb{R})$. Offensichtlich ist $F(r) = P((-\infty, r])$.

Einen anderen, eleganteren Beweis werden wir gleich mit Hilfe der Quantiltransformation führen.
□

1.2.3 Die Quantiltransformation

Der folgende Satz führt die sogenannte Quantilfunktion einer Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ ein und zeigt, wie man im Prinzip Zufallsvariablen mit beliebiger vorgegebener Verteilungsfunktion F simulieren kann.

Satz 1.40 *Sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion im Sinne von (F.1–3).*

***(a)** Für $u \in (0,1)$ ist*

$$F^{-1}(u) \ := \ \min\bigl\{r \in \mathbb{R} : F(r) \geq u\bigr\}$$

wohldefiniert, und man nennt $F^{-1} : (0,1) \to \mathbb{R}$ die Quantilfunktion von F. Für beliebige $r \in \mathbb{R}$ und $u \in (0,1)$ ist

$$F(r) \ \geq \ u \quad \text{genau dann, wenn} \quad r \ \geq \ F^{-1}(u).$$

***(b)** Sei U eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit uniformer Verteilung auf $(0,1)$; das heißt,*

$$P(U \in J) \ = \ \mathrm{Länge}(J) \quad \text{für beliebige Intervalle } J \subset (0,1).$$

Dann ist $X := F^{-1}(U)$ eine reellwertige Zufallsvariable mit Verteilungsfunktion $F_X = F$.

Teil (b) liegt vielen Simulationsprogrammen zugrunde. Die meisten Computerprogramme enthalten einen Zufallsgenerator, der bei jedem Aufrauf eine (Pseudo-) Zufallszahl $U \in (0,1)$ mit uniformer Verteilung auf $(0,1)$ ausspuckt. Ausgehend von einer gegebenen Verteilungsfunktion F und ihrer Quantilfunktion F^{-1} ist dann $X := F^{-1}(U)$ eine (Pseudo-) Zufallszahl mit Verteilungsfunktion F.

Mitunter wird die Einschränkung auf das offene Einheitsintervall als Definitionsbereich von F^{-1} noch aufgehoben, indem man definiert:

$$\begin{aligned} F^{-1}(0) &:= \inf\bigl\{r \in \mathbb{R} : F(r) > 0\bigr\} \in [-\infty, \infty), \\ F^{-1}(1) &:= \inf\bigl\{r \in \mathbb{R} : F(r) < 1\bigr\} \in (-\infty, \infty]. \end{aligned}$$

Dann ergibt sich die Abbildung $F^{-1} : [0,1] \to [-\infty, \infty]$.

Beweis von Satz 1.40. Wir zeigen erst, dass $F^{-1}(u)$ für jedes $u \in (0,1)$ wohldefiniert ist: Wegen $\lim_{r\to\infty} F(r) = 1$ ist die Menge $\{r : F(r) \ge u\}$ nichtleer, und wegen $\lim_{r\to-\infty} F(r) = 0$ ist sie nach unten beschränkt. Aus der Monotonie von F folgt nun, dass $\{r : F(r) \ge u\}$ von der Form (x,∞) oder $[x,\infty)$ mit einem $x \in \mathbb{R}$ sein muss. Doch die rechtsseitige Stetigkeit von F impliziert, dass

$$F(x) = \lim_{r\downarrow x,\, r>x} \underbrace{F(r)}_{\ge u} \ge u.$$

Demnach ist $\{r : F(r) \ge u\}$ gleich $[x,\infty)$ und $F^{-1}(u) = x$.

Die Äquivalenz der Ungleichungen $F(r) \ge u$ und $r \ge F^{-1}(u)$ ergibt sich jetzt direkt aus der Definition von $F^{-1}(u)$.

Was Teil (b) anbelangt, so sei U auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ definiert. Dann gilt für beliebige $r \in \mathbb{R}$:

$$\{\omega \in \Omega : X(\omega) \le r\} = \{\omega \in \Omega : F^{-1}(U(\omega)) \le r\} = \{\omega \in \Omega : U(\omega) \le F(r)\} \in \mathcal{A},$$

und

$$P(X \le r) = P(U \le F(r)) = F(r). \qquad \square$$

Beispiel 1.41 (Verteilungen mit endlichem Träger). Für gegebene Punkte $x_1 < x_2 < \cdots < x_m$ und Wahrscheinlichkeitsgewichte $p_1, p_2, \ldots, p_m > 0$, d.h. $\sum_{i=1}^m p_i = 1$, definieren wir eine Treppenfunktion $F : \mathbb{R} \to [0,1]$ wie folgt:

$$F := \begin{cases} 0 & \text{auf } (-\infty, x_1), \\ \sum_{i=1}^j p_i & \text{auf } [x_j, x_{j+1}), 1 \le j < m, \\ 1 & \text{auf } [x_m, \infty). \end{cases}$$

Dabei handelt es sich um eine Verteilungsfunktion im Sinne von (F.1–3). Die entsprechende Quantilfunktion F^{-1} ist ebenfalls eine Treppenfunktion:

$$F^{-1} := \begin{cases} x_1 & \text{auf } (-\infty, p_1], \\ x_2 & \text{auf } (p_1, p_1 + p_2], \\ x_3 & \text{auf } (p_1 + p_2, p_1 + p_2 + p_3], \\ \vdots & \vdots \\ x_m & \text{auf } (1 - p_m, 1]. \end{cases}$$

Definiert man $X = F^{-1}(U)$, dann ist $P(X = x_i) = p_i$ für $1 \leq i \leq m$.

Beispiel 1.42 (Verteilungsfunktionen mit Dichtefunktion). Sei $f : \mathbb{R} \to [0, \infty)$ uneigentlich Riemann–integrierbar mit

$$\int_{-\infty}^{\infty} f(x)\, dx \;=\; 1.$$

Dann definiert

$$F(r) \;:=\; \int_{-\infty}^{r} f(x)\, dx$$

eine stetige Verteilungsfunktion im Sinne von (F.1–3), und die entsprechende Quantilfunktion F^{-1} lässt sich schreiben als

$$F^{-1}(u) \;=\; \min\bigl\{r \in \mathbb{R} : F(r) = u\bigr\}.$$

Falls die Menge $J := \{x \in \mathbb{R} : f(x) > 0\}$ ein offenes Intervall ist, ist $F : J \to (0, 1)$ stetig, streng monoton wachsend und bijektiv mit Umkehrfunktion F^{-1}. Hier einige Spezialfälle:

Uniforme Verteilung $\mathrm{Unif}(a, b)$**.** Für $-\infty < a < b < \infty$ sei

$$f(x) \;:=\; \frac{1\{x \in (a, b)\}}{b - a}.$$

Hier ist

$$F(r) \;=\; \frac{r - a}{b - a} \quad \text{für } r \in [a, b], \quad \text{und} \quad F^{-1}(u) \;=\; a + (b - a) \cdot u.$$

Exponentialverteilung $\exp(\mu)$**.** Für ein $\mu > 0$ sei

$$f(x) \;:=\; 1\{x > 0\}\, \exp(-x/\mu)/\mu.$$

Dann ist

$$F(r) \;=\; 1 - \exp(-r/\mu) \quad \text{für } r \geq 0, \quad \text{und} \quad F^{-1}(u) \;=\; -\mu \log(1 - u).$$

Logistische Verteilung. Sei

$$f(x) \;:=\; \frac{\exp(x)}{(1 + \exp(x))^2} \;=\; \frac{1}{\exp(x) + \exp(-x) + 2}.$$

Dann ist

$$F(r) \;=\; \frac{\exp(r)}{1 + \exp(r)} \;=\; \frac{1}{\exp(-r) + 1} \quad \text{und} \quad F^{-1}(u) \;=\; \log \frac{u}{1 - u}.$$

Cauchy–Verteilung. Sei

$$f(x) \;:=\; \frac{1}{\pi(1 + x^2)}.$$

Dann ist

$$F(r) \;=\; \frac{1}{2} + \frac{\arctan(r)}{\pi} \quad \text{und} \quad F^{-1}(u) \;=\; \tan(\pi(u - 1/2)).$$

Beispiel 1.43 (Eine seltsame stetige Verteilungsfunktion.) Sei $(\Omega, \mathcal{A}, P)$ unser Modell für den unendlichen Münzwurf mit Parameter $\theta \in (0, 1)$. Dann definiert

$$Y(\omega) \ := \ \sum_{n=1}^{\infty} 3^{-n} 2\omega_n$$

eine reellwertige Zufallsvariable Y auf $(\Omega, \mathcal{A}, P)$. Genauer gesagt, nimmt Y nur Werte in $[0, 1]$ an, und die entsprechende Verteilungsfunktion F_Y hat die folgenden Eigenschaften:

- F_Y ist stetig,
- $F_Y(0) = 0$ und $F_Y(1) = 1$,
- Für beliebige $x \in \mathbb{R}$, die nicht in der Cantormenge C liegen, ist

$$\frac{d}{dx} F_Y(x) \ = \ 0.$$

Der Beweis dieser Tatsachen und die Definition der Cantormenge C sind Bestandteil der Übungen. Insbesondere wird bzw. wurde dort gezeigt, dass das Lebesguemaß von C gleich Null ist. Abbildung 1.2 zeigt F_Y für $\theta = 1/2$ und $\theta = 1/3$.

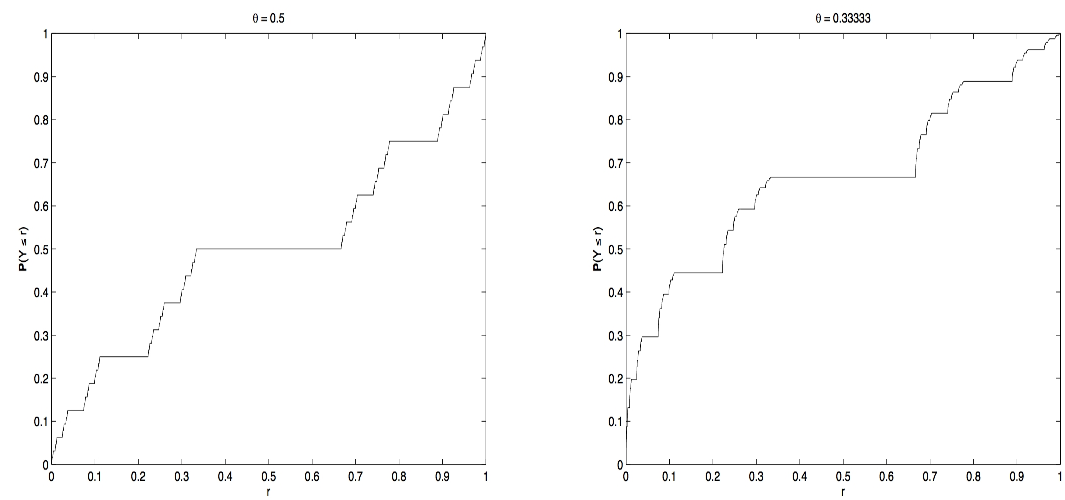

Abbildung 1.2: Verteilungsfunktionen für Beispiel 1.43

1.3 Stochastische Unabhängigkeit

Der Begriff der stochastischen Unabhängigkeit spielt in Wahrscheinlichkeitstheorie und Statistik eine zentrale Rolle. Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \in \mathcal{A}$ heißen bekanntlich *stochastisch unabhängig*, wenn

$$P(A \cap B) \ = \ P(A)P(B).$$

Diese Definition werden wir nun verallgemeinern.

1.3.1 Definitionen der stochastischen Unabhängigkeit und Beispiele

Definition 1.44 (Stochastische Unabhängigkeit von Ereignissen). Für jedes i in einer Indexmenge I sei ein Ereignis A_i gegeben. Diese Ereignisse A_i, $i \in I$, heißen *stochastisch unabhängig*, wenn für jede endliche Indexmenge $J \subset I$ gilt:

$$P\Big(\bigcap_{i \in J} A_i\Big) \;=\; \prod_{i \in J} P(A_i).$$

Anmerkung 1.45 (Komplementärereignisse). Die stochastische Unabhängigkeit von Ereignissen A_i, $i \in I$, bleibt erhalten, wenn man beliebig viele der Ereignisse A_i jeweils durch ihr Komplement A_i^c ersetzt.

Definition 1.46 (Stochastische Unabhängigkeit von Zufallsvariablen). Für jedes i in einer Indexmenge I sei X_i eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Werten in einem messbaren Raum $(\mathcal{X}_i, \mathcal{B}_i)$. Diese Zufallsvariablen X_i, $i \in I$, heißen *stochastisch unabhängig*, wenn für jede endliche Indexmenge $J \subset I$ und beliebige Mengen $B_i \in \mathcal{B}_i$ gilt:

$$P\big(X_i \in B_i \text{ für alle } i \in J\big) \;=\; \prod_{i \in J} P(X_i \in B_i).$$

(Ist I selbst eine endliche Menge, dann genügt es, den Fall $J = I$ zu betrachten. Denn anderenfalls könnte man $B_i := \mathcal{X}_i$ für $i \in I \setminus J$ wählen.)

Anmerkung 1.47 Die stochastische Unabhängigkeit von Ereignissen A_i, $i \in I$, ist gleichbedeutend mit der stochastischen Unabhängigkeit der $\{0,1\}$–wertigen Zufallsvariablen $\omega \mapsto X_i(\omega) := 1\{\omega \in A_i\}$, $i \in I$.

Mit Hilfe des Eindeutigkeitssatzes 1.24 ergeben sich vereinfachte Kriterien für stochastische Unabhängigkeit. Zunächst die beiden wichtigsten Spezialfälle:

Anmerkung 1.48 Reellwertige Zufallsvariablen $X_1, X_2, \ldots, X_n$ auf $(\Omega, \mathcal{A}, P)$ sind stochastisch unabhängig genau dann, wenn

$$P\big(X_1 \le r_1, X_2 \le r_2, \ldots, X_n \le r_n\big) \;=\; \prod_{i=1}^n P(X_i \le r_i) \quad \text{für beliebige } \boldsymbol{r} \in \mathbb{R}^n.$$

Anmerkung 1.49 Zufallsvariablen $X_1, X_2, \ldots, X_n$ auf $(\Omega, \mathcal{A}, P)$ mit Werten in einer abzählbaren Menge $\mathcal{X}$ sind stochastisch unabhängig genau dann, wenn

$$P\big(X_1 = x_1, , X_2 = x_2, \ldots, X_n = x_n\big) \;=\; \prod_{i=1}^n P(X_i = x_i) \quad \text{für beliebige } x_1, x_2, \ldots, x_n \in \mathcal{X}.$$

Beide Anmerkungen 1.48 und 1.49 ergeben sich aus der nachfolgenden Anmerkung:

Anmerkung 1.50 (Kriterium für stochastische Unabhängigkeit). Mit den Bezeichnungen von Definition 1.46 sei $\mathcal{B}_i = \sigma(\mathcal{E}_i)$ für jedes $i \in I$ mit einer $\cap$–stabilen Mengenfamile $\mathcal{E}_i \subset \mathcal{P}(\mathcal{X}_i)$. Dann sind die Zufallsvariablen X_i, $i \in I$, stochastisch unabhängig genau dann, wenn für jede endliche Indexmenge $J \subset I$ und beliebige Mengen $E_i \in \mathcal{E}_i$ gilt:

$$P\big(X_i \in E_i \text{ für alle } i \in J\big) \;=\; \prod_{i \in J} P(X_i \in E_i).$$

Um dies zu zeigen, betrachten wir ohne Einschränkung den Fall, dass $I = \{1, 2, \ldots, n\}$ für ein $n \in \mathbb{N}$. Mit

$$\begin{aligned} M_1(B_1, B_2, \ldots, B_n) \;&:=\; P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n), \\ M_2(B_1, B_2, \ldots, B_n) \;&:=\; \prod_{i=1}^{n} P(X_i \in B_i) \end{aligned}$$

ist zu zeigen, dass $M_1 = M_2$ auf $\mathcal{B}_1 \times \mathcal{B}_2 \times \cdots \times \mathcal{B}_n$. Nach Voraussetzung ist $M_1 = M_2$ auf $\mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_n$. Fixiert man für einen Index $i_o \in \{1, \ldots, n\}$ alle Mengen B_i, $i \neq i_o$, dann ist $B_{i_o} \mapsto M_s(B_1, \ldots, B_n)$ für $s = 1, 2$, ein Maß auf $\mathcal{B}_{i_o}$. Starten wir also mit beliebigen Mengen $B_i \in \mathcal{E}_i$, so können wir nun mit Hilfe des Eindeutigkeitssatzes 1.24 schrittweise jedes erzeugende System $\mathcal{E}_i$ durch $\mathcal{B}_i$ ersetzen …

Beispiel 1.51 (Unendlicher Münzwurf). Sei $\Omega = \{0, 1\}^{\mathbb{N}}$, $\mathcal{A} = \sigma(\mathcal{Z})$ und P das aus den Gewichten

$$p_n(y) \;:=\; \prod_{i=1}^{n} \theta^{y_i} (1 - \theta)^{1 - y_i} \quad (n \in \mathbb{N}, y \in \{0, 1\}^n)$$

konstruierte Wahrscheinlichkeitsmaß auf $\mathcal{A}$. Dann sind die $\{0, 1\}$–wertigen Zufallsvariablen $\omega \mapsto X_i(\omega) := \omega_i$ stochastisch unabhängig.

Beispiel 1.52 (Folgen von unabhängigen, reellwertigen Zufallsvariablen). Wie im Schlüsselbeispiel 2 sei $\Omega = \{0, 1\}^{\mathbb{N}}$, $\mathcal{A} = \sigma(\mathcal{Z})$ und P das aus den speziellen Gewichten

$$p_n(y) \;:=\; 2^{-n} \quad (n \in \mathbb{N}, y \in \{0, 1\}^n)$$

konstruierte Wahrscheinlichkeitsmaß auf $\mathcal{A}$. Für beliebig vorgegebene Verteilungsfunktionen F_1, F_2, F_3, … auf $\mathbb{R}$ existieren stochastisch unabhängige, reellwertige Zufallsvariablen Y_1, Y_2, Y_3, … auf $(\Omega, \mathcal{A}, P)$ mit genau diesen Verteilungsfunktionen.

Ausgangspunkt sind die stochastisch unabhängigen Zufallsvariablen $X_i : \Omega \to \{0, 1\}$ wie im vorigen Beispiel. Diese ordnen wir nun als unendliche Matrix an:

$$\begin{pmatrix} X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} & \cdots \\ X_{2,1} & X_{2,2} & X_{2,3} & \cdots & \\ X_{3,1} & X_{3,2} & \cdots & & \\ X_{4,1} & \cdots & & & \\ \vdots & & & & \end{pmatrix} \;:=\; \begin{pmatrix} X_1 & X_2 & X_4 & X_7 & \cdots \\ X_3 & X_5 & X_8 & \cdots & \\ X_6 & X_9 & \cdots & & \\ X_{10} & \cdots & & & \\ \vdots & & & & \end{pmatrix}$$

Dies liefert uns stochastisch unabhängige, auf $[0,1]$ uniform verteilte Zufallsvariablen

$$U_n \;:=\; \sum_{k=1}^{\infty} 2^{-k} X_{n,k} \quad (n \in \mathbb{N}),$$

und die Zufallsvariablen

$$Y_n \;:=\; F_n^{-1}(U_n)$$

haben dann die gewünschten Eigenschaften.

1.3.2 Das Borel–Cantelli–Lemma und Kolmogorovs 0–1–Gesetz

Ein einfaches Resultat, das recht oft verwendet wird:

Lemma 1.53 ***(Borel–Cantelli).*** *Für Ereignisse $A_1, A_2, A_3, \ldots$ in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ sei*

$$\limsup_{n\to\infty} A_n \;:=\; \bigcap_{N\in\mathbb{N}} \Big(\bigcup_{n>N} A_n \Big) \;=\; \big\{ \omega \in \Omega : \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N} \big\}.$$

(a) *Im Falle von $\sum_n P(A_n) < \infty$ ist $P\big(\limsup_n A_n\big) = 0$.*

(b) *Sind die Ereignisse A_n, $n \in \mathbb{N}$, stochastisch unabhängig, dann folgt aus $\sum_n P(A_n) = \infty$, dass $P\big(\limsup_n A_n\big) = 1$.*

Beweis von Lemma 1.53. Schreibt man

$$\limsup_n A_n \;=\; \bigcap_N C_N \quad \text{mit} \quad C_N \;:=\; \bigcup_{n>N} A_n,$$

dann ist $C_1 \supset C_2 \supset C_3 \supset \cdots$, so dass $P\big(\limsup_n A_n\big) = \lim_N P(C_N)$.

Im Falle von $\sum_n P(A_n) < \infty$ ist

$$\lim_{N\to\infty} P(C_N) \;\le\; \lim_{N\to\infty} \sum_{n>N} P(A_n) \;=\; 0.$$

Im Falle von stochastisch unabhängigen Ereignissen A_n, $n \in \mathbb{N}$, ist

$$\begin{aligned}
P(C_N) &= 1 - P(C_N^c) \\
&= 1 - P\Big(\bigcap_{n>N} A_n^c\Big) \\
&= 1 - \lim_{M\to\infty} P\Big(\bigcap_{n=N+1}^{M} A_n^c\Big) \\
&= 1 - \lim_{M\to\infty} \prod_{n=N+1}^{M} P(A_n^c) \\
&= 1 - \lim_{M\to\infty} \prod_{n=N+1}^{M} (1 - P(A_n)) \\
&= 1 - \lim_{M\to\infty} \exp\Big(\sum_{n=N+1}^{M} \log(1 - P(A_n))\Big) \\
&\geq 1 - \lim_{M\to\infty} \exp\Big(-\sum_{n=N+1}^{M} P(A_n)\Big) \\
&= 1,
\end{aligned}$$

falls $\sum_n P(A_n) = \infty$. $\square$.

Als erste Anwendung des Borel–Cantelli–Lemmas beweisen wir Bernoullis Gesetz der großen Zahlen für den unendlichen Münzwurf.

Korollar 1.54 ***(Bernoulli)***. *Für $n \in \mathbb{N}$ sei $\widehat{\theta}_n := n^{-1} \sum_{i=1}^n X_i$ mit stochastisch unabhängigen Zufallsvariablen $X_1, X_2, X_3, \ldots$ derart, dass $P(X_i = 1) = 1 - P(X_i = 0) = \theta$ für alle $i \in \mathbb{N}$ und ein festes $\theta \in [0, 1]$. Für beliebige $\Delta > 1/2$ ist dann*

$$\mathbb{P}\Big(\big|\widehat{\theta}_n - \theta\big| \geq \sqrt{\frac{\Delta \log n}{n}} \text{ für unendlich viele } n\Big) \;=\; 0.$$

Mit Wahrscheinlichkeit Eins ist demnach $\widehat{\theta}_n - \theta = O\Big((\log(n)/n)^{1/2}\Big)$. Insbesondere hat das in Schlüsselbeispiel 2 definierte Ereignis $A^{(1)}$ Wahrscheinlichkeit Eins. Die Aussage von Korollar 1.54 ist gleichbedeutend mit der Aussage, dass

$$\limsup_{n\to\infty} \sqrt{\frac{2n}{\log n}}\, \big|\widehat{\theta}_n - \theta\big| \;\leq\; 1 \quad \text{fast sicher.}$$

Dabei ist “fast sicher” ein gängiges Synonym für “mit Wahrscheinlichkeit Eins”. Später wurde von A. Kolmogorov bewiesen, dass sogar

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2\log\log n}}\, \big|\widehat{\theta}_n - \theta\big| \;=\; \sqrt{\theta(1-\theta)} \quad \text{fast sicher,}$$

und dies ist wiederum ein Spezialfall des allgemeineren “Gesetzes vom iterierten Logarithmus”.

Beweis von Korollar 1.54. Mit $\epsilon_n := (\Delta \log(n)/n)^{1/2}$ geht es um das Ereignis

$$A := \limsup_{n\to\infty} A_n \quad \text{mit} \quad A_n := \big\{ \big| \widehat{\theta}_n - \theta \big| \geq \epsilon_n \big\}.$$

Nun kann man mit Hilfe der Hoeffding– bzw. Chernov–Ungleichung zeigen, dass

$$P(A_n) \leq 2\exp(-2n\epsilon_n^2) = 2n^{-2\Delta};$$

siehe Übungen. Daher ergibt sich die Behauptung aus dem Borel–Cantelli–Lemma 1.53 und der Konvergenz der Reihe $\sum_{n\in\mathbb{N}} n^{-2\Delta}$. □

Lemma 1.53 zeigt, dass das Ereignis $\limsup_n A_n$ stets Wahrscheinlichkeit Null oder Eins hat, wenn die Einzelereignisse A_n, $n \in \mathbb{N}$, stochastisch unabhängig sind. Diese Feststellung ergibt sich auch aus zwei allgemeineren Resultaten. Um diese präzise zu formulieren benötigen wir noch eine Definition:

Definition 1.55 (Von Zufallsvariablen erzeugte σ–Algebren). Für jedes i in einer Indexmenge I sei X_i eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Werten in einem messbaren Raum $(\mathcal{X}_i, \mathcal{B}_i)$. Die *von den Zufallsvariablen X_i, $i \in I$, erzeugte σ–Algebra*, bezeichnet mit $\sigma(X_i : i \in I)$, ist die kleinste σ–Algebra über Ω, welche alle Ereignisse $X_i^{-1}(B_i)$ mit $i \in I$ und $B_i \in \mathcal{B}_i$ enthält.

Satz 1.56 ***(Kolmogorovs 0–1–Gesetz).*** *Seien $X_1, X_2, X_2, \ldots$ stochastisch unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Dann ist*

$$P(B) \in \{0,1\} \quad \textit{für alle } B \in \bigcap_{n\in\mathbb{N}} \sigma(X_n, X_{n+1}, X_{n+2}, \ldots).$$

Beispiele für Ereignisse in $\bigcap_{n\in\mathbb{N}} \sigma(X_{n+1}, X_{n+2}, X_{n+3}, \ldots)$ sind uns bereits begegnet. Im Falle von $X_i \in \mathbb{R}$ sind beispielsweise

$$\big\{ X_k \geq 0 \text{ für unendlich viele } k \in \mathbb{N} \big\}$$

oder

$$\Big\{ \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} X_i = c \Big\} = \Big\{ \lim_{k\to\infty} \frac{1}{k} \sum_{i=n}^{k} X_i = c \Big\} \quad (\text{für beliebige } n \in \mathbb{N})$$

von diesem Typ.

Satz 1.56 werden wir aus einem Hilfssatz ableiten, der von eigenem Interesse ist:

Hilfssatz 1.57 *Seien X_i, $i \in I$, stochastisch unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Für $J \in I$ sei $\mathcal{A}_J := \sigma(X_i : i \in J)$. Dann gilt für nichtleere disjunkte Mengen $J, K \subset I$:*

$$P(A \cap B) = P(A)P(B) \quad \textit{für alle } A \in \mathcal{A}_J, B \in \mathcal{A}_K.$$

Beweis von Satz 1.57*. Ein $\cap$–stabiles erzeugendes System für $\mathcal{A}_J$ ist die Menge $\mathcal{E}_J$ aller Ereignisse der Form

$$\bigcap_{i \in J} X_i^{-1}(B_i)$$

mit Mengen $B_i \in \mathcal{B}_i$, wobei aber $B_i \neq \mathcal{X}_i$ für höchstens endlich viele $i \in J$.

Nun argumentieren wir ähnlich wie in Anmerkung 1.50: Wir betrachten

$$\begin{aligned} M_1(A, B) &:= P(A \cap B), \\ M_2(A, B) &:= P(A)P(B) \end{aligned}$$

für $A \in \mathcal{A}_J$ und $B \in \mathcal{A}_K$. Zu zeigen ist, dass $M_1 = M_2$ auf $\mathcal{A}_J \times \mathcal{A}_K$. Einerseits folgt aus der Unabhängigkeit aller X_i, dass $M_1 = M_2$ auf $\mathcal{E}_J \times \mathcal{E}_K$. Andererseits ist $A \mapsto M_s(A, B)$ bei festem $B \in \mathcal{A}_K$ ein Maß auf $\mathcal{A}_J$, und $B \mapsto M_s(A, B)$ ist bei festem $A \in \mathcal{A}_J$ ein Maß auf $\mathcal{A}_K$. Zweimaliges Anwenden des Eindeutigkeitssatzes 1.24 ergibt nun, dass $M_1 = M_2$ auf $\mathcal{A}_J \times \mathcal{A}_K$.
$\square$

Beweis von Satz 1.56*. Mit $\mathcal{B}_n := \sigma(X_n, X_{n+1}, X_{n+2}, \ldots)$ und $\mathcal{B}_\infty := \bigcap_{n \in \mathbb{N}} \mathcal{B}_n$ zeigen wir, dass sogar gilt:

$$P(A \cap B) = P(A)P(B) \quad \text{für alle } A \in \mathcal{B}_1, B \in \mathcal{B}_\infty.$$

Insbesondere ist dann $P(B) = P(B \cap B) = P(B)^2$ für beliebige $B \in \mathcal{B}_\infty$, was gleichbedeutend ist mit $P(B) \in \{0, 1\}$.

Aus dem Hilfssatz 1.57 folgt, dass

$$P(A \cap B) = P(A)P(B) \quad \text{für alle } A \in \sigma(X_1, \ldots, X_n), B \in \mathcal{B}_{n+1}, n \in \mathbb{N}.$$

Insbesondere ist

$$P(A \cap B) = P(A)P(B) \quad \text{für alle } A \in \bigcup_{n \in \mathbb{N}} \sigma(X_1, \ldots, X_n), B \in \mathcal{B}_\infty.$$

Doch die linke und rechte Seite dieser Gleichung sind, bei festem $B \in \mathcal{B}_\infty$ und als Funktionen von A, Maße auf $\mathcal{B}_1$. Außerdem ist $\bigcup_{n \in \mathbb{N}} \sigma(X_1, \ldots, X_n)$ ein $\cap$–stabiles erzeugendes System von $\mathcal{B}_1$. Aus Satz 1.24 folgt dann die Behauptung. $\square$

Kapitel 2

Integrale und Erwartungswerte

Sowohl der Erwartungswert $\mathbb{E}(X)$ einer reellwertigen Zufallsvariable X auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) als auch das Riemann-Integral $\int_a^b f(x)\,dx$ einer Funktion $f : [a,b] \to \mathbb{R}$ sind Beispiele für ein *Lebesgue-Integral*, das wir im nächsten Abschnitt beschreiben. Da dies eigentlich Gegenstand der Analysis (3) ist, werden wir aus Zeitgründen auf viele Beweise verzichten.

2.1 Das Lebesgue-Integral

Sei $(\Omega, \mathcal{A}, M)$ ein Maßraum. Im Folgenden betrachten wir messbare Funktionen $f : \Omega \to [0,\infty]$ oder $f : \Omega \to \mathbb{R}$. Das heißt, $f^{-1}(B) \in \mathcal{A}$ für beliebige Intervalle $B \subset [0,\infty]$ bzw. $B \subset \mathbb{R}$.

Das Integral "einfacher Funktionen"

Wir betrachten zunächst die Menge $\mathcal{G}_+ = \mathcal{G}_+(\Omega, \mathcal{A})$ aller Funktionen $f : \Omega \to [0,\infty]$ der Form

$$f \;=\; \sum_{i=1}^{m} 1_{A_i} \lambda_i$$

mit $m \in \mathbb{N}$, $\lambda_i \in [0,\infty]$ und $A_i \in \mathcal{A}$. Dabei ist 1_A die Indikatorfunktion einer Menge A, also $1_A(\omega) = 1\{\omega \in A\}$. Mit anderen Worten, $\mathcal{G}_+$ besteht aus allen messbaren Funktionen $f : \Omega \to [0,\infty]$ mit *endlichem* Wertebereich $f(\Omega)$. Das nachfolgende Lemma benötigen wir, um das Integral einer Funktion $f \in \mathcal{G}_+$ zu defnieren.

Lemma 2.1 *Für zwei Funktionen $f = \sum_{i=1}^m 1_{A_i}\lambda_i$ und $g = \sum_{j=1}^n 1_{B_j}\mu_j$ folgt aus $f \le g$ (punktweise), dass auch*

$$\sum_{i=1}^{m} M(A_i)\lambda_i \;\le\; \sum_{j=1}^{n} M(B_j)\mu_j .$$

Definition 2.2 (Lebesgue-Integral, I). Für eine Funktion $f = \sum_{i=1}^m 1_{A_i}\lambda_i$ definiert man ihr *(Lebesgue-) Integral bezüglich* M als die Zahl

$$\int f\,dM \;:=\; \sum_{i=1}^m M(A_i)\lambda_i \;\in\; [0,\infty].$$

Anmerkung 2.3 (Linearität und Monotonie, I). Aus Lemma 2.1 folgt, dass $\int f\,dM$ wohldefiniert, d.h. von der speziellen Darstellung von f unabhängig ist. Ferner erfüllt dieses Integral folgende Eigenschaften:

$$\int f\,dM \;\le\; \int g\,dM \quad \text{für } f,g \in \mathcal{G}_+ \text{ mit } f \le g;$$

$$\int (\lambda f + \mu g)\,dM \;=\; \lambda \int f\,dM + \mu \int g\,dM \quad \text{für } f,g \in \mathcal{G}_+ \text{ und } \lambda,\mu \ge 0.$$

Beweis von Lemma 2.1. Wie in einer Übungsaufgabe zu σ–Algebren mit endlichem erzeugenden System kann man paarweise disjunkte und nichtleere Mengen $C_1, C_2, \ldots, C_q \in \mathcal{A}$ konstruieren, so dass

$$A_i \;=\; \bigcup_{s:C_s\subset A_i} C_s \quad \text{und} \quad B_j \;=\; \bigcup_{s:C_s\subset B_j} C_s$$

für $1 \le i \le m$ und $1 \le j \le n$. (Dabei ist $q \le 2^{m+n} - 1$.) Folglich ist

$$\begin{aligned} f &= \sum_{s=1}^q 1_{C_s} \sum_{i=1}^m 1\{C_s \subset A_i\}\lambda_i, \\ g &= \sum_{s=1}^q 1_{C_s} \sum_{j=1}^n 1\{C_s \subset B_j\}\mu_j. \end{aligned}$$

Da die Mengen C_s paarweise disjunkt und nichtleer sind, ist $f \le g$ gleichbedeutend mit

$$\sum_{i=1}^m 1\{C_s \subset A_i\}\lambda_i \;\le\; \sum_{j=1}^n 1\{C_s \subset B_j\}\mu_j \quad \text{für } s = 1,2,\ldots,q.$$

Folglich ist

$$\begin{aligned} \sum_{i=1}^m M(A_i)\lambda_i &= \sum_{i=1}^m \Big(\sum_{s=1}^q M(C_s) 1\{C_s \subset A_i\}\Big)\lambda_i \\ &= \sum_{s=1}^q M(C_s) \sum_{i=1}^m 1\{C_s \subset A_i\}\lambda_i \\ &\le \sum_{s=1}^q M(C_s) \sum_{j=1}^n 1\{C_s \subset B_j\}\mu_j \\ &= \sum_{j=1}^n M(B_j)\mu_j. \qquad \Box \end{aligned}$$

Das Integral nichtnegativer Funktionen

Definition 2.4 (Lebesgue-Integral, II). Für eine beliebige messbare Funktion $f : \Omega \to [0, \infty]$ definiert man ihr (Lebesue-) Integral bezüglich M als die Zahl

$$\int f\,dM \;:=\; \sup\Big\{ \int g\,dM : g \in \mathcal{G}_+, g \leq f \Big\}.$$

Anmerkung 2.5 (Monotonie, II). Für $f \in \mathcal{G}_+$ ist die zuletzt gegebene Definition von $\int f\,dM$ mit der ursprünglichen verträglich. Aus der Definition folgt auch direkt die Ungleichung

$$\int f\,dM \;\leq\; \int g\,dM \quad \text{für } f, g : \Omega \to [0, \infty] \text{ messbar, } f \leq g.$$

Wenn man die bisher verwendeten Argumente aufmerksam verfolgt, wird deutlich, dass wir bisher nur ausnutzten, dass (i) $\mathcal{A}$ eine Algebra über Ω und (ii) M ein Inhalt auf $\boldsymbol{A}$ ist. Der nächste Satz macht wirklich von den Eigenschaften eines Maßes Gebrauch:

Satz 2.6 ***(Monotone Konvergenz).*** *Seien $f_1, f_2, f_3, \ldots : \Omega \to [0, \infty]$ messbare Funktionen, so dass $f_1 \leq f_2 \leq f_3 \leq \cdots$. Dann gilt für den (punktweisen) Grenzwert $f := \lim_{n\to\infty} f_n$:*

$$\int f\,dM \;=\; \lim_{n\to\infty} \int f_n\,dM.$$

Anmerkung 2.7 (Approximation durch einfache Funktionen). Zu jeder messbaren Funktion $f : \Omega \to [0, \infty]$ gibt es eine Folge von reellwertigen Funktionen $f_n \in \mathcal{G}_+$, so dass $f_n \uparrow f$ für $n \to \infty$. Hier ist eine von vielen möglichen Konstruktionen:

$$f_n \;:=\; \sum_{k=1}^{n2^n-1} 1\big\{k2^{-n} \leq f < (k+1)2^{-n}\big\} \cdot k2^{-n} + 1\{f \geq n\} \cdot n \;=\; 2^{-n} \sum_{k=1}^{n2^n} 1\big\{f \geq k2^{-n}\big\}.$$

Man unterteilt also $[0, \infty]$ in die $2^n n$ Intervalle $\big[k2^{-n}, (k+1)2^{-n}\big)$ $(1 \leq k \leq n2^n - 1)$ und $[n, \infty]$, und die Funktion f wird zu den linken Randpunkten hin abgerundet.

Anmerkung 2.8 (Linearität, II). Aus Anmerkung 2.7, Anmerkung 2.3 und dem Satz von der monotonen Konvergenz ergibt sich, dass

$$\int (\lambda f + \mu g)\,dM \;=\; \lambda \int f\,dM \;+\; \mu \int g\,dM \quad \text{für } f, g : \Omega \to [0, \infty] \text{ messbar und } \lambda, \mu \geq 0.$$

Anmerkung 2.9 (Markov–Ungleichung). Für jede messbare Funktion $f : \Omega \to [0, \infty]$ und beliebiges $\epsilon > 0$ ist

$$\int f\,dM \;\geq\; M\{f \geq \epsilon\}\epsilon,$$

denn $f \geq 1\{f \geq \epsilon\}\epsilon$. Insbesondere kann man hieraus ableiten, dass

$$\int f\,dM \;=\; 0 \quad \text{genau dann, wenn} \quad M\{f > 0\} \;=\; 0.$$

Das Integral reellwertiger Funktionen

Für eine reelle Zahl $s \in \mathbb{R}$ seien

$$s^+ := \max(s, 0) \quad \text{und} \quad s^- := \max(-s, 0)$$

ihr Positiv- bzw. Negativteil. Dann ist

$$s^\pm \;\geq\; 0, \quad s \;=\; s^+ - s^- \quad \text{und} \quad |s| \;=\; s^+ + s^-.$$

Definition 2.10 (Lebesgue-Integral, III). Eine messbare Funktion $f : \Omega \to \mathbb{R}$ heißt *integrierbar bezüglich* M (oder kurz: M–integrierbar), wenn

$$\int |f|\, dM = \int f^+\, dM + \int f^-\, dM \;<\; \infty.$$

Dann definiert man das (Lebesgue-) Integral von f bezüglich M als die reelle Zahl

$$\int f\, dM \;:=\; \int f^+\, dM \;-\; \int f^-\, dM.$$

Aus dieser Definition folgt direkt, dass

$$\Big| \int f\, dM \Big| \;\leq\; \int |f|\, dM$$

Allgemein sei $f = f_1 - f_2$ mit messbaren Funktionen $f_1, f_2 : \Omega \to [0, \infty)$, so dass $\int f_i\, dM < \infty$. Dann ist $g := f_1 - f^+ = f_2 - f^- \geq 0$, und f ist M–integrierbar mit $\int f\, dM = \int f_1\, dM - \int f_2\, dM$. Mit dieser einfachen Vorüberlegung lassen sich die folgenden Eigenschaften des Integrals sehr leicht beweisen:

Anmerkung 2.11 (Linearität und Monotonie, III). Seien f und g M–integrierbare Funktionen auf Ω. Für reelle Zahlen λ und μ ist dann auch $\lambda f + \mu g$ eine M–integrierbare Funktion, und es gilt:

$$\int (\lambda f + \mu g)\, dM \;=\; \lambda \int f\, dM \;+\; \mu \int g\, dM.$$

Ferner ist

$$\int f\, dM \;\leq\; \int g\, dM \quad \text{falls } f \leq g.$$

Zum Schluss beweisen wir noch ein wichtiges Resultat über die Vertauschung von Integration und Grenzwertbildung:

Satz 2.12 ***(Majorisierte Konvergenz).*** *Seien $f_1, f_2, f_3, \ldots : \Omega \to \mathbb{R}$ messbare Funktionen, so dass der (punktweise) Grenzwert $f := \lim_{n\to\infty} f_n$ existiert. Ist $|f_n| \leq g$ für beliebige $n \in \mathbb{N}$ mit einer M–integrierbaren Funktion g, dann sind alle f_n und f M-integrierbar mit*

$$\lim_{n\to\infty} \int |f_n - f|\, dM \;=\; 0 \quad \textit{und} \quad \lim_{n\to\infty} \int f_n\, dM \;=\; \int f\, dM.$$

Beweis von Satz 2.12. Nach Voraussetzung ist

$$h_n \;:=\; \sup_{m\geq n} |f_m - f|$$

eine nichtnegative und messbare Funktionen mit $2g \geq h_n \downarrow 0$ für $n \to \infty$. Aus dem Satz von der monotonen Konvergenz folgt also, dass

$$\begin{aligned}\limsup_{n\to\infty} \int |f_n - f|\, dM \;&\leq\; \limsup_{n\to\infty} \int h_n\, dM \\ &=\; \int 2g\, dM - \liminf_{n\to\infty} \int \underbrace{(2g - h_n)}_{\uparrow 2g}\, dM \\ &=\; \int 2g\, dM - \int 2g\, dM \;=\; 0. \qquad \square\end{aligned}$$

Zwei zusätzliche Schreibweisen. Für eine Funktion f auf Ω und $A \in \mathcal{A}$ definiert man

$$\int_A f\, dM \;:=\; \int 1_A f\, dM,$$

falls letzteres Integral definiert ist, und spricht von dem Integral von f (bezüglich M) *über der Menge* A. Insbesondere ist $\int f\, dM = \int_\Omega f\, dM$.

Mitunter schreibt man $\int_A f(\omega)\, M(d\omega)$ anstelle von $\int_A f\, dM$.

Anmerkung 2.13 (Lebesgue-Maß und Summen). Summen und Reihen lassen sich als spezielle Lebesgue-Integrale deuten: Sei Ω eine endliche oder abzählbar unendliche Menge, und sei M das Zählmaß auf $\mathcal{P}(\Omega)$; das heißt, $M(A) := \#A$ für $A \subset \Omega$. Dann ist

$$\int f\, dM \;=\; \sum_{\omega\in\Omega} f(\omega)$$

für beliebige Funktionen $f : \Omega \to [0, \infty]$ bzw. Funktionen $f : \Omega \to \mathbb{R}$ mit $\sum_{\omega\in\Omega} |f(\omega)| < \infty$.

Anmerkung 2.14 (Lebesgue-Maß und Riemann-Integrale). Das Lebesgue-Integral ist eine Verallgemeinerung des Riemann-Integrals, und letzteres ist ein essentielles Hilfsmittel für konkrete Berechnungen: Sei Leb das Lebesguemaß auf $\mathrm{Borel}(\mathbb{R})$, also $\mathrm{Leb}(B) = \mathrm{Länge}(B)$ für Intervalle $B \subset \mathbb{R}$. Ist nun f messbar und Riemann-integrierbar auf $[a, b]$, dann ist

$$\int_a^b f(x)\, dx \;=\; \int_{[a,b]} f\, d\,\mathrm{Leb}\,.$$

Anmerkung 2.15 (Wahrscheinlichkeitsdichten). Sei $(\Omega, \mathcal{A}, M)$ ein Maßraum und f eine nichtnegative messbare Funktion auf Ω, so dass $\int f\, dM = 1$. Dann definiert

$$P(A) \;:=\; \int_A f\, dM$$

ein Wahrscheinlichkeitsmaß P auf $\mathcal{A}$, und f nennt man die *Dichtefunktion von P bezüglich M*. Für eine weitere messbare Funktion g auf Ω ist

$$\int g\,dP \;=\; \int gf\,dM.$$

Dabei ist das Integral auf der linken Seite genau dann wohldefiniert, wenn dies für das Integral auf der rechten Seite gilt.

Im Falle von $\Omega = \mathbb{R}$ und $M = \mathrm{Leb}$ spricht man auch einfach von einer *Dichtefunktion* oder *Lebesgue–Dichtefunktion*.

2.2 Erwartungswerte

Sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in $[0, \infty]$ oder in $\mathbb{R}$. Der *Erwartungswert* von X ist das Integral

$$\mathbb{E}(X) \;:=\; \int_\Omega X\,dP$$

von X bezüglich P, falls letzteres wohldefiniert ist. Mitunter verzichtet man auf Klammern und schreibt $\mathbb{E}\,X$ an Stelle von $\mathbb{E}(X)$.

Die Darstellung von $\mathbb{E}(X)$ als Integral über dem Grundraum Ω ist vor allem für theoretische Überlegungen hilfreich. Insbesondere ergibt sich aus den allgemeinen Eigenschaften des Integrals die Tatsache, dass für Zufallsvariablen X, Y und Konstanten λ, μ gilt:

$$\mathbb{E}(\lambda X + \mu Y) \;=\; \lambda\,\mathbb{E}(X) + \mu\,\mathbb{E}(Y)$$

und

$$\mathbb{E}(X) \;\le\; \mathbb{E}(Y) \quad \text{falls } X \le Y.$$

Dabei setzen wir voraus, dass entweder $X, Y, \lambda, \mu \in [0, \infty]$, oder $X, Y, \lambda, \mu \in \mathbb{R}$ mit $\mathbb{E}\,|X| < \infty$ und $\mathbb{E}\,|Y| < \infty$.

Konkrete Berechnung von Erwartungswerten

Für die konkrete Berechnung von Erwartungswerten bedient man sich Summen, Reihen und Riemann-Integralen. Nachfolgend beschreiben wir die wichtigsten Spezialfälle.

Diskret verteilte Zufallsvariablen. Wenn X nur Werte in einer abzählbaren Menge $R \subset \mathbb{R}$ annimmt, dann ist

$$\mathbb{E}(X) \;=\; \sum_{r \in R} P(X = r) \cdot r. \tag{2.1}$$

Zufallsvariablen mit Dichtefunktion. Wenn X nach einer Lebesgue-Dichtefunktion f verteilt ist, also $P(X \in J) = \int_J f(x)\,dx$ für Intervalle $J \subset \mathbb{R}$, dann ist

$$\mathbb{E}(X) \ = \ \int_{\mathbb{R}} x f(x)\,dx. \tag{2.2}$$

Zufallsvariablen mit Werten in $\mathbb{N}_0$. Im Falle von $X : \Omega \to \mathbb{N}_0$ ist

$$\mathbb{E}(X) \ = \ \sum_{k=1}^{\infty} P(X \geq k). \tag{2.3}$$

Nichtnegative Zufallsvariablen. Im Falle von $X : \Omega \to [0, \infty]$ gilt die Formel

$$\mathbb{E}(X) \ = \ \int_0^\infty P(X > r)\,dr \ = \ \int_0^\infty P(X \geq r)\,dr. \tag{2.4}$$

Im Falle von $P(X = \infty) > 0$ sind beide Seiten gleich ∞, und nichts bleibt zu zeigen. Also sei $P(X = \infty) = 0$. Für $k \in \mathbb{N}$ seien

$$\begin{aligned}
X_{k,1} \ &:= \ \sum_{j=1}^{\infty} 1\Big\{X \in \Big[\frac{j}{k}, \frac{j+1}{k}\Big)\Big\}\frac{j}{k} \\
&= \ \sum_{j=1}^{\infty} 1\Big\{X \geq \frac{j}{k}\Big\}\frac{1}{k}, \\
X_{k,2} \ &:= \ \sum_{j=1}^{\infty} 1\Big\{X \in \Big(\frac{j-1}{k}, \frac{j}{k}\Big]\Big\}\frac{j}{k} \\
&= \ \sum_{s=0}^{\infty} 1\Big\{X > \frac{s}{k}\Big\}\frac{1}{k}.
\end{aligned}$$

Dann ist $X_{k,1} \leq X \leq X_{k,2}$ fast sicher, und zusammen mit dem Satz von der monotonen Konvergenz (angewandt auf $X_{k,1}$ und $X_{k,2}$) ergeben sich die Ungleichungen

$$\begin{aligned}
\mathbb{E}(X) \ &\geq \ \mathbb{E}(X_{k,1}) \ = \ \sum_{j=1}^{\infty} P\Big(X \geq \frac{j}{k}\Big)\frac{1}{k} \\
&\geq \ \sum_{j=1}^{\infty} \int_{j/k}^{(j+1)/k} P(X \geq r)\,dr \ = \ \int_{1/k}^{\infty} P(X \geq r)\,dr, \\
\mathbb{E}(X) \ &\leq \ \mathbb{E}(X_{k,2}) \ = \ \sum_{s=0}^{\infty} P(X > s/k)/k \\
&\leq \ 1/k + \sum_{s=1}^{\infty} \int_{(s-1)/k}^{s/k} P(X > r)\,dr \ = \ 1/k + \int_0^\infty P(X > r)\,dr.
\end{aligned}$$

Für $k \to \infty$ folgt hieraus, dass

$$\int_0^\infty P(X \geq r)\,dr \ \leq \ \mathbb{E}(X) \ \leq \ \int_0^\infty P(X > r)\,dr.$$

Da die rechte Seite offensichtlich nicht größer als die linke Seite ist, ergibt sich hieraus (2.4). $\square$

Eine allgemeine Betrachtung. Formel (2.2) kann man als Spezialfall einer allgemeineren Regel auffassen: Für eine Zufallsvariable X auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(\mathcal{X}, \mathcal{B})$ und eine messbare Funktion $h : \mathcal{X} \to [0, \infty]$ oder $h : \mathcal{X} \to \mathbb{R}$ gilt:

$$\mathbb{E}\, h(X) \;=\; \int_\Omega h(X(\omega))\, P(d\omega) \;=\; \int_{\mathcal{X}} h(x)\, P^X(dx). \tag{2.5}$$

Dabei setzen wir voraus, dass $h \geq 0$ oder $\int |h(X)|\, dP < \infty$. Gleichung (2.5) lässt sich in drei Schritten nachweisen. Zunächst sei $h \in \mathcal{G}_+(\mathcal{X}, \mathcal{B})$, also $h = \sum_{i=1}^m 1_{B_i} \lambda_i$ mit $m \in \mathbb{N}$, $B_i \in \mathcal{B}$ und $\lambda_i \geq 0$. Dann ist

$$h(X(\omega)) \;=\; \sum_{i=1}^m 1_{B_i}(X(\omega))\lambda_i \;=\; \sum_{i=1}^m 1_{X^{-1}(B_i)}(\omega)\lambda_i,$$

also $h(X) \in \mathcal{G}_+(\Omega, \mathcal{A})$ mit

$$\begin{aligned} \mathbb{E}\, h(X) \;&=\; \sum_{i=1}^m P(X^{-1}(B_i))\lambda_i \\ &=\; \sum_{i=1}^m P^X(B_i)\lambda_i \\ &=\; \int h(x)\, P^X(dx). \end{aligned}$$

Der Fall einer beliebigen messbaren Funktion $h \geq 0$ ergibt sich durch Approximation von h durch Funktionen in $\mathcal{G}_+(\mathcal{X}, \mathcal{B})$ und den Satz von der monotonen Konvergenz. Der Fall $h \in \mathbb{R}$ ergibt sich dann durch Aufspalten von h in Positiv- und Negativteil.

Beispiel 2.16 (Exponentialverteilungen). Sei X exponentialverteilt mit Parameter $\mu > 0$, also $P(X \in B) = \int_B f(x)\, dx$ mit der Lebesgue-Dichtefunktion

$$f(x) \;=\; 1\{x > 0\} \exp(-x/\mu)/\mu.$$

Nach Formel (2.2) ist

$$\begin{aligned} \mathbb{E}(X) \;&=\; \int_{\mathbb{R}} x f(x)\, dx \\ &=\; \int_0^\infty x \exp(-x/\mu)/\mu\, dx \\ &=\; \mu \int_0^\infty y \exp(-y)\, dy \\ &=\; \mu. \end{aligned}$$

Mit der Verteilungsfunktion $F(r) := P(X \leq r) = 1 - \exp(-r/\mu)$ und Formel (2.4) kommt man zum gleichen Ziel:

$$\begin{aligned} \mathbb{E}(X) \;&=\; \int_0^\infty P(X > r)\, dr \\ &=\; \int_0^\infty \exp(-x/\mu)\, dx \\ &=\; \mu \int_0^\infty \exp(-y)\, dy \\ &=\; \mu. \end{aligned}$$

Beispiel 2.17 (Ordnungsstatisiken). Seien $U_1, U_2, \ldots, U_n$ stochastisch unabhängige, uniform auf $[0,1]$ verteilte Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Nun ordnen wir die Werte U_i der Größe nach und erhalten die *Ordnungsstatistiken* $U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}$. Uns interessiert die Verteilung von $U_{(k)}$ für eine feste Zahl $k \in \{1, 2, \ldots, n\}$. Als erstes berechnen wir die Verteilungsfunktion von $U_{(k)}$: Für $r \in [0,1]$ ist

$$\begin{aligned}
F(r) &:= P(U_{(k)} \le r) \\
&= P\left(\text{mindestens } k \text{ Werte } U_i \text{ sind in } [0,r]\right) \\
&= \sum_{\ell=k}^{n} P\left(\text{genau } \ell \text{ Werte } U_i \text{ sind in } [0,r]\right) \\
&= \sum_{\ell=k}^{n} \binom{n}{\ell} r^\ell (1-r)^{n-\ell}.
\end{aligned}$$

Diese Verteilungsfunktion F ist stetig differenzierbar mit Ableitung

$$f(x) = \frac{d}{dx} F(x) \;=\; n\binom{n-1}{k-1} x^{k-1}(1-x)^{n-k},$$

und $F(0) = 0$, $F(1) = 1$. Also ist f die Lebesgue-Dichtefunktion der Verteilung von $U_{(k)}$. Insbesondere ist $\int_0^1 f(x)\,dx = 1$, und wir erhalten folgende Formel als Nebenprodukt:

$$\int_0^1 x^a (1-x)^b \, dx \;=\; (a+b+1)^{-1} \binom{a+b}{a}^{-1} \quad \text{für } a, b \in \mathbb{N}_0. \tag{2.6}$$

Daraus ergeben sich die Formeln

$$\begin{aligned}
\mathbb{E}(U_{(k)}) &= \int_0^1 x f(x)\,dx \\
&= n\binom{n-1}{k-1} \int_{[0,1]} x^k (1-x)^{n-k}\,dx \\
&= n\binom{n-1}{k-1} (n+1)^{-1} \binom{n}{k}^{-1} \\
&= \frac{k}{n+1},
\end{aligned}$$

und

$$\begin{aligned}
\mathbb{E}(U_{(k)}^2) &= \int_0^1 x^2 f(x)\,dx \\
&= n\binom{n-1}{k-1} \int_{[0,1]} x^{k+1} (1-x)^{n-k}\,dx \\
&= n\binom{n-1}{k-1} (n+2)^{-1} \binom{n+1}{k+1}^{-1} \\
&= \frac{k(k+1)}{(n+1)(n+2)}.
\end{aligned}$$

Insbesondere ist

$$\begin{aligned}
\operatorname{Var}(U_{(k)}) &= \mathbb{E}\big((U_{(k)} - \mathbb{E}\, U_{(k)})^2\big) \\
&= \mathbb{E}(U_{(k)}^2) - (\mathbb{E}\, U_{(k)})^2 \\
&= \frac{k(k+1)}{(n+1)(n+2)} - \frac{k^2}{(n+1)^2} \\
&= \frac{1}{n+2}\frac{k}{n+1}\Big(1 - \frac{k}{n+1}\Big) \;\le\; \frac{1}{4(n+2)}.
\end{aligned}$$

Bei großem n weicht also $U_{(k)}$ nur wenig von $k/(n+1)$ ab.

2.3 Produkte von Funktionen und Zufallsvariablen

Wir wissen, dass das Integral einer (punktweise definierten) Summe von Funktionen identisch ist mit der Summe der Integrale. Eine naheliegende Frage ist, was man über das Integral bzw. den Erwartungswert von Produkten sagen kann. Zunächst beweisen wir ein sehr wichtiges Resultat für Zufallsvariablen:

Satz 2.18 *Seien X und Y stochastisch unabhängige Zufallsvariablen mit Werten in $[0,\infty]$. Dann ist*

$$\mathbb{E}(XY) \;=\; \mathbb{E}(X)\,\mathbb{E}(Y).$$

Diese Produktformel gilt auch im Falle von $X, Y \in \mathbb{R}$ mit $\mathbb{E}\,|X| < \infty$ und $\mathbb{E}\,|Y| < \infty$.

Beweis von Satz 2.18. *Schritt 1:* Zunächst seien die Wertebereiche von X und Y endlich, das heißt,

$$X \;=\; \sum_{i=1}^m 1\{X = \lambda_i\}\lambda_i \quad \text{und} \quad Y \;=\; \sum_{j=1}^n 1\{Y = \mu_j\}\mu_j$$

mit paarweise verschiedenen Zahlen $\lambda_1, \ldots, \lambda_m \in [0,\infty)$ bzw. $\mu_1, \ldots, \mu_n \in [0,\infty)$. Dann ist

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}\Big(\sum_{i=1}^m \sum_{j=1}^n 1\{X = \lambda_i, Y = \mu_j\}\lambda_i\mu_j\Big) \\
&= \sum_{i=1}^m \sum_{j=1}^n P(X = \lambda_i, Y = \mu_j)\lambda_i\mu_j \\
&= \sum_{i=1}^m \sum_{j=1}^n P(X = \lambda_i)P(Y = \mu_j)\lambda_i\mu_j \\
&= \Big(\sum_{i=1}^m P(X = \lambda_i)\lambda_i\Big)\Big(\sum_{j=1}^n P(Y = \mu_j)\mu_j\Big) \\
&= \mathbb{E}(X)\,\mathbb{E}(Y).
\end{aligned}$$

Schritt 2: Im Falle von $X, Y \geq 0$ betrachten wir die Zufallsvariablen

$$X_n \;:=\; 2^{-n} \sum_{k=1}^{n2^n} 1\{X \geq k2^{-n}\} \quad \text{und} \quad Y_n \;:=\; 2^{-n} \sum_{k=1}^{n2^n} 1\{Y \geq k2^{-n}\}.$$

Auch diese sind stochastisch unabhängig und nehmen nur Werte in $\{k2^{-n} : k = 0, 1, \ldots, n2^n\}$ an. Wegen $X_n \uparrow X$ und $Y_n \uparrow Y$ für $n \to \infty$, folgt aus dem Satz von der monotonen Konvergenz und Schritt 1, dass

$$\mathbb{E}(XY) \;=\; \lim_{n\to\infty} \mathbb{E}(X_n Y_n) \;=\; \lim_{n\to\infty} \mathbb{E}(X_n)\,\mathbb{E}(Y_n) \;=\; \mathbb{E}(X)\,\mathbb{E}(Y).$$

Schritt 3: Im Falle von $X, Y \in \mathbb{R}$ mit $\mathbb{E}\,|X|, \mathbb{E}\,|Y| < \infty$ kann man schreiben:

$$XY \;=\; X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+,$$

und nach Schritt 2 ist $\mathbb{E}(X^\pm Y^\pm) = \mathbb{E}(X^\pm)\,\mathbb{E}(Y^\pm) < \infty$. Daher ist

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}(X^+)\,\mathbb{E}(Y^+) + \mathbb{E}(X^-)\,\mathbb{E}(Y^-) - \mathbb{E}(X^+)\,\mathbb{E}(Y^-) - \mathbb{E}(X^-)\,\mathbb{E}(Y^+) \\
&= (\mathbb{E}(X^+) - \mathbb{E}(X^-))(\mathbb{E}(Y^+) - \mathbb{E}(Y^-)) \\
&= \mathbb{E}(X)\,\mathbb{E}(Y). \qquad \square
\end{aligned}$$

Im Allgemeinen kann man für Produkte von Zufallsvaraiblen nur gewisse Ungleichungen garantieren. Solche leiten wir nun im allgemeinem Rahmen her. Dabei definieren wir für eine messbare Funktion $f : \Omega \to \mathbb{R}$ und eine reelle Zahle $p \in [1, \infty)$ die Zahl

$$\|f\|_p = \|f\|_{p,M} \;:=\; \Big(\int |f|^p \, dM\Big)^{1/p}.$$

Im nächsten Abschnitt werden wir sehen, dass $\|\cdot\|_p$ eine *Seminorm* darstellt, doch jetzt betrachten wir sie nur als nützliche Kenngröße dafür, wie “groß” eine Funktion f absolut ist.

Satz 2.19 *Für beliebige $n \in \mathbb{N}$ und messbare Funktionen $f_1, f_2, \ldots, f_n : \Omega \to \mathbb{R}$ sowie Konstanten $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$ mit $\sum_{i=1}^n \lambda_i = 1$ gelten folgende Ungleichungen:*

$$\begin{aligned}
\int \prod_{i=1}^n |f_i| \, dM &\leq \sum_{i=1}^n \lambda_i \int |f_i|^{1/\lambda_i} \, dM \qquad (Minkowski-Ungleichung), \\
\int \prod_{i=1}^n |f_i| \, dM &\leq \prod_{i=1}^n \|f_i\|_{p_i} \quad \text{mit } p_i := \lambda_i^{-1} \qquad (Hölder-Ungleichung).
\end{aligned}$$

Ein wichtiger Spezialfall ist $n = 2$ und $\lambda_1 = \lambda_2 = 1/2$:

Korollar 2.20 ***(Cauchy-Schwarz-Bunjakowski)****.*

(a) *Für messbare Funktionen $f, g : \Omega \to \mathbb{R}$ mit $\|f\|_2, \|g\|_2 < \infty$ ist*

$$\int |fg| \, dM \;\leq\; \|f\|_2 \|g\|_2.$$

(b) *Für reellwertige Zufallsvariablen X, Y mit $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$ ist*

$$\mathbb{E}\,|XY| \;\le\; \sqrt{\mathbb{E}(X^2)\,\mathbb{E}(Y^2)}.$$

Hier noch eine andere Folgerung aus der Hölder-Ungleichung (siehe Übungen):

Korollar 2.21 ***(Lyapunov-Ungleichung).*** *Für eine reellwertige Zufallsvariable X und $0 < s < t$ ist*

$$\mathbb{E}(|X|^s)^{1/s} \;\le\; \mathbb{E}(|X|^t)^{1/t}.$$

Beweis von Satz 2.19. Da die Exponentialfunktion konvex ist, gilt punktweise die Ungleichung

$$\begin{aligned}
\prod_{i=1}^n |f_i| \;&=\; \exp\Big(\sum_{i=1}^n \log|f_i|\Big) \\
&=\; \exp\Big(\sum_{i=1}^n \lambda_i \big(p_i \log|f_i|\big)\Big) \\
&\le\; \sum_{i=1}^n \lambda_i \exp\big(p_i \log|f_i|\big) \\
&=\; \sum_{i=1}^n \lambda_i |f_i|^{p_i}.
\end{aligned}$$

(Die Zwischenschritte mit der Exponentialfunktion sind nur nötig, wenn alle $|f_i| > 0$.) Aufintegrieren beider Seiten bezüglich M liefert die Minkowski-Ungleichung.

Für den Beweis der Hölder-Ungleichung sei zunächst $\|f_i\|_{p_i} = 0$ für ein i. Dann ist $M\{f_i \neq 0\} = 0$, und beide Seiten der Ungleichung sind gleich Null; siehe Anmerkung 2.9. Sei nun also $\|f_i\|_{p_i} > 0$ für alle i. Im Falle von $\|f_k\|_{p_i} = \infty$ für mindestens ein i ist die Ungleichung trivial. Wir betrachten also nur noch den Fall, dass $0 < \|f_i\|_{p_i} < \infty$ für alle i. Mit $g_i := |f_i|/\|f_i\|_{p_i}$ ist nun zu zeigen, dass

$$\int \prod_{i=1}^n g_i \, dM \;=\; 1.$$

Doch nach der Minkowski-Ungleichung ist

$$\int \prod_{i=1}^n g_i \, dM \;\le\; \sum_{i=1}^n \lambda_i \int g_i^{p_i} \, dM \;=\; \sum_{i=1}^n \lambda_i \;=\; 1. \qquad \square$$

2.4 $\mathcal{L}^p$–Räume

Im vorigen Abschnitt tauchte bereits die Größe $\|f\|_p = \|f\|_{p,M}$ einer messbaren Funktion $f : \Omega \to \mathbb{R}$ auf, wobei $1 \le p < \infty$. In diesem Abschnitt betrachten wir die Menge

$$\mathcal{L}^p(M) \;:=\; \big\{ f : \Omega \to \mathbb{R} \text{ messbar mit } \|f\|_p < \infty \big\}.$$

Satz 2.22 *Der Raum $\mathcal{L}^p(M)$ ist ein Vektorraum über $\mathbb{R}$, und $\|\cdot\|_p$ definiert eine Seminorm hierauf. Das heißt, für beliebige $f, g \in \mathcal{L}^p(M)$ und $\lambda \in \mathbb{R}$ gilt:*

$$\begin{aligned} \|f\|_p &\geq 0, \\ \|\lambda f\|_p &= |\lambda| \|f\|_p, \\ \|f+g\|_p &\leq \|f\|_p + \|g\|_p. \end{aligned}$$

Ferner ist $\|f\|_p = 0$ genau dann, wenn $M\{f \neq 0\} = 0$.

Anmerkung 2.23 **(L^p–Räume).** Um einen richtigen normierten Raum zu erhalten, kann man im Raum $\mathcal{L}^p(M)$ die folgende Äquivalenzrelation einführen:

$$f \sim g \quad \text{genau dann, wenn} \quad M\{f \neq g\} = 0.$$

Mit den Äquivalenzklassen $[f] := \{g \in \mathcal{L}^p(M) : g \sim f\}$, $f \in \mathcal{L}^p$, erhält man einen Vektorraum

$$L^p(M) := \{[f] : f \in \mathcal{L}^p\},$$

und

$$\|[f]\|_p := \|f\|_p$$

definiert eine Norm auf $L^p(M)$.

Speziell ist $\|[f]\|_2 = \sqrt{\langle [f], [f] \rangle}$ mit dem Skalarprodukt

$$\langle [f], [g] \rangle = \langle [f], [g] \rangle_M := \int f g \, dM.$$

Somit erhalten wir einen Euklidischen Raum $\big(L^2(M), \langle\cdot,\cdot\rangle, \|\cdot\|_2\big)$.

Die Klammern von $[f]$ werden oft weggelassen. Spricht man also von Funktionen $f \in L^p(M)$, dann wird stillschweigend vereinbart, zwei Funktionen f und g als identisch zu betrachten, wenn $M\{f \neq g\} = 0$.

Beweis von Satz 2.22. Dass $\|f\|_p \geq 0$ ist offensichtlich, und die Äquivalenz von $\|f\|_p = 0$ und $M\{f \neq 0\}$ ergibt sich aus Anmerkung 2.9. Dass $\|\lambda f\|_p = |\lambda| \|f\|_p$ für $\lambda \in \mathbb{R}$ und $f \in \mathcal{L}^p(M)$, kann man direkt nachrechnen. Insbesondere ist auch $\lambda f \in \mathcal{L}^p(M)$.

Weniger offensichtlich ist die Dreiecksungleichung: Ohne Einschränkung betrachten wir nur den Fall, dass $\|f\|_p > 0$ und $\|g\|_p > 0$. Denn beispielsweise folgt aus $\|g\|_p = 0$, dass $f + g = f$ fast überall, also $\|f+g\|_p = \|f\|_p = \|f\|_p + \|g\|_p$.

Im Falle von $\|f\|_p > 0$ und $\|g\|_p > 0$ dürfen wir zusätzlich annehmen, dass $s := \|f\|_p + \|g\|_p = 1$.

Anderenfalls müsste man einfach f und g durch f/s und g/s ersetzen. Nun ist

$$\begin{aligned}
\int |f+g|^p\,dM &\leq \int (|f|+|g|)^p\,dM \\
&= \int \Big(\|f\|_p \frac{|f|}{\|f\|_p} + \|g\|_p \frac{|g|}{\|g\|_p}\Big)^p\,dM \\
&\leq \int \Big(\|f\|_p \frac{|f|^p}{\|f\|_p^p} + \|g\|_p \frac{|g|^p}{\|g\|_p^p}\Big)^p\,dM \\
&= \|f\|_p^{1-p} \int |f|^p\,dM + \|g\|_p^{1-p} \int |g|^p\,dM \\
&= \|f\|_p + \|g\|_p \;=\; 1.
\end{aligned}$$

Bei der letzten Ungleichung verwendeten wir die Konvexität der Funktion $[0,\infty) \ni x \mapsto x^p$. Insbesondere ist auch $f+g \in \mathcal{L}^p(M)$. □

Der folgende Satz impliziert, dass der normierte Raum $(L^p(M), \|\cdot\|_p)$ sogar *vollständig*, also ein *Banach-Raum* ist.

Satz 2.24 *Sei $(f_n)_n$ eine Cauchy-Folge in $(\mathcal{L}^p(M), \|\cdot\|_p)$; das heißt,*

$$\lim_{m,n\to\infty} \|f_m - f_n\|_p \;=\; 0.$$

Dann existiert eine Funktion $f \in \mathcal{L}^p(M)$ mit der Eigenschaft, dass

$$\lim_{n\to\infty} \|f_n - f\|_p \;=\; 0.$$

Beweis von Satz 2.24. Nach Voraussetzung existieren natürliche Zahlen $N_1 < N_2 < N_3 < \cdots$ derart, dass

$$\|f_m - f_n\|_p \;\leq\; 2^{-k} \quad \text{für alle } m,n \geq N_k.$$

Nun untersuchen wir die Teilfolge $(f_{N_k})_k$ auf Konvergenz. Dazu betrachten wir die Hilfsfunktion

$$g \;:=\; \sum_{k=1}^{\infty} |f_{N_{k+1}} - f_{N_k}|$$

mit Werten in $[0,\infty]$. Aus dem Satz von der monotonen Konvergenz und der Dreiecksungleichung für $\|\cdot\|_p$ ergibt sich nun, dass

$$\begin{aligned}
\int g^p\,dM &= \lim_{K\to\infty} \int \Big|\sum_{k=1}^{K} |f_{N_{k+1}} - f_{N_k}|\Big|^p\,dM \\
&= \lim_{K\to\infty} \Big\|\sum_{k=1}^{K} |f_{N_{k+1}} - f_{N_k}|\Big\|_p^p \\
&\leq \lim_{K\to\infty} \Big(\sum_{k=1}^{K} \big\|f_{N_{k+1}} - f_{N_k}\big\|_p\Big)^p \\
&\leq \lim_{K\to\infty} \Big(\sum_{k=1}^{K} 2^{-k}\Big)^p \;=\; 1.
\end{aligned}$$

Insbesondere ist $M\{g = \infty\} = 0$. Doch $g(\omega) < \infty$ impliziert, dass die Folge $(f_{N_k}(\omega))_{k=1}^\infty$ konvergiert. Wir definieren also

$$f(\omega) := \begin{cases} 0 & \text{falls } g(\omega) = \infty, \\ \lim_{k\to\infty} f_{N_k}(\omega) & \text{falls } g(\omega) < \infty. \end{cases}$$

Für beliebige $k \in \mathbb{N}$ und $\omega \in \Omega$ gilt die Ungleichung

$$\bigl|f(\omega) - f_{N_k}(\omega)\bigr| \;\le\; \sum_{\ell=k}^{\infty}\bigl|f_{N_{\ell+1}}(\omega) - f_{N_\ell}(\omega)\bigr|,$$

und wie beim Nachweis von $\int g^p\, dM \le 1$ kann man hieraus ableiten, dass

$$\|f - f_{N_k}\|_p \;\le\; 2^{1-k}.$$

Daher ist

$$\|f - f_n\|_p \;\le\; \|f - f_{N_k}\|_p + \|f_{N_k} - f_n\|_p \;\le\; 3 \cdot 2^{-k}$$

für beliebige $n \ge N_k$ und $k \ge 1$. □

Zuguterletzt zitieren wir noch einen Satz ohne Beweis.

Satz 2.25 ***(Scheffé).*** *Seien f und f_n ($n \in \mathbb{N}$) Funktionen in $\mathcal{L}^p(M)$. Dann ist*

$$\lim_{n\to\infty} \|f_n - f\|_p \;=\; 0$$

genau dann, wenn die folgenden zwei Bedingungen erfüllt sind:

(i) $(f_n)_n$ *konvergiert M–stochastisch gegen f, das heißt, für beliebige $\epsilon > 0$ ist*

$$\lim_{n\to\infty} M\bigl\{|f_n - f| \ge \epsilon\bigr\} \;=\; 0;$$

(ii)

$$\limsup_{n\to\infty} \|f_n\|_n \;\le\; \|f\|_p.$$

Der Spezialfall $p = 1$ wird bzw. wurde in den Übungen behandelt, wobei die “stochastische Konvergenz” (i) durch punktweise Konvergenz ersetzt wurde. Hier noch ein Korollar, das in manchen statistischen Anwendungen nützlich ist:

Korollar 2.26 *Seien f und f_n ($n \in \mathbb{N}$) Wahrscheinlichkeitsdichten bezüglich M. Dann folgt aus $\lim_{n\to\infty} f_n = f$, dass auch*

$$\lim_{n\to\infty} \int |f_n - f|\, dM \;=\; 0.$$

2.5 Varianzen und Kovarianzen

Im Rahmen der diskreten Wahrscheinlichkeitsräume wurden bereits Varianzen und Kovarianzen definiert. An diesen Definitionen und den wesentlichen Eigenschaften dieser Kenngrößen ändert sich nichts:

Definition 2.27 (Kovarianz, Varianz und Standardabweichung). Für zwei Zufallsvariablen $X, Y \in \mathcal{L}^2(P)$ definiert man ihre *Kovarianz* als die Zahl

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

Die *Varianz* von X ist definiert als

$$\mathrm{Var}(X) \ := \ \mathrm{Cov}(X,Y) \ = \ \mathbb{E}\big((X - \mathbb{E}\,X)^2\big) \ = \ \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

und die Standardabweichung von X ist definiert als

$$\mathrm{Std}(X) \ := \ \sqrt{\mathrm{Var}(X)}.$$

Anmerkung 2.28 Aus den allgemeinen Rechenregeln für Erwartungswerte sowie Satz 2.18 und der Cauchy-Schwarz-Bunjakowski-Ungleichung ergeben sich diverse Aussagen über Kovarianzen. Für beliebige Zufallsvariablen $X, Y, Z \in \mathcal{L}^2(P)$ und konstanten $a, b \in \mathbb{R}$ gilt:

- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$,
- $\mathrm{Cov}(X,Y) = 0$ falls X und Y stochastisch unabhängig sind,
- $|\mathrm{Cov}(X,Y)| \le \mathrm{Std}(X)\,\mathrm{Std}(Y)$,
- $\mathrm{Cov}(a + bX, Y) = b\,\mathrm{Cov}(X,Y)$,
- $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$,
- $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$.

Insbesondere ist

- $\mathrm{Var}(a + bX) = b^2\,\mathrm{Var}(X)$ und $\mathrm{Std}(a + bX) = |b|\,\mathrm{Std}(X)$,
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)$,
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ falls X und Y stochastisch unabhängig sind.

Aus Anmerkung 2.28 ergeben sich leicht die folgenden Resultate:

Lemma 2.29 *Für Zufallsvariablen $X_1, X_2, \ldots, X_n \in \mathcal{L}^2(P)$ und Konstanten $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ ist $\mathbb{E}\Big(\sum_{i=1}^n \lambda_i X_i\Big) = \sum_{i=1}^n \lambda_i\,\mathbb{E}(X_i)$, und*

$$\begin{aligned}\mathrm{Var}\Big(\sum_{i=1}^n \lambda_i X_i\Big) \ &= \ \sum_{i,j=1}^n \lambda_i\lambda_j\,\mathrm{Cov}(X_i, X_j) \\ &= \ \sum_{i=1}^n \lambda_i^2\,\mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \lambda_i\lambda_j\,\mathrm{Cov}(X_i, X_j).\end{aligned}$$

Korollar 2.30 *Für stochastisch unabhängige Zufallsvariablen $X_1, X_2, \ldots, X_n \in \mathcal{L}^2(P)$ mit Erwartungswert μ und Standardabweichung σ sei $\bar{X} := n^{-1} \sum_{i=1}^n X_i$. Dann ist*

$$\mathbb{E}(\bar{X}) \;=\; \mu \quad \textit{und} \quad \mathrm{Std}(\bar{X}) \;=\; \frac{\sigma}{\sqrt{n}}.$$

Korollar 2.30 liefert eine Begründung für die wiederholte Durchführung von Messungen und das Bilden des arithmetischen Mittelwertes. Bei n–maliger Wiederholung einer Einzelmessung erhöht sich die Präzision um den Faktor $\sqrt{n}$.

Aus Varianzen und Standardabweichungen lassen sich Ungleichungen für diverse Wahrscheinlichkeiten ableiten:

Lemma 2.31 ***(Tshebyshev-Cantelli-Ungleichungen)****. Für eine Zufallsvariable $X \in \mathcal{L}^2(P)$ und beliebige $\epsilon > 0$ ist*

$$\begin{aligned} P\big(|X - \mathbb{E}\,X| \ge \epsilon\big) \;&\le\; \frac{\mathrm{Var}(X)}{\epsilon^2} \qquad (\textit{Tshebyshev-Ungleichung}) \quad \textit{und} \\ P\big(\pm(X - \mathbb{E}\,X) \ge \epsilon\big) \;&\le\; \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + \epsilon^2} \qquad (\textit{Tshebyshev-Cantelli-Ungleichung}). \end{aligned}$$

Mit anderen Worten, für beliebige Konstanten $c > 0$ ist

$$\begin{aligned} P\big(|X - \mathbb{E}\,X| \ge c \cdot \mathrm{Std}(X)\big) \;&\le\; \frac{1}{c^2}, \\ P\big(\pm(X - \mathbb{E}\,X) \ge c \cdot \mathrm{Std}(X)\big) \;&\le\; \frac{1}{1 + c^2}. \end{aligned}$$

Beweis von Lemma 2.31. Die erstgenannte Ungleichung ergibt sich aus der punktweisen Ungleichung

$$1\big\{|X - \mathbb{E}\,X| \ge \epsilon\big\} \;\le\; \frac{(X - \mathbb{E}\,X)^2}{\epsilon^2}$$

und Bilden des Erwartungswertes beider Seiten.

Für den Beweis des zweiten Teils sei $Y := \pm(X - \mathbb{E}\,X)$. Dann ist

$$1\{Y \ge \epsilon\} \;\le\; \frac{(Y - t)^2}{(\epsilon - t)^2} \quad \text{für beliebige } t < \epsilon,$$

und aus $\mathbb{E}\,Y = 0$ folgt, dass

$$P(Y \ge \epsilon) \;\le\; \frac{\mathbb{E}(Y^2) + t^2}{(\epsilon - t)^2} \quad \text{für beliebige } t < \epsilon.$$

Nun muss man nur noch die rechte Seite bezüglich $t < \epsilon$ minimieren …

Ein alternativer Beweis des zweiten Teils: Mit $Y = \pm(X - \mathbb{E}\,X)$ ergibt sich aus der Cauchy-Schwarz-Ungleichung, dass

$$\begin{aligned} \epsilon &= \mathbb{E}(\epsilon - Y) \\ &\le \mathbb{E}\big((\epsilon - Y)1\{Y \le \epsilon\}\big) \\ &\le \sqrt{\mathbb{E}\big((\epsilon - Y)^2\big)P(Y \le \epsilon)} \\ &= \sqrt{(\mathrm{Var}(X) + \epsilon^2)P(Y \le \epsilon)}, \end{aligned}$$

und die Behauptung ergibt sich durch elementare Umformungen dieser Ungleichung. □

Anwendung: Der Approximationssatz von Weierstrass

Nachdem wir so viele Resultate aus der Analysis verwendet haben, ist es vielleicht an der Zeit, den Analytikern etwas zu schenken, nämlich einen eleganten Beweis der Tatsache, dass jede stetige Funktion $f : [0,1] \to \mathbb{R}$ gleichmäßig durch ganzrationale Funktionen approximiert werden kann. Dabei messen wir Abweichungen in der Supremumsnorm

$$\|h\|_\infty := \sup_{x \in [0,1]} |h(x)| \quad \text{für } h : [0,1] \to \mathbb{R}.$$

Der Beweis ist konstruktiv: Für eine natürliche Zahl n sei $B_n f$ das *n–te Bernsteinpolynom von f*,

$$B_n f(x) := \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\binom{n}{k} x^k (1-x)^{n-k}.$$

Offensichtlich ist dies ein Polynom vom Grad höchstens n. Für $0 \le p \le 1$ kann man $B_n f(p)$ schreiben als

$$B_n f(p) = \mathbb{E}\, f\Big(\frac{Y_p}{n}\Big) \quad \text{mit } Y_p \sim \mathrm{Bin}(n,p).$$

Da $\mathbb{E}(Y_p/n) = p$ und $\mathrm{Var}(Y_p/n) = p(1-p)/n \le 1/(4n)$, sollte $B_n f(p) = \mathbb{E}\, f(Y_p/n)$ nahe an $f(p)$ sein. Der folgende Satz präzsiert diese Aussage.

Satz 2.32 ***(Weierstrass-Bernstein).***

(a) *Sei f eine stetige Funktion auf $[0,1]$. Dann ist*

$$\lim_{n \to \infty} \|B_n f - f\|_\infty = 0.$$

(b) *Sei f sogar Lipschitz-stetig mit Konstante L. Das heißt, $|f(x) - f(y)| \le L|x-y|$ für alle $x, y \in [0,1]$. Dann ist*

$$\|B_n f - f\|_\infty \le \frac{L}{2\sqrt{n}}.$$

(c) *Sei f zweimal differenzierbar auf $[0,1]$ mit Ableitungen f' und f'', so dass $|f''| \le M$ für eine Konstante M. Dann ist*

$$\|B_n f - f\|_\infty \le \frac{M}{8n}.$$

Beweis von Satz 2.32. Mit $Y_p \sim \mathrm{Bin}(n, p)$ und $\widehat{p} := Y_p/n$ ist $B_n f(p) = \mathbb{E}\, f(\widehat{p})$.

Beweis von Teil (a). Für $r > 0$ sei

$$\Delta(f, r) \ := \sup_{y,z\in[0,1]\,:\,|y-z|\le r} |f(y) - f(z)|,$$

die maximale Fluktuation von f auf Intervallen der Länge höchstens r. Für beliebiges $\epsilon > 0$ ist dann

$$\begin{aligned}
\big|B_n f(p) - f(p)\big| &= \Big|\mathbb{E}\big(f(\widehat{p}) - f(p)\big)\Big| \\
&\le \mathbb{E}\big|f(\widehat{p}) - f(p)\big| \\
&= \mathbb{E}\Big(1\{|\widehat{p} - p| < \epsilon\}\,\big|f(\widehat{p}) - f(p)\big|\Big) + \mathbb{E}\Big(1\{|\widehat{p} - p| \ge \epsilon\}\,\big|f(\widehat{p}) - f(p)\big|\Big) \\
&\le \Delta(f, \epsilon) + \mathbb{E}\Big(1\{|\widehat{p} - p| \ge \epsilon\}\Delta(f, 1)\Big) \\
&= \Delta(f, \epsilon) + \Delta(f, 1) P\big(|\widehat{p} - p| \ge \epsilon\big) \\
&\le \Delta(f, \epsilon) + \frac{\Delta(f, 1)}{4n\epsilon^2}.
\end{aligned}$$

Dabei verwendeten wir im letzten Schritt die Tshebyshev-Ungleichung zusammen mit der Schranke $\mathrm{Var}(\widehat{p}) \le 1/(4n)$. Wir wissen also, dass

$$\|B_n f - f\|_\infty \ \le\ \Delta(f, \epsilon) + \frac{\Delta(f, 1)}{4n\epsilon^2}.$$

Setzt man beispielsweise $\epsilon = \epsilon_n := n^{-1/3}$, dann konvergiert die rechte Seite gegen Null für $n \to \infty$. Dabei verwenden wir die Tatsache, dass eine auf dem kompakten Intervall $[0, 1]$ stetige Funktion dort automatisch *gleichmäßig stetig* ist; das heißt, $\lim_{r\to 0} \Delta(f, r) = 0$.

Beweis von Teil (b). Aus der Voraussetzung an f folgt, dass $|f(\widehat{p}) - f(p)| \le L|\widehat{p} - p|$, weshalb

$$|B_n f(p) - f(p)| \ \le\ \mathbb{E}\,(L|\widehat{p} - p|) \ \le\ L\sqrt{\mathrm{Var}(\widehat{p})} \ \le\ \frac{L}{2\sqrt{n}}.$$

Beweis von Teil (c). Nach der Taylorformel ist

$$f(\widehat{p}) \ =\ f(p) + f'(p)(\widehat{p} - p) + \frac{f''(\widehat{\xi})}{2}\,(\widehat{p} - p)^2$$

für eine (zufällige) Zwischenstelle $\widehat{\xi} \in [0, 1]$. Folglich ist

$$\begin{aligned}
|B_n f(p) - f(p)| &= \Big|\mathbb{E}\big(f(\widehat{p}) - f(p)\big)\Big| \\
&= \Big|\mathbb{E}\Big(f'(p)(\widehat{p} - p) + \frac{f''(\widehat{\xi})}{2}\,(\widehat{p} - p)^2\Big)\Big| \\
&= \Big|f'(p)\underbrace{\mathbb{E}(\widehat{p} - p)}_{=0} + \mathbb{E}\Big(\frac{f''(\widehat{\xi})}{2}\,(\widehat{p} - p)^2\Big)\Big| \\
&\le \mathbb{E}\Big(\frac{M}{2}\,(\widehat{p} - p)^2\Big) \\
&= \frac{M}{2}\,\mathrm{Var}(\widehat{p}) \\
&\le \frac{M}{8n}. \qquad \square
\end{aligned}$$

Wir illustrieren die Bernsteinpolynome $B_n f$ für drei verschiedene Funktionen f. In allen folgenden Abbildungen wird der Graph von f durch eine dünne Linie dargestellt, die Stützstellen $(k/n, f(k/n))$ für $B_n f$ werden etwas hervorgehoben, und der Graph von $B_n f$ ist die kräftigere Linie.

Unser erstes Beispiel für f ist

$$f(x) \ := \ |x - 1/2|. \tag{2.7}$$

Die Voraussetzung von Theorem 2.32 (b) ist erfüllt mit $L = 1$. Abbildung 2.1 zeigt zwei verschiedene Bernsteinpolynome.

Das zweite Beispiel für f ist

$$f(x) \ := \ \frac{1}{\exp(5 - 10x) + 1}. \tag{2.8}$$

Nun sind sogar die stärkeren Voraussetzungen von Teil (c) erfüllt. Dies spiegelt sich auch in deutlich kleineren Approximationsfehlern von $B_n f$ wieder; siehe Abbildung 2.2.

Als drittes Beispiel betrachten wir

$$f(x) \ := \ x \sin\Bigl(\frac{4\pi}{\sqrt{x}}\Bigr) \tag{2.9}$$

mit $f(0) := 0$. Diese Funktion ist stetig, erfüllt aber nicht die Glattheitsbedingungen von Teil (b) oder (c). In der Tat benötigt man deutlich größere Zahlen n um eine gute Approximation zu erhalten; siehe Abbildung 2.3.

2.6 Das schwache Gesetz der großen Zahlen

Wir behandeln nun ein erstes zentrales Resultat der Wahrscheinlichkeitstheorie, und zwar in einer speziellen und einer sehr allgemeinen Formulierung:

Satz 2.33 ***(Schwaches Gesetz der großen Zahlen, I).*** *Seien $X_1, X_2, X_3, \ldots$ stochastisch unabhängige, identisch verteilte Zufallsvariablen in $\mathcal{L}^1(P)$ mit Erwartungswert μ. Für den Mittelwert $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i$ ist dann*

$$\lim_{n \to \infty} \mathbb{E} \bigl| \bar{X}_n - \mu \bigr| \ = \ 0.$$

Insbesondere ist

$$\lim_{n \to \infty} P\Bigl(\bigl| \bar{X}_n - \mu \bigr| \ge \epsilon \Bigr) \ = \ 0$$

für beliebige $\epsilon > 0$.

In vielen statistischen Anwendungen betrachtet man komplexere Situationen, für die dieser Satz zu speziell ist. Daher verallgemeinern wir Satz 2.33 wie folgt:

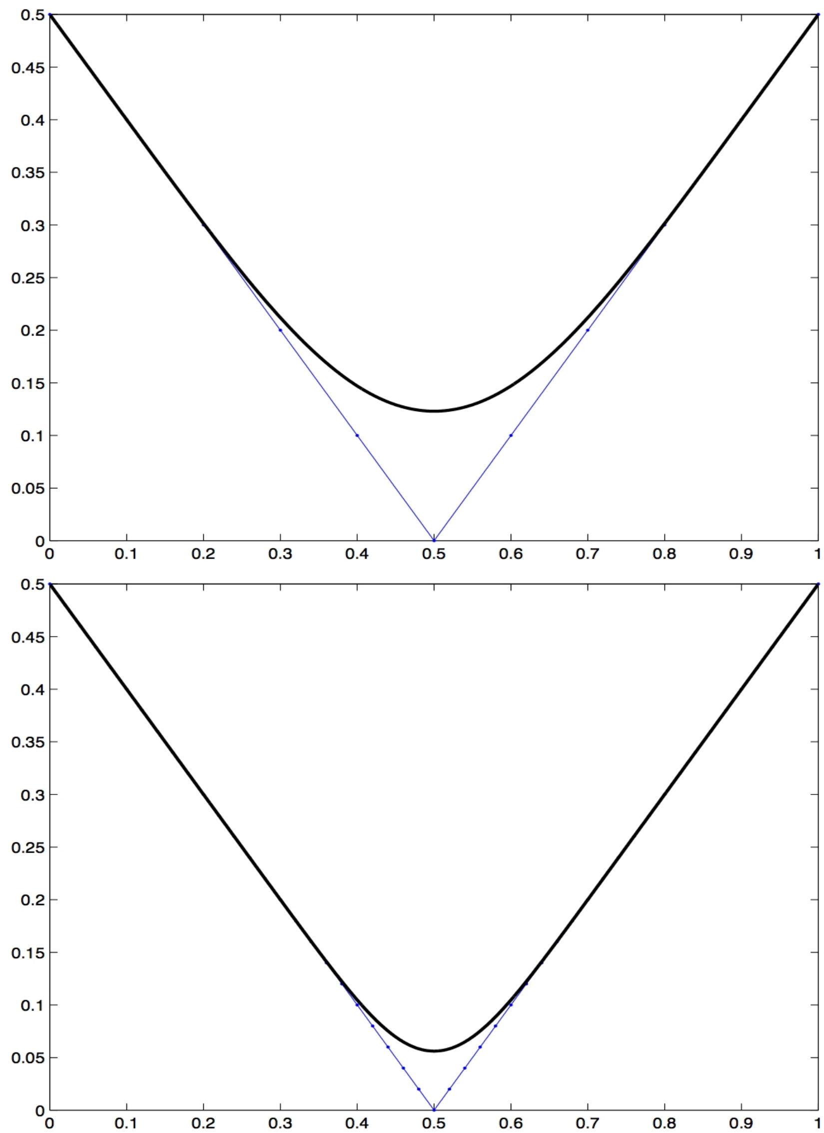

Abbildung 2.1: $B_n f$ für f in (2.7) und $n = 10, 50$

Satz 2.34 ***(Schwaches Gesetz der großen Zahlen, II).*** *Für $n \in \mathbb{N}$ seien $X_{n1}, X_{n2}, \ldots, X_{nn}$ stochastisch unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega_n, \mathcal{A}_n, P_n)$, so dass die folgenden beiden Bedingungen erfüllt sind:*

$$C := \sup_{n \geq 1} \sum_{i=1}^{n} \mathbb{E}\,|X_{ni}| \;<\; \infty; \tag{2.10}$$

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\big(1\{|X_{ni}| > \epsilon\}|X_{ni}|\big) \;=\; 0 \quad \textit{für alle } \epsilon > 0. \tag{2.11}$$

Dann ist

$$\lim_{n\to\infty} \mathbb{E}\Big|\sum_{i=1}^{n}(X_{ni} - \mathbb{E}\,X_{ni})\Big| \;=\; 0.$$

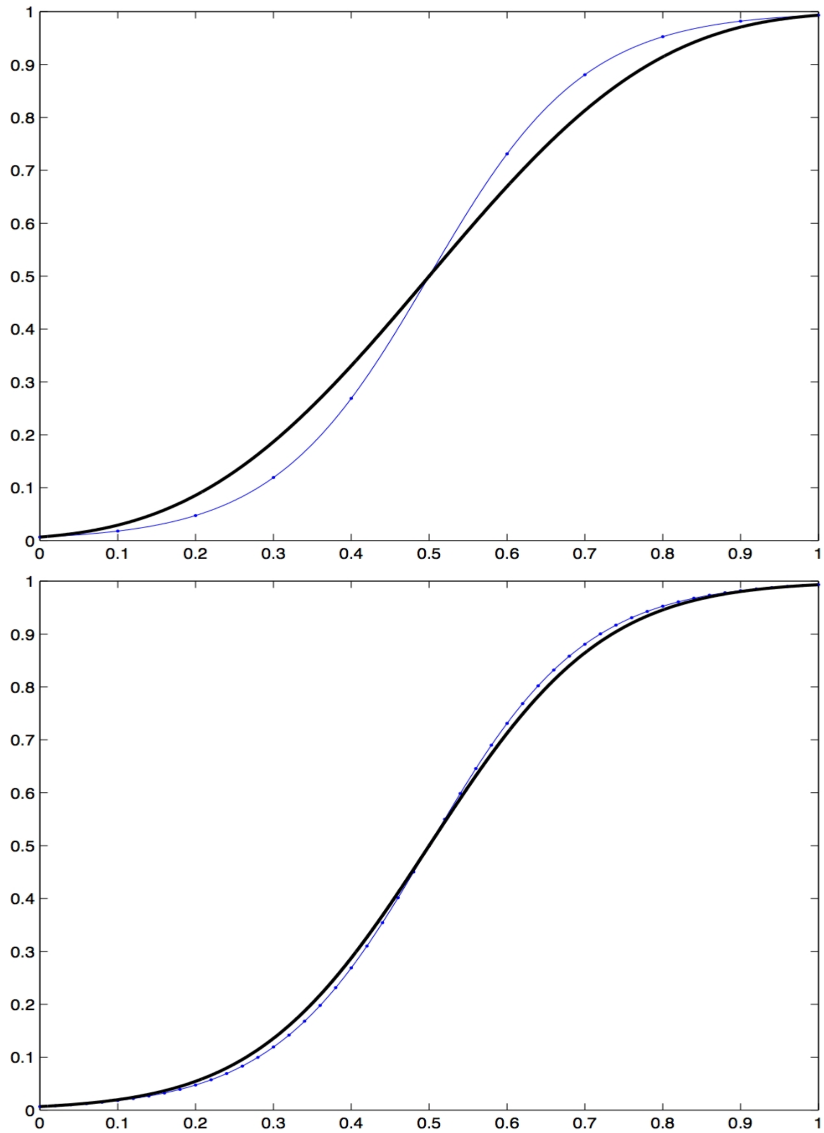

Abbildung 2.2: $B_n f$ für f in (2.8) und $n = 10, 50$

Beweis von Satz 2.33. Für $n \in \mathbb{N}$ und $1 \leq i \leq n$ definieren wir $X_{ni} := n^{-1} X_i$. Dann ist

$$\bar{X}_n - \mu \;=\; \sum_{i=1}^{n} (X_{ni} - \mathbb{E}\, X_{ni}),$$

und die Behauptung ergibt sich aus Satz 2.34. Denn

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}\,|X_{ni}| &= \mathbb{E}\,|X_1|, \\
\sum_{i=1}^{n} \mathbb{E}\big(1\{|X_{ni}| > \epsilon\}|X_{ni}|\big) &= \mathbb{E}\big(1\{|X_1| > n\epsilon\}|X_1|\big) \\
&= \int 1\{|X_1| > n\epsilon\}|X_1|\, dP \;\to\; 0 \quad (n \to \infty).
\end{aligned}$$

Letztere Tatsache folgt aus dem Satz von der majorisierten Konvergenz, angewandt auf die Funktionen $f := 0$ und $f_n := 1\{|X_1| > n\epsilon\}|X_1|$ mit der Majorante $g := |X_1| \in \mathcal{L}^1(P)$. $\square$

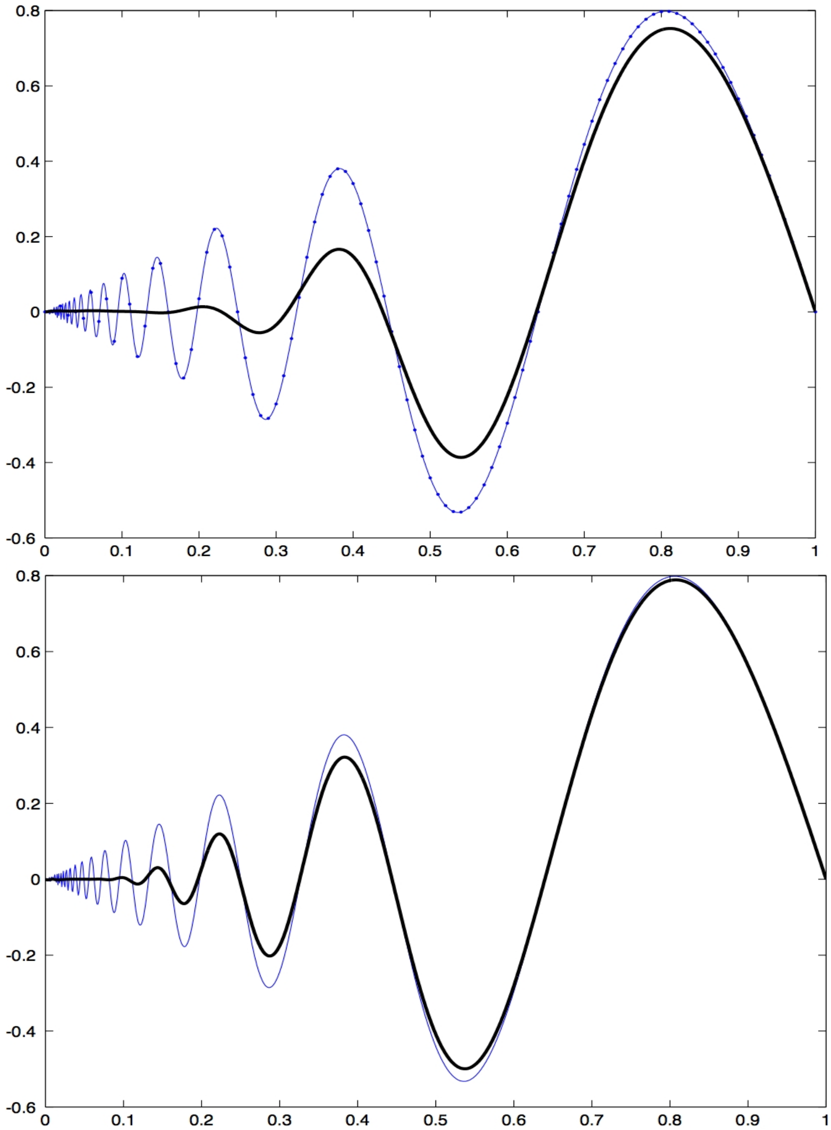

Abbildung 2.3: $B_n f$ für f in (2.9) und $n = 100, 500$

Beweis von Satz 2.34. Für eine beliebig kleine Konstante $\epsilon > 0$ zerlegen wir X_{ni} in $Y_{ni} + Z_{ni}$ mit

$$Y_{ni} \ := \ 1\{|X_{ni}| \le \epsilon\}|X_{ni}| \quad \text{und} \quad Z_{ni} \ := \ 1\{|X_{ni}| > \epsilon\}|X_{ni}|.$$

Dann ist

$$\begin{aligned}
\mathbb{E}\Big|\sum_{i=1}^{n}(X_{ni} - \mathbb{E}\,X_{ni})\Big| \ &= \ \mathbb{E}\Big|\sum_{i=1}^{n}(Y_{ni} - \mathbb{E}\,Y_{ni}) + \sum_{i=1}^{n}(Z_{ni} - \mathbb{E}\,Z_{ni})\Big| \\
&\le \ \mathbb{E}\Big|\sum_{i=1}^{n}(Y_{ni} - \mathbb{E}\,Y_{ni})\Big| + \sum_{i=1}^{n}\big(\mathbb{E}\,|Z_{ni}| + |\,\mathbb{E}\,Z_{ni}|\big) \\
&\le \ \mathbb{E}\Big|\sum_{i=1}^{n}(Y_{ni} - \mathbb{E}\,Y_{ni})\Big| + 2\sum_{i=1}^{n}\mathbb{E}\,|Z_{ni}|.
\end{aligned}$$

Doch nach Voraussetzung (2.11) konvergiert $\sum_{i=1}^n \mathbb{E}\,|Z_{ni}|$ gegen Null für $n \to \infty$. Außerdem folgt aus der Cauchy-Schwarz-Ungleichung und Voraussetzung (2.10), dass

$$\begin{aligned}\mathbb{E}\Big|\sum_{i=1}^n (Y_{ni} - \mathbb{E}\,Y_{ni})\Big| &\le \Big(\mathrm{Var}\Big(\sum_{i=1}^n Y_{ni}\Big)\Big)^{1/2} \\ &= \Big(\sum_{i=1}^n \mathrm{Var}(Y_{ni})\Big)^{1/2} \\ &\le \Big(\sum_{i=1}^n \mathbb{E}(Y_{ni}^2)\Big)^{1/2} \\ &\le \Big(\sum_{i=1}^n \mathbb{E}(\epsilon |X_{ni}|)\Big)^{1/2} \\ &\le \big(\epsilon C\big)^{1/2}.\end{aligned}$$

Also ist

$$\limsup_{n\to\infty} \mathbb{E}\Big|\sum_{i=1}^n (X_{ni} - \mathbb{E}\,X_{ni})\Big| \;\le\; (C\epsilon)^{1/2},$$

und für $\epsilon \downarrow 0$ ergibt sich die Behauptung. □

2.7 Momenterzeugende Funktionen und Exponentialungleichungen

In der diskreten Wahrscheinlichkeitstheorie wurde die *erzeugende Funktion* einer Zufallsvariable X mit Werten in $\mathbb{N}_0$ eingeführt. Für beliebig reellwertige Zufallsvariablen gibt es eine ähnliche Hilfsfunktion:

Definition 2.35 (Momenterzeugende Funktion). Die *momenterzeugende Funktion* einer reellwertigen Zufallsvariable X ist definiert als die Funktion $m_X : \mathbb{R} \to (0, \infty]$ mit

$$m_X(t) \;:=\; \mathbb{E}\exp(tX).$$

Analog definiert man die momenterzeugende Funktion eines Wahrscheinlichkeitsmaßes P auf $\mathbb{R}$ durch

$$m_P(t) \;:=\; \int \exp(tx)\,P(dx).$$

Die Funktion $t \mapsto m_X(-t)$ bzw. $t \mapsto m_P(-t)$ nennt man auch die *Laplace–Transformierte* von X bzw. P.

Anmerkung 2.36 (Existenz beliebiger Momente und deren Rekonstruktion aus m_X). Wenn $m_X(\pm\epsilon) < \infty$ für ein $\epsilon > 0$, dann ist

$$\infty \;>\; \mathbb{E}\big(\exp(-\epsilon X) + \exp(\epsilon X)\big) \;\ge\; \mathbb{E}\exp(\epsilon|X|) \;=\; \sum_{j=0}^\infty \frac{\epsilon^j\,\mathbb{E}(|X|^j)}{j!}.$$

Insbesondere ist $\mathbb{E}(|X|^k) < \infty$ für alle natürlichen Zahlen k, und für $|t| \leq \epsilon$ gilt die Potenzreihenentwicklung

$$m_X(t) \; = \; \sum_{j=0}^{\infty} \frac{\mathbb{E}(X^j)}{j!} \, t^j.$$

Für das k–*te Moment von* X ergibt sich nun die Gleichung

$$\mathbb{E}(X^k) \; = \; \frac{d^k}{dt^k}\Big|_{t=0} m_X(t).$$

Satz 2.37 ***(Eigenschaften momenterzeugender Funktionen).***

(a) *Angenommen,* $m_P(r) < \infty$ *und* $m_P(s) < \infty$ *für reelle Zahlen* $r < s$. *Dann ist* m_P *auf* $[r, s]$ *stetig (insb. reellwertig) und auf* (r, s) *unendlich oft differenzierbar mit*

$$\frac{d^k}{dt^k} m_P(t) \; = \; \int \exp(tx) x^k \, P(dx).$$

(b) *Für zwei stochastisch unabhängige Zufallsvariablen* X *und* Y *auf* $(\Omega, \mathcal{A}, P)$ *ist*

$$m_{X+Y} \; = \; m_X m_Y.$$

(c) *Sind* P *und* Q *Wahrscheinlichkeitsmaße auf* $\mathbb{R}$, *so dass* $m_P = m_Q < \infty$ *auf einem nichtleeren Intervall* (r, s), *dann ist* $P = Q$.

Beispiel 2.38 **(Gamma-Verteilungen)**. Sei $\mathrm{Gamma}(a, \sigma)$ die Gammaverteilung mit Parametern $a, \sigma > 0$. Das heißt, diese Verteilung wird durch die Dichtefunktion

$$f_{a,\sigma}(x) \; := \; 1\{x > 0\} \sigma^{-1} \Gamma(a)^{-1} (x/\sigma)^{a-1} \exp(-x/\sigma)$$

beschrieben, wobei $\Gamma(a) := \int_0^\infty x^{a-1} e^{-x} \, dx$. Hier ist

$$m_{\mathrm{Gamma}(a,\sigma)}(t) \; = \; \begin{cases} (1 - \sigma t)^{-a} & \text{falls } t < 1/\sigma, \\ \infty & \text{falls } t \geq 1/\sigma, \end{cases}$$

denn

$$\begin{aligned} m_{\mathrm{Gamma}(a,\sigma)}(t) \; &= \; \sigma^{-1} \Gamma(a)^{-1} \int_0^\infty \exp(tx) (x/\sigma)^{a-1} \exp(-x/\sigma) \, dx \\ &= \; \sigma^{-1} \Gamma(a)^{-1} \int_0^\infty (x/\sigma)^{a-1} \exp(-(1 - \sigma t) x/\sigma) \, dx \\ &= \; \Gamma(a)^{-1} \int_0^\infty y^{a-1} e^{-(1-\sigma t)y} \, dy. \end{aligned}$$

Dieses Integral divergiert für $t \geq 1/\sigma$, und im Falle von $t < 1/\sigma$ ergibt sich die behauptete Gleichung durch die Variablensubstitution $z = (1 - \sigma t)y$.

Nun ist

$$\frac{d}{dt}(1 - \sigma t)^{-a} \; = \; a\sigma(1 - \sigma t)^{-(a+1)} \quad \text{und} \quad \frac{d^2}{dt^2}(1 - \sigma t)^{-a} \; = \; a(a+1)\sigma^2(1 - \sigma t)^{-(a+2)}$$

für $t < 1/\sigma$, so dass

$$\mathbb{E}(X) \ = \ a\sigma, \quad \mathbb{E}(X^2) \ = \ a(a+1)\sigma^2 \quad \text{und} \quad \mathrm{Var}(X) \ = \ a\sigma^2.$$

Für zwei stochastisch unabhängige Zufallsvariablen X und Y mit Verteilung $\mathrm{Gamma}(a,\sigma)$ bzw. $\mathrm{Gamma}(b,\sigma)$ ist die Summe $X+Y$ verteilt nach $\mathrm{Gamma}(a+b,\sigma)$. Denn für $t < 1/\sigma$ ist

$$\begin{aligned} m_{X+Y}(t) &= m_X(t)m_Y(t) \\ &= (1-t\sigma)^{-a}(1-t\sigma)^{-b} \\ &= (1-t\sigma)^{-(a+b)} \\ &= m_{\mathrm{Gamma}(a+b,\sigma)}(t). \end{aligned}$$

Teil (c) von Satz 2.37 (c) werden wir nur in einem Spezialfall beweisen. Dabei benötigen wir das folgende Lemma:

Lemma 2.39 *Für beliebige reelle Zahlen $a' < a < b < b'$ sowie ganze Zahlen $k \geq 0$ gibt es eine k–mal stetig differenzierbare Funktion f auf $\mathbb{R}$ derart, dass*

$$1_{[a,b]} \ \leq \ f \ \leq \ 1_{(a',b')}.$$

Beweis von Lemma 2.39. Wir betrachten die Hilfsfunktion

$$h(x) \ := \ 1\{0 \leq x \leq 1\}(2k+1)\binom{2k}{k}x^k(1-x)^k.$$

Diese ist nichtnegativ, $(k-1)$–mal stetig differenzierbar, und $\int_{\mathbb{R}} h(x)dx = 1$; siehe (2.6). Nun definieren wir

$$f(x) \ := \ \int_{-\infty}^{x} \Big(\frac{1}{a-a'}h\Big(\frac{t-a'}{a-a'}\Big) - \frac{1}{b'-b}h\Big(\frac{t-b}{b'-b}\Big)\Big)\,dt.$$

Nun kann man sich leicht davon überzeugen, dass f die gewünschten Eigenschaften hat. $\square$

Beweis von Satz 2.37. Teil (a) wird bzw. wurde in den Übungen behandelt. Er folgt im Wesentlichen aus dem Satz von der majorisierten Konvergenz, denn formales k–faches Ableiten von $m_P(\cdot)$ auf (r,s) ergibt

$$\int \frac{d^k}{dt^k}\exp(tx)\,P(dx) \ = \ \int x^k \exp(tx)\,P(dx).$$

Auf der Suche nach geeigneten Majoranten kann man beispielsweise ausnutzen, dass für beliebige Zahlen $r < t < s$ und $k \in \mathbb{N}_0$ gilt:

$$\big|x^k \exp(tx)\big| \ \leq \ \frac{k!}{(t-r)^k}\ \exp(rx) + \frac{k!}{(s-t)^k}\ \exp(sx).$$

Teil (b) ergibt sich aus Satz 2.18. Denn mit X und Y sind auch $\exp(tX)$ und $\exp(tY)$ stochastisch unabhängige, nichtnegative Zufallsvariablen, so dass

$$\begin{aligned} m_{X+Y}(t) &= \mathbb{E}\exp(t(X+Y)) \;=\; \mathbb{E}\big(\exp(tX)\exp(tY)\big) \;=\; \mathbb{E}\exp(tX)\,\mathbb{E}\exp(tY) \\ &= m_X(t)m_Y(t). \end{aligned}$$

Teil (c) beweisen wir nur in dem Spezialfall, dass $P([0,\infty)) = Q([0,\infty)) = 1$ und $m_P = m_Q$ auf $(-\infty, 0]$. Um zu zeigen, dass $P = Q$, genügt es zu zeigen, dass $P([0,b]) = Q([0,b])$ für beliebige reelle Zahlen $b \geq 0$. Nach Lemma 2.39 genügt es sogar zu zeigen, dass

$$\int f\,dP \;=\; \int f\,dQ$$

für eine beliebige stetige Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $\lim_{|x|\to\infty} f(x) = 0$. Definieren wir nun $\widetilde{f}(p) := f(-\log p)$ für $p \in (0,1]$ und $\widetilde{f}(0) := 0$, dann ist $\widetilde{f}$ stetig auf $[0,1]$, und nach dem Weierstrassschen Approximationssatz existiert zu beliebigem $\delta > 0$ ein Polynom der Form $\widetilde{g}(p) := \sum_{k=0}^n a_k p^k$ derart, dass

$$\sup_{p\in[0,1]} |\widetilde{g}(p) - \widetilde{f}(p)| \;\leq\; \delta.$$

Mit $g(x) := \sum_{k=0}^n a_k \exp(-kx)$ ist also

$$\sup_{x\geq 0} |g(x) - f(x)| \;\leq\; \delta.$$

Folglich ist

$$\begin{aligned} \Big|\int f\,dP - \int f\,dQ\Big| &\leq 2\delta + \Big|\int g\,dP - \int g\,dQ\Big| \\ &= 2\delta + \Big|\sum_{k=0}^n a_k m_P(-k) - \sum_{k=0}^n a_k m_Q(-k)\Big| \\ &= 2\delta, \end{aligned}$$

und für $\delta \downarrow 0$ ergibt sich die Behauptung.

Den allgemeinen Fall kann man behandeln, indem man den Definitionsbereich von m_P und m_Q auf die Teilmenge $\big\{z \in \mathbb{C} : r < \mathrm{Re}(z) < s\big\}$ ausdehnt und gewisse Resultate über holomorphe Funktionen (Komplexe Analysis) und charakteristische Funktionen (s. späteres Kapitel) ausnutzt. $\square$

Anwendung: Exponentialungleichungen

Momenterzeugende Funktionen verwendet man häufig, um verfeinerte Abschätzungen von Wahrscheinlichkeiten zu erzielen. Auch hierfür verwendet man die Markov-Ungleichung:

Lemma 2.40 *Sei X eine Zufallsvariable mit Erwartungswert $\mu \in \mathbb{R}$. Dann gilt für beliebige $\eta > 0$:*

$$P\big(\pm(X-\mu) \geq \eta\big) \;\leq\; \inf_{t>0} m_{X-\mu}(\pm t)\exp(-t\eta) \;=\; \inf_{t>0} m_X(\pm t)\exp(-t(\pm\mu+\eta)).$$

Beweis von Lemma 2.40. Für beliebige $t > 0$ gilt punktweise die Ungleichung

$$\begin{aligned} 1\big\{\pm(X-\mu) \geq \eta\big\} &\leq \exp\big(t\big(\pm(X-\mu)-\eta\big)\big) \\ &= \exp\big(\pm t(X-\mu)\big)\exp(-t\eta) \\ &= \exp(\pm tX)\exp\big(-t(\pm\mu+\eta)\big). \end{aligned}$$

Nun ergibt sich die Behauptung durch Bilden der Erwartungswerte. □

Mit Hilfe dieses Lemmas kann man die schon früher verwendete Hoeffdingsche Ungleichung beweisen:

Satz 2.41 ***(Hoeffding).*** *Seien $X_1, X_2, \ldots, X_n$ stochastisch unabhängige Zufallsvariablen mit Werten in $[0,1]$. Dann gelten für $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ die Ungleichungen*

$$\mathbb{E}\exp\big(t(\bar{X} - \mathbb{E}\,\bar{X})\big) \;\leq\; \exp\Big(\frac{t^2}{8n}\Big) \qquad \textit{für alle } t \in \mathbb{R} \tag{2.12}$$

und

$$P\big(\pm(\bar{X} - \mathbb{E}\,\bar{X}) \geq \eta\big) \;\leq\; \exp(-2n\eta^2) \qquad \textit{für alle } \eta \geq 0. \tag{2.13}$$

Beweis von Satz 2.41. Zunächst zeigen wir, dass (2.13) aus (2.12) folgt: Nach Lemma 2.40 und (2.12) ist

$$\begin{aligned} P\big(\pm(\bar{X} - \mathbb{E}\,\bar{X}) \geq \eta\big) &\leq \inf_{t>0} \mathbb{E}\exp\big(\pm t(\bar{X} - \mathbb{E}\,\bar{X})\big)\exp(-t\eta) \\ &\leq \inf_{t>0} \exp\Big(\frac{t^2}{8n} - t\eta\Big) \\ &= \exp\Big(\inf_{t>0} \frac{t^2}{8n} - t\eta\Big) \\ &= \exp(-2n\eta^2) \qquad (\text{mit } t = 4n\eta). \end{aligned}$$

Nun zum Nachweis von Ungleichung (2.12): Für jedes $t \in \mathbb{R}$ folgt aus der stochastischen Unabhängigkeit der Variablen X_i, dass

$$\begin{aligned} \mathbb{E}\exp\big(t(\bar{X} - \mathbb{E}\,\bar{X})\big) &= \mathbb{E}\exp\Big(\frac{t}{n}\sum_{i=1}^n (X_i - \mathbb{E}\,X_i)\Big) \\ &= \mathbb{E}\prod_{i=1}^n \exp\Big(\frac{t}{n}\,X_i\Big) \\ &= \prod_{i=1}^n \mathbb{E}\exp\Big(\frac{t}{n}\,X_i\Big). \end{aligned}$$

Daher genügt es, folgende Ungleichung zu beweisen: Für eine Zufallsvariable X mit Werten in $[0,1]$ und Erwartungswert μ ist

$$\mathbb{E}\exp(s(X-\mu)) \;\leq\; \exp(s^2/8) \quad \text{für alle } s \in \mathbb{R}.$$

Aus der Konvexität der Exponentialfunktion ergibt sich, dass auch $x \mapsto \exp(s(x-\mu))$ konvex ist, so dass

$$\begin{aligned} \exp(s(x-\mu)) &\leq (1-x)\exp(s(0-\mu)) + x\exp(s(1-\mu)) \\ &= (1-x)\exp(-s\mu) + x\exp(s(1-\mu)) \quad \text{für alle } x \in [0,1]. \end{aligned}$$

Bildet man nun den Erwartungswert beider Seiten, ergibt sich die Ungleichung

$$\mathbb{E}\exp(s(X-\mu)) \;\leq\; (1-\mu)\exp(-s\mu) + \mu\exp(s(1-\mu)).$$

Mit

$$h(s) \;:=\; \log\bigl((1-\mu)\exp(-s\mu) + \mu\exp(s(1-\mu))\bigr) \;=\; -\mu s + \log(1-\mu+\mu e^s)$$

ist

$$\begin{aligned} h(0) &= 0, \\ h'(s) &= -\mu + \frac{\mu e^s}{1-\mu+\mu e^s} \;=\; 1-\mu-\frac{1-\mu}{1-\mu+\mu e^s}, \\ h'(0) &= 0, \\ h''(s) &= -\frac{d}{ds}\frac{1-\mu}{1-\mu+\mu e^s} \;=\; \frac{(1-\mu)\mu e^s}{(1-\mu+\mu e^s)^2} \\ &= \frac{\mu e^s}{1-\mu+\mu e^s}\Bigl(1-\frac{\mu e^s}{1-\mu+\mu e^s}\Bigr) \\ &\leq \frac{1}{4}. \end{aligned}$$

Folglich ergibt sich aus der Taylorschen Formel, dass

$$h(s) \;=\; h(0) + h'(0)s + \frac{h''(\xi(s))}{2}\,s^2 \;\leq\; \frac{s^2}{8}. \qquad \square$$

Mitunter kann man die Ungleichungen in eine etwas griffigere Form bringen. Zur Vorbereitung beweisen wir zunächst eine wichtige Eigenschaft der momenterzeugenden Funktion:

Lemma 2.42 *Die Funktion*

$$t \;\mapsto\; \log m_X(t) \in (-\infty, \infty]$$

ist konvex auf $\mathbb{R}$*. Ist* m_X *in einer Nullumgebung endlich, dann gelten die Gleichungen*

$$\frac{d}{dt}\log m_X(t)\Big|_{t=0} \;=\; \mathbb{E}(X) \quad \textit{und} \quad \frac{d^2}{dt^2}\log m_X(t)\Big|_{t=0} \;=\; \mathrm{Var}(X).$$

Nun kann man Lemma 2.40 wie folgt formulieren:

Lemma 2.43 *Sei* X *eine Zufallsvariable mit Erwartungswert* $\mu \in \mathbb{R}$*. Dann gilt für beliebige* $\eta > 0$*:*

$$P\bigl(\pm(X-\mu) \geq \eta\bigr) \;\leq\; \exp(-H(\pm\eta)),$$

wobei

$$H(r) \;:=\; \sup_{t\in\mathbb{R}}\bigl(rt - \log m_{X-\mu}(t)\bigr) \;=\; \sup_{t\in\mathbb{R}}\bigl(rt + \mu t - \log m_X(t)\bigr).$$

Speziell für Mittelwerte von unabhängigen, identisch verteilten Zufallsvariablen ergibt sich dann das folgende Resultat:

Lemma 2.44 *Seien $X_1, X_2, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen mit Erwartungswert $\mu \in \mathbb{R}$. Dann gilt für den Stichprobenmittelwert $\bar{X} := n^{-1}\sum_{i=1}^n X_i$ und beliebige $\eta > 0$:*

$$P\bigl(\pm(\bar{X} - \mu) \geq \eta\bigr) \ \leq \ \exp\bigl(-nH(\pm\eta)\bigr),$$

wobei $H(\cdot)$ wie in Lemma 2.43 definiert ist.

Beispiel 2.45 (Gamma-Verteilungen). Sei X nach $\mathrm{Gamma}(\alpha, 1)$ verteilt. Dann ist $m_X(t) = (1-t)^{-\alpha}$ für alle $t < 1$, so dass

$$H(r) \ = \ \sup_{t<1}\bigl(rt + \alpha t + \alpha\log(1-t)\bigr) \ = \ r - \alpha\log(1 + r/\alpha).$$

Für $\eta > 0$ ist also

$$\begin{aligned} P(X - \alpha \geq \ \eta) \ &\leq \ \exp\bigl(-(\eta - \alpha\log(1 + \eta/\alpha))\bigr), \\ P(X - \alpha \leq -\eta) \ &\leq \ \exp\bigl(-(-\eta - \alpha\log(1 - \eta/\alpha))\bigr) \ \leq \ \exp(-\eta^2/(2\alpha)). \end{aligned}$$

Zum Vergleich: Wegen $\mathrm{Var}(X) = \alpha$ liefern die Tshebyshev-Ungleichungen hier die wesentlich gröberen Schranken

$$P\bigl(\pm(X - a) \geq \eta\bigr) \ \leq \ \frac{a}{a + \eta^2}.$$

Beweis von Lemma 2.42. Sei $h(t) := \log m_X(t)$. Für ein t im Inneren des Intervalles $\bigl\{t \in \mathbb{R} : h(t) < \infty\bigr\}$ ist

$$\begin{aligned} h'(t) \ &= \ \frac{m_X'(t)}{m_X(t)} \ = \ \frac{\mathbb{E}(X\exp(tX))}{\mathbb{E}\exp(tX)}, \\ h''(t) \ &= \ \frac{m_X''(t)}{m_X(t)} - \frac{m_X'(t)^2}{m_X(t)^2} \ = \ \frac{\mathbb{E}(X^2\exp(tX))}{\mathbb{E}\exp(tX)} - h'(t)^2 \\ &= \ \frac{\mathbb{E}\bigl((X - h'(t))^2\exp(tX)\bigr)}{\mathbb{E}\exp(tX)} \ \geq \ 0. \end{aligned}$$

Dabei ergibt sich die letzte Gleichung durch Ausmultiplizieren von $(X - h'(t))^2$ und Einsetzen der vorangehenden Formel für $h'(t)$. Dies beweist die Konvexität von h, und man kann auch leicht ablesen, dass $h'(0) = \mathbb{E}(X)$ und $h''(0) = \mathrm{Var}(X)$.

Es gibt noch eine andere Betrachtungsweise: Sei Q_0 die Verteilung von X, also ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$. Für $t \in \mathbb{R}$ mit $m_X(t) = \int \exp(tx)\,Q_0(dx) < \infty$ definiert dann $f_t(x) := m_X(t)^{-1}\exp(tx)$ eine Wahrscheinlichkeitsdichte bezüglich Q_0. Bezeichnen wir das entsprechende Wahrscheinlichkeitsmaß mit Q_t, dann ist

$$h'(t) \ = \ \int x\,Q_t(dx) \ = \ \mathbb{E}(X_t) \quad \text{und} \quad h''(t) \ = \ \int (x - h'(t))^2\,Q_t(dx) \ = \ \mathrm{Var}(X_t),$$

wobei X_t eine nach Q_t verteilte Zufallsvariable ist. □

Kapitel 3

Starke Gesetze der großen Zahlen

In diesem Kapitel beschäftigen wir uns mit der Konvergenz von Folgen von Zufallsvariablen. In früheren Kapiteln begenete uns bereits die "Konvergenz im q–ten Mittel", d.h. Konvergenz bezüglich $\| \cdot \|_{q,P}$, und als Nebenprodukt die "stochastische Konvergenz". Im Folgenden geht es uns vor allem um die "fast sichere Konvergenz". Insbesondere werden wir am Ende dieses Kapitels das schwache Gesetz der großen Zahlen (Satz 2.33) für Stichprobenmittelwerte durch das sogenannte starke Gesetz der großen Zahlen abrunden.

3.1 Stochastische und fast sichere Konvergenz

Definition 3.1 (Stochastische Konvergenz, fast sichere Konvergenz). Seien $X, X_1, X_2, X_3, \ldots$ reellwertige Zufallvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \boldsymbol{P})$.

(a) Die Folge $(X_n)_{n=1}^{\infty}$ *konvergiert stochastisch gegen* X (oder: *konvergiert in Wahrscheinlichkeit gegen* X), wenn

$$\lim_{n\to\infty} P\big(|X_n - X| \geq \epsilon\big) \;=\; 0 \quad \text{für alle } \epsilon > 0.$$

Schreibweise: $X_n \to_P X \ (n \to \infty)$.

(b) Die Folge $(X_n)_{n=1}^{\infty}$ *konvergiert fast sicher gegen* X, wenn

$$P\Big(\lim_{n\to\infty} X_n = X\Big) \;=\; 1.$$

Schreibweise: $\lim_{n\to\infty} X_n = X$ fast sicher.

Anmerkung 3.2 Aus der Markov–Ungleichung folgt, dass $X_n \to_P X \ (n \to \infty)$, falls

$$\lim_{n\to\infty} \mathbb{E}\big(|X_n - X|^q\big) \;=\; 0$$

für ein $q > 0$, denn

$$P\big(|X_n - X| \geq \epsilon\big) \;\leq\; \epsilon^{-q}\, \mathbb{E}\big(|X_n - X|^q\big).$$

Lemma 3.3 ***(Zusammenhänge zwischen stochastischer und fast sicherer Konvergenz).***

(a) *Aus* $\lim_{n\to\infty} X_n = X$ *fast sicher folgt, dass auch* $X_n \to_P X$ $(n \to \infty)$.

(b) *Angenommen,* $X_n \to_P X$ $(n \to \infty)$. *Dann existiert eine Teilfolge* $(X_{n(k)})_{k=1}^{\infty}$ *von* $(X_n)_{n=1}^{\infty}$, *die fast sicher gegen* X *konvergiert.*

Beweis von Lemma 3.3. Zur Abkürzung schreiben wir $Y_n := X_n - X$.

Beweis von Teil (a). Nach Voraussetzung ist

$$1 \;=\; P\Big(\lim_{n\to\infty} Y_n = 0\Big) \;=\; P\Big(\bigcap_{\epsilon>0}\Big(\bigcup_{N\in\mathbb{N}}\bigcap_{n\geq N}\big\{|Y_n| < \epsilon\big\}\Big)\Big).$$

Aus Monotoniegründen gilt also für beliebige $\epsilon > 0$:

$$\begin{aligned} 1 \;&=\; P\Big(\bigcup_{N\in\mathbb{N}}\bigcap_{n\geq N}\big\{|Y_n| < \epsilon\big\}\Big) \\ &=\; \lim_{N\to\infty} P\Big(\bigcap_{n\geq N}\big\{|Y_n| < \epsilon\big\}\Big) \\ &\leq\; \liminf_{N\to\infty} P\big(|Y_N| < \epsilon\big) \\ &=\; 1 - \limsup_{N\to\infty} P\big(|Y_N| \geq \epsilon\big). \end{aligned}$$

Beweis von Teil (b). Nach Voraussetzung gibt es Indizes $n(1) < n(2) < n(3) < \cdots$ derart, dass

$$P\big(|Y_{n(k)}| \geq k^{-1}\big) \;\leq\; 2^{-k}.$$

Nach dem Borell-Cantelli-Lemma hat das Ereignis

$$\big\{Y_{n(k)} \geq k^{-1} \text{ für unendlich viele } k \in \mathbb{N}\big\}$$

Wahrscheinlichkeit Null, und für jedes ω aus dem komplementären Ereignis konvergiert die Folge $\big(Y_{n(k)}(\omega)\big)_{k=1}^{\infty}$ offensichtlich gegen Null. □

3.2 Konvergenz zufälliger Reihen

In diesem Abschnitt beweisen und verwenden wir eine Verfeinerung der Tshebyshev–Ungleichung.

Lemma 3.4 ***(Kolmogorov).*** *Seien* $Y_1, Y_2, \ldots, Y_n$ *stochastisch unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *derart, dass* $\mathbb{E}(Y_i) = 0$ *und* $\mathbb{E}(Y_i^2) < \infty$. *Für* $1 \leq k \leq n$ *sei* $S_k := \sum_{i=1}^{k} Y_i$. *Dann ist*

$$P\Big(\max_{k=1,2,\ldots,n} |S_k| \geq \eta\Big) \;\leq\; \frac{\mathrm{Var}(S_n)}{\eta^2}$$

für beliebige $\eta > 0$.

Beweis von Lemma 3.4. Sei

$$T \;:=\; \min\left(\{k \le n : |S_k| \ge \eta\} \cup \{\infty\}\right).$$

Deutet man die Indizes von Y_i und S_k als Zeitpunkte, dann ist T derjenige Zeitpunkt k, zu welchem erstmalig der Absolutbetrag von S_k größer oder gleich η ist. Die Wahrscheinlichkeit, dass $\max_{k\le n} |S_k|$ größer oder gleich η ist, kann man nun schreiben als

$$P(T \le n) \;=\; \sum_{k=1}^{n} P(T = k) \;\le\; \sum_{k=1}^{n} \mathbb{E}\left(1\{T = k\}S_k^2\right)/\eta^2,$$

denn $T = k$ impliziert ja, dass $S_k^2 \ge \eta^2$. Aber $Z_k := 1\{T = k\}S_k$ ist eine Funktion von $Y_1, \ldots, Y_k$ mit $\mathbb{E}(Z_k^2) < \infty$. Folglich ist

$$\begin{aligned}
\mathbb{E}\big(1\{T = k\}S_n^2\big) &= \mathbb{E}\big(1\{T = k\}(S_k + (S_n - S_k))^2\big)\\
&\ge \mathbb{E}\big(1\{T = k\}S_k^2\big) + 2\,\mathbb{E}\big(1\{T = k\}S_k(S_n - S_k)\big)\\
&= \mathbb{E}\big(1\{T = k\}S_k^2\big) + 2\sum_{i=k+1}^{n} \mathbb{E}(Z_k Y_i)\\
&= \mathbb{E}\big(1\{T = k\}S_k^2\big),
\end{aligned}$$

denn wegen der stochastischen Unabhängigkeit aller Y_j ist $\mathbb{E}(Z_k Y_i) = \mathbb{E}(Z_k)\,\mathbb{E}(Y_i) = 0$ für $k < i \le n$. Folglich ist $P(T \le n)$ nicht größer als

$$\sum_{k=1}^{n} \mathbb{E}\left(1\{T = k\}S_n^2\right)/\eta^2 \;=\; \mathbb{E}\left(1\{T \le n\}S_n^2\right)/\eta^2 \;\le\; \mathbb{E}(S_n^2)/\eta^2. \qquad \square$$

Lemma 3.4 impliziert, dass gewisse Reihen von Zufallsvariablen mit Wahrscheinlichkeit Eins konvergieren:

Satz 3.5 ***(Konvergenz zufälliger Reihen).*** *Seien $Y_1, Y_2, Y_3, \ldots$ stochastisch unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ mit $\mathbb{E}(Y_i) = 0$ und $\sum_{i=1}^{\infty} \mathbb{E}(Y_i^2) < \infty$. Dann gibt es eine reellwertige Zufallsvariable S auf $(\Omega, \mathcal{A}, P)$, so dass*

$$\lim_{n\to\infty} \sum_{i=1}^{n} Y_i \;=\; S \quad \textit{fast sicher.}$$

Beweis von Satz 3.5. Sei $S_k := \sum_{i=1}^{k} Y_i$. Für ein beliebiges $\omega \in \Omega$ konvergiert die Folge $\sum_{i=1}^{\infty} Y_i(\omega) = (S_n(\omega))_{n=1}^{\infty}$ genau dann, wenn es sich um eine Cauchy-Folge handelt. Das heißt, wenn für beliebige $\epsilon > 0$ ein $N(\omega) \in \mathbb{N}$ existiert, so dass

$$|S_k(\omega) - S_\ell(\omega)| \;\le\; \epsilon \quad \text{für alle } k, \ell \ge N(\omega).$$

Mit dem Ereignis

$$A_\epsilon \;:=\; \bigcup_{N=1}^{\infty} \big\{|S_k - S_\ell| \le \epsilon \text{ für alle} k, \ell \ge N\big\} \;=\; \bigcup_{N=1}^{\infty} \big\{\sup_{k,\ell\ge N} |S_k - S_\ell| \le \epsilon\big\}$$

ist also

$$P\Big(\sum_{i=1}^{\infty} Z_i \text{ konvergiert}\Big) \;=\; P\Big(\bigcap_{\epsilon>0} A_\epsilon\Big) \;=\; \lim_{\epsilon\to 0} P(A_\epsilon).$$

Dabei verwenden wir die Tatsache, dass $A_\epsilon \subset A_\delta$ für $0 < \epsilon < \delta$. Es genügt also zu zeigen, dass $P(A_\epsilon) = 1$ für ein beliebiges $\epsilon > 0$. Man kann schreiben

$$\begin{aligned}
P(A_\epsilon) \;&=\; \lim_{N\to\infty} P\Big(\sup_{k,\ell\geq N} |S_k - S_\ell| \leq \epsilon\Big) \\
&=\; \lim_{N\to\infty} \inf_{M>N} P\Big(\max_{N<k,\ell\leq M} |S_k - S_\ell| \leq \epsilon\Big) \\
&\geq\; \lim_{N\to\infty} \inf_{M>N} P\Big(\max_{N<k\leq M} |S_k - S_N| \leq \epsilon/2\Big) \\
&=\; \lim_{N\to\infty} \inf_{M>N} P\Big(\max_{N<k\leq M} \Big|\sum_{i=N+1}^{k} Y_i\Big| \leq \epsilon/2\Big) \\
&\geq\; \lim_{N\to\infty} \inf_{M>N} \Big(1 - (\epsilon/2)^{-2} \sum_{k=N+1}^{M} \mathbb{E}(Y_i^2)\Big) \\
&=\; 1 - (\epsilon/2)^{-2} \lim_{N\to\infty} \sum_{k=N+1}^{\infty} \mathbb{E}(Y_i^2) \\
&=\; 1.
\end{aligned}$$

Dabei verwendeten wir die Kolmogorov-Ungleichung (Lemma 3.4) und die Konvergenz der Reihe $\sum_{i=1}^{\infty} \mathbb{E}(Y_i^2)$. □

Beispiel 3.6 (Harmonische Reihe mit zufälligen Vorzeichen). Bekanntlich divergiert die harmonische Reihe $\sum_{k=1}^{\infty} 1/k$. Andererseits konvergiert $\sum_{k=1}^{\infty} (-1)^{k-1}/k$ gegen $\log 2$. Aus Satz 3.5 folgt, dass auch eine zufällige Reihe $\sum_{k=1}^{\infty} Z_k/k$ mit stochastisch unabhängigen Vorzeichen $Z_k \in \{-1, 1\}$ fast sicher konvergiert, sofern $P(Z_k = \pm 1) = 1/2$. Denn mit $Y_k := Z_k/k$ ist $\mathbb{E}(Y_k) = 0$ und $\sum_{k=1}^{\infty} \mathbb{E}(Y_k^2) = \sum_{k=1}^{\infty} 1/k^2$ ist endlich.

Beispiel 3.7 (Lineare Prozesse). Seien ϵ_t, $t \in \mathbb{Z}$, stochastisch unabhängige Zufallsvariablen mit

$$\mathbb{E}(\epsilon_t) \;=\; 0 \quad \text{und} \quad \mathbb{E}(\epsilon_t^2) \;=\; \sigma^2 \;<\; \infty.$$

Für feste reelle Zahlen a_n, $n \in \mathbb{N}_0$, betrachten wir nun die Reihe

$$X_t \;:=\; \sum_{n=0}^{\infty} a_n \epsilon_{t-n}.$$

Die Reihe auf der rechten Seite konvergiert fast sicher und im quadratischen Mittel, falls

$$\sum_{n=0}^{\infty} a_n^2 \;<\; \infty.$$

Die fast sichere Konvergenz ergibt sich aus Satz 3.5, angewandt auf $Y_n := a_n \epsilon_{t-n}$, denn $\mathbb{E}(Y_n) = 0$ und $\mathbb{E}(Y_n^2) = a_n^2 \sigma^2$. Die Konvergenz im quadratischen Mittel ergibt sich aus der Tatsache, dass für $0 \le N < M$ gilt:

$$\Big\| \sum_{n=N+1}^{M} a_n \epsilon_{t-n} \Big\|_{2,P}^2 \;=\; \sigma^2 \sum_{n=N+1}^{M} a_n^2 \;\le\; \sigma^2 \sum_{n=N+1}^{\infty} a_n^2 \;\to\; 0 \quad (N \to \infty).$$

Daher ist $\left(\sum_{n=0}^{N} a_n \epsilon_{t-n}\right)_{N=0}^{\infty}$ eine Cauchy–Folge bezüglich $\|\cdot\|_{2,P}$ und die Existenz des Grenzwertes ergibt sich aus Satz 2.24. Dieser stimmt mit dem fast sicheren Grenzwert X_t überein, und

$$\mathbb{E}(X_t^2) \;=\; \sigma^2 \sum_{n=0}^{\infty} a_n^2.$$

Anmerkung: Die Voraussetzungen, dass die ϵ_t Erwartungswert Null haben und stochastisch unabhängig sind, kann man wesentlich abschwächen, wenn die stärkere Bedingung $\sum_{n=0}^{\infty} |a_n| < \infty$ erfüllt ist: Für ein $q \ge 1$ seien ϵ_t, $t \in \mathbb{Z}$, beliebige Zufallsvariablen in $\mathcal{L}^q(P)$ mit $\sup_t \|\epsilon_t\|_{q,P} =: C < \infty$. Dann konvergiert $\sum_{n=0}^{\infty} a_n \epsilon_{t-n}$ im q–ten Mittel gegen eine Zufallsvariable X_t. Denn

$$\sum_{n=0}^{\infty} \big\| a_n \epsilon_{t-n} \big\|_{q,P} \;\le\; \sum_{n=0}^{\infty} |a_n| C \;<\; \infty,$$

so dass $\left(\sum_{n=0}^{N} a_n \epsilon_{t-n}\right)_{N=0}^{\infty}$ eine Cauchy–Folge bezüglich $\|\cdot\|_{q,P}$ ist.

Spezialfall: Lineare Prozesse werden beispielsweise als Modelle für Aktienkursschwankungen verwendet. Genauer gesagt, sei A_t der Tagesendkurs einer bestimmten Aktie am Tag t, und sei $X_t := \log(A_t/A_{t-1})$, ein sogenannter log-return. Modelliert man X_t als

$$X_t \;=\; \sum_{n=0}^{\infty} \theta^n \epsilon_{t-n}$$

mit einer Konstante $\theta \in [0,1)$, dann gilt die Gleichung

$$X_t \;=\; \theta X_{t-1} + \epsilon_t.$$

Der heutige log-return ist also proportional zum gestrigen log-return plus zufälliges Rauschen. Letzteres beschreibt die unvorhersehbaren Einflüsse des heutigen Tages.

3.3 Das Starke Gesetz der Großen Zahlen

Eine wichtige Anwendung der Kolmogorovschen Ungleichung ist das Starke Gesetz der großen Zahlen. Hiervon gibt es verschiedene Varianten, und wir behandeln zunächst eine für statistische Anwendungen interessante Version:

Satz 3.8 ***(Gesetz der großen Zahlen mit expliziter Ungleichung).*** *Seien $X_1, X_2, X_3, \ldots$ stochastisch unabhängige Zufallsvariablen in $\mathcal{L}^2(P)$, jeweils mit Varianz kleiner oder gleich $\sigma^2 < \infty$. Dann ist*

$$\lim_{n\to\infty} (\bar{X}_n - \mathbb{E}\,\bar{X}_n) \;=\; 0 \quad \textit{fast sicher},$$

wobei $\bar{X}_n := n^{-1}\sum_{i=1}^n X_i$. Genauer gesagt, gilt für beliebige $N \in \mathbb{N}$ und $\epsilon > 0$ die Ungleichung

$$P\big(|\bar{X}_n - \mathbb{E}\,\bar{X}_n| \geq \epsilon \text{ für ein } n \geq N\big) \;\leq\; \frac{4\sigma^2}{N\epsilon^2}. \tag{3.1}$$

Im Spezialfall identisch verteilter Zufallsvariablen X_i mit Erwartungswert μ gilt sogar die schärfere Ungleichung

$$P\big(|\bar{X}_n - \mu| \geq \epsilon \text{ für ein } n \geq N\big) \;\leq\; \frac{\sigma^2}{N\epsilon^2}. \tag{3.2}$$

Anmerkung 3.9 (Konvergenzgeschwindigkeit). Wenn man den Beweis von Satz 3.8 geeignet modifiziert und das Borel-Cantelli-Lemma anwendet, kann man auch folgende Aussage beweisen: Unter den allgemeinen Bedingungen von Satz 3.8 gilt für jedes $\gamma > 1/2$:

$$P\Big(|\bar{X}_n - \mathbb{E}\,\bar{X}_n| \geq \frac{\log(n)^\gamma}{\sqrt{n}} \text{ für unendlich viele } n \in \mathbb{N}\Big) \;=\; 0.$$

Für den Fall identisch verteilter Zufallsvariablen X_i kann man die Voraussetzung, dass $X_i \in \mathcal{L}^2(P)$, durch die schwächere Voraussetzung, dass $X_i \in \mathcal{L}^1(P)$, ersetzen und erhält das klassische starke Gesetz der großen Zahlen:

Satz 3.10 ***(Standardversion des Gesetzes der großen Zahlen).*** *Seien $X_1, X_2, X_3, \ldots$ stochastisch unabhängige, identisch verteilte Zufallsvariablen mit Erwartungswert $\mu \in \mathbb{R}$. Dann ist*

$$\lim_{n\to\infty} \bar{X}_n \;=\; \mu \quad \textit{fast sicher.}$$

Beweis von Satz 3.8. Man kann sich schnell davon überzeugen, dass es genügt, die Ungleichungen (3.1) bzw. (3.2) zu beweisen. Mit $Y_i := X_i - \mathbb{E}\,X_i$ und $S_n := \sum_{i=1}^n Y_i$ ergibt sich (3.1) aus der Kolmogorovschen Ungleichung wie folgt:

$$\begin{aligned}
P\big(|\bar{X}_n - \mathbb{E}\,\bar{X}_n| \geq \epsilon \text{ für ein } n \geq N\big) \;&=\; P\big(|S_n| \geq n\epsilon \text{ für ein } n \geq N\big) \\
&\leq\; \sum_{\ell=0}^{\infty} P\big(|S_k| \geq k\epsilon \text{ für ein } k \in [2^\ell N, 2^{\ell+1}N]\big) \\
&\leq\; \sum_{\ell=0}^{\infty} P\Big(\max_{k \leq N2^{\ell+1}} |S_k| \geq N2^\ell\epsilon\Big) \\
&\leq\; \sum_{\ell=0}^{\infty} \frac{N2^{\ell+1}\sigma^2}{N^2(2^\ell)^2\epsilon^2} \\
&=\; \frac{2\sigma^2}{N\epsilon^2}\sum_{\ell=0}^{\infty} 2^{-\ell} \\
&=\; \frac{4\sigma^2}{N\epsilon^2}.
\end{aligned}$$

Um die präzisere Ungleichung (3.2) zu beweisen, müssen wir den Beweis der Kolmogorovschen Ungleichung neu aufrollen: Zunächst gilt für $\bar{Y}_n := n^{-1}S_n$:

$$P\big(|\bar{Y}_n| \geq \epsilon \text{ für ein } n \geq N\big) \ = \ \lim_{M\to\infty} P\big(|\bar{Y}_n| \geq \epsilon \text{ für ein } n \in [N,M]\big),$$

und es genügt zu zeigen, dass die rechte Seite für jedes feste $M > N$ kleiner oder gleich $\sigma^2/(N\epsilon^2)$ ist. Nun betrachten wir die Sequenz $\bar{Y}_N, \bar{Y}_{N+1}, \ldots, \bar{Y}_M$ von hinten nach vorne: Mit

$$G_n \ := \ 1\big\{|\bar{Y}_n| \geq \epsilon, |\bar{Y}_k| < \epsilon \text{ für } n < k \leq M\big\}$$

für $n = N, N+1, \ldots, M$ ist

$$\begin{aligned}
P\big(|\bar{Y}_n| \geq \epsilon \text{ für ein } n \in [N,M]\big) \ &= \ \sum_{n=N}^{M} \mathbb{E}\, G_n \\
&\leq \ \frac{1}{\epsilon^2}\sum_{n=N}^{M} \mathbb{E}(\bar{Y}_n^2 G_n) \\
&\leq \ \frac{1}{\epsilon^2}\sum_{n=N}^{M} \mathbb{E}(\bar{Y}_N^2 G_n) \\
&= \ \frac{1}{\epsilon^2}\,\mathbb{E}\Big(\bar{Y}_N^2 \sum_{n=N}^{M} G_n\Big) \\
&\leq \ \frac{1}{\epsilon^2}\,\mathbb{E}(\bar{Y}_N^2) \\
&= \ \frac{\sigma^2}{N\epsilon^2}.
\end{aligned}$$

Dabei ergibt sich die zweite Ungleichung, $\mathbb{E}(\bar{Y}_n^2 G_n) \leq \mathbb{E}(\bar{Y}_N^2 G_n)$, aus der folgenden Überlegung: Für $N < n \leq M$ ist

$$\begin{aligned}
\mathbb{E}(\bar{Y}_N^2 G_n) \ &= \ \mathbb{E}\Big(\big(\bar{Y}_n + (\bar{Y}_N - \bar{Y}_n)\big)^2 G_n\Big) \\
&\geq \ \mathbb{E}(\bar{Y}_n^2 G_n) + 2\,\mathbb{E}\big((\bar{Y}_N - \bar{Y}_n)\bar{Y}_n G_n\big) \\
&= \ \mathbb{E}(\bar{Y}_n^2 G_n) + 2\Big(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i\,\bar{Y}_n G_n) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i\,\bar{Y}_n G_n)\Big) \\
&= \ \mathbb{E}(\bar{Y}_n^2 G_n).
\end{aligned}$$

Denn $\bar{Y}_n G_n$ ist eine Funktion von $\bar{Y}_n, \bar{Y}_{n+1}, \ldots, \bar{Y}_M$ und somit symmetrisch in $Y_1, Y_2, \ldots, Y_n$. Ferner sind letztere Zufallsvariablen nach Voraussetzung identisch verteilt, so dass

$$\mathbb{E}(Y_i\,\bar{Y}_n G_n) \ = \ \mathbb{E}(Y_1\,\bar{Y}_n G_n) \quad \text{für } 1 \leq i \leq n. \qquad \square$$

Im Beweis von Satz 3.10 verwenden wir ein einfaches Lemma über reelle Zahlenfolgen, dessen Beweis wir als Übungsaufgabe stellen:

Lemma 3.11 *Seien $y_1, y_2, y_3, \ldots$ reelle Zahlen derart, dass die Reihe $\sum_{n=1}^{\infty} n^{-1}y_n$ (bedingt) konvergiert. Dies impliziert, dass*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y_i \ = \ 0.$$

Beweis von Satz 3.10. Ähnlich wie beim schwachen Gesetz der großen Zahlen verwenden wir auch hier “trunkierte Variablen”. Und zwar betrachten wir

$$Y_n \;:=\; 1\{|X_n| \le n\}X_n \quad \text{und} \quad Z_n \;:=\; 1\{|X_n| > n\}X_n.$$

Dann ist

$$|\bar{X}_n - \mu| \;\le\; |\bar{Y}_n - \mathbb{E}\,\bar{Y}_n| + |\,\mathbb{E}\,\bar{Y}_n - \mu| + |\bar{Z}_n|,$$

und es genügt zu zeigen, dass die drei Summanden auf der rechten Seite fast sicher gegen Null konvergieren.

Schritt 1: Aus Satz 3.5 folgt, dass $\sum_{n=1}^{\infty}(Y_n - \mathbb{E}\,Y_n)/n$ fast sicher konvergiert. Denn

$$\begin{aligned}
\sum_{n=1}^{\infty} \operatorname{Var}\Big(\frac{Y_n - \mathbb{E}\,Y_n}{n}\Big) \;&=\; \sum_{n=1}^{\infty} \frac{\operatorname{Var}(Y_n)}{n^2} \\
&\le\; \sum_{n=1}^{\infty} \frac{\mathbb{E}(Y_n^2)}{n^2} \\
&=\; \sum_{n=1}^{\infty} \frac{\mathbb{E}\big(1\{|X_1| \le n\}X_1^2\big)}{n^2} \\
&=\; \mathbb{E}\Big(X_1^2 \sum_{n \ge |X_1|} \frac{1}{n^2}\Big) \\
&\le\; \mathbb{E}\Big(X_1^2 \sum_{n \ge |X_1|} 2\int_n^{n+1} \frac{dx}{x^2}\Big) \\
&\le\; 2\,\mathbb{E}\Big(X_1^2 \int_{|X_1|}^{\infty} \frac{dx}{x^2}\Big) \\
&=\; 2\,\mathbb{E}\,|X_1| \;<\; \infty.
\end{aligned}$$

Zusammen mit Lemma 3.11, angewandt auf $y_n := Y_n(\omega) - \mathbb{E}\,Y_n$, ergibt sich dann die fast sichere Konvergenz von $\big(\bar{Y}_n - \mathbb{E}\,\bar{Y}_n\big)_{n=1}^{\infty}$ gegen Null.

Schritt 2: Nach Definition von Y_n ist

$$\mathbb{E}\,\bar{Y}_n \;=\; \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\big(1\{|X_i| \le i\}X_i\big) \;=\; \mathbb{E}\Big(X_1 \,\frac{1}{n}\sum_{i=1}^{n} 1\{i \ge |X_1|\}\Big)$$

und konvergiert für $n \to \infty$ gegen $\mathbb{E}(X_1) = \mu$. Dies ergibt sich aus dem Satz von der majorisierten Konvergenz. Denn einerseits ist

$$\Big|X_1 \,\frac{1}{n}\sum_{i=1}^{n} 1\{i \ge |X_1|\}\Big| \;\le\; |X_1| \quad \text{mit } \mathbb{E}\,|X_1| < \infty.$$

Andererseits ist $\lim_{n\to\infty} \sum_{i=1}^{n} 1\{i \ge r\}/n = 1$ für jede reelle Zahl r.

Schritt 3: Dass auch $(\bar{Z}_n)_{n=1}^{\infty}$ fast sicher gegen Null konvergiert, ergibt sich aus der Tatsache, dass

$$P\big(Z_n \ne 0 \text{ für unendlich viele } n \in \mathbb{N}\big) \;=\; 0.$$

Diese ist wiederum eine Folgerung aus dem Borel–Cantelli–Lemma (Lemma 1.53), denn

$$\sum_{n=1}^{\infty} P(Z_n \neq 0) \;=\; \sum_{n=1}^{\infty} P\big(|X_1| > n\big) \;\leq\; \int_0^{\infty} P\big(|X_1| > r\big)\, dr \;=\; \mathbb{E}\,|X_1| \;<\; \infty;$$

siehe auch (2.4). □

Beispiel 3.12 (Auf der Suche nach dem idealen Zufallsgenerator: Normale Zahlen). Für jede reelle Zahl x und jede natürliche Zahl $b > 1$ gibt es eine Darstellung der Form

$$x \;=\; x_o + \sum_{k=1}^{\infty} b^{-k} x_{b,k}$$

mit $x_o \in \mathbb{Z}$ und Ziffern $x_{b,k} \in \{0, 1, \ldots, b-1\}$. (Diese Darstellung ist eindeutig, wenn man beispielsweise verlangt, dass unendlich viele Ziffern $x_{b,k}$ von Null verschieden sein sollen.) Nun nennt man eine reelle Zahl x *normal*, wenn für jede Basis $b \in \mathbb{N} \setminus \{1\}$ und jede Ziffer $\ell \in \{0, 1, \ldots, b-1\}$ gilt:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1\{x_{b,k} = \ell\} \;=\; \frac{1}{b}.$$

Rationale Zahlen sind niemals normal, doch empirische Untersuchungen legen z.B. die Vermutung nahe, dass $\sqrt{2}$, π und e normale Zahlen sind. Aus dem Starken Gesetz der großen Zahlen kann man ableiten, dass "fast alle" reellen Zahlen normal sind. Genauer: Die Menge aller nichtnormalen reellen Zahlen hat Lebesguemaß Null. Um dies zu zeigen, betrachten wir eine auf $[0, 1]$ uniform verteilte Zufallsvariable. Diese Zufallszahl ist fast sicher normal. Denn man kann leicht nachweisen, dass für festes b die Ziffern $X_{b,1}$, $X_{b,2}$, $X_{b,3}$, … stochastisch unabhängig und auf $\{0, 1, \ldots, b-1\}$ uniform verteilt sind. Nach Satz 3.10 ist also

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} 1\{X_{b,i} = \ell\} \;=\; \mathbb{E}\,1\{X_{b,1} = \ell\} = \frac{1}{b}$$

fast sicher. Kombiniert man dieses Ereignis für alle abzählbar vielen Paare (b, ℓ), so ergibt sich die fast sichere Normalität von X. Durch Betrachten von $z + X$ mit beliebigem festen $z \in \mathbb{Z}$ ergibt sich die Aussage über alle nichtnormalen reellen Zahlen.

Kapitel 4

Konvergenz von Verteilungen

In diesem Kapitel beschäftigen wir uns mit der Konvergenz von Wahrscheinlichkeitsverteilungen. Im ersten Abschnitt geht es um die Approximation von verallgemeinerten Binomialverteilungen durch Poissonverteilungen. Danach führen wir die sogenannte *schwache Konvergenz* bzw. *Konvergenz in Verteilung* ein und beweisen in diesem Kontext den zentralen Grenzwertsatz, eines der wichtigsten Resultate für die mathematische Statistik.

4.1 Poisson–Approximationen

Wir erinnern an die Definition der Poisson–Verteilung mit Parameter $\lambda \geq 0$: Diese ist definiert als das diskrete Wahrscheinlichkeitsmaß Poiss_λ auf $\mathbb{N}_0$ mit Gewichten

$$\mathrm{Poiss}_\lambda(\{k\}) \;=\; e^{-\lambda}\frac{\lambda^k}{k!} \qquad (k \in \mathbb{N}_0).$$

In der Vorlesung “Kombinatorik und Wahrscheinlichkeit” wurde bereits gezeigt, dass diese Verteilung als Grenzwert der Binomialverteilung $\mathrm{Bin}_{n,p}$ für $n \to \infty$ und $np \to \lambda$ auftritt. Dahinter steckt eine viel allgemeinere Aussage:

Satz 4.1 ***(Hodges–LeCam)****.* *Seien $X_1, X_2, X_3, \ldots$ stochastisch unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit*

$$\mathbb{P}(X_i = 1) \;=\; 1 - \mathbb{P}(X_i = 0) \;=\; p_i$$

und $\lambda := \sum_{i\geq 1} p_i < \infty$. Dann ist $S := \sum_{i\geq 1} X_i$ fast sicher endlich mit $\mathbb{E}(S) = \lambda$, und

$$\sup_{B\subset\mathbb{N}_0} \Big|\mathbb{P}(S \in B) - \mathrm{Poiss}_\lambda(B)\Big| \;\leq\; \sum_{i\geq 1} p_i^2.$$

Anmerkung 4.2 (Anwendungen). Dieser Satz erklärt, warum viele Zufallszahlen näherungsweise poissonverteilt sind. Hier einige Beispiele:

• Die Anzahl von Anfragen bei einer Servicestelle (z.B. Telephon-Hotline einer Firma) in einem bestimmten kurzen Zeitfenster ist näherungsweise poissonverteilt. Denn es gibt eine sehr große Anzahl potentieller Anrufer, doch jeder einzelne nimmt nur mit sehr geringer Wahrscheinlichkeit diesen Service im besagten Zeitintervall in Anspruch.
• Schadensfälle, die bei einer Haftpflichtversicherung mit vielen Kunden in einem bestimmten Zeitraum gemeldet werden. Sofern jeder einzelne Kunde nur mit geringer Wahrscheinlichkeit einen Schadensfall meldet, ist die Gesamtzahl näherungsweise poissonverteilt.
• Radioaktiver Zerfall: Anzahl von Zerfällen einer schwach radioaktiven Substanz in einem kurzen Zeitintervall.

Anmerkung 4.3 Die anfangs erwähnte Approximation der Binomialverteilung $\mathrm{Bin}_{n,p}$ kann man wie folgt reproduzieren: $\mathrm{Bin}_{n,p}$ beschreibt die Verteilung von $\sum_{i\geq 1} X_i$ im Falle von $p_i = p$ für $i \leq n$ und $p_i = 0$ für $i > n$. Folglich ist

$$\sup_{B\subset\mathbb{N}_0} \Big|\mathrm{Bin}_{n,p}(B) - \mathrm{Poiss}_{np}(B)\Big| \;\leq\; np^2 \;=\; \lambda \cdot p.$$

Anmerkung 4.4 Allgemein ist die Schranke in Satz 4.1 nicht größer als

$$\lambda \cdot \max_{i\geq 1} p_i. \tag{4.1}$$

Die Approximation der Verteilung von S durch eine Poisson–Verteilung ist also besonders gut, wenn das Maximum der Einzelwahrscheinlichkeiten p_i sehr klein ist.

Den Faktor λ in (4.1) kann man durch eine verfeinerte Beweistechnik (Chen–Stein–Methode) durch $(1 - e^{-\lambda})/\lambda \leq 1$ ersetzen, was aber über den Rahmen dieser Vorlesung hinausgeht.

Anmerkung 4.5 Im nachfolgenden Beweis verwenden wir an einer Stelle die Tatsache, dass $p(1-\exp(-p)) \leq p^2$ für beliebige $p \geq 0$. Alternativ kann man zeigen, dass $p(1 - \exp(-p)) \leq 1 - \exp(-p^2)$ für $p \geq 0$, und hiermit lässt sich die Ungleichung in Satz 4.1 verfeinern zu

$$\sup_{B\subset\mathbb{N}_0} \Big|\mathbb{P}(S \in B) - \mathrm{Poiss}_\lambda(B)\Big| \;\leq\; 1 - \exp\Big(-\sum_{i\geq 1} p_i^2\Big);$$

siehe Übungen.

Beweis von Satz 4.1. Der Beweis verwendet ein sogenanntes Koppelungsargument. Wir werden einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ angeben, auf dem stochastisch unabhängige Paare $(\widetilde{X}_1, Y_1)$, $(\widetilde{X}_2, Y_2)$, $(\widetilde{X}_3, Y_3)$, ... von Zufallsvariablen definiert sind, so dass für beliebige $i \geq 1$ gilt:

$$\mathbb{P}(\widetilde{X}_i = 1) \;=\; 1 - \mathbb{P}(\widetilde{X}_i = 0) \;=\; p_i, \tag{4.2}$$

$$Y_i \;\sim\; \mathrm{Poiss}_{p_i}, \tag{4.3}$$

$$\mathbb{P}(\widetilde{X}_i \neq Y_i) \;\leq\; p_i^2. \tag{4.4}$$

Definiert man nun $\widetilde{S} := \sum_{i\ge 1} \widetilde{X}_i$ und $T := \sum_{i\ge 1} Y_i$, dann $\widetilde{S}$ wie S verteilt, und T ist nach Poiss_λ verteilt, denn die Summe stochastisch unabhängiger, poissonverteilter Zufallsvariablen ist wieder poissonverteilt. Ferner gilt für beliebige Mengen $B \subset \mathbb{N}_0$:

$$\begin{aligned}
\bigl|\mathbb{P}(S \in B) - \mathrm{Poiss}_\lambda(B)\bigr| &= \bigl|\mathbb{P}(\widetilde{S} \in B) - \mathbb{P}(T \in B)\bigr| \\
&= \bigl|\mathbb{P}(\widetilde{S} \in B, \widetilde{S} \ne T) - \mathbb{P}(T \in B, \widetilde{S} \ne T)\bigr| \\
&\le \mathbb{P}(\widetilde{S} \ne T) \\
&\le \mathbb{P}\bigl(\widetilde{X}_i \ne Y_i \text{ für ein } i \ge 1\bigr) \\
&\le \sum_{i\ge 1} \mathbb{P}(\widetilde{X}_i \ne Y_i) \\
&\le \sum_{i \ge 1} p_i^2.
\end{aligned}$$

Zu zeigen bleibt, wie man Zufallsvariablen $(\widetilde{X}_i, Y_i)$ mit den Eigenschaften (4.2–4.4) konstruiert. Zu diesem Zweck verwenden wir die Quantiltransformation: Seien $U_1, U_2, U_3, \ldots$ stochastisch unabhängige, uniform auf $(0,1)$ verteilte Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann definieren wir

$$\begin{aligned}
\widetilde{X}_i &:= \begin{cases} 0 & \text{falls } U_i < 1 - p_i, \\ 1 & \text{falls } U_i \ge 1 - p_i, \end{cases} \\
Y_i &:= F_i^{-1}(U_i) \begin{cases} = 0 & \text{falls } U_i < \exp(-p_i), \\ = 1 & \text{falls } \exp(-p_i) \le U_i < (1+p_i)\exp(-p_i), \\ \ge 2 & \text{falls } (1+p_i)\exp(-p_i) \le U_i, \end{cases}
\end{aligned}$$

wobei F_i die Verteilungsfunktion von Poiss_{p_i} bezeichnet. Nach Satz 1.40 haben die Paare $(\widetilde{X}_i, Y_i)$ die Eigenschaften (4.2) und (4.3). Was die dritte Eigenschaft (4.4) anbelangt, so ist $1 - p_i \le \exp(-p_i) \le (1+p_i)\exp(-p_i)$, so dass

$$\begin{aligned}
\mathbb{P}(\widetilde{X}_i \ne Y_i) &= \mathbb{P}\bigl(1 - p_i \le U_i < \exp(-p_i) \text{ oder } U_i \ge (1+p_i)\exp(-p_i)\bigr) \\
&= \exp(-p_i) - 1 + p_i \; + \; 1 - (1+p_i)\exp(-p_i) \\
&= p_i(1 - \exp(-p_i)) \\
&\le p_i^2. \qquad \square
\end{aligned}$$

4.2 Schwache Konvergenz und Konvergenz in Verteilung

In Abschnitt 4.1 ging es, allgemein formuliert, um folgenden Abstand zwischen zwei Wahrscheinlichkeitsmaßen P und Q auf einem messbaren Raum $(\mathcal{X}, \mathcal{B})$:

$$\sup_{B \in \mathcal{B}} \bigl|P(B) - Q(B)\bigr|.$$

Für viele Zwecke ist dieser Abstand aber ungeeignet. Vergleicht man beispielsweise die kontinuierliche uniforme Verteilung $P = \mathrm{Unif}[0,1]$ mit der diskreten uniformen Verteilung $P_n =$

$\mathrm{Unif}\{1/n, 2/n, \ldots, 1\}$, dann sind P und P_n für großes $n \in \mathbb{N}$ recht "ähnlich", obwohl

$$P(\{1/n, 2/n, \ldots, 1\}) \;=\; 0 \quad \text{aber} \quad P_n(\{1/n, 2/n, \ldots, 1\}) \;=\; 1.$$

Im vorliegenden Abschnitt behandeln wir ein für diesen Fall geeignetes Konzept. Dabei betrachten wir allgemein Wahrscheinlichkeitsmaße auf einem *metrischen Raum* $(\mathcal{X}, d)$.

4.2.1 Der Fall eines allgemeinen metrischen Raumes

Definition 4.6 (Borelmengen). Die kleinste σ–Algebra über $\mathcal{X}$, die alle offenen und abgeschlossenen Mengen enthält, nennt man die *σ–Algebra der Borelmengen in* $(\mathcal{X}, d)$. Hierfür schreiben wir $\mathrm{Borel}(\mathcal{X}, d)$ oder einfach $\mathrm{Borel}(\mathcal{X})$, wenn klar ist, welche Metrik d gemeint ist.

Im Folgenden wird $\mathcal{X}$ stets mit $\mathrm{Borel}(\mathcal{X}, d)$ versehen, und wir sprechen einfach von Maßen auf $\mathcal{X}$ bzw. Zufallsvariablen mit Werten in $\mathcal{X}$.

Anmerkung 4.7 Jede stetige Funktion $f : \mathcal{X} \to \mathbb{R}$ ist $\mathrm{Borel}(\mathcal{X})$–$\mathrm{Borel}(\mathbb{R})$–messbar. Zwei Wahrscheinlichkeitsmaße P und Q auf $\mathcal{X}$ sind identisch genau dann, wenn

$$\int f\, dP \;=\; \int f\, dQ$$

für jede Lipschitz–stetige Funktion $f : \mathcal{X} \to [0, 1]$. Denn einerseits ist $P = Q$ genau dann, wenn $P(U) = Q(U)$ für jede offene Menge $U \subset \mathcal{X}$. Andererseits gibt es zu U eine Folge von Lipschitz–stetigen Funktionen $f_n : \mathcal{X} \to \mathbb{R}$ derart, dass $0 \le f_n \uparrow 1_U$ für $n \to \infty$; siehe den Beweis des folgenden Satzes 4.8.

Satz 4.8 ***(Portemanteau–Theorem).*** *Seien P und P_n ($n \in \mathbb{N}$) Wahrscheinlichkeitsmaße auf $\mathcal{X}$. Dann sind die folgenden fünf Aussagen äquivalent:*

(SK.1) *Für jede beschränkte, stetige Funktion $f : \mathcal{X} \to \mathbb{R}$ ist*

$$\lim_{n\to\infty} \int f\, dP_n \;=\; \int f\, dP.$$

(SK.1') *Für jede Lipschitz-stetige Funktion $f : \mathcal{X} \to [0, 1]$ ist*

$$\lim_{n\to\infty} \int f\, dP_n \;=\; \int f\, dP.$$

(SK.2) *Für jede abgeschlossene Menge $A \subset \mathcal{X}$ ist*

$$\limsup_{n\to\infty} P_n(A) \;\le\; P(A).$$

(SK.3) *Für jede offene Menge $U \subset \mathcal{X}$ ist*

$$\liminf_{n\to\infty} P_n(U) \;\ge\; P(U).$$

(SK.4) *Für jede Borelmenge $B \subset \mathcal{X}$ mit $P(\partial B) = 0$ ist*

$$\lim_{n\to\infty} P_n(B) \;=\; P(B).$$

Definition 4.9 (Schwache Konvergenz). Seien P und P_n ($n \in \mathbb{N}$) Wahrscheinlichkeitsmaße auf einem metrischen Raum $(\mathcal{X}, d)$. Die Folge $(P_n)_{n=1}^{\infty}$ *konvergiert schwach* gegen P, wenn eine (und damit alle) der Bedingungen (SK.X) in Satz 4.8 erfüllt sind. Wir schreiben auch kurz: $P_n \to_w P$ $(n \to \infty)$.

Beispiel 4.10 Sei $P = \mathrm{Unif}\big([0,1]^d\big)$, das heißt, $P(B) = \mathrm{Vol}(B)$ für jedes Rechteck $B \subset \mathcal{X} := [0,1]^d$, und sei $P_n = \mathrm{Unif}(G_n)$ mit der endlichen Menge $G_n := \{1/n, 2/n, \ldots, 1\}^d\big)$. Dann konvergiert $(P_n)_n$ schwach gegen P. Diese Tatsache verwendet man implizit bei Zufallsgeneratoren. Denn an Stelle von nach P verteilten Zufallsvariablen simuliert der Rechner Zufallsvariablen mit einer diskreten Verteilung, die ähnlich ist zu P_n bei astronomisch großem n.

Nachweis von $(P_n)_n \to_w P$: Für jede stetige Funktion $f : [0,1]^d \to \mathbb{R}$ ist

$$\begin{aligned} \int f\,dP_n &= n^{-d} \sum_{y \in G_n} f(y), \\ \int f\,dP &= \sum_{y \in G_n} \int_{B_n(y)} f\,dP, \end{aligned}$$

wobei

$$B_n(y) \;=\; (y_1 - 1/n, y_1] \times (y_2 - 1/n, y_2] \times \cdots \times (y_d - 1/n, y_d].$$

Wegen $P(B_n(y)) = n^{-d}$ ist

$$\begin{aligned} \Big| \int f\,dP_n - \int f\,dP \Big| &= \Big| \sum_{y \in G_n} \int_{B_n(y)} (f(y) - f(x))\,P(dx) \Big| \\ &\le \sup_{x,y \in [0,1]^d \,:\, d(x,y) \le n^{-1}} \big| f(x) - f(y) \big| \\ &\to 0 \quad (n \to \infty), \end{aligned}$$

wobei $d(x,y) := \max_{j=1,\ldots,d} |x_j - y_j|$. Daher ist (SK.1) erfüllt.

Eng verwandt mit dem Konzept der schwachen Konvergenz ist ein Konzept für Zufallsvariablen.

Definition 4.11 (Konvergenz in Verteilung). Seien X und X_n ($n \in \mathbb{N}$) Zufallsvariablen mit Werten in $\mathcal{X}$ (auf beliebigen Wahrscheinlichkeitsräumen). Die Folge $(X_n)_n$ *konvergiert in Verteilung* gegen X, wenn

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

für jede beschränkte, stetige Funktion $f : \mathcal{X} \to \mathbb{R}$. Mit anderen Worten, die Folge der *Verteilungen* der Zufallsvariablen X_n konvergiert schwach gegen die *Verteilung* von X.

Gängige Kurzschreibweisen hierfür sind: $X_n \to_d X$ oder $X_n \to_{\mathcal{L}} X$ für $(n \to \infty)$. (Dabei stehen 'd' und '$\mathcal{L}$' für 'distribution' bzw. 'law'.)

Beweis von Satz 4.8. Dass (SK.1') aus (SK.1) folgt, ist offensichtlich.

Angenommen, es gilt (SK.1'). Für eine beliebige abgeschlossene Menge $A \subset \mathcal{X}$ und $\epsilon > 0$ definieren wir nun
$$f_\epsilon(x) \ := \ \max\Big(1 - \frac{\inf_{y \in A} d(x,y)}{\epsilon}, 0\Big).$$
Man kann zeigen, dass f_ϵ Lipschitz-stetig ist mit Konstante ϵ^{-1}, und
$$1 \ \geq \ f_\epsilon \ \downarrow \ 1_A \quad (\epsilon \downarrow 0).$$

Daher ist
$$\limsup_{n\to\infty} P_n(U) \ \leq \ \limsup_{n\to\infty} \int f_\epsilon \, dP_n \ = \ \int f_\epsilon \, dP$$
für beliebige feste $\epsilon > 0$. Doch für $\epsilon \downarrow 0$ konvergiert die rechte Seite gegen $\int 1_A \, dP = P(A)$. Dies ergibt Aussage (SK.2).

Der vorangehende Beweisschritt ist übrigens für den Namen 'Portemanteau'–Theorem verantwortlich: Die stetige Funktion f_ϵ ist ein 'Mantel', den man über den 'Kleiderständer' 1_A hängt.

Dass die Aussagen (SK.2) und (SK.3) äquivalent sind, ergibt sich durch Betrachtung von komplementären Mengen. Denn das Komplement einer abgeschlossenen (bzw. offenen) Menge ist offen (bzw. abgeschlossen).

Nun seien die Aussagen (SK.2–3) gültig. Dann gilt für jede Borelmenge $B \subset \mathcal{X}$ und ihren Abschluss $\overline{B} \supset B$ bzw. ihr Inneres $B^o \subset B$:
$$\begin{aligned} \limsup_{n\to\infty} P_n(B) \ &\leq \ \limsup_{n\to\infty} P_n(\overline{B}) \ \leq \ P(\overline{B}), \\ \liminf_{n\to\infty} P_n(B) \ &\geq \ \liminf_{n\to\infty} P_n(B^o) \ \geq \ P(B^o). \end{aligned}$$
Wenn nun $P(\partial B) = P(\overline{B} \setminus B^o) = P(\overline{B}) - P(B^o)$ gleich Null ist, dann ist $P(B^o) = P(\overline{B}) = P(B)$ der Grenzwert der Folge $(P_n(B))_n$. Dies beweist Aussage (SK.4).

Zuletzt nehmen wir an, dass Aussage (SK.4) gilt. Nun sei $f : \mathcal{X} \to \mathbb{R}$ stetig und beschränkt. Durch Betrachten von $(f-a)/(b-a)$ mit reellen Zahlen $a < \inf(f)$ und $b > \sup(f)$ können wir uns auf den Fall $0 < f < 1$ zurückziehen. Da die Mengen $\{f = r\}$, $r \in (0,1)$, paarweise disjunkt sind, ist $P(f = r) > 0$ für höchstens abzählbar viele $r \in \mathbb{R}$. Zu jedem $\epsilon > 0$ gibt es also Zahlen $0 = t_0 < t_1 < t_2 < \cdots < t_m = 1$, so dass
$$t_i - t_{i-1} \ \leq \ \epsilon \quad \text{und} \quad P(f = t_i) \ = \ 0 \quad \text{für } 1 \leq i \leq m.$$

Wegen der Stetigkeit von f ist $\{t_{i-1} < f < t_i\}$ eine offene und $\{t_{i-1} \leq f \leq t_i\}$ eine abgeschlossene Teilmenge von $\mathcal{X}$. Also ist $B_i := \{t_{i-1} < f < t_i\}$ eine Borelmenge mit der Eigenschaft, dass $P(\partial B_i) = 0$. Zusammen mit den Ungleichungen
$$\sum_{i=1}^m 1_{B_i} \cdot t_{i-1} \ \leq \ f \ \leq \ \sum_{i=1}^m 1_{B_i} \cdot t_{i-1} + \epsilon$$

ergibt sich also, dass

$$\begin{aligned}
\limsup_{n\to\infty} \int f\,dP_n \;&\leq\; \limsup_{n\to\infty} \sum_{i=1}^{m} P_n(B_i)t_{i-1} + \epsilon \;=\; \sum_{i=1}^{m} P(B_i)t_{i-1} + \epsilon \\
&\leq\; \int f\,dP + \epsilon, \\
\liminf_{n\to\infty} \int f\,dP_n \;&\geq\; \liminf_{n\to\infty} \sum_{i=1}^{m} P_n(B_i)t_{i-1} \;=\; \sum_{i=1}^{m} P(B_i)t_{i-1} \\
&\geq\; \int f\,dP - \epsilon.
\end{aligned}$$

Für $\epsilon \downarrow 0$ ergibt sich dann Aussage (SK.1). □

4.2.2 Der Spezialfall $\mathcal{X} = \mathbb{R}$

Für Wahrscheinlichkeitsmaße auf $\mathbb{R}$ kann man die schwache Konvergenz mit Hilfe von Verteilungs– oder Quantilfunktionen charakterisieren.

Satz 4.12 *Seien P und P_n Wahrscheinlichkeitsmaße auf $\mathbb{R}$ mit Verteilungsfunktionen F bzw. F_n und Quantilfunktionen F^{-1} bzw. F_n^{-1} ($n \in \mathbb{N}$). Dann sind die folgenden vier Aussagen äquivalent:*

(FQ.1) *$(P_n)_n$ konvergiert schwach gegen P;*

(FQ.2) *Für jede unendlich oft differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit beschränktem Träger ist*

$$\lim_{n\to\infty} \int f\,dP_n \;=\; \int f\,dP.$$

(FQ.3) *Für beliebige Zahlen $-\infty < a < b < \infty$ ist*

$$\limsup_{n\to\infty} P_n[a,b] \;\leq\; P[a,b] \quad \textit{und} \quad \liminf_{n\to\infty} P_n(a,b) \;\geq\; P(a,b).$$

(FQ.4) *Für jede Stetigkeitsstelle x von F ist*

$$\lim_{n\to\infty} F_n(x) \;=\; F(x).$$

(FQ.5) *Für jede Stetigkeitsstelle $u \in (0,1)$ von F^{-1} ist*

$$\lim_{n\to\infty} F_n^{-1}(u) \;=\; F^{-1}(u).$$

Anmerkung 4.13 Angenommen, $(P_n)_n$ konvergiert schwach gegen P, und die Verteilungsfunktion F sei stetig. Nach (FQ.3) ist dies gleichbedeutend damit, dass $(F_n)_n$ punktweise gegen F konvergiert. Nun kann man zeigen, dass sogar

$$\sup_{x\in\mathbb{R}} \bigl|F_n(x) - F(x)\bigr| \;\to\; 0 \quad (n \to \infty);$$

siehe Übungen.

Anmerkung 4.14 Angenommen, $(P_n)_n$ konvergiert schwach gegen P. Für eine auf $(0,1)$ uniform verteilte Zufallsvariable U definieren dann $X := F^{-1}(U)$ und $X_n := F_n^{-1}(U)$ Zufallsvariablen mit Verteilung P bzw. P_n, und aus (FQ.4) ergibt sich, dass

$$\lim_{n\to\infty} X_n \ = \ X \quad \text{fast sicher,}$$

denn F^{-1} hat als monotone Funktion höchstens abzählbar viele Unstetigkeitsstellen.

Beispiel 4.15 Sei P_n die Verteilung von Y_n/n, wobei Y_n geometrisch verteilt ist mit Parameter p_n und $np_n \to 1/\lambda$ für ein $\lambda > 0$. Das heißt,

$$\mathbb{P}(Y_n = k) \ = \ (1-p_n)^{k-1}p_n \quad \text{und} \quad \mathbb{P}(Y_n \geq k) \ = \ (1-p_n)^{k-1}$$

für beliebige $k \in \mathbb{N}$. Für die Verteilungsfunktion F_n von P_n gilt also: $F_n \equiv 0$ auf $(-\infty, 0]$, und

$$\begin{aligned} F_n(r) &= 1 - \mathbb{P}(Y_n > nr) \\ &= 1 - (1-p_n)^{\lfloor nr \rfloor} \\ &= 1 - \exp\Big(\frac{\lfloor nr \rfloor}{n}\, n \log(1-p_n)\Big) \\ &\to 1 - \exp(-r/\lambda) \quad (n \to \infty) \end{aligned}$$

für jedes feste $r \geq 0$. Also konvergiert $(F_n)_n$ punktweise gegen die Verteilungsfunktion F der Exponentialverteilung $\mathrm{Exp}(\lambda)$.

Beweis von Satz 4.12. Dass (FQ.2) aus (FQ.1) folgt, ist offensichtlich.

Um zu zeigen, dass (FQ.3) aus (FQ.2) folgt, betrachten wir die Funktion

$$h(x) \ := \ 1\{0 < x < 1\} C \exp\Big(-\frac{1}{x(1-x)}\Big)$$

mit einer Konstante $C > 0$ derart, dass $\int_0^1 h(x)\,dx = 1$. Offensichtlich ist h nichtnegativ, und man kann leicht nachweisen, dass h unendlich oft differenzierbar ist mit beschränkten Ableitungen beliebiger Ordnung. Mit $H(y) := \int_{-\infty}^y h(x)\,dx$ betrachten wir für beliebige reelle Konstanten $a'' < a' < b' < b''$ die Funktion

$$f(x) \ := \ H\Big(\frac{x-a''}{a'-a''}\Big) - H\Big(\frac{x-b'}{b''-b'}\Big).$$

Diese Funktion erfüllt die Voraussetzungen von (FQ.2) sowie die Ungleichungen

$$1_{[a',b']} \ \leq \ f \ \leq \ 1_{(a'',b'')}.$$

Ähnlich wie im Beweis des Portemanteau–Theorems kann man nun schließen, dass (FQ.3) aus (FQ.2) folgt.

Angenommen, es gilt (FQ.3). Dann gilt für feste reelle Zahlen $r < x < s$:

$$\begin{aligned} \liminf_{n\to\infty} F_n(x) &\geq \liminf_{n\to\infty} P_n(r,x) \ \geq \ P(r,s) \quad \text{und} \\ \limsup_{n\to\infty} F_n(x) &\leq 1 - \liminf_{n\to\infty} P_n(x,s) \ \geq \ 1 - P(x,s). \end{aligned}$$

Für $r \to -\infty$ und $s \to \infty$ ergibt sich hieraus, dass

$$\liminf_{n\to\infty} F_n(x) \;\geq\; F(x-) \quad \text{und} \quad \limsup_{n\to\infty} F_n(x) \;\leq\; F(x).$$

Insbesondere ist $\lim_{n\to\infty} F_n(x) = F(x)$, wenn $F(x-) = F(x)$. Dies ergibt Aussage (FQ.4).

Den Nachweis, dass (FQ.4) auch (FQ.5) impliziert, stellen wir als Übungsaufgabe.

Angenommen, es gilt (FQ.5). Für eine auf $(0,1)$ uniform verteilte Zufallsvariable U sind dann $X := F^{-1}(U)$ und $X_n := F_n^{-1}(U)$ nach P bzw. P_n verteilt, und $\lim_{n\to\infty} X_n = X$ fast sicher. Für eine beschränkte stetige Funktion $f : \mathbb{R} \to \mathbb{R}$ ist dann

$$\lim_{n\to\infty} f(X_n) \;=\; f(X) \quad \text{fast sicher,}$$

und $|f(X_{(n)})| \leq \sup_{x\in\mathbb{R}} |f(x)| < \infty$. Nach dem Satz von der majorisierten Konvergenz ist also

$$\lim_{n\to\infty} \int f\,dP_n \;=\; \lim_{n\to\infty} \mathbb{E}\, f(X_n) \;=\; \mathbb{E}\, f(X) \;=\; \int f\,dP,$$

was Aussage (FQ.1) beweist. □

4.3 Normalapproximationen

In diesem Abschnitt möchten wir den Zentralen Grenzwertsatz von Lindeberg beweisen. Um dies vorzubereiten, behandeln wir zunächst Wahrscheinlichkeitsdichten auf dem $\mathbb{R}^d$ und multivariate Normalverteilungen.

4.3.1 Multivariate Dichtefunktionen

In den folgenden Abschnitten betrachten wir das Lebesgue–Maß Leb auf $\mathbb{R}^d$ bzw. $\mathrm{Borel}(\mathbb{R}^d)$, das heißt, $\mathrm{Leb}(B) = \mathrm{Vol}(B)$ für Elementarmengen $B \subset \mathbb{R}^d$. An Stelle von $\int_B f\,d\,\mathrm{Leb}$ verwenden wir die gängigere, an das Riemann-Integral angelehnte Schreibweise $\int_B f(\boldsymbol{x})\,d\boldsymbol{x}$ für messbare Funktionen $f : \mathbb{R}^d \to \mathbb{R}$. Für die konkrete Berechnung solcher Integrale erinnern wir an zwei Hilfsmittel der Analysis:

Der Satz von Fubini (Cavalieri–Prinzip)

Für jede nichtnegative und messbare bzw. Leb–integrierbare Funktion f auf einem Rechteck $B = B_1 \times B_2 \times \cdots \times B_d$ im $\mathbb{R}^d$ ist

$$\begin{aligned} &\int_B f(\boldsymbol{x})\,d\boldsymbol{x} \\ &\quad = \int_{B_{\sigma(d)}} \int_{B_{\sigma(d-1)}} \cdots \int_{B_{\sigma(2)}} \int_{B_{\sigma(1)}} f(x_1, x_2, \ldots, x_p)\,dx_{\sigma(1)}\,dx_{\sigma(2)} \cdots dx_{\sigma(d-1)}\,dx_{\sigma(d)}. \end{aligned}$$

Dabei werden die Integrale von innen nach außen berechnet. Man fixiert also zunächst alle Komponenten von $\boldsymbol{x}$ mit Ausnahme von $x_{\sigma(1)}$ und integriert $f(\boldsymbol{x})$ bezüglich letzterer Komponente über $B_{\sigma(1)}$. Dadurch erhält man eine Funktion von $(x_i)_{i\neq\sigma(1)}$. Diese wird dann bezüglich $x_{\sigma(2)}$ über $B_{\sigma(2)}$ integriert, und so weiter.

Die Transformationsformel

Dieses Hilfsmittel bezieht sich auf glatte Transformationen des Integrationsbereiches. Seien Ω und $\widetilde{\Omega}$ offene Teilmengen des $\mathbb{R}^d$, und sei $T : \Omega \to \widetilde{\Omega}$ eine bijektive und stetig differenzierbare Abbildung mit nichtsingulärer Ableitung (Jacobi–Matrix)

$$DT(\boldsymbol{x}) \;=\; \Bigl(\frac{\partial T_i(\boldsymbol{x})}{\partial x_j}\Bigr)_{i,j=1}^d \;\in\; \mathbb{R}^{d\times d}$$

für alle $\boldsymbol{x} \in \Omega$. Dann gilt für beliebige messbare Funktionen $f : \widetilde{\Omega} \to \mathbb{R}$ die Gleichung

$$\int_{\widetilde{\Omega}} f(\boldsymbol{y})\, d\boldsymbol{y} \;=\; \int_{\Omega} f(T(\boldsymbol{x}))\, |\det DT(\boldsymbol{x})|\, d\boldsymbol{x},$$

sofern $f \geq 0$ oder $\int_{\widetilde{\Omega}} |f(\boldsymbol{y})|\, d\boldsymbol{y} < \infty$.

Heuristische Begründung. Sei $(B_\lambda)_{\lambda\in\Lambda}$ eine Partition von Ω in abzählbar viele und paarweise disjunkte Borelmengen mit kleinem Durchmesser. Wählt man Punkte $\boldsymbol{x}_\lambda \in B_\lambda$, dann ist

$$\int_{\Omega} f(T(\boldsymbol{x}))|\det DT(\boldsymbol{x})|\, d\boldsymbol{x} \;\approx\; \sum_{\lambda\in\Lambda} \mathrm{Leb}(B_\lambda) \cdot f(T(\boldsymbol{x}_\lambda))\, |\det DT(\boldsymbol{x}_\lambda)|.$$

Mit $C_\lambda := T(B_\lambda)$ und $\boldsymbol{y}_\lambda := T(\boldsymbol{x}_\lambda) \in C_\lambda$ ist $(C_\lambda)_{\lambda\in\Lambda}$ eine Partition von $\widetilde{\Omega}$, und

$$\int_{\widetilde{\Omega}} f(\boldsymbol{y})\, d\boldsymbol{y} \;\approx\; \sum_{\lambda\in\Lambda} \mathrm{Leb}(C_\lambda) \cdot f(\boldsymbol{y}_\lambda).$$

Wählt man speziell eine Partition von Ω in kleine *Rechtecke* B_λ, dann ist

$$\mathrm{Leb}(C_\lambda) \cdot f(\boldsymbol{y}_\lambda) \;\approx\; \mathrm{Leb}(B_\lambda) \cdot f(T(\boldsymbol{x}_\lambda)\, |\det DT(\boldsymbol{x}_\lambda)|.$$

Denn auf B_λ kann man T durch die affin lineare Funktion

$$\boldsymbol{x} \;\mapsto\; S_\lambda(\boldsymbol{x}) := T(\boldsymbol{x}_\lambda) + DT(\boldsymbol{x}_\lambda)(\boldsymbol{x} - \boldsymbol{x}_\lambda)$$

approximieren, und aus der linearen Algebra ist bekannt, dass $\mathrm{Vol}(S_\lambda(B)) = |\det DT(\boldsymbol{x}_\lambda)| \cdot \mathrm{Vol}(B)$ für beliebige Rechtecke $B \subset \mathbb{R}^d$.

Aus dieser heuristischen Begründung kann man einen richtigen Beweis machen, indem man die Behauptung durch Approximationsargumente auf den Fall einer stetigen Funktion $f : \widetilde{\Omega} \to \mathbb{R}$ mit kompaktem Träger zurückführt …

Ein erstes Anwendungsbeispiel

Als Anwendung beider Hilfsmittel zeigen wir, dass

$$J := \int_{\mathbb{R}} \exp(-x^2/2)\,dx \;=\; \sqrt{2\pi}.$$

Zu diesem Zweck betrachten wir die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(\boldsymbol{x}) \;:=\; \exp(-\|\boldsymbol{x}\|^2/2) \;=\; \exp(-x_1^2/2)\exp(-x_2^2/2).$$

Aus dem Satz von Fubini ergibt sich einerseits, dass

$$\int_{\mathbb{R}^2} f(\boldsymbol{x})\,d\boldsymbol{x} \;=\; \Big(\int_{\mathbb{R}} \exp(-x^2/2)\,dx\Big)^2 \;=\; J^2.$$

Andererseits ist

$$\int_{\mathbb{R}^2} f(\boldsymbol{x})\,d\boldsymbol{x} \;=\; \int_{\widetilde{\Omega}} f(\boldsymbol{y})\,d\boldsymbol{y},$$

wobei

$$\widetilde{\Omega} \;:=\; \mathbb{R}^2 \setminus \big\{(x,0) : x \geq 0\big\} \;=\; T(\Omega)$$

mit

$$\Omega \;:=\; (0,\infty)\times(0,2\pi) \quad \text{und} \quad T(r,\gamma) \;:=\; (r\cos\gamma, r\sin\gamma).$$

Hier ist

$$DT(r,\gamma) \;=\; \begin{pmatrix} \cos\gamma & -r\sin\gamma \\ \sin\gamma & r\cos\gamma \end{pmatrix},$$

also $\det DT(r,\gamma) = r$, und $f(T(r,\gamma)) = \exp(-r^2/2)$. Folglich ist

$$\begin{aligned}
\int_{\widetilde{\Omega}} f(\boldsymbol{y})\,d\boldsymbol{y} \;&=\; \int_{\Omega} f(T(r,\gamma))\,|\det DT(r,\gamma)|\,d(r,\gamma) \\
&=\; \int_{(0,\infty)\times(0,2\pi)} \exp(-r^2/2)\,r\,d(r,\gamma) \\
&=\; \int_0^\infty \int_0^{2\pi} \exp(-r^2/2)\,r\,d\gamma\,dr \\
&=\; 2\pi \int_0^\infty \exp(-r^2/2)\,r\,dr \\
&=\; 2\pi\Big(-\exp(-r^2/2)\Big)\Big|_{r=0}^{\infty} \\
&=\; 2\pi.
\end{aligned}$$

Wahrscheinlichkeitsdichten

Eine messbare Funktion $f : \mathbb{R}^d \to [0,\infty)$ mit

$$\int_{\mathbb{R}^d} f(\boldsymbol{x})\,d\boldsymbol{x} \;=\; 1$$

heißt *(Lebesgue–) Wahrscheinlichkeitsdichte(funktion)*. Sie induziert ein Wahrscheinlichkeitsmaß P auf $\mathbb{R}^d$, nämlich

$$P(B) \;:=\; \int_B f(\boldsymbol{x})\,d\boldsymbol{x}.$$

Aus dem Satz von Fubini ergibt sich ein einfaches Kriterium für stochastische Unabhängigkeit reellwertiger Zufallsvariablen:

Lemma 4.16 ***(Produktdichten)***. *Für $d \in \mathbb{N}$ seien $P_1, P_2, \ldots, P_d$ Wahrscheinlichkeitsmaße auf $\mathbb{R}$ mit Dichtefunktionen $f_1, f_2, \ldots, f_d$. Dann definiert*

$$f(\boldsymbol{x}) \;:=\; f_1(x_1)f_2(x_2)\cdots f_d(x_d)$$

eine Wahrscheinlichkeitsdichte auf $\mathbb{R}^d$ und beschreibt die Verteilung eines Zufallsvektors $\boldsymbol{X}$ mit stochastisch unabhängigen Komponenten, wobei X_i nach P_i verteilt ist.

Beweis von Lemma 4.16. Für ein beliebiges Rechteck $B := B_1 \times B_2 \times \cdots \times B_d$ im $\mathbb{R}^d$ folgt aus dem Satz von Fubini, dass

$$\int_B f(\boldsymbol{x})\,d\boldsymbol{x} \;=\; \prod_{i=1}^d \int_{B_i} f_i(x)\,dx \;=\; \prod_{i=1}^d P_i(B_i).$$

Setzt man $B_i = \mathbb{R}$ für alle i, dann zeigt sich, dass f eine Wahrwscheinlichkeitsdichte ist. Für einen Zufallsvektor $\boldsymbol{X} = (X_i)_{i=1}^d$ mit Dichtefunktion f ist also

$$\mathbb{P}(\boldsymbol{X} \in B) \;=\; \prod_{i=1}^d P_i(B_i).$$

Fixiert man nun noch einen beliebigen Index j und setzt $B_i := \mathbb{R}$ für alle $i \neq j$, dann ergibt sich die Identität $\mathbb{P}(X_j \in B_j) = P_j(B_j)$. Also ist X_j tatsächlich nach P_j verteilt. □

Eine einfache Anwendung der Transformationsformel bezieht sich auf *affin lineare* Transformationen eines Zufallsvektors.

Lemma 4.17 ***(Affin lineare Transformationen)***. *Sei $\boldsymbol{X}$ ein Zufallsvektor mit Werten in $\mathbb{R}^d$ und Dichtefunktion f. Für einen festen Vektor $\boldsymbol{a} \in \mathbb{R}^d$ und eine feste nichtsinguläre Matrix $\boldsymbol{B} \in \mathbb{R}^{d\times d}$ sei*

$$\boldsymbol{Y} \;:=\; \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}.$$

Dann hat dieser Zufallsvektor $\boldsymbol{Y}$ die Dichtefunktion

$$\boldsymbol{y} \;\mapsto\; g(\boldsymbol{y}) := |\det \boldsymbol{B}|^{-1} f(\boldsymbol{B}^{-1}(\boldsymbol{y} - \boldsymbol{a})).$$

Beweis von Lemma 4.17. Wir schreiben $\boldsymbol{Y} = T(\boldsymbol{X})$ mit der bijektiven Transformation $T : \mathbb{R}^d \to \mathbb{R}^d$, $T(\boldsymbol{x}) := \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}$. Die Umkehrabbildung ist gegeben durch $T^{-1}(\boldsymbol{y}) := \boldsymbol{B}^{-1}(\boldsymbol{y} - \boldsymbol{a})$, und $\det DT(\boldsymbol{x}) = \det \boldsymbol{B}$. Folglich gilt für beliebige Borelmengen $C \subset \mathbb{R}^d$:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y} \in C) &= P(T(\boldsymbol{X}) \in C) \\
&= \int_{\mathbb{R}^d} 1\{T(\boldsymbol{x}) \in C\} f(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} 1\{T(\boldsymbol{x}) \in C\} f(T^{-1}(T(\boldsymbol{x})))\, d\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} 1\{T(\boldsymbol{x}) \in C\} |\det \boldsymbol{B}|^{-1} f(T^{-1}(T(\boldsymbol{x})))\, |\det DT(\boldsymbol{x})|\, d\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} 1\{\boldsymbol{y} \in C\} |\det \boldsymbol{B}|^{-1} f(T^{-1}(\boldsymbol{y}))\, d\boldsymbol{y} \\
&= \int_C g(\boldsymbol{y})\, d\boldsymbol{y}. \qquad \square
\end{aligned}$$

Aus der zuletzt betrachteten Transformationsformel ergibt sich eine Aussage über die Summe von stochastisch unabhängigen Zufallsvariablen:

Lemma 4.18 ***(Faltung zweier Dichten).*** *Seien X und Y stochastisch unabhängige Zufallsvariablen mit Dichtefunktionen f bzw. g auf $\mathbb{R}$. Dann ist $Z := X + Y$ nach der Dichtefunktion $f * g$ verteilt, wobei*

$$f * g\,(z) \;:=\; \int_{\mathbb{R}} f(x)g(z-x)\, dx \;=\; \int_{\mathbb{R}} f(z-y)g(y)\, dy.$$

Die in Lemma 4.18 auftretende Dichtefunktion $f * g$ nennt man auch die *Faltung von f und g*.

Beweis von Lemma 4.18. Der Zufallsvektor $\boldsymbol{W} := (X, Y) \in \mathbb{R}^2$ ist nach der Dichtefunktion $h(x,y) := f(x)g(y)$ verteilt. Nun definiert $T(x,y) := (x, x+y)$ eine bijektive lineare Abbildung $T : \mathbb{R}^2 \to \mathbb{R}^2$ mit Umkehrabbildung $T^{-1}(x,z) := (x, z-x)$, und

$$DT(x,y) \;=\; \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

also $DT(x,y) = 1$. Daher ist $(X, X+Y)$ nach der Dichtefunktion

$$g(x,z) \;:=\; h(T^{-1}(x,z)) \;=\; f(x)g(z-x)$$

verteilt. Insbesondere folgt aus dem Satz von Fubini, dass

$$\begin{aligned}
\mathbb{P}(Z \in B) &= \mathbb{P}\big((X,Z) \in \mathbb{R} \times B\big) \\
&= \int_{\mathbb{R}\times B} f(x)g(z-x)\, d(x,z) \\
&= \int_B \int_{\mathbb{R}} f(x)g(z-x)\, dx\, dz \\
&= \int_B f * g\,(z)\, dz
\end{aligned}$$

für beliebige Borelmengen $B \subset \mathbb{R}$.

Die alternative Formel für die Faltung ergibt sich durch Betrachten der Transformation $S(x, y) := (x + y, y)$ mit Umkehrabbildung $S^{-1}(z, y) = (z - y, y)$. $\square$

Beispiel 4.19 Seien X und Y stochastisch unabhängig und auf $[0, 1]$ uniform verteilt. Dann ist $X + Y$ nach der Dichtefunktion

$$h(z) \ := \ \max(1 - |z - 1|, 0)$$

verteilt. Denn X und Y sind nach der Dichtefunktion $f(x) := 1\{0 < x < 1\}$ verteilt, und

$$\begin{aligned} f * f\,(z) &= \int_{\mathbb{R}} f(x) f(z - x)\, dx \\ &= \int_{\mathbb{R}} 1\{0 < x < 1, 0 < z - x < 1\}\, dx \\ &= \mathrm{Leb}\big\{x \in \mathbb{R} : \max(0, 1 - z) < x < \min(1, z)\big\} \\ &= \max(1 - |z - 1|, 0). \end{aligned}$$

4.3.2 Normalverteilungen

Im vorherigen Abschnitt über Wahrscheinlichkeitsdichten auf dem $\mathbb{R}^p$ haben wir einige Hilfsmittel erarbeitet, mit denen wir nun die Normalverteilungen auf dem $\mathbb{R}^d$ einführen können.

Definition 4.20 (Standardnormalverteilung). Die *Standardnormalverteilung auf* $\mathbb{R}$ ist definiert als das Wahrscheinlichkeitsmaß auf $\mathrm{Borel}(\mathbb{R})$ mit Dichtefunktion

$$z \ \mapsto \ \phi(z) \ := \ (2\pi)^{-1/2} \exp(-z^2/2).$$

Die entsprechende Verteilungsfunktion bezeichnen wir mit Φ, also

$$\Phi(r) \ := \ \int_{-\infty}^{r} \phi(t)\, dt.$$

Eine Zufallsvariable Z mit dieser Verteilung heißt *standardnormalverteilt* und erfüllt die Gleichungen

$$\mathbb{E}(Z) \ = \ 0 \quad \text{und} \quad \mathrm{Var}(Z) \ = \ 1. \tag{4.5}$$

Begründung und Verallgemeinerung von (4.5). Einerseits ist

$$\mathbb{E}(Z) \ = \ \int_{-\infty}^{\infty} x\phi(x)\, dx \ = \ -\int_{-\infty}^{\infty} \frac{d}{dx}\phi(x)\, dx \ = \ -\Big(\phi(x)\Big)\Big|_{x=-\infty}^{\infty} \ = \ 0.$$

Für $k \in \mathbb{N} \setminus \{1\}$ erhält man mit partieller Integration die Formel

$$\begin{aligned}
\mathbb{E}(Z^k) &= \int_{-\infty}^{\infty} x^k \phi(x)\,dx \\
&= -\int_{-\infty}^{\infty} x^{k-1} \frac{d}{dx}\phi(x)\,dx \\
&= -\Big(x^{k-1}\phi(x)\Big)\Big|_{x=-\infty}^{\infty} + (k-1)\int_{-\infty}^{\infty} x^{k-2}\phi(x)\,dx \\
&= (k-1)\,\mathbb{E}(Z^{k-2}).
\end{aligned}$$

Speziell für $k = 2$ ergibt sich, dass $\mathbb{E}(Z^2) = 1$. Induktiv kann man aus der Formel $\mathbb{E}(Z^k) = (k-1)\,\mathbb{E}(Z^{k-2})$ ableiten, dass

$$\mathbb{E}(Z^{2m-1}) = 0 \quad \text{und} \quad \mathbb{E}(Z^{2m}) = \prod_{i=1}^{m}(2i-1) \quad \text{für } m \in \mathbb{N}. \qquad \square$$

Definition 4.21 (Allgemeine Normalverteilungen). Für reelle Zahlen μ und $\sigma \geq 0$ definiert man die *Normalverteilung mit Mittelwert* μ *und Varianz* σ^2 als die Verteilung von $X := \mu + \sigma Z$, wobei Z standardnormalverteilt ist. Insbesondere ist

$$\mathbb{E}(X) = \mu \quad \text{und} \quad \mathrm{Var}(X) = \sigma^2.$$

Als Symbol für diese Verteilung verwendet man auch $\mathcal{N}(\mu, \sigma^2)$.

Anmerkung 4.22 Im Falle von $\sigma > 0$ wird $\mathcal{N}(\mu, \sigma^2)$ durch die Dichtefunktion

$$x \mapsto \frac{1}{\sigma}\phi\Big(\frac{x-\mu}{\sigma}\Big) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)$$

beschrieben, was sich aus Lemma 4.17 ergibt. Diese Dichtefunktion ist um μ symmetrisch mit Maximum bei μ und Wendepunkten $\mu \pm \sigma$.

Wir beenden diesen Abschnitt mit zwei bemerkenswerten Eigenschaften normalverteilter Zufallsvariablen:

Lemma 4.23 ***(Rotationsinvarianz).*** *Für $d \in \mathbb{N}$ sei $\boldsymbol{Z} = (Z_i)_{i=1}^d$ ein Zufallsvektor mit stochastisch unabhängigen, standardnomalverteilten Komponenten Z_i, und $\boldsymbol{B} \in \mathbb{R}^{d\times d}$ sei eine feste orthonormale Matrix, d.h. $\boldsymbol{B}^{-1} = \boldsymbol{B}^\top$. Dann ist $\boldsymbol{BZ}$ genauso verteilt wie $\boldsymbol{Z}$.*

Lemma 4.24 ***(Summen unabhängiger, normalverteilter Zufallsvariablen).*** *Für $d \in \mathbb{N}$ seien $X_1, X_2, \ldots, X_d$ stochastisch unabhängige Zufallsvariablen, wobei $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Dann ist*

$$\sum_{i=1}^d X_i \sim \mathcal{N}\Big(\sum_{i=1}^d \mu_i, \sum_{i=1}^d \sigma_i^2\Big).$$

Beweis von Lemma 4.23. Nach Voraussetzung ist $\boldsymbol{Z}$ nach der Dichtefunktion

$$\phi(\boldsymbol{z}) \ := \ \prod_{i=1}^{d} \phi(z_i) \ = \ (2\pi)^{-d/2} \exp\bigl(-\|\boldsymbol{z}\|^2/2\bigr)$$

auf $\mathbb{R}^d$ verteilt. Gemäß Lemma 4.17 ist der Zufallsvektor $\boldsymbol{Y} := \boldsymbol{B}\boldsymbol{Z}$ nach der Dichtefunktion

$$g(\boldsymbol{y}) \ = \ |\det \boldsymbol{B}|^{-1} \phi(\boldsymbol{B}^{-1}\boldsymbol{y}) \ = \ \phi(\boldsymbol{B}^\top \boldsymbol{y}) \ = \ \phi(\boldsymbol{y})$$

verteilt. Dabei verwendeten wir die Tatsachen, dass $\det \boldsymbol{B} \in \{-1, 1\}$ und $\|\boldsymbol{B}^{-1}\boldsymbol{y}\| = \|\boldsymbol{y}\|$ für beliebige $\boldsymbol{y} \in \mathbb{R}^d$. $\square$

Beweis von Lemma 4.24. Wir schreiben $X_i = \mu_i + \sigma_i Z_i$ mit stochastisch unabhängigen, standardnormalverteilten Zufallsvariablen $Z_1, Z_2, \ldots, Z_d$. Dann ist

$$\sum_{i=1}^{d} X_i \ = \ \sum_{i=1}^{d} \mu_i + \sum_{i=1}^{d} \sigma_i Z_i \ = \ \sum_{i=1}^{d} \mu_i + \sigma \sum_{i=1}^{d} b_i Z_i$$

mit $\sigma := \bigl(\sum_{i=1}^d \sigma_i^2\bigr)^{1/2}$ und einem geeigneten Einheitsvektor $\boldsymbol{b} = (b_i)_{i=1}^d$. Es genügt nun zu zeigen, dass $\sum_{i=1}^d b_i Z_i = \boldsymbol{b}^\top \boldsymbol{Z}$ standardnormalverteilt ist. Mit Hilfe des Basisergänzungssatzes und des Gram–Schmidt–Verfahrens findet sich eine Orthonormalbasis $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_d$ des $\mathbb{R}^d$ derart, dass $\boldsymbol{b}_1 = \boldsymbol{b}$. Dann definiert $\boldsymbol{B} := [\boldsymbol{b}_1\, \boldsymbol{b}_2\, \ldots\, \boldsymbol{b}_d]^\top$ eine orthogonale Matrix, und $\boldsymbol{b}^\top \boldsymbol{Z}$ ist die erste Komponente des Zufallsvektors $\boldsymbol{B}\boldsymbol{Z}$. Doch diese ist nach Lemma 4.23 genauso wie Z_1 verteilt, also standardnormalverteilt. $\square$

Ein alternativer Beweis von Lemma 4.24 mit Hilfe momenterzeugender Funktionen wird in den Übungen behandelt.

4.3.3 Der zentrale Grenzwertsatz

In diesem Abschnitt präzisieren und beweisen wir die folgende Aussage: *"Eine Summe stochastisch unabhängiger Zufallsvariablen ist näherungsweise normalverteilt, wenn jeder einzelne Summand nur geringen Einfluss auf die Gesamtsumme hat."*

Satz 4.25 ***(Lindeberg).*** *Sei $S = \sum_{i=1}^n Y_i$ mit stochastisch unabhängigen Zufallsvariablen Y_1, $Y_2, \ldots, Y_n$, wobei $\mathbb{E}(Y_i) = 0$ für alle i, und*

$$\mathrm{Var}(S) = \sum_{i=1}^{n} \mathbb{E}(Y_i^2) \ = \ 1.$$

Ferner sei

$$L \ := \ \sum_{i=1}^{n} \mathbb{E} \min(Y_i^2, |Y_i|^3) \ \in \ (0, 1].$$

Dann gilt für jede dreifach differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $|f''| \leq 1$ und $|f'''| \leq 1$ die Ungleichung

$$\Big| \mathbb{E} f(S) - \int f(x)\phi(x)\,dx \Big| \;\leq\; 2L.$$

Korollar 4.26 *Sei $S = \sum_{i=1}^n Y_i$ wie in Satz 4.25. Dann ist*

$$\sup_{r \in \mathbb{R}} \big| \mathbb{P}(S \leq r) - \Phi(r) \big| \;\leq\; C L^{1/4}$$

mit einer universellen Konstante $C < \infty$.

Anmerkung 4.27 Die obere Schranke in Korollar 4.26 kann man mit einer verfeinerten, aber aufwändigen Beweistechnik (Steinsche Methode) durch die deutlich präzisere Schranke

$$(32/\pi)^{1/2}\, L^{1/2}$$

ersetzen. In Spezialfällen gibt es weitere Verfeinerungen, beispielsweise die Ungleichungen von Berry–Esseen.

Anmerkung 4.28 (Interpretation von L). Die Kenngröße L ist ein Maß dafür, wie stark sich einzelne Summanden Y_i auf die Gesamtsumme S auswirken. Wenn beispielsweise alle Zufallsvariablen $|Y_i|$ durch eine Konstante $\kappa < 1$ nach oben beschränkt sind, dann ist

$$L \;\leq\; \sum_{i=1}^n \mathbb{E}(\kappa Y_i^2) \;=\; \kappa \sum_{i=1}^n \mathbb{E}(Y_i^2) \;=\; \kappa.$$

Beispiel 4.29 (Binomialverteilungen). Sei X binomialverteilt mit Parametern $n \in \mathbb{N}$ und $p \in (0,1)$, also inbesondere $\mathbb{E}(X) = np$ und $\mathrm{Std}(X) = \sigma_{n,p} := \sqrt{np(1-p)}$. Dann konvergiert

$$\sup_{r \in \mathbb{R}} \Big| P\Big(\frac{X - np}{\sigma_{n,p}} \leq r \Big) - \Phi(r) \Big|$$

gegen Null für $\sigma_{n,p} \to \infty$.

Denn X ist verteilt wie $\widetilde{X} = \sum_{i=1}^n X_i$ mit unabhängigen, $\{0,1\}$-wertigen Summanden X_i mit $\mathbb{E}\, X_i = p$. Also ist

$$\frac{\widetilde{X} - np}{\sigma_{n,p}} \;=\; \sum_{i=1}^n Y_i \quad \text{mit} \quad Y_i \;:=\; \frac{X_i - p}{\sigma_{n,p}},$$

und $|Y_i| \leq \sigma_{n,p}^{-1}$ für alle i. Nun ergibt sich die Behauptung aus Korollar 4.26 und Anmerkung 4.28.

Beispiel 4.30 (Vorzeichen-Teststatistiken). In der nichtparametrischen Statistik tauchen mitunter Summen $S = \sum_{i=1}^n b_i \xi_i$ mit einem festen Einheitsvektor $\boldsymbol{b} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ und unabhängigen, auf $\{-1, 1\}$ uniform verteilten Zufallsvariablen $\xi_1, \xi_2, \ldots, \xi_n$ auf. Hier ist

$$\mathbb{E}(S) \;=\; 0 \quad \text{und} \quad \mathrm{Var}(S) \;=\; 1.$$

Aus Satz 4.25, angewandt auf $Y_i := b_i\xi_i$, folgt die Aussage, dass

$$\sup_{r\in\mathbb{R}} \bigl|P(S \le r) - \Phi(r)\bigr| \ \to\ 0 \quad \text{für} \ \max_{i=1,\ldots,n} |b_i| \to 0.$$

Manche Autoren verstehen unter dem Zentralen Grenzwertsatz das folgende Resultat:

Korollar 4.31 ***(Stichprobenmittelwerte).*** *Seien $X_1, X_2, X_3, \ldots$ stochastisch unabhängige, identisch verteilte Zufallsvariablen mit Mittelwert μ und endlicher Standardabweichung $\sigma > 0$. Für den Stichprobenmittelwert $\bar{X}_n := n^{-1}\sum_{i=1}^n X_i$ ist dann*

$$\lim_{n\to\infty} \sup_{r\in\mathbb{R}} \left| P\Bigl(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le r\Bigr) - \Phi(r)\right| \ =\ 0.$$

Beweis von Korollar 4.31. Die standardisierte Zufallsvariable

$$\frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\mathrm{Std}(\bar{X}_n)} \ =\ \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

ist gleich $\sum_{i=1}^n Y_{n,i}$ mit

$$Y_{n,i} \ :=\ \frac{X_i - \mu}{\sqrt{n}\,\sigma}.$$

Hier ist $L = L_n$ gleich

$$L_n \ =\ n\,\mathbb{E}\bigl(Y_{n,1}^2 \min(|Y_{n,1}|, 1)\bigr) \ =\ \mathbb{E}\left(\frac{(X_1-\mu)^2}{\sigma^2} \min\Bigl(\frac{|X_1-\mu|}{\sqrt{n}\,\sigma}, 1\Bigr)\right),$$

und nach dem Satz von der majorisierten Konvergenz konvergiert dies gegen Null für $n \to \infty$. □

Beweis von Satz 4.25. Da es nur auf die gemeinsame Verteilung von $Y_1, \ldots, Y_n$ ankommt, wählen wir einen Wahrscheinlichkeitsraum, auf dem unabhängige Zufallsvariablen $Y_1, \ldots, Y_n$ und $Z_1, \ldots, Z_n$ definiert sind, wobei Z_i nach $\mathcal{N}(0, \sigma_i^2)$ verteilt ist mit $\sigma_i := \mathrm{Std}(Y_i)$. Also ist

$$\mathbb{E}(Y_i) \ =\ \mathbb{E}(Z_i) \ =\ 0, \quad \mathbb{E}(Y_i^2) \ =\ \mathbb{E}(Z_i^2) \ =\ \sigma_i^2.$$

Im Folgenden vergleichen wir die Verteilungen von $S = \sum_{i=1}^n Y_i$ und $T := \sum_{j=1}^n Z_j$. Gemäß Lemma 4.24 ist T standardnormalverteilt. Zu zeigen ist also, dass

$$\bigl|\mathbb{E}\,f(S) - \mathbb{E}\,f(T)\bigr| \ \le\ 2L. \tag{4.6}$$

Die wesentliche Idee des Beweises besteht darin, die Summanden Y_i von S nacheinander durch Z_i zu ersetzen, bis man schließlich bei der normalverteilten Summe T landet. Wir betrachten also die $n+1$ Summen

$$\begin{aligned} S = T_0 \ &:=\ Y_1 + Y_2 + Y_3 + \cdots + Y_n, \\ T_1 \ &:=\ Z_1 + Y_2 + Y_3 + \cdots + Y_n, \\ T_2 \ &:=\ Z_1 + Z_2 + Y_3 + \cdots + Y_n, \\ &\ \ \vdots \qquad \vdots \\ T = T_n \ &:=\ Z_1 + Z_2 + Z_3 + \cdots + Z_n. \end{aligned}$$

Je zwei aufeinanderfolgende Summen haben $n-1$ gemeinsame Summanden. Genauer gesagt, ist

$$T_{k-1} \;=\; W_k + Y_k \quad \text{und} \quad T_k \;=\; W_k + Z_k$$

mit

$$W_k \;:=\; \sum_{i\,:\,i<k} Z_i + \sum_{i\,:\,i>k} Y_i$$

für $1 \le k \le n$. Hieraus ergibt sich die Gleichung

$$\mathbb{E}\, f(S) - \mathbb{E}\, f(T) \;=\; \sum_{k=1}^{n} \mathbb{E}\big(f(W_k + Y_k) - f(W_k + Z_k)\big),$$

und es genügt zu zeigen, dass

$$\big|\mathbb{E}\big(f(W_k + Y_k) - f(W_k + Z_k)\big)\big| \;\le\; 2\,\mathbb{E}\min(Y_k^2, |Y_k|^3). \tag{4.7}$$

Aus der Taylorformel folgt, dass für beliebige $w, y, z \in \mathbb{R}$ gilt:

$$f(w+y) \;=\; f(w) + f'(w)y + \frac{f''(w)}{2}\,y^2 + \begin{cases} \dfrac{f''(\xi) - f''(w)}{2}\,y^2 \\ \dfrac{f'''(\eta)}{6}\,y^3 \end{cases}$$

mit geeigneten Zwischenstellen $\xi = \xi(w,y)$ und $\eta = \eta(w,y)$. Also ist

$$f(w+y) - f(w+z) \;=\; f'(w)(y-z) + \frac{f''(w)}{2}\,(y^2 - z^2) + R(w,y,z),$$

wobei

$$|R(w,y,z)| \;\le\; \min(y^2, |y|^3/6) + \min(z^2, |z|^3/6).$$

Setzt man nun (W_k, Y_k, Z_k) anstelle von (s, y, z) ein und bildet Erwartungswerte, dann folgt aus der stochastischen Unabhängigkeit von W_k und (Y_k, Z_k), dass

$$\begin{aligned} \mathbb{E}\big(f'(W_k)(Y_k - Z_k)\big) \;&=\; \mathbb{E}(f'(W_k))\,\underbrace{\mathbb{E}(Y_k - Z_k)}_{=\,0} \;=\; 0, \\ \mathbb{E}\Big(\frac{f''(W_k)}{2}\,(Y_k^2 - Z_k^2)\Big) \;&=\; \mathbb{E}\Big(\frac{f''(S_k)}{2}\Big)\,\underbrace{\mathbb{E}(Y_k^2 - Z_k^2)}_{=\,0} \;=\; 0. \end{aligned}$$

(Dabei ist noch zu beachten, dass $\mathbb{E}(f'(W_k))$ existiert, denn wegen $|f''| \le 1$ ist $|f'(W_k)| \le |f'(0)| + |W_k|$.) Also ist

$$\begin{aligned} &\big|\mathbb{E}\big(f(W_k + Y_k) - f(W_k + Z_k)\big)\big| \\ &\quad=\; |\,\mathbb{E}\, R(W_k, Y_k, Z_k)| \;\le\; \mathbb{E}\min(Y_k^2, |Y_k|^3) + \mathbb{E}(|Z_k|^3/6). \end{aligned}$$

Die Behauptung (4.7) ist somit bewiesen, wenn wir zeigen können, dass

$$\mathbb{E}(|Z_k|^3) \;\le\; 6\,\mathbb{E}\min(Y_k^2, |Y_k|^3).$$

Zum einen ist $\sigma_k^{-1} Z_k$ standardnormalverteilt, also

$$\mathbb{E}(|Z_k|^3) \;=\; \frac{\sigma_k^3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |y|^3 \exp(-y^2/2)\, dy \;=\; \sqrt{\frac{8}{\pi}}\,\sigma_k^3 \;\le\; 2\sigma_k^3.$$

Zu zeigen bleibt also, dass $\sigma_k^3 \le 3\,\mathbb{E}\min(Y_k^2, |Y_k|^3)$. Für beliebige $0 < \epsilon \le 1$ ist

$$\sigma_k^2 = \mathbb{E}(Y_k^2) \ \le\ \epsilon^2 + \mathbb{E}(1\{|Y_k| > \epsilon\}Y_k^2) \ \le\ \epsilon^2 + \frac{\mathbb{E}\big(Y_k^2 \min(|Y_k|, 1)\big)}{\epsilon},$$

und für $\epsilon := (\mathbb{E}\min(Y_k^2, |Y_k|^3))^{1/3}$ ergibt sich, dass σ_k^3 nicht größer ist als $\sqrt{8}\ \mathbb{E}\min(Y_k^2, |Y_k|^3)$. □

Beweis von Korollar 4.26. Mit den Bezeichnungen im Beweis von Satz 4.25 betrachten wir eine dreimal differenzierbare und monoton fallende Funktion $H : \mathbb{R} \to [0,1]$ mit $H(x) = 1$ für $x \le 0$ und $H(x) = 0$ für $x \ge 1$ mit beschränkter dritter Ableitung. Es gibt also eine reelle Konstante D mit $|h''| \le D$ und $|h'''| \le D$. Für beliebige reelle Zahlen $q < r$ mit $r - q \le 1$ folgt dann aus (4.6), dass

$$\Big|\mathbb{E}\,H\Big(\frac{S-q}{r-q}\Big) - \mathbb{E}\,H\Big(\frac{T-q}{r-q}\Big)\Big| \ \le\ 2D(r-q)^{-3}L,$$

denn die Funktion $f(x) := D^{-1}(r-q)^3 H\big((x-q)/(r-q)\big)$ erfüllt die Voraussetzungen von Satz 4.25. Doch

$$1\{x \le q\} \ \le\ H\Big(\frac{x-q}{r-q}\Big) \ \le\ 1\{x < r\},$$

so dass

$$\begin{aligned} \mathbb{P}(S \le q) - \mathbb{P}(T \le q) \quad &\le \quad \mathbb{E}\,H\Big(\frac{S-r}{s-r}\Big) - \mathbb{E}\,H\Big(\frac{T-r+\epsilon}{\epsilon}\Big) + \mathbb{P}(q < T < r) \\ &\le \quad 2D(r-q)^{-3}L + \phi(0)(r-q). \end{aligned}$$

Wählt man $r = q + L^{1/4}$, dann ist diese Schranke gleich $CL^{1/4}$ mit $C = 2D + \phi(0)$. Analog ergibt sich auch, dass $P(S \le r) - P(T \le r) \ge -CL^{1/4}$. □

Schlussbemerkung. Die ersten bekannten Normalapproximationen bezogen sich auf spezielle Verteilungen. Beispielsweise zeigten de Moivre und Laplace mit Hilfe der Stirlingschen Formel, dass

$$\max_{k=0,1,\ldots,n} \Big|\sigma_{n,p}\binom{n}{k}p^k(1-p)^{n-k} - \phi\Big(\frac{k-np}{\sigma_{n,p}}\Big)\Big| \ \to\ 0$$

falls $\sigma_{n,p} = \sqrt{np(1-p)} \to \infty$. Ähnliche “lokale Grenzwertsätze” lassen sich für Poisson– oder Gammaverteilungen beweisen; siehe Übungen.

Kapitel 5

Produktmaße und charakteristische Funktionen

In diesem Kapitel behandeln wir noch zwei Themen, die vielerorts in der Stochastik eine wichtige Rolle spielen.

5.1 Produktmaße und der Satz von Fubini

In Abschnitt 4.3 verwendeten wir bereits den Satz von Fubini für Lebesgue–integrierbare Funktionen $f : \mathbb{R}^d \to \mathbb{R}$. Dahinter stecken viel allgemeinere Tatsachen, die wir nun erklären und illustrieren werden.

5.1.1 Produktmaße

Im Folgenden seien $(\Omega_1, \mathcal{A}_1, M_1)$ und $(\Omega_2, \mathcal{A}_2, M_2)$ Maßräume, wobei wir voraussetzen, dass beide Maße M_i σ*–endlich* sind. Das heißt, es gibt Mengen $B_{i,1} \subset B_{i,2} \subset B_{i,3} \subset \cdots$ aus $\mathcal{A}_i$ derart, dass $M_i(B_{i,k}) < \infty$ für alle $k \in \mathbb{N}$ und $\Omega_i = \bigcup_k B_{i,k}$. Diese Eigenschaft ist trivial für Wahrscheinlichkeitsmaße. Die Verallgemeinerung ist vor allem wichtig, um auch das Zählmaß auf einer abzählbar unendlichen Menge und das Lebesguemaß auf dem $\mathbb{R}^d$ einzubeziehen.

Definition 5.1 (Produkt–σ–Algebra). Das Produkt $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ der σ–Algebren $\mathcal{A}_1$ und $\mathcal{A}_2$ ist definiert als

$$\mathcal{A} \;:=\; \sigma\Big(\big\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\big\}\Big),$$

also eine σ–Algebra über $\Omega := \Omega_1 \times \Omega_2$.

Beispiel 5.2 Identifiziert man den Raum $\mathbb{R}^d$ mit $\mathbb{R}^{d(1)} \times \mathbb{R}^{d(2)}$ für natürliche Zahlen $d(1), d(2)$

mit $d(1) + d(2) = d$, dann ist

$$\mathrm{Borel}(\mathbb{R}^d) \;=\; \mathrm{Borel}(\mathbb{R}^{d(1)}) \otimes \mathrm{Borel}(\mathbb{R}^{d(2)}).$$

Nun möchten wir auch das Produkt $M = M_1 \otimes M_2$ der beiden Maße M_1 und M_2 definieren. Dieses Maß M soll folgende Eigenschaft haben:

$$M(A_1 \times A_2) \;=\; M_1(A_1)M_2(A_2) \quad \text{für beliebige } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2. \tag{5.1}$$

Der folgende Satz besagt, dass genau ein solches Maß existiert, und er zeigt auch, wie man Integrale bezüglich dieses Maßes auf Integrale bezüglich M_1 und M_2 zurückführen kann:

Satz 5.3 ***(Satz von Fubini).*** *Es gibt genau ein Maß $M = M_1 \otimes M_2$ auf $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ mit Eigenschaft (5.1). Für eine beliebige messbare Funktion $f : \Omega \to [0, \infty]$ bzw. eine M–integrierbare Funktion $f : \Omega \to \mathbb{R}$ ist*

$$\begin{aligned}\int_\Omega f\, dM \;&=\; \int_{\Omega_1} \Big(\int_{\Omega_2} f(\omega_1, \omega_2)\, M_2(d\omega_2) \Big)\, M_1(d\omega_1) \\ &=\; \int_{\Omega_2} \Big(\int_{\Omega_1} f(\omega_1, \omega_2)\, M_1(d\omega_1) \Big)\, M_2(d\omega_2).\end{aligned}$$

Insbesondere gilt für $A \in \mathcal{A}$ die Formel

$$\begin{aligned}M(A) \;&=\; \int_{\Omega_1} M_2\big(\{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}\big)\, M_1(d\omega_1) \\ &=\; \int_{\Omega_2} M_1\big(\{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}\big)\, M_2(d\omega_2).\end{aligned}$$

Sowohl Definition 5.1 als auch Satz 5.3 kann man induktiv auf $n \geq 2$ Maßräume $(\Omega_i, \mathcal{A}_i, M_i)$, $1 \leq i \leq n$, übertragen. Dies liefert dann den Maßraum

$$(\Omega, \mathcal{A}, M) \;=\; \Big(\Omega_1 \times \cdots \times \Omega_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n, M_1 \otimes \ldots \otimes M_n\Big),$$

und für jede messbare Funktion $f : \Omega \to [0, \infty]$ bzw. M–integrierbare Funktion $f : \Omega \to \mathbb{R}$ kann man schreiben

$$\int f\, dM \;=\; \int_{\Omega_{\sigma(n)}} \cdots \int_{\Omega_{\sigma(1)}} f(\omega_1, \omega_2, \ldots, \omega_n)\, M_{\sigma(1)}(d\omega_{\sigma(1)}) \cdots M_{\sigma(n)}(d\omega_{\sigma(n)})$$

mit einer beliebigen Permutation σ von $\{1, 2, \ldots, n\}$.

5.1.2 Anwendungen

Konstruktion spezieller Wahrscheinlichkeitsräume. Mit Hilfe von Satz 5.3 kann man endlich viele Wahrscheinlichkeitsräume $(\Omega_i, \mathcal{A}_i, P_i)$, $1 \leq i \leq n$, zu einem Wahrscheinlichkeitsraum

$$(\Omega, \mathcal{A}, P) \;=\; \Big(\Omega_1 \times \cdots \times \Omega_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n, P_1 \otimes \ldots \otimes P_n\Big)$$

zusammenzufassen. Deutet man jeden Wahrscheinlichkeitsraum $(\Omega_i, \mathcal{A}_i, P_i)$ als ein Zufallsexperiment, dann beschreibt $(\Omega, \mathcal{A}, P)$ die "unabhängige" Durchführung dieser n Experimente.

Konkrete Berechnung von Erwartungswerten. Seien X und Y stochastisch unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathcal{X}, \mathcal{B})$ beziehungsweise $(\mathcal{Y}, \mathcal{C})$. Nun betrachten wir eine Zufallsvariable der Form

$$h(X, Y)$$

mit einer messbaren Abbildung h von $\mathcal{X} \times \mathcal{Y}$ nach $[0, \infty]$ bzw. $\mathbb{R}$. Dabei wird $\mathcal{X} \times \mathcal{Y}$ mit der σ–Algebra $\mathcal{B} \otimes \mathcal{C}$ versehen. Für einen festen Punkt $x \in \mathcal{X}$ sei

$$g(x) \ := \ \mathbb{E}\, h(x, Y).$$

Dann gilt die Gleichung

$$\mathbb{E}\, h(X, Y) \ = \ \mathbb{E}\, g(X). \tag{5.2}$$

Mitunter schreibt man auch $g(x) = \mathbb{E}(h(X, Y) \,|\, X = x)$ oder einfach $g(X) = \mathbb{E}(h(X, Y) \,|\, X)$.

Formel (5.2) ergibt sich aus Satz 5.3 wie folgt: Wegen der stochastischen Unabhängigkeit von X und Y ist $\mathbb{P}^{(X,Y)} = \mathbb{P}^X \otimes \mathbb{P}^Y$, so dass

$$\begin{aligned}
\mathbb{E}\, h(X, Y) \ &= \ \int_{\mathcal{X} \times \mathcal{Y}} h \; d\,\mathbb{P}^{(X,Y)} \\
&= \ \int_{\mathcal{X} \times \mathcal{Y}} h \; d\,\mathbb{P}^X \otimes \mathbb{P}^Y \\
&= \ \int_{\mathcal{X}} \Big(\int_{\mathcal{Y}} h(x, y) \ \mathbb{P}^Y(dy) \Big) \ \mathbb{P}^X(dx) \\
&= \ \int_{\mathcal{X}} g(x) \ \mathbb{P}^X(dx) \\
&= \ \mathbb{E}\, g(X).
\end{aligned}$$

Beispiel 5.4 (Ein Inspektionsparadoxon). Eine bestimmte Buslinie fährt laut Fahrplan zu Zeitpunkten $k \in \mathbb{Z}$ in einer geeigneten Zeiteinheit. Angenommen man kommt zu einem “rein zufälligen” Zeitpunkt T an die Haltestelle. Genauer gesagt, sei $T \bmod 1 = T - \lfloor T \rfloor$ uniform verteilt auf $[0, 1)$; siehe auch die nachfolgende Anmerkung 5.5. Wenn alle Busse pünktlich abfahren, dann ist die Wartezeit gleich

$$W \ = \ 1 - (T \bmod 1),$$

und die mittlere Wartezeit beträgt

$$\mathbb{E}(W) \ = \ \int_0^1 (1 - t)\, dt \ = \ 1/2.$$

Nun seien die tatsächlichen Abfahrtszeiten gleich $k + V_k$ mit stochastisch unabhängigen, identisch verteilten Verspätungen $V_k \in [0, 1)$, $k \in \mathbb{Z}$. Wir nehmen außerdem an, dass T von $(V_k)_{k \in \mathbb{Z}}$ stochastisch unabhängig ist. Wie lange muss man jetzt im Mittel auf den Bus warten? Intuitiv würde man vielleicht denken, dass immer noch $\mathbb{E}(W) = 1/2$. Tatsächlich ist aber

$$\mathbb{E}(W) \ = \ 1/2 + \mathrm{Var}(V_0).$$

Um dies nachzuweisen, nehmen wir ohne Einschränkung an, dass T uniform verteilt ist auf $[0,1)$. Dann ist die Wartezeit gleich

$$\begin{aligned} W &= 1\{T \le V_0\}(V_0 - T) + 1\{T > V_0\}(1 + V_1 - T) \\ &= 1\{T \le V_0\}V_0 + 1\{T > V_0\}(1 + V_1) - T. \end{aligned}$$

Betrachtet man vorübergehend (V_0, V_1) als festes Paar und bildet den Erwartungswert bezüglich T, dann ergibt sich die neue Größe

$$\begin{aligned} g(V_0, V_1) &:= \mathbb{E}(W \mid V_0, V_1) \\ &= \mathbb{E}\big(1\{T \le V_0\}V_0 + 1\{T > V_0\}(1 + V_1) - T \,\big|\, V_0, V_1\big) \\ &= \mathbb{P}(T \le V_0 \mid V_0, V_1) \cdot V_0 + \mathbb{P}(T > V_0 \mid V_0, V_1) \cdot (1 + V_1) - \mathbb{E}(T \mid V_0, V_1) \\ &= V_0^2 + (1 - V_0)(1 + V_1) - 1/2 \\ &= 1/2 + V_0^2 - V_0 + V_1 - V_0 V_1. \end{aligned}$$

Bildet man nun den Erwartungswert der rechten Seite, dann folgt aus der stochastischen Unabhängigkeit und identischen Verteilung von V_0 und V_1, dass

$$\begin{aligned} \mathbb{E}(W) &= 1/2 + \mathbb{E}(V_0^2) - \mathbb{E}(V_0) + \mathbb{E}(V_1) - \mathbb{E}(V_0 V_1) \\ &= 1/2 + \mathbb{E}(V_0^2) - \mathbb{E}(V_0)^2 \\ &= 1/2 + \mathrm{Var}(V_0). \end{aligned}$$

Hier ist eine heuristische Erklärung für dieses Phänomen: Die Zeitintervalle zwischen den Abfahrtszeiten sind unterschiedlich lang. Bei zufälliger Ankunft an der Haltestelle sind die Chancen, in einem bestimmten Zeitintervall $[k + V_k, k + 1 + V_{k+1})$ zu landen, umso größer, je länger dieses Intervall ist. Bei gegebenem Ankunftsintervall ist die erwartete Wartezeit gleich der halben Intervalllänge. Alles in allem spielt also das *Quadrat* der Intervalllänge eine Rolle.

Anmerkung 5.5 (Rundungsreste). Ist T eine Zufallsvariable mit Dichtefunktion f auf $\mathbb{R}$, dann ist $T \bmod 1 := T - \lfloor T \rfloor$ eine Zufallsvariable mit Dichtefunktion

$$u \mapsto \widetilde{f}(u) := \sum_{z \in \mathbb{Z}} f(z + u)$$

auf $[0,1)$. Falls f unimodal ist, das heißt, für ein $\mu \in \mathbb{R}$ ist f monoton wachsend auf $(-\infty, \mu]$ und monoton fallend auf $[\mu, \infty)$, dann ist

$$\sup_{u,v \in [0,1)} \big|\widetilde{f}(u) - \widetilde{f}(v)\big| \le f(\mu)$$

(Beweis als Übungsaufgabe). Insbesondere ist $|\widetilde{f} - 1| \le f(\mu)$. Wenn also die Verteilung von T “breit gestreut” ist im Sinne von $f(\mu) << 1$, so ist $T \bmod 1$ approximativ uniform verteilt auf $[0,1)$. Beispielsweise ist $f(\mu) = (2\pi\sigma^2)^{-1/2} << 1$, falls $T \sim \mathcal{N}(\mu, \sigma^2)$ mit $\sigma >> 1$.

Beispiel 5.6 (Polyas Urnenmodell). Dieses Modell wurde bereits in Abschnitt 1.1.1 (Schlüsselbeispiel 2) beschrieben: In einer Urne befinden sich zunächst zwei Kugeln, eine mit '1' und eine mit '0' beschriftet. Zum Zeitpunkt $n \in \mathbb{N}$ zieht man rein zufällig eine Kugel und notiert die daraufstehende Zahl ($\rightsquigarrow X_n$). Danach legt man diese Kugel zusammen mit einer Kopie hiervon zurück. Dieses Gedankenmodell führt zu einer Zufallsfolge $\boldsymbol{X} = (X_n)_{n=1}^{\infty}$ in $\{0,1\}$, und in den Übungen wurde gezeigt, dass für beliebige $n \in \mathbb{N}$ und $\boldsymbol{x} \in \{0,1\}^n$ gilt:

$$p_n(\boldsymbol{x}) := \mathbb{P}\big((X_i)_{i=1}^n = \boldsymbol{x}\big) \ = \ \frac{s(\boldsymbol{x})!(n-s(\boldsymbol{x}))!}{(n+1)!} \ = \ (n+1)^{-1}\binom{n}{s(\boldsymbol{x})}^{-1}$$

mit $s(\boldsymbol{x}) := \sum_{i=1}^n x_i$. Simulationen dieses Urnenmodells suggerieren, dass der relative Anteil

$$V_n \ := \ \frac{1+\sum_{i=1}^n X_i}{n+2}$$

von Kugeln des Typs '1' für $n \to \infty$ fast sicher gegen eine Zufallsvariable V mit Werten in $[0,1]$ konvergiert. In der Tat ist die Folge $\boldsymbol{X}$ genauso verteilt wie die folgende Zufallssequenz $\tilde{\boldsymbol{X}}$:

Seien $V, U_1, U_2, U_3, \ldots$ stochastisch unabängig und auf $[0,1]$ uniform verteilt. Dann definieren wir $\tilde{\boldsymbol{X}} = (\tilde{X}_n)_{n=1}^{\infty}$ durch

$$\tilde{X}_n \ := \ 1\{U_n \le V\}.$$

Mit anderen Worten, wir wählen einen zufälligen Parameter $V \sim \mathrm{Unif}[0,1]$, und danach erzeugen wir bei gegebenem V einen unendlichen Münzwurf mit $\mathrm{Ws}(\text{'Zahl'}) = V$.

Um nachzuweisen, dass $\boldsymbol{X}$ und $\widetilde{\boldsymbol{X}}$ tatsächlich identisch verteilt sind, müssen wir nur zeigen, dass $\mathbb{P}\big((\tilde{X}_i)_{i=1}^n = \boldsymbol{x}\big) = p_n(\boldsymbol{x})$ für beliebige $n \in \mathbb{N}$ und $\boldsymbol{x} \in \{0,1\}^n$. Nach dem Satz von Fubini ist

$$\begin{aligned}
\mathbb{P}\big((\tilde{X}_i)_{i=1}^n = \boldsymbol{x}\big) \ &= \ \mathbb{E}\,\mathbb{P}\big((\tilde{X}_i)_{i=1}^n = \boldsymbol{x} \,\big|\, V\big) \\
&= \ \mathbb{E}\prod_{i=1}^n V^{x_i}(1-V)^{1-x_i} \\
&= \ \mathbb{E}\big(V^{s(\boldsymbol{x})}(1-V)^{n-s(\boldsymbol{x})}\big) \\
&= \ \int_0^1 v^{s(\boldsymbol{x})}(1-v)^{n-s(\boldsymbol{x})}\,dv \\
&= \ (n+1)^{-1}\binom{n}{s(\boldsymbol{x})}^{-1};
\end{aligned}$$

siehe Formel (2.6).

5.1.3 Beweisskizze für Satz 5.3*

Reduktion auf endliche Maße M_i. Angenommen, der Satz wurde bereits für den Fall, dass $M_1(\Omega_1) < \infty$ und $M_2(\Omega_2) < \infty$ bewiesen. Dann kann man im allgemeinen Fall die Maße

$$M_i^{(k)}(A_i) \ := \ M_i(A_i \cap B_{i,k})$$

und $M^{(k)} := M_1^{(k)} \otimes M_2^{(k)}$ betrachten. Die Behauptungen ergeben sich dann im Wesentlichen durch den Grenzübergang $k \to \infty$ und dem Satz von der monotonen Konvergenz.

Im Folgenden seien also $M_1(\Omega_1)$ und $M_2(\Omega_2)$ endlich.

Eindeutigkeit von M. Da das System aller Mengen $A_1 \times A_2$ mit $A_i \in \mathcal{A}_i$ durchschnittstabil ist, ergibt sich die Eindeutigkeit von M aus dem Eindeutigkeitssatz 1.24.

Die Integralformel. Falls die behauptete Darstellung von $M(A)$ als Doppelintegral bewiesen wurde, dann gilt diese Darstellung auch für einfache messbare Funktionen $f : \Omega \to [0, \infty]$, und mit dem Satz von der monotonen Konvergenz ergibt sich der allgemeine Fall einer messbaren Funktion $f : \Omega \to [0, \infty]$. Eine M–integrierbare Funktion f wird dann, wie üblich, in Positiv– und Negativteil aufgespalten.

Dynkin–Systeme. Eine Familie $\mathcal{A}$ von Teilmengen einer Menge Ω heißt *Dynkin–System über* Ω, wenn die folgenden drei Bedingungen erfüllt sind:

(D.1) $\Omega \in \mathcal{A}$;
(D.2) mit $A \in \mathcal{A}$ ist auch $\Omega \setminus A \in \mathcal{A}$;
(D.3) sind $A_1, A_2, A_3, \ldots$ paarweise disjunkte Mengen aus $\mathcal{A}$, dann ist auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Jede σ–Algebra über Ω ist auch ein Dynkin–System. Umgekehrt ist ein Dynkin–System über Ω eine σ–Algebra, falls es durchschnittstabil ist. Diverse Beweise in Wahrscheinlichkeits– und Maßtheorie stützen sich auf das folgende Schlüsselresultat:

Satz 5.7 ***(Dynkin).*** *Sei $\mathcal{A}$ ein Dynkin–System über Ω, welches eine durchschnittstabile Mengenfamilie $\mathcal{D} \subset \mathcal{P}(\Omega)$ enthält. Dann ist $\sigma(\mathcal{D}) \subset \mathcal{A}$.*

Die Konstruktion von $M(\cdot)$. Wir betrachten die Menge $\mathcal{A}_*$ aller Mengen $A \subset \mathcal{A}$, für welche

$$M(A) \;:=\; \int_{\Omega_1} M_2\big(\big\{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\big\}\big)\, M_1(d\omega_1)$$

wohldefiniert ist. Das heißt, für jedes $\omega_1 \in \Omega_1$ gehört die Menge $\big\{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\big\}$ zu $\mathcal{A}_2$, und die Funktion $\omega_1 \mapsto M_1\big(\big\{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\big\}\big)$ ist messbar.

Man kann leicht verifizieren, dass $\mathcal{A}_*$ die Familie $\mathcal{D}$ aller kartesischen Produkte $A_1 \times A_2$ mit $A_i \in \mathcal{A}_i$ enthält, wobei

$$M(A_1 \times A_2) \;=\; M_1(A_1) M_2(A_2).$$

Andererseits kann man aus den allgemeinen Eigenschaften des Lebesgueintegrals ableiten, dass $\mathcal{A}_*$ ein Dynkin–System ist, wobei $M(\emptyset) = 0$ und

$$M\Big(\bigcup_{n=1}^{\infty} A_n\Big) \;=\; \sum_{n=1}^{\infty} M(A_n)$$

für paarweise disjunkte Mengen $A_1, A_2, A_3, \ldots$ aus $\mathcal{A}_*$.

Da unsere spezielle Familie $\mathcal{D}$ durchschnittstabil ist, ergibt sich aus Satz 5.7, dass $\mathcal{A} = \sigma(\mathcal{D})$ in $\mathcal{A}_*$ enthalten ist, also $\mathcal{A}_* = \mathcal{A}$. Insbesondere ist M ein Maß auf $\mathcal{A}$ mit der gewünschten Eigenschaft (5.1).

Analog kann man zeigen, dass

$$M(A) := \int_{\Omega_2} M_1\bigl(\bigl\{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\bigr\}\bigr)\, M_2(d\omega_2)$$

ein Maß auf $\mathcal{A}$ mit Eigenschaft (5.1) definiert. □

Beweis von Satz 5.7. Der Durchschnitt beliebig vieler Dynkinsysteme ist ebenfalls ein Dynkinsystem. Daher ersetzen wir $\mathcal{A}$ durch den Durchschnitt aller Dynkinsysteme, welche $\mathcal{D}$ enthalten. Mit anderen Worten, $\mathcal{A}$ sei das *kleinste* Dynkinsystem, welches $\mathcal{D}$ enthält. Wenn wir zeigen können, dass $\mathcal{A}$ durchschnittstabil ist, dann ist $\mathcal{A}$ selbst eine σ–Algebra und enthält folglich $\sigma(\mathcal{D})$. (Tatsächlich ist dann $\mathcal{A} = \sigma(\mathcal{D})$, denn nach Definition von $\mathcal{A}$ ist $\mathcal{A} \subset \sigma(\mathcal{D})$.)

Für eine beliebige Menge $B \in \mathcal{A}$ sei

$$\mathcal{A}(B) := \bigl\{A \in \mathcal{A} : A \cap B \in \mathcal{A}\bigr\}.$$

Dies definiert ein Dynkinsystem, welches in $\mathcal{A}$ enthalten ist. Im Falle von $B \in \mathcal{D}$ folgt aus der Durchschnittstabilität von $\mathcal{D}$ und der Inklusion $\mathcal{D} \subset \mathcal{A}$, dass $\mathcal{D} \subset \mathcal{A}(B)$. Wegen der Minimalität von $\mathcal{A}$ ist also $\mathcal{A}(B) = \mathcal{A}$. Diese Überlegungen zeigen, dass

$$A \cap B \in \mathcal{A} \quad \text{für alle } A \in \mathcal{A}, B \in \mathcal{D}.$$

Insbesondere ist $\mathcal{D} \subset \mathcal{A}(A)$ für jedes $A \in \mathcal{A}$, und dies zeigt wiederum, dass $\mathcal{A}(A) = \mathcal{A}$ für alle $A \in \mathcal{A}$. Mit anderen Worten,

$$A' \cap A \in \mathcal{A} \quad \text{für alle } A, A' \in \mathcal{A}.$$ □

5.2 Charakteristische Funktionen

Wir lernten bereits momentenerzeugende Funktionen kennen, um Wahrscheinlichkeitsverteilungen auf $\mathbb{R}$ zu charakterisieren. Der Nachteil dieser Methode ist, dass nicht für jedes Wahrscheinlichkeitsmaß P die momentenerzeugende Funktion m_P auf einem offenen Intervall endliche Werte annimmt. Als Ersatz verwendet man gerne charakteristische Funktionen. Diese sind eng verwandt mit der *Fouriertransformation*, die in der Analysis und ihren Anwendungen eine wichtige Rolle spielt.

Um charakteristische Funktionen zu definieren, benötigen wir Integrale und Erwartungswerte von komplexwertigen Funktionen bzw. Zufallsvariablen. Diese sind komponentenweise zu verstehen, das heißt, für $f : \Omega \to \mathbb{C}$ setzen wir

$$\int f\,dM \;:=\; \int \mathrm{Re} f\,dM + \boldsymbol{i}\int \mathrm{Im} f\,dM,$$

wobei wir voraussetzen, dass $\mathrm{Re} f$ und $\mathrm{Im} f$ M–integrierbar sind. Man kann leicht nachrechnen, dass für zwei solche Funktionen f und g sowie Konstanten $\lambda, \mu \in \boldsymbol{C}$ gilt:

$$\begin{aligned} \int (\lambda f + \mu g)\,dM \;&=\; \lambda \int f\,dM + \mu \int g\,dM, \\ \Big|\int f\,dM\Big| \;&\leq\; \int |f|\,dM. \end{aligned}$$

Außerdem gilt für $\mathbb{C}$–wertige Zufallsvariablen X und Y mit $\mathbb{E}\,|X| < \infty$ und $\mathbb{E}\,|Y| < \infty$ nach wie vor die Produktregel:

$$\mathbb{E}(XY) \;=\; \mathbb{E}(X)\,\mathbb{E}(Y) \quad \text{wenn } X \text{ und } Y \text{ stochastisch uanbängig sind.} \tag{5.3}$$

5.2.1 Definition und einige Eigenschaften charakteristischer Funktionen

Definition 5.8 (Charakteristische Funktion). Die *charakteristische Funktion* einer Zufallsvariable $\boldsymbol{X}$ mit Werten in $\mathbb{R}^d$ ist definiert als die Funktion $\varphi_{\boldsymbol{X}} : \mathbb{R}^d \to \mathbb{C}$ mit

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) \;:=\; E\exp(\boldsymbol{i}\langle \boldsymbol{t}, \boldsymbol{X}\rangle) \;=\; \mathbb{E}\cos(\langle \boldsymbol{t}, \boldsymbol{X}\rangle) + \boldsymbol{i}\,\mathbb{E}\sin(\langle \boldsymbol{t}, \boldsymbol{X}\rangle).$$

Dabei ist $\langle \boldsymbol{x}, \boldsymbol{y}\rangle := \sum_{i=1}^d x_i y_i$, das Standardskalarprodukt auf dem $\mathbb{R}^d$. Analog definiert man die charakteristische Funktion eines Wahrscheinlichkeitsmaßes P auf $\mathbb{R}^d$:

$$\varphi_P : \mathbb{R}^d \to \mathbb{C}, \quad \varphi_P(\boldsymbol{t}) \;:=\; \int \exp(\boldsymbol{i}\langle \boldsymbol{t}, \boldsymbol{x}\rangle)\,P(d\boldsymbol{x}).$$

Für $d = 1$ entspricht die charakteristische Funktion φ_P der Funktion $t \mapsto m_P(\boldsymbol{i}t)$. Von daher ist es nicht überraschend, dass charakteristische Funktionen ähnliche Eigenschaften wie momentenerzeugende Funktionen haben.

Satz 5.9 ***(Eigenschaften charakteristischer Funktionen).***

(a) *Die Funktion $\varphi_{\boldsymbol{X}}$ erfüllt die Ungleichung*

$$|\varphi_{\boldsymbol{X}}| \;\leq\; \varphi_{\boldsymbol{X}}(\boldsymbol{0}) \;=\; 1.$$

Sie ist gleichmäßig stetig auf $\mathbb{R}^d$, nämlich

$$\big|\varphi_{\boldsymbol{X}}(\boldsymbol{t}+\boldsymbol{h}) - \varphi_{\boldsymbol{X}}(\boldsymbol{t})\big| \;\leq\; \mathbb{E}\min\big(2, \|\boldsymbol{h}\|\|\boldsymbol{X}\|\big) \text{ für beliebige } \boldsymbol{t}, \boldsymbol{h} \in \mathbb{R}^d.$$

(b) *Ist $\mathbb{E}(\|\boldsymbol{X}\|^m) < \infty$ für ein $m \in \mathbb{N}$, so ist $\varphi_{\boldsymbol{X}}$ m–fach stetig differenzierbar auf $\mathbb{R}^d$. Genauer gesagt, gilt für beliebige Vektoren $\boldsymbol{t}, \boldsymbol{h} \in \mathbb{R}^d$ die Taylorentwicklung*

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}+\boldsymbol{h}) \;=\; \sum_{k=0}^m \frac{\boldsymbol{i}^k}{k!}\,\mathbb{E}\big(\exp(\boldsymbol{i}\langle \boldsymbol{t}, \boldsymbol{X}\rangle)\langle \boldsymbol{h}, \boldsymbol{X}\rangle^k\big) + R_{\boldsymbol{X},m}(\boldsymbol{t}, \boldsymbol{h})$$

wobei

$$\big|R_{\boldsymbol{X},m}(\boldsymbol{t},\boldsymbol{h})\big| \;\le\; \|\boldsymbol{h}\|^m\,\mathbb{E}\min\Big(\frac{2\|\boldsymbol{X}\|^m}{m!},\,\frac{\|\boldsymbol{h}\|\|\boldsymbol{X}\|^{m+1}}{(m+1)!}\Big).$$

(c) *Für zwei stochastisch unabhängige, $\mathbb{R}^d$–wertige Zufallsvariablen $\boldsymbol{X}$ und $\boldsymbol{Y}$ auf $(\Omega,\mathcal{A},\mathbb{P})$ ist*

$$\varphi_{\boldsymbol{X}+\boldsymbol{Y}} \;=\; \varphi_{\boldsymbol{X}}\varphi_{\boldsymbol{Y}}.$$

Teil (c) dieses Satzes ergibt sich direkt aus der Produktformel (5.3). Die Teile (a-b) ergeben sich im Wesentlichen aus der folgenden Ungleichung:

Lemma 5.10 *Für beliebige $m \in \mathbb{N}_0$ und $x \in \mathbb{R}$ ist*

$$\Big|\exp(\boldsymbol{i}x) - \sum_{k=0}^m \frac{(\boldsymbol{i}x)^k}{k!}\Big| \;\le\; \min\Big(\frac{2|x|^m}{m!},\,\frac{|x|^{m+1}}{(m+1)!}\Big).$$

Beweis von Lemma 5.10. Mit $r_m(x) := \exp(\boldsymbol{i}x) - \sum_{k=0}^m (\boldsymbol{i}x)^k/k!$ ist zunächst

$$r_0(x) \;=\; \exp(\boldsymbol{i}x) - 1 \;=\; \boldsymbol{i}\int_0^x \exp(\boldsymbol{i}t)\,dt,$$

und dies zeigt, dass

$$|r_0(x)| \;\le\; \min(2,|x|).$$

Allgemein gilt für $m \in \mathbb{N}_0$ die Darstellung

$$r_{m+1}(x) \;=\; \boldsymbol{i}\int_0^x r_m(t)\,dt.$$

Wenn also $|r_m(x)| \le |x|^{m+1}/(m+1)!$ für beliebige $x \in \mathbb{R}$, dann ist

$$|r_{m+1}(x)| \;\le\; \int_0^{|x|} |r_m(\pm t)|\,dt \;\le\; \int_0^{|x|} \frac{t^{m+1}}{(m+1)!}\,dt \;=\; \frac{|x|^{m+2}}{(m+2)!}.$$

Ferner ist $r_{m+1}(x) = r_m(x) - (\boldsymbol{i}x)^{m+1}/(m+1)!$, so dass

$$|r_{m+1}(x)| \;\le\; |r_m(x)| + \frac{|x|^{m+1}}{(m+1)!} \;\le\; \frac{2|x|^{m+1}}{(m+1)!}. \qquad \square$$

Beispiel 5.11 (Normalverteilte Vektoren). Sei $\boldsymbol{Z} = (Z_j)_{j=1}^d$ ein standardnormalverteilter Zufallsvektor; das heißt, seine Komponenten Z_j seien stochastisch unabhängig und standardnormalverteilt. Dann ist

$$\varphi_{\boldsymbol{Z}}(\boldsymbol{t}) \;=\; \exp(-\|\boldsymbol{t}\|^2/2) \quad \text{für } \boldsymbol{t} \in \mathbb{R}^d.$$

Ferner gilt für feste Vektoren $\boldsymbol{\mu} \in \mathbb{R}^d$ und Matrizen $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ die Formel

$$\varphi_{\boldsymbol{\mu}+\boldsymbol{A}\boldsymbol{Z}}(\boldsymbol{t}) \;=\; \exp\big(\boldsymbol{i}\langle\boldsymbol{t},\boldsymbol{\mu}\rangle\big)\exp\big(-\|\boldsymbol{A}^\top\boldsymbol{t}\|^2/2\big).$$

Denn aus der stochastischen Unabhängigkeit der Z_j folgt, dass

$$\begin{aligned}\varphi_{\boldsymbol{Z}}(\boldsymbol{t}) &= \mathbb{E}\prod_{j=1}^{d}\exp(\boldsymbol{i}t_jZ_j)\\ &= \prod_{j=1}^{d}\mathbb{E}\exp(\boldsymbol{i}t_jZ_j)\\ &= \prod_{j=1}^{d}\exp(-t_j^2/2)\\ &= \exp(-\|\boldsymbol{t}\|^2/2).\end{aligned}$$

Dabei ergibt sich der vorletzte Schritt aus der Tatsache, dass für eine standardnormalverteilte Zufallsvariable Z gilt:

$$m_Z(s) = \mathbb{E}\exp(sZ) \ = \ \exp(s^2/2) \quad \text{für beliebige } s \in \mathbb{R},$$

woraus die Formel

$$\mathbb{E}(Z^k) \ = \ \begin{cases} 0 & \text{für ungerades } k \\ \dfrac{(2m)!}{2^m m!} & \text{für } k = 2m, m \in \mathbb{N}_0 \end{cases}$$

resultiert; siehe Übungen. Zusammen mit Satz 5.9 (b) ergibt sich dann, dass

$$\begin{aligned}\mathbb{E}\exp(\boldsymbol{i}tZ) &= \sum_{k=0}^{\infty}\frac{(\boldsymbol{i}t)^k}{k!}\,\mathbb{E}(Z^k)\\ &= \sum_{m=0}^{\infty}\frac{(-t^2)^m}{(2m)!}\,\frac{(2m)!}{2^m m!}\\ &= \sum_{m=0}^{\infty}\frac{(-t^2/2)^m}{m!}\\ &= \exp(-t^2/2).\end{aligned}$$

Die charakteristische Funktion des Zufallsvektors $\boldsymbol{X} := \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z}$ ergibt sich aus

$$\begin{aligned}\varphi_{\boldsymbol{X}}(\boldsymbol{t}) &= \mathbb{E}\exp\bigl(\boldsymbol{i}\langle\boldsymbol{t},\boldsymbol{\mu}+\boldsymbol{A}\boldsymbol{Z}\rangle\bigr)\\ &= \exp\bigl(\boldsymbol{i}\langle\boldsymbol{t},\boldsymbol{\mu}\rangle\bigr)\,\mathbb{E}\exp\bigl(\boldsymbol{i}\langle\boldsymbol{A}^\top\boldsymbol{t},\boldsymbol{Z}\rangle\bigr)\\ &= \exp\bigl(\boldsymbol{i}\langle\boldsymbol{t},\boldsymbol{\mu}\rangle\bigr)\varphi_{\boldsymbol{Z}}(\boldsymbol{A}^\top\boldsymbol{t})\\ &= \exp\bigl(\boldsymbol{i}\langle\boldsymbol{t},\boldsymbol{\mu}\rangle\bigr)\exp\bigl(-\|\boldsymbol{A}^\top\boldsymbol{t}\|^2/2\bigr).\end{aligned}$$

5.2.2 Eineindeutigkeit und schwache Konvergenz

Eine essentielle Eigenschaft charakteristischer Funktionen ist, dass ein Wahrscheinlichkeitsmaß P auf $\mathbb{R}^d$ durch φ_P eindeutig festgelegt wird.

Satz 5.12 ***(Eineindeutigkeit von*** $P \mapsto \varphi_P$***).*** *Für Wahrscheinlichkeitsmaße* P, Q *und* P_n $(n \in \mathbb{N})$ *auf* $\mathbb{R}^d$ *gilt:*

(a) *$P = Q$ genau dann, wenn*

$$\varphi_P(\boldsymbol{t}) \ = \varphi_Q(\boldsymbol{t}) \quad \text{für alle } \boldsymbol{t} \in \mathbb{R}^d.$$

(b) *$(P_n)_n$ konvergiert schwach gegen P genau dann, wenn*

$$\lim_{n\to\infty} \varphi_{P_n}(\boldsymbol{t}) \ = \ \varphi_P(\boldsymbol{t}) \quad \text{für alle } \boldsymbol{t} \in \mathbb{R}^d.$$

Anmerkung 5.13 (Charakterisierung durch Projektionen). Satz 5.12 impliziert, dass die Verteilung eines Zufallsvektors $\boldsymbol{X} \in \mathbb{R}^d$ eindeutig charakterisiert wird durch die Verteilung der reellwertigen Zufallsvariablen $\langle \boldsymbol{v}, \boldsymbol{X}\rangle$, $\boldsymbol{v} \in \mathcal{V}$, wenn $\mathcal{V}$ eine dichte Teilmenge der Einheitssphäre $\{\boldsymbol{v} \in \mathbb{R}^d : \|\boldsymbol{v}\| = 1\}$ ist.

Denn Kenntnis der Verteilungen von $\langle \boldsymbol{v}, \boldsymbol{X}\rangle$, $\boldsymbol{v} \in \mathcal{V}$, bedeutet, dass wir

$$\varphi_{\langle \boldsymbol{v},\boldsymbol{X}\rangle}(r) \ = \ \mathbb{E}\exp\bigl(\boldsymbol{i}r\langle \boldsymbol{v}, \boldsymbol{X}\rangle\bigr) \ = \ \varphi_{\boldsymbol{X}}(r\boldsymbol{v})$$

für alle $r \in \mathbb{R}$ und $\boldsymbol{v} \in \mathcal{V}$ kennen. Doch die Menge $\bigl\{r\boldsymbol{v} : r \in \mathbb{R}, \boldsymbol{v} \in \mathcal{V}\bigr\}$ ist eine dichte Teilmenge des $\mathbb{R}^d$. Auf Grund der Stetigkeit von $\varphi_{\boldsymbol{X}}$ kennen wir also (im Prinzip) die gesamte Funktion $\varphi_{\boldsymbol{X}}$. Daher kennen wir auch die Verteilung von $\boldsymbol{X}$.

5.2.3 Beweis von Satz 5.12

Es gibt verschiedene Beweise. Die hier vorgestellten Argumente beinhalten auch eine explizite Methode, wie man P aus φ_P rekonstruieren kann. Ein wesentliches Hilfsmittel ist die Standardnormalverteilung auf dem $\mathbb{R}^d$ mit Dichtefunktion

$$\phi(\boldsymbol{x}) \ = \ (2\pi)^{-d/2}\exp(-\|\boldsymbol{x}\|^2/2).$$

Ist $\boldsymbol{Z}$ ein Zufallsvektor mit dieser Verteilung, dann wird für $\sigma > 0$ die Verteilung von $\sigma\boldsymbol{Z}$ durch die Dichtefunktion

$$\phi_\sigma(\boldsymbol{x}) \ := \ \sigma^{-d}\phi\bigl(\sigma^{-1}\boldsymbol{x}\bigr)$$

beschrieben.

Lemma 5.14 ***(Glättung von Verteilungen und Funktionen).*** *Seien $\boldsymbol{X}$ und $\boldsymbol{Z}$ stochastisch unabhängige Zufallsvektoren, wobei $\boldsymbol{X}$ nach P verteilt und $\boldsymbol{Z}$ standardnormalverteilt ist. Für $\sigma > 0$ und messbare, beschränkte (oder nichtnegative) Funktionen $h : \mathbb{R}^d \to \mathbb{R}$ ist dann*

$$\mathbb{E}\, h(\boldsymbol{X} + \sigma\boldsymbol{Z}) \ = \ \int_{\mathbb{R}^d} h(\boldsymbol{y})f_\sigma(\boldsymbol{y})\, d\boldsymbol{y} \ = \ \int_{\mathbb{R}^d} h_\sigma(\boldsymbol{x})\, P(d\boldsymbol{x}),$$

mit

$$\begin{aligned} f_\sigma(\boldsymbol{y}) \ &:= \ \mathbb{E}\,\phi_\sigma(\boldsymbol{y} - \boldsymbol{X}) \ = \ \int_{\mathbb{R}^d} \phi_\sigma(\boldsymbol{y} - \boldsymbol{x})\, P(d\boldsymbol{x}), \\ h_\sigma(\boldsymbol{x}) \ &:= \ \mathbb{E}\, h(\boldsymbol{x} + \sigma\boldsymbol{Z}) \ = \ \int_{\mathbb{R}^d} h(\boldsymbol{x} + \sigma\boldsymbol{z})\phi(\boldsymbol{z})\, d\boldsymbol{z}. \end{aligned}$$

Insbesondere ist f_σ die Dichtefunktion der Verteilung von $\boldsymbol{X} + \sigma\boldsymbol{Z}$.

Lemma 5.15 ***(Inversionsformel).*** *Die Dichtefunktion f_σ in Lemma 5.14 lässt sich schreiben als*

$$\begin{aligned} f_\sigma(\boldsymbol{y}) &= (2\pi)^{-d} \int_{\mathbb{R}^d} \exp(-\boldsymbol{i}\langle \boldsymbol{y}, \boldsymbol{t}\rangle) \varphi_{\boldsymbol{X}+\sigma\boldsymbol{Z}}(\boldsymbol{t})\, d\boldsymbol{t} \\ &= (2\pi)^{-d} \int_{\mathbb{R}^d} \exp(-\boldsymbol{i}\langle \boldsymbol{y}, \boldsymbol{t}\rangle) \exp\Big(-\frac{\sigma^2\|\boldsymbol{t}\|^2}{2}\Big) \varphi_P(\boldsymbol{t})\, d\boldsymbol{t}. \end{aligned}$$

Beweis von Lemma 5.14. Nach dem Satz von Fubini ist $\mathbb{E}\, h(\boldsymbol{X} + \sigma\boldsymbol{Z})$ gleich $\mathbb{E}\, h_\sigma(\boldsymbol{X})$ mit $h_\sigma(\boldsymbol{x}) = \mathbb{E}\, h(\boldsymbol{x} + \sigma\boldsymbol{Z}) = \int h(\boldsymbol{x} + \sigma\boldsymbol{z})\phi(\boldsymbol{z})\, d\boldsymbol{z}$. Andererseits ist

$$\begin{aligned} \mathbb{E}\, h(\boldsymbol{X} + \sigma\boldsymbol{Z}) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h(\boldsymbol{x}+\boldsymbol{z})\, \phi_\sigma(\boldsymbol{z})\, d\boldsymbol{z}\; P(d\boldsymbol{x}) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h(\boldsymbol{y})\, \phi_\sigma(\boldsymbol{y}-\boldsymbol{x})\, d\boldsymbol{y}\; P(d\boldsymbol{x}) \qquad (\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{z}) := \boldsymbol{x}+\boldsymbol{z}) \\ &= \int_{\mathbb{R}^d} h(\boldsymbol{y}) \int_{\mathbb{R}^d} \phi_\sigma(\boldsymbol{y}-\boldsymbol{x})\, P(d\boldsymbol{x})\, d\boldsymbol{y} \\ &= \int_{\mathbb{R}^d} h(\boldsymbol{y}) f_\sigma(\boldsymbol{y})\, d\boldsymbol{y} \end{aligned}$$

nach der Transformationsformel und dem Satz von Fubini. □

Beweis von Lemma 5.15. Auch hier kommt der Satz von Fubini mehrfach zum Einsatz, wobei sich alle auftretenden Integrale über den $\mathbb{R}^d$ erstrecken:

$$\begin{aligned} &(2\pi)^{-d} \int \exp(-\boldsymbol{i}\langle \boldsymbol{y}, \boldsymbol{t}\rangle) \exp\Big(-\frac{\sigma^2\|\boldsymbol{t}\|^2}{2}\Big) \varphi_P(\boldsymbol{t})\, d\boldsymbol{t} \\ &= (2\pi)^{-d/2} \int \exp(-\boldsymbol{i}\langle \boldsymbol{y}, \boldsymbol{t}\rangle) \phi(\sigma\boldsymbol{t}) \varphi_P(\boldsymbol{t})\, d\boldsymbol{t} \\ &= (2\pi)^{-d/2} \int \exp(-\boldsymbol{i}\langle \boldsymbol{y}, \boldsymbol{t}\rangle) \phi(\sigma\boldsymbol{t}) \int \exp(\boldsymbol{i}\langle \boldsymbol{t}, \boldsymbol{x}\rangle)\, P(d\boldsymbol{x})\, d\boldsymbol{t} \\ &= (2\pi)^{-d/2} \int\int \exp(\boldsymbol{i}\langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{t}\rangle) \phi(\sigma\boldsymbol{t})\, d\boldsymbol{t}\, P(d\boldsymbol{x}) \\ &= (2\pi)^{-d/2}\sigma^{-d} \int\int \exp\big(\boldsymbol{i}\langle \sigma^{-1}(\boldsymbol{x}-\boldsymbol{y}), \boldsymbol{z}\rangle\big) \phi(\boldsymbol{z})\, d\boldsymbol{z}\, P(d\boldsymbol{x}) \quad (\boldsymbol{z} := \sigma\boldsymbol{t}, d\boldsymbol{t} = \sigma^{-d} d\boldsymbol{z}) \\ &= (2\pi)^{-d/2}\sigma^{-d} \int \varphi_{\boldsymbol{Z}}(\sigma^{-1}(\boldsymbol{x}-\boldsymbol{y}))\, P(d\boldsymbol{x}) \\ &= (2\pi)^{-d/2}\sigma^{-d} \int \exp\big(-\|\sigma^{-1}(\boldsymbol{x}-\boldsymbol{y})\|^2/2\big)\, P(d\boldsymbol{x}) \\ &= \int \phi_\sigma(\boldsymbol{y}-\boldsymbol{x})\, P(d\boldsymbol{x}) \\ &= f_\sigma(\boldsymbol{y}). \qquad \square \end{aligned}$$

Beweis von Satz 5.12 (a). Aus $\varphi_P \equiv \varphi_Q$ folgt nach Lemma 5.15 und Lemma 5.14, dass

$$\int h_\sigma\, dP = \int h_\sigma\, dQ$$

für jede stetige Funktion $h : \mathbb{R}^d \to [0,1]$ und beliebige $\sigma > 0$. Doch nach dem Satz von der majorisierten Konvergenz ist einerseits

$$\lim_{\sigma \downarrow 0} h_\sigma(\boldsymbol{x}) \;=\; \lim_{\sigma \downarrow 0} \int h(\boldsymbol{x} + \sigma \boldsymbol{z}) \phi(\boldsymbol{z}) \, d\boldsymbol{z} \;=\; h(\boldsymbol{x}) \quad \text{für alle } \boldsymbol{x} \in \mathbb{R}^d$$

und andererseits

$$\lim_{\sigma \downarrow 0} \int h_\sigma \, dM \;=\; \int h \, dM$$

für $M = P, Q$. Also ist $\int h \, dP = \int h \, dQ$ für jede stetige Funktion $h : \mathbb{R}^d \to [0,1]$, was gleichbedeutend ist mit $P = Q$. □

Beweis von Satz 5.12 (b). Für festes $\sigma > 0$ betrachten wir die Dichtefunktionen f_σ und

$$f_{\sigma,n}(\boldsymbol{y}) \;:=\; \int \phi_\sigma(\boldsymbol{y} - \boldsymbol{x}) \, P_n(d\boldsymbol{x}).$$

Nach Lemma 5.14 handelt es sich bei f_σ und $f_{\sigma,n}$ um Wahrscheinlichkeitsdichten, und Lemma 5.15 impliziert, dass

$$\lim_{n \to \infty} f_{\sigma,n}(\boldsymbol{y}) \;=\; f_\sigma(\boldsymbol{y}) \quad \text{für alle } \boldsymbol{y} \in \mathbb{R}^d.$$

Nach dem Satz von Scheffé (einfacher Spezialfall von Satz 2.25) ist sogar

$$\lim_{n \to \infty} \int \bigl| f_{\sigma,n}(\boldsymbol{y}) - f_\sigma(\boldsymbol{y}) \bigr| \, d\boldsymbol{y} \;=\; 0.$$

Für eine Lipschitzstetige Funktion $h : \mathbb{R}^d \to [0,1]$ mit Lipschitzkonstante L ist

$$\begin{aligned} \bigl| h_\sigma(\boldsymbol{x}) - h(\boldsymbol{x}) \bigr| \;&=\; \bigl| \mathbb{E}\, h(\boldsymbol{x} + \sigma \boldsymbol{Z}) - h(\boldsymbol{x}) \bigr| \\ &\le\; L\sigma \, \mathbb{E}\, \|\boldsymbol{Z}\| \\ &\le\; L\sigma \sqrt{\mathbb{E}(\|\boldsymbol{Z}\|^2)} \\ &=\; L\sigma\sqrt{d}. \end{aligned}$$

Diese Überlegungen zeigen, dass

$$\begin{aligned} \limsup_{n \to \infty} \Bigl| \int h \, dP_n - \int h \, dP \Bigr| \;&\le\; 2L\sqrt{d}\,\sigma + \limsup_{n \to \infty} \Bigl| \int h_\sigma \, dP_n - \int h_\sigma \, dP \Bigr| \\ &=\; 2L\sqrt{d}\,\sigma + \limsup_{n \to \infty} \Bigl| \int h(\boldsymbol{y}) \bigl(f_{\sigma,n}(\boldsymbol{y}) - f_\sigma(\boldsymbol{y}) \bigr) \, d\boldsymbol{y} \Bigr| \\ &\le\; 2L\sqrt{d}\,\sigma + \limsup_{n \to \infty} \int \bigl| f_{\sigma,n}(\boldsymbol{y}) - f_\sigma(\boldsymbol{y}) \bigr| \, d\boldsymbol{y} \\ &=\; 2L\sqrt{d}\,\sigma. \end{aligned}$$

Für $\sigma \downarrow 0$ ergibt sich dann die Aussage, dass $\int h \, dP_n$ für $n \to \infty$ gegen $\int h \, dP$ konvergiert. Dies impliziert die schwache Konvergenz von $(P_n)_n$ gegen P. □